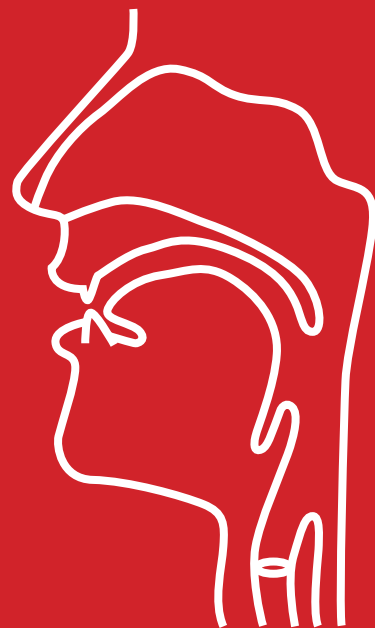


Propuestas en fonética experimental: enfoques metodológicos y nuevas tecnologías

**Editores: Beatriz Blecua, Jordi Cicres, Marina Espejel
y María J. Machuca.**



**Propuestas
en fonética experimental:
enfoques metodológicos
y nuevas tecnologías**

**Beatriz Blecua
Jordi Cicres
Marina Espejel
María J. Machuca
(editores)**



Aviso legal

Esta obra està sujeta a una licencia Reconocimiento 4.0 de Creative Commons. Está permitida la reproducción, distribución, comunicación pública y transformación para generar una obra derivada, sin restricción siempre que se cite al titular de los derechos (Universitat de Girona). La licencia completa se puede consultar en:
<https://creativecommons.org/licenses/by-sa/4.0/deed.es>

© de los textos: los autores

Edita: Universitat de Girona-Servei de Publicacions

ISBN: 978 84 8458 590 9

Girona, Marzo de 2022

Índice

PRESENTACIÓN	7
CATEGORIZACIÓN DE ERRORES PROSÓDICOS EN LA ENSEÑANZA DE LA PRONUNCIACIÓN L2 ASISTIDA POR ORDENADOR. Lourdes Aguilar, Eva Estebas-Vilaplana	11
LA IMPORTANCIA DE LA INTENSIDAD COMO PISTA ACÚSTICA EN LA IDENTIFICACIÓN DE ORACIONES INTERROGATIVAS. María Cristina Armero, Ignacio Moreno-Torres, Paolo Roseano.	16
INTRA- AND INTER-SPEAKER VARIATION OF MULTILINGUAL DATA: A FORENSIC PHONETICS APPROACH. José Vicente Benavent Cháfer	21
LA IMPORTANCIA DE LAS VOCALES EN LA ADQUISICIÓN Y LA ENSEÑANZA DEL ESPAÑOL HABLADO POR CHINOS. Yongfa Cao, Agnès Rius-Escudé	26
DOES METALINGUISTIC KNOWLEDGE ABOUT THE L2 ENHANCE THE EFFECTIVENESS OF L2 PERCEPTUAL TRAINING? Angélica Carlet, Juli Cebrian, Núria Gavaldà, Celia Gorba	31
LECTURA EN VEU ALTA, PAUSES SILENCIOSES I IOGA NIDRA. Josefina Carrera-Sabaté, Jesús Bach Marquès	36
LA ENTONACIÓN DEL ESPAÑOL DE LOS ÁNGELES. UNIDAD VS. DIVERSIDAD. Yolanda Congosto Martín	41
POSITIONALLY CONSTRAINED ACCENTEDNESS CUES. A PILOT STUDY ON BARI ITALIAN. Giovina Angela del Rosso	46
ALLOPHONES OF /tr/ AFTER CODA /s/ ASPIRATION IN WESTERN ANDALUSIAN SPANISH. María del Saz	51
SENTENCE ACCENT IN ENGLISH AS A FOREIGN LANGUAGE: EXCEPTIONS TO THE LAST LEXICAL ITEM RULE. María del Saz, Ximena Grau	56
ESTUDIO ACÚSTICO Y DIALECTOMÉTRICO DE LA DURACIÓN VOCÁLICA: RELACIÓN DE PROXIMIDAD DEL OCCIDENTE, CENTRO Y ORIENTE CUBANO. Chaxiraxi Díaz, Josefa Dorta, Carolina Jorge	61
LA PROSODIA DE ACTOS DEL HABLA EN EL ESPAÑOL DE ASUNCIÓN, BUENOS AIRES Y CIUDAD DE MÉXICO. Natalia dos Santos Figueiredo, Carolina Gomes da Silva	66
DETERMINING SPEAKERS' ORIGIN ON THE BASE OF THEIR INTONATION: A PRELIMINARY STUDY. Wendy Elvira-García, Paolo Roseano, Ana Ma. Fernández Planas	71
UMBRAL DE INTELIGIBILIDAD EN AUDIOMETRÍAS VERBALES: VALORES DE REFERENCIA EN NIÑOS NORMOYENTES DE 5, 6 Y 7 AÑOS. Wendy Elvira-García, Paolo Roseano, Ana Ma. Fernández Planas, Silvia Marro, Pilar Calvo, M. Antonia Claveria	76
APLICACIONES DEL ULTRASOUND SOUND IMAGE (UTI) A LOS ESTUDIOS DE FONÉTICA. Wendy Elvira-García, Paolo Roseano, Lourdes Romera Barrios, Victoria Marrero Aguiar	80

ChaLeT: DETERMINING THE CHANCE LEVEL IN PERCEPTION EXPERIMENTS BY MEANS OF BINOMIAL TESTS. Wendy Elvira-García, Paolo Roseano, Assumpció Rost Bagudanch	85
USO DE SISTEMAS DE CLASIFICACIÓN SUPERVISADA PARA LA EVALUACIÓN DE LA PROSODIA: APLICACIONES AL HABLA DE PERSONAS CON SÍNDROME DE DOWN. David Escudero, Mario Corrales-Astorgano, Yolanda Martín de San Pablo, Alfonso Rodríguez de Rojas, Valle Flores, César González-Ferreras, Valentín Cardeñoso, Lourdes Aguilar	90
UNA HERRAMIENTA DE FONÉTICA APLICADA: MATERIALES DE LOGOAUDIOMETRÍAS INFANTILES EN CATALÁN. Ana Ma. Fernández Planas, Paolo Roseano, Wendy Elvira-García, Silvia Marro, Pilar Calvo, M. Antonia Clavería	95
EVIDENCIA ACÚSTICA Y ESTADÍSTICA DE UNA CLASE NATURAL DE EYECTIVAS EN KAWÉSQAR. Mauricio A. Figueroa Candia, Daniela A. Mena Sanhueza	101
¿DÓNDE ESTÁ LA ENTONACIÓN? PROPUESTA DIDÁCTICA INTEGRADORA DE LA ENTONACIÓN EN EL NIVEL A1. Andrea García Coronado, Beatriz Bleuca	106
PROPUESTA METODOLÓGICA PARA LA EVALUACIÓN DE FACTORES PROSÓDICOS EN EXPOSICIONES ORALES DE ALUMNOS DE SECUNDARIA. Clara Luna García García de León, Juan María Garrido Almiñana	110
APROXIMACIÓN AL ANÁLISIS MELÓDICO DE LA ENTONACIÓN NEUTRA DEL ESPAÑOL HABLADO POR RUSOHABLANTES. Zoya Garmátina, Empar Devís	114
DISEÑO DE UNA BATERÍA DE PRUEBAS PARA LA EVALUACIÓN DE LA PROSODIA DEL ESPAÑOL EN EL MARCO DEL PROYECTO PROA. Juan María Garrido Almiñana, Eva Estebas-Vilaplana, María J. Machuca, Lorraine Baqué, Wendy Elvira-García, Victoria Escandell Vidal	118
INVESTIGATING THE ROLE OF COGNATE STATUS IN L2 PRODUCTION AND PERCEPTION. Celia Gorba, Susana Cortés, Juli Cebrian, Núria Gavaldà	122
PERCEPTION OF L1 SPEECH OF LATE CZECH-FRENCH BILINGUALS BY CZECH MONOLINGUALS. M. Hevrova, B. Köpke, T. Bořil	127
FRICATIVAS ALVEOLARES DEL EUSKERA: APROXIMACIÓN A SU CARACTERIZACIÓN ARTICULATORIA. Alexander Iribar, Rosa Miren Pagola, Itziar Túrrez	132
MODELOS MIXTOS Y CRITERIOS DE INFORMACIÓN EN FONÉTICA. Miguel Jiménez-Bravo	135
RELEVANCIA DE LA DURACIÓN EN EL ESPAÑOL DE CANARIAS: ETIQUETAJE Y DIALECTOMETRÍA A PARTIR DE DATOS ACÚSTICOS. Carolina Jorge Trujillo, Josefa Dorta Luis, Chaxiraxi Díaz Cabrera	140
DISTANCIA DEL COSENO ENTRE VOCALES CEPSTRALES COMO MÉTODO PARA CUANTIFICAR LA DIFERENCIA ACÚSTICA ENTRE MARCAS EN CONFLICTO. José María Lahoz-Bengochea, Marianela Fernández Trinidad	145
SPEECH PRODUCTION IN NOISE BY CHINESE LEARNERS OF ENGLISH. Bin Li, Qianwen Guan, Zixia Fan, Mosi He	150
LA PRODUCCIÓN DE LAS OCLUSIVAS CHINAS POR HISPANOABLANTES. Zhao Liu, María J. Machuca	154
LA FUNCIÓN DE LAS HESITACIONES EN LA IDENTIFICACIÓN DEL HABLANTE. Joaquim Llisterri, María J. Machuca, Antonio Ríos	160

LAS TÉCNICAS DE NEUROIMAGEN Y LA FONÉTICA EXPERIMENTAL. Victoria Marrero Aguiar	165
DESEMPEÑO FONÉTICO-ACÚSTICO DE OCLUSIVAS ÁFONAS EN HABLANTES ADULTOS MAYORES DEL ESPAÑOL CHILENO, DE LA CIUDAD DE CHILLÁN, CHILE. Renato Martínez-Cifuentes, Jaime Soto-Barba, Mauricio Figueroa-Candia, Katia Sáez-Carrillo	170
LEXICAL EFFECTS IN THE USE OF WORD-LEVEL STRESS FOR WORD RECOGNITION BY ENGLISH AND KOREAN LEARNERS OF SPANISH. María Teresa Martínez-García, Dong-Jin Shin, Goun Lee	174
UN CAMBIO ENTONATIVO EN EL HABLA DE LOS JÓVENES DE MAÓ Y ALAIOR (MENORCA). Ignasi Mascaró, Paolo Roseano	178
LA ENTONACIÓN DE LAS PREGUNTAS ABSOLUTAS DEL ESPAÑOL DE SAN JUAN (ARGENTINA) EN EL HABLA ESPONTÁNEA. Miguel Mateo Ruiz	182
UNA APROXIMACIÓN AL ANÁLISIS PROSÓDICO DE LA CONVERSACIÓN COLOQUIAL. Miguel Mateo Ruiz, Francisco José Cantero Serena	187
P-CENTRES AND NUCLEAR VOWEL DURATION IN THE CONTEXT OF IRISH (GAELIC) INITIAL MUTATION. Connor McCabe	192
“JUST COMPARE THAT!” THE USE OF INSTANT AUDIO MESSAGES IN FORENSIC PHONETICS. Chiara Meluzzi, Sonia Cenceschi, Nicholas Nese	197
THE ROLE OF ATTENTION CONTROL AND STIMULI LEXICAL STATUS IN HIGH VARIABILITY PHONETIC TRAINING. Joan C. Mora, Mireia Ortega, Íngrid Mora-Plaza, Jonás Fouz-González	201
SECUENCIAS VOCÁLICAS EN FRONTERA DE PALABRAS EN DATOS DEL ESPAÑOL MEXICANO. João Paulo Moraes Lima dos Santos	206
PROSODIA DEL ASTURIANO CENTROMERIDIONAL: RASGOS CARACTERÍSTICOS Y ANOTACIÓN EN ASTUR_ToBI. Carmen Muñiz Cachón, Paolo Roseano	210
VARIACIÓN SOCIAL EN LA REALIZACIÓN DE LA FRICATIVA /s/ EN EL ESPAÑOL DE ANTIOQUIA, COLOMBIA. Diana Muñoz-Builes, Wendy Elvira-García	215
ESTUDIO PRELIMINAR DE LA VOCAL ROTIZADA EN EL HABLA ESPONTÁNEA DE ANDALUCÍA Y ASTURIAS. Núria Ortiz-de-Pinedo Sánchez	220
L2 PHONETIC ACCURACY DEVELOPMENT IN A FOREIGNLANGUAGE SETTING: A LONGITUDINAL STUDY. Alejandra Pesantez, Volker Dellwo	225
PERCEPTUAL ASSIMILATION OF TASHLHIYT CONSONANTS BY SPANISH-DOMINANT BILINGUALS. Lucrecia Rallo Fabra, Mustapha Achichaou, Michael D. Tyler	229
FUNDAMENTACIÓN ARTICULATORIA Y AERODINÁMICA DE LOS PROCESOS FONOLÓGICOS. LAS ASIMILACIONES. Daniel Recasens	234
LOS RASGOS PROSÓDICOS DE LOS NEOLECTORES: UN ESTUDIO MULTIDIMENSIONAL CON UN ENUNCIADO INTERROGATIVO. Sara Recio-Pineda, Alicia Sola, Francisco José Cantero Serena	240
EMERGÈNCIA FONÈTICA: CAL APRENDRE I ENSENYAR A PRONUNCIAR LA [ʎ]. Agnès Rius-Escudé, Francina Torras	244
RHYTHMIC TRANSFER IN GALICIAN LEARNERS OF ENGLISH. Rosalía Rodríguez Vázquez, Paolo Roseano	249

TraFo: AN ONLINE G2P TOOL FOR SPANISH. Francesco Rodriquez	253
EL ESTUDIO DE LOS VOCATIVOS DEL ESPAÑOL EN EL MARCO DEL PROYECTO AMPER-CAT. Lourdes Romera Barrios, Paolo Roseano, Ana Ma. Fernández Planas, Francesco Rodriquez, Diana Muñoz Builes, Carolina Fabra Navarro, Eugenio Martínez Celdrán	257
THE LINGUISTIC RHYTHM OF FRIULAN: FIRST DATA. Paolo Roseano, Wendy Elvira-García, Francesco Rodriquez	262
A PHONETICALLY BASED VALIDATION OF THE CLASSIFICATION OF FRIULIAN DIALECTS. Paolo Roseano, Franco Finco	266
IS CATALAN SYLLABLE-TIMED? AN ANSWER BASED ON CLUSTER ANALYSIS. Paolo Roseano, Patrícia Marsà Morales, Laura Alañá Vilas	270
ACTOS DE HABLA INDIRECTOS Y SU INTERPRETACIÓN EN EL MARCO PROSÓDICO Y GESTUAL (EN LA BASE DE INTERACCIONES EN ESPAÑOL DE APRENDICES RUSOS). Albina Sarymsakova	275
WORD-FINAL [Z] IN ENGLISH PLURALIA-TANTUM AND REGULAR-PLURAL NOUNS: A COMPARISON OF DURATION. Marcel Schlechtweg, Greville G. Corbett	279
MELODÍAS ENFÁTICAS EN PORTUGUÉS DE BRASIL. Raquel Sena Mendes, Dolors Font-Rotchés	283
UN ESTUDIO PILOTO DE EJERCICIOS DE LA PRONUNCIACIÓN DE LAS PALABRAS ÁTONAS Y TÓNICAS CON AYUDA DE MÚSICA. Hirotaka Sensui, Takuya Kimura y Miyuki Takasawa	288
ANÁLISIS COMPARATIVO DEL F0 RANGO Y DE LA FLUIDEZ ORAL ENTRE LOS HABLANTES CHINOS Y NATIVOS DEL ESPAÑOL. Peizhu Shang	293
PROSODIC ORGANIZATION OF ENGLISH FOLK RIDDLES AND THEIR EMOTIONAL AND PRAGMATIC POTENTIALS. Larysa Taranenko	298
SENSIBILIDAD PROSÓDICA ACENTUAL DE NIÑOS CON TRASTORNO DEL DESARROLLO DEL LENGUAJE DE SEGUNDO GRADO DE PRIMARIA. Valeska Torres Bustos, Bernardo Riffo Ocares, Katia Sáez Carrillo	302
¿ES MI PREGUNTA UNA INTERROGATIVA EN ESPAÑOL? LA ENTONACIÓN DEL ESPAÑOL HABLADO POR POLACOS. Weronika Urbanik-Pęk	307
ADQUISICIÓN DE LA ENTONACIÓN DE ELE: LA PRODUCCIÓN DEL FOCO ESTRECHO POR LOS SINOHABLANTES. Weiqi Li	313
INTERROGATIVAS Y ENTONACIÓN EN EL ESPAÑOL HABLADO POR CHINOS. Tianshu Zhao, Dolors Font-Rotchés	318

PRESENTACIÓN

En este volumen se recogen los trabajos presentados en el VIII Congreso Internacional de Fonética Experimental (CIFE), que tuvo lugar en la Facultat de Lletres de la Universitat de Girona entre el 28 y el 30 de junio de 2021. En el Congreso participaron 127 fonetistas y se presentaron un total de 81 comunicaciones (además de las dos conferencias plenarias).

Las circunstancias excepcionales que se vivieron a partir del mes de marzo del año anterior (con confinamientos y restricciones en los viajes y los actos públicos) recomendaron hacer un congreso con un formato híbrido, que combinó asistentes presenciales (que viajaron a Girona y pudieron presentar en vivo sus trabajos) con asistentes virtuales (que siguieron el congreso mediante sistemas de videoconferencia y presentaron sus comunicaciones virtualmente). Aun así, el Congreso tuvo una alta participación.

El Congreso se estructuró en cinco sesiones monográficas: Herramientas aplicadas al estudio experimental del habla, organizada por la Dra. Sandra Schwab; Evaluación de la prosodia, organizada por el Dr. Juan María Garrido; Nuevas metodologías para el estudio fonético: qué nos aportan y qué nos exigen, organizada por la Dra. Victoria Marrero; Adquisición y enseñanza de lenguas, organizada por el Dr. Jordi Cicres;

y Estudios en fonética forense, organizada por la Dra. Núria Gavaldà. Estas sesiones monográficas se complementaron con paneles con comunicaciones de temáticas diversas. Además, se contó con una mesa redonda moderada por el Dr. José María Lahoz sobre "La fonética experimental actual: conocimiento experto o habilidades tecnológicas", con los Dres. Wendy Elvira, Joaquim Llisterrí y Jordi Cicres, como participantes. Hay que destacar también la participación del Dr. Daniel Recasens, que pronunció la ponencia de apertura ("Fundamentación articuladora y aerodinámica de los procesos fonológicos: las asimilaciones") y Volker Dellwo, que se encargó de la conferencia de clausura ("The dynamics of indexical information").

A pesar de esta organización temática del congreso, los editores hemos considerado pertinente presentar los textos de los trabajos por orden alfabético de sus primeros autores, puesto que muchos de los trabajos son interdisciplinarios.

Finalmente, los editores queremos dar las gracias a los participantes al Congreso por la calidad de sus trabajos y la buena predisposición que mostraron para que esta nueva edición del CIFE (en plena pandemia de la Covid-19 y con un formato diferente a las anteriores ediciones de CIFE) fuera un éxito. También queremos agradecer el trabajo de los miembros del comité colaborador, de los organizadores de las sesiones monográficas, de los conferenciantes de las sesiones plenarias, de los participantes en la mesa redonda, de los presidentes de las sesiones y de

los miembros del comité científico. Igualmente, agradecemos el apoyo económico y logístico de la Universitat de Girona (en especial, de la Facultat de Lletres, el Departament de Filologia i Comunicació y el Institut de Recerca Educativa).

CATEGORIZACIÓN DE ERRORES PROSÓDICOS EN LA ENSEÑANZA DE LA PRONUNCIACIÓN L2 ASISTIDA POR ORDENADOR

Lourdes Aguilar, Eva Estebas-Vilaplana

Departament de Filologia Espanyola, Universitat Autònoma de Barcelona
Departamento de Filologías Extranjeras y sus Lingüísticas, Universidad Nacional de Educación a Distancia
lourdes.aguilar@uab.cat, eestebas@flog.uned.es

ABSTRACT

El objetivo del estudio es avanzar en la definición del concepto de “buena formación prosódica” a partir de la identificación de las principales dificultades de categorización de los errores prosódicos en un sistema automático CAPT (*Computer Assisted Pronunciation Training*), con el fin último de mejorar la enseñanza y aprendizaje de la entonación en L2. Los datos proceden de la revisión manual de un corpus transcrito con segmentos de habla identificados como errores prosódicos y el correspondiente mensaje de retroalimentación. Los resultados de coherencia entre los transcripores demuestran la complejidad de la tarea de identificar errores en el componente prosódico: el valor de consistencia $Kappa=0.432$ señala un acuerdo moderado. El análisis de las divergencias a la luz del modelo de la fonología de la entonación permite identificar las principales fuentes de desacuerdo y proponer mejoras en el método de evaluación automática.

Palabras clave: evaluación prosódica, etiquetado ToBI, corrección prosódica, enseñanza de la pronunciación asistida por ordenador

The aim of this study is to define the concept of "prosodic well-formedness" by identifying the main difficulties of categorization of prosodic mistakes in an automatic CAPT (*Computer Assisted Pronunciation Training*) system, with the goal of improving the teaching and learning of L2 intonation. The data include the results of a manual revision by two experts in prosody of a corpus transcribed with the speech segments identified as prosodic errors and the corresponding feedback message. The consistency results between the two annotators demonstrate the complexity of error identification in prosody: the consistency value $Kappa = 0.432$ indicates a moderate agreement. The analysis of the divergences among the reviewers with respect to intonational phonology allows us to identify the main sources of disagreement and propose improvements for the automatic evaluation method.

Keywords: prosodic evaluation, ToBI labelling, prosodic well-formedness, computer assisted pronunciation training

1. INTRODUCCIÓN

Podemos convenir con el *Plan Curricular del Instituto Cervantes* (PCIC) que muchos fracasos comunicativos de los hablantes del español como lengua extranjera (en adelante, ELE) se deben a dificultades relacionadas con aspectos suprasegmentales. Sin embargo, a pesar de reconocer la primacía del plano suprasegmental sobre el segmental en la comunicación, el PCIC dedica un único epígrafe a la pronunciación y la prosodia (capítulo 3: subsistema de lo fonético y fonológico), centrandó la descripción en el estudio de la base articulatória general del español. Del mismo modo, el

tiempo dedicado al entrenamiento prosódico en los cursos de L2 es limitado, y a menudo inexistente. Como consecuencia, los alumnos, incluidos los que tienen un dominio más alto de la lengua, conservan la entonación propia de su L1 cuando hablan una L2 (cf. Chen et al., 2016, de-la-Mota, 2019, entre otros).

La falta de trabajo específico en la corrección prosódica puede atribuirse, además de a otros factores mencionados en Cortés Moreno (2000), a la indefinición del concepto de “buena formación prosódica”, dado el alto grado de variación que presentan las distintas formas entonativas, y la falta de correspondencia directa entre categorías prosódicas e implementación fonética en las distintas lenguas. Por un lado, la delimitación de “error”

prosódico en el aprendizaje de una L2 no es una tarea fácil. Por otro, lo que un estudiante incluye en su competencia fonológica a propósito de la prosodia varía de una lengua a otra. Dada esta complejidad, los intentos de evaluar de forma automática la prosodia en L2, en general, se abordan desde la señal acústica, usando rasgos de F0, energía o duración. Por el contrario, en Escudero *et al.* (2017) se propone incorporar el etiquetado simbólico, en particular, el sistema ToBI (*Tones and Break Indices*, Beckman *et al.*, 2005).

El presente estudio pretende avanzar en la definición del concepto de “buena formación prosódica” a partir de la identificación de las principales dificultades de categorización de los errores prosódicos en un sistema automático CAPT (*Computer Assisted Pronunciation Training*). Partiendo de la hipótesis de que es posible dar cuenta de los errores prosódicos con un etiquetado prosódico simbólico planteamos las siguientes preguntas de investigación: 1) ¿Cómo se incluye la variabilidad prosódica en el concepto de error prosódico? y 2) ¿Cómo se informa a los aprendices sobre los errores prosódicos que ha cometido?

2. PROCEDIMIENTO EXPERIMENTAL

2.1. Corpus

El corpus contiene quince oraciones del corpus SAMPLE (Escudero-Mancebo *et al.*, 2014b), leídas por catorce estudiantes de español L2. Para disponer de pronunciaciones de referencia de hablantes nativos de español, se usan los enunciados procedentes del subcorpus de noticias de Glissando leídos por ocho profesionales de la radio y la publicidad (Garrido *et al.*, 2013). Las grabaciones del corpus SAMPLE se realizaron en una sala de estudio silenciosa, utilizando una grabadora digital a una frecuencia de muestreo de 48 kHz y un micrófono de estudio profesional. Las condiciones de grabación del corpus Glissando se describen en Garrido *et al.* (2013).

2.2. Etiquetado prosódico e identificación de errores

Las frases del corpus se transcribieron ortográficamente, se alinearon con la señal y se segmentaron las palabras, sílabas y sonidos. Para el etiquetado prosódico, se usó el sistema descrito en González-Ferreras *et al.* (2012) y en Escudero-Mancebo *et al.* (2014a). El sistema se basa en las etiquetas Sp_ToBI propuestas en Estebas-Vilaplana y Prieto (2008, 2010), pero con una agrupación de aquellas clases que tienen una frecuencia de aparición baja y un grado de similitud alto. Una vez se dispone de las transcripciones prosódicas, se comparan las de los hablantes no nativos con las de los hablantes de

referencia con el objetivo de identificar aquellas partes de la cadena hablada que se han producido incorrectamente desde el punto de vista prosódico. Como criterio general, una palabra se pronuncia con una prosodia incorrecta cuando las etiquetas asociadas no se corresponden con ninguna de las halladas (o con un mínimo asignado previamente) en los enunciados de los hablantes nativos.

Aunque defendemos la idea de que el etiquetado simbólico con ToBI es eficaz para identificar errores prosódicos en ELE, descartamos su uso directo para incorporarlo en un sistema CAPT que informe a los estudiantes sobre sus limitaciones prosódicas. Como alternativa, el sistema usa los siguientes mensajes de retroalimentación:

(1) *Uncommon intonation* (UI). El tipo de acento tonal o tono de frontera asignado por el sistema de etiquetado automático a una palabra de un hablante no nativo rara vez fue utilizado por los hablantes nativos.

(2) *Unexpected boundary* (UB). El hablante no nativo sitúa un tono de frontera donde ninguno de los hablantes nativos lo ha producido.

(3) *Over accentuation* (OA). El hablante no nativo ha acentuado una palabra que carece de acento tonal.

(4) *Lack of boundary* (LB). En la producción del hablante no nativo se omite una frontera en un lugar donde aparece frecuentemente asignada en el conjunto de referencia (hablantes nativos).

(5) *Lack of accent* (LA). El hablante no nativo ha pronunciado como átona una palabra comúnmente acentuada por los hablantes nativos.

2.3. Revisión manual del corpus de errores prosódicos

Dos expertas en prosodia (R1, R2) revisaron de forma independiente los resultados de la evaluación automática del conjunto de oraciones de ocho de los locutores (en total, 3535 palabras en 60 oraciones). Las revisoras usaron el programa Praat, con el que disponían del audio de la oración, el espectrograma y la curva melódica, y un TextGrid con la información de los errores prosódicos asignada por el sistema automático. En un TextGrid independiente las revisoras podían borrar, añadir o cambiar el mensaje de retroalimentación.

3. RESULTADOS

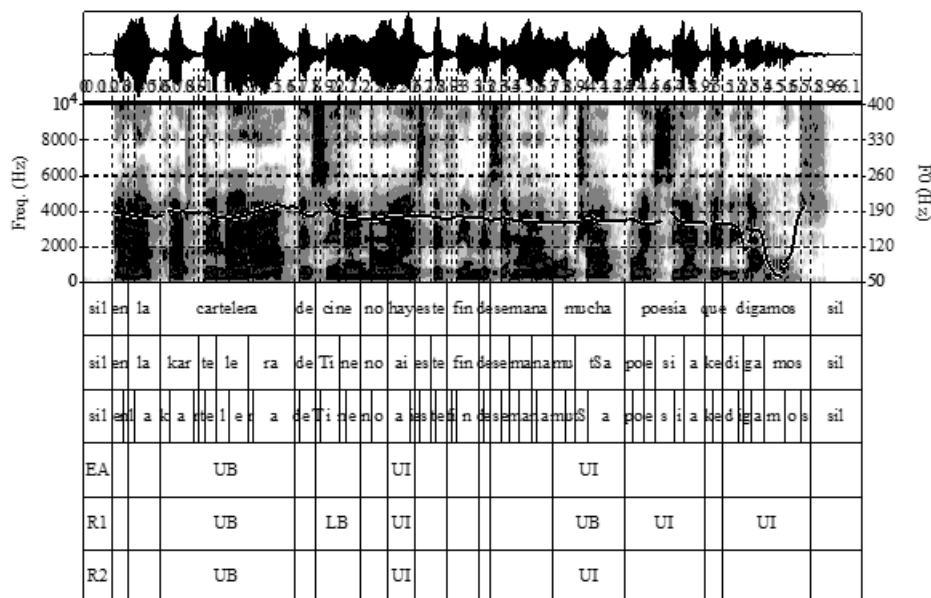


Figura 1: Análisis de la frase “en la cartelera de cine no hay este fin de semana mucha poesía que digamos” pronunciada por un hablante de inglés americano en la que pueden compararse las transcripciones de errores prosódicos del sistema automático (EA) con las de las revisoras (R1 y R2). UB=Unexpected Boundary, UI=Uncommon Intonation y LB=Lack of Boundary.

Los resultados de coherencia de la revisión manual confirman la complejidad de la tarea de identificar errores prosódicos, con un grado moderado de acuerdo entre las dos evaluadoras (Kappa=0.432). La Tabla 1 permite visualizar el número de transcripciones diferentes. Hay un alto grado de coincidencia en las categorías que informan sobre la ausencia de acento tonal (LA) y de tono de frontera (LB), mientras que la entonación poco común (UI), la aparición imprevista de una frontera prosódica (UB) y la sobreacentuación (OA) son los fenómenos con un mayor número de desacuerdos (88, 17 y 17 sobre 122, 55 y 29, respectivamente).

El parámetro UI es el más complejo porque abarca varias posibilidades de error, desde un evento tonal que nunca se usaría en un contexto determinado hasta combinaciones tonales inusuales pero aceptables. En general, R2 ha interpretado UI como una producción claramente errónea, mientras que R1 ha incluido también casos de entonación poco común. Además, la valoración UI no siempre corresponde a un mismo índice prosódico. A modo de ejemplo, en la Figura 1 la etiqueta UI de R1 en “poesía” indica un contorno melódico inesperado, ya que el hablante produce un tono suspensivo frente a un tono descendente, mientras que la misma etiqueta UI en “digamos” corresponde a la ausencia de desacentuación tonal. Como la desacentuación no se

contemplaba en la lista de errores, R1 lo etiquetó como UI y R2 no lo indicó como error.

Otra de las fuentes de desacuerdo está relacionada con la presencia/ausencia de fronteras prosódicas. Según vemos en la Figura 1, ambas evaluadoras coinciden en que hay una UB en “cartelera”, pero R1 añade también una LB en “cine”, donde se esperaría un tono ascendente indicando el final del grupo entonativo. De nuevo, aunque esto sea lo esperado, no implica que la ausencia de frontera prosódica sea algo erróneo.

Tabla 1. Tabla de contingencia de las etiquetas asignadas por los evaluadores humanos. En cada celda, las columnas presentan los datos de R1 y las filas, los datos de R2. NM=No Mistake; UI=Uncommon intonation; UB=Unexpected boundary; OA=Over accentuation LB=Lack of boundary; LA=Lack of accent.

R1/ R2	NM	UI	UB	OA	LB	LA
NM	723	20	11	36	1	1
UI	79	34	4	4	1	0
UB	10	5	38	2	0	0
OA	13	2	1	12	1	0
LB	12	0	0	0	5	0
LA	2	0	0	0	0	2

Finalmente, un gran número de transcripciones diferentes se explica según consideremos los alargamientos como parte de la estructura prosódica o del componente segmental. En la Figura 1, la palabra “muchu” se produce con un alargamiento en la sílaba postónica. Mientras que el transcriptor automático (EA) y R2 lo interpretan como UI, R1 lo valora como UB con final suspensivo, ya que las sílabas anteriores a una frontera prosódica suelen ser más largas (Cruttenden, 1997). El alargamiento y sus implicaciones en la modelización prosódica es un tema de debate en la actualidad, ya que se ha observado cómo distintos dialectos del español no difieren tanto en el movimiento melódico final como en la duración de la última sílaba acentuada (Estebas-Vilaplana, 2010). Posiblemente, se debería incorporar el criterio “duración” entre posibles errores y desvincularlo de otras categorías como UI o UB. Lo mismo se aplica a la categoría OA, que puede interpretarse como diferencias de duración o tonales.

4. DISCUSIÓN DE LOS RESULTADOS Y CONCLUSIONES

Uno de los principales desafíos en la evaluación de la competencia prosódica de los hablantes L2 es la identificación de los errores prosódicos, debido a una falta de definición del concepto de “buena formación” o corrección prosódica. Para captar toda su complejidad, necesitamos unidades descriptivas que nos permitan un análisis sistemático, y la fonología de la entonación, con el sistema de etiquetado ToBI, las proporciona al explicar los fenómenos suprasegmentales según tres categorías entonativas: acento (*accent*), organización prosódica (*phrasing*) y entonación (*intonation*), y separar la estructura métrica de la entonativa (Ladd, 1996). Aunque los resultados de coherencia en la revisión manual del corpus señalan una falta de coherencia importante entre los evaluadores manuales, el análisis de las divergencias a la luz de la fonología de la entonación permite identificar las principales fuentes de desacuerdo y proponer mejoras en el método de evaluación automática.

Así, se ha podido dar respuesta a las preguntas planteadas al inicio de la investigación:

(1) *¿Cómo se incluye la variabilidad prosódica en el concepto de error prosódico?* En el modelo AM, para dar cuenta de las categorías entonativas, se proponen como unidades básicas la manera en que se realiza la prominencia tonal (acento melódico o *pitch accent*) y la manera en que termina un determinado grupo prosódico (tono de frontera o *boundary tone*). Desde este punto de vista, los patrones entonativos están formados por secuencias de uno o varios

acentos tonales más los tonos de frontera. Lo que permite explicar la variedad de formas entonativas de una misma función prosódica es el nexo que se establece entre acentuación (estructura métrica) y entonación (estructura prosódica). Así es el propio modelo adoptado el que ofrece herramientas para abordar la variación.

Asimismo, el sistema automático, al usar un conjunto de referencia de varios hablantes nativos, tiene en cuenta la variación interindividual. El sistema compara las etiquetas de la transcripción prosódica de la misma oración pronunciada por los hablantes nativos con las etiquetas de la transcripción prosódica de dicha oración pronunciada por el hablante no nativo. Se considera error prosódico cuando las etiquetas asociadas al hablante no nativo no se corresponden con ninguna de las asignadas (o con un mínimo requerido) en los enunciados de referencia por los hablantes nativos.

(2) *¿Cómo se informa a los aprendices de las categorías de errores prosódicos?* Si bien el etiquetado simbólico, de orientación fonológica, es eficaz para identificar los principales errores prosódicos, la interpretación de las etiquetas ToBI requiere de una formación en fonética y fonología de la entonación que queda lejos de los requisitos de un estudiante ELE. El estudiante necesita que se le informe no solo sobre sus errores, sino sobre qué debe hacer para mejorar. Por ello, en el sistema automático se opta por transformar el contenido de las etiquetas ToBI en mensajes de retroalimentación. Ahora bien, en la revisión manual se pone de manifiesto que los mensajes de retroalimentación han simplificado el contenido de las categorías entonativas. Esta simplificación, por un lado, afecta a la coherencia de los evaluadores, dado que induce a juicios que pueden estar basados en los parámetros de tono, energía o duración; y por otro lado, a los estudiantes, dado que la retroalimentación debe orientar al estudiante para la acción comunicándole qué debe hacer para mejorar, lo que no se cumple con mensajes de contenido genérico como los usados en el sistema CAPT. También se hace patente la necesidad de aclarar el límite entre lo erróneo y lo inesperado para conseguir una mayor coincidencia de juicios entre los evaluadores y definir de forma más precisa la corrección prosódica.

5. BIBLIOGRAFÍA

- Beckman, M., Hirschberg, J., y Shattuck-Hufnagel, S., (2005). The original ToBI system and the evolution of the ToBI framework. En Jun, S.-A. (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford University Press, New York, pp. 9–54.
- Chen, N. F., Wee, D., Tong, R., Ma, B., y Li, H. (2016). Large-scale characterization of non-native Mandarin

- Chinese spoken by speakers of European origin: Analysis on iCALL. *Speech Communication*, 84, 46-56.
- Cortés Moreno, M. C. (2000). Sobre la adquisición de la prosodia en lengua extranjera: Estado de la cuestión. *Didáctica (Lengua y Literatura)*, 2000(12), 91-119.
- Cruttenden, A. (1997). *Intonation*. Cambridge: CUP.
- de-la-Mota, C. (2019). Improving non-native pronunciation: Teaching prosody to learners of Spanish as a second/foreign language. En *Key Issues in the Teaching of Spanish Pronunciation* (pp. 163-198). Routledge.
- Escudero-Mancebo, D. González-Ferreras, C., Aguilar, L. y Estebas-Vilaplana, E. (2017). Automatic assessment of non-native prosody by measuring distances on prosodic label sequences. En *Proceedings of Interspeech*, p. 1442-1447. ISCA, Agosto 2017.
- Escudero-Mancebo, D., González-Ferreras, C., Vivaracho-Pascual, C., y Cardeñoso-Payo, V. (2014a). A fuzzy classifier to deal with similarity between labels on automatic prosodic labeling. *Computer Speech & Language*, 28(1), 326-341.
- Escudero-Mancebo, David César González-Ferreras, and Valentín Cardeñoso Payo (2014b). Assessment of non-native spoken spanish using quantitative scores and perceptual evaluation. En Calzolari, N. et al., (eds). *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, p. 3967-3972, Reykjavik.
- Estebas-Vilaplana, E. (2010). The role of duration in intonational modelling: a comparative study of Peninsular and Argentinean Spanish. *RESLA*, 23: 153-174.
- Estebas-Vilaplana, E. y Prieto, P. (2008). La notación prosódica del español: una revisión del Sp-ToBI. *Estudios de fonética experimental*, 17, 264-283.
- Estebas-Vilaplana, E. y Prieto, P. (2010). Castilian Spanish intonation. Prieto, P. y Roseano, P. (eds.). *Transcription of intonation of the Spanish language*, München: Lincom Europa. 17-48.
- Garrido, J. M., Escudero, D., Aguilar, L., Cardeñoso, V., Rodero, E., De-La-Mota, C., y Laplaza, Y. (2013). Glissando: a corpus for multidisciplinary prosodic studies in Spanish and Catalan. *Language Resources and Evaluation*, 47(4), 945-971.
- González-Ferreras, C., Escudero-Mancebo, D., Vivaracho-Pascual, C., y Cardeñoso-Payo, V. (2012). Improving automatic classification of prosodic events by pairwise coupling. *IEEE transactions on audio, speech, and language processing*, 20(7), 2045-2058.
- Ladd, D. R. (2008). *Intonational phonology*. Cambridge University Press.

LA IMPORTANCIA DE LA INTENSIDAD COMO PISTA ACÚSTICA EN LA IDENTIFICACIÓN DE ORACIONES INTERROGATIVAS

María Cristina Armero¹, Ignacio Moreno-Torres¹, Paolo Roseano^{2,3}

¹Universidad de Málaga, ²Universitat de Barcelona, ³University of South Africa
mararmper@uma.es, imoreno@uma.es, paolo.roseano@ub.edu

ABSTRACT

La evidencia en varias lenguas sugiere que los usuarios de implante coclear no usan la F0 como una señal acústica para la identificación de las características prosódicas y, por lo tanto, confían en otras señales acústicas (por ejemplo, intensidad, duración de las vocales). Con el objetivo de determinar en qué medida estas señales secundarias proporcionan información confiable sobre el contraste entre oraciones declarativas e interrogativas, se diseñó un experimento de percepción en el que los oyentes tenían que clasificar un conjunto de oraciones naturales y manipuladas (por ejemplo, para simular el implante coclear) como aseveraciones o preguntas. Los resultados indican que la intensidad puede ayudar a los pacientes con implantes cocleares a diferenciar entre afirmaciones e interrogativas totales.

Palabras clave: fonética clínica, intensidad, prosodia, entonación, implantes cocleares

Evidence from various languages suggests that cochlear implant users fail to recognize the F0 as an acoustic cue for the recognition of prosodic characteristics, and therefore rely on other acoustic cues (e.g. intensity, vowel duration). With the aim of determining to what extent these secondary cues provide reliable information about the declarative / interrogative contrast, a perception experiment was designed in which listeners had to classify a set of natural and manipulated (e.g. to simulate cochlear implant) sentences as statements or questions. The results indicate that intensity can help patients with cochlear implants to differentiate between statements and yes-no questions.

Keywords: clinical phonetics, intensity, prosodic, intonation, cochlear implant

1. INTRODUCCIÓN

Los hablantes suelen asociar diversas pistas acústicas a un mismo fenómeno fonológico (Lisker, 1986; Giezen, Escuero y Baker, 2010). Las interrogativas absolutas en español se caracterizan por un patrón entonacional con una subida de la frecuencia fundamental (F0) al final del enunciado (Navarro Tomás, 1944: 135; Quilis, 1993: 414), mientras que en las declarativas se produce una bajada. De acuerdo con Dorta, Hernández y Cabrera (2009), la F0 es la pista acústica fundamental y, por esa razón, también es la más estudiada. Por otra parte, la duración e la intensidad serían pistas secundarias (Candía, Urrutia y Fernández, 2006). La hipótesis de este trabajo es, justamente, que existen otras pistas acústicas, independientes de la F0, que permiten diferenciar desde el punto de vista

prosódico las oraciones interrogativas totales de las declarativas.

Conocer la aportación de las diferentes pistas acústicas a la hora de vehicular información lingüística tiene interés no solo desde el punto de vista descriptivo, sino desde un punto de vista aplicado. En concreto, es sabido que las personas con implante coclear tienen dificultades para acceder a la F0 (Zeng, Tang y Lu, 2014), por lo que sería importante saber si hay otra información alternativa a la que pueden acceder. Este trabajo tiene el objetivo concreto de estudiar el papel de la intensidad como pistas acústica complementaria en la percepción de la modalidad oracional por parte de los normo-oyentes en condiciones de degradación espectral, para en un segundo momento extrapolar las conclusiones y aplicarlas a la población con implante coclear.

El estudio de las pistas secundarias de la modalidad oracional, sin embargo, es complejo por varios motivos. En primer lugar, en la medida en que la F0 es una pista acústica necesaria (Candía, Urrutia y Fernández, 2006) y presente en todas las emisiones con voz modal, que son las más habituales en el habla. Para poder garantizar que los oyentes no la están utilizando y que solo están usando las pistas secundarias, es necesario manipular la señal acústica mediante un proceso de vocodizado (v. apartado 2.3) que permite eliminar la F0 aun conservando otras propiedades espectrales y temporales de la señal acústica (DiNino, Wright, Winn y Bierer (2016). La vocodización permite, además, simular el proceso de percepción en implantados cocleares.

En segundo lugar, puede que la presencia de estas pistas secundarias sea variable entre locutores, y que igualmente varíe en función de factores como la variedad dialectal, contexto, etc. Es más, es posible que un mismo locutor las emplee de forma variable.

El presente estudio, de carácter preliminar, pretende comprobar si un grupo de hablantes efectivamente emplea la intensidad para diferenciar enunciados declarativos e interrogativos. Además, en el caso de que la empleen, nos preguntamos si la utilizan todos ellos.

Para ello se ha grabado a un grupo de sujetos de la localidad andaluza de Álora (apartado 2.1), y se han obtenido señales vocodizadas (apartado 2.2) que se han utilizado para llevar a cabo un experimento de percepción (apartado 2.3). Los resultados del experimento (detallados en el apartado 2.4) permiten comprobar el papel que desempeñan las pistas acústicas secundarias en la percepción de la modalidad oracional.

2. METODOLOGÍA

2.1. Locutores y grabaciones

Para la grabación de las frases que se han utilizado como estímulos en el experimento de percepción, se contó con la participación de 12 sujetos (4 niños, 4 hombres y 4 mujeres). Todos los sujetos han crecido y viven en Álora, una localidad de la comarca del Guadalhorce perteneciente a la provincia de Málaga. Para diferenciar los locutores les hemos asignado un número identificativo (101,102, 103, etc.).

Los estímulos se obtuvieron pidiendo a los locutores que leyeran el enunciado escogido en las diferentes modalidades de la forma más natural posible.

El dispositivo empleado para la grabación fue una grabadora ZOOM H4n Pro conectada mediante conector XLR a un micrófono dinámico Shure WH20XLR. Se siguieron las recomendaciones de grabación aportadas por el producto.

El enunciado de las frases que se han grabado es, tanto para las interrogativas como para las declarativas, "Marina limpiaba la nevera". El criterio seguido para la elección de esta oración fue el de que existiese el mayor número posible de segmentos sonoros. Utilizando este enunciado se grabó a los 12 sujetos produciendo en dos ocasiones el enunciado, tanto en modalidad declarativa como interrogativa, y se escogió la que resultaba más natural.

Los locutores fueron grabados a campo libre en una habitación sin interferencias acústicas del exterior, en mono, con una frecuencia de muestreo de 44.100 Hz, que posteriormente se redimensionó utilizando Praat (Boersma y Weenink, 2014) a 22.050 Hz y una resolución de 16 bits. Los archivos se generaron en formato WAV.

2.2. Estímulos vocodizados (VV)

Siguiendo DiNino et al. (2016), el proceso de codificación de voz se diseñó para simular el procesamiento del implante IC Fidelity 120 con las mismas asignaciones de banda de frecuencia que las utilizadas en muchos dispositivos Advanced Bionics (Advanced Bionics Corp., Valencia, CA). Se obtuvieron un total de 15 bandas de frecuencia contiguas pseudo-logarítmicamente espaciadas de 250 a 8700 Hz de cada estímulo original. La raíz cuadrada de la energía total en cada canal se obtuvo para calcular la envolvente y la señal resultante se filtró bajo los 68 Hz. La envolvente de cada canal se utilizó para modular una banda de ruido con una frecuencia central igual a la del canal correspondiente. Las pendientes de salida del filtro se ajustaron a 30 dB/octava.

2.3. Prueba de percepción

La prueba de percepción se administró mediante un script MFC de Praat. El script presentaba a los jueces las grabaciones (tanto interrogativas como declarativas) en orden aleatorio, por duplicado y con la condición de que no se reprodujesen dos grabaciones iguales seguidas. Para cada enunciado, los jueces debían hacer clic en la pantalla en la casilla de la modalidad oracional correspondiente (declarativa o interrogativa). Los jueces tenían la opción de hacer una pausa cada 16 ítems.

Con el fin de comprobar la estabilidad de los resultados, se dividió la tarea en dos pruebas (prueba 1 y prueba 2), donde cada prueba tenía 48 audios (6 locutores x 2 modalidades x 2 voces x 2 repeticiones).

Un total de 70 alumnos de primer curso de Grado de Logopedia de la Universidad de Málaga participaron en el estudio, donde 7 de ellos eran hombres y 63 mujeres. Todos los participantes tenían el español como lengua materna y residían en Andalucía. La realización del experimento formó parte de una práctica académica. Al Grupo 1 de alumnos (N = 26) se les pasó la prueba 1, al Grupo 2 (N = 24) la prueba 2 y al Grupo 3 (N = 20) se le pasaron las dos pruebas. En total, se recogieron 8640 respuestas.

2.4. Medidas acústicas y análisis de correlaciones

Usando Praat, de cada grabación se tomaron las siguientes medidas acústicas relacionadas con la intensidad: intensidad máxima de la antepenúltima, penúltima y última sílaba. A partir de estas medidas se obtendrían otras relativas como el incremento de intensidad entre la sílaba tónica y la sílaba postónica (ΔI).

Los datos de las respuestas de los jueces se pasaron a formato SPSS (versión 25). Con SPSS se calculó el índice de correlación de Spearman R entre el porcentaje de aciertos en estímulos vocodizados (VV) en interrogativas y las medidas acústicas señaladas anteriormente.

3. RESULTADOS

3.1. Resultados de la prueba de percepción con VM

Antes de calcular las medidas indicativas de posibles pistas acústicas se comprobó si el porcentaje de aciertos en VM era superior al 95% (con objeto de descartar posibles deficiencias auditivas, atencionales, etc.). Ello llevó a eliminar las respuestas de 8 jueces. La muestra final constó de 7600 respuestas válidas de 62 oyentes (7 hombres y 55 mujeres).

El porcentaje de aciertos fue muy similar en hombres (79,2%) y mujeres (80%), lo que indica que la variable sexo del oyente no era determinante en este estudio.

Centrándonos en la voz vocodizada, el porcentaje total de aciertos medio en declarativas fue del 88,9% y en interrogativas del 66,7%, tendencia que se observó por igual en los tres grupos. Ello muestra que los jueces no respondieron al azar, e indica que

la prueba puede servir para analizar las pistas acústicas empleadas por los jueces.

3.2. Resultados de la prueba de percepción con VV

En voz vocodizada, en 11 de los 12 locutores el porcentaje de aciertos en las declarativas era próximo al 100% y superior al porcentaje en las interrogativas. En uno de los locutores (con el código 203 en la figura 1), sin embargo, el porcentaje de aciertos fue superior en las interrogativas.

El porcentaje de aciertos en las interrogativas se localizaba, en la mayoría de los casos, entre el 30% y 70%, y existían 5 casos con valores superiores a 50%. El hecho de que haya un nivel de aciertos tan alto en las interrogativas en voz vocodizada (o sea, sin F0), indica claramente que los oyentes deben haber accedido a otras pistas acústicas (v. apartado 3.3).

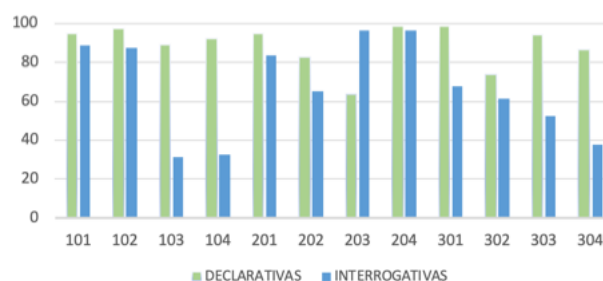


Figura 1: Porcentaje de aciertos por locutor en voz vocodizada en declarativas e interrogativas.

3.3. Correlación entre porcentaje de aciertos y medidas acústicas

Se calculó el índice de correlación de Spearman R, entre la variable “porcentaje de aciertos en VV en interrogativas” y las medidas acústicas indicadas en el apartado 2.4. Se observó una correlación negativa con la intensidad media de la penúltima sílaba del enunciado ($\rho = -0,60$; S.: $< 0,05$), y una correlación positiva con la intensidad media de la última sílaba del enunciado ($\rho = 0,64$; S.: $< 0,05$). Además, se observó una correlación positiva muy significativa con el incremento de intensidad entre las dos sílabas finales del enunciado ($\rho = 0,78$; S.: $< 0,01$). Esto quiere decir que aquellos locutores que resultaron más fáciles de reconocer redujeron la intensidad de la penúltima sílaba e incrementaron la intensidad de la última sílaba.

Con el fin de comprobar esta correlación observada entre ΔI y ratio de aciertos, se examinaron las curvas de intensidad de las interrogativas de los locutores con mayor y menor porcentaje de aciertos en VV. Los resultados mostraron que en los primeros hay un pico de intensidad en la postónica (101 y 204 en Fig. 3) que no se aprecia en las curvas de intensidad de los segundos (103 y 304 en Fig. 3).

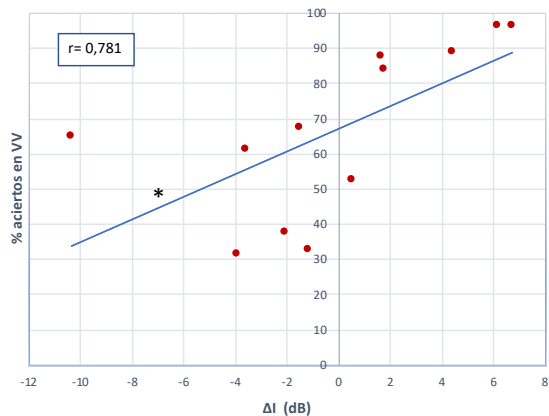


Figura 2: Correlación entre ΔI (entre la sílaba tónica y postónica) en las interrogativas y el porcentaje de aciertos en VV.

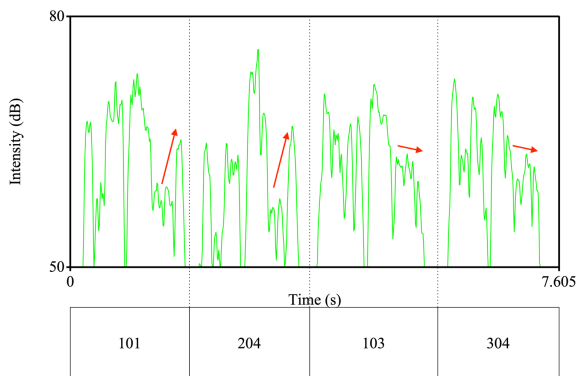


Figura 3: Los cambios de intensidad en las sílabas tónica y postónica de las interrogativas de los locutores con mayor y peor porcentaje de aciertos en VV.

4. DISCUSIÓN Y CONCLUSIONES

Son tres los resultados más destacables de este estudio. En primer lugar, y con la excepción del locutor 203, las interrogativas vocodizadas son más difíciles de reconocer que las declarativas vocodizadas.

En segundo lugar, el hecho de que los oyentes logren reconocer la modalidad en la mayoría de los casos (y con porcentajes cercanos al 100% en algunos) indica que debe haber pistas acústicas alternativas a la curva melódica. En análisis de resultados nos ha permitido observar que cuando se

produce un incremento final de intensidad entre la tónica y la postónica, los oyentes identifican la oración como interrogativa, lo que implicaría que el ΔI entre estas sílabas podría ser una pista acústica secundaria. Este resultado es compatible con otros trabajos (Dorta et al., 2009) que han observado cómo la intensidad puede tener una función prosódica relevante en español, por ejemplo, para marcar posiciones de frontera de la estructura sintagmática de interrogativas absolutas en Canarias.

En tercer lugar, este estudio muestra que existen diferencias individuales entre locutores. Todos los locutores usan la subida final de F0 para denotar una interrogativa, pero esto no ocurre con el ΔI : no todos los sujetos usan ΔI como pista acústica secundaria para denotar una interrogativa, lo que indica que es una pista variable. Este resultado parece compatible con el hecho de que las pistas secundarias no se observan de forma sistemática y pueden variar en función de factores de diversa índole (contextuales, individuales, etc.) (Candía, Urrutia y Fernández, 2006).

En general, los datos obtenidos en este estudio confirman la hipótesis sobre la existencia de una pista, diferente de la F0, que emplean algunos sujetos para marcar la modalidad interrogativa. Tal y como se ha visto, la pista en cuestión es el incremento de intensidad en la última sílaba del enunciado.

Este estudio, por lo tanto, puede ser el comienzo de nuevas líneas de investigación. Desde un punto de vista lingüístico se podría profundizar en la forma en la que varía ΔI entre dialectos, entre locutores, o incluso, en la variabilidad del uso que hace de este incremento un mismo locutor. Esta línea de investigación puede tener implicaciones también en otros ámbitos, como el de la logopedia, puesto que existe un grupo de hablantes, los usuarios de implante coclear, para los que –debido a la dificultad de acceder a la F0 (Driscoll, Gfeller, Kliethermes y Oleson, 2013) las pistas acústicas secundarias podrían ser esenciales. Futuros trabajos, vueltos a profundizar en las aplicaciones prácticas de los resultados obtenidos en población normo-oyente, se tendrán que plantear la pregunta de si y cómo es posible enseñar a utilizar el incremento de intensidad para vehicular la modalidad oracional interrogativa en el habla dirigida a personas con implante coclear.

5. REFERENCIAS

- Candía, L., Urrutia, H. y Fernández, T. (2006) Rasgos acústicos de la prosodia acentual del español. *Boletín de Filología*, 41, 11-14.
- DiNino, M., Wright, R. A., Winn, M. B. y Bierer, J. A. (2016). Vowel and consonant confusions from spectrally-manipulated stimuli designed to simulate poor cochlear implant electrode-neuron interfaces. *Acoustical Society of America*, 140(6), 4404–4418.
- Dorta, J., Hernández, B. y Díaz, C., (2009). Interrogativas absolutas: relación entre F0, duración e intensidad. *Estudios de Fonética Experimental*, 18, 123-144.
- Lisker, L. (1986). “Voicing” in English: A catalogue of acoustic features for Signaling /b/ Versus /p/ in Trochees. *Language and Speech*, 29(1), 1-13.
- Navarro Tomás, T. (1944): *Manual de entonación española*. Nueva York: Hispanic Institute in the United States.
- Quilis, A. (1981): *Manual de fonética acústica española*. Madrid: Editorial Gredos.
- Quilis, A y Fernández, J.A. (1989). *Curso de fonética y fonología españolas*. Madrid: Consejo Superior de Investigaciones Científicas.
- See, R.L., Driscoll, V.D., Gfeller, K., Kliethermes, S. y Oleson, J. (2013). Speech intonation and melodic contour recognition in children with cochlear implants and with normal hearing. *Otol Neurotol* 34(3), 490-498. doi: 10.1097/MAO.0b013e318287c985.
- Zeng, F., Tang, Q. y Lu, T. (2014). Abnormal Pitch Perception Produced by Cochlear Implant Stimulation. *PLoS ONE*, 9(2), p.e88662. doi:10.1371/journal.pone.0088662

INTRA- AND INTER-SPEAKER VARIATION OF MULTILINGUAL DATA: A FORENSIC PHONETICS APPROACH

José Vicente Benavent Cháfer

Universitat de Barcelona (UB)
jbenchbe7@alumnes.ub.edu

ABSTRACT

Previous findings on forensic phonetics suggest employing acoustic parameters with low intra-speaker variation and high inter-speaker variation for an effective forensic voice comparison. Unlike traditional research in this area of expertise, the present paper purposely includes adverse conditions (recordings of short duration, uncontrolled segmental phenomena) originating from semi-spontaneous interviews in three distinct languages (English, Spanish, and Dutch) to perform an acoustic-phonetic analysis with a more realistic view. Global pitch, global sound intensity, and pausing-related measures are registered within the suprasegmental domain for the purposes of this study, whilst segmental features are concerned with voiced [b, d, g] and voiceless [k, p, t] plosives, on the one hand, and the voiceless/voiced alveolar sibilants [s, z] on the other. Results revealed that the variables of choice were robust enough in inter-speaker variation, whereas similarities were wrongly matched in terms of intra-speaker variation, thus exhibiting error rates (false alarms) between 0% and 16.21%.

Keywords: forensic phonetics, forensic voice comparison, acoustic-phonetic analysis, inter-speaker variation, intra-speaker variation

1. INTRODUCTION

Forensic voice comparison tasks are typically undertaken in Spain by highly specialised research groups such as those belonging to the *Guardia Civil/ Policía Científica*, private companies, and/or academic institutions like the Polytechnic University of Madrid through the development and usage of biometric-based speaker recognition software like ASIS, SIBMATI, BATVOX, BS3, and FASR, among others (Jiménez et al. 2014: 37). Biometric features refer to the co-occurring physiological and behavioural cues involved in speech production, which appear to be quite idiosyncratic (Farrús 2011: 42), thus enhancing the chances of success. Despite its increasing use, biometric data may not be accessible to the average phonetician.

Far from the aforementioned (semi)automated methods employed in forensic phonetics research, this study builds up a fictitious voice line-up for the sake of testing inter-speaker and intra-speaker variation by conducting an acoustic-phonetic analysis on three distinguishable data sets (English, Spanish, and Dutch corpora). In this vein, the former examines the speech samples referring to both the distractors and the suspects in disguise, whereas the latter deals with the appointed suspects' excerpts.

The acoustic variables considered in the present paper take into account the proven discriminatory potential suggested by previous research, both at the segmental (Gordon et al. 2002, Koenig et al. 2013, Whiteside et al. 2004) and suprasegmental (Künzel 1997, Rose 2006) level. Besides being based on such empiric foundations, the selected acoustic units of measurement display high inter-speaker variation and low intra-speaker variation (Nolan 1983). Admittedly, previous studies reflect a high degree of control upon recorded speech, usually involving a rehearsal or reading of pre-defined passages, which is intended to control the number of segmental units produced by the intended speaker (Köster & Schiller 1997). In this study, however, segmental phenomena remain uncontrolled, which renders a more realistic object of study.

2. METHODOLOGY

2.1. Informants

As previously stated, several corpora have been selected depending on the informants' first language: English, Spanish, and Dutch input. For the first group of speakers, the British Library Sound Archive (2016) registers the experiences of wildlife sound recordists through semi-directed interviews taking place

between 2013 and 2016. Spanish informants were extracted from the ESLORA corpus, which consists of a series of semi-spontaneous interviews recorded in Galicia between 2007 and 2015 (Barcala et al. 2018). Lastly, the IFADV corpus (Van Son et al. 2009) is a free dialog video corpus of conversational speech in Dutch. As noted above, the exchanges were recorded in similar time periods, and thus linguistic input should expect little diachronic variation.

From all the available informants of said corpora, 7 speakers were chosen according to the criteria put forward by the existing guidelines on voice line-ups (Broeders & van Amelsvoort 2001, De Jong-Lendle et al. 2015, and Hollien 2012). As a result, every constituent of the line-up ought to resemble as closely as possible in terms of age and sociolinguistic background, and thus the following groups were created: English group, 7 males aged 47-64 (*mean age* = 54.14, *sd* = 6.4), Spanish speakers, 7 females aged 20-32 (*mean age* = 25.86, *sd* = 4.8), and Dutch informants, 7 females aged 20-28 years old (*mean age* = 21.00, *sd* = 3.3).

2.2. Material

As a reflection of the actual exchanges occurring in real-life scenarios where a voice parade would be needed, this paper decides to analyse excerpts of short duration (4-14 sec.). Even though segmental phenomena remain uncontrolled in said audio files, a suprasegmental criterion is formulated for the sake of challenging the acoustic-phonetic analysis: the excerpts obtained from distractors (aimed at testing inter-speaker variation) display similar intonation contours (rising intonation), whereas the intended suspects' speech includes sentences with rising and falling intonation patterns to gauge levels of intra-speaker variation. Additionally, only 1 excerpt is examined for each informant, with the exception of extracting 2 samples from the speaker appointed as the suspect, due to the already mentioned criterion. In total, 24 excerpts were annotated and subsequently analysed.

2.3. Procedure

The software Camtasia Studio 8.1.2 (TechSmith 2013) was employed to crop the recordings to comply with the criterion on audio duration mentioned above (4-14 sec.). Besides that, the encoding settings of every excerpt were set at 44.100kHz, Mono, 128kBits/sec, since the material was obtained from three different sources, and therefore the resulting files differed slightly in this respect. Apart from the aforementioned criteria (short duration, and intonation contours), excerpts were also taken from

the first sections of said semi-spontaneous interviews, which customarily covered similar topics (early-career experiences, childhood memories, etc.).

2.4. Acoustic-phonetic analysis

The software Praat 6.0.25 (Boersma & Weenink 2019) is used to carry out an acoustic-phonetic segmentation at the word and phonetic level with SAMPA codes. Regarding the extraction of segmental features, plosives [b, d, g] and [k, p, t]'s *VOT* and *release burst intensities* are calculated manually. Fricatives [s] and [z] (the latter being added only to English recordings to compensate the absence of [s] realisations in some samples), on the other hand, tap into the use of Praat scripts to extract the values of each variable: *COG*, *noise duration*, and *noise amplitude* are calculated with the script made by Elvira-García (2014). *F1-F3* values are obtained through Kawahara's (2010) script. Despite sounding counterintuitive due to the inharmonicity of fricative consonants (Martínez 2007: 69), formants are hereby conceived as a concentration of energy in the spectrogram, much in line with Univaso et al.'s (2014) definition. Lastly, the variable *spectral peak location* is estimated with a 25ms Hamming window alongside LPC and FFT spectral slices, according to the procedure specified by Jongman et al. (2000: 1255).

In the suprasegmental department, Lennes' (2013) script is run to calculate the measures of global pitch (Hz), namely $P\bar{x}$ (mean pitch), 25%, 50%, and 75% *pitch* quantiles. Said script's F0 threshold has been re-adjusted depending on the speaker's gender, according to Boersma's (2019) suggestions. Consequently, the range of values for males is adjusted to 75-300 Hz, whereas the threshold is increased in female speakers to 100-500 Hz. Concerning global sound intensity (dB), the variables $I\uparrow$ (max. intensity), $I\downarrow$ (min. intensity), and $I\bar{x}$ (mean intensity) are extracted with built-in Praat commands. Thirdly, De Jong & Wempe's (2008) script registers the numerical values of the following pausing-related measurements: *DurPause* (duration of pauses), *N_pause/min.* (n° of pauses per minute), *Pause_%* (percentage of pauses per excerpt), *N_pause* (number of pauses per excerpt), *speech rate*, *articulation rate*, and *ASD* (Average Syllable Duration).

2.5. Statistical analysis

This study employs IBM SPSS statistics 25 (IBM Corp. 2017) to calculate Euclidean distances of the suprasegmental variables of choice through a dissimilarity matrix with standardised z-scores, which are converted to p-values with the significance

level set at 0.05. This first analytical step assumes a single point of data per variable. In contrast, segmental variables register several realisations across the studied audio files. Once the distribution of the data is determined, the most suitable statistical measures are applied to test acoustic parameters against the dependent variable *speaker*: Kruskal-Wallis tests for non-normal distributions, and ANOVA tests for normally distributed data. The application of ANOVAs would be switched to Welch's tests in cases where the assumption of homogeneity of variances is violated, which is verified beforehand with a Levene test. Likewise, said statistical tests are conducted to measure significant deviation between voiced and voiceless consonants (plosives and fricatives) so as to examine them separately or together as a group in each segmental variable employed.

3. RESULTS AND DISCUSSION

Results on inter-speaker variation have proved that every variable considered in this study has yielded at least 1 statistically significant pairwise dissimilarity within each group of informants (at the 0.05 level). Nevertheless, two criteria have been formulated for the purpose of refining the findings even further: 1) variables which report the highest number of dissimilarities and 2) variables that find at least 1 case of pairwise dissimilarities for each constituent of the line-up. The list of variables that comply with such criteria goes as follows:

- English group: $I\uparrow$, $I\downarrow$, *DurPause*, *N_pause/min.*, *Pause_%*, *speech rate*, *articulation rate*, *ASD*, [k, p, t] *release burst intensity*, [s, z] *noise amplitude*, and [s] *F1* are grouped within the first criterion. The second criterion includes $I\downarrow$, *DurPause*, *Pause_%*, *N_pause*, *articulation rate*, *ASD*, and [s] *F1*.
- Spanish group: Criterion n°1 comprises $I\downarrow$, *speech rate*, *N_pause*, [b, d, g] & [k, p, t] *release burst intensity*, and [s] *noise amplitude*. On the other hand, criterion n°2 considers the variables $P\bar{x}$, *25%-75% pitch*, *speech rate*, *N_pause*, and [s] *noise amplitude*.
- Dutch group: $P\bar{x}$, $I\uparrow$, $I\bar{x}$, *DurPause*, *Pause_%*, [b, d, g] & [k, p, t] *release burst intensity*, and [s] *noise amplitude* are found within criterion n° 1. Additionally, the variables $I\uparrow$, *DurPause*, *Pause_%*, *N_pause*, *speech rate*, *articulation rate*, *ASD*, and [b, d, g] & [k, p, t] *release burst intensity* match the conditions for the second criterion.

When adopting said criteria as a standardisation procedure, it seems that the suprasegmental variable related to the number of pauses per extract (*N_pause*)

complies with the second criterion throughout every group of informants. In the segmental domain, *release burst intensity* and *noise amplitude* seem to fulfil the conditions for criterion n° 1, irrespective of the group of speakers examined. A notable distinction in this respect is that the English group considers only voiceless plosives [k, p, t] as efficient factors in the inter-speaker variation of *release burst intensity*, and thus voiced plosives [b, d, g] are deemed statistically different from their voiceless counterparts, as a Kruskal-Wallis test confirms ($H= 15.657$, $p= 0.008$, $df= 5$).

As far as intra-speaker variation is concerned, feasible variables refer to those which did not find statistically significant differences between voice samples belonging to the same speaker, albeit with differing intonation patterns. Such variables can be consulted in the following list:

- English group: $P\bar{x}$, *25%-75% pitch*, $I\uparrow$, $I\downarrow$, $I\bar{x}$, *DurPause*, *N_pause/min.*, *Pause_%*, *speech rate*, *articulation rate*, *ASD*, [b, d, g] & [k, p, t] *VOT*, *release burst intensity*, [s, z] *noise duration*, *noise amplitude*, *F1* and *F2*.
- Spanish group: $P\bar{x}$, *25% & 50% pitch*, $I\uparrow$, $I\downarrow$, $I\bar{x}$, *N_pause/min.*, *N_pause*, [b, d, g] & [k, p, t] *VOT*, [s] *spectral peak location*, *COG*, *noise duration*, *noise amplitude*, and *F1-F3*.
- Dutch group: $P\bar{x}$, *25%-75% pitch*, *DurPause*, *N_pause/min.*, *Pause_%*, *speech rate*, *articulation rate*, *ASD*, [b, d, g] & [k, p, t] *VOT*, *release burst intensity*, [s] *spectral peak location*, *COG*, *noise amplitude*, and *F1-F3*.

With the purpose of standardising the results, the variables appearing in every single language group are classified as *robust* acoustic measures in intra-speaker variation: $P\bar{x}$, *25% pitch*, *50% pitch*, *N_pause/min.*, *VOT* in plosives, and sibilants' *noise amplitude*, *F1-F2* values. It could be argued that *noise amplitude* matches the ideal conditions for forensic phonetics research, since it shows low intra-speaker variation and high inter-speaker variation in every language group studied in the present paper.

The results exhibited above refer to an experiment on intra-speaker variation consisting only of the excerpts made by the suspects alongside their recordings as distractors. However, when adding the rest of distractors of the voice line-up to the equation, error rates can be calculated as a result. Statistical tests have been re-run to examine again the similarities/dissimilarities of speakers' suprasegmental and segmental features in relation to the suspects' voice samples. Percentages of error rates/false alarms (dissimilarities found between recordings belonging to the same speaker) are illustrated in figure 1 below:

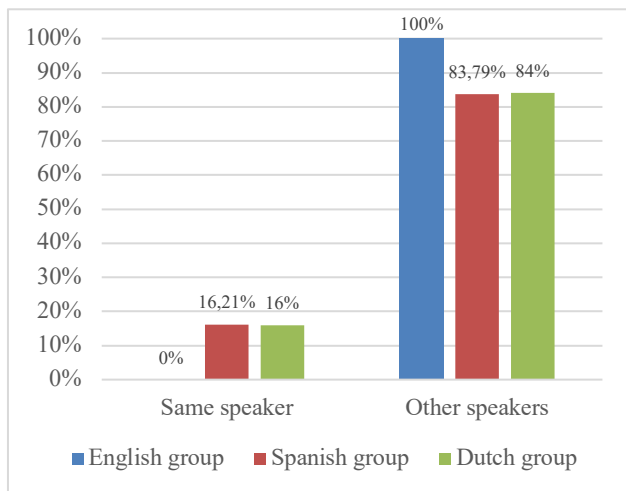


Figure 1: Percentages on the number of dissimilarities found between the same speaker (false alarm) and other speakers across English, Spanish, and Dutch recordings.

As inferred from figure 1, no dissimilarities were found between the recordings with falling and rising intonation of the English suspect when adding the rest of speakers to the voice comparison procedure ($n=0$, 0%). Conversely, error rates appear higher in the case of Spanish ($n=6$, 16.21%) and Dutch ($n=4$, 16%) recordings under these conditions.

It must be stressed that findings should be interpreted with caution, given the limited amount of material being analysed. Despite the short length of the voice samples examined, all variables found at least 1 pairwise difference among the 7 speakers of each line-up, and the rate for false alarms ranges from 0% to 16.21%. Two main observations may derive from such findings: 1) some acoustic variables within this study maintained their robustness despite the inconveniences of the data set, and 2) the addition of distractors to the voice comparison procedure may have improved the results, even for intra-speaker variation (the English group displays 0% error rates, whereas a few variables did find differentiations between the same speaker's samples when the rest of distractors were left out of the equation). In light of the foregoing, building dialectal databases to be used as reference corpora could entail greater success in forensic voice comparisons (Fernández Planas 1998: 165), as higher numbers of comparisons may lead to a more accurate verdict.

4. CONCLUSIONS

It could be concluded that, even under limited conditions, acoustic-phonetic analysis is still a feasible option when performing forensic voice comparisons, especially for phoneticians that may lack access to more sophisticated analytical means

such as biometrics-based software. In spite of its downsides, this paper offered a view on speech leaning towards real-life circumstances and its analysis, which is far from the traditional controlled laboratory setting. Much in line with previous research, suprasegmental features related to global pitch/intensity and pausing-related measurements are efficient parameters in forensic phonetics, even with a multilingual data set. The same is true for segmental phenomena concerned with voiceless/voiced plosives and sibilants, most notably sibilants' *noise amplitude* due to its performance both in inter-speaker and intra-speaker variation across the three language groups. Further research on speech databases is encouraged to unveil their potential as referential material, which could ultimately aid forensic voice comparison procedures and other linguistics-related investigations alike.

5. BIBLIOGRAPHY

- Barcala, M., Domínguez, E., Fernández, A., Rivas, R., Santalla, M., Vázquez, V., Villapol, R. (2018). El corpus ESLORA de español oral. Diseño, desarrollo y explotación. *CHIMERA: Romance Corpora and Linguistic Studies*, 5(2), 217-237.
- Boersma, P. (2019). *Intro 4.2. Configuring the pitch contour*. Retrieved from: https://www.fon.hum.uva.nl/praat/manual/Intro_4_2__Configuring_the_pitch_contour.html [11/08/2018].
- Boersma, P., & Weenink, D. (2019). Praat (v. 6.0.25) [Computer Program]. Retrieved from: www.praat.org [02/03/2017].
- British Library Sound Archive (2016). *Interviews with wildlife sound recordists*. Retrieved from: <https://sounds.bl.uk/Environment/Interviews-with-wildlife-sound-recordists-> [07/01/2017].
- Broeders, A., & van Amelsvoort, A. (2001). A practical approach to forensic earwitness identification: Constructing a voice line-up. *Problems of Forensic Sciences*, 47(1), 237-245.
- De Jong, N., & Wempe, T. (2008). Syllable Nuclei (v.2) [Praat script]. Retrieved from: <https://sites.google.com/site/spechrates/Home/praat-script-syllable-nuclei-v2> [04/05/2018]. Modified by Quené, H., Persoon, I., & De Jong, N., 2010.
- De Jong-Lendle, G., Nolan, F., McDougall, K., & Hudson, T. (2015). *Voice lineups: A practical guide*. Paper presented at the 18th International Congress of Phonetic Sciences. Glasgow: IPA, 1-5.
- Elvira-García, W. (2014). Zero crossing and spectral moments (v. 1.3) [Praat script]. Retrieved from: <http://stel3.ub.edu/labfon/sites/default/files/zero-crossing-and-spectral-moments13.praat> [10/06/2019].
- Farrús, M. (2011). La prosòdia com a identificador biomètric. *Llengua, Societat i Comunicació: Revista de Sociolingüística de la Universitat de Barcelona*, 9(1), 42-48.
- Fernández Planas, A. (1998). Fonètica forense. Lanàlisi pericial de la veu com una aplicació de la fonètica. In:

- Pradilla, M. (Ed.) *El món dels sons*. Benicarló: Alambor, 153-166.
- Gordon, M., Barthmaier, P., & Sands, K. (2002). A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association*, 32(2), 141-174.
- Hollien, H. (2012). On earwitness lineups. *Investigative Sciences Journal*, 4(1), 1-17.
- IBM Corp. (2017). IBM SPSS Statistics for Windows (v. 25.0) [Computer Program]. Retrieved from: <https://www.ibm.com/support/pages/downloading-ibm-spss-statistics-25> [04/02/2017].
- Jiménez, M., Reigosa, M., & Garayzábal, E. (2014). La lingüística forense: Licencia para investigar la lengua. In: Garayzábal, E., Jiménez, M., & Reigosa, M. (Eds.) *Lingüística forense: La lingüística en el ámbito legal y policial* (2ª ed.). Madrid: Euphonía Ediciones, 27-48.
- Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America*, 108(1), 1252-1263.
- Kawahara, S. (2010). Get F1, F2, F3 (averages) [Praat script]. Retrieved from: http://user.keio.ac.jp/~kawahara/scripts/get_formants.praat [21/07/2019].
- Koenig, L., Shadle, C., Preston, J., & Mooshammer, C. (2013). Toward improved spectral measures of /s/: Results from adolescents. *J Speech Lang Hear Res.*, 56(4), 1175-1189.
- Köster, O., & Schiller, N. (1997). Different influences of the native language of a listener on speaker recognition. *Forensic Linguistics*, 4(1), 18-28.
- Künzel, H. (1997). Some general phonetic and forensic aspects of speaking tempo. *Forensic Linguistics*, 4(1), 48-83.
- Lennes, M. (2013). Draw_pitch_histogram_from_sound [Praat script]. Retrieved from: https://github.com/FieldDB/Praat-Scripts/blob/master/draw_pitch_histogram_from_sound.praat [16/06/2018].
- Martínez, E. (2007). *Análisis espectrográfico de los sonidos del habla* (2ª ed.). Barcelona: Ariel.
- Nolan, F. (1983). *The phonetic bases of speaker recognition*. Cambridge: Cambridge University Press.
- Rose, P. (2006). Technical forensic speaker recognition: Evaluation, types and testing of evidence. *Computer Speech and Language*, 20(1), 159-191.
- TechSmith (2013). Camtasia Studio (v. 8.1.2) [Computer Program]. Retrieved from: <https://www.techsmith.com> [01/01/2017].
- Univaso, P., Martínez, M., & Gurlekian, J. (2014). Variabilidad intra- e inter-hablante de la fricativa sibilante /s/ en el español de Argentina. *Estudios de Fonética Experimental*, 23(1), 96-124.
- Van Son, R., Wesseling, W., Sanders, E., Van Den Heuvel, H. (2009). Promoting free dialog video corpora: the IFADV corpus example. In: Heylen, D., Kipp, M., Martin, J., & Paggio, P. (Eds.) *Multimodal corpora: From models of natural interaction to systems and applications*. Berlin: Springer-Verlag, 18-37.
- Whiteside, S., Henry, L., & Dobbin, R. (2004). Sex differences in voice onset time: A developmental study of phonetic context effects in British English. *The Journal of the Acoustical Society of America*, 116(2), 1179-1183.

LA IMPORTANCIA DE LAS VOCALES EN LA ADQUISICIÓN Y LA ENSEÑANZA DEL ESPAÑOL HABLADO POR CHINOS

Yongfa Cao, Agnès Rius-Escudé

Universitat de Barcelona
ycaocaox8@alumnes.ub.edu, agnesrius@ub.edu

ABSTRACT

En esta comunicación presentamos los resultados obtenidos en una investigación basada en la caracterización acústica de las vocales de la interlengua español hablado por chinos, con el objetivo de realizar una descripción acústica de cada vocal. El corpus es de habla semiespontánea y consta de 725 sonidos vocálicos producidos por 21 estudiantes chinos. Se ha comprobado que hay diferencias significativas entre las vocales producidas por hombres y mujeres; y que no las hay entre las vocales tónicas y las átonas, ni en el F_1 ni en el F_2 , a excepción del F_2 de la vocal anterior alta, [i]. Además se ha realizado una comparación con las vocales de habla de laboratorio, espontánea y la interlengua. A partir de las conclusiones obtenidas, se ha planteado algunas actividades didácticas en la enseñanza-aprendizaje del vocalismo del español para aprendices chinos.

Palabras claves: vocales españolas, fonética acústica, hablante chino, fonética aplicada

In this communication, we present the results of the research on the acoustic characterization of the vowels when Chinese speak Spanish, as interlanguage. The research aims at making an acoustic description of each vowel. We have worked with a semi-spontaneous speak corpus, which contains 725 vowel sounds produced by 21 Chinese students. It has been proven that there are differences between the vowels produced by men and wowed; and there are no significant difference between the stressed and unstressed vowels, neither in the F_1 nor in the F_2 , with the exception of the F_2 of the high anterior vowel , [i]. In addition, we have compared with those of Spanish in laboratory-speaking, spontaneous speaking and interlanguage. Based on the conclusions obtained, we have proposed some possible phonetic didactics that are useful for Chinese Spanish teachers and Chinese learners.

Keywords: Spanish vowels, acoustic phonetics, Chinese speaker, applied phonetics

1. INTRODUCCIÓN

Sabemos que China juega un papel importante en la Globalización Mundial y es el segundo socio comercial del mundo. China mantiene relaciones económicas y comerciales con 16 países hispanohablantes. Un escenario así propicia el aumento de español. A principios de 2018 el Gobierno chino estableció un nuevo diseño curricular de la enseñanza secundaria. El español es desde entonces, por impulso gubernamental, la nueva gran opción como lengua extranjera para los estudiantes y entra en el sistema público de educación junto al francés y el alemán. Así pues, se ha convertido en una de las seis lenguas extranjeras que se evalúan en el examen de ingreso universitario

(Ministerio de Educación, 2017: 16; y Ministerio de Educación, 2018: 10-55). El currículum actual de la enseñanza del español como Lengua Extranjera en China considera la pronunciación como una parte necesaria de la instrucción en el aula y uno de los objetivos más importantes para la formación de docentes profesionales (Grupo de Español del Comité de Didáctica de Lenguas Extranjeras en Escuelas Superiores Chinas, 1998). En este contexto, conocer los rasgos de la pronunciación del español e investigar sobre la interlengua español hablado por chinos es cada vez más necesario.

La pronunciación, como materialización del habla, no puede tener un papel secundario de subdestreza para la articulación del discurso, sino que debe considerarse fundamental para la negociación de los

interlocutores (Gil Fernández, 2007). Desde un punto de vista sociolingüístico, para los aprendices de una lengua extranjera tener una correcta pronunciación no sólo causa una buena impresión cuando se comunican con los nativos, sino que es motivo de autosatisfacción porque potencia su confianza (Laroya, 1995:3). Particularmente, es importante la pronunciación de las vocales, que son el núcleo de una palabra, porque es donde predomina la resonancia musical (Quilis, 1999: 141-142). Gil Fernández (2007:436) ya constató que el análisis acústico resulta menos simplificador y más fiable que el articulatorio, en cualquier caso y desde el punto de vista de la didáctica del español como segunda lengua, es imprescindible señalar algunas de las propiedades de carácter acústico, así como las propiedades preceptivas o auditivas de ellas derivadas, pues que el profesor de español ha de tener en cuenta tales rasgos a la hora de diseñar los ejercicios correctivos.

Aunque se han logrado grandes progresos en el campo de la fonética (Quilis y Esgueva, 1983; Bradlow, 1995; Martínez Celdrán, 1995; Albalá et al., 2008; Alfonso, 2014.), son escasos los estudios en relación al sistema vocálico de la interlengua español hablado por aprendices chinos, debido a la tardía introducción de este idioma en China. Entre los más recientes, hay el trabajo realizado por Jiménez & Tang (2018), que han estudiado la producción del sistema vocálico del español por hablantes chinos, basándose en un corpus de seis mujeres chinas.

La realización de este estudio permitirá acercarnos a las características acústicas de las vocales producidas por varones y mujeres chinos, y aportará los valores del primer formante (F_1) y segundo formante (F_2) de cada vocal, que son indispensables para la descripción de cada vocal y que son su característica vocálica (Quilis, 1981; Martínez Celdrán, 1989).

Según Delattre (1948 y 1951), el F_1 depende directamente del grado de abertura de la cavidad oral y del grado de constricción de la faringe, cuando más alta es la frecuencia de F_1 , más grande es la abertura total de la cavidad; mientras que el F_2 depende del grado de constricción linguovelar y de la labialización, cuando hay más constricción linguovelar y más labialización, más bajo es el formante y a la inversa, cuando hay menos constricción y menos labialización, más alto. De este modo se podrá conocer y comprender cómo son y qué sucede con los sonidos vocálicos del español hablado por chinos. Esta investigación será de

utilidad para la didáctica de la lengua española como LE para chinos.

2. OBJETIVOS

El objetivo principal de esta investigación es realizar una descripción de las características acústicas de las cinco vocales españolas habladas por chinos. Como objetivos secundarios, vamos a comprobar si hay diferencias entre hombres y mujeres en la pronunciación de dichas vocales y si hay diferencias entre átonas y tónicas en el español hablado por chinos. Además, compararemos los resultados de nuestro corpus con los estudios del español en habla de laboratorio y espontánea, y con los de la interlengua español hablado por chinos. Finalmente, plantear algunas propuestas didácticas para mejorar la enseñanza-aprendizaje de las vocales del español de los aprendices chinos.

3. CORPUS Y METODOLOGÍA

El corpus está constituido 46 minutos de grabaciones en habla semiespontánea de 21 alumnos chinos, 8 hombres, 35%, y 13 mujeres, 62%, entre 19 y 23 años de edad, de 8 provincias diferentes de China. Tienen un nivel entre A2 y B1 de español, porque han estudiado dos años de español en la Universidad Beihua en China y, en el momento de la grabación, llevaban dos meses como estudiantes internacionales en Barcelona. De los materiales grabados, hemos seleccionado 150 enunciados con más de 1500 palabras y hemos obtenido para nuestra investigación un total de 725 vocales, entre tónicas y átonas:

- 138 vocales anteriores altas, [i], que constituyen un 19% del corpus total.
- 174 vocales anteriores medias, [e], que constituyen un 24% del corpus total.
- 178 vocales centrales bajas, [a], que constituyen un 24,6% del corpus total.
- 164 vocales posteriores medias, [o], que constituyen 22,6% del corpus total.
- 71 vocales posteriores altas, [u], que constituyen un 9,8% del corpus total.

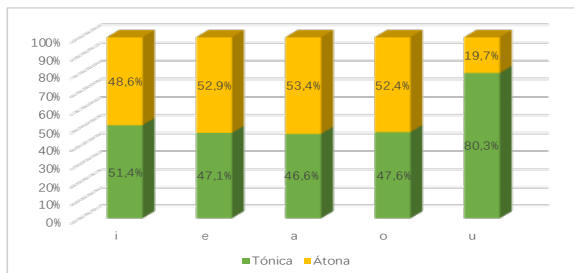


Figura 1: Porcentaje de presencia de cada vocal en el corpus.

Cada vocal fue analizada con el software *Praat* (Boersma y Weenink, 1992-2019), el cual genera un oscilograma y un sonograma de cada sonido que facilitan la representación visual de las características acústicas de las vocales. Con este programa, analizamos las vocales de forma aislada, con el fin de obtener las características acústicas de cada una (los valores de F_1 y F_2). Con los valores extraídos, calculamos los valores medios de cada vocal, así como también creamos los distintos gráficos y de los campos de dispersión. Finalmente, aplicamos la prueba estadística con la aplicación SPSS Statistics 17.0. Realizamos la prueba Anova para comprobar si hay diferencias significativas entre las variantes átona y tónica de cada vocal, y entre hombres y mujeres en el F_1 y en el F_2 .

4. RESULTADOS Y DISCUSIÓN

A partir de los tres objetivos que nos planteamos, exponemos los resultados según el sexo de los informantes, según las diferencias entre átonas y tónicas y comparamos nuestros resultados con el de otras investigaciones.

4.1. Resultados según el sexo de informantes

Como hemos constatado mediante prueba Anova, las cinco vocales españolas [i], [e], [a], [o] y [u] producidas por estudiantes chinos presentan una diferencia significativa ($p < 0,05$) entre hombres y mujeres en el F_1 y F_2 .

Hemos representado los resultados en campos de dispersión según el sexo de los informantes (Véase en los figuras 2 y 3) y tanto en el de los hombres como en el de las mujeres, coinciden con el dibujo del "Triángulo articulatorio" elaborado por Quilis (1999). En concreto, la [i] es anterior alta, la [e] es anterior media, la [a] es media baja, la [o] es posterior media y la [u] es posterior alta.

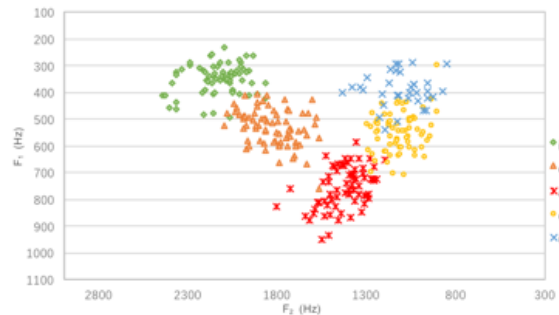


Figura 2: Campo de dispersión de las vocales del español hablado por informantes chinos en voz masculina.

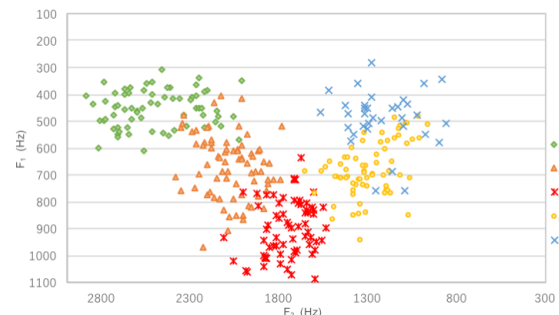


Figura 3: Campo de dispersión de las vocales del español hablado por informantes chinos en voz femenina.

Si comparamos las posiciones de las cinco vocales [i], [e], [a], [o] y [u] en la carta de formantes, la agrupación de las cinco vocales producidas por mujeres chinas se sitúa en una posición más baja e izquierda frente a la de hombres chinos. Además, comprobamos que las vocales de las mujeres ocupan un espacio más amplio que las emitidas por hombres porque tanto los valores de F_1 como los de F_2 de las mujeres son más altos.

4.2. Resultados entre átonas y tónicas

En cambio, como hemos comprobado que las diferencias entre átona y tónica no son significativas ($p > 0,05$) ni en hombres ni en mujeres, a excepción del F_2 de la [i]: para la voz masculina en el F_2 ($p = 0,005$) y para la voz femenina en el F_2 ($p = 0,001$), hemos propuesto un solo valor para cada vocal (Véase en la tabla 1).

Tabla 1: Valores medios de F_1 y F_2 de las vocales del español producidas por informantes chinos en voz masculina y femenina.

Sonidos	Hombre	Mujer	Hombre	Mujer
	F ₁	F ₁	F ₂	F ₂
[i]	351	457	2127	2486
[e]	530	666	1826	2089
[a]	757	908	1425	1761
[o]	550	679	1112	1296
[u]	392	491	1106	1240

4.3. Análisis comparativo

En la comparación con los resultados de otros autores, hemos constatado que los valores medios de las cinco vocales [i], [e], [a], [o] y [u] emitidas por chinos de nuestra investigación en el F₁ son más altos tanto en ambos sexos, como en átonas y tónicas o en ambas juntas, que en habla controlada (Quilis y Esgueva, 1983; Bradlow, 1995; Martínez Celdrán, 1995; Albalá et al., 2008) y en habla espontánea (Alfonso, 2014.). Esto significa que los informantes chinos realizan más abertura bucal y más constricción de la faringe que los otros hablantes.

En cuanto al F₂, las diferencias con el habla controlada (Quilis y Esgueva, 1983; Bradlow, 1995; Martínez Celdrán, 1995; Albalá et al., 2008), la tendencia es que los valores medios en las vocales posteriores en nuestra investigación son más altos, es decir, hay menos labialización y menos constricción linguovelar. Mientras que si los comparamos con el habla espontánea (Alfonso, 2014) los valores medios que hemos obtenido son más elevados en las anteriores, hay menos labialización y menos constricción linguovelar.

5. ACTIVIDADES DIDÁCTICAS

Considerando que los chinos no tienen diferencias acústicas significativas entre las vocales tónicas y átonas, proponemos algunos ejercicios:

- Discriminación auditiva: sensibilización y familiarización con los patrones acentuales del español, así como la diferenciación de sílabas tónicas y átonas en palabras, en frases.
- Práctica controlada: imitación y repetición de los patrones acentuales del español, tales como imitación de los modelos propuestos por el mismo profesor o bien utilizando el software *Praat* para que el estudiante chino sea consciente de la diferencia entre vocales átonas y tónicas, además de tener en cuenta el tono.

En relación con la dificultad que tienen los aprendices chinos con la labialización, proponemos

una serie de actividades didácticas para aumentarla y/o disminuirla:

- Observar e imitar el movimiento de los labios que realizan los españoles cuando hablan en habla espontánea. A continuación, deben grabar en video algunas frases y, posteriormente, analizar su movimiento labial.
- Practicar poniendo el dedo índice pegado ligeramente a los labios e indicar la acción de hacer silencio.

6. CONCLUSIONES

Los resultados que hemos obtenido en esta investigación serán útiles para lograr una aproximación de los rasgos del español hablado por chinos y para desarrollar nuevas herramientas que permitan al aprendiz chino de español visualizar y concretar las características del español como LE. Por otra parte, estos resultados son una nueva vía abierta para los que quieran investigar la pronunciación del español para extranjeros, y para crear propuestas didácticas más adecuadas para la enseñanza-aprendizaje de la pronunciación del español LE.

7. BIBLIOGRAFÍA

- Albalá, M., Battaner, E., Carranza, M., Gil, J., Llisterri, J., & Machuca, M. (2008). VILE: Nuevos datos acústicos sobre vocales del español, en A. Pamies y E. Melguizo (eds.), *Language Design. Journal of Theoretical and Experimental Linguistics. Special Issue 1: New Trends in Experimental Phonetics: Selected Papers From the IV International Conference on Experimental Phonetics, 11-14 february 2008*, Granada, 1-14.
- Alfonso Lozano, R. (2014). El vocalismo del español en el habla espontánea. *PHONICA*, vol. 9-10, 3-12.
- Boersma, P., & Weenink, D. (1992-2019). *PRAAT: Doing phonetics by computer*. <http://www.fon.hum.uva.nl/praat/>, [30/10/2019].
- Delattre, P. (1948). Un triangle acoustique des voyelles orales du français. *The French Review*, 21(6), 477-484.
- (1951). La interpretación fisiológica de los espectrogramas de sonido. *PMLA*, 864-875.
- Fant, G. (1960). *Acoustic Theory of Speech Production*. The Hague: Mouton.
- Gil Fernández, J. (2007). *Fonética para profesores de español: de la teoría a la práctica*. Madrid: Arco/Libros, Golpe.
- Grupo de Español del Comité de Didáctica de Lenguas Extranjeras en Escuelas Superiores Chinas. (1998). *Programa de Enseñanza para Cursos Básicos de las*

- Especialidades de Lengua Española en Escuelas Superiores Chinas*. Shanghai: 上海外语教育出版社.
- Quilis, A. (1981). *Fonética acústica de la lengua española*. Madrid: Gredos.
- (1999): *Tratado de fonología y fonética españolas*. Madrid: Gredos.
- Quilis, A., & Esgueva, M. (1983). Realización de los fonemas vocálicos españoles en posición fonética normal». In M. Esgueva, & M. Cantarero (eds.), *Estudios de fonética*. Madrid: Inst. "Miguel de Cervantes", Consejo Superior de Investigaciones Científicas, 159-252.
- Laroya, C. (1995). *Pronunciación*. Oxford: Oxford University Press.
- Martínez Celadrán, E. (1989). *Fonética: con especial referencia a la lengua castellana*. Barcelona: Teide.
- (1995): En torno a las vocales del español: análisis y reconocimiento. *Estudios de fonética experimental*, 195-218.
- Ministerio de Educación, P.R.C. (2017). *普通高中课程方案 [Proyecto de Currículum de Bachillerato Normal]*. Beijing: 教育部.
- (2018): *普通高中西班牙语课程标准 [Normativas de la asignatura de español para bachillerato normal]*. Beijing: 人民教育出版社.
- Quilis, A. (1988). Estudio comparativo entre la entonación portuguesa (de Brasil) y la española. *Revista de Filología Española*, 68 (1-2), 33-65.

DOES METALINGUISTIC KNOWLEDGE ABOUT THE L2 ENHANCE THE EFFECTIVENESS OF L2 PERCEPTUAL TRAINING?

Angélica Carlet¹, Juli Cebrian², Núria Gavalda³, Celia Gorba²

¹Universitat Internacional de Catalunya, ²Universitat Autònoma de Barcelona, ³Universidad Internacional de La Rioja

acarlet@uic.es, juli.cebrian@uab.cat, nuria.gavalda@unir.net, celia.gorba@uab.cat

ABSTRACT

This paper examines the role of metalinguistic knowledge in the ability to perceive L2 vowel sounds, generalize and retain knowledge as a result of high variability phonetic training (HVPT). Native Catalan/Spanish speakers with different levels of L2 experience underwent identification (ID) or discrimination (DIS) training regimes. All participants were tested on vowel identification immediately before, after and some months after training. Perceptual improvement was observed with all trained participants, showing beneficial effects of training for learners with varying levels of L2 experience (Iverson, Pinet & Evans, 2012). However, the more experienced learners showed more improvement in the identification of vowels in non-sense words, showing that metacognitive learning may play a relevant role in L2 vowel learning.

Keywords: Metalinguistic knowledge, phonetic training, L2 vowel learning, L2 perception

1. INTRODUCTION

Experience in a second language (L2) is a prerequisite for learning novel sounds, and the greater the exposure to the target language, the more likely it is that acquisition may take place (Flege, Munro, & MacKay, 1995). Therefore, the limited exposure to authentic input that is commonly found in a L2 instructional setting hinders learners' ability to perceive non-native sounds. In this problematic context, specialized phonetic training emerges as a potential alternative to authentic input. In particular, high variability phonetic training (HVPT), that is, perceptual training involving multiple talkers and stimuli, has been found to enhance L2 learners' ability to perceive target language sounds (Logan, Lively & Pisoni, 1991; Flege, 1995; Shinohara & Iverson, 2018; Carlet & Cebrian, 2019, among many more). Besides, previous studies have provided evidence of generalization to untrained structures (new stimuli, phonetic contexts, speakers, dimensions) and of retention of knowledge beyond the training period, showing that robust learning may take place as a result of training (Logan & Pruitt, 1995).

Interestingly, a positive effect of HPVPT has been identified for learners with varying levels of L2 experience (Iverson, Pinet & Evans, 2012). The

authors explored whether training was equally effective for French learners of English differing in proficiency level and amount of exposure. A group of beginner learners in their home country and a group of mid-intermediate learners living in the UK underwent a vowel identification training regime and were tested on the identification, discrimination and production of 14 English vowels and diphthongs. Both groups showed a slight effect of training on discrimination ability, and improved significantly in identification and production as a result of training, indicating that amount of experience is not an underlying factor for improvement to take place.

In order to explore this issue further, this paper evaluated the performance of metalinguistically naïve learners with no prior knowledge of English phonetics (henceforth, "naïve" learners) and metalinguistically experienced learners who had been attending an English phonetics and phonology course (henceforth, "experienced" learners), in order to test the potential role of metalinguistic knowledge on the perception of L2 vowels. Moreover, this study aimed at assessing whether specialized instruction had an impact on the generalization of learning to new stimuli, and/or on the retention of L2 vowels knowledge over time.

2. METHODS

This paper reports and contrasts the results of two different studies. A few methodological aspects that are shared by the two studies are explained first, followed by individual sections explaining the peculiarities of each individual study.

Both studies followed a pre-test/post-test/delayed post-test design. The target sounds were five English vowels sounds (i.e. /æ/-/ʌ/, /ɪ/-/i/, /ɜ:/), and the stimuli consisted of CVC non-words (for testing and training) and CVC real words to test generalization to novel stimuli. Moreover, stimuli from different talkers were used for testing and training so that generalization to novel talkers was assessed along with the main effect of training.

The stimuli were produced by six speakers of Standard Southern British English (3 females, 3 males), who were recorded in a soundproof chamber at University College London. Stimuli were embedded in a carrier sentence and each word was recorded three times, with additional repetitions whenever necessary.

The testing task was a 7-alternative forced-choice identification task (ID) with no feedback. Participants were trained either with a 7-alternative forced-choice (ID) task or a categorical discrimination task (DIS). The number of trials was doubled during the ID task, so that both trained groups had the same amount of exposure to the stimuli (ID: 576 trials; DIS: 288 trials). Immediate feedback after each trial and global feedback at the end of the training sessions were provided. Training and testing were implemented with the software TP (Rauber, Rato, Kluge & Santos, 2011).

2.1. Study 1

The participants of Study 1 were sixty “meta-linguistically experienced” learners, who were second-year English majors, with current knowledge of English phonetics and phonology (Carlet, 2017). They were divided into three groups (ID, DIS, and control group (CG)). Participants underwent five 30-minute sessions of perceptual training and were tested on the identification of L2 vowels before, after and two months after training. Due to participant mortality, only 38 participants completed this very last phase: 9 controls, 17 ID trainees and 12 DIS trainees. L2 production was also tested, but it will not be reported here.

2.2. Study 2

The participants of Study 2 were forty-five “meta-linguistically naïve” learners, who were first-year

English majors, with no prior knowledge of English phonetics and phonology. They were also divided into three groups (ID, DIS, CG). Participants underwent six 30-minute sessions of training and were tested on the identification of L2 vowels before, after and four months after training. L2 discrimination and production were also tested, but they will not be reported here. Some of the original participants were discarded from the study because they did not complete all the training sessions or the post-tests. The final number of participants per group was 13, 14 and 11 for the ID, DIS and control group, respectively.

3. RESULTS AND DISCUSSION

3.1. Study 1

3.1.1. L2 vowel perception and generalization effects

The percentage correct identification of non-words and real words (generalization) at pre-test and post-test, and gain scores (i.e. difference in percentage points between post-test and pre-test) for the experimental groups and the controls were calculated and are shown in Table 1.

Table 1. Percent correct identification of English vowels in non-words and real words at pre-test and post-test and gain scores per group (Study 1).

NON WORDS	CG % (SD)	DIS % (SD)	ID % (SD)
PRE-TEST	54.1 (9.9)	55.5 (6.5)	52.9 (9.5)
POST-TEST	57.8(10.2)	65.3 (9.7)	79.1 (13.3)
GAIN	3.7	9.8	26.3
REAL WORDS	CG % (SD)	DIS % (SD)	ID % (SD)
PRE-TEST	72.2 (11)	78.2 (9.7)	73.1 (11.2)
POST-TEST	79.5(10.3)	79.7(11.1)	88.5 (9.5)
GAIN	7.3	1.5	15.4

Since the three groups did not differ statistically at pre-test ($F(2, 51) = .37, p = .68$), the effect of training was analysed by comparing the amount of gain by means of a generalized linear mixed model (GLMM) with group as fixed factor. Regarding the identification of non-words, a significant main effect of group was found ($F(2, 51) = 53.29, p <.01$), and pairwise comparisons revealed that both experimental groups (DIS and ID) significantly outperformed the controls ($p <.01$ in both cases). Moreover, the model revealed that the ID trainees significantly outperformed the DIS trainees ($p <.01$). These results suggest that even

though both training methods (ID and DIS) are efficient when training L2 vowels, the ID method may be superior to the DIS method.

Regarding the identification of L2 vowel sounds embedded in real words (generalization), the GLMM results yielded a significant main effect of group, ($F(2, 51) = 9.16, p < .001$). Sequential Bonferroni pairwise comparisons confirmed that only the ID trainees outperformed the CG, $p < .05$. Moreover, the ID group outperformed the DIS group, indicating that generalization to L2 vowels embedded in real words only occurred after receiving identification training ($p < .01$). This result confirms the robustness of an ID training method when training L2 vowels (Logan & Pruitt, 1995).

3.1.2 Retention results

Two months after the completion of the training regime, a total of 38 out of the 60 original participants completed a delayed post-test. The results of the three groups at this time were numerically higher than or similar to the post-test results. Analyses with test time as a fixed effect (pre-test, post-test, delayed post-test) for the control group revealed no significant effect of time ($F(2, 48) = 1.84, p > .05$), confirming that the controls performed similarly across all three testing times. Regarding the experimental groups, results in each case yielded a significant effect of time (ID: $F(2, 48) = 51.35, p < .01$; DIS: $F(2, 33) = 7.62, p < .01$). Importantly, pairwise comparisons confirmed that the performance at pre-test significantly differed from the performance at post-test and delayed post-test ($p < .001$ in both cases). Moreover, the results of the delayed post-test did not differ from the results of the post-test, indicating that the L2 vowel learning acquired through HVPT training was retained over a period of two months.

3.2. Study 2

3.2.1. L2 vowel perception and generalization effects

The percent correct identification for the three groups in Study 2 at pre-test and post-test are given in Table 2. The pattern of results is very similar to the one observed in Study 1 in that the group that seemed to have the greatest improvement was the ID group. However, while the overall scores are similar across stimulus type in Study 2, the identification scores for real words were higher than for non-words for the experienced learners in Study 1.

In the same fashion as in Study 1, a GLMM was conducted on the amount of gain (difference in percent points between post-test and pre-test scores) with group as the fixed factor. There was a significant

effect of group both in the case of non-word stimuli ($F(2, 37) = 6.57, p = .004$) and real word stimuli ($F(2, 37) = 4.55, p = .017$). In both cases, post-hoc comparisons showed that the ID group outperformed the DIS and the control group. Separate statistical analyses revealed that ID and DIS's score at post-test were significantly higher than at pre-test both for non-words and for real words. The difference was not significant for the untrained controls in the case of the non-words, but they also showed significant improvement with real words, which may be the result of general exposure to English in the course of the duration of the training regime. Nevertheless, the greatest improvement was obtained by the ID group in all conditions.

Table 2. Percent correct identification of English vowels in non-words and real words at pre-test and post-test and gain scores per group (Study 2).

NON WORDS	CG % (SD)	DIS % (SD)	ID % (SD)
PRE-TEST	55.1 (14)	50 (10)	53.6 (13)
POST-TEST	57.5 (13)	59 (15)	69.9 (9)
GAIN	2.4	9	16.3
REAL WORDS	CG % (SD)	DIS % (SD)	ID % (SD)
PRE-TEST	56.2 (19)	56.7 (10)	55.6 (9)
POST-TEST	66.5 (12)	65.1 (16)	78.3 (11)
GAIN	10.2	9.4	22.6

3.2.2 Retention results

As in Study 1, retention of knowledge was tested with a delayed post-test. In this case, however, the delayed test occurred even later, four months after training had ended (as opposed to two months later in Study 1). Ten participants did not complete the delayed test. Thus, the total number of participants at this time was 28 (10, 11, and 7 for ID, DIS and control group, respectively). The results for the pre-test, post-test and delayed test were subjected to a GLMM analysis with testing time and group as fixed factors. The results for the trained groups did not differ significantly from the post-test results and differed from the pre-test results. Thus, despite the long interval since the post-test, the improvement observed from pre- to post-test was maintained four months later.

3.3. Assessing metalinguistic knowledge

In order to compare the results of the two studies, a GLMM analysis was conducted on the gain scores

obtained in the identification of L2 vowels for the trained groups in both studies (ID and DIS “naïve” and “experienced” groups). In the case of the non-words, the statistical analysis yielded a significant effect of group ($F(3,61) = 15.08, p = .000$). As can be observed in Figure 1, the ID groups obtained better results than the DIS groups and the more experienced ID trainees outperformed all the other groups. This result suggests that identification training may be the most efficient training regime for L2 vowel learning particularly when combined with metalinguistic knowledge.

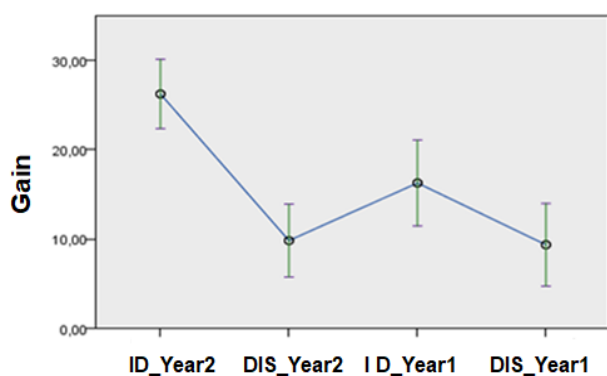


Figure 1. Gain scores of L2 vowel identification in non-words by naïve (year 1) and experienced (year 2) learners.

Regarding generalization to real words (see Fig. 2), the statistical analysis also revealed a significant effect of group ($F(3,61) = 10.82, p = .000$). However, despite the numerical differences, the two ID groups did not differ significantly. The same was true for the two DIS groups, showing that experience did not play a role when identifying real word stimuli. The group differences here are connected to the fact that generalization to real words took place for the ID trained groups only, showing again a greater effect with ID training compared to DIS training.

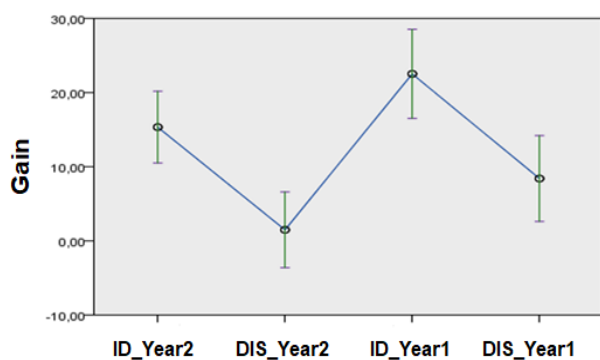


Figure 2. Gain scores of L2 vowel identification in real words by naïve (year 1) and experienced (year 2) learners.

4. GENERAL DISCUSSION AND CONCLUSIONS

This paper investigated the relationship between learners’ metalinguistic knowledge and the effectiveness of high variability phonetic training (HVPT). The results of two parallel studies involving L1 Spanish/Catalan learners of English were contrasted: a group of undergraduate students who had completed a course on English phonetics, which included a detailed description of the English vowel system (Study 1), and a group of students with no such metalinguistic knowledge (Study 2). Learners were trained by means of either identification (ID) or discrimination (DIS) tasks, or remained untrained (controls). In addition, the study investigated the effect of training on trained items (non-word stimuli) and untrained items (real words), and also examined retention of knowledge over time (2 and 4 months after the completion of the training regime).

The results indicate that short phonetic training regimes are effective for different levels of L2 experience, in line with Iverson et al. (2012). However, the comparison of the two populations (Study 1 and Study 2) suggests that HVPT may work best when combined with the metalinguistic knowledge acquired during English phonetics instruction, as the group that showed the greatest improvement was one of the “metalinguistically experienced” groups. Regarding the type of training task, the ID trainees improved the most, and were able to generalize the knowledge to real word stimuli to a greater extent. Thus, ID training proved to be overall more beneficial for training L2 vowels than DIS training. This advantage may be accounted for by an effect of task familiarity, as the ID group was trained and tested with the same type of task (although the stimuli and talkers were different). Still, similar studies on consonant perception have reported comparable improvements for ID and DIS groups (Carlet, 2017; Flege 1995; Nozawa, 2015), suggesting that different tasks may be more suited for training different types of segments. Importantly, the fact that retention of knowledge was found for all trained groups, two and four months after training had ended, together with the evidence of generalization of learning (to new stimuli and new voices) highlights the efficacy of the training regimes and suggests that robust learning may occur as a result of HVPT (Logan & Pruitt, 1995).

This research was supported by research grants FFI2017-88016-P and FFI2013-46354-P from the Spanish Ministry of Economy and Competitiveness, and by the research group Grant 2017SGR34 from the Catalan Government.

6. REFERENCES

- Carlet, A. (2017) L2 perception and production of English consonants and vowels by Catalan speakers: The effects of attention and training task in a cross-training study". Unpublished PhD dissertation. Universitat Autònoma de Barcelona.
- Carlet, A. & Cebrian, J. (2019) Assessing the effect of perceptual training on L2 vowel identification, generalization and long-term effects. In Anne Mette Nyvad, Michaela Hejná, Anders Højen, Anna Bothe Jespersen & Mette Hjortshøj Sørensen (eds.): *A Sound Approach to Language Matters – In Honor of Ocke-Schwen Bohn*. Dept. of English, School of Communication & Culture, Aarhus University, pp. 91-119.
- Flege, J. E. (1995). Two procedures for training a novel second language phonetic contrast. *Applied Psycholinguistics*, 16, 425-442.
- Flege, J., Munro, M., & MacKay, I. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, 97(5), 3125-3134.
- Iverson, P., Pinet, M., & Evans, B. G. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics*, 33(01), 145-160.
- Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America*, 89(2), 874-886.
- Logan, J., & Pruitt, J. (1995). Methodological issues in training listeners to perceive non- native phonemes. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross Language Research* (pp. 351-378). Timonium, MD: York Press.
- Nozawa, T. (2015). Effects of training methods and attention on the identification and discrimination of American English coda nasals by native Japanese listeners. In the Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow. ISBN 978-0-85261-941-4
- Rauber, A., Rato, A., Kluge, D., & Santos, G. (2011). TP (Version 3.1). [Software]. Brazil: Worken. [http://www.worken.com.br/tp_regfree.php?l=i].
- Shinohara, Y., & Iverson, P. 2018 High variability identification and discrimination training for Japanese speakers learning English /r/-/l/. *Journal of Phonetics* 66, 242-251.

LECTURA EN VEU ALTA, PAUSES SILENCIOSES I IOGA NIDRA

Josefina Carrera-Sabaté, Jesús Bach Marquès

Universitat de Barcelona
jcarrera@ub.edu, jbach93@gmail.com

ABSTRACT

Nombrosos estudis corroboren que l'entonació i les pauses són especialment rellevants per a la comprensió de textos llegits. D'altres constaten que la pràctica de la meditació comporta beneficis cognitius i, en el cas del ioga, la respiració s'alenteix. Per conèixer si en la lectura en veu alta hi ha modificacions que es poden relacionar amb respiracions més lentes, s'ha dissenyat un experiment per analitzar possibles diferències en les pauses silencioses d'un text oral llegit abans i després d'una pràctica de ioga nidra.

Els participants d'aquest estudi són 19 joves que han llegit un text abans i després d'una sessió de ioga. Els resultats mostren un augment de pauses en la lectura post-ioga, així com un increment general en la durada d'aquestes pauses i en la durada dels textos llegits. Tot sembla indicar que la pràctica de ioga nidra pot ser útil per a la lectura en veu alta.

Paraules clau: pauses, ioga, lectura, comprensió oral

Many studies state that intonation and pauses are especially relevant for the comprehension of read-aloud texts. Others prove that the practice of meditation brings with it cognitive benefits, and in the case of Yoga, breathing slows down. To find out whether there are any modifications when reading aloud that could be related to slower breathing, an experiment has been designed to analyse possible differences in the silent pauses of a text read aloud before and after a practice of Yoga Nidra.

The participants of this study are 19 youngsters who read a text before and after a yoga session. The results show an increase of pauses in the post-yoga reading, as well as a general lengthening of the duration of the pauses and of the read texts. Hence, it seems clear that reading after doing Yoga Nidra can be useful for read-aloud tasks.

Keywords: pauses, yoga, reading, oral comprehension

1. INTRODUCCIÓ

En la lectura d'un text en veu alta, sembla que els tipus de pauses i el canvi de F0 al final de les frases són crucials per comprendre un text llegit (vegeu Benjamin i Schwanenflugel, 2010; Miller i Schwanenflugel, 2006, 2008; Schanenflugel et. al, 2004).

En un altre ordre de coses, diversos estudis han demostrat els efectes positius de la meditació i el ioga sobre diferents aspectes de la cognició com ara l'atenció, la memòria o el funcionament executiu (vegeu Chiesa, Calati i Serretti, 2011; Vago, Gupta i Lazar, 2019). També s'ha comprovat que la meditació té efectes directes sobre la respiració atès que les persones que mediten tenen un ritme més lent de

respiració que les que no practiquen meditació (vegeu Wielgosz, Schuyler, Lutz, Davidson, 2016).

A l'hora de parlar i de llegir en públic, un dels reptes principals és la gestió de la respiració i de l'estrès. Per saber si els canvis de respiració descrits després de les pràctiques meditatives també poden tenir impacte en la lectura en veu alta s'ha dissenyat un experiment relacionat amb la pràctica de *ioga nidra*. En aquest experiment, s'analitzen les diferències en la lectura d'un mateix text en dues situacions acadèmiques diferents: l'una després d'haver escoltat unes explicacions a classe i l'altra després d'haver fet una pràctica d'aquest tipus de ioga.

El *ioga nidra* és una pràctica de meditació profunda que es defineix, en un principi, com un mètode de

relaxació física, emocional i mental i una eina de desenvolupament de l'atenció plena. Mitjançant aquest ioga, l'impacte dels sentits es va desconectar de la ment i aquesta es focalitza en l'experiència interna. Així, el còrtex cerebral no rep informació dels òrgans motors i la ment conscient passa a un segon pla i, com que no es donen ni seleccions ni judicis de les informacions que es reben, la ment és molt més receptiva (vegeu Satyananda Saraswati, 2003).

El propòsit d'aquest treball és determinar fins a quin punt la pràctica del ioga nidra pot influir en la durada dels textos oralitzats i en la durada de les pauses silencioses (així com en el nombre que se'n fan).

Abans d'abordar el tema, seguint Gil (2007), podem definir les pauses silencioses com a silencis intercalats en el discurs que poden estar relacionats amb la respiració i la interrupció de la fonació; també poden tenir funcions demarcatives o estilístiques. Seguint amb la classificació de les pauses silencioses en lectura en veu alta, i ara partint de Campione i Véronis (2002), aquestes pauses es poden dividir en tres grups segons la durada que tinguin: breus (<200ms), mitjanes (200-1000 ms) i llargues (>1000 ms). A més, pel que fa a les durades, hi ha unes similituds interlingüístiques entre l'anglès, el francès, l'alemany, l'italià i l'espanyol: les pauses breus es donen al voltant de 100-150 ms i les pauses mitjanes, al voltant de 500-600 ms.

2. METODOLOGIA

2.1. Corpus i procediment

El corpus analitzat consisteix en les dades obtingudes de dues lectures en veu alta del text *La Tramuntana i el sol* (de 104 paraules) per part de 19 joves estudiants. El text es va llegir el mateix dia en dos contextos diferents: després d'unes explicacions per part de la professora i després d'una pràctica de *ioga nidra*. Els mateixos lectors de les frases van enregistrar-se amb els dispositius mòbils de què disposaven. La durada total dels 38 enregistraments del conte ha estat de 22 minuts.

El corpus s'ha etiquetat i transcrit amb el programa Praat (Boersma Weenink, 2013) mitjançant un script (Kawahara, 2010) modificat *ad hoc* per calcular les pauses silencioses (en algun cas, el límit entre el final d'una pausa i el principi d'una oclusiva sorda ha estat problemàtic). L'anàlisi estadística s'ha dut a terme amb l'ajut de la llibreria *glmmTMB*, en l'entorn del

programa R, i l'anàlisi del nombre de pauses, amb un model lineal generalitzat mixt, amb distribució *poisson*. L'anàlisi de la durada dels diferents elements, en canvi, s'ha realitzat amb models lineals mixtos. La variable dependent durada ha estat prèviament normalitzada per a un millor ajustament (en els casos dels fragments oralitzats, a partir d'un logaritme neperà; en els casos de les pauses silencioses, a partir de l'arrel quadrada).

2.2. Participants

Els lectors són catalanoparlants de diferents zones dialectals del català sense cap problema auditiu ni de pronúncia i tenen entre 21 i 24 anys. Són 18 dones i 1 home.

3. RESULTATS

3.1. Durada dels textos i dels fragments entre pauses

La durada total mitjana dels textos ha estat significativament més gran en la lectura post-ioga (38,3 segons) que en la pre-ioga (33,7 segons), tal com es pot observar en la figura 1 ($\chi^2(1) = 32.56, p < 0,001, d = 1,851$).

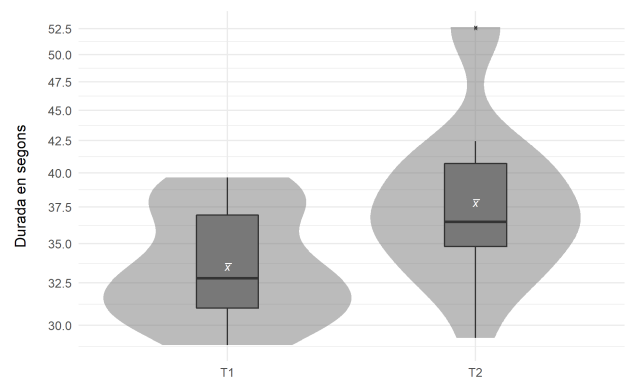


Figura 1: Durada total dels textos en lectura pre-ioga (T1) i post-ioga (T2)

Al mateix temps, s'han analitzat les diferències de durada dels sis diferents fragments en què es pot dividir el text atenent als elements delimitadors de la puntuació escrita (punts i seguit i punts i coma). Aquests fragments es mostren en la taula 1:

Taula 1: Pauses més llargues en les lectures del conte *La tramuntana i el sol*

F1. La tramuntana i el sol es discutien sobre qui era el més fort, quan, de sobte, van veure un viatger que s'acostava embolicat en una gran capa.
F2. Van convenir que el primer que aconseguiria que el viatger es tragués la capa seria considerat el més fort.

F3. La tramuntana es posà a bufar amb gran fúria, però com més bufava, més s'abrigava amb la capa el viatger;
F4. a la fi, va deixar córrer l'empresa.
F5. Aleshores el sol començà a brillar i, immediatament, el viatger, ben escalfat, es va treure la capa.
F6. Així la tramuntana va haver de reconèixer que el sol era el més fort.

La durada mitjana de cadascun d'aquests fragments també ha estat significativament superior en la lectura postioiga (5,7 segons) que en la preioga (5,1 segons), tal com es pot observar en la figura 2 ($\chi^2(1) = 28,467, p < 0,001, d = 1,441$).

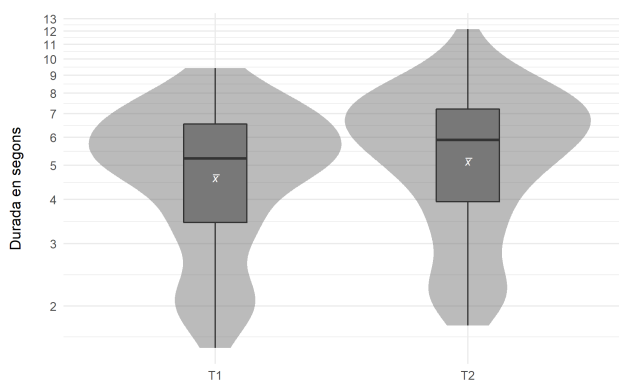


Figura 2: Durada dels fragments entre pauses en lectura preioga (T1) i postioiga (T2)

3.2. Nombre de pauses

S'ha analitzat, també, si en la lectura postioiga s'havien produït més pauses que en la preioga, i així ha resultat ser, tot i la diferència de durada mitjana dels textos entre preioga i postioiga (vegeu la secció 3.1). En tot el corpus, en la lectura preioga es van comptar un total de 412 pauses, mentre que en la lectura postioiga se'n van comptar 441. Dels 19 lectors, 13 van incrementar el nombre de pauses d'un temps a l'altre.

Aquests valors totals impliquen que, mentre que el nombre mitjà de pauses del preioga ha estat de 21,8, el del postioiga ha augmentat fins a 23,3. Aquests valors s'han analitzat estadísticament tenint en compte la diferència de la durada dels textos com a factor de ponderació. Un altre cop, malgrat que els valors són propers, el model estadístic ha determinat que l'augment en el nombre de pauses és estadísticament significatiu ($\chi^2(1) = 30,80, p < 0,001, d = 0,063$) (vegeu la taula 2). Si bé en alguns casos els informants han produït el mateix nombre de pauses en la lectura preioga i en la postioiga, o fins i tot alguna pausa menys, la tendència observada és un augment de pauses silencioses en la lectura postioiga.

Taula 2: Diferència entre el nombre de pauses en les lectures preioga i postioiga dels dos primers fragments (F1 i F2 de la taula 1)

Context de la pausa	Nombre de pauses preioga	Nombre de pauses postioiga	Diferència
tramuntana i	1	0	-1
sol es	9	12	+3
discutien sobre	1	4	+3
sobre qui	0	3	+3
fort quan	19	19	0
quan de	8	7	-1
sobte van	19	18	-1
viatger que	4	6	+2
s'acostava embolicat	1	1	0
embolicat en	8	9	+1
capa van	19	19	0
convenir que	12	15	+3
primer que	0	1	+1
aconseguiria que	12	14	+2
viatger es	2	1	-1
tragués la	0	1	+1
la capa	1	0	-1
capa seria	13	15	+2
seria considerat	0	1	+1
considerat el	4	4	0
fort la	19	19	0

3.3. Pauses silencioses

Al marge de l'augment de durada dels textos i fragments, així com del nombre de pauses en la lectura postioiga, la durada global de les pauses silencioses també ha estat diferent entre les dues lectures (vegeu la figura 3): en la lectura preioga ha estat de 320 ms i, en la postioiga, de 422 ms., valors significativament superiors als de la preioga ($\chi^2(1) = 23,46, p < 0,001, d = 0,607$) (vegeu la figura 3).

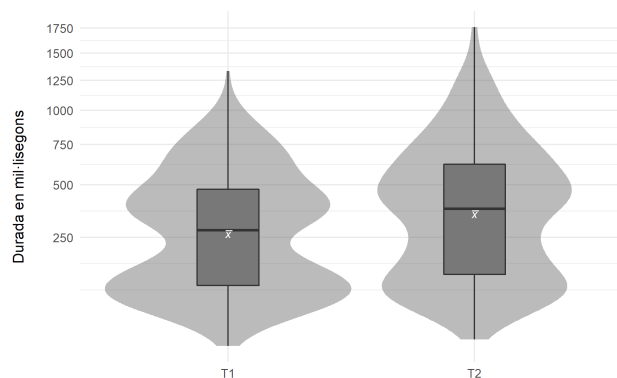


Figura 3: Durada de les pauses en lectura preioga (T1) i postioiga (T2)

Endemés, analitzant amb una mica més de detall les pauses de les dues lectures, hi ha algunes diferències entre les preioga i postioga. Així, en la lectura preioga, s'observen dos pics: el primer, que correspon a pauses breus, es troba al voltant dels 90 ms, i el segon, que correspon a pauses mitjanes, es troba al voltant dels 700 ms. (Vegeu la figura 4). Aquests resultats coincideixen amb els observats en lectura en veu alta de textos en anglès, alemany, francès, italià i espanyol (vegeu Campione i Véronis, 2002).

En la lectura postioga, les pauses més freqüents es troben al voltant de 475 ms, i coincideixen amb les pauses de durada mitjana que classifiquen Campione i Véronis (2002). Les pauses més llargues tenen una durada propera a 900 ms, i s'acosten als valors de les pauses llargues classificats per Campione i Véronis (2002). Vegeu la figura 4.

Les pauses que han resultat més llargues es corresponen amb els finals de frases que estan ortografiats amb punt i seguit i, en un cas, amb punt i coma (vegeu la taula 1). Si ens focalitzem en aquest tipus de pauses llargues (que serien les que tindrien una motivació estructural més clara des del punt de vista lingüístic), s'observa que la durada mitjana en lectura preioga ha estat de 671 ms i en la postioga, de 854 ms (vegeu la figura 4).

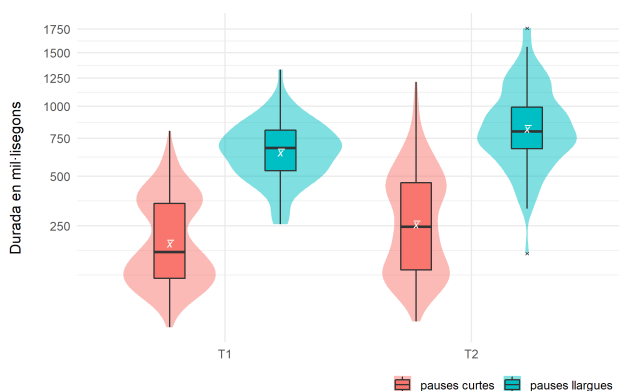


Figura 4: Durada de les pauses curtes (generalment intratoracionals) i de les llargues (aquelles relacionades amb un punt i seguit o amb un punt i coma en l'escriptura). Lectura preioga (T1) i postioga (T2)

4. DISCUSSIÓ

L'anàlisi de 38 lectures en veu alta del conte *La tramuntana i el sol* ha permès observar, d'entrada, que la durada de les pauses en lectura en veu alta en català i sense haver estat precedides per una pràctica de *ioga nidra* és similar a la durada observada en

altres llengües romàniques i germàniques (vegeu Campione i Véronis, 2002).

La durada dels textos i dels fragments de textos, així com la distribució i la durada de les pauses en les lectures dutes a terme abans i després de la pràctica de *ioga nidra* ens permeten observar durades dels textos i pauses significativament majors en les lectures en veu alta postioga. Les lectures postioga, a més, tenen un major nombre de pauses silencioses. Aquestes modificacions en la segmentació del conte i també en la durada de les pauses es poden explicar pels canvis fisiològics i respiratoris que comporta la meditació i l'atenció plena (Wielgosz, Schuyler, Lutz, Davidson, 2016). D'aquesta manera, la pràctica de *ioga nidra* s'albira com una eina útil per a la millora de la locució en veu alta.

5. CONCLUSIONS

Aquest és un estudi exploratori que combina l'observació de les pauses silencioses en lectura en veu alta amb la pràctica de *ioga nidra*. L'impacte que el *ioga* té sobre el cos i la ment dels lectors sembla expandir-se a la locució, de manera que les lectures en veu alta després de la pràctica de *ioga nidra* tenen una major durada i més pauses silencioses (que també són més llargues), cosa que pot ajudar a millorar la dicció i comprensió dels textos. Caldrien més estudis per confirmar aquestes observacions inicials.

6. BIBLIOGRAFIA

- Benjamin, R. G.; Schwanenflugel, P. J. (2010). Text complexity and oral reading prosody in young readers. *Reading Research Quarterly*, 45, 388–404.
- Boersma, P.; & Weenink, D. (2013). Praat: doing phonetics by computer, Version 6.1.03 [Computer program].
- Campione, E.; Véronis, J. (2002). A large-Scale Multilingual Study of Silent Pause Duration. *Speech Prosody*, Aix-en-Provence. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.2.6135&rep=rep1&type=pdf>
- Chiesa A, Calati R, Serretti A. (2011). Does mindfulness training improve cognitive abilities? A systematic review of neuropsychological findings. *Clinical Psychology Review*, 31, 449–464.
- Gil, J. (2007). *Fonètica para profesores de español: De la teoría a la práctica*. Madrid: Arco/Libros.
- Kawahara, S. (2010). *Get duration* Praat script. http://user.keio.ac.jp/~kawahara/scripts/duration_getter.praat
- Miller, J.; Schwanenf lugellugel, P. J. (2006). Prosody of syntactically complex sentences in the oral reading of young children. *Journal of Educational Psychology*, 98(4), 839–853.

- Miller, J.; Schwanenflugel, P. J. (2008). A longitudinal study of the development of reading prosody as a dimension of oral reading fluency in early elementary school children. *Reading Research Quarterly*, 43(4), 336–354.
- Satyananda Saraswati, Swami (2003). *Yoga Nidra*. Munger: Yoga Publications Trust.
- Schwanenflugel, P. J., Hamilton, A. M, Kuhn, M. R., Wisenbaker, J. M., & Stahl, S. A. (2004). Becoming a fluent reader: Reading skill and prosodic features in the oral reading of young readers. *Journal of Educational Psychology*, 96(1), 119–129.
- Vago DR, Gupta RS, Lazar SW. (2018). Measuring Cognitive Outcomes in Mindfulness-based Intervention Research: A Reflection on Confounding Factors and Methodological Limitations. *Current Opinion in Psychology*, 28, 143–150.
- Wielgosz, J.; Schuyler, B.; Lutz, A.; Davidson, R.J. (2016). Long-term mindfulness training is associated with reliable differences in resting respiration rate. *Scientific Reports*. *Nature research*, 6. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4895172/>

LA ENTONACIÓN DEL ESPAÑOL DE LOS ÁNGELES. UNIDAD VS. DIVERSIDAD

Yolanda Congosto Martín

Universidad de Sevilla
ycongosto@us.es

ABSTRACT

El objetivo fundamental de este trabajo es mostrar los resultados del estudio contrastivo realizado sobre la entonación de la población hispana que reside en LA (mexicana, salvadoreña y guatemalteca), con la intención de comprobar si el proceso de inmersión al que se ha visto sometida en el país que los acoge y la convivencia entre ellos ha incidido de alguna manera en sus rasgos prosódicos y valorar en qué medida.

Para llevar a cabo esta investigación se parte de los estudios previos realizados en el proyecto internacional AMPER con informantes femeninas de las tres comunidades de habla mencionadas (Congosto Martín 2011, 2012, 2015, 2016, 2017, 2018, 2019, 2020 y 2021 en prensa). El estudio tiene también un carácter geoprosódico.

Palabras clave: prosodia, entonación, español de Los Ángeles, español mexicano, guatemalteco y salvadoreño

The main objective of this paper is to offer the results of the contrastive study conducted on the intonation of the Hispanic population residing in LA (Mexican, Salvadoran and Guatemalan), with the intention of checking whether the immersion process to which they have been subjected in the country that welcomes them and the coexistence between them has had some impact on their prosodic features and assess the extent to which they are.

To carry out this research, we start from the previous studies carried out within the AMPER international project with female informants from the three mentioned speaking communities (Congosto Martín 2011, 2012, 2015, 2016, 2017, 2018, 2019, 2020 and 2021 in the press). The study also holds a geoprosodic character.

Keywords: prosody, intonation, Los Angeles Spanish, Mexican, Guatemalan and Salvadoran Spanish

1. INTRODUCCIÓN

La comunidad hispana que reside en el condado de Los Ángeles procede fundamentalmente de tres países americanos: México, Guatemala y El Salvador, en proporciones distintas. La población de origen mexicano, la mayoritaria, supone el 36% del total de la población, 3 510 677 individuos, según el último censo oficial, mientras que la salvadoreña y guatemalteca se reducen a unos 126 000 y 66 000 habitantes, respectivamente.

Los estudios previos realizados, especialmente sobre el español mexicano de hablantes nativos que residen a uno y otro lado de la frontera entre EE.UU. y México¹, han dejado entrever que, más allá de los

rasgos propios de cada individuo y los relativos a la zona dialectal a la que pertenecen, la entonación mexicana supone un continuo entre ambos países. No ocurre igual con la entonación de los hablantes guatemaltecos que residen en LA (Congosto Martín 2020), pues presentan diferencias con respecto al nativo de Guatemala que no ha salido de su país; diferencias que afectan especialmente a la zona nuclear de la curva entonativa de las declarativas y que las acerca a la curva mexicana de LA. En este trabajo se añade a lo ya manifestado los resultados de la investigación realizada sobre la entonación salvadoreña con el objetivo de comprobar si realmente se ha producido o no un proceso de convergencia entre los hablantes de las tres

¹ Congosto Martín 2011, 2012, 2015, 2016, 2017, 2018, 2019 y 2021 en prensa.

comunidades hispanas de LA y de darse ver qué dirección ha seguido.

2. METODOLOGÍA

2.1. El corpus

Las muestras de habla digitalizadas han sido extraídas del corpus fijo del proyecto AMPER-español, en su versión adaptada al español de América, compuesto por 126 frases de foco amplio (63 declarativas y 63 interrogativas absolutas), que multiplicadas por sus tres repeticiones hacen un total de 378 enunciados por informante. Dichos enunciados respetan las tres posibilidades acentuales del español en sujeto y en el objeto, mientras que el verbo siempre es paroxítono. Ello nos ha permitido obtener información no solo del comportamiento de la curva en el tonema y en el pretonema, sino también de las particularidades geoprosódicas de los hablantes y del grado de adquisición de los rasgos prosódicos, lo que ha hecho efectiva la posibilidad de realizar un estudio contrastivo riguroso y exhaustivo.

2.2. Los informantes

De los 27 informantes grabados y analizados que conforman de momento la base de datos del proyecto AMPER-Los Ángeles, se han seleccionado para este estudio cuatro mujeres: dos de origen o ascendencia mexicana, una guatemalteca y otra salvadoreña, que presentan diferencias de carácter sociolingüístico. Dichas variables se entienden necesarias si realmente se quieren extraer consideraciones de carácter geoprosódico, de ahí que se hayan tenido en cuenta y valoradas en los resultados de la investigación.

2.3. El análisis de los datos

Para el análisis acústico instrumental y la descripción fonética de las curvas melódicas se han seguido las pautas metodológicas establecidas en el proyecto AMPER-español (Fernández Planas 2005 y López Bobo et al., 2007). Por su parte, para el etiquetaje fonológico se han seguido las consideraciones teóricas establecidas dentro del modelo métrico-autosegmental o modelo AM (Pierrehumbert 1980, Ladd 1996) y, en consecuencia, la propuesta revisada de etiquetaje prosódico del español de análisis entonativo Sp_ToBI presentada por Estebas y Prieto (2008, 2010). En este sentido, conviene puntualizar que la propuesta que se hace de acentos tonales y

tonos de frontera tiene una base fonética, para la que se ha considerado el umbral de percepción establecido en 1,5 semitonos (Pamies Bertrán et al., 2002).

3. ANÁLISIS ACÚSTICOS

Las investigaciones hasta ahora realizadas con la población de origen mexicano que vive a uno y otro lado de la frontera política entre EE.UU. y México (Congosto Martín 2019) nos han permitido establecer el grado de similitud y diferencias que existen entre ambas comunidades de habla y determinar su idiosincrasia entonativa.

En líneas generales, las declarativas presentan un primer pico tonal anclado en la sílaba postónica del sujeto [L+>H*], seguido de un progresivo descenso de F0, más o menos escalonado, hasta el final del enunciado [L*L%]. De las cinco informantes que se muestran en la Figura 1 (dos de ellas residentes en México –AX– y tres en Los Ángeles –AO–), solo la mujer angelina de segunda generación presenta un segundo pico bastante marcado e igualmente desplazado [L+>H*]. En la sección nuclear, todas las curvas tienden a terminan con un tono bajo [L*L%], si bien no son extrañas en estas variedades soluciones seguidas de tonos de frontera [M%] e incluso [H%].

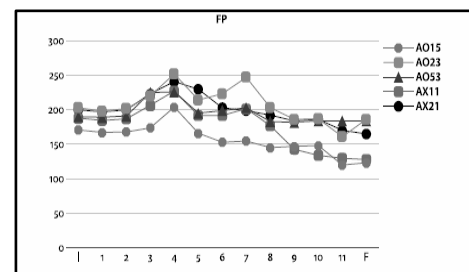


Figura 1: *La guitarra se toca con paciencia.* Datos comparativos de hablantes de México y Los Ángeles. Fuente: Congosto Martín (2019).

En lo que atañe a las interrogativas absolutas (Figura 2), todas las curvas presentan una trayectoria muy similar, con dos picos tonales en la zona prenuclear, el primero generalmente con valores más altos que el segundo y ambos posrealizados [L+>H*], a excepción de la informante de Ciudad de México que presenta un único pico máximo situado en la primera postónica del enunciado. En cuanto al núcleo, todas ellas terminan con progresión ascendente: FO [L+H*] / FP [L+H*H%] / FPP [L*HH%].

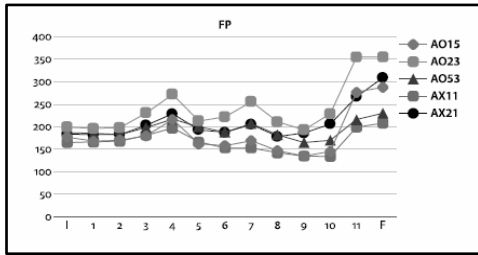


Figura 2: ¿La guitarra se toca con paciencia? Datos comparativos de hablantes de México y Los Ángeles. Fuente: Congosto Martín (2019).

Por su parte, las investigaciones previas realizadas igualmente con hablantes guatemaltecos que residen en Los Ángeles vs Guatemala (Congosto Martín 2020) han dejado entrever también puntos de encuentro y desencuentro entre ambas comunidades y, a su vez, entre estas y las de origen mexicano.

En el caso de las declarativas guatemaltecas de LA, las curvas melódicas presentan un primer pico máximo, de nuevo posrealizado [L+>H*], seguido de un descenso continuado del tono hasta el final del enunciado, que puede ser en algunos casos escalonado. A diferencia de esta, la curva entonativa del español de Guatemala presenta claramente tres movimientos tonales, en correlación con la estructura acentual del enunciado (Figuras 3 y 4).

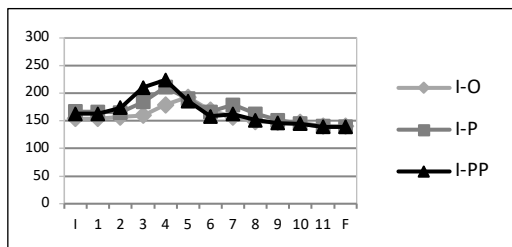


Figura 3: Declarativas. Español guatemalteco de LA. Fuente: Congosto Martín (2020).

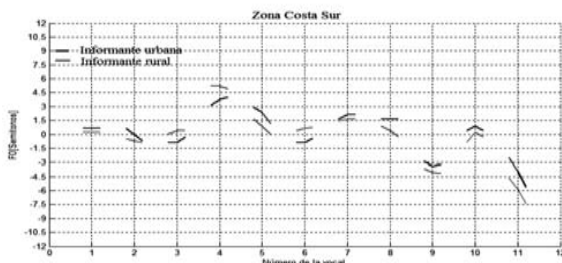


Figura 4: Declarativas. Español de Guatemala. *La cantante platica con paciencia*. Fuente: Utgård (2014).

En cuanto a las interrogativas absolutas (Figuras 5 y 6), todas las curvas, las de uno y otro punto geográfico, presentan dos picos tonales en la zona prenuclear: el primero, desplazado a la

postónica en el español de Guatemala [L+>H*] y alineado en el español guatemalteco de LA [L+>H*]; el segundo, también desplazado [L+>H*], mayoritariamente. En el núcleo, un movimiento tonal en anticadencia con movimientos similares a los realizados por la población de origen mexicano FO [L+>H*] / FP [L+>H*H%] / FPP [L*HH%].

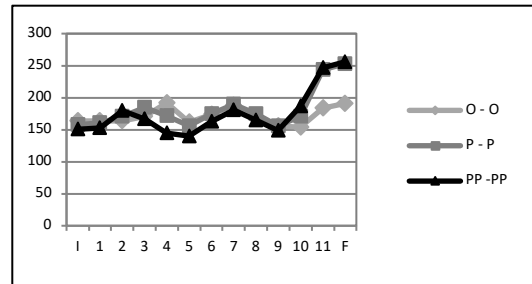


Figura 5: Interrogativas. Español guatemalteco de Los Ángeles. Fuente: Congosto Martín (2020).

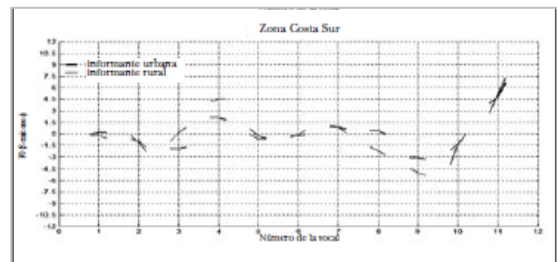


Figura 6: Interrogativas. Español de Guatemala. *La cantante platica con paciencia*. Fuente: Utgård (2014).

Por último, lo hasta ahora analizado sobre la entonación del español salvadoreño de Los Ángeles (Congosto Martín 2018), tiende a presentar comportamientos similares a los encontrados en las otras dos comunidades de hispanos que residen en LA: declarativas en cadencia, con un primer pico tonal posrealizado [L+>H*], seguido de un descenso de F0 que, al igual que en los casos previos, a veces puede ser escalonado (Figura 7).

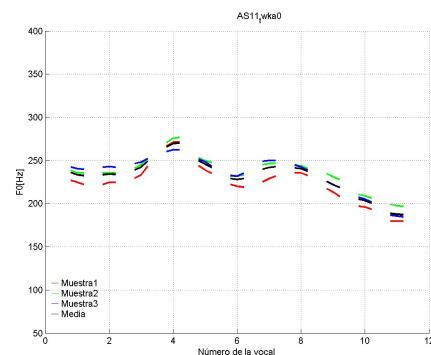


Figura 7: Declarativas. Español angelino salvadoreño. *La guitarra se toca con obsesión.*
Fuente: Congosto Martín (2018).

En cuanto a las interrogativas absolutas, de nuevo las curvas entonativas presentan dos picos tonales en la zona prenuclear, ambos posrealizados [L+>H*] y una configuración tonemática en progresión ascendente FO [L+H*] - FP [L+H*H%] - FPP [L*HH%].

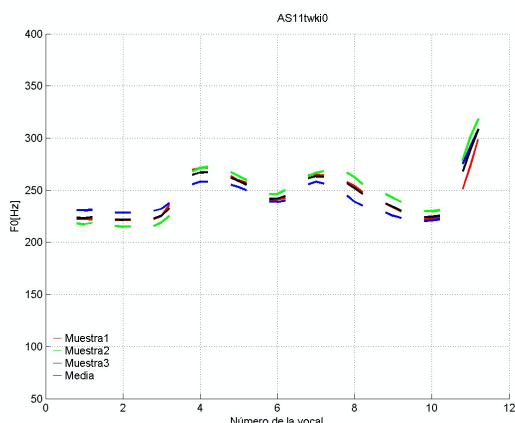


Figura 8: Interrogativas. Español angelino salvadoreño. *¿La guitarra se toca con obsesión?*
Fuente: Congosto Martín (2018).

4. RESULTADOS

Los análisis acústicos realizados demuestran que, más allá de las similitudes geoprosódicas de carácter general asociadas al patrón entonativo de las declarativas en español, existen rasgos entonativos propios de determinados ámbitos dialectales del español de México que perviven en el español mexicano de Los Ángeles, tanto en hablantes de primera como de segunda generación, que afectan tanto a la altura tonal como al desarrollo de la curva en su fase pretonemática (picos tonales) y tonemática, con tonos de frontera M% y H% presentes en diversas partes de México (Martín Butragueño y Mendoza 2018). Del mismo modo, se ha podido apreciar igualmente cómo la hablante guatemalteca de primera generación de LA produce un acento nuclear [L*L%] frente al contorno circunflejo [L+H*L%] que caracteriza a la hablante guatemalteca de Guatemala, acercándose así al comportamiento entonativo de las otras dos

comunidades de habla con las que convive, la mexicana y la salvadoreña.

En cuanto a las interrogativas absolutas, todas las curvas entonativas analizadas muestran un segundo pico tonal en el pretonema, independientemente de cual sea origen de la informante; perfil melódico también presente en otros puntos geográficos de Hispanoamérica: Mérida (México), Cuba, Medellín, Guayaquil, Lima, Santiago de Chile², Costa Rica (Congosto Martín 2009) y Bolivia (Congosto Martín 2007), por ejemplo, y del que no hay apenas constancia en el español de España.

5. CONCLUSIÓN

Todo lo expuesto permite entrever ciertos lazos de unidad versus diversidad desde el punto de vista entonativo (declarativas e interrogativas absolutas) entre el español de Los Ángeles y el español de México, Guatemala y El Salvador: por un lado, rasgos de continuidad geoprosódica entre el español mexicano de uno y otro lado de la frontera, con tonos más agudos y finales [M%] y [H%] en la entonación declarativa de ciertos hablantes; por otro, una tendencia unificadora materializada en la pérdida de ciertos rasgos propios de la entonación declarativa del español de Guatemala con finales [L+H*L%] (detectados igualmente en otras zonas de México) que han evolucionado hacia [L*L%], haciéndola coincidir con la configuración tonemática más generalizada en la zona. En cuanto a las interrogativas, todas las curvas melódicas presentan dos picos tonales en el pretonema, contorno también registrado en otros países hispanoamericanos, pero ajeno al español de España. En definitiva, procesos en curso de convergencia que, favorecidos por el contacto sociocultural y lingüístico, tienden a la unificación y homogenización entonativa del español de Los Ángeles, sin llegar a perderse del todo, evidentemente, ciertos rasgos identitarios.

6. AGRADECIMIENTOS

Esta investigación queda enmarcada dentro de dos proyectos dirigidos en calidad de Investigadora Principal por quien suscribe estas páginas dentro del Plan Nacional de I+D:

² Cfr. el mapa interactivo de Amper-español en http://stel3.ub.edu/labfon/amper/cast/amperespana_m_apas_interactivos.html

- Proyecto PROESLA. *Características prosódicas del español de los Estados Unidos, Los Ángeles, California*. (Ref. FFI2011-23292).
- Proyecto GEOPROSODIC. *Estudio geoprosódico y sociodialectal del español americano septentrional*. (Ref. FFI2015-68508)

7. BIBLIOGRAFÍA

- Congosto Martín, Y. (2007). Dialectología de la entonación. Interrogativas absolutas en el español de España (Sevilla, Badajoz) y en el español de América (Bolivia). En: *La prosodia en el ámbito lingüístico románico*, ed. J. Dorta, 389-415. Santa Cruz de Tenerife: La Página Ediciones.
- Congosto Martín, Y. (2009). L'Atlas Multimédia Prosodique de l'Espace Roman (AMPER) au Costa Rica, *Géolinguistique*, 11, 119-147.
- Congosto Martín, Y. (2011). Lenguas en contacto y prosodia. La entonación del español en hablantes bilingües mexicanoamericanos, *Lingüística Española Actual* XXXIV/1: 137-166.
- Congosto Martín, Y. (2012). Acento y entonación en el español hablado en los Estados Unidos (Los Ángeles, California). En: *La lengua, lugar de encuentro. Actas del XVI Congreso Internacional de la ALFAL*, eds. Ana M^a. Cestero Mancera, et al., 1907-1916. Alcalá de Henares, Universidad.
- Congosto Martín, Y. (2015). Entonación idiomática y fonética forense: la información dialectal y sociolingüística de la entonación en español, *Normas* (Revista de Estudios Lingüísticos Hispánicos) 5: 31-49.
- Congosto Martín, (2016). Un paso más en la configuración del mapa geoprosódico del español actual. En: *53 reflexiones sobre aspectos de la fonética y otros temas de lingüística*, ed. Ana M.^a Fernández Planas, 161-170. Barcelona: Laboratorio de Fonética de la Universidad de Barcelona.
- Congosto Martín, Y. (2017). Estudio geoprosódico y socio-dialectal del español de la comunidad lingüística chicana de Los Ángeles (CA)". En: *Estudios de variación geolingüística*, coords., Leonor Orozco y Alonso Guerrero, 169-199. Ciudad de México: Instituto Nacional de Antropología e Historia.
- Congosto Martín, Y. (2018). La prosodia en contextos urbanos multilingües: el caso del español de Los Ángeles. En: *Xè Workshop sobre la prosòdia del català*. Laboratori de Fonètica Aplicada UB. Crup de Recerca en Entonació i Parla (GREP).
- Congosto Martín, Y. (2019). Political vs. Linguistic Borders: The Spanish intonation of Mexicans and Mexican Americans living on either side of the border, *Spanish in Context*, 16/3, 390-418. <https://doi.org/10.1075/sic.00044.con>
- Congosto Martín, Y. (2020). El español guatemalteco en contacto: aspectos entonativos, *Estudios de Fonética Experimental*, 29, 153-194.
- Congosto Martín, Y. (2021 en prensa). Rasgos entonativos del español mexicano, guatemalteco y salvadoreño de Los Ángeles, California. En: *La entonación del español americano septentrional: estudios geoprosódicos*, eds. Y. Congosto, P. Martín Butragueño. Frankfurt am Main: Peter Lang Edition.
- Estebas Vilaplana, E, / Prieto Vives, P. (2008). La notación prosódica del español: una revisión del Sp_ToBI, *Estudios de Fonética Experimental* 17: 265-283.
- Estebas Vilaplana, E, / Prieto Vives, P. (2010). Castilian Spanish Intonation. En: *Transcription of Intonation of the Spanish Language*, ed. Pilar Prieto & Paolo Roseano, 17-48. Lincom GmbH.
- Fernández Planas, A. M. (2005). Aspectos generales acerca del Proyecto Internacional AMPER en España, *Estudios de Fonética Experimental* 14: 13-27.
- Ladd, D. R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.
- López Bobo, M. J., et al. (2007). Análisis y representación de la entonación. Replanteamiento metodológico en el marco del proyecto AMPER. En *La prosodia en el ámbito lingüístico románico*, eds. J. Dorta y B. Fernández, 17-34. Santa Cruz de Tenerife: La Página Ediciones.
- Martín Butragueño, P. & Mendoza Vázquez, É. (2018). Prosodic nuclear patterns in narrow and broad focus utterances: pragmatic and social factors in Central Mexican Spanish. En: *Focus Realization in Romance and Beyond*, eds. M. Uth y M. García, 131-172. Amsterdam - Philadelphia: John Benjamins.
- Pamies Bertrán, A., et. al. (2002). Umbrales tonales en español peninsular. En: *Actas del II Congreso Nacional de Fonética Experimental*, ed. J. Díaz García, 272-278. Sevilla, Universidad.
- Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. Tesis doctoral, MIT.
- Utgård, K. (2014). La investigación prosódica del español de Guatemala. Situación geográfica, social y lingüística. En: *Fonética experimental, educación superior e investigación*. III. *Prosodia*, eds. Congosto Martín, Y., et al. 377-400. Madrid, Arco/Libros, S.L.

POSITIONALLY CONSTRAINED ACCENTEDNESS CUES. A PILOT STUDY ON BARI ITALIAN

Giovina Angela del Rosso

Roma Tre University – Sapienza University of Rome
giovinaangela.delrosso@uniroma3.it

ABSTRACT

In the crowded vowel inventory of Bari Dialect (BD), quality alternations are governed by syllable type, stress position, and gender differences (Manzari, 2020). The vowel systems of BD and Bari Italian (BI) are in substantial contact (Stehl, 1990; Cerruti, 2011). Yet, in BI only [æ] has been identified as a dialect-oriented variant of the Italian /a/, appearing in open syllables (Canepari, 1983; Filipponio & Cazzorla, 2015). In this paper, the /a/ patterns of BI are furtherly investigated through the acoustic analysis of semi-spontaneous materials. The results highlight that not only palatalized but also velarized variants of /a/ occur in BI, in open and closed positions respectively, lawfully matching the [æ]-[ɐ] patterns of BD (Manzari, 2020). Hence, the substratum interference exerted by BD on BI results in positionally constrained accentedness cues that shape the /a/ targets (Marotta, 2014) and challenge the traditional classifications of the RI vowel systems (Schmid 1999).

Keywords: Bari Italian, accentedness, language contact, vowels, acoustic phonetics

1. INTRODUCTION

1.1. Aim

This paper investigates to what extent Bari Italian (BI) resembles the positionally constrained /a/ patterns of Bari Dialect (BD), due to language contact phenomena. Indeed, previous research has shown that Italian speakers tend to transfer features of their native local dialects into regional varieties. This results in shibboleth-like patterns cueing Regional Italians (RI) and conveying indexical information through accentedness (Berruto, 2005; Marotta, 2014).

While most studies on the Italian vowel systems focus on the status of /e, ε, o, ɔ/ (Calamai, 2003; Clemente et al., 2006), here the realizations of the /a/ targets are acoustically examined and discussed. Sampling a larger number of speakers and testing the perceptual effects of the /a/ patterns are the two main purposes for future analysis.

1.2. Terminological caveat

Preliminary terminological disambiguation is due, for the sake of clarity. Definitory overlaps may arise discussing socio-phonetic issues while borrowing concepts traditionally rooted in dialectology.

The term *dialect* calls for clarification since its broad acceptance may refer to both geographical and social varieties. In its broad meaning, “a dialect may predominate as the official or standard language”

(Crystal, 2008: 142). In Italian dialectology, however, dialects are traditionally considered in a narrow geographical sense, as local varieties, often genetically related to the standard variety (Beccaria, 2004: 229-230). Differently, the blanket term *variety* skates around the debated language-dialect dichotomy (Beccaria, 2004: 792-793), referring to “any system of linguistic expression whose use governed by variables” (Crystal, 2008: 509). In this paper, the narrow meaning of dialect will be used to refer to BD, whereas BI and RIs will be labeled as varieties, though historically they are primary and tertiary dialects (Berruto, 2005; Cerruti, 2011).

1.3. Repertoire

1.3.1. Regional Italian (RI)

RIs are spoken varieties that differ from Standard Italian (SI) especially in phonetics and phonology (Canepari, 1983; Bertinetto & Loporcaro, 2005). The standard norm does not exert much pressure on these language levels and, even in school, non-standard variants are tolerated, when non conflicting with distinctions conveyed in writing, regardless of the style, also due to the general unawareness of substratum interference and regional markedness (Cerruti, 2011: 20). Since there are no SI natives, RIs may be considered as originated from imperfect learning (Thomanson, 2003). Nonetheless, the interaction between context and individual variability

plays a crucial role in pronunciation. According to the repertoire, competence, attitude, and context, speakers exhibit a more standard- or dialect-oriented speech. The more dialect-oriented, the more accented RI will result (Marotta, 2014).

1.3.2. Bari dialect (BD)

Bari is the capital of Apulia, a south-eastern Italian region. In Bari, most speakers born before the '60s are native of BD and BI and were formally taught SI in school; younger speakers use primarily BI but retain at least passive competence and reduced active competence of BD (Stehl, 1990; Sobrero, 1993).

BD belongs to the south-eastern Italo-Romance varieties. Manzari (2020) describes the development of the BD vowel system, varying according to linguistic factors (stress position, syllable type) and sociolinguistic parameters (age, gender). Following Carosella (2005), Manzari (2020) argues the vowel quality is fundamentally conditioned by its position, as summarized in Tab. 1 for /a/.

Table 1: The historical change of A for male speakers according to open and closed positions.

position	stress	syllable	A	context
open	parox.	open	[ɛ, æ]	nasal
			[a, ɑ]	-
closed	proparox.	closed	[ɔ]	-
	oxytones	-		

1.3.3. Bari Italian (BI)

BI is the variety of Apulian RI spoken in Bari. The peculiar features of BI can be described in a contrastive perspective since they originate from the contact between BD and (taught) SI (Carosella, 2005; Golovko, 2012).

At the segmental level, BI is under-investigated. Accordingly, three types of stressed vowel systems are traditionally discerned among RIs:

- I. phonologically eptavocalic, /a, ɛ, e, i, ɔ, o, u/ with 7 phones [a, ɛ, e, i, ɔ, o, u];
- II. phonologically pentavocalic /a, e, i, o, u/ with the 7 phones [a, ɛ, e, i, ɔ, o, u];
- III. phonologically pentavocalic /a, e, i, o, u/ with the 5 phones (Calamai, 2008; Clemente et al., 2006; Schmid 1999).

It is noteworthy that the Apulian RI is not explicitly included in any of these three groups. Yet, according to Canepari (1983: 75-76): the variability of /a, i, u/ is low; the vowels may diphthongize in strongly accented speech; the unstable /ɛ, ɔ/~e, o/ contrasts can depend on the syllable type, on the speaker or undergo neutralization; /a/ may be raised in stressed open syllables, whereas the unstressed /a/ are raised and centralized.

Filipponio and Cazzorla (2015) conducted the first analysis devoted to the stressed vowels of BI. They claim that the phonological distinction among mid vowels has been suppressed BI. In addition, BI has [+tense] vowels in stressed open syllables (/i, e, a, o, u/), [-tense] in closed syllables (/i, ɛ, a, ɔ, u/) – but for the tenseness feature see, Ladefoged and Maddieson (1996: 302-306). Thus, the distribution of mid vowels can be predicted relying on the syllable type. They also found that /a/ palatalization affects F2 in open syllables (Filipponio & Cazzorla, 2015: 64-69).

According to these features and recalling the traditional taxonomy, the BI vowel system would be classified in type II: pentavocalic with 7 phones.

2. MATERIALS AND METHOD

Audio recordings in .wav format have been collected using a Tascam DR-40 portable recorder (sample rate: 44.1k Hz, 24 bit). The average duration of each recording is approximately 30 minutes.

The speech materials have been elicited via semi-guided interviews about three key issues for the citizens of Bari: sea pollution, green areas, and public transportation. These data collection technique and topics have been chosen for the following reasons: reducing the interviewer-interviewee distance; acquiring spontaneous speech in the digressions; indirectly inducing token repetitions, enhancing comparability without undermining spontaneity; incremental nature. However, the (expected) drawbacks are overlapping speech, suprasegmental variation, unbalanced data.

In all, 10 speakers were recorded. This paper focuses on the sociolinguistically most homogeneous subgroup, made of 4 male speakers (age: 54-60). They have always dwelled in Bari, are BD-Italian bilingual, stopped studying after high school, and belong to the working-class (employed at the local university in non-academic jobs).

The interviews have been orthographically transcribed. Subsequently, a database containing both comparable and hapax tokens has been created. Only content words entered the analysis. A sample of 40 words per speaker has been acoustically analyzed (160 tokens). The analyzed tokens have been manually segmented by means of waveforms and spectrograms in Praat (Boersma & Weenink, 2021). The results have been furtherly processed in R (R Core Team, 2021)

In this study the following variables have been taken into account: vowel target and realization; stress position; syllable type; position (stress position combined with syllable type, following Carosella, 2005 and Manzari, 2020); F1 and F2 values of the

stressed vowels, measured at the steady-state. The dataset is shown in Tab. 2.

Table 2: Pretonic, stressed and postonic items of /a/ in the dataset, according to syllable type and, for stressed tokens, positional criteria.

	all	open σ	closed σ	open p.	closed p.
total	142	88	54	NA	NA
pretonic	41	25	16	NA	NA
stressed	69	31	38	30	39
postonic	32	32	0	NA	NA

Thus, in the first part of the analysis the overall vowel system of BI has been inspected, to detect non-standard and dialectal-like features characterizing it. Afterward, the pattern constraining the occurrence of different realizations of /a/ was furtherly examined.

3. RESULTS

3.1. Overall segmental phonetics

In a contrastive perspective, only the elements characterizing BI will be commented, as usually in accounts about RI only features differing from SI are discussed. Relying on the positional criterion open vs closed, the mean F2-F1 values of stressed vowels have been plotted and are shown in Fig. 1.

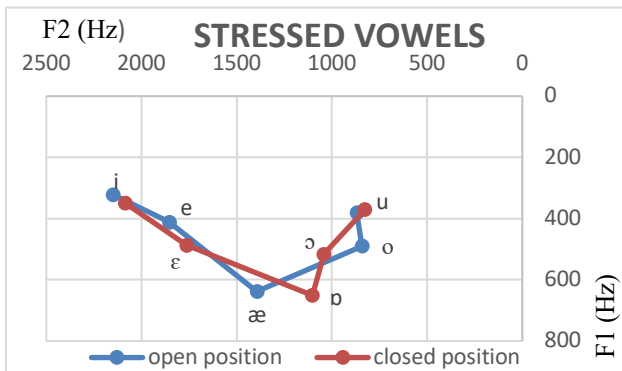


Figure 1: F2-F1 scatterplot of stressed vowel targets /a, e, i, o, u/ of BI, showing the variations related to the positional criterion open vs closed.

It can be observed that in the pentavocalic system of BI, especially mid and low vowels are sensitive to the position. The difference in the distributional pattern impacting /a/ realizations is larger than the difference in /e, o/. The realizations transcribed as [æ] occur in open positions, while those labeled as [ɔ] occur in closed positions. The allophonic processes regulating the [a] alternations are non-standard and, thus, have been furtherly investigated.

3.2. The positional constraint on /a/ variants

The stressed /a/ items have been reconsidered in the light of their auditory categorizations. Consequently, part of the items in open position has been categorized as palatalized ([æ]), part of the items in closed position have been judged as velarized ([ɔ]), and part of the items occurring in both positions have been considered as sufficiently standard-like [a]. Thus, the position cannot be regarded as a parameter blindly predicting which variant will occur. Instead, the vowel position seems to act as a constraint to predict if dialect-like variants, i.e. [æ] or [ɔ], can occur in a certain context.

In Tab. 3, an overview of the results is provided along with the mean formant frequencies (and related standard deviations) of /a/ realizations. The formant values for [a] realizations in both open and closed positions are similar. F1 values of both [æ] and [ɔ] are lower than F1 of [a], and F1 of [æ] is slightly lower than F1 of [a]. F2 of [æ] is similar to F2 of [a], while F2 of [ɔ] is considerably lower, consistently with its back and round articulation.

Table 3: Mean formant values (in Hz) of stressed /a/ targets, auditorily labeled as [æ], [a], or [ɔ].

stressed /a/	69 items	F1	F2
[æ]	15	618±63	1439±138
[a] (open p.)	14	688±80	1418±123
[a] (total)	20	680±93	1431±170
[a] (closed p.)	6	661±125	1462±262
[ɔ]	34	657±53	1060±101
stressed /a/	69 items	open p.	closed p.
stressed [æ]	15	15 (100%)	0
stressed [a]	20	14 (70%)	6 (30%)
stressed [ɔ]	34	2 (6%)	32 (94%)

Thus, a subtle difference in the height dimension has been found between [a] and [æ] in open positions, while a larger difference between [a] and [ɔ] has been found to operate, particularly in the horizontal axis.

Velarized realizations seem not only more detectable, but also more frequent than palatalized realizations since the instances of standard-like [a] have mostly been found in open positions.

4. DISCUSSION

BI is characterized by palatalized and velarized variants of /a/, whose realizations are constrained by the position, that results from the syllable type and stress position combined. However, the positional criterion itself does not imply raised and backed realizations of the vowel target. Instead, it acts more like a constraint: if the type and openness of the stressed syllable structurally support the allophonic

process of the model pattern of BD, then it may occur in BI and thus characterize it.

Table 4: Mean formant values (in Hz) for stressed mid and low vowels of Italian for male speakers. Data reported from other studies refer to eptavocalic systems, while the means for BI in this table refer to the positional criterion.

‘V	Cosi et al. 1995		Albano Leoni & Maturi 2002		Zmarich & Bonifacio 2003		Bari Italian	
	F1	F2	F1	F2	F1	F2	F1	F2
e	394	2082	375	2028	390	1951	411 o.	1855 o.
ɛ	513	1989	500	1844	562	1780	487 c.	1764 c.
a	742	1420	708	1466	715	1274	638 o.	1393 o.
ɒ	NA	NA	NA	NA	NA	NA	651 c.	1103 c.
ɔ	447	856	554	1055	518	847	516 c.	1042 c.
o	552	949	409	1001	382	774	490 o.	842 o.

Table 4 displays the average formant values found in the literature for mid and low vowels of Italian. F1 values of /a/ targets in both open and closed positions of BI are lower than those reported by Zmarich and Bonifacio (2003), Albano Leoni and Maturi (2002), and Cosi *et al.* (1995). Filipponio and Cazzorla (2015: 69) refer to F2 median values for /a/: 1632 Hz in open positions, 1174 Hz in closed positions. The median values they report are higher than the mean value found here. They comment on the palatalization as unexpected and do not mention velarization effects.

Regardless of the position, the low vowel target is higher in BI than in other varieties (lower F1). In closed positions, also F2 is lower: this is related to both backness and lip protrusion (Harrington, 2010). The difference in F2 is larger than in F1, as predicted by the quantal theory of speech (Stevens & Hanson, 2010). These variants resemble for quality and context the sounds of BD inventory (Manzari, in press). Since they are also alien to SI (Table 3), it can be argued that [æ] and especially [ɒ] and their constraint pattern characterize dialect-oriented BI. These features work as acoustic shibboleths. Their perceptual salience needs to be tested.

The emergence of variants of /a/ can be interpreted as a case of ongoing contact-induced language change in BI, caused by the substratum influence and consequent pattern replication of BD. This is a case of source language agentivity or interference because the source language pattern is stronger than the one in the recipient language (Golovko, 2012; Aalberse & Myusken, 2018), though BD is not the most prestigious variety in the linguistic repertoire of the community.

According to the linguistic parameters of the Solid Sociophonetic Model proposed by Marotta (2014), the process investigated has frequent shape, since it involves stressed /a/, and small size, since it affects only /a/, but no functional load and low phonological value, because the surface phonetic variation does not conflict (yet) with any phonological category, even if this pattern makes the distance between [ɒ] and [ɔ] decrease. In the current vowel systems, there are no merges, but the balance has changed.

Interlinguistically the [ɒ] variant carries two-fold markedness: its low back articulation and the introduction of a new contrast in the front-back dimension (Ladefoged & Maddieson, 1996).

When this positional process operates, the stressed vowel system of BI becomes more BD-like and is not anymore ascribable to type II systems. The fourth type of system emerges. Indeed, BI testifies a pentavocalic system with more than 7 phones.

5. CONCLUSIONS

In Bari Italian evidence of a non-standard pattern involving /a/ was found, cueing speech as dialectal-accented. Due to this substratum interference, the expanded inventory of Bari Italian does not fit in the current vowel system typology of Regional Italians. Thus, it is argued that the traditional taxonomies of RIs (Schmid, 1999; Calamai, 2008) ought to be reconsidered.

6. REFERENCES

- Aalberse, S., & Muysken, P. (2018). Bilingualism and Language Contact. In A. De Houwer & L. Ortega (Eds.), *The Cambridge Handbook of Bilingualism* (pp. 524–543). Cambridge University Press.
- Albano Leoni, F., & Maturi, P. (2002). *Manuale di fonetica*. Carocci.
- Beccaria, G. L. (Ed.). (2004). *Dizionario di linguistica e di filologia, metrica, retorica*. Einaudi.
- Berruto, G. (2005). Dialect/standard convergence, mixing, and models of language contact: The case of Italy. In P. Auer, F. Hinskens, & P. Kerswill (Eds.), *Dialect Change* (pp. 81–95). Cambridge University Press.
- Bertinetto, P. M., & Loporcaro, M. (2005). The sound pattern of Standard Italian, as compared with the varieties spoken in Florence, Milan and Rome. *Journal of the International Phonetic Association*, 35(2), 131–151.
- Boersma, P., & Weenink, D. (2021). *Praat: Doing phonetics by computer* (6.1.43) [Computer software].

- Calamai, S. (2003). Vocali d'Italia. Una prima rassegna. In P. Cosi, E. Magno Caldognetto, & A. Zamboni (Eds.), *Voce canto parlato. Studi in onore di Franco Ferrero* (pp. 49–58). Unipress.
- Calamai, S. (2008). *L'italiano: Suoni e forme*. Carocci.
- Canepari, L. (1983). *Italiano standard e pronunce regionali*. CLUEP.
- Carosella, M. (2005). Sistemi vocalici tonici nell'area garganica settentrionale fra tensioni diatopiche e dinamiche variazionali. Edizioni Nuova Cultura.
- Cerruti, M. (2011). Regional varieties of Italian in the linguistic repertoire. *International Journal of the Sociology of Language*, 210, 9–28.
- Cerruti, M., & Regis, R. (2014). Standardization patterns and dialect/standard convergence: A northwestern Italian perspective. *Language in Society*, 43(1), 83–111.
- Clemente, G., Savy, R., & Calamai, S. (2006). Sistemi vocalici in diatopia. *Analisi prosodica. Teorie, modelli e sistemi di annotazione*, 439–460.
- Cosi, P., Ferrero, F., & Vagges, K. (1995). Rappresentazioni acustiche e uditive delle vocali italiane. *Proceedings of the XXIII National Conference of the Italian Acoustic Association*, 151–156.
- Crystal, D. (2008). *A dictionary of linguistics and phonetics*. Blackwell Publishing Ltd.
- Dal Negro, S., & Vietti, A. (2011). Italian and Italo-Romance dialects. *International Journal of the Sociology of Language*, 2011(210).
- Filipponio, L., & Cazzarola, S. (2015). The vowels of Bari. A comparison between local dialect and regional Italian. *Il Farsi e Il Disfarsi Del Linguaggio. Acquisizione, Mutamento e Destutturazione Della Struttura Sonora Del Linguaggio*, 1, 59–71.
- Golovko, E. (2012). The Formation of Regional Italian as a Consequence of Language Contact. The Salentino Case. *Journal of Language Contact*, 5(1), 117–143.
- Harrington, J. (2010). Acoustic Phonetics. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The Handbook of Phonetic Sciences* (pp. 81–129). Blackwell Publishing.
- Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Blackwell Publishers.
- Manzari, G. (2020). Il vocalismo tonico del dialetto parlato nel borgo antico di Bari: Un esempio di differenziazione secondo il genere sessuale in una varietà apul-barese. *L'Italia Dialettale: Rivista Di Dialettologia Italiana*, 81, 371.
- Marotta, G. (2014). New parameters for the sociophonetic indexes: Evidence from the Tuscan varieties of Italian. In C. Celata & S. Calamai (Eds.), *Advances in sociophonetics* (pp. 137–168). John Benjamins Publishing Company.
- R Core Team. (2021). *R: A language and environment for statistical computing* (4.0.4) [Computer software]. R Foundation for Statistical Computing.
- Schmid, S. (1999). *Fonetica e fonologia dell'italiano*. Paravia.
- Sobrero, A. (Ed.). (1993). *Introduzione all'italiano contemporaneo*, 2. Laterza.
- Stehl, T. (1990). Il problema di un italiano regionale in Puglia. In M. A. Cortelazzo & A. Mioni (Eds.), *L'italiano regionale* (pp. 265–280). Bulzoni.
- Stevens, K. N., & Hansen, H. M. (2010). Articulatory-Acoustic Relations as the Basis of Distinctive Contrasts. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The Handbook of Phonetic Sciences* (pp. 424–453). Blackwell Publishing.
- Thomason, S. G. (2003). Contact as a Source of Language Change. In B. D. Joseph & R. D. Janda (Eds.), *The Handbook of Historical Linguistics* (pp. 687–712). Blackwell Publishing.
- Zmarich, C., & Bonifacio, S. (2003). Sui piani formantici acustici e uditivi delle vocali di infanti, bambini, e adulti maschi e femmine. In P. Cosi, E. Magno Caldognetto, & A. Zamboni (Eds.), *Voce, Canto, Parlato. Studi in onore di Franco Ferrero* (pp. 311–320). Unipress.

ALLOPHONES OF /tr/ AFTER CODA /s/ ASPIRATION IN WESTERN ANDALUSIAN SPANISH

María Del Saz

Universidad de Santiago de Chile
maria.delsaz@usach.cl

ABSTRACT

Western Andalusian Spanish coda /s/ aspiration is a well-known phenomenon, and there is vast research on its implications concerning /s/ + /t/ sequences in terms of post-aspiration and affrication. Nevertheless, little is known on its effects on /tr/ clusters. This study aims to shed light on this context, by analyzing /s/ + /tr/ sequences in terms of duration of the stop closure, VOT, and frication or vibration of /r/. Results indicate that there are four general outcomes: a) canonical /s/ + short-lag /t/ + vibration; b) long-lag VOT /t/ + vibration; c) long-lag /t/ + frication; and d) short-lag /t/ + frication. No significant effects of stress pattern were found; however, the front high vowel /i/ seems to trigger frication.

Keywords: Western Andalusian Spanish, /tr/ clusters, coda /s/ aspiration

1. INTRODUCTION

In Western Andalusian Spanish (WAS), coda /s/ is generally aspirated, as in many other varieties of Spanish. Nevertheless, unlike other varieties, an outstanding characteristic of WAS is that aspiration is realized as post-aspiration rather than pre-aspiration; therefore, it has certain effects on the following segments. Widely studied has been the effect of aspiration of coda /s/ on voiceless stops, which become post-aspirated [p^h, t^h, k^h] (Torreira, 2007), as in English. Furthermore, special attention has been paid to the context of aspiration before /t/, in which a new phonetic change is ongoing among the younger population (especially females); i.e. post-aspiration in this context becomes affrication [t^h] → [tʃ]. Although it is a common phenomenon in all vowel contexts, it seems to be particularly triggered by the high front vowel /i/ (Ruch, 2013). Nevertheless, to our knowledge, no study has explored the influence of coda /s/ aspiration on the following /tr/ cluster, such as in *estrella* (star). This study provides evidence of different allophones of /tr/ clusters in WAS after aspiration.

In a language such as English, voiceless stops present aspiration in onset position on stressed syllables. The aspiration in the stops causes the devoicing of the following approximants /l, r, j, w/. Therefore, a word like “tray” is pronounced as [t_r̥eɪ]. In this instance of [t^h] + [l̥], the stop undergoes a process of posteriorization, i.e. the alveolar stop becomes postalveolar by the approximant.

In our study, we explore the effects of WAS coda /s/ weakening on the following /tr/ cluster, while also considering the position of the stressed syllable with respect to the cluster and its vowel context.

2. METHODS

2.1. Participants

Thirty-two speakers of Western Andalusian Spanish (age range 19-33; Mage = 22.08) from different towns in Seville, Cadiz, Huelva, and Malaga took part in this study, carried out at the Universidad de Sevilla. Participants signed informed consent forms and filled out background questionnaires prior to the recordings.

2.2. Materials

For this study, we analyzed words that contained /s/+/tr/ sequences in a speeded task. Twelve words were included in the task (out of 90 words): four words with /tr/ clusters in pre-tonic position (*extranjero, estresado, estridente, astronómico*), four in tonic position (*extraño, estrella, industrial, destrozado*), and four in post-tonic position (*menestra, postre, industria, desastre*). A total of 366 items were analyzed.

2.3. Procedure

Recordings were carried out at the Phonetics Laboratory of the English Language Department at the Universidad de Sevilla (Spain). We employed a Zoom H6 hand-recorder to collect the data and an ASUS laptop to conduct the speeded task. We then used Praat (Boersma & Weenik, 2001) to segment the target words and to label the corresponding segments.

Target words were classified based on the /s/ realization, closure, VOT, and friction/vibration of /r/, by means of auditory classification, spectrograms, and waveforms. We took into account Hall, Hamann, & Zigys's (2006) classification of assimilation types: spiratization /t/ → /s/, affrication /t/ → [tʰ] or [tʰs], and posteriorization /t/ → [ɟ]. Additionally, context was considered, i.e., the stress pattern of the syllable that contained /tr/: pre-tonic, tonic, post-tonic, as well as the vowel that contained the nucleus of the target syllables.

The measurements obtained were analyzed by means of Generalized Linear Models (GLM).

3. RESULTS

3.1. Variants

From the data analyzed, we were able to find 4 main groups of variants¹: 1) canonical /s/ + short-lag /t/ + vibration (16.67%); 2) long-lag VOT /t/ + vibration (14.75%); 3) long-lag /t/ + friction (32.77%); and 4) short-lag /t/ + friction (35.79%).

An example of each category (waveform and spectrogram) is provided in the following Figures 1, 2, 3, and 4 (Praat script by Elvira García, 2018):

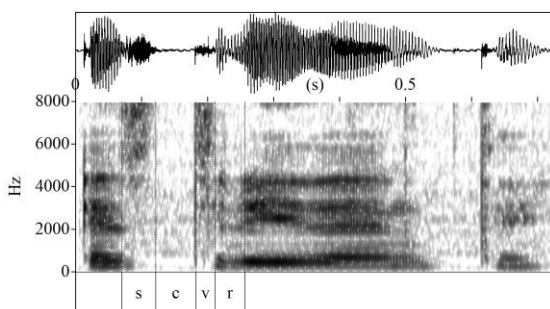


Figure 1: Canonical /s/ + /tr/ realization of the word *estridente* (strident)

¹ We also found 7 tokens of the canonical variant with friction of /r/ instead of vibration. We believe the target was variant 3, but the imposition of an orthographic <s> and the formality of the task prompted participants to produce the sibilant. Additionally, one further token was produced with friction of /r/ without stop occlusion.

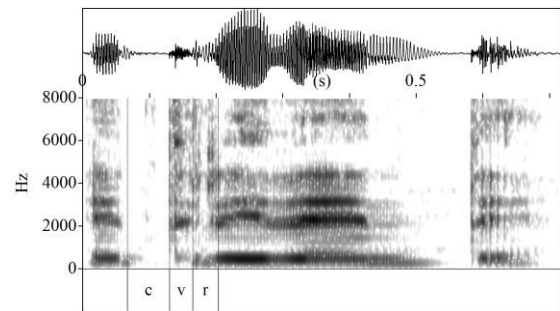


Figure 2: Affricated /s/ + /tr/ realization of the word *estridente* (long-lag /t/ + vibration)

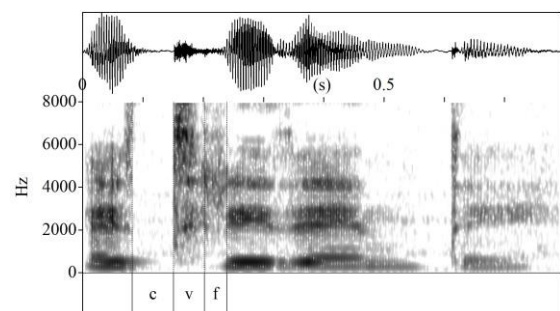


Figure 3: Affricated and posteriorized /s/ + /tr/ realization of the word *estridente* (long-lag /t/ + friction)

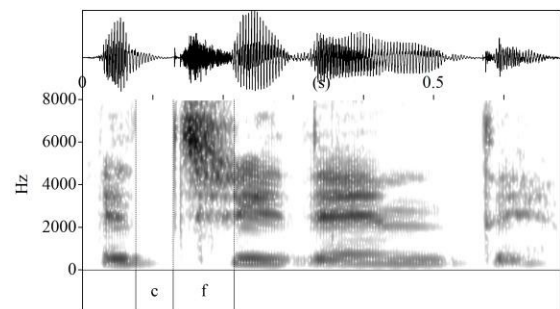


Figure 4: Posteriorized /s/ + /tr/ realization of the word *estridente* (short-lag /t/ + friction)

We then measured the duration of the stop closure in the four variants, the stop VOT in variants 1-3, the friction or vibration of /r/, and the duration of the whole /tr/ cluster in each case.

As we can see in Table 1, concerning the whole duration of the sequences, variants 2 and 3 were longer than variants 1 and 4 ($\beta = 148.9$; $SE = 1.65$; $z = 9.18$; $p < 0.001$). Variants 2 and 3 (diff: - 2.08; $SE = 4.80$; $z = - 0.43$; $p = 1.00$) and variants 1 and 4 (diff: - 11.45; $SE = 4.54$; $z = - 2.52$; $p = 0.07$) did not present statistically significant differences between them.

Table 1: Duration of closure, VOT, frication/vibration of /r/, /tr/ cluster, and whole sequence for the four variants of /s/ + /tr/

	type	M	SD	SE
closure	1	64.3	19.55	2.503
	2	76.1	21.24	2.891
	3	69.9	20.36	1.858
	4	59.6	22.02	1.924
VOT	1	26.4	7.09	0.908
	2	44.0	15.42	2.098
	3	49.1	13.54	1.236
	4	0.0	0.00	0.000
f/v	1	40.9	11.95	1.530
	2	39.4	10.75	1.463
	3	42.5	11.50	1.050
	4	83.4	15.36	1.342
/tr/	1	67.3	13.58	1.739
	2	83.4	19.50	2.653
	3	91.6	17.63	1.609
	4	83.4	15.36	1.342
total	1	131.6	22.86	2.927
	2	159.4	35.32	4.806
	3	161.5	30.47	2.782
	4	143.1	28.06	2.452

In terms of closure duration, variant 2 showed the longest duration, while variant 4 displayed the shortest ($\beta = 67.46$; $SE = 1.18$; $z = 57.04$; $p < 0.001$). Variants 1 and 3 show similar values (diff: -5.57 ; $SE = 3.30$; $z = -1.69$; $p = 0.55$). Concerning VOT duration, only variants 1-3 were considered. Variants 2 and 3, with long-lag VOTs, were longer than variant 1 ($\beta = 56.5$; $SE = 0.91$; $z = 61.85$; $p < 0.001$). Nevertheless, VOT values for variant 3 were longer than for variant 2 (diff: -5.09 ; $SE = 1.67$; $z = -3.06$; $p = 0.013$).

With respect to the overall duration of the /tr/ clusters, variant 3 is the longest one, while variant 1 is the shortest one ($\beta = 81.4$; $SE = 0.93$; $z = 87.38$; $p < 0.001$). Variants 2 and 4 show similar values (diff: 0.09 ; $SE = 2.67$; $z = -0.03$; $p = 1.00$). In terms of frication or vibration of the /r/ segment, the duration of the frication in variant 4 was significantly longer than in the other 3 variants ($\beta = 51.56$; $SE = 0.73$; $z = 7.037$; $p < 0.001$). Values were similar among the other three variants.

3.2. Context

Having determined the main variants of /str/, we subsequently analyzed the data according to their

position with respect to the stressed syllable (pre-

tonic, tonic, post-tonic). Figure 5 summarizes the occurrence of each variant for every syllable type.

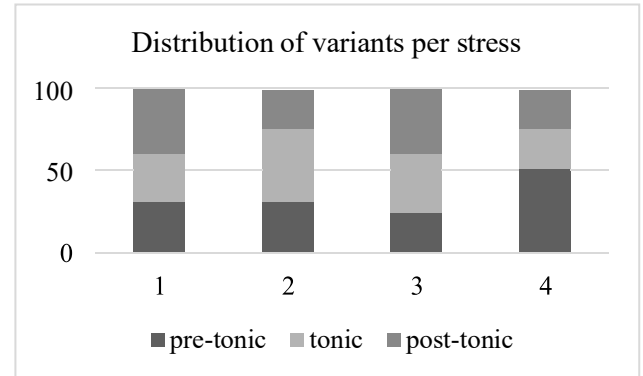


Figure 5: Distribution of variants according to stress patterns

The occurrence of variant 1 was evenly distributed among the three types of stress patterns; variant 2 was more frequently found in tonic position, variant 3 in post-tonic position, and variant 4 in pre-tonic position. Indeed, duration of closure, VOT, and whole cluster was observed in the following manner: post-tonic > tonic > pre-tonic. Nevertheless, the duration of friction/vibration was not dependent on stress pattern. Overall, there were no significant effect of stress among the four types of variants.

In terms of vowel context, Figure 6 shows the distribution of the variants across the contexts.

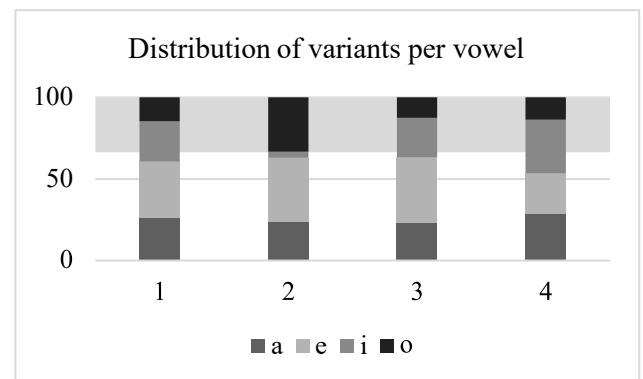


Figure 6: Distribution of variants according to vowel context

As shown, there is a high incidence of variants 1-3 before /e/, while much lower for variant 4. Variant 2 occurs before /o/ to a greater extent than the other variants, while minimally before /i/. For variant 4, /i/ is the context of higher occurrence. Indeed, frication seems to be significantly longer before /i/ than before the other vowel contexts ($\beta = 56.21$; SE

² In our target words, vowel /i/ and glide /j/.

= 1.27; $z = 44.34$; $p < 0.001$), which seems to be the trigger for this posteriorized variant of /s/ + /tr/ in WAS.

4. DISCUSSION

In light of these findings, we can say that there are four main realizations of /s/ + /tr/ sequences in Western Andalusian Spanish: a) canonical /s/ + short-lag /t/ + vibration; b) long-lag VOT /t/ + vibration; c) long-lag /t/ + frication; and d) short-lag /t/ + frication. There seems to be a tendency towards the frication of /r/, and a high number of posteriorized items. This phenomenon happens in English /tr/ clusters on stressed syllables, by which the aspiration in the stop devoices the approximant, while the postalveolar nature of [ɹ] retracts the stop to the same place of articulation, even to the point of affecting the place of articulation of initial /s/, from alveolar to postalveolar [ʃ], as in “street” [ʃtɹ.ɪt] (Gylfadottir, 2015).

The tendency towards affrication and posteriorization, particularly before front vowels, is also documented in other languages (Hall & Hamann, 2006). For example, in Japanese (Hume, 2018), dental /t/ is not only affricated, but also posteriorized before /i/, i.e. it becomes postalveolar /t/ → [tʃ]. In Brazilian Portuguese /t, d/ before /i/ also undergo this process (Barboza, 2015). Interestingly, another variety of Spanish that presents posteriorized allophones of /tr/ clusters is Chilean Spanish (Figueroa Candia, Soto-Barba, & Ñanculeo Raquileo, 2010), although these are not the result of coda /s/ aspiration. In this variety, a posteriorized allophone of /tr/ is present, i.e. long voiceless frication after the stop burst, with diverse ratios of closure and frication (even loss of closure) and sociolinguistic implications.

For Western Andalusian Spanish, the ongoing process of affrication of /t/ → [tʃ] after coda /s/ aspiration seems to have been triggered by /i/ as well (Ruch, 2013). Kim (2001) proposes that stop assibilation has “its phonetic origin in the brief period of turbulence that occurs at the release of a plosive into a high vocoid” (p. 81). In our study, while there is a higher number of instances of affrication and posteriorization before the high front vowel, these phenomena already seem to be spreading to all vowel contexts and stress patterns.

A possible venue for further research could consider stop + /tr/ sequences in Western Andalusian Spanish. For example, we found that both post-aspiration and affrication of /t/ could be emerging in stop + /tr/ sequences, such as in *exacto* (exact) (Del Saz, 2019), in which the coda stop is aspirated or affricated in the same manner as coda /s/. Therefore,

it is worth investigating whether affrication and posteriorization also take place in stop + /tr/ sequences, as in *eléctrico* (electric) or *actriz* (actress). Additionally, one can also analyze coda /s/ + voiceless stops /p, k/, as in *desprecio* (d disdain) and *escrito* (written), or even across word boundaries, as in *son las tres* (it’s 3 o’clock). Finally, sociolinguistic variables should be worth looking into to better understand the phenomenon.

5. CONCLUSIONS

This paper accounts for the effect of coda /s/ aspiration on /tr/ clusters in Western Andalusian Spanish. Four main variants have emerged, with a higher frequency of the frication of /r/ and posteriorization of the cluster. In line with other studies and languages, the front high vowel seems to trigger these phenomena. Further research on the implications of aspiration in emerging and ongoing processes is still needed.

6. REFERENCES

- Barboza, C. (2015). Palatalization in Brazilian Portuguese and its effects on the phonological learning of English as a foreign language. *Organon*, 30(58), 49-68.
- Boersma, P. & Weenik, D. (2001). Praat, a system for doing phonetics by computer. *Glott International* 5(9/10), 341-345.
- Del Saz, M. (2019). From postaspiration to affrication: New phonetic contexts in Western Andalusian Spanish. In Sasha Calhoun, Paola Escudero, Marija Tabain, & Paul Warren (eds.) *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019* (pp. 760-764). Canberra, Australia: Australasian Speech Science and Technology Association Inc.
- Elvira García, W. (2014). Zero-crossings-and-spectral-moments, v.1.3 [Praat script] (Retrieved from <http://stel.ub.edu/labfon/en/praat-scripts>).
- Elvira García, W. (2018). Create pictures with tiers v.4.5. [Praat script]. (Retrieved from <http://stel.ub.edu/labfon/en/praat-scripts>).
- Figueroa Candia, M., Soto-Barba, J., & Ñanculeo Raquileo, M. (2010). Los alófonos del grupo consonántico /t en el castellano de Chile. *Onomázein*, 22, 11-42.
- Gylfadottir, D. (2015). Shtreets of Philadelphia: An acoustic study of /str/-retraction in a naturalistic speech corpus. *University of Pennsylvania Working Papers in Linguistics*, 21(2), 87-97.
- Hall, T. A. & Hamann, S. (2006). Towards a typology of stop assibilation. *Sociology*
- Hall, T. A., Hamann, S., & Zygis, M. (2006). The phonetic motivation for phonological stop assibilation. *Journal of the International Phonetic Association*, 36(1), 59-81.
- Hume, E. V. (2018). *Front Vowels, Coronal Consonants and Their Interaction in Nonlinear Phonology*.

- Kim, H. (2001). A phonetically based account of phonological stop assibilation. *Phonetics in Phonology*, 18(1), 81-108.
- Ruch, H. (2013). Investigating a gradual metathesis: Phonetic and lexical factors on /s/- aspiration in Andalusian Spanish. *University of Pennsylvania Working Papers in Linguistics*, 19:(2), 169-180.
- Torreira, F. (2007). Coarticulation between aspirated-s and voiceless stops in Spanish: An interdialectal comparison. In N. Sagarra & A. J. Toribio (Eds.), *Selected Proceedings of the 9th Hispanic Linguistics Symposium*, (pp. 113–120). Somerville, MA: Cascadilla Press.

SENTENCE ACCENT IN ENGLISH AS A FOREIGN LANGUAGE: EXCEPTIONS TO THE LAST LEXICAL ITEM RULE

María Del Saz, Ximena Grau

Universidad de Santiago de Chile
maria.delsaz@usach.cl, ximena.grau@usach.cl

ABSTRACT

Research shows L2 segments are easier to acquire than suprasegmentals. In an SLA context, great amounts of input and early exposure seem to be needed for ultimate attainment of L2 features. In an EFL context, L1 influence seems to be minimized as learners gain L2 proficiency. Our study explores the effect of L1 transfer from Chilean Spanish and L2 proficiency in English on the production of sentence accent in utterances with exceptions to the Last Lexical Item rule, by comparing the productions of elementary and advanced learners of English in Chile. Results indicate improvement with L2 proficiency, although still far from native speakers. Nucleus placement improves and pitch range worsens with higher proficiency, with an overreliance on duration to mark the nucleus, to the detriment of pitch range. Further research on L2 English prosody is warranted for effective teaching and learning in EFL contexts.

Keywords: sentence accent, L1 Spanish, L2 English, acoustic parameters, prosody

1. INTRODUCTION

Research on ultimate attainment of the L2 phonetics and phonology in a SLA context suggests that L2 learners can achieve native-like proficiency provided some requirements are met. With a great amount of L2 input, high proficiency can be attained in terms of the acoustic and articulatory characteristics of L2 segments that are also present in the L1 but need some readjustment, and also in terms of pauses and reformulations. Only if exposed to the L2 at an early age can high proficiency be achieved concerning the acoustic and articulatory characteristics of L2 segments that are not present in the L1, and also concerning prosody (word stress, sentence accent, and intonation). Finally, fluency (articulation, speech rate) seems to be highly difficult regardless of age of L2 exposure (Saito, 2018).

More recent research focuses on prosody and intonation beyond the segmental level. So & Best (2010) attempt to adapt PAM to the perception of intonation in L2. As a consequence, in production L2 learners tend to generalize their L1 specific rules to their L2, giving way to miscommunication problems and even unintelligibility. Prosody is not only related to the rhythm of language, but is also connected to attitudes and emotion expressed through speech. In fact, the biggest problems

encountered when there is interference from the L1 with the L2 concern prosody and tone, followed by insufficient duration of vowels and slow speech rate (Wei, Idemaru, & Gubbins, 2015).

A stress-timed language like English presupposes a problem for speakers of syllable-timed languages like Spanish and mora-timed languages like Japanese. English rhythm implies that utterances are composed of a succession of strong and weak syllables. The vowels or diphthongs in strong syllables are realized in their full form, while the ones in weak syllables are reduced. This notion of strong/weak relates to one of the correlates of prominence, i.e. quality. The other three parameters that make a syllable prominent are pitch (sounds perceived as high or low), duration (sounds perceived as long or short), and loudness (sounds perceived as loud or soft). While they work in combination to make up prominence, the most important one is pitch.

Nuclear accents are the pitch movement initiators in an utterance. It was Halliday (1967) that introduced the concept of nuclear accent and defined the Last Lexical Item (LLI) rule, which states that unmarked utterances place their nucleus in the last lexical item (usually nouns, verbs, adjectives, and adverbs). In Spanish, the LLI rule is also the unmarked (default) rule, especially given that syntactic movements such as right dislocation of the

subject are possible –sometimes mandatory, as in the case of event sentences- in this language (Klassen, 2013; Ortiz, 1995, 2000; Véliz, 2001).

Nevertheless, with certain structures, despite being broad focus utterances, the LLI rule does not follow, i.e., the last lexical item is not the one that contains the nucleus. For example, in event sentences (subject + intransitive verbs), the sentence accent falls on the subject, as in “The BUS is coming”.

English word order can become an obstacle for Spanish learners of English, especially at the beginning stages of learning (Derwin & Munro, 2009; Major, 2001), precisely because of the aforementioned tendency in Spanish to accent the last lexical item. The exceptions to the LLI rule definitely imply an additional difficulty, even for more advanced learners (Finch & Ortiz Lira, 1982).

Concerning these exceptions, a previous study by Ortiz (1995) reports nucleus placement by L1 Chilean Spanish speakers in L2 English sentences with these types of structures. There are two issues with this piece of research: a) the author reported that speakers were recorded after receiving a training course in the topic; nevertheless, their level of L2 English competence is not specified; b) no acoustic or statistical analyses were realized or at least reported in the study. Therefore, our study aims at filling this gap by a) taking into consideration two groups of L1 Chilean Spanish speakers: first-year students of an English Pedagogy university program (elementary) and students who finished this program (advanced); and b) carrying out acoustic and statistical analyses of their utterances in both L2 English and L1 Spanish.

2. METHODS

2.1. Participants

For this study, we recorded 3 groups of participants: twenty L1 Chilean Spanish speakers (ages 18-20; 16 females and 4 males), first-year students in an English major from a state-funded university, with elementary level of English; twelve L1 Chilean Spanish speakers (ages 24-28; 8 females and 4 males) who graduated from the same program, with advanced level of English; and twelve L1 English speakers (ages 25-32; 8 females, 4 males), TEFL teachers in Chile. None of them reported speech or hearing disorders, and they all participated voluntarily in this study.

2.2. Materials

We employed 10 English sentences taken or adapted from Ortiz (2000), which comprise diverse

exceptions to the LLI rule, divided into two main groups: focus on the last noun and focus on other lexical item prior to the last lexical item. The first group was comprised of nouns (accented) followed by an adjective, intransitive verb, or infinitive (unaccented). The second group contained lexical items –not necessarily nouns- (accented) followed by adverbials or by objects of general reference (unaccented). Additionally, we also employed their 10 Spanish counterparts in the analysis. Sentences with final vocatives, softening phrases, and reporting clauses are said to present similar patterns in English and Spanish (Ortiz, 1995); thus, we did not consider these categories in this study (see Véliz, 2001, for further comments on these types of structures).

Table 1: List of English sentences

English sentences	
adverbial	There’s a FLY in my soup. It’s a bit CHILly in here.
general reference	It’s NICE to help people. You’re iMAG ining things.
intransitive verb	My BAT tery’s going to die. I wonder how long the VIDE o lasts.
infinitive	I’ve got a couple of BOOKS to read. I don’t know which BUS to take.
adjective	There’s a lot of MON ey involved. He left the DOOR open.

2.3. Procedure

Sentences were extracted and tagged using Praat (Boersma & Weenik, 2001), and subsequently normalized. Peak amplitude was scaled to 5000Hz for males and 5500Hz for females. Vowels were then labeled using TextGrid, and nuclear focus was marked (N). Nuclear accents in English sentences were confirmed by two experts and a native speaker of English; nucleus placement in Spanish was confirmed by two native speakers of Chilean Spanish and an expert. We then ran a Praat script (Elvira-García, 2014) to obtain the following information: overall pitch range of each utterance (st) and pitch range from the nuclear accent to the last point in each utterance (st), as well as duration (ms) and intensity (dB) of each nuclear accent. All data was analyzed statistically by means of generalized linear models (GLM). Our target was

to compare the elementary and the advanced groups of learners with native speakers of English, and also to compare the L2 English utterances with their productions in Spanish, to determine the influence of their L1 and their L2 proficiency level.

3. RESULTS

From the sentences analyzed, elementary-level participants placed the nuclear accent on the correct syllable in only 9% of the instances. On the other hand, advanced-level participants did so in 51.67% of their utterances.

Concerning the acoustic parameters of the nuclear accents, we first made a comparison between L1 English sentences and their L1 Spanish counterparts. As we can see in Table 2, English nuclear accents display longer duration [diff. = -24.0; SE = 3.91; $z = -6.13$; $p < .001$], higher intensity [diff. = -1.99; SE = 0.452; $z = -4.39$; $p < .001$], and wider pitch range from focus [diff. = 2.53; SE = 0.284; $z = 8.91$; $p < .001$] than Spanish. Conversely, Spanish shows a wider overall pitch range than English [diff. = -2.04; SE = 0.902; $z = -2.26$; $p = 0.024$].

Table 2: L1 English and L1 Spanish acoustic parameters.

	duration	intensity	max-min (N)	max-min
Eng	104.1	83.24	-5.42	-10.06
Spa	76.42	80.98	-2.7	-11.91

Upon comparing the L1 and L2 English utterances with correct focus and tone (Table 3), we see that both groups of L2 learners displayed similar overall pitch range to the native speakers, but lower intensity [elementary: diff. = -3.95; SE = 0.970; $z = -4.07$; $p < .001$; advanced: diff. = -2.63; SE = 0.578; $z = -4.55$; $p < .001$]. Interesting results were obtained concerning pitch range from nucleus and duration: While advanced learners' utterances displayed lower values for the first factor [diff. = 1.94; SE = 0.505; $z = 3.84$; $p < .001$] and higher for the second than the native speakers [diff. = 23.2; SE = 8.12; $z = 2.854$; $p = 0.013$], elementary learners showed similar values to those in L1 English.

Table 3: Acoustic parameters for L1 and L2 English (correct nucleus).

	duration	intensity	max-min (N)	max-min
nat	104.1	83.24	-5.42	-10.06
adv	124.2	80.78	-3.53	-11.24
elem	98.06	80.19	-6.38	-13

Finally, we compared the values for nuclear accents in L1 Spanish with those in L2 English that presented incorrect nucleus placement (i.e., nuclear accent on last content word), to understand the extent of L1 transfer other than the location of the nucleus. As shown in Tables 4 and 5, both groups of learners show similar overall pitch range and intensity in their English and L1 Spanish utterances. Nevertheless, pitch range from nucleus is significantly higher in their L1 Spanish than in their L2 English productions [elementary: diff. = -1.5890; SE = 0.188; $z = -8.432$; $p < .001$; advanced: diff. = -1.0198; SE = 0.297; $z = -3.431$; $p = 0.004$]. Likewise, duration is significantly longer in their L2 English utterances than in their L1 Spanish sentences [elementary: diff. = -44.15; SE = 4.56; $z = -9.69$; $p < .001$; advanced: diff. = -71.70; SE = 7.19; $z = -9.98$; $p < .001$].

Table 4: Acoustic parameters for L2 English (incorrect nucleus).

	duration	intensity	max-min (N)	max-min
elem	130.82	79.44	-1.43	-12.86
adv	135.86	81.08	-1.39	-10.48

Table 5: Acoustic parameters for L1 Spanish.

	duration	intensity	max-min (N)	max-min
elem	84.09	80.56	-2.95	-13.27
adv	65.06	81.6	-2.33	-9.89

It seems that in the production of L2 English sentences, L1 Spanish speakers transfer their L1 patterns into the L2 in terms of nucleus placement, at least in the initial stages of learning, when L1 influence is higher. However, despite incorrect placement, nuclear accents show higher duration than in their Spanish productions, although they fail to reach a higher pitch range from nucleus.

4. DISCUSSION

Concerning our findings in terms of prosody, we have seen that L2 English sentences that present exceptions to the last lexical item rule initially pose a problem for L1 Spanish learners of English in terms of nucleus placement, as Spanish tends to place the focus of the utterance on the last lexical item and can also resort to syntactic movement to draw attention onto a lexical item. With more competence in the L2, their productions start to become more accurate, as shown by the advanced learners of English.

In terms of the acoustic characteristics of nuclear accents (i.e., duration, intensity, pitch range), L1 English nuclear accents show longer duration, higher intensity, and wider pitch range from the nucleus than Spanish. Conversely, Spanish productions (at least Chilean Spanish) display an overall wider range throughout the whole utterance than English. In cases where L1 and L2 English utterances coincided in nucleus placement, elementary learners displayed similar parameters to native speakers. Advanced learners, however, differed in terms of longer duration of the nuclear accent and lower pitch range from nucleus than native speakers (Ramírez Verdugo, 2002). Although not measured, we speculate that advanced learners' higher fluency in the L2 –as attested by the experts and native speaker who assessed their productions- may cause a tradeoff with pitch range which is detrimental to their proficiency, and prompts them to resort to duration to signal pitch accent. Furthermore, comparing those L2 utterances that showed incorrect nucleus placement, i.e. on the last lexical item, with their L1 Spanish productions, both display similar pitch range from nucleus and intensity, while overall pitch range was higher in Spanish, and duration was higher in English. Here is where we can see that L1 Spanish learners certainly did not use pitch as an essential correlate of prominence in English; instead, they used duration. Indeed, several studies indicate that L1 Spanish learners of L2 English resort to duration as a cue to distinguish vowels in the target language despite not being an acoustic cue in their native language (Jiang, 2018) and, thus, not the result of L1 transfer but rather of psychoacoustic salience (Strange & Shafer, 2008). Concerning the acoustic correlates of nuclear pitch accents in our study, this also seems to be the case.

We have seen how ultimate attainment of prosody in SLA contexts seems to be linked to an early exposure to the language (Saito, 2018), and how suprasegmental errors hinder intelligibility much more than segmental errors (Rasier & Hiligsmann, 2007). Our findings could be applied to

improving teaching techniques, especially in a context of English as a foreign language (EFL) in which immersion is out of the picture, taking into account the caveat that suprasegmentals are less teachable and learnable than segmentals (Benrabah, 1997).

5. CONCLUSIONS

This paper shed some light on the production of L2 English nuclear accents in sentences with exceptions to the LLI. In terms of nucleus placement, the influence from L1 Spanish seems to be greater at the early stages of learning. However, L2 proficiency seems to be detrimental to the accurate production of the acoustic parameters of the nuclear accents, with a seemingly overreliance on duration instead of pitch range. Further research into prosody, fluency, and rhythm of L2 English seems to be necessary, particularly for its potential application in teaching and learning in EFL contexts.

6. REFERENCES

- Benrabah, M. (1997). Word stress – A source of unintelligibility in English. *International Review of Applied Linguistics*, 35(3), 157-165. <https://doi.org/10.1515/iral.1997.35.3.157>.
- Boersma, P. & Weenik, D. (2001). Praat, a system for doing phonetics by computer. *Glott International* 5(9/10), 341-345.
- Derwing, T. M. & Munro, M. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching*, 42, 476-490. doi:10.1017/S026144480800551X.
- Elvira-García, W. (2014). *Prosodic-data-extraction v2.1*. [Praat script] (Retrieved from <http://stel.uh.edu/labfon/en/praat-scripts>) Distributed under GNU General Public License.
- Finch, D. F. & Ortiz Lira, H. (1982). *A course in English phonetics for Spanish speakers*. London, UK: Heinemann Educational Books Ltd.
- Jiang, N. (2018). *Second language processing: An introduction*. New York & Abingdon: Routledge.
- Klassen, J. (2013). Second language acquisition of English focus prosody: Evidence from Spanish native speakers. In J. Cabrelli Amaro et al. (eds.), *Proceedings of the 12th Generative Approaches to Second Language Acquisition Conference* (pp. 76-84). Somerville, MA: Cascadilla Proceedings Project.
- Major, R. C. (2001). *Foreign accent: The ontogeny and phylogeny of second language phonology*. Mahwah, New Jersey: Lawrence Erlbaum Associates Publishers.
- Ortiz, H. (1995). Nucleus placement in English and Spanish: A pilot study of patterns of interference. In Lewis J. Windsor (ed.), *Studies in General and English Phonetics: Essays in Honor of Professor J. D. O'Connor* (pp. 255-265). London & New York: Routledge.

- Ortiz, H. (2000). *Word stress and sentence accent*. Santiago: UMCE Manuscript.
- Rasier, L. & Hiligsmann, P. (2007). Prosodic transfer from L1 to L2. Theoretical and methodological issues. *Nouveaux cahiers de linguistique française*, 28, 41-66.
- Ramírez Verdugo, D. (2002). Non-native interlanguage intonation systems: a study based on a computerized corpus of Spanish learners of English. *ICAME Journal* 26, 115-132.
- Saito, K. (2018). Advanced Second Language Segmental and Suprasegmental Acquisition. In Paul A. Malovrh & Alessandro G. Benatti (eds.), *The Handbook of Advanced Proficiency in Second Language Acquisition* (pp. 282-303). Hoboken, NJ: John Wiley & Sons, Inc. <https://doi.org/10.1002/9781119261650.ch15>.
- So, C. & Best, C. (2010). Cross-language perception of non-native tonal contrasts: effects of non-native phonological and phonetic influences. *Language and Speech* 53(2), 273 - 293. DOI: 10.1177/0023830909357156
- Strange, W. & Shafer, V. L. (2008). Speech perception in second language learners: The re-education of selective perception. *Phonology and second language acquisition*, 36, 153-192. <https://doi.org/10.1075/sibil.36.09str>.
- Véliz, M. (2001). A contrastive study of English and Spanish post-nuclear patterns. *Onomázein*, 6, 53-68.
- Wei, P., Idemaru, K., & Gubbins, L. (2015). Strong influence of prosody on the perception of foreign accent. In The Scottish Consortium for ICPHS 2015 (ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: The University of Glasgow. ISBN 978-0-85261-941-4. Paper number 185.1-4 retrieved from <http://www.internationalphoneticassociation.org/icphs/proceedings/ICPhS2015/Papers/ICPHS1041.pdf>

ESTUDIO ACÚSTICO Y DIALECTOMÉTRICO DE LA DURACIÓN VOCÁLICA: RELACIÓN DE PROXIMIDAD DEL OCCIDENTE, CENTRO Y ORIENTE CUBANO

Chaxiraxi Díaz, Josefa Dorta, Carolina Jorge

Universidad de La Laguna-Laboratorio de Fonética
chadiaz@ull.es, jdorta@ull.es, cjorgetr@ull.es

ABSTRACT

Este trabajo tiene como objetivo comparar la duración de tres puntos de encuesta cubanos. Se realizó el etiquetado de las vocales tónicas considerando el umbral perceptivo de 33.33% para tratar de reflejar las relaciones entre esta y sus adyacentes. Posteriormente, se intentó establecer relaciones de distancia y proximidad (ProDis 2) entre las diferentes zonas dialectales cubanas: esto nos permitirá comprobar la coherencia de los resultados obtenidos en el análisis acústico.

Palabras clave: duración, diferencias mínimas perceptibles, dialectometría, español cubano

This work has as an objective to compare the duration of the surveys done in three different locations in Cuba. In the first part of the work, the tonic vowels were tagged taking into consideration the perceptual threshold of 33,33%. This was done in an attempt to show the relationships between this and its adjacent (pretonic and posttonic). In the second part of the work, was made a dialectometric study (ProDis v2) between the different Cuban dialectal zones. The aim of this second part was to check the coherence of the obtained results in the acoustic analysis.

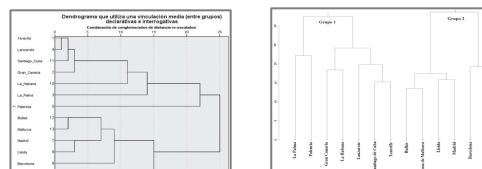
Keywords: duration, minimum perceptible differences, dialectometry, Cuban Spanish

1. INTRODUCCIÓN

En el marco de *El español en Cuba* (García Riverón, 1988) se establecieron tres zonas geolectales (Occidente, Centro y Oriente) y se señalaron pequeñas diferencias prosódicas regionales entre el occidente y el oriente de la isla. Posteriormente, la inclusión de esta variedad del español dentro de AMPER (*Atlas Multimedia Prosodique de l'Espace Roman*)¹ permitió que se impulsara el análisis de la entonación en las tres zonas de la isla (v. gr. Dorta ed., 2013; Dorta ed., 2018). Estos trabajos de tipo descriptivo o comparativo con otras variedades del español, como la venezolana o la canaria, evidencian que, en relación con el parámetro F0, las declarativas mantienen el patrón más general en español, pero

las interrogativas se caracterizan por compartir un mismo patrón final alto-descendente o circunflejo.

En cuanto al análisis dialectométrico de los datos de Cuba y la proximidad o distancia con otras variedades del español, ha sido objeto de estudio en algunos trabajos (v. gr. Fernández Planas *et al.*, 2015). Los resultados evidenciaron que las interrogativas (al contrario que las declarativas) permiten clasificar los datos en dos grupos de conglomerados o *clusters*.



¹ AMPER nace en el año 2002 en el Centre de Dialectologie de l'Université Stendhal-Grenoble III (Francia) con el propósito de realizar un atlas multimedia en el que plasmar la prosodia del espacio románico. AMPER fue coordinado por Michel Contini (Université

Stendhal-Grenoble III) y por Antonio Romano (Università di Torino) hasta principios de 2015. Este último es su coordinador en la actualidad.

Figura 1-2: Dendrogramas de declarativas e interrogativas conjuntamente obtenido en CalcuDista (SPSS) y en DiaTech (Fernández Planas *et al.*, 2015), respectivamente.

En Dorta *et al.*, 2018 se volvió a analizar el español cubano (La Habana, Santa Clara y Santiago de Cuba) con otras zonas afines del español como Canarias (Tenerife, El Hierro, La Gomera, La Palma, Gran Canaria, Fuerteventura y Lanzarote), Venezuela (Bolívar, Mérida, Barinas, Zulia y Aragua), Colombia (Medellín, Cali, Barranquilla, Bogotá) y San Antonio de Texas, esta vez con la herramienta ProDis v1. Los resultados mostraron que los diferentes patrones delimitados en el análisis acústico (alto-descendente o circunflejo, descendente y ascendente) son índices muy potentes para la clasificación en grupos coherentes o esperables según los resultados del análisis acústico efectuado (figura 2).

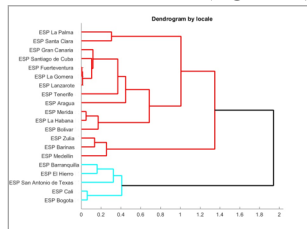


Figura 3: Dendrograma de declarativas vs. interrogativas por zonas (eje vertical) y valores de distancia (disimilitud) en el eje horizontal (Dorta *et al.*, 2018).

Por último, el parámetro duración ha sido abordado en habla femenina (Dorta *et al.*, 2013), considerando el umbral del 36% definido para el español (Pamies Bertrán y Fernández Planas, 2006). En este estudio, al considerar la relación entre la tónica y las adyacentes, se observó que la duración de la acentuada se veía afectada por su mayor o menor proximidad respecto las fronteras sintagmáticas. De este modo, la tónica de los esdrújulos, al estar más distanciada de la posición final del sintagma, era más corta que la de llanos y agudos. Tal diferencia se incrementaba en el SPrep puesto que la proximidad con la frontera final determinaba que la tónica de los tres acentos fuera más larga que en el sintagma inicial, aunque el porcentaje diferencial respecto de este no superaba siempre el umbral perceptivo.

2. OBJETIVO

Este trabajo tiene como objetivo comparar la duración de tres puntos de encuesta cubanos (La Habana, Santa Clara y Santiago de Cuba) desde dos perspectivas a partir de los datos acústicos obtenidos con AMPER2006 (Brezmes Alonso, 2007).

- Etiquetaje: para tratar de reflejar las relaciones entre la tónica y sus adyacentes

(pretónica y postónica) en los distintos puntos de encuesta siguiendo la propuesta de Muñetón Ayala, Díaz y Dorta (2018).

- Dialectométrica: para establecer las relaciones de distancia y proximidad entre las diferentes zonas dialectales cubanas y comprobar la coherencia de los resultados obtenidos en el análisis acústico.

3. METODOLOGÍA

3.1. Informantes

En el presente estudio hemos elegido seis informantes de Cuba, con edades comprendidas entre los 25 y los 60 años, sin estudios superiores que viven en zonas urbanas de los puntos de encuesta reflejados en la figura 4. Para la elección de los puntos se ha seguido la propuesta de división dialectal de García Riverón (1988).

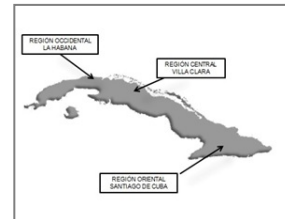


Figura 4: Puntos de encuesta.

3.2. Corpus

El corpus de análisis es el denominado *formal sin expansión en los sintagmas de frontera*, emitidas mediante elicitación textual, del tipo SN + SV + SPrep como, por ejemplo: *el saxofón se toca con obsesión o la cítara se toca con pánico*. Como puede verse en los ejemplos, el núcleo de los sintagmas de frontera lo forma una palabra trisílaba de diferente tipología acentual. Estas combinaciones dan como resultado 9 oraciones declarativas y otras tantas interrogativas absolutas que, a su vez, han sido repetidas tres veces, lo que supone 324 oraciones teniendo en cuenta el conjunto de los seis informantes considerados en este estudio.

3.3. Extracción de los datos

3.3.1. Grabación

Las grabaciones se realizaron en el lugar de procedencia de los informantes con una grabadora portátil Zoom H4n. Las señales fueron optimizadas con una frecuencia de muestreo de 16 Khz (GoldWave Digital Audio Editor versión 4.25).

3.3.2. Análisis acústico

El análisis acústico se hizo con rutinas desarrolladas en Matlab por la Universidad de Oviedo (Brezmes Alonso, 2007)² a partir de las originales creadas por Antonio Romano (1995). El análisis acústico tiene diferentes fases y cada frase es pasada por una serie de subrutinas. El punto de partida es el oscilograma de la frase en el que se hace una segmentación de las vocales lo que nos permite obtener una media de sus valores de duración.

3.4. Análisis dialectométrico

Los datos acústicos han sido sometidos al análisis dialectométrico con la herramienta ProDis v2 diseñada por el Laboratorio de Fonética de Barcelona (Elvira García *et al.* 2018). Según los autores, ProDis (Fernández Planas *et al.*, 2019) se inspira en Stat-Distances (Romano y Miotti, 2008; Rilliard y Lai, 2008 y Romano *et al.*, 2011), funciona en el entorno MatLab y calcula la media y la mediana de correlación (Pearson) de informantes y área de encuesta. A partir de estos datos hace un análisis de clúster y un análisis MDS que permite clasificarlos en distintos grupos según su semejanza. Este análisis se llevó a cabo por separado en las oraciones declarativas, interrogativas y conjuntamente atendiendo a zona e informante.

3.5. Umbral perceptivo y etiquetaje

Nos centramos en los valores medios de duración de la pretónica, tónica y postónica (pos-postónica, además, en los acentos esdrújulos) para establecer, en primer lugar, las diferencias mínimas perceptibles (DMP) a partir del umbral diferencial de 33,33% (Fernández Planas y Martínez Celdrán, 2003)³. A continuación siguiendo la propuesta de Muñetón Ayala, Díaz y Dorta (2018) en el que se contrasta largas (L) y breves (B) se marca la tónica con un asterisco (*) con el fin de poder diferenciarla de las vocales contiguas. Esta base permitirá generar las estructuras de duración a partir de dos contrastes pretónica-tónica (Pre-T) y tónica-postónica (T-Pos) o pos-postónica (T-Pos-Pos solo en acentos esdrújulos). La definición y posición de cada estructura en la citada propuesta es la siguiente.

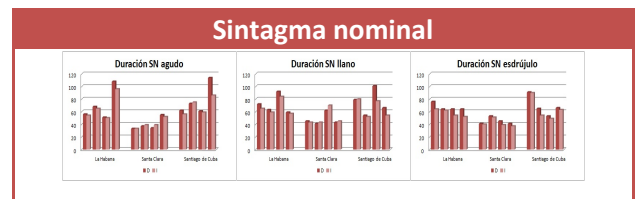
Tabla 1: Estructuras de duración según la propuesta de Muñetón Ayala, Díaz y Dorta (2018).

Estructuras de duración	
L*	la tónica es significativamente más larga que la pretónica y postónica
B*	no hay diferencias significativas entre la pretónica-tónica, ni entre tónica-postónica
BL*	solo hay diferencias entre pretónica-tónica, siendo la tónica la más larga
L*B	solo hay diferencias entre tónica-postónica, siendo la tónica la más larga
LB*	solo hay diferencias entre la pretónica-tónica, siendo la pretónica la más larga
B*L:	solo hay diferencias entre tónica-postónica, siendo la postónica la más larga
BL*L	la tónica es más larga que la pretónica y a su vez la postónica es más larga que la tónica

4. RESULTADOS

4.1. Descripción acústica

De manera general la relación entre la duración y el acento deja ver un comportamiento temporal más regular en el sintagma inicial con núcleo agudo y llano de las tres zonas cubanas y en las dos modalidades analizadas, teniendo en cuenta que la tónica siempre destaca de las vocales contiguas. En los SN esdrújulos, salvo en Santa Clara que sigue la tendencia de los sintagmas anteriores, no se observa una única tendencia temporal. En las figuras 5-7 se puede ver la media de las vocales correspondientes a la voz femenina de las tres zonas.



Figuras 5-7: Medias de duración del SN en voz femenina.

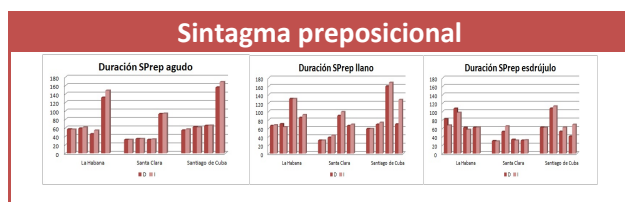
En el SPrep de declarativas e interrogativas, la tónica destaca claramente de las adyacentes pretónica y postónica, sobre todo si se encuentra próxima a la posición final, es decir, la duración suele ser mayor si el sintagma es agudo, seguido de los llanos y, en último lugar, los esdrújulos.

Para ello, se mide, en ms, en la pretónica, tónica y postónica y, posteriormente se etiquetan las vocales silábicas como G o P tras aplicar el umbral de un tercio de diferencia basado en el resultado obtenido por Pamies Bertrán y Fernández Planas (2006) que situaba el umbral diferencial en 36%.

² La licencia de Matlab en el Laboratorio de Fonética de la ULL es la nº 256105. Ver el replanteamiento de las nuevas rutinas en López Bobo *et al.*

³ En Fernández Planas y Martínez Celdrán (2003) se investiga la relación de la duración y el tono fundamental.

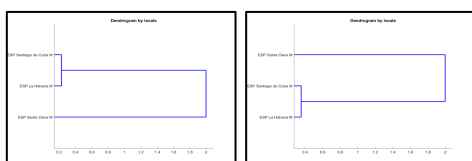
A modo de ejemplo, mostramos a continuación la media de la voz femenina en el sintagma final.



Figuras 8-10: Medias de duración del SPrep en voz femenina.

4.2. Descripción dialectométrica

En relación con el análisis dialectométrico, el dendograma de la figura 11 para la modalidad declarativa señala dos grupos prosódicos claramente diferenciados: el formado por Santa Clara y el formado por La Habana y Santiago. El dendograma de la figura 12 vuelve a distanciar en las interrogativas a Santa Clara de los otros dos puntos de encuesta que, como en las declarativas, aparecen muy próximos entre sí.



Figuras 11-12: Dendrogramas de los puntos de encuesta a partir de los datos de duración de las declarativas (izquierda) e interrogativa (derecha).

5. CONCLUSIONES

En este trabajo hemos estudiado el parámetro duración en tres zonas de la variedad de español de Cuba: La Habana, en el occidente de la isla; Santa Clara, en el centro y Santiago de Cuba, en el oriente. En la primera parte, el etiquetado de las vocales tónicas considerando diferencias iguales o superiores al 33,33% entre esta y las contiguas nos ha permitido ver el comportamiento de este parámetro temporal en las diferentes zonas; en la segunda parte, el análisis dialectométrico con el programa ProDis señala dos grupos prosódicos claramente diferenciados en las dos modalidades oracionales. En el trabajo final podremos corroborar si efectivamente se podrían considerar estos dos grupos y si la separación se relaciona con el etiquetaje efectuado.

6. BIBLIOGRAFÍA

- Brezmes Alonso, D. (2007). Desarrollo de una aplicación software para el análisis de características fundamentales de la voz. (Proyecto de fin de carrera). Universidad de Oviedo. Oviedo.
- Dorta, J. (ed.) (2013). *Estudio comparativo preliminar de la entonación de Canarias, Cuba y Venezuela*. Madrid–Santa Cruz de Tenerife: La Página ediciones S/L, Colección Universidad.
- Dorta, J. (ed.) (2018). La entonación declarativa e interrogativa en cinco zonas fronterizas del español: Canarias, Cuba, Venezuela, Colombia y San Antonio de Texas. Peter Lang Edition. STUDIEN ZUR ROMANISCHEN SPRACHWISSENSCHAFT UND INTERKULTURELLEN KOMMUNIKATION. Herausgegeben von Gerd Wotjak.
- Elvira García, W., Balocco, S., Roseano, P., y Fernández-Planas, A. M. (2018). ProDis: A dialectometric tool for acoustic prosodic data. *Speech Communication*, 97, 9-18.
- Fernández Planas, A. M., Dorta, J., Roseano, P., Díaz, C., Elvira García, W., Martín Gómez, J. A., y Martínez Celadrán, E. (2015). Distancia y proximidad prosódica entre algunas variedades del español: un estudio dialectométrico a partir de datos acústicos. *RLA. Revista de Lingüística Teórica y Aplicada*, 53 (2), 13-45.
- Fernández Planas, A. M., Elvira García, W., Roseano, P., y Balocco, S. (2019). Análisis dialectométrico con ProDis: Un paso más en los estudios prosódicos de AMPER. En J. Dorta (ed.), *Investigación geoprosódica. Amper: análisis y retos* (pp. 119-135). Madrid/Frankfurt am Main: Iberoamericana Vervuert.
- Fernández Planas, A. M., y Martínez Celadrán, E. (2003). El tono fundamental y la duración: dos aspectos de la taxonomía prosódica en dos modalidades de habla (enunciativa e interrogativa) del español. *Estudios de Fonética Experimental*, 12, 166-200.
- García Riverón, R. (1988). *Atlas Lingüístico de Cuba. Cuestionario*. La Habana: Academia de Ciencias de Cuba.
- López, M., Muñoz, C., Díaz, L., Corral, N., Brezmes, D., y Alvarellos, M. (2007). Análisis y representación de la entonación. Replanteamiento metodológico en el marco del proyecto AMPER. En J. Dorta (ed.), *La*

prosodia en el ámbito lingüístico románico (pp. 17–34). Madrid–Santa Cruz de Tenerife: La Página ediciones S/L, Colección Universidad.

- Muñetón Ayala, M., Díaz, C., y Dorta, J. (2018). La duración en oraciones sin expansión en la voz femenina de dos países fronterizos: Colombia (Bogotá -Medellín) y Venezuela (Caracas-Mérida)”, *Literatura y Lingüística*, 37, 401-423.
- Pamies Bertrán, A., y Fernández Planas, A. M^a. (2006). La percepción de la duración vocálica en español. En J. D. Luque Durán (Ed.), *Actas del V Congreso Andaluz de Lingüística General. Homenaje al Profesor José Andrés de Molina Redondo I* (pp. 501-512). Granada: Lingüística-Ediciones Método.
- Rilliard, A., y Lai, J. P. (2008). Outils pour le calcul et la comparaison prosodique dans le cadre du projet AMPER, l'exemple des variétés Occitane et Sarde. En A. Turculeț (org.), *La variation diatopique de l'intonation dans le domaine roumain et roman* (pp. 217-229). Iași: Editura Universității Al. I. Cuza.
- Romano, A. (1995). Développement d'un environnement de travail pour l'étude des structures sonores et intonatives de la parole. (Mémoire de DEA En Sciences Du Langage). ICP. Univ. Stendhal.
- Roseano, P., Elvira García, W., y Fernández Planas, A. M. (2017). Calcu-Dista: A Tool for Dialectometric Analysis of Intonational Variation. *Revista de Lingüística Teórica y Aplicada*, 55(2), 63-86.
- Romano, A., y Contini, M., Lai, J. P., y Rilliard, A. (2011). Distancias prosódicas entre variedades románicas en el marco del proyecto AMPER. *Revista internacional de lingüística iberoamericana*, 17, 13-26.
- Romano, A., y Miotti, R. (2008). Distanze prosodiche tra varietà friulane, romene e ispaniche. En A. Turculeț (org.), *La variation diatopique de l'intonation dans le domaine roumain et roman* (pp. 231-249). Iași: Editura Universității Al. I. Cuza.

LA PROSODIA DE ACTOS DEL HABLA EN EL ESPAÑOL DE ASUNCIÓN, BUENOS AIRES Y CIUDAD DE MÉXICO

Natalia dos Santos Figueiredo, Carolina Gomes da Silva

UNILA (Brasil), UFPB (Brasil)
natalia.figueiredo@unila.edu.br, carolinagsufpb@gmail.com

ABSTRACT

El objetivo de este trabajo es presentar la descripción de la estructura entonativa de 3 actos de habla: pregunta, pedido, y súplica, producidos en un *corpus* de habla experimental, en tres variedades del español: Asunción, Buenos Aires y Ciudad de México, considerando la teoría de la Fonología Entonacional (LADD, 2008; SOSA, 1999; PRIETO, 2003; ESTEBAS-VILAPLANA & PRIETO, 2009; PRIETO & ROSEANO, 2010).

Para tal, se grabó 117 enunciados en voz y video de locutores, del sexo femenino, como del masculino, considerando los contextos: “Apagar la tele”, para la variedad mexicana y “Cerrar la puerta”, para las variedades asuncena y porteña, en los 3 actos de habla mencionados, A través del análisis multimodal (Moraes y Rilliard, 2014), describimos los contornos de F0, los valores de duración y los movimientos gestuales. Los resultados revelan que sólo la entonación no es suficiente para contrastar los actos de habla y las variedades del español.

Palabras clave: prosodia, Análisis Multimodal, Actos de Habla

The objective of this work is to present the description of the intonative structure of 3 speech acts: question, request, and supplication, produced in an experimental speech corpus, in three varieties of Spanish: Asunción, Buenos Aires and Ciudad de México, considering the theory of the Intonational Phonology (LADD, 2008; SOSA, 1999; PRIETO, 2003; ESTEBAS-VILAPLANA & PRIETO, 2009; PRIETO & ROSEANO, 2010).

For this, 117 statements were recorded in voice and video of broadcasters, of the women, as of the men, considering the contexts: “Apagar la tele”, for the Mexican variety and “Cerrar la puerta”, for the Asunción and Buenos Aires varieties, In the 3 speech acts mentioned, Through multimodal analysis (Moraes and Rilliard, 2014), we describe the contours of F0, duration values and gestural movements. The results reveal that only intonation is not enough to contrast speech acts and varieties of Spanish.

Keywords: prosody, Multimodality, Speech Acts

1. INTRODUCCIÓN

Los actos de habla directivos corresponden pragmáticamente al intento del hablante de indicar al oyente una acción por realizarse. Este es abordaje lingüístico reciente de la teoría de los actos de habla, sin embargo, lo encontrábamos más comunmente dentro del campo de la filosofía del lenguaje.

En su investigación, Moraes et al. (2010) y Moraes y Rilliard (2014), analizan un repertorio de seis actos de habla directivos para el portugués brasileño, marcados por actitudes proposicionales, realizados en los modos declarativo, interrogativo e imperativo: *desafío*, *pedido*, *sugerencia*, *súplica* y *consejo*; y un repertorio de seis actos de habla directivos marcados

por actitudes sociales y realizados siempre en modo imperativo: *arrogancia*, *autoridad*, *desprecio*, *enojo*, *cortesía* y *seducción*. Los resultados demuestran que las actitudes proposicionales, consideradas más lingüísticas, se reconocen mejor auditivamente. En cambio, las actitudes sociales, consideradas menos lingüísticas, se reconocen mejor visualmente.

Con base en esos datos, este trabajo describe la estructura entonativa de los actos de habla en tres variedades del español de América Latina: Asunción, Buenos Aires y Ciudad de México. Para ellos, trabajamos con dos actos de habla directivos: *pedido* y *súplica*, y uno considerado “neutro”, *pregunta*.

Nos preguntamos si la entonación constituye un rasgo distintivo entre los actos de habla directivos y

entre las tres variedades del español, y cuáles serían esos rasgos. Además, desde una perspectiva multimodal, buscamos saber cuáles son los elementos que interactúan con la prosodia y componen el repertorio de pistas visuales más frecuentes en cada acto de habla.

En síntesis, los objetivos de este trabajo son:

- Describir, fonética y fonológicamente, las curvas entonativas de los actos de habla directivos;
- Contrastar los rasgos prosódicos entre los actos de habla directivos en las variedades de Asunción, Buenos Aires y Ciudad de México;
- Describir los elementos que actúan en la prosodia auditiva y componen la prosodia visual.

2. METODOLOGÍA DE ANÁLISIS

En esta etapa describimos el proceso de recolección y análisis de los datos desarrollados para cada nivel de análisis experimental, es decir, las precisiones metodológicas, acústicas y visuales.

2.1. Producción

La recolección de los datos se realizó en audio y video, a partir de la interacción de las investigadoras con los participantes de las tres ciudades investigadas: Asunción, Buenos Aires y Ciudad de México. En total, fueron grabados de 4 a 5 locutores en cada ciudad, considerando tanto hombres como mujeres, conforme explicitado en la tabla 1.

Tabla 1: Distribución de los actos de habla por ciudad.

Actos de Habla	Asunción	Buenos Aires	Ciudad de México
Pregunta	12	12	15
Pedido	12	12	15
Súplica	12	12	15
Total	36	36	45

2.1.1. Estímulos

Los contextos producidos por los participantes corresponden a representación de 3 actos de habla: una pregunta total, que genera una respuesta “Sí” o “no” (*yes/no question*), un pedido cortés, y una súplica. Trabajamos con dos situaciones de interacción para cada acto de habla:

- Apagar la tele (para los enunciados de Ciudad de México)

- Cerrar la puerta (para los enunciados de Asunción y Buenos Aires)

Todos los contextos están pensados para tratamiento lineal, es decir, dirigidos a personas próximas en su convivencia social, conforme instrucción a seguir:

Interacción: “Cerrar la puerta”

INVESTIGADOR

Situación Querés saber qué pasó, si María cerró la puerta.

Contexto: Están en la facultad María, Pedro y vos. Al final de la clase María se encargó de cerrar la puerta de la clase. Como no la viste cerrando la puerta, le preguntas a Pedro:

INFORMANTE

Pregunta (enunciado meta) - ¿Cerró la puerta?
(FIGUEIREDO, 2018)

Situación 3: *Pedido - El hablante le está pidiendo educadamente al oyente que haga tal cosa: hazlo, te lo pido.*

El ruido de la tele te molesta, por eso le pides a tu amigo, educadamente, que la apague. ¿Cómo se lo pedirías?

APAGA LA TELE.

(GOMES DA SILVA, 2019)

2.2. Análisis Acústico

Los enunciados grabados fueron separados con programa *Audacity* (2014), y analizados por el programa de análisis acústico PRAAT (BOERSMA & WEENINK, 2015). La segmentación de los datos se hizo con el uso del script *Easy Align* (GOLDMAN, 2011), en 4 niveles: transcripción fonética SAMPA (WEEELS, 1997), sílabas, palabras y enunciado.

Esta etapa de análisis acústico consiste en un descripción fonético-fonológica de las curvas entonativas, observando principalmente las variaciones en el tonema de los enunciados en acto de habla y variedad del español. En síntesis:

- Descripción fonética: variaciones de F0 a lo largo de los enunciados y de duración de sílabas para cada acto de habla y variedad del español.
- Descripción fonológica: a través de la notación Sp_ToBI (ESTEBAS VILAPLANA & PRIETO, 2009), contrastando las configuraciones en los tonemas para cada acto de habla y variedad del español.

2.3. Análisis de la variación visual

Para este análisis, se realizó la descripción y sistematización de los elementos que componen el sistema de pistas visuales y que interactúan con la prosodia auditiva. Utilizamos como base el repertorio del sistema FACS – *Facial Affect Coding System* (EKMAN *et alii*, 2002), que presenta movimientos de ojos, de labios y movimientos de cabeza. También consideramos los movimientos de hombros y manos, además de las sonrisas, conforme descrito en Figueiredo (2018).

A seguir, presentamos un ejemplo de la descripción, la musculatura facial utilizada para el movimiento y una figura que ilustra el movimiento, en Fig. 1.


AU	Descripción	Músculo Facial	Ejemplo en imagen
1	Elevación de la parte interna de la ceja	<i>Frontalis, pars medialis</i>	

Figura 1: descripción traducida (traducción propia) de los movimientos de faz (EKMAN & FRIESEN, 1978).

3. RESULTADOS Y DISCUSIONES

El análisis entonativo revela hay más diversidad de contornos entonativos entre los actos de habla directivos, si consideramos las posibilidades de descripciones del tonema de nos enunciados a través de la notación Sp_ToBI.

Tabla 2: Descripción fonológica Sp_ToBI por acto de habla y variedad del español.

Actos de Habla	Tonema
Pregunta	
Asunción	H+L*HH%
Buenos Aires	L+H*HL%
Ciudad de México	L*LH%
Pedido	
Asunción	H+L*HH% L*HH% H+L*L%
Buenos Aires	L*L% L+>H*L% L+H*L%
Ciudad de México	L*H% H+L*L%

Actos de Habla	Tonema
Súplica	
Asunción	H+L*L% L*L% H+L*HH%
Buenos Aires	H*L% L*L% H*L% H+L*L% H+L*HL% H+L*HH%
Ciudad de México	L+H*L% L+H*L% H+L*L%

En el análisis de duración de sílabas en el tonema, se generaron gráficos que indicaron la mediana y la dispersión de valores de duración para cada acto de habla y cada variedad del español analizada. Se normalizaron las medidas de duración por las variaciones intrínsecas de los fonemas, utilizando la metodología propuesta por Barbosa, Camargo y Madureira (2017), a través de la fórmula:

$$(1) \quad dsil^{z-score} = \frac{\sum_{i=1}^n dfon_i^{ms} - \sum_{i=1}^n \mu_i^{ms}}{\sqrt{\sum_{i=1}^n var_i^{ms}}}$$

Observamos que el acto de habla súplica se distingue de los demás por el alargamiento de las sílabas tónicas y postónicas, tanto del tonema como del pretonema.

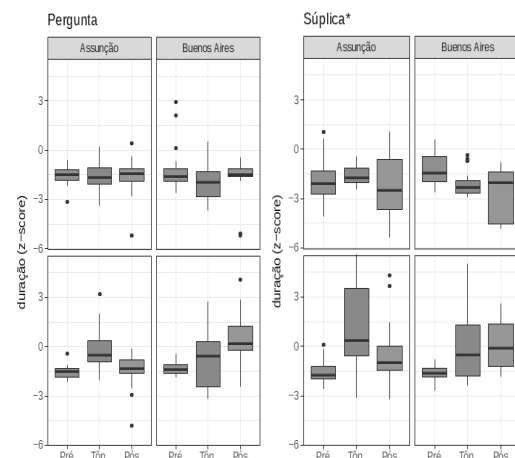
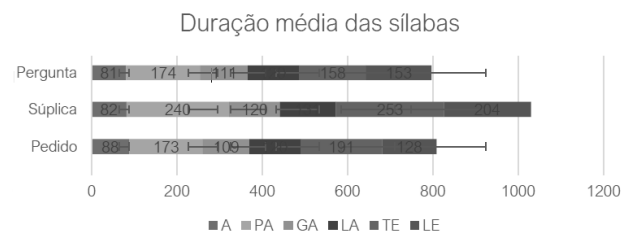


Figura 2: variación de duración de sílabas entre los actos de habla

Con el análisis de la variación visual, observamos que:

1. Para el acto de habla pregunta de Ciudad de México, las unidades de acción más frecuentes son “fruncir de cejas” y “movimiento de cabeza hacia delante”. Para la variedad de Buenos Aires: “Estiramiento de la comisura de los labios” (*Dimpler*) y “Parpadeo” (*Blink*). Para la variedad de Asunción: “Labios estrechados” (*Lip Funneler*), “Nariz y labios aumentados” (*Nasallabial Deepener*) y “Distanciamiento de los labios/relajación del mentón” (*Lipspart*).
2. Para el acto de habla pedido de Ciudad de México, las unidades de acción más frecuentes son “movimiento de cabeza hacia delante” y “sonrisa”. Para Asunción y Buenos Aires, “Movimiento da cabeza hacia delante” (*Head forward*).
3. Para el acto de habla pedido de Ciudad de México, las unidades de acción más frecuentes son “sonrisa”, seguido de “decaimiento del párpado”, “mirada hacia abajo” y “elevación de manos”. Para Asunción y Buenos Aires “Elevación de la parte interna de la ceja” (*Inner Brow Raiser*), “Elevación de los labios” (*Lip Corner Puller*) y “Ojos casi cerrados” (*Slit*). Los actos de habla súplica y pedido, son los que más utilizan la “sonrisa”, actuando como una estrategia de atenuación, y más seductora, en el caso del acto de habla súplica.



Figura 3: Ejemplo de producción de sonrisa para el acto de habla súplica

4. CONSIDERACIONES FINALES

Con relación al análisis prosódico de los actos de habla podemos considerar que:

En el nivel da entonación, el acto de habla pregunta no presenta variación de patrón fonológico en las tres variedades de español analizadas. Para los demás actos de habla, encontramos variaciones, con lo que no se puede definir un único patrón para cada acto. En lo que se refiere a la duración de sílabas, destacamos alargamientos más significativos en el acto de habla súplica, para todas as variedades. En el contraste dialectal, Buenos Aires se destaca por presentar mayores alargamientos en las sílabas postónicas del tonema de los enunciados analizados. La combinación entre el análisis acústico y visual

revelan que el acto de habla pregunta es más lingüístico y, por lo tanto, más perceptibles solamente por lo auditivo. El contraste entre los demás actos de habla – pedido y súplica – son menos lingüísticos y necesitan los datos visuales para su mejor interpretación.

5. BIBLIOGRAFÍA

- BARBOSA, Plinio A., Camargo, Zuleica A., & Madureira, Sandra (2017). Acoustic-based tools and scripts for the automatic analysis of speech in clinical and non-clinical settings. Berlín: De Gruyter.
- BOERSMA, Paul & Weenink, David (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.37, retrieved 14 March 2018 from <http://www.praat.org/>
- EKMAM, P.; FRIESEN, W.; HAGER, J. (2002). Facial Action Coding System. Salt lake City, UT: The Manual on CD ROM., Research Nexus Division of Network Information Research Corporation.
- EKMAN, P., & Friesen, W. V. (1978). Manual for the facial action coding system. Palo Alto, CA: Consulting Psychologists Press.
- ESTEBAS VILAPLANA, Eva y Prieto, Pilar (2009). La notación prosódica del español: una revisión del Sp_ToBI. In: Estudios de fonética experimental XVIII. Barcelona: Laboratorio de Fonética de la Universidad de Barcelona. pp. 263-83.
- FACIAL Action Coding System (2017). Disponible em: <http://www.paulekman.com/product-category/facs/> Consultado em: 17/09/2017
- FIGUEIREDO, N. S (2018). Variação pragmática e ecologia das línguas: análise multimodal de atos de fala no espanhol do Paraguai e da Argentina. Tese de Doutorado em Língua Espanhola. Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro.
- GOLDMAN, Jean-Philippe (2011). EasyAlign: an automatic phonetic alignment tool under PRAAT. Proceedings of InterSpeech, Firenze, Italy.
- GOMES DA SILVA, C. (2019). A prosódia de Atos de fala no espanhol da Cidade do México. Tese de Doutorado em Língua Espanhola. Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro.
- LADD, Robert (2008). Intonational Phonology. Cambridge: Cambridge University Press, 1996.
- MORAES, J.A. and Rilliard, A. (2014). Illocution, attitudes and prosody: A multimodal analysis. In: Raso, T. and Mello, H. (eds.) Spoken Corpora and Linguistic Studies. Amsterdam: John Benjamins, pp. 233-70.
- MORAES, Joao Antonio de, Rilliard, A., Mota, B. and Shochi, T. (2010). Multimodal perception and production and of attitudinal meaning in Brazilian Portuguese. Proceedings of Speech Prosody 2010: Fifth Conference on Speech Prosody, Chicago, paper 340.
- PRIETO, P.; Roseano, P. (eds.) (2010). Transcription of Intonation of the Spanish Language. Lincom Europa: Munchen, 2010.
- PRIETO, Pilar (2003). Teorías de la Entonación. Barcelona: Ariel.

SOSA, Juan Manuel (1999). *La entonacion del espanol. Su estructura fonica, variabilidad y dialectologia*. Madrid: Catedra.

WEELS, J. C (1997). SAMPA computer readable phonetic alphabet. In: Gibson, D., Moore, R. & Winski, R. (eds.). *Handbook of Standards and Resources for Spoken Language Systems*. Berlin and New York: Mouton de Gruyter. Part IV, section B.

DETERMINING SPEAKERS' ORIGIN ON THE BASE OF THEIR INTONATION: A PRELIMINARY STUDY

Wendy Elvira-García¹, Paolo Roseano^{2,3}, Ana Ma. Fernández Planas²

¹Universidad Nacional de Educación a Distancia, ²Universitat de Barcelona, ³University of South Africa
welvira@flog.uned.es, paolo.roseano@ub.edu, anamariafernandez@ub.edu

ABSTRACT

Este trabajo presenta AMPER_Forensic, una herramienta diseñada para ayudar a los lingüistas en la tarea de identificar el origen de un hablante. La herramienta usa datos exclusivamente prosódicos y propone al usuario una lista de lugares en los que se usan contornos entonativos parecidos a los que aparecen en la grabación objeto de estudio. AMPER_Forensic usa técnicas conocidas que se utilizan normalmente en estudios de dialectometría, aplicándolas a la tarea de atribución de dialecto, una parte del LADO que ha ganado popularidad en los últimos años.

Palabras clave: entonación, fonética forense, fonética judicial, AMPER

This paper presents AMPER_Forensic, a tool aimed at helping linguists in the task of assessing speaker's origin. The tool uses prosodic data exclusively and suggests to the user a list of places that use intonational patterns similar to those who appear in the target recording. AMPER_Forensic uses well-known dialectometric techniques and applies them to the dialect attribution task, which is a part of LADO that has been gaining popularity in recent years.

Keywords: intonation, forensic phonetics, AMPER

1. INTRODUCTION

Due to its importance in court cases, language analysis for the determination of origin (LADO), and specifically dialect attribution, has been the object of increasing research (Allen, DeLima, Freed, & Nielsen, 2019). However, determining speakers' origin is a difficult task even for trained linguists (Cambier-Langeveld, 2010).

For this reason, recent studies use *automatic* methods in order to identify dialects (Brown & Watt, 2014; Ferragne & Pellegrino, 2007; Hanani, Russell, & Carey, 2013; Vidya Prasad et al., 2019). No matter if the above-mentioned studies use textual data or spoken data, they tend to use a holistic approach, without giving the expert a clear indication concerning which parameters contribute the most to the result of the analysis they carry out. Only in few cases, automatic systems act "human-like" insofar as they choose some parameters that are believed to be qualitatively more significant than others. For example, Biadisy's (2011) system performs a

classification relying only on phones that are known to be realized differently in different dialects.

On the other hand, methods based on *human* expertise do not perform holistic analyses but choose some parameters that, according to the knowledge of the experts, can perform better than others can. Van Bezooijen and Gooskens (1999) explore the contribution of different linguistic levels to the identification of language varieties by human listeners and they conclude that "pronunciation" (sounds and prosody altogether) is more significant than the rest of levels. However, despite the existence of studies on the contribution of intonation to speaker identification (Cicres, 2007), the research about the contribution of intonation to dialect detection is not abundant.

This paper aims at testing whether prosody can be an important feature in semi-automatic origin determination. In order to do so, we have designed a tool that can help experts in their task of determining speaker's origin in Romance languages.

2. MATERIALS

2.1. Corpora

Determining speaker's origin requires a dataset that includes data from several survey point. To date, there are two intonational databases that could serve that purpose for Romance languages: the *Interactive Atlas of Romance Intonation* (IARI; Prieto, Borràs-Comes, & Roseano, 2010-2014) and the *Atlas Multimédia de la Prosodie de l'Espace Roman* (AMPER; Contini, 1992; Romano & Contini, 2001; Romano, 2003). This paper uses the latest.

The AMPER database consists of several corpus containing speech samples of different formality grades, ranging from spontaneous speech to laboratory read speech, including map tasks and discourse completion tasks.

In order to work with intonation, this study requires that the F0 contours are aligned among all samples. Therefore, the read speech corpus has been chosen. Specifically, this study uses read yes-no questions in three languages: Italian, Friulian (Roseano & Fernández Planas, 2009-2013) and Catalan (Martínez Celdrán & Fernández Planas, 2003-2019).

Every question of our corpus is formed by three constituents with lexical stress in the same positions (all of them are paroxytones). Examples are provided in (1) for Italian, (2) for Friulian and (3) for Catalan.

- (1) It. *La bambina mangiava la banana?*
- (2) Fri. *La ghitare si sunie cun dolcece?*
- (3) Cat. *El copista no porta la caputxa?*

2.2 Speakers

2.2.1. Accents database

The accents database used in this study includes recordings from 18 speakers of Catalan (i.e. 2 from each of the following survey points: Alicante, Andorra, Barcelona, Castelló de la Plana, Fraga, Girona, Tarragona, Tortosa, and Valencia), 8 speakers of Friulian (2 in each of the following survey points: Agrons, Beivars, Gradisca and Tesis), and 5 speakers of Italian from the following points: Venezia, Arezzo, Siena, Firenze, and Perugia.

In total 31 speakers of 18 survey points were included in the study.

2.2.2. Test database

In order to test if the attribution of accent performs correctly, we need a set of recordings that can be compared with the accent database. Therefore, we chose a set of recordings that could serve as "accent unknown recording" (AUR), which we will call test database.

In order to create the test database, we have used the following approach. In the survey points where we recorded two speakers, each recording has been excluded once from the training set and has been analysed as if it were the AUR. This allows us to test the performance of the program 26 times with different recordings.

2.3 Extraction and processing of prosodic data

The tool that we are presenting uses prosodic data, therefore sound files must be pre-processed in order to be correctly analysed.

To that end, the recordings have been analysed with the program AMPER06 that allows the user to extract and stylize F0 contours (López Bobo et al., 2007). The program saves in a txt, for each sentence, three F0 values in semitones per syllable (measures taken at the starting point, at the mid point, and at the end of each vowel), a value of vowel duration per syllable and the value of the mean intensity of each syllable.

3. METHODS

The tool created for helping the linguist in the task of determining speakers origin is a Python script called AMPER_Forensic. Its main goal is providing a rank of survey points that match the intonational contour used in the AUR. In order to do so, AMPER_Forensic takes the intonational contour of a given recording and compares it to the contours of previously recorded sentences that show prototypical accents of each region. In following sections, the pipeline of AMPER_Forensic is described in detail.

3.1 Input

The input of the program, as we have stated in section 2, is a set of txt files containing prosodic data. Such txt files (see Figure 1 for an example) are generated with AMPER06, but the program can be adapted to any kind of input as long as the input keeps the syllable/stress position correspondence in all the languages/varieties used in the dataset.

```

C:\Amper\ficherostxtSt\8001005p0p0a0u0e03x6xAx11i1.txt
size: 29609
29-Oct-2010

zona      duration [ms]      energy [dB]      fo1
fo2      fo3 [ST]

1         31         101             -0.32 -0.07 0.25
2         47         104             1.71  2.58  3.33
3         96         104             3.51  0.73  -0.57
4         47         99              -1.52 -2.02 -2.47
5         47         93              -2.43 -1.61 -1.08
6         51         99              1.42  2.51  3.46
7         68         99              3.04  2.65  2.15
8         46         96              0.49  -0.07 -0.45
9         44         99              -2.70 -2.70 -2.62
10        96         100             -2.91 -2.91 -2.91
11        114        98              -1.90 0.33  3.13

values at:
1687 1935 2182 3095 3473 3851 5389 6158 6927 7527 7905 8283
10030 10408 10786 12480 12884 13288 14149 14696 15244 16573
16938 17303 18503 18855 19206 21735 22504 23274 25516 26428
27341

```

Figure 1. Example of an input file. It includes data of F0 in semitones, duration and intensity of the sentence.

3.2 Computing intonation differences

Once the AUR sentence is loaded, the system compares it with the sentences in the database (in the case of this paper, the corpora explained in section 2.2.2). This comparison is performed by computing the similarities and differences between the AUR and the rest of them. In order to do so, the program computes Pearson correlation, as this is the metric that has proved to be more useful for measuring similarities among intonation contours (Hermes, 1998). The approach is similar to what is done in intonational dialectometry (Elvira-García et al., 2018).

Correlation values range from -1 to 1, where 1 is the perfect correlation. Thus, the higher values of correlation, the more likely that the two sentences analysed are from the same dialectal area.

4. RESULTS

The results of the program contain all the correlations of the AUR with the contours in the Accents database. Such results are offered to the user in two ways: numerically (in a txt) and visually (as depicted in Figure 2) ordered from highest correlation coefficient to lowest correlation coefficient. Therefore, survey points that share intonational patterns with the recording appear in the highest part of the rank (left side of Figure 2). On the contrary, cities with other patterns appear with lower correlation indices (in right part of the figure).

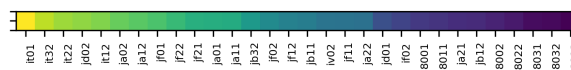


Figure 2. Heatmap of ordered correlations created by AMPER_Forensic.

For the dataset analysed, the results have correlated each AUR with recordings of the same linguistic area (Friulian, Catalan and Italian) in 100% of the cases. However, the speaker with the highest correlation with the AUR is not always a speaker from the same survey point. We have assessed classifications as correct when the two speakers of the same city have a correlation higher than 0.9 (this is true in 42% of the cases).

5. LIMITATIONS AND FUTURE WORK

The results obtained are satisfactory but not perfect, insofar as AMPER_Forensic identifies correctly to which large geoproscopic area the AUR belongs to, but most of the times it does not identify correctly the survey point. In this section, we try to explain possible reasons for that.

The poor performance of the tool for locality identification was not unexpected. In fact, it would be rather obvious for any dialectologist. Whereas some dialectal areas have different intonational patterns from town to town, others, especially varieties with a strong standard or areas with great mobility are levelling their accents making them almost disappear. For those places, we could never achieve a good performance. If speakers do not have a “survey point accent”, neither a machine nor a linguist can find it (Dyer, 2002). This is most likely the case for many of the Catalan dialectal areas, since previous research has shown that cities like Andorra and Barcelona share the same intonational pattern in SVO yes-no questions (Fernández Planas et al., 2015).

However, this is not the only factor that could have had an effect on the outcome of the analysis we present in this paper. The approach adopted to test the data (see Section 2.2.1) has made that for each trial we had only a speaker of the same survey point as the AUR in the database. Therefore, if the two speakers of the same survey point used different intonational patterns (which is something that can happen due to the fact that some variation is always observed in intonation), AMPER_Forensic has not been able to identify correctly the survey point. In order to overcome this shortcoming, more speakers

of each survey point should be included in the database.

Furthermore, a more comprehensive model should be created. In this paper, we have compared the contours one to one. Nevertheless, in an ideal scenario, when a survey point presents more than one intonational pattern, the AUR should be compared to every contour attested in the area. In order to do so, the most straightforward way is changing the approach and constructing a model that includes all the features attested in the dialect. For doing so, the next step we are working on is using a neural network classifier.

6. CONCLUSIONS

In this paper we have presented AMPER_Forensic, a tool aimed at helping linguists in the task of assessing speaker's origin using prosody. The results show that intonation can be used in dialect identification tasks insofar as AMPER_Forensic is able to assign an AUR to its geoproisodic area. However, the system does not recognize the specific survey point from which an AUR comes from. This limitation can be explained by the natural intra-speaker and inter-speaker variation and can be solved by means of a more comprehensive classification method, like new machine learning techniques.

7. REFERENCES

- Allen, C. O., DeLima, R., Freed, A. R., & Nielsen, R. L. (2019). *U.S. Patent Application No. 10/275,446*.
- Biadys, F. (2011). *Automatic dialect and accent recognition and its application to speech recognition*. PhD thesis, Columbia University.
- Brown, G. and Watt D. 2014. Performance of a novel automatic accent classifier system using geographically proximate accents. Poster session *BAAP*, University of Oxford, 7th-9th April.
- Cambier-Langeveld, T. (2010). The role of linguists and native speakers in language analysis for the determination of speaker origin. *International Journal of Speech, Language & the Law*, 17(1), 67-93.
- Cicres, J. (2007). *Aplicació de l'anàlisi de l'entonació i de l'alineació tonal a la identificació de parlants en fonètica forense*. PhD thesis, Universitat Pompeu Fabra.
- Contini, M. (1992). Vers une géoprosodie romane. In G. Aurrekoetxea & X. Bidegain (Eds.). *Actas del Nazioarteko Dialektologia Biltzarra Agiriak*, Bilbao 1991. Bilbao: Publicaciones de la Real Academia de la Lengua Vasca, 83-109.
- Dyer, J. (2002). 'We all speak the same round here': Dialect levelling in a Scottish-English Community. *Journal of Sociolinguistics*, 6(1), 99-116.
- Elvira-García, W., Balocco, S., Roseano, P., & Fernández Planas, A. M. (2018). ProDis: A dialectometric tool for acoustic prosodic data. *Speech Communication* 97, 9-18.
- Fernández Planas, A. M., Roseano, P., Elvira-García, W., Cerdà-Massó, R., Romera-Barrios, L., Carrera-Sabaté, J., Szmídt, D., Labraña, S., & Martínez Celdrán, E. (2015). Cap a un nou mapa dialectal del català? Consideracions a partir de dades prosòdiques tractades dialectomètricament. *Estudios de Fonética Experimental* 24, 257-286.
- Ferragne, E., & Pellegrino, F. (2007). Automatic dialect identification. A study of British English. In Müller, C. (Ed.). *Speaker classification*. Berlin: Springer, 243-257.
- Hanani, A., Russell, M., & Carey, M. (2013). Human and computer recognition of regional accents and ethnic groups for British English speech. *Computer Speech and Language* 27, 59-74.
- Hermes, D. J. (1998). Measuring the perceptual similarity of pitch contours. *Journal of Speech, Language, and Hearing Research*, 41(1), 73-82.
- López Bobo, M. J., Muñiz Cachón, C., Díaz Gómez, L., Corral Blanco, N. Brezmes Alonso, D., & Alvarellos Pedrero, M. 2007. Análisis y representación de la entonación: Replanteamiento metodológico en el marco del proyecto AMPER. En Dorta, J. (Ed.). *La prosodia en el ámbito lingüístico románico*. Santa Cruz de Tenerife: La Página, 17-34.
- Martínez Celdrán, E., & Fernández Planas, A. M. (Eds.) (2003-2019). *Atlas multimèdia de la prosòdia de l'Espai romànic*. http://stel.ub.edu/labfon/amper/cast/index_ampercat.html (2019-12-12).
- Prieto, P., Borràs-Comes, J., & Roseano, P. (Eds.) (2010-2014). *Interactive atlas of Romance intonation*. <http://prosodia.upf.edu/iari/> (2019-12-12).
- Romano, A. (2003). Un projet d'atlas multimèdia prosodique de l'espace roman (AMPER). In Sánchez Miret, F. (Ed.). *Actas del XXIII Congreso Internacional de Lingüística y Filología Románica*. Salamanca, 24-30 septiembre 2001, vol. I: Discursos inaugurales. Conferencias plenarias. Sección 1: Fonética y fonología. Sección 2: Morfología. Índices: Índice de autores, Índice general. Tübingen, Niemeyer, 279-294.
- Romano, A., & Contini, M. (2001). Un progetto di atlante geoproisodico multimediale delle varietà linguistiche romanze. In Magno Caldognetto, E. & Cosi, P. (Eds.). *Multimodalità e multimedialità nella comunicazione. Atti delle XI Giornate di Studio del Gruppo di Fonetica Sperimentale dell'Associazione Italiana di Acustica*. Padova: Unipress, 121-126.
- Roseano, P., & Fernández Planas, A. M. (Eds.) (2009-2013). *Atlant multimèdiâl de prosodie des varietâts romanichis*. <http://stel.ub.edu/labfon/amper/friul/index.html/> (2019-04-04).
- Van Bezooijen, R., & Gooskens, C. (1999). Identification of language varieties: The contribution of different

linguistic levels. *Journal of language and social psychology*, 18(1), 31-48.

Vidya Prasad, K., Akarsh, S., Vinayakumar, R., & Soman, K. P. (2019). A Deep Learning Approach for Similar Languages, Varieties and Dialects. *arXiv preprint arXiv:1901.00297*.

Acknowledgements. This paper has been developed within the framework of the project *Technologies derived from AMPER-CAT and analysis of complementary corpora* (FFI2015-64859-P) funded by the Spanish *Ministerio de Economía y Competitividad*. We thank the Faculty of Philology and Communication of the University of Barcelona for its financial support for the acquisition of the GPU (NVIDIA GTX 1060 6GB) that has been used in this study.

UMBRAL DE INTELIGIBILIDAD EN AUDIOMETRÍAS VERBALES: VALORES DE REFERENCIA EN NIÑOS NORMOYENTES DE 5, 6 Y 7 AÑOS

Wendy Elvira-García¹, Paolo Roseano^{2,3}, Ana Ma. Fernández Planas³, Silvia Marro⁴, Pilar Calvo⁵, M. Antonia Claveria⁶

¹Universidad Nacional de Educación a Distancia, ²Universitat de Barcelona, ³University of South Africa, Equipo de Asesoramiento y Orientación Psicopedagógica (EAP) B-34 Sabadell-Sant Quirze, ⁵CREDA Jordi Perelló, ⁶Hospital Sant Joan de Déu

welvira@flog.uned.es; paolo.roseano@ub.edu, anamariafernandez@ub.edu, smarro@xtec.cat, pcalvo@xtec.cat, maclaveria@sjdhospitalbarcelona.org

ABSTRACT

El siguiente trabajo examina el umbral de intensidad en tareas de audiometría verbal o logaudiometría para niños de 5 a 7 años. El trabajo tuvo 52 participantes (niños y niñas de Sabadell escolarizados en catalán). Se realizaron audiometrías verbales con palabras bisílabas y frases como estímulos. Los estímulos se reprodujeron con intensidad descendente de 15dB, 20dB, 30dB, 40dB, 45dB. Para los niños de 5 años, el rango fue de 45dB a 15dB. La tasa del 100% de repetición se consiguió a 45dB en niños de 5 años, a 40dB en niños de 6 años y a 30dB en niños de 7 años.

Palabras clave: audiometría verbal, logaudiometría, curva inteligibilidad, población infantil

This paper examines the threshold intensity level for speech in verbal audiometry applied to a normal hearing children population (5, 6 and 7 years old). Fifty-two scholars took part in the test. The materials used were balanced lists of bisyllabic words and sentences. The stimuli were played in an intensity descendent order in the following steps: 15dB, 20dB, 30dB, 40dB, 45dB. For 5 years-old, the range of the stimuli was from 45dB to 15dB. The 100% repetition rate is achieved by 5-year-olds at 45dB; by 6-year-olds, at 40dB and by 7-year-olds, at 30dB.

Keywords: verbal audiometry, speech threshold, normal-hearing children, intensity threshold

1. INTRODUCCIÓN

Las logaudiometrías (*speech audiometry*) son una herramienta habitual de evaluación audiológica en niños con pérdida auditiva (hipoacusia) (Fernández Planas et al. 2017). En ellas, los pacientes tienen que escuchar una grabación (o la voz del audioprotesista) y repetir (o señalar en una lámina) la palabra que acaban de oír. En una audiometría completa esas palabras se presentan a diferentes intensidades tanto para determinar el umbral de inteligibilidad verbal, es decir, a qué intensidad los pacientes son capaces de repetir/comprender una palabra, como la máxima inteligibilidad verbal que es el porcentaje más alto de repetición de palabras a una determinada intensidad.

El resultado de estas evaluaciones se suele mostrar en un audiograma (*speech audiogram*), un gráfico de línea, como el que se observa en la figura 1. En él, se muestra el tanto por ciento de palabras repetidas (eje y), en las diferentes intensidades que se han usado en la prueba (eje x). En estos gráficos, la curva (función) del paciente (en la Figura 1 en gris) se presenta junto con la curva de referencia para población normoyente (en la figura 1 en negro). Los valores de audición normales dan como resultado un gráfico en forma de S.

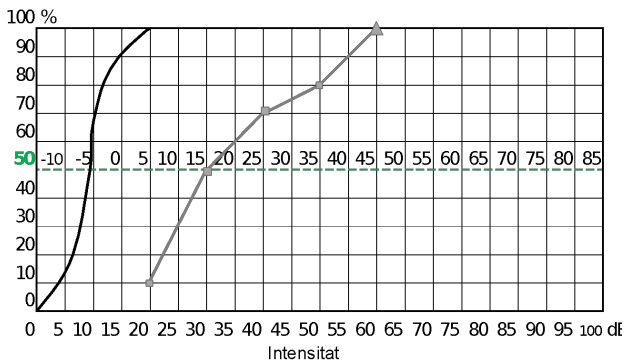


Figura 1: Gráfico de una audiometría verbal con la función normal y de un paciente. La línea discontinua verde horizontal representa el umbral de recepción verbal (a 30dB).

Siguiendo los ejemplos de la American Speech-Language-Hearing Association (ASHA), las hojas de evaluación con las que cuentan los audiólogos de los *Centres de Recursos per a Deficients Auditius*¹ presentan una curva de referencia tomada de los valores de población adulta (ASHA, 1988).

Sin embargo, los audioprotesistas de los CREDA trabajan con población infantil y no población adulta. Pero sabemos que el umbral de inteligibilidad del habla difiere entre niños de diferentes edades debido a factores relacionados con sus facultades receptoras (Jerger, Jerger y Abrams, 1983). Y que los valores de la función de inteligibilidad difieren según la tarea (reconocimiento de sílabas aisladas o de palabras) e incluso dependiendo de la lista de palabras que se use o el lugar donde se pase la prueba (Bradley y Sato, 2004; Garlock, Walley y Metsala, 2001).

Por todo lo dicho, parece injusto usar como punto de comparación para niños con implante coclear o audífono los valores de población adulta ya que, además, un niño con una edad cronológica de, 6 años, puede tener una edad auditiva de, por ejemplo, 4 años, dependiendo de la edad a la que empezara a usar el audífono con asiduidad o se hubiera realizado la activación del implante coclear.

El objetivo de este estudio es proporcionar a los profesionales el umbral y las funciones de inteligibilidad del habla (usando unas listas concretas) para niños normoyentes de 5, 6 y 7 años con el propósito de que estos se puedan usar como valor de control cuando se evalúa el habla de niños con hipoacusia.

¹ Los *Centres de Recursos per a Deficients Auditius* o CREDA son servicios educativos públicos del Departamento de Educación formados por un equipo

Lo que sigue del artículo presenta, en §2, la metodología (oyentes, materiales y pruebas) seguida para conseguir las funciones de inteligibilidad normales. La sección 3, presenta las funciones de inteligibilidad normales para niños de 5, 6 y 7 años y, en la sección 4, se resumen los resultados del estudio.

2. METODOLOGÍA

2.1. Participantes

En los tests participaron 52 niños de entre 5 y 7 años (edad cronológica). Todos ellos eran alumnos de la Escola Samuntada en Sabadell sin problemas de audición diagnosticados (antes de realizar la logaudiometría se practicó una otoscopia y una impedanciometría para descartar posibles problemas auditivos que pudieran alterar los resultados de la prueba). De ellos, 17 niños (9 niños y 8 niñas) tenían 5 años. 17 niños (8 niños y 9 niñas), tenían 6 años y 18 niños (12 niños y 6 niñas) tenían 7 años en el momento en el que se pasaron los test. Todos ellos proceden de un área bilingüe catalán-español y estaban escolarizados en inmersión en lengua catalana. Ningún alumno presentaba problemas atencionales conocidos. Todas las pruebas se realizaron en horario matinal para evitar, en la medida de lo posible, efectos de la fatiga.

2.2. Materiales

Los materiales usados para el test se explican con detalle en Fernández Planas et ál., 2017. Se realizaron 2 pruebas, la primera con palabras aisladas y la segunda, con frases.

Para la primera prueba se usaron listas de 20 palabras bisilábicas fónicamente equilibradas (balanced) en catalán central (Figura 2). La tarea consistía en repetir correctamente la palabra que habían oído. En la segunda prueba se tenían que repetir frases. Estas frases también estaban equilibradas fónicamente y contenían léxico de alta frecuencia.

En especial, se cuidó que todas las palabras formaran parte del vocabulario infantil y se tuvieron en cuenta las siguientes características: a) familiaridad del vocabulario, b) similitud fonética-semántica y c)

interdisciplinar de audiólogos, psicopedagogos y logopedas que proporcionan intervención directa al alumno y soporte en centros educativos para niños con sordera y con trastornos del lenguaje.

homogeneidad en la audibilidad. También se realizó un análisis de la frecuencia de aparición de los fonemas, la división silábica y el tipo de palabra. Para la selección de las palabras que integran las listas de la logaudiometría con soporte visual además se tuvo en cuenta: tamaño del dibujo, uso del color, relación palabra dibujo y convencionalidad del mismo.

Todas las listas están adaptadas a la población infantil, lo que permitió usar las mismas listas para los niños de 5, 6 y 7 años para obtener la curva logaudiométrica a diferentes intensidades (dB), de esta manera, se garantizaba la consistencia interna de la prueba.

Los estímulos usados fueron pregrabados con una voz femenina y sin ruido. Y se presentaron con altavoces coaxiales mediante el campo libre (FF) a través del reproductor de audio conectado al audiómetro. Los estímulos están también disponibles en voz masculina e infantil, con ruido blanco y ruido *party*. Estos no se usaron en la prueba para evitar la fatiga de los estudiantes.

PARAULES BISÍL-LABIQUES
Fernández Planas et al. (2016)

LISTA 1

NDIM: A. H.							
data:	data:	data:	data:	data:	data:	data:	data:
<input type="checkbox"/> veu home	<input type="checkbox"/> veu home	<input type="checkbox"/> veu home	<input type="checkbox"/> veu home	<input type="checkbox"/> veu home	<input type="checkbox"/> veu home	<input type="checkbox"/> veu home	<input type="checkbox"/> veu home
<input type="checkbox"/> veu dona	<input type="checkbox"/> veu dona	<input type="checkbox"/> veu dona	<input type="checkbox"/> veu dona	<input type="checkbox"/> veu dona	<input type="checkbox"/> veu dona	<input type="checkbox"/> veu dona	<input type="checkbox"/> veu dona
<input type="checkbox"/> veu nen	<input type="checkbox"/> veu nen	<input type="checkbox"/> veu nen	<input type="checkbox"/> veu nen	<input type="checkbox"/> veu nen	<input type="checkbox"/> veu nen	<input type="checkbox"/> veu nen	<input type="checkbox"/> veu nen
pròtesi: IC	pròtesi:	pròtesi:	pròtesi:	pròtesi:	pròtesi:	pròtesi:	pròtesi:
dB: 50	dB:	dB:	dB:	dB:	dB:	dB:	dB:
% encerts: 20/20	% encerts: /20	% encerts: /20	% encerts: /20	% encerts: /20	% encerts: /20	% encerts: /20	% encerts: /20
% encerts: 100	% encerts:	% encerts:	% encerts:	% encerts:	% encerts:	% encerts:	% encerts:
imoto	+						
flatxes	+						
calbes	+						
quince	+						
nina	+						
cebes	+						
plata	+						
cabina	+						
figre	+						
bideu	+						
patres	+						
sturo	+						
salut	+						
serra	+						
roda	+						
mida	+						
gomes	+						
pinjol	+						
madai	+						
lissa	+						

Figura 2: Ejemplo del protocolo de recogida de datos usado en la prueba.

2.3. Las audiometrías verbales

Las audiometrías se realizaron en 2015 en una cabina audiológica en el CREDA de Sabadell. Los estímulos se pasaron en orden descendente de 40dB to 10dB en pasos de 5dB. En el caso de los niños de 5 años, el rango de los estímulos fue de 45dB a 15db. En alumnos de menor edad poder obtener una curva con tres intensidades provocaba una cierta fatiga por lo que al realizar la prueba se proponía a los alumnos poder realizar descansos entre las diferentes listas presentadas.

3. RESULTADOS

3.1. Resultados para niños de 5 años

Los niños de 5 años empiezan a repetir palabras a los 20dB con una tasa de repetición del 38% de las palabras. La inteligibilidad a los 30dB fue del 85% a los 40dB a del 97% y, finalmente se obtuvo el 100% de repetición a los 40dB. Los resultados se ilustran en la figura 2.

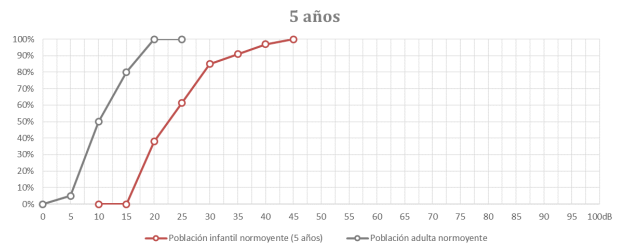


Figure 3: Gráfica de los resultados de la logaudiometría o prueba verbal con la función de inteligibilidad para 5 años (rojo) y los valores de referencia para adultos (gris).

3.2. Resultados para niños de 6 años

Los niños de 6 años empezaron el test en 10dB y repitieron el 0% de las palabras a esta intensidad. A 20dB consiguieron repetir el 54% de las palabras y a 30dB el 96%. A 40dB pudieron repetir el 99% de las palabras que escucharon.

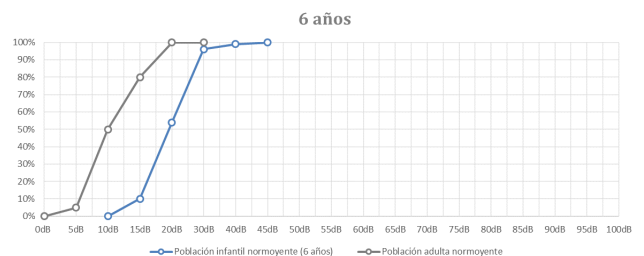


Figura 4: Gráfica verbal con la función de inteligibilidad para 6 años (azul) y los valores de referencia para adultos (gris).

3.3. Resultados para niños de 7 años

En niños de 7 años la tasa de repetición a 15dB es del 25%. A 20dB es del 74% a 30dB es del 98% y a 30dB del 100%.

Aquí, tal y como se puede observar en la figura 5, la función de inteligibilidad es muy similar a la de los adultos. Además, la función se acerca a la característica forma de S que se espera encontrar en los audiogramas.

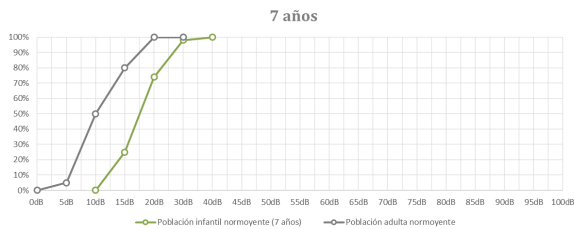


Figure 5: Gráfica de los resultados de la prueba verbal con la función de inteligibilidad para 7 años (verde) y los valores de referencia para adultos (gris).

3.4. Resultados en frases

Los resultados de inteligibilidad para frases administradas sin ruido muestran una mejoría en la inteligibilidad de las mismas según aumenta la edad cronológica.

En este caso los índices de repetición son menores (89% el más alto) ya que la memoria y el cansancio juegan un papel importante en la repetición, pero se puede observar (Figura 6) cómo se produce un mayor salto de los 5 a los 6 años (probablemente debido al desarrollo de las habilidades cognitivas y la memoria), mientras que entre 6 y 7 años no hay prácticamente cambios en los índices de repetición.

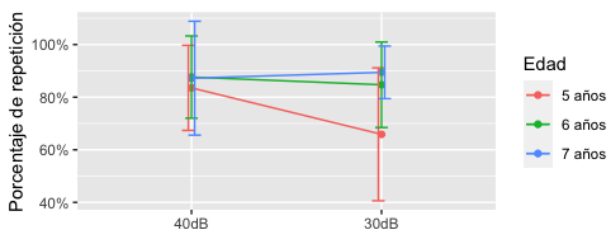


Figure 6: Resultados de inteligibilidad (a 30dB y 40dB) de listas de frases para niños de 5, 6 y 7 años.

4. CONCLUSIONES

Los resultados muestran que a medida que la edad cronológica aumenta, la curva se acerca a los niveles de inteligibilidad de los adultos (Figura 7). Estos resultados son consistentes con la poca literatura existente para niños normoyentes donde se constata que, aunque los niños consiguen un nivel de inteligibilidad equiparable al de los adultos con 4 años, esta sigue mejorando ligeramente hasta los 7 (Chin, Tsai, Gao, 2003).

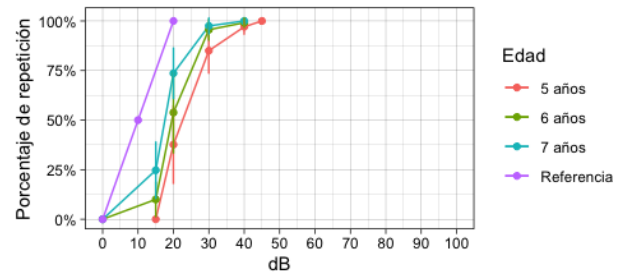


Figura 7: Gráfico que ilustra las funciones de inteligibilidad para niños de 5, 6 y 7 años y en adultos.

En este trabajo se han presentado las funciones de inteligibilidad del habla para niños normoyentes de 5, 6 y 7 años con el propósito de que estos se puedan usar como valor de control cuando se evalúa el habla de niños usuarios de prótesis auditivas.

5. BIBLIOGRAFÍA

- American Speech-Language-Hearing Association. (1988). Determining threshold level for speech [Guidelines]. Available from www.asha.org/policy.
- Bradley, J. S. y Sato, H. (2004). Speech intelligibility test results for grades 1, 3 and 6 children in real classrooms. Proceedings of ICA, Kyoto.
- Chin, S. B., Tsai, P. L. y Gao, S. (2003). Connected speech intelligibility of children with cochlear implants and children with normal hearing.
- Jerger, S.; Jerger J.; Abrams, S. (1983) Speech audiometry in the young child. Ear and Hearing. Jan-Feb;4(1): 56-66.
- Fernández Planas, Ana M.; Marro, S.; Roseano, Paolo; Calvo, Pilar; Elvira-García, Wendy; Claveria, Antonia. (2017). Logaudiometries infantils en català oriental central. Barcelona: Horsori.
- Garlock, V. M.; Walley, A. C. y Metsala, J. L. (2001). Age-of-acquisition, word frequency, and neighborhood density effects on spoken word recognition by children and adults. Journal of Memory and Language, 45(3), 468-492.

Agradecimientos. Este trabajo se ha llevado a cabo gracias a la ayuda económica de la UB y del Servicio ORL del Hospital Materno-Infantil Sant Joan de Déu. También queremos agradecer la ayuda prestada a la escuela Samuntada de Sabadell y a nuestros pequeños participantes y sus padres que tan desinteresadamente ofrecieron sus oídos y su tiempo para baremar esta prueba.

APLICACIONES DEL ULTRASOUND SOUND IMAGE (UTI) A LOS ESTUDIOS DE FONÉTICA

Wendy Elvira-García¹, Paolo Roseano^{2,3}, Lourdes Romera Barrios³, Victoria Marrero Aguiar¹

¹Universidad Nacional de Educación a Distancia, ²Universitat de Barcelona, ³University of South Africa
welvira@flog.uned.es, paolo.roseano@ub.edu, lourdesromera@ub.edu, vmarrero@flog.uned.es

ABSTRACT

Este texto, presentado a la sesión monográfica de *Nuevas metodologías para el estudio fonético: qué nos aportan y qué nos exigen*, pretende ofrecer una panorámica de las características y de las aplicaciones de la visualización de la lengua por ultrasonido (UTI) al ámbito de la fonética. El trabajo presenta el equipo necesario para llevar a cabo una grabación de ultrasonido y algunos análisis posibles.

Palabras clave: ultrasound tongue imaging, técnicas fonéticas, instrumentos

This paper aims to give an overview of the ultrasound tongue imaging (UTI) characteristics and applications in phonetics. It is been submitted to the specific panel *Nuevas metodologías para el estudio fonético: qué nos aportan y qué nos exigen*. The paper presents the equipment needed to perform an ultrasonic recording and some of the possible analyses.

Keywords: ultrasound tongue imaging, phonetic techniques, tools

1. INTRODUCCIÓN

A lo largo de las décadas se han ido desarrollado varias técnicas de análisis articulatorio, como la pletismografía, la electromiografía, la fotoglografía, la electrolaringología, la resonancia magnética, la radiografía, la articulografía electromagnética, la aerometría, la endoscopia, la electropalatografía o la videoquimografía.

En los años 70 del siglo pasado se empezó a usar una nueva técnica para visualizar el aparato fonador, la ecografía. Al principio, se usó para visualizar las articulaciones faríngeas (Kelsey et al., 1969; Skolnick et al., 1975; Zagzebski, 1975; MacKay, 1977, etc.) y, más tarde, en los años 80 se aplicó a la visualización de la lengua (Morrish et al., 1984). Esta última técnica, se llamó *Ultrasound Tongue Imaging*, o 'visualización de la lengua mediante ultrasonido' (UTI, por sus siglas en inglés) (Wrench y Scobbie, 2006; Gick et al., 2013; Lulich et al., 2018).

Pese a que desde los 80 se venían efectuando estudios lingüísticos, el principal uso del UTI era el ámbito clínico. Son abundantes los trabajos sobre alteraciones de la deglución (por ejemplo, Shawker, 1983), la succión en bebés (Bosma et al., 1990) o la

posición de la lengua durante la anestesia (Abernethy et al., 1990), entre otros.

Sin embargo, en los 2000 se produjeron dos hechos que ayudaron a impulsar los estudios lingüísticos realizados con UTI. En primer lugar, nació una iniciativa, el UltraFEST, en la que se reúnen investigadores que usan el UTI para la investigación lingüística. En segundo lugar, la aparición de unidades de ultrasonido portátiles y de menor precio y la mejora de los softwares de análisis de imágenes han hecho que aumente su adquisición por parte de los laboratorios de fonética.

El UTI en la actualidad es una técnica no invasiva para realizar imágenes dinámicas y permite a los investigadores ver la forma, la posición y los movimientos de la lengua (desde la raíz hasta el ápice) en tiempo real durante el habla. Esta herramienta está disponible, hasta donde sabemos, en tres laboratorios de fonética del estado español: el de la Universidad de Navarra (UN), el del Institut d'Estudis Catalans (IEC) y el de la Universitat de Barcelona (UB).

El objetivo de este trabajo es mostrar como el uso de una técnica de análisis articulatorio de última generación puede arrojar nueva luz sobre temas controvertidos del ámbito fonético-fonológico, a la

vez que se anima a los asistentes habituales al CIFE a empezar por cuestiones aún abiertas en la lingüística española, y más en general, iberorrománica y románica. Estaríamos hablando, por ejemplo, de temas tan discutidos como la variación dialectal de las realizaciones de la /s/ en español (Alvar, 1996; Cano, 2005), el punto de articulación de la /tʃ/ en catalán y en español peninsular central (Fernández Planas, 2000; Martínez Celdrán y Fernández Planas, 2001; Recasens, 2019), el papel del rasgo ATR en la armonía vocálica del andaluz o del valenciano (Lloret y Jiménez, 2009; Herrero y Jiménez, 2013), entre otros.

2. EQUIPO NECESARIO

Como casi todas las técnicas de fonética articuladora, el UTI requiere un equipo específico. Básicamente, se trata de un ecógrafo y una sonda, un elemento de sujeción de la sonda, así como de los controladores y del programario correspondientes.

2.1. El ecógrafo

Los laboratorios de fonética han pasado en los últimos tiempos del uso de grandes equipos (como los Digital Sona-Graph 7800 y DSP Sona-Graph 5500) al uso de un ordenador personal. La historia de los equipos de ecografía ha mostrado un desarrollo similar.

En la actualidad, se ha pasado de la existencia de grandes ecógrafos de pantalla integrada (equivalentes en tamaño a los Sona-Graph) a equipos que se pueden conectar a un ordenador mediante USB (Figura 1).



Figura 1: Ecógrafo de bolsillo de Telemed y sonda.

2.2. La sonda

A este equipo se conecta una sonda, que es la encargada de producir el sonido de alta frecuencia que sirve para crear la imagen del ecógrafo.

Las características de la sonda dependen del lugar donde se quiera aplicar el ecógrafo. En los estudios de fonética, la sonda se suele colocar sobre la parte

baja de la mandíbula, orientada de detrás-adelante para conseguir una imagen sagital de la lengua. Como es una zona pequeña, se suelen usar para ello sondas pediátricas como la MC4-2R20S-3 de TELEMED. Esta sonda en concreto, disponible para el uso de investigadores en el laboratorio de fonética de la UB (Figura 1), tiene una frecuencia de 2-4MHz, 20mm de radio y un campo máximo campo de visión de 92º, perfecto para cortes sagitales.

2.3. La monitorización o sujeción de la sonda

Existen varios modos de colocar la sonda al informante. La primera manera es que el propio informante sujete la sonda sin que haya un sistema de sujeción. Este método es el más cómodo de usar y es el utilizado para efectuar, por ejemplo, demostraciones sobre articulaciones en clase (Mozafari et al., 2008). Sin embargo, tiene inconvenientes. El principal entre ellos es que el ultrasonido ofrece imágenes de la lengua, pero no su posición respecto una referencia absoluta. Para calcular la posición de la lengua en la boca, se usa como referencia el centro de la imagen que se graba. Por lo tanto, para mantener esa referencia fija, la sonda no se puede mover durante la grabación.

Por esa razón, aunque es cierto que se puede confiar en el pulso del informante para que él mismo sujete la sonda, lo más probable es que en algún momento de la grabación el informante mueva la cabeza o la sonda se mueva. Si eso ocurre, la grabación queda inutilizable.

Para evitar perder los puntos de referencia durante las grabaciones, se han creado diferentes métodos de monitorización del movimiento y sujeción de la sonda.

Para evitar que el informante se mueva demasiado cuando está sujetando la sonda, hay investigadores que usan un puntero láser que apunta desde la sonda a un cuadro en una pared cercana y ordenan al informante que mantenga el puntero dentro del cuadro mientras habla. Este método hace que tanto el investigador como el informante puedan monitorizar cuando se ha movido la posición de la sonda y, si esto ocurre, la grabación se debería desechar y empezar de nuevo, ya que se habrían movido los puntos de referencia para el análisis.

Un método que requiere menos confianza en la habilidad del informante es la monitorización del movimiento de la cabeza respecto a la sonda. Si se dispone de la información sobre los movimientos que ha efectuado el informante, después se puede reorientar la imagen de ultrasonido usando esa información. Para implementar este sistema, existen

estándares como el de Haskins Optically Corrected Ultrasound System (HOCUS), en el que se usan diodos emisores de infrarrojos en la sonda y en diversos puntos de la cara del informante (Whalen et al., 2005).

Por último, existen diversos modelos de cascos de sujeción para la sonda. Estos cascos, aunque son la solución más limpia en términos de análisis, tienen algunos detractores, ya que la inmovilización parcial de la mandíbula que se lleva a cabo con el casco puede tener efectos en la producción con naturalidad de los estímulos.

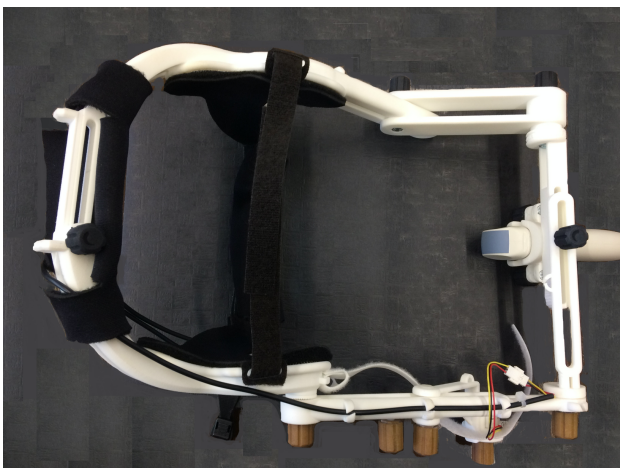


Figura 2. Casco de sujeción UltraFit.

Sin embargo, el diseño de los cascos también ha mejorado en los últimos años y se ha pasado de modelos de metal (Stone, 1995; Scobbie et al., 2008) a modelos de nylon como el UltraFit, que pesa 350gr (Spreafico et al., 2018) (Figura 2).

2.4. La sincronización

Para que las imágenes obtenidas con el ultrasonido puedan tener una interpretación lingüística, el investigador necesita oír el sonido que está sonando a la vez que la imagen que se reproduce.

Este hecho, obvio para los fonetistas, en realidad no es una cuestión técnica trivial, ya que la imagen y el sonido se transmiten a diferente velocidad. Quienes hayan estudiado procesamiento del lenguaje saben que los humanos tardamos 30ms en percibir un estímulo sonoro y 80-95ms en procesar un estímulo visual y, sin embargo, cuando vemos caer un vaso al suelo oímos como se rompe a la vez; esto es gracias a los cálculos y predicciones que hace nuestro cerebro.

Pero, como es habitual, programarlo en una máquina no es fácil. Para ello, los *settings* de ultrasonido para fonética usan diferentes sistemas de sincronización en los que una unidad externa se

encarga de que el ordenador reciba prácticamente a la vez la señal de video y audio. Todos ellos son aproximaciones lo más exactas posible, pero ninguno puede ofrecer una sincronización perfecta (al menos en un laboratorio ordinario, con un ordenador común y con la técnica disponible en la actualidad) (Wrench y Scobbie, 2006).

2.5. Controladores y software

El software de análisis del UTI más usado es el Advanced Articulate Analysis (AAA; también útil para analizar datos de EPG) (Figura 3). Este programa permite realizar grabaciones, cambiar los ajustes de grabación del ultrasonido (como la frecuencia de muestreo de la grabación o la profundidad del ultrasonido), anotar los datos y exportarlos en diferentes formatos.

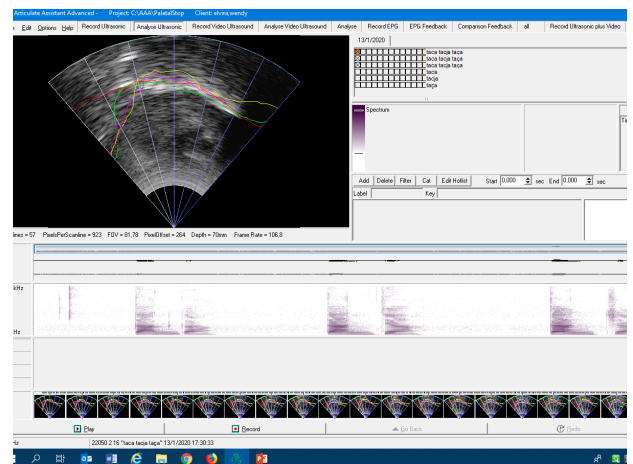


Figura 2. Interfaz del software de grabación y análisis AAA.

Una grabación estándar se realiza usando una profundidad de 70mm y un campo de visión de 90%, con la potencia al mínimo posible para reducir la exposición del informante al ultrasonido y la ganancia al máximo.

Si el objetivo es obtener un video para reproducir a modo de demo (por ejemplo, en una clase de fonética articuladora) es necesario tener en cuenta que las pantallas europeas reproducen un máximo de 24 frames por segundo. Ahora bien, para un análisis estadístico en R, por ejemplo, podemos grabar a mucho más muestreo y captar movimientos que incluso serían imposibles de ver para el ojo humano. La sincronización entre las señales de audio y video se realiza la primera vez que se analiza la grabación. Para ello, el usuario solo necesita ordenar al software que sincronice y él lo hará automáticamente a partir de una pequeña señal que se emite durante la sesión de grabación.

Además, el AAA permite al usuario efectuar la grabación mientras en pantalla aparecen las palabras que tiene que pronunciar el informante.

Con el mismo software se pueden analizar datos mediante la creación de abanicos que delimitan las zonas del paladar. También es en este software donde el investigador deberá reseguir el perfil de la lengua a mano alzada para crear los perfiles de la lengua en los diferentes *frames*. Estos perfiles (llamados *splines*) son los que normalmente se exportan y usan posteriormente para el análisis cuantitativo de los datos. AAA también permite la importación de TextGrids de Praat y el uso de los espectrogramas incluidos en la visualización (v. Figura 2).

3. USOS DEL ULTRASONIDO EN FONÉTICA

3.1. Determinación del punto de articulación

El uso más transparente del UTI es la determinación del punto de articulación de un sonido. Para efectuar ese tipo de análisis, se crean unos abanicos que se superponen a la imagen de la lengua en los que se definen los límites de cada zona del paladar (alveolar, palatal, velar) al modo de los estudios de EPG.

En el caso del catalán, Recasens y Rodríguez (2016, 2017) establecen los límites entre estas zonas a partir de las articulaciones de los siguientes sonidos del propio informante:

- Límite alveolar-palatal: [r]
- Límite velar-palatal [içi]
- Zona velar: 1,26cm o 1,51cm (según sexo, basado en Fitch & Giedd, 1999)

Una vez conseguidos esos puntos de referencia, se realiza la articulación que se quiere estudiar y se mide en el momento central de la articulación el punto donde se produce la máxima constricción. Así se obtendría de manera estática el punto de articulación de un sonido.

3.2. Comparación de dos articulaciones

Cuando el problema que se plantea es si dos articulaciones son iguales o diferentes (por ejemplo, consonante en ataque y en coda), se puede recurrir a un análisis que determina si la forma de la lengua es diferente en las dos condiciones.

Para ello se analizan dos grupos de *splines*¹ de los que se obtiene el perfil medio mediante Smoothing spline SSANOVA (Davidson, 2006). Después, se puede determinar si hay diferencias significativas entre los *splines* que se ajustan mejor (*fit*) a dos conjuntos de datos que se comparan.

La interacción puede ser significativa o no, pero en el caso de que lo sea, queda la duda de qué parte de la lengua ha hecho que la articulación sea diferente. Para saber qué parte (sección de la lengua) ha cambiado, se usan intervalos de confianza bayesianos.

Aunque este es posiblemente uno de los análisis más extendidos, no es el único posible: otros métodos como el Análisis de Componentes principales se han usado para ver, por ejemplo, las diferencias entre [l] y [ɫ] (Turton, 2015).

3.3. Comparación de dos articulaciones en el tiempo

Hasta ahora los análisis que se han planteado son estáticos. Es decir, de cada producción de un sonido se ha elegido el momento central (como una fotografía) y se han comparado sets de fotos. Sin embargo, los datos que proporciona el ultrasonido son de video y, por tanto, dinámicos.

El análisis de datos dinámicos en fonética no es nuevo. De hecho, las trayectorias de los formantes en diptongos lo son, y precisamente para ello se aplicó por primera vez el tipo de análisis que vamos a exponer a continuación: el Generalised Additive Modelling (GAM) (Sóskuthy, 2017). Además de a trayectorias de formantes, estos modelos se han usado en EMA, a partir de lo cual se han aplicado a datos de imagen de ultrasonido. El GAM crea una función para predecir cómo van a cambiar los datos usando como predictores las relaciones entre los datos que varían en el tiempo. Sin embargo, como en todos los modelos, de ahí no se puede obtener un valor de significación “p” directo, sino que hay que efectuar una comparación de dos modelos usando *Likelihood Ratios*.

4. CONCLUSIONES

Este trabajo quiere mostrar el funcionamiento y las posibles aplicaciones de un equipo de UTI al campo de la fonética, con la voluntad de que en un futuro próximo se puedan ver más estudios con UTI realizados en España.

¹ Hay que tener en cuenta que el UTI no permite ver el primer centímetro del ápice de la lengua.

Las técnicas de fonética articuladora, como el UTI, contribuyen revitalizar los laboratorios de fonética como espacios físicos, pero sobre todo permiten un avance significativo en nuestro estado de conocimientos sobre los procesos necesarios para la emisión de sonidos, en condiciones cada vez más naturales, con un coste menor y más fácil interpretación.

5. BIBLIOGRAFÍA

- Alvar, M. (1996). *Manual de dialectología hispánica. El español de España*. Barcelona: Ariel Lingüística.
- Cano, R. (2005). *Historia de la lengua española*. Barcelona: Ariel Lingüística.
- Fernández Planas, A.M. (2000). Estudio electro-palatográfico de la coarticulación vocálica en estructuras VCV en castellano, Tesis doctoral, Universitat de Barcelona.
- Gick, B., Wilson, I. y Derrick, D. (2013). *Articulatory Phonetics*. Wiley: Blackwell.
- Herrero, R. y Jiménez, J. (2013). De la coarticulación a la armonía vocálica en valenciano. En: F. Sánchez Miret, D. Recasens (eds.): *Studies in phonetics, phonology and sound change in Romance* (pp. 43-63). München: Lincom,
- Kelsey, C.A., Woodhouse, R.J. y Minifie, F.D. (1969). Ultrasonic observations of coarticulation in the pharynx. *Journal of the Acoustical Society of America* 46(4), 1016-1018.
- Lloret, M.R. y Jiménez, J. (2009). Un análisis óptimo de la armonía vocálica del andaluz. *Verba. Anuario Galego de Filoloxía*, 36, 293-325.
- Lulich, S. M., Berkson K.H., y de Jong, K. (2018). Acquiring and visualizing 3D/4D ultrasound recordings of tongue motion, *Journal of Phonetics*, 71, 410-424.
- Mackay, I.R.A. (1977). Tenseness in vowels: An ultrasonic study. *Phonetica*, 34, 325- 351.
- Martínez Celdrán, E. y Fernández Planas, A.M. (2001). Propuesta de transcripción para la africada palatal sonora del español, *Estudios de Fonética Experimental*, 11, p. 173-190.
- Mozaffari, M.H., Guan, S., Wen, S., Wang, N., & Lee, W. (2018). Guided Learning of Pronunciation by Visualizing Tongue Articulation in Ultrasound Image Sequences. En: *Proceedings of IEEE 2018*, 1-5.
- Recasens, D. (2019). *Configuració lingual de vocals, consonants i grups de consonants del català: dades d'ultrasons*. Barcelona: Universitat Autònoma de Barcelona.
- Scobbie, J.M., Wrench, A.A., van der Linden, M. (2008). Head-probe stabilisation in ultrasound tongue imaging using a headset to permit natural head movement. En: *Proceedings of the 8th International Seminar on Speech Production*, 373-376.
- Skolnick, M.L., Zagzabski, J.A. y Watkin, K.L. (1975). Two dimensional ultrasonic demonstration of lateral pharyngeal wall movement in real time: A preliminary report. *Cleft Palate Journal*, 12, 299-303.
- Spreafico, L., Pucher, M. y Matosova, A. (2018). UltraFit: A speaker-friendly headset for ultrasound recordings in speech science. En: *Proceedings of Interspeech 2018, 19th Annual Conference of the International Speech Communication Association*, Hyderabad, India, 2-6 September 2018.
- Stone, M. y Davis, E.P. (1995). A head and transducer support system for making ultrasound images of tongue/jaw movement. *The Journal of The Acoustical Society of America*, 98(6), 3107-3112.
- Turton, D. (2015). Determining categoricity in English /l/-darkening: A principal component analysis of ultrasound spline data. En: *Proceedings of ICPhS 2015*.
- Whalen, D.H., Iskarous, K., Tiede, M.K. y Ostry, D.J. (2005). The Haskins Optically Corrected Ultrasound System (HOCUS). *Journal of Speech, Language, and Hearing Research*, 48, 543-553.
- Wilson, I. (2014). Using ultrasound for teaching and researching articulation. *Acoustical Science and Technology*, 35(6), 285-289.
- Wrench, A.A. y Scobbie, J.M. (2006). Spatio-temporal inaccuracies of video-based ultrasound images of the tongue, En: *Proceedings of the 7th International Seminar on Speech Production*. Ubatuba, Brazil, December 2006.
- Zagzebski, J.A. (1975). Ultrasonic measurement of lateral pharyngeal wall motion at two levels in the vocal tract. *Journal of Speech and Hearing Research*, 18, 308-318.

ChaLeT: DETERMINING THE CHANCE LEVEL IN PERCEPTION EXPERIMENTS BY MEANS OF BINOMIAL TESTS

Wendy Elvira-García¹, Paolo Roseano^{2,3}, Assumpció Rost Bagudanch⁴

¹Universidad Nacional de Educación a Distancia, ²Universitat de Barcelona, ³University of South Africa,

⁴Universitat de les Illes Balears

welvira@flog.uned.es, paolo.roseano@ub.edu, assumpcio.rost@uib.es

ABSTRACT

This paper presents *ChaLeT*, an *R* script that can be used to determine statistically whether answers given by listeners in a perception test are above chance level. The script has been designed for perception tests where listeners have to choose between two categories. The script makes use of the binomial test of statistical significance to determine the chance level. It then plots the results of the perception test on a chart where coloured ribbons signal the interval where answers are below the chance level.

Keywords: perception test, chance level, binomial test, *R*

1. INTRODUCTION

In perceptual phonetics, researchers often use identification tests where listeners have to choose between two possible answers. In studies about intonation, for example, listeners can be asked to listen to an utterance (or part of an utterance) and to decide whether it is a statement or a yes-no question. Once the researchers have collected and counted the answers provided by the listeners, they have to decide whether the percentage of listeners that have given an answer is above the chance level or not. This step is crucial for further interpretation of the results.

In spite of being a fundamental methodological decision, the determination of the chance level is not always carried out with a method generally accepted by the scientific community. This paper puts forward a statistical tool that allows the researchers to decide whether the answers obtained in a perception test are significantly above the chance level. The tool is called *ChaLeT* (acronym of Chance Level Test) and consists of an *R* script, which runs binomial tests of statistical significance.

In the following sections, we shall first show that one of the dominant models in perceptual phonetics, the Signal Detection Theory, does not address explicitly the question of chance level (Section 1.1). Section 1.2 sets the specific goals of this paper. In Section 2, we describe the method that we propose.

Namely, we give details about the input that the script needs (Section 2.1), about the statistical test it runs (Section 2.2), about the availability of the script (Section 2.3) and about the output (Section 2.4). In Section 3, we offer an example of how the script can be used in identification tests. Section 4 contains the conclusions and envisages some future developments.

1.1. Chance level in Signal Detection Theory

When facing perceptual phonetics, the important issue is to determine whether a participant may identify a signal as a particular category or not (an identification task) or to establish if he/she can distinguish two or more elements through different matching procedures (a discrimination test). There are different ways to process the results of such experiments, being one of them the Signal Detection Theory model (known as SDT), which was not developed for phonetics use (or at least, not only, since it is employed in many scientific fields) but is widely accepted in analysing phonetic perceptual data (Elman, 1979; Goldinger, 1998; Connell, 2000).

Within this framework, the researcher presents the participants some stimuli, which include both the signal (the object of study) and noise. The participants' responses to these may be hits (correct identification of the signal), misses (omission of the signal), false alarms (incorrect identification of the noise as the signal) or correct rejections (correct

detection of the noise). SDT is concerned with the hits and the false alarms to set an adequate panorama of the perception process. From these data, it can provide two basic parameters: the sensitivity of the signal (its strength) and the strategy of the participant in adopting a decision. Their combination allows displaying a picture for the strength of a signal and its degree of perceptiveness by the listeners (Wickens, 2001; Macmillan & Creelman, 2005).

Although this model provides very valuable information, it offers no clear and direct information about chance level. Of course, one of the representations of the results within SDT, the ROC curves, show in a graphic way the chance threshold, which is assumed to be at 50%. Nevertheless, presuming that chance level implies only this limit is problematic: why 50% of hits corresponds to decisions made by chance and not 49% or 55%? This seem to be a gap in the model, which lacks an essential complementary figure to the great amount of details it can provide.

The lack of a methodological consensus is even more evident if one reviews recent studies in the field. For example, if we focus on studies that present the results of identification tests (like Ladd & Morton, 1997; Remijnsen & van Heuven, 1999; Post, 2000; Chen, 2003; Schneider & Linfert, 2003; Cummins, Doherty, & Dilley, 2006; Falé & Hub Faria, 2006; Schneider, Dogil, & Möbius, 2009; Dilley, 2010; Vanrell et al., 2013; Vanrell, Armstrong & Prieto, 2017, among several others), we observe that the authors have taken different methodological decisions. In some cases, chance level is not considered. In other cases, it is established in non-statistical ways. Other studies (e.g. Roseano et al., 2015) use statistics, but do not use the most appropriate test.

1.2. Goals

The general goal of this paper is helping researchers who run perception tests to establish when the results obtained pass chance level. Our specific objectives are 1) putting forward a standard test which can be used to determine chance level in tests where listeners have to choose between two categorical answers, 2) visualizing the results of such test in a clear and easily interpretable way.

2. METHOD

2.1. Input

The results of the perception test must be collected in an Excel sheet, which will serve as input for the script that runs statistical analysis. Table 1 shows the format of the Excel sheet, and a template is available online and can be downloaded from the webpage https://github.com/wendyelviragarcia/chance_level_s_for_perceptual_tests.

It has to be pointed out that the input must be the *number* of answers given, not the *percentage*, since the statistical test we use (Section 2.2) is sensitive to the size of the sample. Another important requirement is that the number of answers is the same for all stimuli.

Table 1: Expected input of the pipeline.

Stimuli	Answer	nAnswers
1	Question identified as question	90
2	Question identified as question	91
3	Question identified as question	85
4	Question identified as question	77
5	Question identified as question	57
6	Question identified as question	47
7	Question identified as question	38
8	Question identified as question	20
9	Question identified as question	18
10	Question identified as question	20
11	Question identified as question	9
1	Statement identif. as statement	6
2	Statement identif. as statement	11
3	Statement identif. as statement	7
4	Statement identif. as statement	24
5	Statement identif. as statement	46
6	Statement identif. as statement	65
7	Statement identif. as statement	78
8	Statement identif. as statement	84
9	Statement identif. as statement	85
10	Statement identif. as statement	90
11	Statement identif. as statement	91

2.2. Statistical test

In order to determine the chance level, we consider every question asked to a judge as a Bernoulli trial (also known as binomial trial) (Papoulis, 1984), which in statistics is a random experiment with exactly two possible outcomes, and where the probability of each outcome is the same every time the experiment

is conducted. The result of a sequence of Bernoulli trials is called Bernoulli process. The expected outcome of a Bernoulli process is a Bernoulli distribution, which is a special case of a binomial distribution. The binomial distribution, on its turn, is the basis for the *binomial test* of statistical significance, which is the standard test used when the null hypothesis is that two categories are equally likely to occur (such as “head” and “tail” in a coin toss).

The classical statistical example for which a binomial test is used is determining whether a coin used for tossing is fair. If the coin is fair, the probability of getting heads is 50%. If I toss the coin 100 times and I get 100 “heads”, the coin is evidently not fair. The same holds if I get 98 heads, or 97, 96, 95... Nevertheless, if I toss the coin 100 times and I get 51 heads and 49 tails, it does not necessarily imply that the coin has been tampered. In other words, if I get 51 heads and 49 tails, the result is still within the chance level. The question, therefore, is: how many heads/tails do I have to get in order to be sure that the coin is not fair?

This is the same question we were asking ourselves in Section 1.1. In perception tests, we need to know how many “correct” answers we need to get in order to be sure that listeners do not answer random.

The *binomial test* gives an answer to such questions. We set the significance level at 1% or 5%, we run a binomial test, and if the p-value we get is below .05, we reject the null hypothesis and we conclude that the coin is not fair. In perception tests, we would conclude that answers are above chance level.

2.3. ChaLeT

We designed an R (R Core Team, 2019) script called *ChaLeT* (Chance Level Test), which is available online (https://github.com/wendylviragarcia/chance_levels_for_perceptual_tests) and that implements a binomial test (2.2) on the results of an identification test (2.1).

2.4. Output

The output of *ChaLeT* is a chart like the ones commonly used to plot the result of identification tests (like those in Figures 3 and 4 below). On the horizontal axis, one finds the stimuli and on the vertical axis the number of “correct” answers. In addition to this traditional elements of the plot,

ChaLeT draws two coloured ribbons which signal the interval where answers are below the chance level according to binomial tests with a significance level of $p < .01$ (outer mauve ribbon) or $p < .05$ (inner plum-coloured ribbon) (Figure 5). If the points corresponding to a stimulus fall within this ribbon, judges’ identification is below chance level. Researchers will then need to draw the phonological consequences of this.

3. EXAMPLIFICATION

In order to give an example of how *ChaLeT* can help the researchers interpret their results, in this section, we replicate in a simplified way an identification test presented in a well-known study about Catalan intonation (Vanrell et al., 2013) and we apply *ChaLeT* to the results.

Vanrell et al. (2013) show that Majorcan Catalan listeners distinguish information and confirmation seeking questions by means of two distinct nuclear falling pitch accents, which are represented schematically in Figure 1.

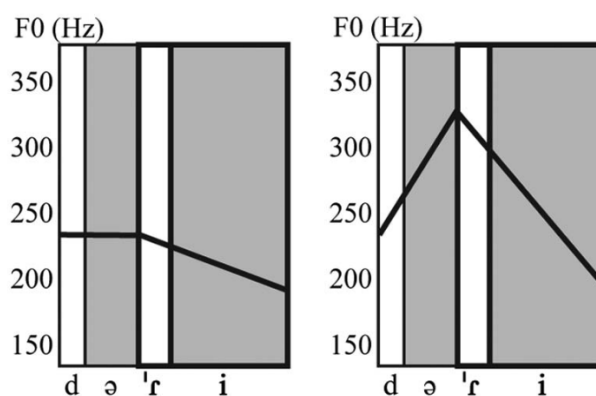


Figure 1: Simplified F0 contours of nuclear pitch accents of a confirmation-seeking (left) and an information-seeking yes-no question (right) (Source: Vanrell et al., 2013).

In order to carry out an identification test, the researchers manipulated F0 to create a continuum between the two stimuli. Figure 2 represents the original contours (solid lines) and contours obtained by means of manipulation in the two directions (dotted lines). In this case, the total number of steps was 11 and the distance between each step and the following was always the same.

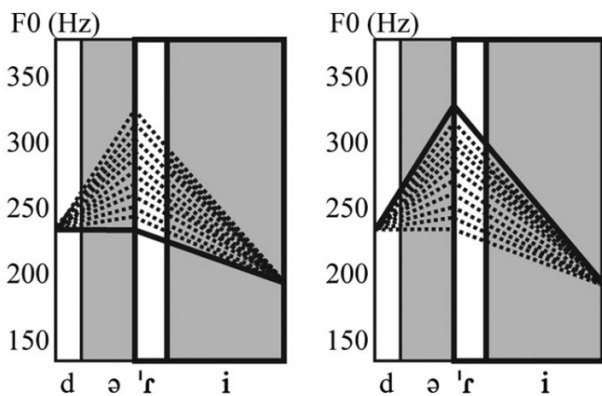


Figure 2: Simplified F0 contours of pitch accents obtained from a confirmation-seeking (left) and an information-seeking yes-no question (right) (Source: Vanrell et al., 2013).

In the identification test, listeners listened to all 22 stimuli and had to relate each stimulus to one of two categories (i.e. they had to decide if what they listened to was a confirmation-seeking or an information-seeking yes-no question). The results of the perception test were plotted on a diagram (Figures 3) where the horizontal axis contained the stimuli. On the vertical axis the researchers represented the percentage of “correct” answers, where “correct” means that “the judge considers that the stimulus corresponds to the category of the recording used to create the stimulus the judge hears”.

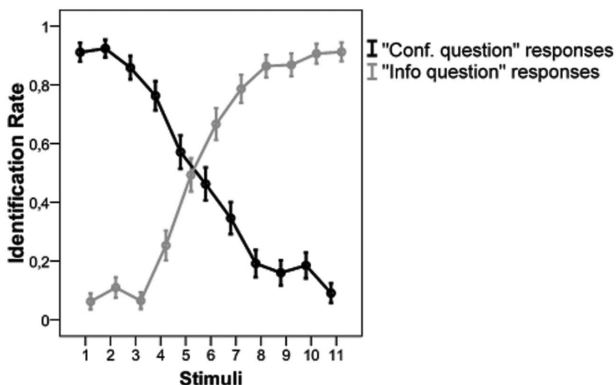


Figure 3: Results for the identification presented in Vanrell et al. (2013) (Source: Vanrell et al., 2013).

If we run *ChaLeT* with a dataset (Table 1) that approximately corresponds to the results obtained in the identification test presented above, we obtain a graph like the one in Figure 4. The inner coloured ribbons indicates the fringes where the *binomial test* says that answers are not different from chance level, at different significance levels (see Section 2.4).

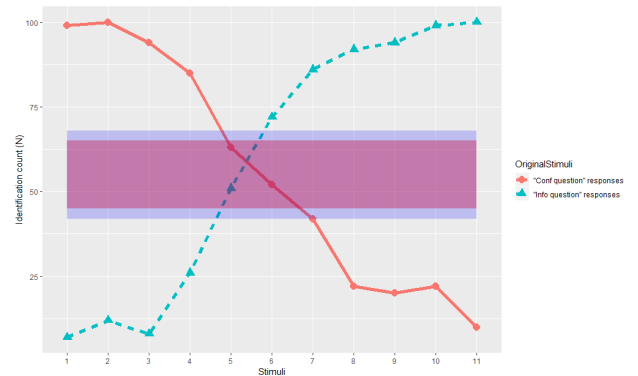


Figure 4: Results for an identification test similar to Vanrell et al.’s (2013) plotted by means of *ChaLeT*, with coloured ribbons marking chance level.

In comparison with the graphs traditionally used in perception studies (like the one in Figure 3), the graph in Figure 4 adds an important piece of information: it tells the researchers for which stimuli perception is at chance level. It goes without saying that this piece of information is relevant to draw conclusions about perception.

4. CONCLUSIONS

This paper describes a method for establishing the chance level in a perception test where listeners have to choose between two categories. The method is based on binomial test and is implemented by means of an *R* script called *ChaLeT*. The output of the script is a graph that represents the results of the identification task and marks chance level graphically by means of coloured ribbons.

Future versions of *ChaLeT* will need to use multinomial tests, so that the tool can be employed in perception tests with more than two possible answers.

5. REFERENCES

- Chen, A. (2003). Reaction time as an indicator of discrete intonational contrasts in English. *Proceedings of Eurospeech*, Geneva, 97–100.
- Connell, B. 2000. The perception of lexical tone in Mambila, *Language & Speech*, 43, 2163–2182.
- Cummins, F., Doherty, C., & Dille, L. (2006). Phrase-final pitch discrimination in English. In R. Hoffmann & H. Mixdorff (Eds.), *Proceedings of Speech Prosody* (pp. 5467–5470). Dresden: TUDpress Verlag der Wissenschaften.
- Dille, L. C. (2010). Pitch range variation in English tonal contrasts: Continuous or categorical?, *Phonetica*, 67, 63–81.
- Elman, J. (1979). Perceptual origins of the phoneme boundary effect and selective adaptation to speech: A

- signal detection theory analysis, *The Journal of the Acoustical Society of America*, 65, 190–207.
- Falé, I., & Hub Faria, I. (2006). Categorical perception of intonational contrasts in European Portuguese. In R. Hoffmann & H. Mixdorff (Eds.), *Proceedings of Speech Prosody* (pp. 69–72). Dresden: TUDpress Verlag der Wissenschaften.
- Goldinger, S. D. (1998). Signal detection comparisons of phonemic and phonetic priming: The flexible-bias problem, *Perception & Psychophysics*, 60, 952–965.
- Ladd, D. R., & Morton, R. (1997). The perception of intonational emphasis: Continuous or categorical?, *Journal of Phonetics*, 25, 313–342.
- Macmillan, N. A. & Douglas Creelman, C. (2005). *Detection Theory: A User's Guide*. London: Erlbaum.
- Papoulis, A. (1984). Bernoulli Trials. In A. Papoulis & S. Unnikrishna Pillai (Eds.), *Probability, Random Variables, and Stochastic Processes* (pp. 57–63). New York: McGraw-Hill.
- Post, B. (2000). *Tonal and Phrasal Structures in French Intonation*. The Hague: Holland Academic Graphics.
- R Core Team (2019). *R: A language and environment for statistical computing* (version 3.6.2). Retrieved from <http://www.R-project.org/>
- Remijsen, B., & van Heuven, V. (1999). Categorical pitch dimensions in Dutch: Diagnostic test. In J. J. Ohala, Y. Hasagawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), *Proceedings of the XIVth International Congress of Phonetic Sciences* (pp. 1865–1868). San Francisco: University of California.
- Roseano, P., Fernández Planas, A. M., Elvira-García, W., Cerdà Massó, R., & Martínez Celdrán, E. (2015). Diferencias perceptivas entre los acentos tonales prenucleares en catalán. In A. Cabedo Nebot (Ed.), *Perspectivas actuales en el análisis fónico del habla: tradición y avances en la fonética experimental*, (pp. 163–173). València: Universitat de València.
- Schneider, K., & Linfert, B. (2003). Categorical perception of boundary tones in German. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the XVth International Congress of Phonetic Sciences* (pp. 631–634). Barcelona: Causal Productions.
- Schneider, K., Dogil, G., & Möbius, B. (2009). German boundary tones show categorical perception and a perceptual magnet effect when presented in different contexts. *Proceedings of Interspeech* (pp. 2519–2522). Brighton, UK, September 6–10.
- Vanrell, M. M., Mascaró, I., Torres-Tamarit, F., & Prieto, P. (2013). Intonation as an encoder of speaker's certainty: Information and confirmation yes-no questions in Catalan. *Language and Speech*, 56(2), 163–190.
- Vanrell, M. M., Armstrong, M., & Prieto, P. (2017). Experimental evidence for the role of intonation in evidential marking. *Language and Speech*, 60 (2): 242–259.
- Wickens, T. D. (2001). *Elementary Signal Detection Theory*. Oxford: Oxford University Press.

USO DE SISTEMAS DE CLASIFICACIÓN SUPERVISADA PARA LA EVALUACIÓN DE LA PROSODIA: APLICACIONES AL HABLA DE PERSONAS CON SÍNDROME DE DOWN

David Escudero¹, Mario Corrales-Astorgano¹, Yolanda Martín de San Pablo⁴, Alfonso Rodríguez de Rojas⁵, Valle Flores², César González-Ferreras¹, Valentín Cardeñoso¹, Lourdes Aguilar³

Departamento de Informática de la Universidad de Valladolid¹, Departamento de Psicología de la Universidad de Valladolid², Spanish Philology Department de la Universitat Autònoma de Barcelona, Fundación Personas de Valladolid⁴, Asociación Down de Valladolid⁵
descuder@infor.uva.es

ABSTRACT

El entrenamiento de las competencias de producción prosódica es una necesidad para personas con síndrome de Down (SD). Se dispone de un videojuego educativo para el entrenamiento de la producción oral adaptado a personas con SD. Para ofrecer la posibilidad de juego autónomo, se ha desarrollado un módulo automático de predicción. En este trabajo se presenta la arquitectura de dicho sistema de evaluación automática poniendo el acento en las variables acústicas que se analizan y en el contraste del rendimiento del sistema automático empleado a la hora de predecir diversas dimensiones prosódicas.

Training prosody is a need for people with Down syndrome. Our team has developed a video game aimed to improve the prosody of persons with DS but it was not designed for use as a self-learning resource. To offer the possibility of autonomous play, this study presents the advances achieved to build a component for the automatic prediction of prosodic quality. We present the system architecture with emphasis on the selection of the analyzed acoustic variables and the prosodic aspects of interest considering the target audience of the game. We contrast the performance of the system for predicting different prosodic dimensions.

Keywords: evaluation of prosodic quality, automatic assessment of prosodic quality, supervised classification, down syndrome voice, computer assisted pronunciation training

1. INTRODUCCIÓN

La evaluación de la prosodia es un ámbito de interés que, abordando la complejidad de la prosodia desde perspectivas enriquecedoras, puede ofrecer soluciones a necesidades reales en el ámbito de la enseñanza de la pronunciación asistida por ordenador, tanto para el aprendizaje de idiomas como para el tratamiento del habla patológica. El trabajo que aquí se presenta se enmarca dentro del proyecto Protoaucal (Ministerio de Ciencia, Innovación y Universidades and the European Regional Development Fund FEDER TIN2017-88858-C2-1-R) para el desarrollo de un módulo de evaluación de la calidad prosódica que permita el uso autónomo de un videojuego para la mejora de la comunicación oral por parte de hablantes con síndrome de Down, un

colectivo con dificultades en producción y percepción prosódicas (Stojanovik, 2011; O'Leary, 2020).

El enfoque consiste en entrenar clasificadores supervisados (entrenados con muestras que han sido previamente clasificadas por expertos, Bishop, 2006) que se especializan en aspectos relacionados con la calidad de la prosodia previamente definidos en una rúbrica de evaluación y que comparten los etiquetadores del corpus.

La sección 2 presenta el videojuego en el que se integra el módulo de evaluación automática y la arquitectura del nuevo sistema; la sección 3, la información sobre las variables de calidad y los parámetros acústicos. En la sección 4 se comparan los distintos clasificadores y se muestran las tasas de clasificación, para acabar en la sección 5 con una discusión sobre las limitaciones de la aproximación.

2. EVOLUCIÓN DEL VIDEOJUEGO PARA EL ENTRENAMIENTO AUTÓNOMO

2.1. Descripción del juego

El potencial de los videojuegos para el entrenamiento de personas con discapacidad es indiscutible, según demuestran estudios como los de Prena (2018), Cano (2019). Como resultado de la labor investigadora del equipo, se dispone de un videojuego educativo para la práctica de la comunicación oral (pragmática y prosodia principalmente) de personas con SD desarrollado en el marco de dos proyectos de investigación financiados por Recercaixa-ACUP y BBVA. Su novedad es que el jugador debe superar una serie de desafíos y misiones en una aventura desarrollada en un entorno de juego estimulante mediante el uso adecuado de los rasgos prosódicos. Estos hitos se ubican en un entorno social de situaciones cotidianas que favorecen la mejora de la competencia comunicativa (González-Ferreras, 2017; Aguilar, 2019).

El videojuego ha demostrado su utilidad al ser capaz de motivar a los usuarios en la realización de ejercicios en compañía de un terapeuta, profesor o familiar. También ha revelado su potencial para recoger corpus de voz de un colectivo de usuarios particularmente difícil. Durante las sesiones de juego se almacena automáticamente tanto información sobre la interacción del usuario con el juego y los resultados en las diferentes actividades como todos los enunciados producidos durante las actividades en que se le pide al jugador que construya, repita o lea una frase. Las grabaciones realizadas durante las sesiones de juego permiten construir un corpus de voz para el estudio de la prosodia en personas con síndrome de Down que ha sido utilizado en varios trabajos previos, como en Corrales-Astorgano (2018), para caracterizar el habla de personas con síndrome de Down.

2.2. Arquitectura para el juego autónomo

En la versión existente del videojuego, un adulto (típicamente el profesor, terapeuta o familiar) se sienta junto al jugador y decide, mediante un teclado accesorio, la corrección o incorrección de los enunciados, de manera que pueda continuar o deba repetir las actividades. Estos juicios se han empleado como medidas de calidad de las producciones orales en trabajos previos como en Corrales-Astorgano (2019), donde se concluye que los juicios de los expertos dependen en gran medida de cada uno de los individuos con SD, tanto de sus capacidades cognitivas como de su estado anímico. A modo de ejemplo, el terapeuta puede permitir continuar en el

juego a alguno de los jugadores simplemente para no provocar situaciones de estrés. En cambio, en otras ocasiones, el terapeuta pide que el jugador repita una actividad si cree que el jugador tiene suficiente habilidad para hacerlo mejor. En una evaluación automática, los factores externos implicados en el desarrollo del juego (nivel de frustración, entre otros) deben dejarse de lado en beneficio del examen de las variables prosódicas.

En una nueva versión del juego, se pretende incluir un módulo de evaluación automática de la calidad del habla para permitir el juego autónomo, de modo que los jugadores puedan practicar la prosodia de forma autónoma empleando sus dispositivos móviles, a la vez que su trabajo queda monitorizado y puede ser supervisado fuera de línea por el terapeuta.

Para disponer de datos enriquecidos sobre los juicios de los expertos, hemos realizado una nueva campaña de grabación en la que los evaluadores no sólo toman la decisión de dejar continuar o no al usuario en la partida, sino que también evalúan con una rúbrica consensuada previamente la calidad de los turnos orales. El sistema automático aprovecha esta información para, además de emitir juicios sobre si el usuario debe o no repetir la actividad, informar sobre aquellos aspectos que el usuario está realizando peor.

3. DESARROLLO DEL MÓDULO DE EVALUACIÓN PROSÓDICA

La figura 1 muestra el esquema de clasificación supervisada seguido. En una primera fase, se debe configurar un corpus de entrenamiento que incluya, junto con las muestras de audio, los juicios de calidad emitidos por los expertos. En una segunda fase de entrenamiento, el clasificador analiza las muestras de entrenamiento para especializarse en imitar los juicios de los expertos. En la etapa de test, el clasificador puede ser empleado para hacer el trabajo de los expertos.

En trabajos previos, (Corrales-Astorgano, 2019), hemos utilizado tres tipos diferentes de clasificadores: las redes neuronales, los árboles de decisión y las máquinas de soporte vectorial. Se trata de tres tecnologías que emplean enfoques diferentes para ajustar los parámetros de clasificación dando resultados que pueden ser complementarios facilitando enfoques posteriores de fusión de expertos. Seguimos aquí un enfoque similar, pero en este caso empleamos los clasificadores en el análisis de los juicios de calidad obtenidos en las rúbricas de evaluación.

En este trabajo, los clasificadores se especializan en la decisión de dejar continuar o no al

usuario en el juego y también en el establecimiento de un juicio crítico sobre la calidad de la locución. El uso práctico de este segundo juicio es, por un lado, enriquecer la realimentación que el sistema puede dar al usuario, y de otro lado generar informes para su análisis a posteriori por parte de los expertos.

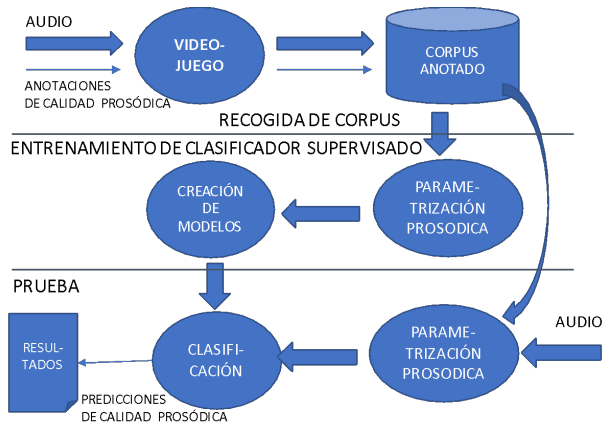


Figura 1: Esquema de entrenamiento de sistemas de clasificación supervisados.

3.1. Variables de calidad

Con el fin de disponer de un procedimiento de evaluación de la calidad oral de los enunciados que permita desarrollar el módulo automático de predicción, se ha acordado con los agentes educativos un mapa de criterios de evaluación. Una vez definidos, puede procederse a la recogida de datos para el corpus de entrenamiento. Mediante un dispositivo hardware, el terapeuta valora, en tiempo de juego, las siguientes variables relacionadas con la calidad de la producción oral en general y prosódica en particular:

Inteligibilidad: valora la capacidad del evaluador para transcribir la locución en un mensaje.

Adecuación: indica si el mensaje emitido por el jugador es adecuado al contexto comunicativo del juego.

Omisiones: número de palabras que el locutor ha omitido. Es necesario porque las incorrecciones gramaticales son frecuentes.

Fluidez: mide si la locución contiene puntos de interrupción inapropiados, repeticiones y disfluencias.

Velocidad: Evalúa la adecuación de la velocidad del habla y penaliza las producciones orales demasiado lentas.

Curva melódica: penaliza construcciones poco expresivas o con una entonación inapropiada.

Continuar: indica si la calidad de la producción es lo suficientemente buena para continuar con el juego o, por el contrario, tiene que repetir la actividad.

Inteligibilidad, adecuación y continuar son variables binarias (SI/NO). Para el resto de variables, se emplea una escala Likert de 1 a 4 y se asignan valores sólo cuando la producción oral se considera inteligible y correcta.

3.2. Parámetros acústicos

Los clasificadores utilizan un conjunto reducido de parámetros prosódicos extraídos empleando el extended Geneva Minimalistic Acoustic Parameter Set (eGeMAPS), descrito en Eyben et al. (2015) que obtiene con la herramienta openSmile (Eyben et al., 2013). En particular, estos parámetros están relacionados con F0: frecuencia fundamental y jitter; con la energía: intensidad y shimmer; y con la dimensión temporal: tasa de picos de intensidad por segundo, longitud media y desviación estándar de los segmentos sonoros y no sonoros y la tasa de segmentos sonoros. Se eliminan en este estudio las variables espectrales empleadas en Corrales-Astorgano (2018). Para cada coeficiente se calcula la media y el coeficiente de variación. Para la frecuencia fundamental y la intensidad se mide el percentil 20, 50 y 80, la media y desviación estándar de la pendiente ascendente o descendente, la media global y la diferencia entre el percentil 80 y el 20. A mayores, se incluye el porcentaje de silencio dentro de la locución, la tasa de pausas por segundo y la duración media de las pausas. En total, se utilizan 34 características.

4. RESULTADOS

Empleamos los datos recogidos en la última campaña de evaluación realizada durante el curso 2018/19 en el colegio Tórtola y en la Asociación Síndrome Down Valladolid. Son 601 frases evaluadas de 17 locutores diferentes. Dos profesores anotaron las variables prosódicas. La tasa de consistencia, entre los etiquetadores va desde 0.33 a 0.9, usando Kendall (1990), dependiendo de la dimensión analizada. La tabla 1 muestra los resultados de clasificación obtenidos para cada una de las variables de clasificación, excepto la inteligibilidad y la adecuación, que no se pueden medir utilizando parámetros acústicos:

Tabla 1. Tasa de clasificación (TC) y porcentaje de falsos negativos (FN) por cada dimensión y clasificador. Las muestras en las que no se ha realizado la evaluación se han eliminado de este estudio. DT significa Decision Trees, SVM, Support Vector Machines y MLP, Multilayer Perceptron.

	DT		SVM		MLP	
Variable	TC	FN	TC	FN	TC	FN
Fluidez	0.63	0.46	0.69	0.41	0.64	0.42
Velocidad	0.68	0.25	0.7	0.27	0.65	0.35
Curva	<i>0.61</i>	0.15	<i>0.61</i>	0.17	<i>0.62</i>	0.3
Continuar	0.69	0.14	0.73	0.02	0.69	0.18

Los mejores resultados de clasificación se obtienen utilizando el clasificador SVM, salvo en el caso de la curva melódica, en la que los resultados son similares en los tres clasificadores. Con respecto a la tasa de falsos negativos utilizando el clasificador SVM, las tasas más bajas se obtienen para las dimensiones de continuar (0.02) y curva (0.17). Las tasas de falsos negativos son más altas en las dimensiones de velocidad (0.27) y fluidez (0.41).

5. DISCUSIÓN

El uso de métodos de clasificación supervisada para predecir la calidad prosódica permiten alcanzar altas tasas con un número de variables relativamente pequeño. En próximos desarrollos se pondrá en evidencia si esta tasa es lo suficientemente alta para el propósito que se persigue, que no es otro que el de permitir el juego autónomo de los usuarios. Hemos puesto el acento en presentar bajas tasas de falsos negativos porque pueden provocar en el usuario una sensación de frustración deteriorando su predisposición hacia el entrenamiento con el videojuego.

Las tasas de predicción de las variables son útiles para ofrecer indicadores a los terapeutas sobre el rendimiento de los alumnos y para ser relacionadas con configuraciones anómalas en la producción oral, por ejemplo, alargamiento excesivo de pausas, o monotonía en los contornos de F0.

Apuntamos los siguientes aspectos que deben considerarse en trabajos futuros que pretendan mejorar las tasas de predicción:

1. Usar corpus con un número elevado de grabaciones y con categorías equilibradas. Se trata de un punto débil importante, porque disponer de un corpus de habla mayor de hablantes con discapacidad intelectual es muy complicado, como se discute en Corrales-Astorgano (2018).
2. Asegurar en la medida de lo posible la consistencia de los juicios de los expertos que evalúan la calidad prosódica. Evaluar prosodia siempre es difícil porque un mismo mensaje puede emitirse con diferentes realizaciones prosódicas, todas diferentes y correctas.

3. Tener en cuenta la heterogeneidad de los locutores. En Corrales-Astorgano (2019) mostramos que el impacto del locutor es importante por la gran diversidad de perfiles que se encuentran en el colectivo con síndrome de Down.

4. Emplear múltiples dominios de evaluación para focalizar el entrenamiento en uno de los ámbitos con el fin de lograr mayor efectividad.

6. CONCLUSIONES

Se ha presentado un trabajo de implementación de un módulo de predicción de calidad prosódica que se va a incluir en un videojuego educativo para el entrenamiento de la producción oral, en general, y prosódica, en particular, por parte de usuarios con síndrome de Down. La precisión del clasificador valora su futura utilidad en este dominio con la inclusión de sesgos para penalizar la presencia de falsos negativos. Es trabajo futuro estudiar la relación entre las tasas de predicción de las variables de calidad prosódica y las posibles configuraciones de parámetros acústicos anómalas. Se destacan como puntos clave para mejorar las tasas de predicción el uso de corpus grandes equilibrados con anotaciones fiables sobre múltiples dimensiones y la consideración como variable de la elevada diversidad de los informantes.

6. BIBLIOGRAFÍA

- Aguilar, L. (2019). Learning Prosody in a Video Game-Based Learning Approach. *Multimodal Technologies and Interaction*, 3(3), 51.
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- Cano, A. R., García-Tejedor, Á. J., Alonso-Fernández, C., & Fernández-Manjón, B. (2019). Game Analytics Evidence-Based Evaluation of a Learning Game for Intellectual Disabled Users. *IEEE Access*, 7, 123820-123829.
- Corrales-Astorgano, M., Martínez-Castilla, P., Escudero-Mancebo, D., Aguilar, L., González-Ferreras, C., & Cardeñoso-Payo, V. (2019). Automatic Assessment of Prosodic Quality in Down Syndrome: an Analysis of the Impact of Speaker Heterogeneity. *Applied Sciences*, 9(7), 1440.
- Corrales-Astorgano, M., Escudero-Mancebo, D., & González-Ferreras, C. (2018). Acoustic characterization and perceptual analysis of the relative importance of prosody in speech of people with Down syndrome. *Speech Communication*, 99, 90-100.
- Eyben, F., Scherer, K. R., Schuller, B. W., Sundberg, J., André, E., Busso, C., ... & Truong, K. P. (2015). The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. *IEEE transactions on affective computing*, 7(2), 190-202.
- Eyben, F., Weninger, F., Gross, F., & Schuller, B. (2013, October). Recent developments in opensmile, the munich open-source multimedia feature extractor. In *Proceedings of the 21st ACM international conference on Multimedia* (pp. 835-838).
- González-Ferreras, C., Escudero-Mancebo, D., Corrales-Astorgano, M., Aguilar-Cuevas, L., & Flores-Lucas, V. (2017). Engaging adolescents with Down syndrome in an educational video game. *International Journal of Human-Computer Interaction*, 33(9), 693-712.
- Kendall, M. G., & Gibbons, J. D. (1990). Rank correlation methods. New York, NY : Oxford University Press
- O’Leary, D., Lee, A., O’Toole, C., & Gibbon, F. (2020). Perceptual and acoustic evaluation of speech production in Down syndrome: A case series. *Clinical linguistics & phonetics*, 34(1-2), 72-91.
- Prena, K., & Sherry, J. L. (2018). Parental perspectives on video game genre preferences and motivations of children with Down syndrome. *Journal of Enabling Technologies*.
- Stojanovik, V. Prosodic deficits in children with Down syndrome. *J. Neurolinguist.* 2011, 24, 145–155.

UNA HERRAMIENTA DE FONÉTICA APLICADA: MATERIALES DE LOGOAUDIOMETRÍAS INFANTILES EN CATALÁN

Ana Ma. Fernández Planas¹, Paolo Roseano^{1,2}, Wendy Elvira García³, Silvia Marro⁴, Pilar Calvo⁵, M. Antonia Claveria⁶

¹Universitat de Barcelona, ²University of South Africa, ³Universidad Nacional de Educación a Distancia, ⁴Equipo de Asesoramiento y Orientación Psicopedagógica (EAP) B-34 Sabadell-Sant Quirze, ⁵CREDA Jordi Perelló, ⁶Hospital Sant Joan de Déu

anamariafernandez@ub.edu, paolo.roseano@ub.edu, welvira@flog.uned.es, smarro@xtec.cat, pcalvo@xtec.cat, maclaveria@sjdhospitalbarcelona.org

ABSTRACT

Este trabajo, que se desarrolla en el marco de la fonética clínica, presenta una herramienta recientemente publicada para trabajar logoaudiometrías infantiles en catalán oriental central con niños de diferentes rangos etarios que sufren hipoacusia. El material es relevante porque viene a llenar un hueco en el trabajo diario de otorrinos y audioprotesistas para trabajar con pacientes de corta edad hablantes nativos de la variedad oriental central del catalán.

Palabras clave: logoaudiometrías, logopedia, hipoacusia infantil, catalán oriental central

This paper belongs to the field of clinical phonetics and presents a set of materials for speech therapies for children who speak Central Eastern Catalan. The materials are designed for children of different age ranges with hearing loss. The set of materials is relevant because it fills a gap in the daily work of otolaryngologists and hearing care professionals who work with young patients who are native speakers of the Central Eastern variety of Catalan.

Keywords: speech audiometry, speech therapy, childhood hearing impairment, Central Catalan

1. INTRODUCCIÓN

La incidencia de la hipoacusia infantil susceptible de tratamiento es de 3/1000 niños, mientras que la hipoacusia severa y profunda afecta a 1/1000 niños (Departament de Salut, 2010). La presencia de una hipoacusia en edades tempranas puede provocar una alteración en el procesamiento frecuencial y temporal del estímulo sonoro, que provoca a su vez una percepción alterada en el reconocimiento de palabras. Este conjunto de factores puede, en última instancia, afectar negativamente al desarrollo del habla y el lenguaje.

En niños con hipoacusia severa y profunda, las prótesis auditivas compensan la pérdida auditiva, pero no restauran la audición normal. Es por ello que la mayoría de niños con prótesis auditivas requieren de un proceso de rehabilitación auditiva, lingüística y comunicativa. Esta rehabilitación consiste en

recodificar las informaciones que provienen de las señales acústicas y/o electroacústicas de las prótesis auditivas. Las tareas de rehabilitación auditiva, comunicativa y lingüística tienen que empezar en una edad muy temprana, después del diagnóstico de la hipoacusia, ya que actuar en el período crítico hace ganar tiempo en el proceso habilitador (Salguero Santana et ál., 2015). Un elemento clave del proceso de rehabilitación es la realización de evaluaciones de la capacidad lingüística de los niños.

El objetivo general de este trabajo es presentar una herramienta en forma de libro dirigida a la realización de dichas evaluaciones de la capacidad lingüística. En concreto, se quiere dar a conocer una herramienta de valoración del rendimiento protésico verbal en el campo de la audiología infantil adaptada al catalán (Fernández Planas et ál. 2017).

La novedad de la herramienta en cuestión respecto a otros materiales radica en cinco puntos: 1) está diseñada para trabajar población infantil; 2)

se centra en una variante dialectal del catalán, la oriental central; 3) contiene materiales de trabajo material fonéticamente equilibrados; 4) incluye diferentes tipos de materiales auditivos y visuales; y 5) incorpora unos casos prácticos que permiten ilustrar su uso.

2. PRUEBAS DE VALORACIÓN AUDIOLÓGICA

Hoy en día los profesionales cuentan con varias herramientas a su disposición para evaluar el lenguaje. Pruebas como el PLON-R, el ITPA, el MacArthur/ELI (Kirk et ál., 1984; Fenson, 2002; Aguinaga, et ál., 2004) se han convertido en una parte esencial de la valoración del lenguaje tanto en niños neurotípicos como con sordera trastornos del lenguaje (Wiig et al. 2013). Además, pruebas como el CELF-5 están diseñadas para detectar y diagnosticar trastornos del lenguaje. Y la prueba fonológica de Bosch (Bosch, 1983) se centra especialmente en el plano fónico.

En el caso de la valoración y el seguimiento audiológico estas pruebas deben ser de dos tipos: pruebas audiológicas objetivas y valoraciones audiológicas subjetivas. Las pruebas objetivas utilizadas para la valoración audiológica son: 1) electrofisiológicas y 2) no electrofisiológicas. Entre las electrofisiológicas se encuentran: a) PEATC (Potenciales Evocados Auditivos de Tronco Cerebral); b) PEAee (Potenciales Evocados Auditivos de estado estable); c) OEA (Otoemisiones Acústicas Transitorias); d) PD (Otoemisiones acústicas para productos de Distorsión); y f) electrococleografía. Entre las no-electrofisiológicas, podemos recordar la impedanciometría, el reflejo estapedial y la timpanometría.

Las pruebas subjetivas utilizadas para complementar la valoración audiológica objetiva – sobre las que se volverá más detenidamente en el § 2.1– son las siguientes: 1) la audiometría tonal observacional o de distracción; 2) la audiometría tonal de respuesta de orientación condicionada (VRA); 3) la audiometría tonal de participación o lúdica; y 4) la audiometría verbal.

Es en estas pruebas subjetivas –y, sobre todo, en las logaudiometrías– donde los lingüistas participan especialmente en el diseño, puesto que en ellas la evaluación se realiza con estímulos lingüísticos que requieren un trabajo interdisciplinar entre diferentes profesionales: 1) médicos ORL infantiles; 2) lingüistas –concretamente fonetista y fonólogos–; y 3) profesionales especializados en logopedia y audiología infantil.

2.1. Las audiometrías

Tal y como se ha expuesto anteriormente, existen dos tipos de audiometrías: las audiometrías tonales (§ 2.1.1) y las audiometrías verbales, también conocidas como logaudiometrías (§ 2.1.2).

2.1.1. Audiometría tonal infantil

La audiometría tonal es la más utilizada en el ámbito clínico. Estas audiometrías son fáciles de realizar y de medir, sencillas de contestar y sus resultados son fiables aunque son difíciles de realizar en niños de corta edad, por tanto, se requiere experiencia en valoración audiológica infantil y adaptar tanto la técnica como la metodología en función de la edad del niño y grado de colaboración. En el ámbito clínico y logopédico, las audiometrías tonales permiten valorar las capacidades auditivas de los niños y determinar su grado de pérdida auditiva

En concreto, en una audiometría tonal el audiómetro emite tonos puros de diferentes frecuencias y con varios niveles de intensidad. El paciente indica al profesional cuándo puede oír la señal. Los resultados se representan en un audiograma, que es una ilustración gráfica que muestra los umbrales de audición de cada paciente. La audiometría tonal también se utiliza para determinar el grado de rendimiento protésico cuantitativo.

2.1.2. Audiometría verbal (o logaudiometría) infantil

A diferencia de una audiometría tonal, en una logaudiometría el sujeto no escucha tonos puros, sino palabras o frase producidas por diferentes tipos de voz (masculina, femenina o infantil, que corresponden a franjas frecuenciales diferentes) y con varios niveles de intensidad. A continuación, tiene que repetir las o indicar el objeto correspondiente en una lámina.

Las pruebas se pueden realizar de viva voz y/o con un audiómetro (con altavoces de campo libre o con auriculares para valorar cada oído). Como en la audiometría tonal, en la logaudiometría también se pide la participación del niño, señalando o repitiendo palabras, frases o elementos fónicos y hay que tener en cuenta que la repetición de la palabra no siempre implica su comprensión. Además, hay que tener en cuenta que la familiarización de los niños con las pruebas verbales se realiza tan pronto se observa una ganancia protésico-tonal favorable para la recepción de los estímulos de habla.

Uno de los aspectos cruciales de las logaudiometrías es, justamente, que los estímulos (es decir, las palabras o frases) están formadas por sonidos reales del habla, no por tonos puros. Eso permite evaluar no sólo la capacidad auditiva del sujeto desde el punto de vista meramente fisiológico, sino también su capacidad de percibir y reproducir los contrastes entre sonidos que desempeñan un papel importante en la comunicación oral (es decir, permite evaluar sus habilidades en un nivel más abstracto o psíquico, es decir, fonológico).

Esto, unido a que los niños suelen reaccionar mejor a las logaudiometrías que a las audiometrías tonales, hace de ella una herramienta especialmente interesante en casos de hipoacusia infantil. Por ello, es importante disponer de diferentes pruebas verbales para valorar el rendimiento auditivo del niño, adaptadas a su maduración etaria, así como para valorar la necesidad de modificar la adaptación protésica y / o determinar las áreas en las que puede haber dificultades y así definir los objetivos del programa de rehabilitación auditiva.

En concreto, la logaudiometría infantil contribuye a lo siguiente: 1) la evaluación y el seguimiento de las capacidades perceptivas necesarias para la adquisición de las habilidades auditivas; 2) la indicación de un programa de rehabilitación auditiva; 3) la mejora en la optimización de la tecnología disponible, ya sea prótesis de estimulación por vía ósea, prótesis de estimulación por vía aérea, prótesis implantable de estimulación coclear, prótesis implantables de estimulación neural, equipamiento móvil de FM, entre otros; y 4) la evolución de la adaptación protésica y el control de su eficacia.

En la actualidad existen materiales para la aplicación de logaudiometrías en español (De Cárdenas y Marrero, 1994; Huarte et ál., 1996; Marrero, Maggio y Calvo, 2014; Widex Audífonos SA., 1993) y para el catalán (Tolosa et al., 2000; Calvo y Marro, 2016; Widex, 1993), pero ninguna de ellas contaba con materiales fonéticamente equilibrados para el catalán central.

3. MATERIALES PROPUESTOS

La diversidad en los niños con pérdida auditiva conlleva la necesidad de elaborar diferentes materiales que se ajusten a sus individualidades. Es decir, para la realización de esta prueba es necesario adaptarse a las capacidades y el nivel lingüístico del niño.

Así, a la hora de seleccionar la prueba más adecuada para cada uno hay que tener en cuenta varios factores: 1) su edad; 2) el tiempo de uso de su prótesis; 3) el rendimiento protésico tonal; 4) sus habilidades auditivas; y 5) su nivel lingüístico y de habla.

Los elementos léxicos de los materiales presentan una frecuencia alta en el léxico habitual de los niños y, por tanto, les resultan familiares. El vocabulario elegido ha seleccionado y distribuido en diferentes listas en función de la dificultad y de sus características fónicas y léxico-semánticas. Todas las listas de las pruebas están equilibradas fónicamente tomando como referencia para el equilibrio fónico en catalán central Rafel (1980, 1996).

Además, se ha tenido en cuenta que el habla humana tiene una gran variación frecuencial dependiendo de la voz del locutor. A grandes rasgos podemos establecer tres grandes grupos de voces: voces graves (normalmente de hombre), voces agudas (normalmente de mujer) y voces muy agudas (correspondientes a niños y niñas en edades infantiles). Esta variación se ha tenido en cuenta en los materiales incluyendo estos tres tipos de voces para cada estímulo.

El libro también incluye diferentes tablas de protocolos para la recogida de datos, los plafones visuales y todos los audios en CD grabados en formato digital con los tres tipos de voces mencionadas (masculina, femenina e infantil). Además, las listas de frases se grabaron con tres condiciones de escucha: 1) sin ruido; 2) con ruido blanco; y 3) con ruido *party*. Los materiales de evaluación se encuentran disponibles en formato digital en la web del Laboratori de Fonètica (UB).

El libro que se presenta propone diferentes materiales para que el audiólogo pueda seleccionar, en cada momento del proceso de valoración y/o seguimiento audioprotésico del niño, aquellos que considere más adecuados. Los estímulos verbales que se usan en las diferentes pruebas son: logatomos, palabras y frases. Y están optimizados para las tareas que se detallan en las subsecciones siguientes. Además, se ha comprobado la efectividad de las pruebas con casos y sujetos de control.

3.1. Identificación de palabras bisílabas con soporte visual

El libro incluye listas de palabras bisílabas con apoyo visual (para los niños más pequeños y/o con bajo nivel de habla (Figura 1).

Las listas constan de 20 palabras equilibradas fonéticamente. Sin embargo, estas listas son demasiado largas para los niños más pequeños por lo que se presentan en 3 secciones de 9 estímulos cada una.

Para cada una de las secciones el niño/a cuenta con una lámina dónde debe señalar la palabra que ha oído. La lámina cuenta con 6 palabras objetivo (que el niño oirá) y 3 distractores, que no se oirán en la grabación, pero que se parecen a las palabras objetivo (por ejemplo, *fila/pila* o *cotxe/cotxes*). En la Figura 1, se pueden observar las láminas con el contenido que se oye en la grabación y los distractores escritos en rojo.

Lista 1 - plafó 1 (Fernández Planas et al. 2016)



 pila	 cames	 fila	 cotxes
 cotxe	 gorro	 moto	 núvol
 fada	 tassa	 caca	 pala

Llista	paraula	index de freqüència
1	pila	1373
2	oli	4952
3	cabells	5087
4	cames	5205
5	cebes	1152
6	cinta	596
7	cotxes	1033
8	fada	158
9	gatet	554
10	gorro	15
11	lluna	1456
12	mitjons	74
13	moto	192 (motocicleta)
14	neda	1098
15	nina	300
16	núvol	1183
17	pala	285
18	peres	312
19	pluja	2215
20	roda	1011
21	suro	1246
22	tassa	418
23	teles	1107 (televisió)
24	tigre	146
25	caca	15
26	vermell	1372
27	vestit	3446

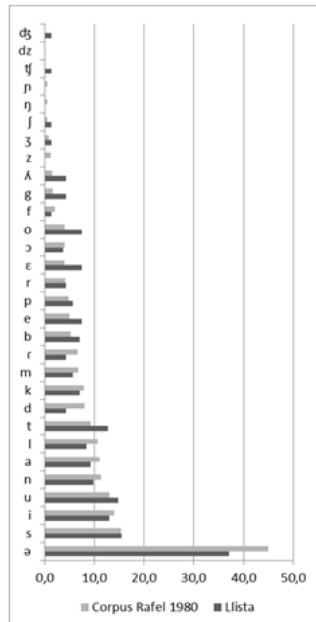


Figura 1: Lista de palabras bisílabas con su equilibrio fónico y el plafón de una de las secciones.

3.2. Repetición de palabras bisílabas sin soporte visual

Los materiales están preparados para valorar el grado de inteligibilidad del habla con 5 listas de palabras bisílabas.

En este caso, la tarea consiste en repetir la palabra que acaban de oír. Las listas tienen una dificultad creciente, así las listas 1 y 2 contienen palabras de vocabulario más básico o cercano para los niños como *fada* 'hada, mientras que palabras abstractas como *pena* o *burla*, se incluyen en las listas 4 o 5.

PARAULES BISÍLLABES
Fernández Planas et al. (2016)



Llista 1		NOM: A. H.							
data:	data:	data:	data:	data:	data:	data:	data:	data:	data:
<input type="checkbox"/> veu home	<input type="checkbox"/> veu home	<input type="checkbox"/> veu home	<input type="checkbox"/> veu home	<input type="checkbox"/> veu home	<input type="checkbox"/> veu home	<input type="checkbox"/> veu home	<input type="checkbox"/> veu home	<input type="checkbox"/> veu home	<input type="checkbox"/> veu home
<input type="checkbox"/> veu dona	<input type="checkbox"/> veu dona	<input type="checkbox"/> veu dona	<input type="checkbox"/> veu dona	<input type="checkbox"/> veu dona	<input type="checkbox"/> veu dona	<input type="checkbox"/> veu dona	<input type="checkbox"/> veu dona	<input type="checkbox"/> veu dona	<input type="checkbox"/> veu dona
<input type="checkbox"/> veu nen	<input type="checkbox"/> veu nen	<input type="checkbox"/> veu nen	<input type="checkbox"/> veu nen	<input type="checkbox"/> veu nen	<input type="checkbox"/> veu nen	<input type="checkbox"/> veu nen	<input type="checkbox"/> veu nen	<input type="checkbox"/> veu nen	<input type="checkbox"/> veu nen
pròtesi: IC	pròtesi: IC	pròtesi: IC	pròtesi: IC	pròtesi: IC	pròtesi: IC	pròtesi: IC	pròtesi: IC	pròtesi: IC	pròtesi: IC
dB: 60	dB: 60	dB: 60	dB: 60	dB: 60	dB: 60	dB: 60	dB: 60	dB: 60	dB: 60
encerts: 20/20	encerts: /20	encerts: /20	encerts: /20	encerts: /20	encerts: /20	encerts: /20	encerts: /20	encerts: /20	encerts: /20
% encerts: 100	% encerts: --	% encerts: --	% encerts: --	% encerts: --	% encerts: --	% encerts: --	% encerts: --	% encerts: --	% encerts: --
moto									
fletxes	+								
caixes	+								
quinze	+								
nina	+								
cebes	+								
nata	+								
calma	+								
tigre	+								
bidell	+								
peres	+								
suro	+								
salut	+								
sera	+								
roda	+								
mida	+								
pomes	+								
pinyol	+								
Nadal	+								
tassa	+								

Figura 2: Ejemplo del protocolo de recogida de datos de una lista de palabras bisílabas.

3.3. Repetición de palabras polisílabas

Los materiales preparados para determinar el umbral de recepción verbal son 25 palabras polisílabas (como *senyora* 'señora', *llimona* 'limón', *germaneta* 'hermanita') de alta frecuencia y equilibradas fónicamente, como es habitual.

3.4. Repetición de frases

El libro también incluye 6 listas de frases para valorar los efectos perceptivos de la coarticulación y de la prosodia habitual.

De ellas las 5 primeras listas están diseñadas con una dificultad lo más baja posible. Las listas de frases se han equilibrado en grupos de 5 frases y tienen diferentes estructuras y longitudes.

La sexta lista contiene cambios de entonación declarativa/interrogativa y palabras repetidas para aumentar su dificultad. En este caso el equilibrio de la lista se ha realizado con 12 frases, pensadas para

ser administradas en dos tareas. Estas frases son del tipo *La Caputxeta Vermella juga amb la iaia*.

3.5. Repetición de logatomos

Por último, los materiales incluyen logatomos (*non-sense words*) para poder rellenar una matriz de confusiones de consonantes, que nos indica las dificultades perceptivo-auditivas que pueden comprometer la representación fonológica de los sonidos del habla a través de las prótesis auditivas. La figura 3 muestra la lista de logatomos para la matriz de confusiones de las consonantes.

Posició intervocàlica (VCV)			Posició inicial (CV)			Posició final (CVC)		
<apa>	/'apa/	[ˈapa]	<pa>	/'pa/	[ˈpa]	<pap>	/'pap/	[ˈpap]
<ata>	/'ata/	[ˈata]	<ta>	/'ta/	[ˈta]	<pat>	/'pat/	[ˈpat]
<aca>	/'aka/	[ˈaka]	<ca>	/'ka/	[ˈka]	<pac>	/'pak/	[ˈpak]
<aba>	/'aba/	[ˈaba]	<ba>	/'ba/	[ˈba]			
<ada>	/'ada/	[ˈada]	<da>	/'da/	[ˈda]			
<aga>	/'aga/	[ˈaga]	<ga>	/'ga/	[ˈga]			
<afa>	/'afa/	[ˈafa]	<fa>	/'fa/	[ˈfa]	<paf>	/'paf/	[ˈpaf]
<assa>	/'asa/	[ˈasa]	<sa>	/'sa/	[ˈsa]	<pas>	/'pas/	[ˈpas]
<aixa>	/'aia/	[ˈaia]	<xa>	/'xa/	[ˈxa]	<paix>	/'paɪ/	[ˈpaɪ]
<aza>	/'aza/	[ˈaza]	<za>	/'za/	[ˈza]			
<aja>	/'aja/	[ˈaja]	<ja>	/'ja/	[ˈja]			
<atsa>	/'atsa/	[ˈatsa]	<tza>	/'tza/	[ˈtza]	<pats>	/'pats/	[ˈpats]
<atxa>	/'atxa/	[ˈatxa]	<txa>	/'txa/	[ˈtxa]	<patx>	/'patx/	[ˈpatx]
<adza>	/'adza/	[ˈadza]	<dza>	/'dza/	[ˈdza]			
<adja>	/'adja/	[ˈadja]	<dja>	/'dja/	[ˈdja]			
<ala>	/'ala/	[ˈala]	<la>	/'la/	[ˈla]	<pal>	/'pal/	[ˈpal]
<alla>	/'alla/	[ˈalla]	<lla>	/'lla/	[ˈlla]	<pall>	/'paɫ/	[ˈpaɫ]
<ara>	/'ara/	[ˈara]						
<arra>	/'arra/	[ˈarra]	<ra>	/'ra/	[ˈra]	<par>	/'par/	[ˈpar]
<ama>	/'ama/	[ˈama]	<ma>	/'ma/	[ˈma]	<pam>	/'pam/	[ˈpam]
<ana>	/'ana/	[ˈana]	<na>	/'na/	[ˈna]	<pan>	/'pan/	[ˈpan]
<anya>	/'anya/	[ˈanya]	<nya>	/'nya/	[ˈnya]	<pany>	/'panj/	[ˈpanj]
						<pang>	/'pang/	[ˈpang]

Figura 3: Logatomos para crear las matrices de confusiones de las consonantes del catalán oriental central.

4. CONCLUSIONES

Los materiales preparados y ofrecidos: 1) facilitan la valoración de la inteligibilidad verbal en niños que presentan diferentes tipos y grados de pérdida auditiva; y 2) presentan para llevarlo a cabo un material verbal fónicamente equilibrado y grabado en formato digital.

5. BIBLIOGRAFÍA

Aguinaga, G., Armentia, M. A., Fraile, A., Olangua, P., y Uriz, N. (2004). *Prueba de Lenguaje Oral de Navarra-Revisada (PLON-R)*. Pamplona: Fondo de Publicaciones del Gobierno de Navarra.

Bosch Galceran, L. (1983). El desarrollo fonológico infantil: una prueba para su evaluación. *Anuario de Psicología*, 28, 85-114.

Calvo, P., y Marro, S. (2016). *Adaptació al català d'EARS, MED-EL*, Innsbruck. [Allum-Mecklenburg, D. (2008): EARS, MED-EL, Innsbruck].

De Cárdenas, M. R., y Marrero, V. (1994). *Cuaderno de logaudiometría*. Madrid: Universidad Nacional de Educación a Distancia.

Departament de Salut (2010). *Document de cribatge auditiu neonatal universal*. Departament de Salut.

Fenson, L. (2002). *MacArthur Communicative Development Inventories: User's guide and technical manual*. Paul H. Brookes.

Fernández Planas, A. M.; Marro, S.; Roseano, P.; Calvo, Pilar; Elvira-García, W.; Claveria, M. A. (2017). *Logaudiometries infantils en català oriental central*. Barcelona: Horsori.

Huarte, A., Molina, M., Manrique, M., y Olleta, I. (1996). Protocolo para la valoración de la audición y el lenguaje en lengua española en un programa de Implantes Cocleares. *Acta Otorrinolaringologica Española*, 47(1), S1-S14.

Kirk, S. A., McCarthy, J. J., & Kirk, W. D. (1984). Test Illinois de aptitudes psicolingüísticas (ITPA). TEA.

Marrero, V., Maggio, M., y Calvo, C. (2014). *Prueba de audiometria verbal en ruido PAVER, PIP*. Madrid: Programa Infantil Phonak / UNED.

Rafel i Fontanals, J. (1980). Dades sobre la freqüència de les unitats fonològiques en català. *Estudis Universitaris catalans*, 25, 473-406.

Rafel i Fontanals, J. (1996). *Diccionari de freqüències: Llengua no literària*. Barcelona: Institut d'Estudis Catalans.

Tolosa, F., Montoya, B., Dols, N., y Ramis, M. M. (2000). *Material lingüístic en català per a exploracions logaudiomètriques (central, valencià i balear)*. Palma: Edicions de la Universitat de les Illes Balears / Conselleria d'Educació i Cultura, Govern de les Illes Balears.

Salguero Santana, M. S., Álvarez Arrieta, Y., Verane Dubalón, D., y Santelices Jiménez, B. Y. (2015). El desarrollo del lenguaje. Detección precoz de los retrasos/trastornos en la adquisición del lenguaje. *Revista Cubana de Tecnología de la Salud*, 6(3), 43-57.

Widex Audífonos SA. (1993). *Audiometría vocal. Prueba de inteligibilidad. Discriminación acústica. Confusión consonántica. (Listas elaboradas por el Departamento de Filología Española de la UAB – J. M. Garrido y J. Llisterri. Medida del poder separador temporal Test Leman-Renard. Elaborado por el Dr. Leman y X. Renard. Locutor: Jordi Royo) [CD]*. Barcelona: Widex Audífonos SA.

Wiig, E. H., Secord, W. A., & Semel, E. (2013). *Clinical evaluation of language fundamentals: CELF-5*. Pearson.

Agradecimientos. Este material se ha podido llevar a cabo gracias a la ayuda económica de la UB y del Servicio ORL del Hospital Materno-Infantil Sant Joan de Déu. Asimismo, agradecemos la ayuda de la escuela *Samuntada* de Sabadell, de los locutores que desinteresadamente ofrecieron sus voces y su tiempo para las grabaciones y de diferentes páginas web para la parte gráfica de los plafones y para obtener el ruido *party*.

EVIDENCIA ACÚSTICA Y ESTADÍSTICA DE UNA CLASE NATURAL DE EYECTIVAS EN KAWÉSQR

Mauricio A. Figueroa Candia, Daniela A. Mena Sanhueza

Universidad de Concepción
maufigueroa@udec.cl, danielamena@udec.cl

ABSTRACT

El presente estudio presenta primera evidencia acústica y estadística para evaluar la hipótesis de una serie natural de consonantes eyectivas en la lengua kawésqar. Para lograr este objetivo, se trabajó con señales provenientes de 4 hablantes, de las que se extrajeron promedios de intensidad de consonantes no eyectivas (oclusivas sordas y africadas) y de las eyectivas de los mismos puntos articulatorios, así como datos del f_0 promedio de la vocal que sigue a cada consonante, dado que ambos correlatos acústicos han demostrado ser eficientes en codificar el contraste de estos modos de articulación. Los resultados de análisis de Modelos Lineales Generalizados Mixtos sugieren que los promedios de intensidad son un buen correlato del contraste, y favorecen la hipótesis de una serie natural de eyectivas para la lengua kawésqar.

Palabras clave: kawésqar, consonantes eyectivas, tipología, fonética acústica

This study presents first acoustic and statistical evidence to evaluate the hypothesis of a natural series of ejective consonants in Kawesqar. In order to do so, signals from 4 subjects were analysed, from which mean intensity values of non-ejective consonants (voiceless stops and affricates) and from ejectives of the same place of articulation were extracted, as well as mean f_0 values of the vowels following each consonant, given that both acoustic correlates have been shown as effective in encoding the contrast between these manners of articulation. The results of Generalized Linear Mixed Models analyses suggest that mean intensity is a good acoustic correlate for the contrast, favouring the hypothesis of a natural series of ejectives in Kawesqar.

Keywords: Kawesqar, ejective consonants, typology, acoustic phonetics

1. INTRODUCCIÓN

1.1. Sobre la etnia kawésqar

Los kawésqar (o “alacalufes”) solían habitar el territorio comprendido entre la costa del Golfo de Penas (latitud 47° 30' S) hasta las islas del oeste de Tierra del Fuego (Gusinde, 1924, 1926; Emperaire, 1963). El clima de estos territorios se caracteriza por tener temperaturas bajas todo el año y por ser una de las zonas más lluviosas del mundo (Gusinde, 1924). Además, se encuentra conformado por montañas, campos de hielo, y por pasajes de agua, lo que, sumado a una densa vegetación, explica la preferencia nómada y marítima por parte de los kawésqar de generaciones anteriores (Bird, 1946). En la actualidad, prácticamente la totalidad de los kawésqar se encuentra radicada en Puerto Edén (49° 80' de latitud sur y 74° 27' 10" de longitud oeste), un poblado situado en la costa occidental de la isla Wellington (Aguilera, 1978), y al que solo se puede

acceder por vía marítima (Clairis, 1978). Aunque en el pasado la etnia era nómada, y subsistía de la caza y recolección (Aguilera, 1978), en la actualidad, los kawésqar de Puerto Edén subsisten a través de la venta de artesanías y de otras actividades similares.

Con respecto al estado de conservación de la etnia y de su lengua, desde una perspectiva cultural, puede afirmarse que los kawésqar han sufrido un proceso de transculturación (Lipschutz, 1962), el que ha tenido como consecuencias la pérdida del nomadismo, el abandono de la mayoría de las prácticas de caza y el aparente cese de ceremonias tradicionales (Clairis, 1972, 1978). En cuanto a la lengua, hacia fines de la década de los 70 solo existían alrededor de 45 hablantes de la lengua (Aguilera, 1978). En la actualidad los hablantes no superan la media decena, y la mayoría de edad muy avanzada (Crevels, 2012). En consecuencia, el kawésqar es una lengua que se encuentra en un claro proceso de extinción (Viegas Barros, 1990, 1991; Sánchez, 1994; Salas, 1998), lo

que releva la urgencia del trabajo con la lengua y sus hablantes.

1.2. La lengua kawésqar

El kawésqar es una lengua aglutinante y polisintética, de acento libre, que carece de género, artículos y diferenciación de número (Aguilera, 1978, 1982b). Algunas propiedades inusuales de la lengua son la ausencia de una serie de consonantes oclusivas sonoras, la presencia de variantes fónicas glotalizadas y uvulares, y un orden de palabra SVO (Campbell, 2012). Con respecto a su filiación genética, se han defendido propuestas que la clasifican como una lengua aislada (e.g., Viegas Barros, 1991; Aguilera, 1982b; Campbell, 2012), y propuestas que la emparentan con otras lenguas fueguinas (e.g., Lehmann-Nitsche, 1919; Greenberg, 1960; Key, 1979).

Con respecto al nivel fonético y fonológico, nivel en el que se concentra la presente propuesta, el estado del arte se caracteriza por la existencia de considerables diferencias entre las descripciones de la lengua. Como ejemplos del grado de desacuerdo, en lo que respecta a las consonantes fricativas y africadas, existen autores que le dan estatus fonémico a /tʃ/, /tʃʰ/, /f/, /x/ y /s/, pero con [tʃʰ] como variante de /tʃ/, [ɸ] como variante de /f/, [x] como variante de [h] (a la que no se le da estatus fonológico), [χ] como variante de /x/, [ʃ] como variante de /s/, y [h] como variante de /x/, /s/ y /qʰ/ (Clairis, 1977a, 1977b, 1978, 1980, 1985). Otros autores, en cambio, le dan estatus fonológico a /tʃ/, /f/, /x/, /s/ y /h/, y consignan a [χ] como variante de /q/, [ʃ] y [z] como variantes de /s/, y [ʒ] como variante de /j/ (Aguilera, 1978, 1982a, 1982b; Aguilera y Tonko, 2006).

1.3. Serie de oclusivas, eyectivas y aspiradas

El estatus fonológico de las series naturales de las consonantes oclusivas, aspiradas y eyectivas es uno de los debates más importantes dentro de los estudios de la lengua kawésqar. Resolver el estatus de estas series de consonantes es tipológicamente interesante, pues existe un escaso número de lenguas con contrastes de este tipo y que al mismo tiempo carezcan de una serie de consonantes oclusivas sonoras (Maddieson, 1984; Burquest y Payne, 2001).

En el caso de Clairis, el autor identifica como unidades fonológicas contrastivas a las oclusivas sordas /p/, /t/ y /q/, a las eyectivas /pʰ/, /tʰ/ y /qʰ/, y a las oclusivas sordas con aspiración /pʰ/, /tʰ/ y /qʰ/ (Clairis, 1977a, 1977b, 1978, 1985, 1997). Es importante consignar también que indica que [k] es

en ocasiones una variante alofónica de /q/ (Clairis, 1980), que [pʰ] y [pʰ] pueden ser alófonos de /p/, que [tʰ] es en ocasiones variante de /t/, que [q] puede ser variante de /qʰ/, y que [k] puede ser variante alofónica de /qʰ/ (Clairis, 1985). Como puede apreciarse, no se proponen unidades fonológicas para el punto articulatorio velar, y existe una superposición entre las variantes combinatorias de distintas unidades subyacentes.

Aguilera, en cambio, da estatus fonológico a la serie de unidades oclusivas sordas /p/, /t/, /k/ y /q/, y a una serie de eyectivas, /pʰ/, /tʰ/, /tʰ/ (alveopalatal eyectiva sorda) y /kʰ/, no incluyendo a /qʰ/ (Aguilera, 1982a, 1982b; Aguilera y Tonko, 2006). También propone que existen variantes alofónicas aspiradas e implosivas sordas para ambas series de consonantes, dentro del correspondiente punto articulatorio, pero no para las consonantes eyectivas uvulares (Aguilera, 1982a).

1.4. La presente propuesta

El presente estudio se propone proveer una primera evidencia acústica y estadística tendiente a contribuir a la discusión sobre el estatus fonológico de la serie de eyectivas en kawésqar. Para lograr este cometido, se obtendrán datos de correlatos acústicos que han demostrado ser capaces de codificar las diferencias relevantes entre las categorías, y se los evaluará estadísticamente para determinar si existe evidencia en favor del contraste entre los modos de articulación no eyectivos (oclusivo sordo y africado) y eyectivo.

2. MÉTODOS

2.1. Participantes, tareas y señales

Los datos provienen de señales grabadas a partir del habla de cuatro participantes (2 de sexo femenino y 2 de sexo masculino), de la localidad de Puerto Edén. La grabación se realizó en una fecha indeterminada durante los últimos años de la década de 1970. En ese entonces, las edades aproximadas de los sujetos rondaba los 40 años. Las grabaciones corresponden a la elicitación de una lista de palabras, con el estímulo presentado en castellano, y seguido de tres repeticiones de la palabra en kawésqar. A pesar de que las señales se encontraban en cintas magnéticas ya deterioradas, y a que las condiciones de grabación no fueron las ideales, estas presentan una relación señal-ruido aceptable para estudios del tipo de variables acústicas que se discutirán más abajo ($\bar{x} = 21.9$ dB, $\sigma = 1.02$ dB).

2.2. Segmentado y etiquetado de señales

Las señales fueron inspeccionadas y todas las instancias de consonantes oclusivas sordas, africadas y eyectivas fueron segmentadas y etiquetadas en TextGrids del programa Praat (Boersma y Weenink, 2019), con ayuda de las formas de onda, espectrogramas y mediante análisis auditivos. Además de cada instancia consonántica, se identificaron los límites temporales de las vocales que seguían a las consonantes de interés.

2.3. Mediciones acústicas

El promedio de intensidad de cada consonante y los valores de f_0 de las vocales que las siguen fueron medidos para todos los fonos de interés mediante scripts para el programa Praat, sobre objetos *Intensity* y *Pitch* creados con valores por defecto. Estas dos medidas acústicas han demostrado ser buenos correlatos para las diferencias entre el modo de articulación eyectivo y los modos oclusivo sordo y africado. En particular, las consonantes eyectivas suelen presentar valores de intensidad promedio más altos que los de las oclusivas sordas y africadas, y las vocales que siguen a las consonantes eyectivas suelen mostrar mayores valores de f_0 que sus contrapartes de otros modos (Bennet, 2010).

3. RESULTADOS Y ANÁLISIS

3.1. El corpus

El corpus recogido contiene 1073 consonantes, de las cuales 968 corresponden a oclusivas sordas o africadas y 105 a eyectivas (para detalles, véase la Tabla 1). Es importante consignar que solamente se identificaron dos casos de la eyectiva [p'] y ningún caso de [q'], por lo que se excluyó el punto de articulación uvular de los subsecuentes análisis.

Tabla 1: Instancias de consonantes en el corpus separadas por punto y modo de articulación.

Punto	Modo	Unidad	n
Bilabial	Oclusivo	[p]	98
	Eyectivo	[p']	2
Dental	Oclusivo	[t]	286
	Eyectivo	[t']	29
Alveopalatal	Africado	[tʃ]	100
	Eyectivo	[tʃ']	32
Velar	Oclusivo	[k]	428
	Eyectivo	[k']	42
Uvular	Oclusivo	[q]	56

3.2. Diferencias en la intensidad promedio

Los datos provenientes del corpus fueron importados a R (R Core Team, 2018), donde se construyó un Modelo Generalizado Lineal Mixto (MGLM) para evaluar los efectos de las variables *modo articulatorio*, *punto articulatorio* y *segmento* –y sus interacciones– en la variable dependiente *intensidad promedio*. El modelo fue creado usando la función *lmer* del paquete *lmerTest* (Kuznetsova, Brockhoff y Christensen, 2017). La variable *sujeto* fue incluida como un efecto al azar. Para la construcción del modelo, se utilizó un procedimiento aditivo. Las tablas con los efectos principales y las interacciones de cada modelo, con análisis de varianza de Tipo II, fueron producidas mediante la función *Anova* del paquete *car* (Fox y Weisberg, 2011), y usando la función *ranova* de *lmerTest*.

El MGLM con mejor ajuste presentó un efecto principal de *modo articulatorio* ($\chi^2(1) = 62.95, p < 0.001$), *punto articulatorio* ($\chi^2(3) = 29.4, p < 0.001$) y *segmento* ($\chi^2(3) = 8.29, p < 0.05$), pero no presentó interacciones estadísticamente significativas (también se detectó un efecto significativo de la variable al azar *sujeto*). Los datos que explican estos efectos pueden ser observados en los dos paneles de la Figura 1. Como se aprecia allí, se observan diferencias tanto entre segmentos pertenecientes a un mismo punto de articulación (pero diferenciados por modo), como entre los puntos de articulación, y, si se observa el panel derecho de la misma figura, existe una mayor intensidad promedio en las consonantes eyectivas.

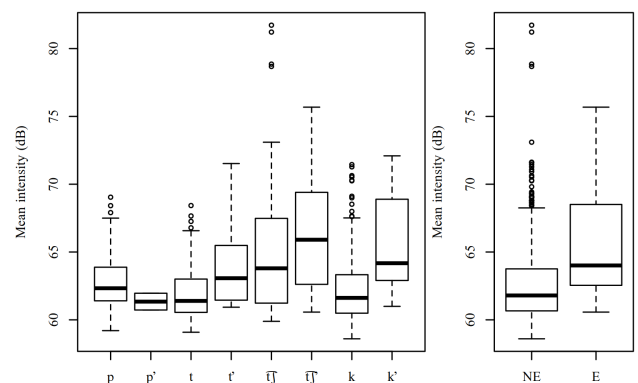


Figura 1: Panel izquierdo: valores de intensidad promedio de cada segmento. Panel derecho: valores de intensidad de los modos articulación no eyectivos (“NE”) y para el modo eyectivo (“E”).

3.3. Diferencias en el f_0 de la vocal siguiente

Al igual que en el análisis anterior, se construyó un MGLM para evaluar el efecto de las variables independientes ya mencionadas en los valores de la

variable dependiente *promedio de f_0 de la vocal siguiente*. Para este análisis, se excluyeron los casos en los que la vocal siguiente se encontraba ensordecida (en total, 323 casos).

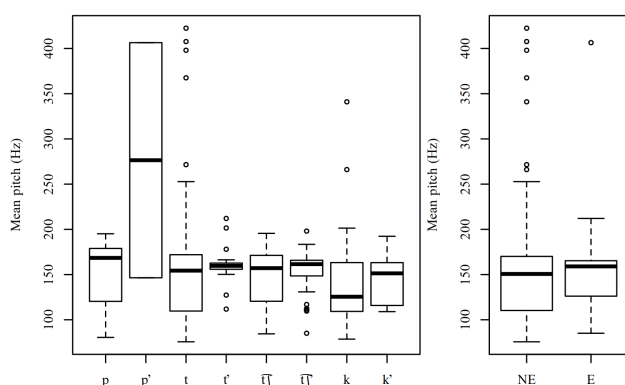


Figura 2: Panel izquierdo: valores de f_0 promedio de cada segmento. Panel derecho: valores de f_0 para los modos articulación no eyectivo (“NE”) y para el eyectivo (“E”).

El modelo con mejor ajuste presentó un efecto principal de *segmento* ($\chi^2(3) = 31.18, p < 0.001$), pero no de *modo articulatorio* ($\chi^2(1) = 2.58, p = 0.1081$) ni de *punto articulatorio* ($\chi^2(3) = 1.77, p = 0.621$), ni interacciones (sí se detectó un efecto significativo de la variable al azar *sujeto*). Los datos que explican estos resultados pueden ser inspeccionados en la Figura 2. Como puede observarse en el panel izquierdo, existe una considerable superposición entre los valores promedio de f_0 de cada segmento, con la excepción de /p'/, que solo está representado por 2 casos en el corpus. El panel derecho de la Figura 2 también muestra un grado considerable de superposición, lo que explica la ausencia de un efecto principal de la variable *modo de articulación*. En el caso del efecto principal de la variable *segmento*, este puede ser, quizás, explicado por los rangos de la distribución de /p'/, y por los datos atípicos de /t/.

4. DISCUSIÓN Y CONCLUSIONES

4.1. Discusión

Los procedimientos de extracción y análisis de datos permitieron recoger evidencia estadística que revela que, en primer lugar, existen diferencias con significatividad estadística entre los valores de intensidad promedio entre las consonantes no eyectivas (oclusivas sordas y africada sorda) y las eyectivas, además de efectos principales de las variables *punto de articulación* y *segmento*. Estos resultados sugieren que existe evidencia acústica para defender la hipótesis de una clase natural de eyectivas en la lengua. Dado lo observado en la Figura 1, sin

embargo, sorprende la ausencia de interacciones entre las variables *modo de articulación* y *punto de articulación*, y entre *modo de articulación* y *segmento*. Estas interacciones eran esperables dada, por ejemplo, la magnitud de las diferencias entre las consonantes no eyectivas y las eyectivas en los puntos de articulación bilabial y dental (en las que hay considerable superposición de los valores), versus las diferencias observadas en el modo africado o en las consonantes velares, en las que las diferencias de intensidad promedio son mayores. Una explicación posible para la ausencia de interacciones como las mencionadas es la baja cantidad de datos para algunos puntos articulatorios.

En segundo lugar, los resultados de los promedios de f_0 de las vocales que siguen a las consonantes sugieren que no existe evidencia suficiente de esta variable acústica para defender la hipótesis de una serie natural de eyectivas. Esto se justifica tanto por la superposición de los valores acústicos dentro de cada modo articulatorio (con la excepción del punto bilabial, ya discutido más arriba), y por la ausencia de efectos principales de las variables *modo de articulación* y *punto de articulación*, además de ausencia de las interacciones. Una vez más, es posible que una mayor cantidad de datos contribuya a precisar estos resultados estadísticos, o que señales de mayor calidad contribuyan a esclarecer este asunto. También es posible que algún tipo de normalización de los valores de f_0 para controlar el efecto de la variable *sexo* tenga un impacto en los resultados que puedan obtenerse para esta medición acústica.

4.2. Conclusiones

En su conjunto, proponemos que existe evidencia acústica en favor de la hipótesis de una serie natural de eyectivas, y que el correlato acústico de intensidad promedio es capaz de capturar este contraste, confirmando en este plano estudios anteriores (Bennet, 2010). Sin embargo, los promedios de f_0 en las vocales que siguen a las consonantes no confirman esta tendencia, lo que: (a) abre cuestionamientos respecto de la efectividad de este correlato para la detección del contraste, y (b) invita a replicar estudios como el presente con corpus más amplios, y con señales de mayor calidad. Como queda en evidencia, estudios de este tipo –los que deben realizarse de manera urgente– pueden contribuir a proveer información tipológicamente interesante y relevante, que complemente nuestro conocimiento de lo que es posible en las lenguas del mundo.

5. BIBLIOGRAFÍA

- Aguilera F., Ó. E., y Tonko P., J. S. (2006). *Manual para la enseñanza de la lengua kawésqar. Nivel básico Ib*. Santiago/Puerto Edén/Punta Arenas.
- Aguilera F., O. E. (1978). Léxico Kawéskar-Español, Español-Kawésqar (Alacalufe Septentrional). *Boletín de Filología*, 29, 7-149.
- Aguilera F., O. E. (1982a). Fonemas kawesqar. Esbozo de la fonología de la palabra. I Parte. *Boletín Indigenista de Chile*, 3, 16-30.
- Aguilera F., O. E. (1982b). Notas sobre la lengua kawésqar. *Boletín Indigenista de Chile*, 2, 21-33.
- Bennett, R. (2010). Contrast and laryngeal states in Tz'utujil. *The UCSC Linguistics Research Center*, 93-120.
- Bird, J. (1946). The Alacaluf, en Julian H. Stewart (ed.), *Handbook of South American Indians. Vol I. The marginal tribes*. New York: Cooper Square Pub.
- Boersma, P, y Weenink, D. (2019). Praat: doing phonetics by computer [Programa computacional]. Versión 6.0.40, obtenida de <http://www.praat.org/>
- Burquest, D. A., y Payne, D. L. (2001). *Phonological analysis: A functional approach*. Dallas, TX: Summer Institute of Linguistics.
- Campbell, L. (2012). Typological characteristics of South American indigenous languages, en L. Campbell y V. Grondona (eds.), *The indigenous languages of South America: a comprehensive guide. Volume 2*, 259-330, Germany: Walter de Gruyter.
- Clairis, C. (1972). Les Alacalufs de Puerto Edén (1971). *Objets et Mondes*, 12(2), 197-200.
- Clairis, C. (1977a). Lingüística fueguina. *Boletín de Filología de la Universidad de Chile*, 28, 29-47.
- Clairis, C. (1977b). Première approche du qawasqar: identification et phonologie. *La Linguistique*, 13(1), 145-152.
- Clairis, C. (1978). La lengua qawasqar (alacaluf). VICUS. *Cuadernos de Lingüística*, 2, 29-44.
- Clairis, C. (1980). José Emperaire et les Qawasqar. Notes linguistiques inédites sur "Les nomades de la mer". *Journal de la Société des Américanistes*, 67, 359-380.
- Clairis, C. (1985). El qawasqar. Lingüística fueguina. Teoría y descripción. *Estudios Filológicos*, Anejo 12. Valdivia: Universidad Austral de Chile.
- Clairis, C. (1997). Lingüística fueguina 1997. *Onomazein*, 2, 421-450.
- Crevels, M. (2012). Language endangerment in South America: The clock is ticking, en L. Campbell y V. Grondona (eds.), *The indigenous languages of South America: a comprehensive guide. Volume 2*, 167-233, Germany: Walter de Gruyter.
- Emperaire, J. (1963). *Los nómades del mar*. Santiago de Chile: Ediciones de la Universidad de Chile.
- Fox, J., y Weisberg, S. (2011). *Multivariate linear models in R. An R Companion to Applied Regression*. Los Angeles: Thousand Oaks.
- Greenberg, J. H. (1960). The general classification of Central and South American languages. *Selected papers of the 5th International Congress of Anthropological and Ethnological Sciences. Philadelphia*, 791-794.
- Gusinde, M. (1924). Cuarta expedición a la Tierra del Fuego. *Publicaciones del Museo de Etnología y Antropología de Chile*, 4(1-2), 7-67.
- Gusinde, M. (1926). Das Lautsystem der feuerländischen Sprachen. *Anthropos*, 21(5/6), 1000-1024.
- Key, M. R. (1979). *The grouping of South American Indian Languages*. Tübingen: Gunter Narr.
- Kuznetsova, A., Brockhoff, P. B., y Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, 82(13).
- Lehmann-Nitsche, R. (1919). El grupo lingüístico alacaluf de los canales magallánicos. *Revista del Museo de La Plata*, 25, 15-69.
- Lipschutz, A. (1962). Los últimos fueguinos: transculturación y desculturación, extinción y exterminación. *Genus*, 18(1-4), 3-29.
- Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.
- R Core Team (2018). R: A language and environment for statistical computing [Programa computacional]. Versión 3.4.4. Obtenida de <https://www.R-project.org/>
- Salas, A. (1988). Reseña a Clairis 1985. *Revista de Lingüística Teórica y Aplicada*, 26, 143-153.
- Sánchez, G. (1994). Estado actual de las lenguas aborígenes de Chile. *Boletín de la Academia Chilena de la Lengua*.
- Viegas Barros, J. P. (1990). Dialectología qawasqar. *Amerindia*, 15, 43-73.
- Viegas Barros, J. P. (1991). Fonología histórica qawasqar, en A. Gerzenstein (coord.), *Temas de lingüística aborígen*, 71-99, Universidad de Buenos Aires: Facultad de Filosofía y Letras.

Estudio financiado por el Proyecto FONDECYT Iniciación 11180356, titulado "Reanálisis de aspectos controversiales del kawésqar y nuevos aportes para su conocimiento: descripción acústica, exploración del dominio perceptivo y procesamiento estadístico multivariable para la actualización de su inventario fonético y fonológico", dirigido por el Dr. Mauricio A. Figueroa Candia.

¿DÓNDE ESTÁ LA ENTONACIÓN? PROPUESTA DIDÁCTICA INTEGRADORA DE LA ENTONACIÓN EN EL NIVEL A1

Andrea García Coronado¹, Beatriz Blecua Falgueras²

¹Universität Konstanz, ²Universitat de Girona
andregarciacoronado14@gmail.com, beatriz.blecua@udg.edu

ABSTRACT

In recent years, the research in intonation and prosody has taken on a prominent importance in the field of study of foreign language teaching. These many studies highlight the fact that the practice of intonation is essential and that it should be integrated in the foreign language curriculum, but how to do it? This paper presents a didactic proposal that includes some practical suggestions and activities to teach intonation within the curriculum of level A1. It shows that the integration of the practice of intonation is a possible challenge to achieve as well as how to do it. This proposal was applied to a real A1 intensive course.

Keywords: *intonation, prosody, pronunciation, oral communication, communicative competence, teaching materials, pronunciation activities, Spanish teaching, Spanish foreign language*

1. INTRODUCCIÓN

La ausencia de contenidos relacionados con la entonación y los demás elementos suprasegmentales en los manuales didácticos de español como lengua extranjera (ELE) es evidente. Este hecho pone de relieve la necesidad de crear materiales que integren la prosodia dentro de las demás destrezas lingüísticas (Villaescusa 2009). Esto, sumado a la importancia y auge que han adoptado el estudio de la entonación y la prosodia en los últimos años, ha motivado esta investigación.

Partimos de la hipótesis de que es posible integrar la entonación al plan curricular de un nivel A1 sin dejar de lado el contenido teórico y las destrezas a las que está asociado dicho nivel. Así, el objetivo de la investigación es diseñar una propuesta didáctica que integre la práctica de la entonación en un curso de nivel A1 y trasladarla al aula.

¿Por qué un nivel A1? La bibliografía pone de relieve la necesidad de empezar a trabajar los elementos suprasegmentales desde el inicio del aprendizaje de una segunda lengua¹ siguiendo el modelo de adquisición de la lengua materna. En este campo destaca Lahoz (2007), que parte de algunas consideraciones sobre la entonación en el ámbito de la adquisición de la lengua materna para ver su

aplicación en el aula de ELE y, de este modo, intentar reproducir el mismo orden de adquisición en el aula. En el diseño de la propuesta, además, se incluyen objetivos concretos centrados en los estudiantes. A partir de ella, se espera que adquieran nociones básicas sobre la entonación que les permitan desenvolverse exitosamente en situaciones comunicativas. Dentro de la propuesta, se abordan temas como la melodía de las lenguas, los patrones melódicos del español y el significado afectivo de la entonación. El objetivo último es que estos conocimientos les sirvan como base para comprender el funcionamiento de la entonación del español.

2. METODOLOGÍA

2.1 Diseño

En el diseño de la propuesta didáctica se busca integrar la práctica de la entonación a los demás contenidos propios de un nivel A1 según el MCER. Para ello, en primer lugar, se han establecido unos objetivos siguiendo las indicaciones que se ofrecen en el PCIC en relación con la entonación en la fase de aproximación (A1-A2).

Para el diseño de los ejercicios se han puesto en práctica aportaciones teóricas y propuestas metodológicas de distintos lingüistas, entre ellos

¹ Segundas lenguas: a partir de ahora y en adelante, L2.

Cortés (2002), Gil Fernández (2007), Santamaría Busto (2007) y Lahoz (2012). En cuanto a la estructuración de la progresión de los ejercicios diseñados, se han seguido las indicaciones que ofrece Lahoz (2012).

2.2. Contexto de implementación

Una vez diseñada la propuesta, se trasladó al contexto adecuado, esto es, se puso en práctica en un curso de español A1.

2.2.1. El curso

La propuesta didáctica se integró en el curso intensivo de *Spanisch II*, un curso de español que se ofrece en el centro de lenguas de la Universität Trier, Alemania. Su duración fue de 48 horas presenciales durante el semestre de verano (del 19 de abril al 19 de julio de 2017), distribuidas en un total de 24 clases.

2.2.2 Los estudiantes

El curso estaba destinado a estudiantes de todas las facultades de la Universität Trier. Participaron un total de 12 personas de entre 19 a 26 años que ya tenían conocimiento de, por lo menos, dos lenguas

extranjerías. La mayoría habían estudiado inglés como segunda lengua y francés o italiano como tercera. Algunos, asimismo, ya tenían nociones básicas de español, pero para la mayoría el curso fue su primera toma de contacto con esta lengua. Sus objetivos e intereses respecto al aprendizaje del español eran muy concretos: viajar, realizar un Erasmus y el prestigio laboral.

3. PROPUESTA DIDÁCTICA

La propuesta didáctica incluye un total de once ejercicios divididos en tres fases o bloques temáticos. Todas las actividades están incluidas en la planificación de 9 clases distribuidas desde el principio y hasta el final del curso intensivo. La entonación se trabajó, por lo tanto, el 37,5% de las clases del curso.

La tabla 1 muestra un ejemplo de la planificación didáctica de una de las clases en la que se destacan en sombreado las actividades dedicadas a la práctica de la entonación.

Se ha intentado, en la medida de lo posible, que estos ejercicios estén directamente relacionados con los conceptos que se tratan en la unidad.

Tabla 1: Ejemplo del desarrollo y la planificación de una clase

Desarrollo de la clase (planificación didáctica y metodológica)					
Profesora/Practicante: Andrea García Coronado	Fecha: 31.05.2017	Asignatura: Curso intensivo de español	Centro de lenguas: Universidad de Tréveris	Condiciones de aprendizaje: (Nivel A1)	Tiempo: 90 minutos
Unidad pedagógica: 2. <i>Mi ciudad ideal</i>			Sesión: 12		
Condiciones de aprendizaje:					
<ul style="list-style-type: none"> ❖ Los alumnos quieren obtener un nivel A1 ❖ Los alumnos comprenden frases simples, dichas despacio y con claridad, que se refieren a su persona sus intereses o sus actividades de clase. ★ Los alumnos entienden una conversación, si se trata de un tema conocido y se habla despacio. 					
Metas de la clase: (Meta a), b), c), d), e), f), g), h), i)					
Tiempo	Fase/Método	Actividades	Dinámica	Medio	Objetivo
5'	Corrección deberes	- Se corrigen los deberes de la sesión anterior y se resuelven dudas.	Global	- VR/LE ej. 5 A pág. 19 y ej. 12 pág. 22	a) Los alumnos afianzan su conocimiento sobre el uso de los posesivos y sobre la descripción de ciudades.
15'	Elaboración II	- Los alumnos realizan ejercicios relacionados con el vocabulario de la familia.	Individual	- Ejercicio extraído de: https://www.profdesle.es/vocabulario/la-familia/	b) Los alumnos afianzan su conocimiento sobre el vocabulario de la familia y los posesivos.
15'	Elaboración III	- Los alumnos realizan una expresión oral y escrita que consiste en presentar a un miembro de su familia.	Individual	- Ninguno	c) Los alumnos son capaces de realizar textos sencillos describiendo a su familia.
5'	Introducción	- Entonación, los estados de ánimo.	Global	- Ninguno	d) Los alumnos entienden el hecho de que el estado de ánimo está ligado a la entonación del habla.
10'	Percepción	- Los alumnos realizan un ejercicio de percepción de los estados de ánimo mediante una actividad.	Individual	- Ficha de ejercicios - Video: https://www.youtube.com/watch?v=KfnP201W1Iq	e) Los alumnos son capaces de identificar el estado de ánimo del hablante según su forma de entonar.
15'	Producción controlada	- Los alumnos realizan un ejercicio sobre la producción de enunciados según el estado de ánimo del hablante.	Individual/ Parejas	- Ficha de ejercicios 2 y 3.	f) Los alumnos son capaces de expresar sus estados de ánimo mediante la entonación.
10'	Expresión oral	- Lectura de los diálogos que los alumnos prepararon.	Grupos	- Diálogos de los alumnos.	g) Los alumnos leen sus diálogos en clase procurando entonar correctamente.

Todas las actividades divididas en fases que se realizaron durante el curso fueron ensayo y preparación para la tarea final que engloba todos los aspectos sobre la entonación trabajados durante el curso y que culmina con la realización de unas grabaciones. A continuación, se resume en qué consiste cada una de las fases.

3.1. Fase 1: ¿Cómo suena el español? La melodía de las lenguas

El objetivo marcado en esta fase es concienciar al estudiante sobre la importancia y significado de la entonación y la variación lingüística del español. En ella, se diseñaron ejercicios de sensibilización y reflexión hacia la melodía de las lenguas y hacia la importancia de la entonación en la comunicación.

3.2. Fase 2: Las estructuras entonativas del español

En esta fase, se pretende que los estudiantes perciban los patrones entonativos del español y entiendan su funcionamiento. Así pues, este bloque está compuesto por ejercicios de sensibilización y práctica de diferentes estructuras entonativas y de las características de los patrones melódicos correspondientes a distintos tipos de entonación.

3.3. Fase 3: El significado afectivo del español

El objetivo de esta fase de práctica es que los estudiantes sean capaces, por un lado, de entender el papel que juega la entonación en la expresión de sentimientos y, por el otro, de reconocer el estado de ánimo de un hablante a través de la entonación que utilice para expresarse. Este objetivo se consiguió mediante ejercicios de percepción de estados de ánimo y emociones según la entonación.

3.4. Tarea final

Mediante la tarea final se pretende comprobar que los objetivos establecidos en el planteamiento inicial de la propuesta didáctica se han cumplido al terminar el curso. La tarea incluye ejercicios de todas las fases de la práctica de la entonación llevadas a cabo durante el curso, la realización de unas grabaciones y la inclusión de *Feedback* visual y audiovisual para la autocorrección de la entonación. La tarea final ocupó una sesión de 90 minutos. Esta es su cronología:



Figura 2: Cronología de la sesión dedicada a la tarea final

5. CONCLUSIONES Y REFLEXIONES

El objetivo de este estudio era diseñar una propuesta didáctica que integrara la entonación al plan curricular de un nivel A1 sin dejar de lado el contenido teórico y las destrezas a las que está asociado dicho nivel y, además, ponerla en práctica.

La programación presentada demuestra que se ha conseguido introducir este elemento suprasegmental sin distorsionar el resto del contenido del curso, por lo tanto, la hipótesis planteada al inicio de la investigación se confirma.

A pesar de haber cumplido el objetivo, surgieron algunas dificultades. Por un lado, posiblemente el tiempo y el tipo de curso en el que se inscribía la propuesta no eran los adecuados. El curso intensivo se realizó en 48 horas y en él se debía dar cabida a todos los aspectos gramaticales y léxicos que le corresponden a un nivel A1 para que los alumnos aprobaran el examen final. Por otro lado, se tenía que intentar introducir un elemento suprasegmental que ni siquiera se contemplaba entre los objetivos y la descripción del curso del centro de lenguas de la universidad ni en el manual del curso.

Tras haber puesto en práctica la propuesta didáctica en un contexto real, consideramos que un curso mayor duración habría sido más adecuado, ya que el trabajo de la entonación precisa de tiempo e insistencia. La entonación no es un aspecto que se logre de un día para otro, se consiguen resultados a largo plazo y, cuanto más tiempo, más posibilidad de práctica y, por ende, más posibilidades de asimilar el trabajo y de mejorar y obtener resultados positivos.

Los estudiantes respondieron bien a la incorporación de la entonación en el plan del curso y siempre mostraron interés y curiosidad hacia la práctica de esta, pero ¿se han cumplido los objetivos de la propuesta? Parece que sí, tal como demuestran los ejercicios en clase, las encuestas de ejercicios de percepción y las grabaciones realizadas por los alumnos. Hay que tener en cuenta que más allá de que los estudiantes sepan entonar correctamente una interrogativa, en un nivel A1, lo importante es que sean conscientes de que existe una entonación española y que esta es relevante en la comunicación, y que conozcan su funcionamiento y sus características. Por ende, es recomendable que en un nivel A1 predominen los ejercicios de percepción, identificación y concienciación por encima de la producción.

Lo que se consiguió en este curso intensivo de nivel A1 no es baladí: los estudiantes, que antes no tenían en cuenta este elemento propio y tan importante de la lengua, ahora lo considerarán dentro de las demás competencias necesarias para conocer y aprender una lengua con éxito.

El experimento finalizó al acabar el curso intensivo y el mismo grupo clase no se mantuvo en el siguiente nivel. Por tanto, no se pudo continuar con el trabajo de la entonación con los mismos alumnos en un nivel A2. Tampoco se ha podido llevar a cabo un seguimiento y controlar si a lo largo del tiempo se ha mantenido este progreso.

Una futura propuesta de trabajo sería investigar si alumnos que han seguido un curso con la entonación integrada en el currículum, a largo plazo, tienen una mejor entonación que otros estudiantes que no hayan trabajado nunca este aspecto suprasegmental. Para poderlo comprobar, sería necesario contar con un grupo piloto control del mismo nivel. Analizando y comparando ambos grupos se obtendrían unos resultados que consideramos que serían muy interesantes y reveladores.

En la introducción se hablaba de la necesidad de crear actividades para la práctica de la entonación y la pronunciación en general. La realidad es que ya existe gran variedad de propuestas didácticas que pueden aplicarse a las clases de ELE, pero hay pocos docentes dispuestos a emplearlas en sus clases. Es necesario que, además de contar con estas propuestas, se utilicen en el aula. Del mismo modo, es imprescindible que se realicen investigaciones relacionadas con la aplicación de estas propuestas didácticas y metodológicas en el aula de ELE, que aporten resultados empíricos donde se demuestre la evolución y el perfeccionamiento de las destrezas comunicativas de los estudiantes gracias a la práctica de la entonación.

6. BIBLIOGRAFÍA

- Cortés, M. (2002). Didáctica de la entonación: una asignatura pendiente. *Didáctica (Lengua y literatura) España*, vol. 14, 65-67.
- Gil Fernández, J. (2007). *Fonética para profesores de español: de la teoría a la práctica*. Madrid: Arco/Libros.
- Lahoz, J.M. (2007). La enseñanza de la entonación en el aula de ELE: cómo, cuándo y por qué. En E. Balmaseda Maestu (ed.), *Las destrezas orales en la enseñanza del español L2-LE: XVII Congreso Internacional de la Asociación del Español como lengua extranjera (ASELE)*, vol.2, 705-719. Logroño: Universidad de La Rioja.
- Lahoz, J.M. (2012). La enseñanza de la entonación, el ritmo y el tempo. En Gil Fernández, J. (2012) *Aproximación a la enseñanza de la pronunciación en el aula de español*, 93-132. Madrid: Edinumen.
- Santamaría Busto, E. (2007). Enseñar prosodia en el aula: reflexiones y propuestas. En E. Balmaseda Maestu (Ed.) *Las destrezas orales en la enseñanza del español L2 -LE: XVII Congreso Internacional de la Asociación del Español como lengua extranjera (ASELE)*, vol.2, 1237-1250. Logroño: Universidad de La Rioja.
- Villaescusa, I. (2009). La enseñanza de la pronunciación en la clase de ELE. *El currículum de E/LE en Asia-Pacífico. I Congreso de Español como Lengua Extranjera en Asia-Pacífico (CE/LEAP)*. Manila: Instituto Cervantes de Manila y Embajada de España en Filipinas.

PROPUESTA METODOLÓGICA PARA LA EVALUACIÓN DE FACTORES PROSÓDICOS EN EXPOSICIONES ORALES DE ALUMNOS DE SECUNDARIA

Clara Luna García García de León, Juan María Garrido Almiñana

Laboratorio Antonio Quilis, Universidad Nacional de Educación a Distancia
claralunagarcia@gmail.com, jmgarrido@flog.uned.es

ABSTRACT

El objetivo de este trabajo es presentar una propuesta de rúbrica de evaluación de las competencias prosódicas de estudiantes de Secundaria, y los resultados de la evaluación de un conjunto de producciones orales de estudiantes llevada a cabo utilizando esta metodología. La rúbrica incluye diez descriptores de otros tantos parámetros fonéticos relacionados con la expresión de la prosodia, que fueron empleados para preparar una prueba en línea con la que 12 jueces evaluaron 16 producciones orales de 8 alumnos de Secundaria de un centro de la provincia de Madrid. Los resultados de la prueba muestran importantes divergencias en la evaluación de las muestras por parte de los jueces, que pueden achacarse a una falta de formación específica en el ámbito de la prosodia, pero también a problemas en la definición de los descriptores. Los resultados muestran también la satisfacción de los jueces por el formato y estructura de la rúbrica.

Palabras clave: prosodia, evaluación, enseñanza de lenguas, español

The goal of this work is to present a proposal of evaluation rubrics of the prosodic competences of Secondary students, and the results of the evaluation of a set of oral productions of students carried out using this methodology. The rubrics includes ten descriptors of ten phonetic parameters related to prosody expression, which were used to prepare an on-line test which was completed by 12 judges, who evaluated 16 oral productions of 8 Secondary school students of an education center in Madrid. The results of the test show important differences in the evaluation of the samples by the judges, which can be related to a lack of specific training in the field of prosody, but also to problems in the definition of the descriptors. The results show also the judges' satisfaction by the format and structure of the rubrics.

Keywords: prosody, evaluation, language teaching, Spanish

1. INTRODUCCIÓN

La enseñanza de la expresión oral en las aulas de Secundaria y Bachillerato presenta una serie de retos particulares que pueden complicar la tarea del docente y el aprendizaje del alumnado. La adquisición de competencias orales forma parte del currículo oficial de la asignatura de Lengua castellana y Literatura y es, además, una parte fundamental en la formación del alumnado, porque a lo largo de su vida académica, laboral y personal, deberán hacer un uso competente de su expresión oral en diversas situaciones comunicativas. Por ello, preparar a los alumnos para mejorar su competencia lingüística en el ámbito de la oralidad debe ser un objetivo académico de primer orden.

Tanto en Enseñanza Secundaria Obligatoria como en Bachillerato, el currículo oficial para la asignatura de Lengua castellana y Literatura incluye entre sus contenidos específicos el dominio de una serie de destrezas propias de la comunicación oral (BOE, 2015). Por ello, más allá de los beneficios que de por sí reporta al alumnado la adquisición de un nivel de competencia adecuado, existe una motivación normativa para atender estos aspectos educativos en el aula.

El desarrollo de las competencias de expresión oral requiere practicar el uso de los diversos componentes de la lengua en su faceta hablada (fonética, fonología, sintaxis, léxico, morfología y pragmática), además de aspectos sociolingüísticos y normativos. En el caso concreto del nivel fónico, el desarrollo de habilidades prosódicas (uso de entonación, ritmo, inserción de pausas...) para la

expresión oral es un aspecto fundamental, pero recibe hoy en día poca atención en las aulas. Uno de los motivos para este hecho, además de su falta de tratamiento en los materiales didácticos que se emplean en las aulas, es la dificultad que implica su evaluación objetiva. El profesorado se siente a menudo perdido en la maraña terminológica de los manuales académicos, y carece de instrumentos adecuados para abordar esta evaluación, en forma de rúbricas que les orienten en los elementos en que deben fijarse. Esto deriva en dificultades para los docentes a la hora de evaluar de manera objetiva y consistente la expresión oral en aspectos como la prosodia, que es el objeto de este estudio.

La definición de criterios para la evaluación de la prosodia se ha abordado en el ámbito del aprendizaje del español como lengua extranjera, pero no en el de la enseñanza obligatoria. Se incluyen los aspectos prosódicos, por ejemplo, entre los criterios de evaluación del DELE (Instituto Cervantes, 2014), aunque solo a partir del nivel B2, y en alguna otra propuesta de evaluación de la expresión oral en el aula de ELE (Carbó y Mora, 2008; Bores y Camacho, 2016). En general, los descriptores relacionados con la prosodia que se incluyen en estas propuestas son bastante generales y difíciles de aplicar de forma fácil y objetiva.

El objetivo de este trabajo es presentar una propuesta de rúbrica de evaluación de las competencias prosódicas de estudiantes de Secundaria, y los resultados de la evaluación de un conjunto de producciones orales de estudiantes llevada a cabo utilizando esta metodología.

2. RÚBRICA DE EVALUACIÓN

La propuesta de rúbrica que aquí se presenta incluye una serie de diez descriptores de evaluación que intentan ir más allá de propuestas anteriores, proporcionando pautas lo más concretas y objetivas posible para evaluar una serie de aspectos relacionados con el componente prosódico en la expresión oral, pero intentando, a la vez, ofrecer un instrumento manejable y no demasiado técnico, que pueda ser empleado por docentes de Secundaria sin formación específica en Fonética. Para ello, los descriptores intentan hacer unívoca la interpretación del elemento evaluado y deben, asimismo, facilitar su comprensión a los evaluadores para que estos puedan realizar ponderaciones coherentes y precisas. Los descriptores presentan dos tipos de información: una información técnica y una información simplificada, que permita comprender cada elemento prosódico sin tener conocimientos técnicos de fonética. Asimismo, las preguntas mantienen una terminología escogida para evitar que haya más de una interpretación, pero,

al mismo tiempo, mantienen la tendencia de los descriptores a transponer la información para hacerla fácilmente comprensible.

La lista de descriptores propuestos evalúa los siguientes aspectos de la prosodia de un hablante:

1. Rango tonal / variación melódica máxima
2. Intensidad / volumen de la voz
3. Longitud de las pausas vacías (silencios)
4. Cantidad de pausas vacías (silencios)
5. Cantidad de pausas llenas (vocalizaciones como: “eeeh”, “mmm”)
6. Cantidad de alargamientos
7. Pausas vacías en lugares inadecuados
8. Velocidad de elocución/habla
9. Repetición de patrones entonativos
10. Adecuación de los patrones entonativos al tipo de enunciado

La evaluación de cada uno de estos parámetros debe hacerse utilizando una escala de tres niveles (“defecto”, “bien”, “exceso”), para señalar si los valores de un determinado parámetro son demasiado bajos/escasos, si son adecuados o si son demasiado altos/excesivos.

A partir de la valoración por separado de cada uno de estos parámetros prosódicos, se puede obtener una calificación global de la prosodia de la muestra asignando un punto a cada una de las respuestas ‘bien’ obtenidas y sumando la puntuación total. El resultado es una calificación numérica en una escala 0-10.

Para comprobar la utilidad de esta rúbrica para la evaluación de las competencias prosódicas de estudiantes de enseñanza secundaria se llevó a cabo una prueba con docentes de español y grabaciones reales de presentaciones orales de alumnos de ESO.

3. EVALUACIÓN DE LA RÚBRICA

3.1. Muestras

Las muestras utilizadas fueron 16 grabaciones de presentaciones orales llevadas a cabo por 8 alumnos de cuarto curso de ESO del (dos grabaciones de cada alumno), recogidas por la primera autora de este trabajo en un centro educativo de enseñanza secundaria de la provincia de Madrid. Los alumnos tenían entre 15 y 16 años, eran todos hablantes monolingües de español, sin patologías de habla y audición aparentes o diagnosticadas y sin patologías psicológicas declaradas. Las grabaciones se llevaron a cabo en un aula del centro, como evaluación inicial y final de una actividad para mejorar las competencias de expresión oral llevada a cabo por la primera autora.

3.2. Jueces

Las muestras recogidas fueron evaluadas por 12 jueces, profesionales de la educación (profesores de la asignatura de Lengua castellana y Literatura en Educación Secundaria y Bachillerato y de español como lengua extranjera). Los jueces no recibieron ninguna instrucción o formación previa a la realización de la evaluación.

3.3. Procedimiento

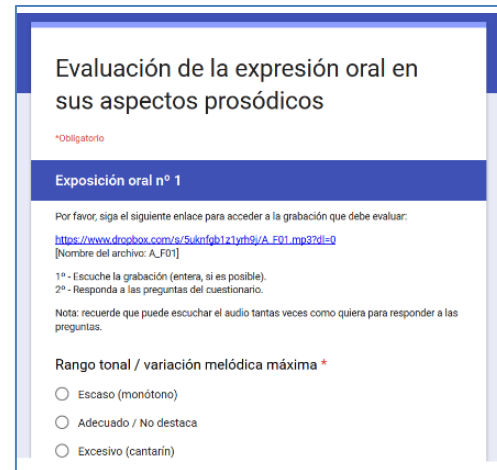
Se preparó una prueba de evaluación en línea con las 16 muestras de habla de los alumnos y la rúbrica descrita en el apartado anterior, utilizando *Google Forms*. La prueba incluía tres partes:

1. Introducción, instrucciones y descriptores de los parámetros
2. Estímulos y cuestionarios de evaluación (16 estímulos x 10 descriptores: 160 preguntas)
3. Cuestionario para recabar impresiones del evaluador (5 preguntas)

En la primera parte de la prueba, los jueces recibían una serie de instrucciones generales para realizar de manera adecuada el cuestionario: leer con detenimiento las indicaciones, leer con atención los descriptores de los elementos evaluados, mantener constante el volumen de su dispositivo (o tener en cuenta una variación de volumen, si fuese necesario realizarla) y tener presente que entre las respuestas a cada pregunta siempre hay una que indica que el elemento está en niveles adecuados y dos que señalan que requiere una corrección por exceso o por defecto.

En la segunda parte de la prueba, se presentaban a los participantes las diferentes muestras que debían evaluar utilizando la rúbrica desarrollada. La figura 1 muestra un ejemplo de pantalla de esta segunda parte.

Al finalizar la evaluación, se pidió a cada participante que completara un cuestionario de siete preguntas para recabar su opinión sobre la validez y utilidad del cuestionario. Se trataba de un cuestionario de respuesta cerrada en el que se daban varias opciones de respuesta para cada pregunta (dos o tres, dependiendo de la pregunta), para facilitar su posterior procesado.



Evaluación de la expresión oral en sus aspectos prosódicos

*Obligatorio

Exposición oral n° 1

Por favor, siga el siguiente enlace para acceder a la grabación que debe evaluar:
https://www.dropbox.com/s/5uknfgb121yfh5j/A_F01.mp3?dl=0
[Nombre del archivo: A_F01]

1º - Escuche la grabación (entera, si es posible).
2º - Responda a las preguntas del cuestionario.

Nota: recuerde que puede escuchar el audio tantas veces como quiera para responder a las preguntas.

Rango tonal / variación melódica máxima *

Escaso (monótono)

Adecuado / No destaca

Excesivo (cantarín)

Figura 1: Ejemplo de pantalla de la segunda parte del formulario de la prueba de evaluación.

3.4. Análisis

Los datos perceptivos obtenidos fueron analizados utilizando el paquete estadístico SPSS v. 22.0.0.0. Se han procesado de manera independiente los resultados del estudio perceptivo (cuestionarios de evaluación de la prosodia) y los cuestionarios de impresiones del evaluador.

3.5. Resultados y discusión

La figura 2 muestra las calificaciones globales otorgadas a cada muestra de forma separada para cada juez. Como puede observarse, los jueces calificaron las muestras de forma bastante desigual. El análisis estadístico de estos datos reveló que las evaluaciones de las muestras por parte de los jueces eran significativamente diferentes (chi cuadrado de Pearson = 0,001). Esto parece indicar una cierta heterogeneidad en los criterios de evaluación que aplicaron los jueces.

Con el uso de descriptores, se intentó homogeneizar el criterio de los jueces en el estudio perceptivo; sin embargo, a la vista de los resultados, es posible que una sesión de entrenamiento previa hubiese sido útil para unificar algo más sus estrategias y criterios de evaluación. Es decir, aun suponiendo que todos los jueces comprendiesen los contenidos de los parámetros, la falta de práctica en las evaluaciones técnicas, por elementos, de aspectos prosódicos ha podido dar lugar a una notable disparidad en su comportamiento.

Nota de los alumnos en función del evaluador o juez													
Sesión	Alumno	J1	J2	J3	J4	J5	J6	J7	J8	J9	J10	J11	J12
1	AF01	7	7	4	10	4	7	4	5	3	4	1	7
2	AF01	6	6	4	5	4	6	9	9	4	5	5	6
1	AF02	6	6	5	4	7	7	7	8	4	7	2	7
2	AF02	6	8	3	10	9	8	9	10	7	4	7	9
1	AM01	9	10	5	6	9	10	10	10	8	6	7	10
2	AM01	8	10	8	10	10	10	4	10	6	4	10	9
1	AM02	6	9	5	7	7	7	4	4	3	6	6	7
2	AM02	9	6	7	2	7	7	3	9	7	6	6	8
1	BF01	7	6	5	7	8	7	4	7	5	6	4	8
2	BF01	8	5	8	8	6	7	6	9	7	8	8	7
1	BF02	9	5	7	10	9	7	9	10	8	6	5	8
2	BF02	6	10	7	10	9	9	6	10	8	9	8	8
1	BMO1	5	9	5	6	7	10	6	10	8	6	4	9
2	BMO1	8	6	6	4	8	9	6	7	7	5	7	8
1	BMO2	5	5	5	3	6	6	1	9	5	5	3	7
2	BMO2	6	7	6	7	8	9	4	6	6	5	6	8

Figura 2: Resultados de la prueba de evaluación.

La heterogeneidad en la evaluación podría achacarse, pues, a diversos factores, no excluyentes entre sí:

- el nivel de conocimiento de los jueces sobre prosodia
- el nivel de comprensión de los descriptores por parte de los jueces
- el criterio de evaluación (más conservador / más liberal)
- el grado de modificación de sus criterios por efecto del cansancio o del entrenamiento
- sus propias experiencias como hablante
- sus capacidades perceptivas
- su capacidad general para realizar evaluaciones

Por su parte, el análisis de las respuestas de los evaluadores reveló que todos ellos consideraban que la evaluación sería más objetiva con el uso de un cuestionario como el que se les propuso y que la rúbrica propuesta sería positiva para el aprendizaje en el aula. El sistema de descriptores ha tenido una buena acogida (el 83,7% piensa que sería útil para el docente) y la presencia de información técnica e información simplificada ha sido la opción preferida por los evaluadores (91,7%). Puede decirse que, en general, la herramienta de evaluación ha sido valorada positivamente y que los descriptores son considerados útiles y necesarios.

4. CONCLUSIONES

Los resultados de la encuesta realizada al profesorado que llevó a cabo la evaluación parecen indicar que la herramienta descrita en este trabajo sería útil en el aula para evaluar las competencias prosódicas de los alumnos de Secundaria. Sin embargo, las divergencias en los resultados de la evaluación permiten realizar algunas reflexiones sobre las dificultades que implica la evaluación de la prosodia para el profesorado de Secundaria. La disparidad de

los resultados que surge del cotejo de sus respuestas podría deberse a una cierta carencia teórica, o bien, una cierta desatención de la prosodia en su labor profesional, lo cual haría que realizar evaluaciones como las del presente estudio les supusiera alguna dificultad. No se puede saber con certeza qué conocimientos tienen los evaluadores sobre prosodia; sin embargo, el hecho de que la respuesta mayoritaria a qué tipo de evaluaciones realizan sea la de una evaluación global con atención a algún parámetro específico lleva a la conclusión de que la desatención puede haber sido un factor de peso en esta cuestión.

En todo caso, las diferencias en las respuestas también pueden ser debidas a una falta de concreción o claridad en la definición de los descriptores, que debería ser explorada en futuros trabajos.

5. BIBLIOGRAFÍA

- Boletín Oficial del Estado (2015). Real Decreto 1105/2014, de 26 de diciembre, por el que se establece el currículo básico de la Educación Secundaria Obligatoria y del Bachillerato, sábado 3 de enero de 2015.
- Bores, M., Camacho, L (2016). Criterios para evaluar la expresión oral y escrita en la clase de E/LE. Actas del LI Congreso Internacional de la AEPE, Palencia.
- Carbó, C. & Mora, M. A. (2008). La evaluación de la expresión oral en la prueba única de español de las escuelas oficiales de idiomas en la Comunidad Valenciana. En Pastor, S., Roca, S (coords). XVIII Congreso Internacional de la Asociación para la Enseñanza del Español como lengua Extranjera (ASELE): Alicante, 19-22 de septiembre de 2007 / 2008, 213-227
- Instituto Cervantes (2014). Guía del examen DELE B2, Instituto Cervantes.

APROXIMACIÓN AL ANÁLISIS MELÓDICO DE LA ENTONACIÓN NEUTRA DEL ESPAÑOL HABLADO POR RUSOHABLANTES

Zoya Garmátina, Empar Devís

Universitat de Barcelona
zoya-ga14@mail.ru, devis@ub.edu

ABSTRACT

The objective of this work is to determine phrases in Spanish which present most difficulties for the Russian-speakers who speak this foreign language from an intonation point of view. The results obtained will be useful from the didactic point of view. With the results of this study it will be possible to elaborate a number of exercises which could help Russian-speakers speak Spanish with more similar intonation.

We may note, nowadays the results of Spanish studies by the Russian speakers show that they obtain lexical and syntactical knowledge rather successfully, but they still have some problems with the intonation.

To do the analysis, we recorded spontaneous-speaking conversations of native Russian speakers who have lived in Spain for at least two years, with an intermediate or advanced level of Spanish. In total, we have 15 hours of recording, of which we segment 70 neutral sentences. We use the method of Melodic Speech Analysis (Cantero 2002; Cantero & Font-Rotchés 2009; Cantero & Mateo 2011), which allows segmenting the intonation curve in tonal centers and analyzing the intonation independent of other levels of linguistic analysis.

The results obtained allow us a preliminary approach to the characterization of the Spanish intonation and emphatic features spoken by Russian speakers.

Keywords: intonation, Spanish intonation analysis, melodic analysis, Spanish and Russian, tonal segments, communication, language didactic

1. Introducción

Durante los últimos años ha habido un creciente interés en Rusia por el aprendizaje del español y, conocedores de la importancia de la prosodia en el habla espontánea, hemos centrado nuestro estudio en este nivel lingüístico. Como es bien sabido, la comunicación oral consiste en establecer contacto con los demás, y la entonación nos facilita establecer este contacto. Entre códigos compatibles, la posibilidad de intercomprensión oral es muy elevada, y la negociación entre los interlocutores es fácil, pero entre códigos ajenos, como es el caso del ruso y del español, no es tan sencillo.

En esta comunicación presentamos los primeros resultados de un estudio pionero sobre el análisis melódico de la entonación neutra del español hablado por rusohablantes. Este estudio forma parte de un estudio más amplio sobre la melodía

del español hablado por rusos que pretende identificar los rasgos melódicos distintivos de la entonación del español hablado por rusos y observar las diferencias, respecto a la entonación del español hablado por españoles, que podrían llevar a dificultar la intercomprensión entre hablantes de los dos idiomas (Zoya Garmatina, Tesis Doctoral en curso).

2. Corpus

El corpus está obtenido al grabar audios de conversaciones espontáneas de rusohablantes nativos que hablan español. Hemos hecho entrevistas a 10 personas de varios niveles y edades. Todas las personas tienen un nivel de español intermedio como mínimo y viven en España al menos desde hace dos años.

Para hacer el análisis, grabamos conversaciones de habla espontánea de diez rusohablantes nativos que llevan viviendo en España al menos dos años, con un nivel de español intermedio o avanzado. En total, contamos con 15 horas de

grabación, de las que segmentamos 70 enunciados neutros. Empleamos el método de Análisis Melódico del Habla (Cantero, 2002; Cantero & Font-Rotchés, 2009; Cantero & Mateo, 2011), que permite segmentar la curva entonativa en centros tonales y analizar la entonación independiente de otros niveles de análisis lingüístico.

2.1 Entrevistas

Las entrevistas se realizaban y se grababan personalmente en un sitio silencioso o a través de llamadas por Skype. Hablamos con cada uno sobre varios temas: música, arte, trabajo, estudios, comida, filosofía de vida, sus problemas de cada día. Cada entrevista duraba desde 30 minutos hasta 2 horas. En total obtuvimos 15 horas de grabaciones.

En algunos casos era bastante difícil que durante la conversación el interlocutor hablara como en la vida cotidiana; es decir, sin estrés, utilizando un vocabulario regular y expresando sus ideas con espontaneidad. El hecho de ser conscientes de que se les estaba grabando les cambiaba el estilo de habla. Para evitar este problema ofrecimos a los informantes participar en juegos intelectuales.

Los participantes contestaban a preguntas curiosas sobre varios temas y tenían que adivinar si algunas afirmaciones eran verdaderas o falsas. A continuación, se ven algunos ejemplos:

“El órgano más pesado del cuerpo humano es el cerebro. ¿Es verdad o no? Inglaterra y Reino Unido es lo mismo. ¿Es verdad o no?”.

Otro juego que organizamos se basaba en adivinar las circunstancias y causas de varios sucesos inventados. El presentador explicaba una situación, por ejemplo, “Un señor se marchó al trabajo y se puso la máscara. ¿Por qué lo hizo?”. Los que adivinaban la causa podían hacer varias preguntas, y el presentador contestaba solamente “sí o no” hasta que se adivinara la situación con todos sus detalles. De este modo, las personas estaban tan entretenidas con las preguntas que se les olvidaba que estaban participando en un estudio y así hablaban de manera más espontánea, expresando sus emociones. Como resultado, obtuvimos enunciados más completos (y no solamente con flexión final) y se pudo elaborar de nuevo una parte del corpus.

2.1 Establecimiento del corpus e informantes

De 15 horas totales de las grabaciones, segmentamos 70 enunciados neutros de 10 informantes: hombres y mujeres que habían vivido en España más de 2 años y que tenían un nivel B1 de idioma como mínimo.

3. Análisis melódico del habla

Para el análisis de los enunciados, empleamos el método de Análisis Melódico del Habla (Cantero, 2002) y su protocolo Cantero y Font-Rotchés (2009, 2020), que permite segmentar la curva entonativa en centros tonales y analizar la entonación independientemente de otros niveles de análisis lingüístico.

Una vez identificados los centros tonales de cada enunciado (que suelen corresponderse con las vocales), se etiquetaron en un textgrid de Praat (Boersma y Weenink, 2019) y se procedió al análisis, empleando el Script de Praat elaborado por Mateo (2010a, 2010b), que extrae los valores absolutos en Hz, los relativiza y dibuja el gráfico de la melodía estandarizada.

Obtenidas las melodías, para establecer los rasgos del perfil melódico, nos basamos en la propuesta de Cantero y Mateo (2011).

4. Resultados

Como resultado, hemos obtenido varios tipos de melodías neutras en el español hablado por rusohablantes que presentamos clasificadas según los rasgos entonativos.

Hemos identificado que los enunciados neutros del español hablado por rusohablantes tienen tres variantes de los cuerpos melódicos:

- plano,
- ondulado,
- con rama tensiva (primera parte suspensa).

Los enunciados los dividimos en dos tipos.

Tipo I – con final plano. Puede contener leves ascensos o descensos melódicos hasta 15%.

Dentro del tipo I hay tres subtipos:

- tipo Ia cuerpo plano (figura 1),

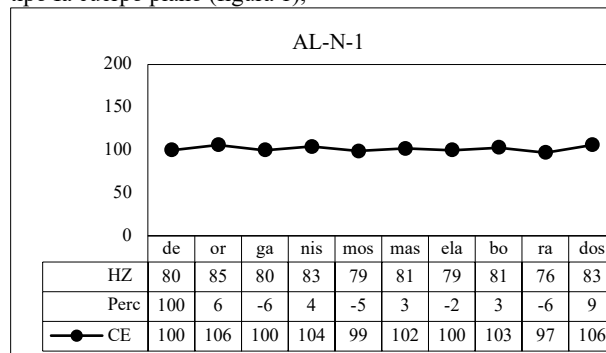


Figura 1: gráfico de la entonación neutra tipo Ia cuerpo plano.

- tipo Ib cuerpo ondulado (figura 2),

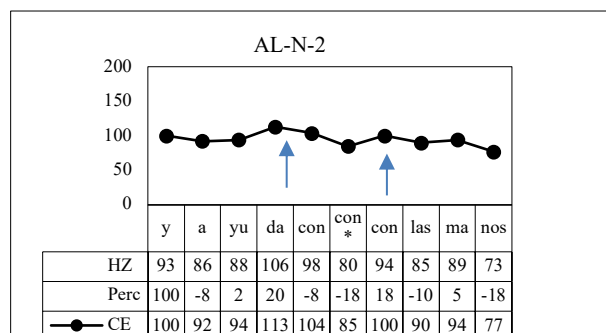


Figura 2: gráfico de la entonación neutra tipo Ib cuerpo ondulado.

- tipo Ic con rama tensiva (primera parte suspensa) (figura 3)

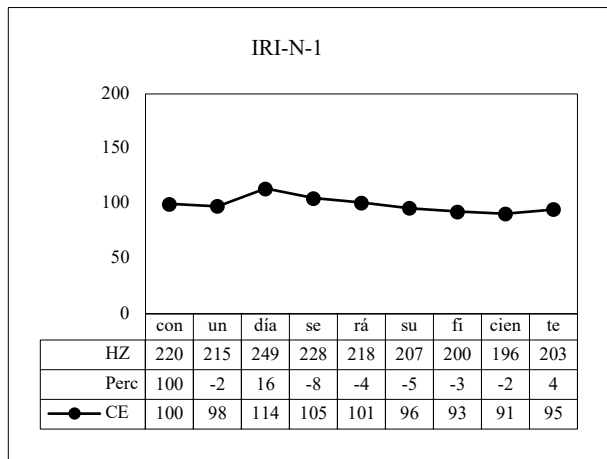


Figura 3: gráfico de la entonación neutra tipo Ic con rama tensiva (primera parte suspensa).

Tipo II – con final descendente más del +20%.
Dentro del II tipo hay tres subtipos:

- tipo IIa descenso desde núcleo elevado (figura 4),

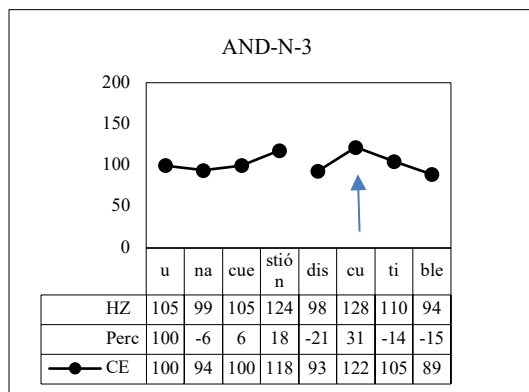


Figura 4: gráfico de la entonación neutra tipo IIa descenso desde núcleo.

- tipo IIb descenso desde pico central (figura 5),

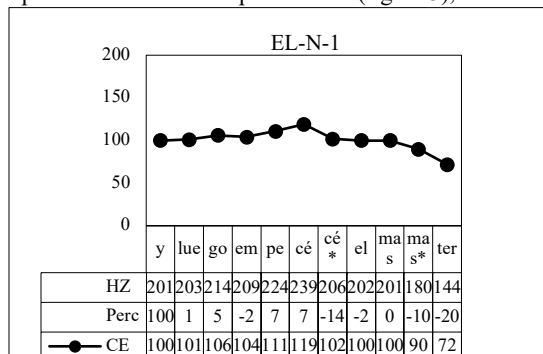


Figura 5: gráfico de la entonación neutra tipo IIb descenso desde pico central.

- tipo IIc descenso desde la última vocal tónica (figura 6).

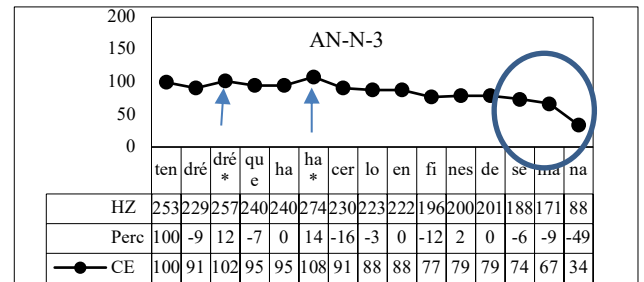


Figura 6: IIc descenso desde la última vocal tónica.

5. Conclusiones

Según el experimento que hemos realizado, tenemos dos tipos melódicos y cada uno de ellos tiene tres subtipos. Como podemos observar, no aparecen muchos rasgos de transferencia lingüística en los casos en los que los rusohablantes pronuncian las frases neutras en castellano, aunque evidentemente en los gráficos obtenidos están presentes las inflexiones internas que son muy características de la entonación rusa. Tenemos una breve descripción de cada tipo y el número y porcentajes obtenidos en la tabla 1.

Tipos	Descripción	Cantidad obtenida	% de los 70 enunciados
la cuerpo plano	Los enunciados con subidas y bajadas hasta 17%	22	31,43
lb cuerpo ondulado	Los enunciados con dos o más subidas y bajadas hasta 42%	11	15,71
Ic con rama tensiva	Los enunciados neutros con la primera parte suspensa	23	32,86
IIa descenso desde núcleo elevado	La entonación sube, y la vocal tónica o pretónica se eleva para después poder hacer el descenso	3	4,29
IIb descenso desde pico central	Los enunciados que tienen un pico central a partir de cuál comienza un descenso	7	10

IIC descenso desde la última vocal tónica	Los enunciados con las subidas y bajadas hasta 24% que se acaban es subidas desde la última vocal tónica	4	5,71
---	--	---	------

Tabla 1: tipología de los enunciados neutros.

Referencias bibliográficas

- [1] Cantero Serena, F. J. (2002): *Teoría y análisis de la entonación*. Barcelona, Edicions de la Universitat de Barcelona.
- [2] Cantero Serena, F.J. y Font Rotchés, D. (2009): Protocolo para el análisis melódico del habla, *Estudios de Fonética Experimental*, núm XVIII, 17-32.
- [3] Cantero Serena F.J. y Devís Herraiz, E. (2011): Análisis melódico de la interlengua, con Hidalgo, A.; Congosto, Y.; Quilis, M. (eds.), *El estudio de la prosodia en España en el s. XXI: perspectivas y ámbitos*. Valencia, Universitat de València, 285-299.
- [4] Cantero, F.J., Alfonso, R., Bartolí, M., Corrales, A., Vidal, M (2005): Rasgos melódicos de énfasis en español, *Phonica*
- [5] Cantero Serena, F. J. & Mateo Ruiz, M. (2011). Análisis melódico del habla. Complejidad y entonación en el discurso. *Oralia 14: 105-128*.
- [6] Devís Herraiz, E. y Cantero Serena, F. J. (2011): La entonación del español hablado por italianos. *Entonaciones del español acentos dialectales y acentos extranjeros*. Barcelona, Octaedro.
- [7] Goglova, T. (2005). La enseñanza del español a rusohablantes. En Cuadernos Cervantes [en línea]. Disponible en la web: <http://www.espanol.kiev.ua/rusohablantes.html>
- [8] Iruela, A. (2004): Adquisición y enseñanza de la pronunciación en lenguas extranjeras. Tesis doctoral. Barcelona, Universidad de Barcelona.
- [9] Mateo Ruiz, M. (2011): *Análisis melódico del habla: protocolo para la automatización de la obtención de los datos de la curva estándar*. Trabajo final de Master [en línea]. Disponible en la web: <http://diposit.ub.edu/dspace/handle/2445/16062>
- [10] Yen Hui Liu y Cantero Serena, F J. (2001): La entonación prelingüística del español hablado por taiwaneses: establecimiento de un corpus. *Actas del II Congreso de Fonética Experimental*. Sevilla: Universidad de Sevilla.

DISEÑO DE UNA BATERÍA DE PRUEBAS PARA LA EVALUACIÓN DE LA PROSODIA DEL ESPAÑOL EN EL MARCO DEL PROYECTO PROA

Juan María Garrido Almiñana¹, Eva Estebas-Vilaplana¹, María J. Machuca², Lorraine Baqué²,
Wendy Elvira-García¹, Victoria Escandell Vidal³

¹Laboratorio Antonio Quilis, Universidad Nacional de Educación a Distancia

²*Servei de Tractament de la Parla i el So*, Universidad Autónoma de Barcelona

³Universidad Complutense de Madrid

jmgarrido@flog.uned.es, eestebas@flog.uned.es, mariajesus.machuca@uab.cat, lorraine.baque@uab.cat,
welvira@flog.uned.es

ABSTRACT

El objetivo de este trabajo es presentar los criterios generales de diseño y las principales características de la batería de pruebas desarrolladas para la evaluación de la prosodia en español en el marco del proyecto ProA (*Prosody Assessment*). Las pruebas están organizadas en tres niveles de dificultad, según la complejidad de la función comunicativa evaluada, e incluyen tareas de lectura de enunciados cortos, de lectura de textos y de producción de enunciados no leídos. El objetivo final del proyecto es desarrollar una herramienta automatizada para la evaluación de la prosodia que implemente las pruebas desarrolladas.

The goal of this work is to present the general design criteria and main features of the set of tests for the evaluation of prosody in Spanish developed in the framework of the ProA (*Prosody Assessment*) project. The tests are organized in three difficulty levels, depending on the complexity of the evaluated communicative function, and include tasks of short-utterance reading, text reading and non-read material production. The final goal of the project is to develop an automated tool for prosody evaluation implementing these tests.

Keywords: prosodia, evaluación, enseñanza de lenguas, patologías del lenguaje, tecnologías del habla, español

1. INTRODUCCIÓN

Dentro del componente fónico de las lenguas, el nivel prosódico desempeña un papel fundamental en la comunicación lingüística. Los fenómenos (entonación, acento, velocidad de elocución, ritmo) y unidades suprasegmentales (sílabas, palabras fonológicas, grupos entonativos) son cruciales para la inteligibilidad de los mensajes, ya que transmiten información lingüística (estructura sintáctica, modalidad oracional, foco, énfasis, información nueva/conocida) o paralingüística (emociones, actitudes). La prosodia es también importante en los actos comunicativos por otras cuestiones no lingüísticas: se ha señalado, por ejemplo, que los oyentes tienden a evaluar el habla de los locutores de los medios audiovisuales de comunicación de forma positiva o negativa (como ‘agradable’ o ‘desagradable’, como ‘divertido’ o ‘aburrido’ o como ‘creíble’ o ‘poco creíble’) teniendo principalmente en cuenta su prosodia (Rodero, 2001, entre muchos otros). También se ha demostrado la importancia de

los parámetros prosódicos en la evaluación positiva o negativa del habla en los discursos políticos (Rosenberg y Hirschberg, 2005; Strangert y Gustafson, 2008).

La evaluación de la corrección o adecuación de los usos prosódicos es una tarea de interés innegable en varias disciplinas, como el aprendizaje/enseñanza de lenguas, la práctica de la expresión oral en público, la rehabilitación o terapia del habla, o las tecnologías del habla. La evaluación es de gran importancia como elemento inicial o final de un proceso de aprendizaje o rehabilitación, para establecer la calidad de las producciones prosódicas, normalmente por parte de un evaluador externo (un profesor, un logopeda o incluso un sistema automático) de la forma más objetiva posible (habitualmente por medio de una calificación numérica), pero también como parte del proceso en sí mismo, como medio de proporcionar a una persona referencias lo más concretas posible sobre qué aspectos de sus realizaciones prosódicas debería mejorar para obtener mejores resultados comunicativos.

La evaluación de las producciones prosódicas debería cumplir idealmente una serie de requisitos:

- 1) Ser global, es decir, cubrir todos los fenómenos y unidades prosódicos.
- 2) Ser adecuado para diferentes entornos de evaluación (enseñanza de lenguas, terapia lingüística, tecnologías del habla).
- 3) Basarse en criterios objetivos y fáciles de aplicar por parte de un evaluador humano o automático.
- 4) Utilizar un sistema de calificación apropiado y objetivo, que refleje de una forma ponderada el resultado de la evaluación de los diferentes criterios considerados.
- 5) Proporcionar a la persona evaluada una retroalimentación adecuada, que le permita identificar aquellos rasgos prosódicos de su habla que dificultan su comunicación. Esto implica que no debería limitarse a una evaluación global de su efectividad en términos comunicativos o estilísticos, sino que debería incluir también una evaluación diagnóstica, que analice por separado la producción fonética de los diferentes fenómenos prosódicos (melodía, duración segmental, intensidad, pausas) y que permita orientar a la persona evaluada sobre qué aspectos concretos de su producción prosódica debería modificar para mejorar.

A pesar de su innegable interés, es difícil actualmente encontrar métodos y herramientas, ya sean manuales o automáticos, para evaluar la prosodia y sus usos comunicativos, y menos aún que cumplan estos requisitos, al menos en lo que se refiere al español y al resto de lenguas de España. Así, en el ámbito del aprendizaje de lenguas, las pruebas estandarizadas para estudiantes de español como lengua extranjera no tienen en cuenta normalmente la evaluación de los aspectos prosódicos, y cuando lo hacen (en general como parte de la evaluación global de las producciones orales), los criterios de evaluación que se incluyen son muy generales y difíciles de aplicar de forma objetiva. Por ejemplo, el DELE (examen oficial de español lengua extranjera del Instituto Cervantes, Instituto Cervantes, 2014) solo incluye la entonación entre los criterios para evaluar la fluidez de la expresión oral a partir del nivel B2, y los descriptores de evaluación que se proporcionan para hacerlo son genéricos y difícilmente objetivables. Otras pruebas utilizadas también en el ámbito de la enseñanza del español (Carbó y Mora, 2008) o propuestas de rúbricas para la evaluación docente de la prosodia (Bores y Camacho, 2016; Verano-Tacoronte et al., 2016)

presentan problemas similares. Las propuestas probablemente más desarrolladas para la evaluación específica de la prosodia del español provienen del campo de la logopedia (Abarzúa et al 2005, Martínez-Castilla y Peppé, 2008; Zac, 2016), pero su cobertura no es global, no ofrecen criterios de evaluación y no están diseñadas para una evaluación diagnóstica detallada.

2. EL PROYECTO PROA

Teniendo en cuenta este marco, el proyecto ProA (*Prosody Assessment*) nace con el propósito de desarrollar métodos y herramientas que permitan evaluar la prosodia de producciones orales en español, ya sea de forma manual o automática. Sus objetivos generales son:

- 1) El desarrollo de una batería completa de pruebas para la evaluación diagnóstica de una serie de fenómenos suprasegmentales (entonación, acento, velocidad de elocución, organización de los enunciados en unidades prosódicas) en producciones orales del español que puedan utilizarse en diferentes entornos (aprendizaje de lenguas o de la expresión oral, terapia lingüística, tecnologías del habla).
- 2) El desarrollo de un prototipo de herramienta que permita la realización de las pruebas de forma lo más automatizada posible, que proporcione a la persona una calificación objetiva y comprensible para alguien no experto en prosodia y que ofrezca la retroalimentación adecuada para mejorar sus habilidades prosódicas.

Estos dos objetivos generales se concretan en siete objetivos más específicos:

- 1) El desarrollo de materiales para la evaluación de la corrección y adecuación de diferentes tipos de producciones orales en español.
- 2) La recogida de un corpus de producciones prosódicas 'normales' y 'desviadas' en español utilizando los materiales desarrollados.
- 3) La identificación de parámetros acústicos adecuados para la evaluación de estos fenómenos prosódicos.
- 4) El desarrollo o adaptación de herramientas para la extracción y procesamiento automáticos de estos parámetros.
- 5) La definición de modelos de referencia adecuados para la evaluación de dichos parámetros.

- 6) El desarrollo de métricas y criterios adecuados para la evaluación.
- 7) El desarrollo de herramientas automáticas que integren todos los elementos desarrollados previamente (métodos, modelos, métricas y criterios), para obtener una evaluación lo más objetiva posible y proporcionar una retroalimentación útil a la persona evaluada.

Para conseguir estos objetivos, se ha reunido un equipo de trabajo formado por lingüistas con intereses y experiencia en ámbitos como la comunicación audiovisual, la logopedia, la enseñanza de lenguas o las tecnologías del habla, además de un interés común por la prosodia y sus diferentes usos comunicativos, formado por investigadores del Laboratorio de Fonética Antonio Quilis de la Universidad Nacional de Educación a Distancia, del *Servei de Tractament de la Parla i del So* de la Universidad Autónoma de Barcelona, de la Universidad Complutense de Madrid y de la Universidad Pompeu Fabra.

3. DESCRIPCIÓN DE LAS PRUEBAS

Dentro de este marco general, el propósito de esta comunicación es presentar la batería de pruebas para la evaluación de la prosodia del español desarrollada en el ámbito de este proyecto y los criterios que han regido su proceso de diseño.

Las pruebas presentan una serie de características que las hacen diferentes de otras propuestas anteriores:

- a) A diferencia de otras pruebas existentes, el objetivo de cada prueba es evaluar la prosodia en función de un **uso comunicativo** determinado (expresión de diferentes tipos de actos de habla, emociones, actitudes, gestión del fondo común, contrastes informativos...).
- b) Los usos comunicativos evaluados están distribuidos en tres **niveles de dificultad**, inspirados de forma parcial en el plan curricular del Instituto Cervantes para la enseñanza de ELE (Instituto Cervantes, 2006).

Las pruebas se han desarrollado teniendo en cuenta criterios de usabilidad, como el hecho de que no contengan demasiados enunciados para que las personas que las realicen (especialmente en el caso de pacientes con patologías o estudiantes principiantes) no lleguen a acusar el esfuerzo de pruebas demasiado largas. También se han diseñado de la forma lo más modular posible, para que puedan utilizarse juntas o

por separado, según los intereses del evaluador o de la propia persona evaluada.

La batería de pruebas desarrollada teniendo en cuenta estos criterios incluye tres tipos de tareas:

- **Lectura de oraciones:** la persona evaluada debe leer en voz alta una serie de oraciones cortas con la prosodia más adecuada a la situación comunicativa que se le proporciona en cada caso.
- **Lectura de textos:** la persona evaluada debe leer en voz alta una serie de textos cortos, representativos de un tipo de lectura determinado.
- **Habla no leída:** la persona evaluada debe realizar una serie de tareas de expresión oral a partir de un material que se le presenta.

3.1. Pruebas de nivel básico

En el nivel básico de dificultad prosódica se ha incluido la evaluación de las siguientes funciones comunicativas:

- Expresión de modalidades oracionales básicas.
- Marcado prosódico de estructuras sintácticas básicas en oraciones simples.
- Realización prosódica del acento léxico.

Las pruebas desarrolladas son actividades de lectura de textos, que implican la lectura en voz alta por parte de la persona evaluada de enunciados cortos con la prosodia que considere más adecuada para la situación que se le plantea para cada enunciado, salvo en el caso de la prueba dedicada al acento léxico, en la que los enunciados que deben producir son autoexplicativos.

3.2. Pruebas de nivel intermedio

Las funciones comunicativas consideradas en el nivel intermedio son:

- Expresión de contrastes informativos.
- Expresión de actos de habla.
- Expresión de emociones básicas.
- Marcado prosódico de estructuras sintácticas en el ámbito de la oración compuesta.

Como en el nivel básico, las pruebas son actividades de lectura en voz alta de enunciados aislados en el marco de la situación que se plantea en cada caso a la persona evaluada.

3.3. Pruebas de nivel avanzado

Los usos comunicativos considerados en este nivel son:

- Expresión de actitudes.
- Corrección y adecuación prosódicas en la lectura de textos formales.
- Corrección y adecuación prosódicas en la lectura de textos expresivos.
- Corrección y adecuación prosódicas en exposiciones formales en público.
- Corrección y adecuación prosódicas en producciones informales en público.

En este caso, las pruebas incluyen:

- Actividades de lectura de enunciados aislados (expresión de actitudes).
- Actividades de lectura de textos (formales e informales) en el marco de una situación comunicativa concreta.
- Actividades de producción de habla no leída en el marco de una situación comunicativa concreta.

4. VALIDACIÓN DE LAS PRUEBAS

La utilidad de estas pruebas se validará en fases posteriores del proyecto, mediante su utilización como material de base para la recogida de un corpus de producciones orales con prosodia ‘normal’ y ‘alterada’ con personas de diferentes colectivos (aprendices de español como lengua extranjera, personas con patologías del habla y personas con español como lengua extranjera y sin dificultades detectadas en sus producciones orales) y la evaluación perceptiva de las producciones recogidas por parte de jueces expertos. Esta evaluación se centrará tanto en la determinación de hasta qué punto se cumple con éxito la función comunicativa esperada (evaluación comunicativa) como en la determinación del uso correcto o adecuado de los diferentes parámetros acústicos relacionados con la expresión de dicha función (evaluación diagnóstica).

5. RECONOCIMIENTOS

Este trabajo ha sido llevado a cabo con el apoyo financiero del Programa Estatal de Generación de Conocimiento y Fortalecimiento Científico y Tecnológico del Sistema de I+D+i, Proyectos de I+D de Generación de Conocimiento, del Gobierno de España (referencia PGC2018-094233-B-C21).

6. BIBLIOGRAFÍA

Abarzúa, R., Caradeux, M., Jeria, P., Viano, M., Zamorano, M. A. (2005). Diseño y aplicación de tareas para evaluar velocidad y ritmo, como componentes de

la fluidez, en niños preescolares entre 4 años o mes y 5 años 11 meses de edad, Universidad de Chile.

Bores, M., Camacho, L (2016). Criterios para evaluar la expresión oral y escrita en la clase de E/LE. Actas del LI Congreso Internacional de la AEPE, Palencia.

Carbó, C. & Mora, M. A. (2008). La evaluación de la expresión oral en la prueba única de español de las escuelas oficiales de idiomas en la Comunidad Valenciana. En Pastor, S., Roca, S (coords). XVIII Congreso Internacional de la Asociación para la Enseñanza del Español como lengua Extranjera (ASELE): Alicante, 19-22 de septiembre de 2007 / 2008, 213-227

Instituto Cervantes (2006). Plan curricular del Instituto Cervantes. Niveles de referencia para el español. Madrid: Instituto Cervantes, Biblioteca Nueva.

Instituto Cervantes (2014). Guía del examen DELE B2, Instituto Cervantes.

Martínez-Castilla, P., & Peppé, S. (2008). Intonation features of the expression of emotions in Spanish: preliminary study for a prosody assessment procedure, *Clinical Linguistics & Phonetics*, 22:4-5, 363-370.

Rodero, E. (2001). Los principales errores que debe evitar todo locutor de informativos radiofónicos: un estudio práctico. In Martínez, P. (ed.). *Reinventar la Radio*. Pamplona: Ediciones Eunat, 307-315.

Rosenberg, A., & Hirschberg, J. (2005). Acoustic/prosodic and lexical correlates of charismatic speech. *Interspeech 2005 proceedings*, 513–516. ^[1]_{SEP}

Strangert, E. & Gustafson, J. (2008). What makes a good speaker? subject ratings, acoustic measurements and perceptual evaluations. *Proceedings of Interspeech 2008*, 1688 -1691.

Verano-Tacoronte, D., González-Betancor, S. Bolívar-Cruz, A., Fernández-Monroy, M., & Galván-Sánchez, I. (2016). Valoración de la competencia de comunicación oral de estudiantes universitarios a través de una rúbrica fiable y válida, *Revista Brasileira de Educacao* 21(64), 39-60.

Zac, M. (2016). Test de evaluación prosódica en el habla, Editorial Akadia.

INVESTIGATING THE ROLE OF COGNATE STATUS IN L2 PRODUCTION AND PERCEPTION

Celia Gorba¹, Susana Cortés², Juli Cebrian¹, Núria Gavalda³

¹Universitat Autònoma de Barcelona, ²Universitat de les Illes Balears, ³Universidad Internacional de La Rioja
celia.gorba@uab.cat, susana.cortes@uib.cat, juli.cebrian@uab.cat, nuria.gavalda@unir.net

ABSTRACT

This paper presents the results of two exploratory studies on the effect of cognate status on the perception and production of L2 sounds. Previous studies show that cognate words are more likely to be mispronounced by L2 speakers than non-cognates, but whether the influence of the L1 is modulated by L2 experience is unclear. In addition, few studies have explored the presence of cognate effects (CE) in perception. A first study involving Spanish/Catalan learners of English of different levels of proficiency found evidence of CE in perception more evident with the more advanced learners. However, no CE were found in production, possibly due to the nature of the production task. A second study failed to replicate CE in perception but found more evidence of CE in production with a less controlled task. The discrepant results are discussed in terms of the role of proficiency, metalinguistic knowledge, word frequency and task characteristics.

Keywords: cognates, production, perception, second language speech

1. INTRODUCTION

Cognate words are words that share semantic, phonological and orthographic properties in two languages, such as the Spanish words *estación*, *pétalo* and *hábito* and the English words *station*, *petal* and *habit*, respectively. The study of how cognate words are processed and produced by bilingual speakers can shed light on the relationship between the phonological systems of a bilingual speaker's two languages. This is so because when bilingual speakers use one of their languages, the lexicon of their language-not-in-use continues to be accessed (Costa, 2006). In particular, when words are cognates in the speakers' L1 and L2, the word in the language-not-in-use may affect pronunciation and lexical decision accuracy.

The effects of cognate status can be both beneficial and detrimental. On the one hand, a greater level of L1 activation takes place when processing L2 cognate words, which leads to faster vocabulary learning and access to lexical meaning (Tessel et al., 2018). On the other hand, the similarity in the form of these words has been reported to have a detrimental effect on pronunciation accuracy as cognate words tend to show a greater influence from the L1 during L2 production. For example, Spanish-dominant Spanish-Catalan bilinguals have been found to produce Catalan /ɔ/ and /ɛ/ (not found in Spanish) as Spanish /o/ and /e/ more often with cognate than with

non-cognate words (Amengual, 2016; Mora & Nadeu, 2012). Similarly, Spanish/English bilinguals tend to produce stop VOT less accurately in their less dominant language in cognates (Goldrick, Runnqvist & Costa, 2014; Flege & Munro, 1994).

Regarding the relationship between L2 proficiency levels and cognate effects in production, studies show inconsistent results. While some previous research shows no clear relationship (Amengual, 2012), others suggest that cognate effects decrease as proficiency increases (Kroll et al., 2002; Jacobs et al., 2016). Finally, no studies have analysed cognate effects in purely perceptual tasks (cf. Tessel et al., 2018, which is an EEG study). Therefore, the goal of this paper is to explore further the relationship between cognate words and accuracy in production as well as the possible effect of cognate status in the perception of L2 sounds.

2. STUDY 1

In this first study, the perception and production of a set of English sounds in cognate and non-cognate words by speakers with different proficiency levels in L2 English was assessed.

2.1. Methodology

2.1.1. Stimuli

The target sounds in this study involved the English consonants /v, p, b, ɹ/. The segments examined differed phonetically or phonologically in the L1 and L2, as described in Table 1. The target sounds appeared in words that are cognates in English and Spanish/Catalan and words that are not. A trained

Phonological differences	/v/	An En. phoneme that does not exist in Sp.	C: video NC: viewer
Phonetic differences	/p/	+VOT in En., 0 VOT in Sp.	C: petal NC: pillow
	/b/	Spirantisation in Sp. /b/ > [β]/V_V	C: debate NC: above
Control	/ɹ/	Obvious difference between Sp. /r/ and En. /ɹ/	C: radio NC: runner

bilingual English-Catalan speaker produced two versions of each target word: L1 (Spanish)-accented versions and native-like (English) pronunciations of the target words. The L1-accented versions involved [b] for target /v/, unaspirated /p/, spirantised /b/ and [r] for target /ɹ/.

Table 1: Phonological variables with examples. (Sp. = Spanish, En. = English, C = cognate, NC = non-cognate)

2.1.2. Participants and tasks

A total of 26 L1 Spanish-Catalan L2 English speakers and 7 native English speakers participated in the study. The L2 English speakers were divided into three groups according to their proficiency level: intermediate, advanced and proficient. The intermediate group was made up of 7 2nd-year English Studies students who had taken a 6-credit course in English phonetics and phonology, the advanced group consisted of 10 4th-year students of the same degree who had taken 18 credits in English phonetics and phonology, and participants in the proficient group were faculty members in an English Department. Participants performed a forced choice goodness task in which listeners were presented with two versions of the same word and had to indicate whether the first or second word was clearly a more English-like pronunciation, or if they were equally fine. Thus, trials differed in terms of the word combination presented: accented + native, native + accented, or native + native words. Half the trials involved cognate words shared by English and Spanish and Catalan (e.g., *video*, *panic*, *debate* and

robot), and the other half were non-cognate words. The tasks were created and administered using Praat (Boersma & Weenink, 2016). A subset of the participants, the group of advanced learners, also completed a picture naming task in which they produced a number of words, including cognate and non-cognate L1 (Spanish) and L2 words containing initial stops and /v/.

2.2. Results of Study 1

Regarding the perception task, the percentage of correct responses was calculated for each group and each segment. The results are given in Figure 1 and 2. Participants tended to be more successful at identifying the accented word with non-cognate words than with cognate words, illustrating a cognate effect on L2 perception. Statistical analyses on the percentage of correct responses for each segment yielded a significant difference between cognates and non-cognates for /v/, /b/ and /p/ [$t(32) = -3.15$; $p = .003$; /b/ $t(32) = -2.43$; $p = .020$; /p/: $t(32) = -4.73$; $p = .000$]. As for /ɹ/, which acted as a control variable predicted to be easily identified when mispronounced, no cognate effect was found [$t(32) = 1.99$; $p = .055$].

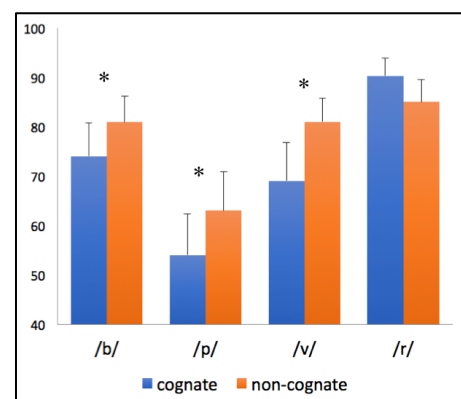


Figure 1: Perception task results. Percentage of accurate responses per segment.

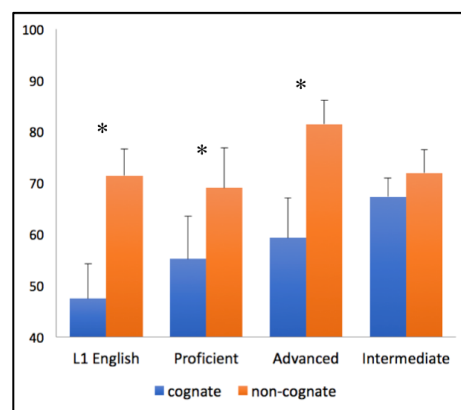


Figure 2: Perception task results. Percentage of accurate responses per group (accented + native combinations).

When comparing groups (see Figure 2), cognate effects were found for the advanced and proficient groups, and even for the native speaker group. That is, these groups displayed a higher percentage of correct responses in non-cognates than in cognates [advanced: $p = .000$; proficient: $p = .012$; L1 English: $p = .003$]. Therefore, cognate status seems to have a negative effect on the ability to detect mispronunciations in the L2. This effect appears to increase with L2 proficiency and is found even for native English speakers who have lived in a Spanish-Catalan environment for a long while. This is in contrast with previous findings that showed greater effects with lower proficiency learners (Kroll et al., 2002; Jacobs et al., 2016). The cognate effect found in native speakers of English could be explained by the fact that their knowledge of Spanish may have affected the perception of their L1. Alternatively, having exposure to Spanish/Catalan-accented English may have resulted in greater tolerance toward typical Spanish mispronunciations.

The production data was analysed acoustically (VOT duration for initial voiceless stops) and impressionistically (accuracy of /v/ production). Speakers tended to produce L1 and L2 initial stops with relatively target-like values in both languages, regardless of cognate status. In addition, speakers showed an overall accurate production of English /v/. This lack of cognate effects does not replicate previous findings (Amengual, 2012; Flege & Munro, 1994), and may be accounted for by the type of production task and the metalinguistic knowledge on the part of the advanced learners. First, this group of participants (4th-year English majors who had completed three courses on English phonetics and pronunciation) was metalinguistically aware of differences between the L1 and the L2. Second, the production of isolated words, and the inclusion of both L1 and L2 words, may have favoured a careful pronunciation of English sounds, particularly those known to be challenging for Spanish learners of English. Thus, a further study, Study 2, was designed to examine further possible cognate effects in perception and to evaluate production by means of a less controlled task.

3. STUDY 2

An exploratory follow-up perception and production study was carried out in order to refine the testing of specific segments in cognates and non-cognates by a group of advanced learners of English, which was comparable to the advanced group in Study 1 (found to show cognate effects in perception). In this second study, an attempt was made to elicit production with a task that allowed for less attention to form.

3.1. Methodology

3.1.1. Stimuli

The variables involved in this second study were the same as in study 1: (/p/, /b/, /v/ and /r/), with the addition of a further control variable, which was the production or omission of an epenthetic vowel before a consonant cluster beginning with /s/. A trained phonetician, who had a native-like command of both English and Spanish, produced both the English native-like productions and the Spanish accented productions. A native speaker of English listened to all the stimuli and detected correct and accented pronunciations in all instances.

3.1.2. Participants and tasks

Twelve advanced learners of English who had taken 18 credits in English phonetics and pronunciation have completed the study, which is still in progress. Participants performed both a perception and a production test. The perception test consisted in a forced choice goodness task with the new stimuli where listeners had to choose which pronunciation of a given word, the first or the second, sounded more accurate and English-like. In this case, there was no “equally good” option, but participants indicated for each trial how confident they were about their response. The production task consisted of reading a story that contained a number of words with the phonological variables under study, including some of the words in the perception test. The story reading was especially designed for this study in order to avoid the greater attention to form derived from the isolated word production in Study 1.

3.2. Results of Study 2

Preliminary results of the perceptual task reveal that performance was at ceiling for a number of individuals (83-100% accuracy), yielding a less clear pattern of overall cognate effects. This outcome runs counter to the results obtained for the advanced learners in Study 1. This may be due to methodological differences between the tasks, and the effect of metalinguistic knowledge, as discussed below.

Descriptive results for production show variation depending on the type of variable: speakers show no cognate effect on /v/ production, while they do with /p/ and /b/. On the one hand, overall results on /p/ VOT duration show that non-cognates tend to be produced with longer – i.e., more English-like – VOT values than non-cognates, although cognate effects seem to vary depending on the speaker, since 5 speakers show cognate effects whereas 7 do not.

Moreover, the cognate effect is more evident with high frequency words, such as ‘pizza’, which tend to be produced with more Spanish-like VOT values than a less frequent word like ‘panda’. Hence, further data are needed to explore the role of word frequency on cognate effects. On the other hand, regarding /b/, the analysis of the intensity difference between /b/ and the following vowel (a greater difference indicates a more stop-like production) show a slight cognate effect, since the intensity ratio is slightly lower – more approximant-like – with cognate words than with non-cognates. A closer inspection of individual differences reveals that cognate effects also vary depending on the speaker, since 7 speakers show cognate effects whereas 5 do not.

5. GENERAL DISCUSSION AND CONCLUSIONS

The results of the first study show that participants were more accurate at identifying the accented word in non-cognate words than with cognate words, which indicates an overall cognate effect on L2 perception. These results provide evidence of cognate effects in perception, which complements similar outcomes regarding production and supports previous findings about the activation of the lexicon(s) of the languages-not-in-use in multilingual speakers (Costa, 2006). Further, the comparisons between groups in the first study show proficiency effects in perception, since cognate effects were found in more advanced learners, proficient learners, and even in native speakers who are living in a Spanish-Catalan environment. These results contrast with previous findings that suggest that cognate effects decrease with proficiency (Kroll et al., 2012; Jacobs et al., 2016). Besides, a closer inspection of individual segments reveal that the percentage of accurate responses was lower for /p/ than for /v/, which might indicate that L2 speakers may be worse at detecting accentedness at the phonetic level than at the phonological level.

Results of the second study show ceiling effects in perception and greater individual variation in production, resulting in a less clear pattern of overall cognate effects. Several factors may account for this difference. On the one hand, the advanced learners in Study 2 were undergraduate students who had been selected for their high mark in the oral exam of an oral skills course. This may imply a greater and more recent metalinguistic knowledge than the most advanced groups in Study 1. On the other hand, methodological differences in the perception task may also account for the discrepant results. In Study 2 participants were forced to choose which of two words sounded more native-like. In contrast, Study 1

involved the possibility of both words being native-like, so listeners did not know that one option was necessarily more native-like.

In terms of production, results from the two studies reveal no cognate effects with /v/. Speakers were overall accurate in the production of this variable regardless of cognate status. The type of task does not seem to have an effect, since both the isolated words in Study 1 and the words in context in Study 2 show a similar lack of cognate effects for this variable. On the other hand, the production of /p/ revealed no cognate effects in Study 1, since speakers tended to produce English initial stops with target-like values regardless of cognate status. This may be explained by the greater attention-to-form allowed in the production of words in isolation. In contrast, in Study 2 there was a general tendency for cognates to have shorter VOT durations than non-cognates, in line with previous studies (Amengual, 2012; Flege & Munro, 1994). Thus, a more naturalistic task seemed to favour cognate effects in this case. Yet, cognate effects may be affected by word frequency, with greater L1 influence with more frequent words. Future studies should explore this issue further. As regards the results for /b/, this variable also showed a general cognate effect, since intensity ratios were slightly lower (which indicates more spirant-like realisations) in cognates words than in non-cognates, although more data and inferential tests are needed to confirm this tendency. Finally, the results obtained so far seem to indicate that cognate effects may be more evident in phonetic traits such as /p/ and /b/ that do not involve a phonological difference, whereas, at least in these more metalinguistically-aware participants, cognate effects are not present in a trait that has phonological status such as /v/. Further data on phonetic and phonological variables, with a larger sample and different proficiency levels are necessary to evaluate and confirm these results.

This research was supported by research grants FFI2017-88016-P from the Spanish Ministry of Economy and Competitiveness, and by the research group Grant 2017SGR34 from the Catalan Government.

6. REFERENCES

- Amengual, M. (2012). Interlingual influence in bilingual speech: Cognate status effect in a continuum of bilingualism. *Bilingualism*, 15(3), 517–530.
- Amengual, M. (2016). Cross-linguistic influence in the bilingual mental lexicon: Evidence of cognate effects in the phonetic production and processing of a vowel contrast. *Frontiers in Psychology*, Apr;7:617.

- Boersma, P., & Weenink, D. (2016). Praat: *Doing phonetics by computer* (version 5.3.56) [computer program].
- Costa, A. (2006). Speech Production in Bilinguals (Chapter 8). *The Handbook of Bilingualism*, 201–223.
- Flege, J.E., & Munro, M.J. (1994). The word unit in second language speech. *Studies in Second Language Acquisition*, 16, 381–411.
- Goldrick, M., Runnqvist, E., & Costa, A. (2014). Language switching makes pronunciation less native-like. *Psychological Science*, 25, 1031-1036
- Jacobs, A., Fricke, M., and Kroll, J. (2016). Cross-language activation begins during speech planning and extends into second language speech. *Language Learning*, 66, 324–353.
- Kroll, J.F., Dussias, P.E., Bogulski, C.A., & Valdes-Kroff, J. (2012). Juggling two languages in one mind: What bilinguals tell us about language processing and its consequences for cognition. In B. Ross (Ed.), *The Psychology of Learning and Motivation*, Volume 56 (pp. 229-262). San Diego: Academic Press.
- Mora, J.C., & Nadeu, M. (2014). L2 effects on the perception and production of a native vowel contrast in early bilinguals. *International Journal of Bilingualism*, 16(4), 484-500.
- Tessel, C.A., Levy, E. S., Gitterman, M., & Shafer, V. L. (2018). Neurophysiological indices of the effect of cognates on vowel perception in late Spanish-English bilinguals. *Journal of Phonetics*, 68, 117–137.

PERCEPTION OF L1 SPEECH OF LATE CZECH-FRENCH BILINGUALS BY CZECH MONOLINGUALS

M. Hevrova^{1,2}, B. Köpke¹, T. Bořil²

¹ Octogone-Lordat, University of Toulouse UT2 (France)

² Institute of Phonetics, Charles University (Czech Republic)
marie.hevrova@univ-tlse2.fr

ABSTRACT

This study examines the perception of L1 speech of 14 late Czech-French bilinguals who have been living in France for at least two months by Czech monolinguals. We wanted to know whether native speakers judge the L1 production of late Czech-French bilinguals differently from the L1 production of Czech monolingual. The perceptual experiment involved samples taken from L1 production of the bilinguals and Czech monolinguals in a reading task and a semi-spontaneous speech task and was submitted to 17 Czech students of phonetics in the Charles University in Prague. The results show a significant difference between the perception of L1 speech of bilinguals and monolinguals and between the two production tasks.

Keywords: L1 attrition, cross-linguistic influence, perception, late Czech-French bilinguals

1. INTRODUCTION

In the field of Second Language Acquisition (SLA), Sharwood Smith (1983) introduced the term of cross-linguistic influence (CLI) to refer to the influence of the learner's L1 on the L2 as well as the one of the L2 on the L1. The first case, i.e., CLI of the L1 on the L2, was and remains a topic of numerous studies at the phonetic level (see, e.g., Aoyama & Guion, 2007; Colantoni & Steele, 2007; Curtin, Goad & Pater, 1998; Flege, 1987b; Kijak, 2009; Major, 1986). However, the second case (the influence of the L2 on the L1) has received less attention and is investigated mostly in research on L1 attrition, defined as "the natural consequence of decrease in the use of a language" (Köpke, 2019: 365). Following Major & Baptista (2009), there are now several studies on phonetic language attrition and CLI in L1 production (see, e.g., Bullock, Dalola, & Gerfen, 2006; Chang, 2012; de Leeuw, Mennen & Scobbie, 2012; Stoehr, Benders, Van Hell, & Fikkert, 2018) while studies of L2 impact on L1 perception are less frequent (cf. Cancila, Celata, & Giannini, 2005; Major & Baptista, 2009; Ventureyra, Pallier, & Yoo, 2004).

Moreover, the language combinations investigated are still limited. For example, there is no study on the influence of L2 French on L1 Czech at the phonetic level, although several interesting phonetic contrasts exist between these languages

both at segmental and suprasegmental levels (cf. Paillereau, 2015; Skarnitzl, Šturm, & Volín, 2016). Hence, our study focuses on phonetic attrition and CLI of L1 Czech of late Czech-French bilinguals (hereafter CF) living in France. In this paper, we investigate whether there are perceptual differences between the L1 speech produced by CF and Czech monolingual speakers, i.e., Czechs living in Czechia (hereafter C).

Studies on phonetic CLI and attrition are generally based on acoustic measures (cf. Dmitrieva, Jongman, & Sereno, 2010; Lord, 2008; Major, 1992; Mayr, Price & Mennen, 2012; Mennen, 2004, among others) as well as perceptual experiments (e.g., de Leeuw, 2009; Sancier & Fowler, 1997). The latter are often used prior to acoustic analysis in order to test the correspondence between perceptual and acoustic findings (de Leeuw, 2009). Similarly, the present study reports on the results of a Perceptual Test (hereafter PT), which will be enriched by the results of acoustic analyses later on.

2. METHOD

2.1. Material

The speech samples used in the PT have been elicited in a reading task (RT) and in semi-spontaneous speech (SS). In the RT, the speakers read a short paragraph chosen from the book *Jak se co dělá*. O

lidech (Čapek, 1960). In the SS, they talked for one minute and a half in Czech about one or more proposed topics such as plans for holidays or the next weekend, describing a typical day, job, studies, family, hobbies, etc. Bilinguals were asked not to mention that they live in France. Reading and semi-spontaneous speech are commonly used in studies of phonetic CLI and attrition (e.g., Bullock *et al.*, 2006; de Leeuw, 2009; Major, 1992; Mennen, 2004) and their use here will facilitate comparison across studies. It also allows us to examine whether phonetic attrition and CLI are more obvious in semi-spontaneous speech than in reading, as observed by Major (1992).

Flege's study (1984) showed that native speakers are highly accurate at identifying non-natives in segments of only 30ms. Schmid and Hopp (2014) report that full phrases or sentences are usually used for perceptual experiments; however, there seems to be no study investigating whether rating is more accurate in longer samples. Therefore, we used for our PT speech samples containing full sentences or clauses varying in duration from 1.2 to 13.28 seconds (average 5.23s). The samples were extracted from the production of two groups of speakers, C and CF. The distractors were extracted from the production of speech synthesis "Amazon Polly" (Text-to-Speech on AWS, <https://aws.amazon.com/polly/>) and one female French native speaker who both read a Czech text with segmental characteristics as close as possible to Common Czech but with French prosody.

CF and French native speaker were recorded in a quiet recording studio (PETRA) at University Toulouse Jean-Jaurès using a Neumann TLM 49 microphone and sound card MOTU ULMk3. They received a small reward for participation. Audio files obtained from speech synthesis were played in high-quality loudspeakers in this studio and recorded with the same material. This procedure aimed at reducing the slightly artificial sound background of speech synthesis sound files and rendering these sound files more authentic. C were recorded in a quiet, comfortably furnished office with a low level of ambient noise and short natural reverberation in Prague. A head-mounted condenser microphone (Bayerdynamic Opus 55) was plugged directly into a pocket recorder set to uncompressed 48 kHz 16-bit mode.

All recordings were orthographically transcribed in Praat (Boersma & Weenink, 2019). The original 44 100 kHz recordings of D and CF were resampled to 48 kHz according to the original 48 Hz C' recordings. This resampling was necessary for running the

experiment in Praat. We also adjusted the loudness of samples in Audacity in a way that all samples were perceptually similar in loudness.

The PT comprised 77 samples. There were 56 samples from CF (38 samples from SS and 18 from RT), 14 samples from C (9 from SS and 5 from RT), and 7 samples from D (5 from SS and 2 from RT). For each speaker, we extracted at least one sample and at most six samples. From the production of CF, we chose the samples with phonetic phenomena unusual for Common Czech but not rare in the CF' production. To the best of our knowledge, there are no studies investigating whether the proportion of bilinguals' and monolinguals' samples in perceptual experiments influences the perception of phonetic attrition and CLI. For SLA, Flege and Fletcher (1992) showed that, with a higher proportion of native samples, the non-native samples create an impression on the listener to be heavily foreign-accented. This is why the proportion of bilinguals' versus monolinguals' samples in perceptual experiments in studies of phonetic attrition is usually varied although there is no rule indicating the right proportion (cf. de Leeuw, 2009; Schmid & Hopp, 2014). For the present study, we followed Schmid and Hopp (2014), who used four times more bilinguals' than monolinguals' samples.

2.2. Speakers

As shown by Hollien & Ship (1972) and Chevrie-Muller, Dodart, Seguir-Dermier & Salomon (1971), the age from 20 to 50 years matches the period of f0 stability (Chevrie-Muller *et al.*, 1971; Hollien & Ship, 1972). Hence, we included only speakers aged from 20 to 50 years in our study.

2.2.1. Control group

The monolingual group (C) comprised 11 native Common Czech speakers (2 males, 9 females, mean=34.83 years) living in the Central Bohemian region of the Czech Republic.

2.2.2. Experimental group

The bilingual group (CF) included 14 native Common Czech speakers (1 male, 13 females, mean=34.43 years) living in the Toulouse area ($n=13$) or Paris ($n=1$). Following Lang & Davidson (2017), we divided CF speakers into two sub-groups: five of them (E) had a Length of Residence (LOR) in France of less than 5 years (mean=2,32 years) and nine (A) for 5 years and more (mean=14,55 years).

2.3. Listeners

The PT was administered to 17 native speakers of Czech, students at Charles University in Prague. Sixteen of them were students in the first year of the bachelor of phonetics, and one was a Ph.D. student in phonetics. None reported any hearing or speech disorders.

2.4. PT procedure

The total duration of the PT (each sample presented only once) is approximately 25 min. The testing took place in a quiet room using headphones.

A forced-choice identification test (cf. McGuire, 2010), was created in the Praat multiple forced-choice (MFC) environment. Participants had to rate whether the sample seemed to be “absolutely Czech” or “absolutely French” on a 5-degree scale (1=“absolutely Czech,” 5=“absolutely French”). They were allowed to replay each sample five times, pass over to the next sample with the ‘Next’ button, or correct their response with the ‘Oops’ button. Samples were randomized and separated by a desensitization beep. Every twenty samples, the participants were invited to take a short break, where they could listen to a short song (approx. 30s).

The PT was preceded by a training session involving seven items in order to check comprehension of the instruction asking them to judge samples of Czech that may be produced either by a French speaker, a Czech living in the Czech Republic or a Czech living in France on a scale of 1 to 5 (see above). If they noticed phenomena unusual for Czech, they were invited to list them in an Excel sheet (only for ratings 2 to 4). These observations helped us enrich the list of phonetic phenomena to be examined in the acoustic study.

2.5. Analysis

The data were analyzed in RStudio (R Core Team, 2019) using the packages *lme4* (Bates, Maechler, Bolker & Walker, 2015), *dplyr* (Wickham, François, Henry & Müller, 2019), *rPraat* (Bořil, & Skarnitzl, 2016), and *ggplot2* (Wickham, 2016). We computed a mean value of the responses provided for each speaker (meanResponse) in one or both tasks as follows taken into account (1) means per item, (2) means per speaker. The significance level was set at $\alpha = 0.05$ (confidence interval=95%).

In order to test our hypothesis, we performed two linear mixed-effects models. In the first model, we analysed the relationship between group and

meanResponse and between task and meanResponse. In the second, we analysed the relationship between sub-group and meanResponse and between task and meanResponse. For both models, we had intercepts for speakers and subjects as random effects. As fixed effects, we entered group and task into the first model and sub-group and task into the second model. In both models, the fixed effects were tested with and without interaction.

Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality. P-values were obtained by likelihood ratio tests of the full model with the effect in question against the model without the effect in question. The distractors were excluded from all analyses.

3. RESULTS

Fig. 1 shows a visible difference in meanResponse depending on the group, task, and sub-group factors confirmed statistically by the two models. In the first model, the analysis showed that meanResponse was affected by group ($\chi^2(1)=10.644$, $p=0.0011$) and task ($\chi^2(1)=13.566$, $p=0.0002$), but there was no significant inter-dependence between group and task ($p=0.1385$). In the second model, the analysis

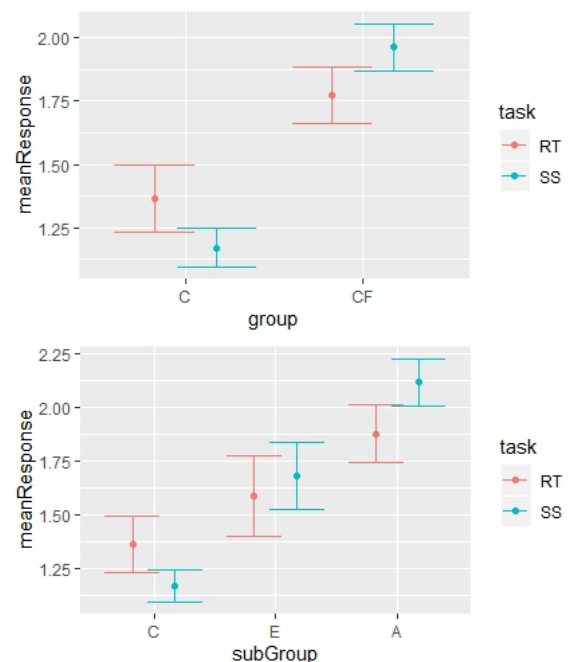


Figure 1: MeanResponse by group and by sub-group. RT=reading task, SS=semi-spontaneous speech, C=Czech monolinguals, CF=late Czech-French bilinguals, E=CF with LOR < 5 years, A=CF with LOR > 5 years.

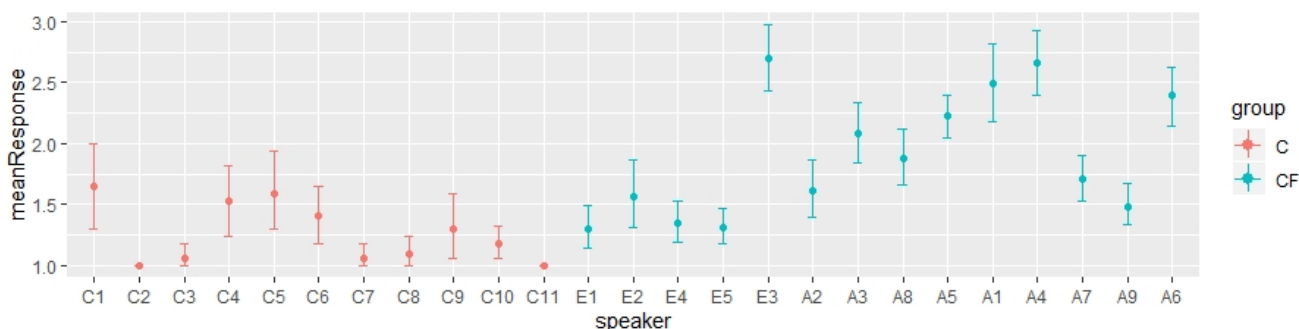


Figure 2: MeanResponse by speaker. CF are ordered from left to right with increasing LOR. C=Czech monolinguals, CF=late Czech-French bilinguals, E=CF with LOR < 5 years, A=CF with LOR > 5 years.

showed that meanResponse was affected by sub-group ($\chi^2(2)=13.219$, $p=0.0013$), and task ($\chi^2(1)=13.401$, $p=0.0003$), but there was no significant inter-dependence between sub-group and task ($p=0.1027$).

In brief, as expected, monolinguals speakers (C) were evaluated as speakers whose production was closest to typical Czech pronunciation, then came the bilingual Czechs with LOR < 5 years (E), and then the bilingual Czechs with LOR > 5 years (A). The distractors' production was evaluated as "absolutely French" with meanResponse of 4.91 ± 0.10 for RT and of 4.93 ± 0.06 for SS. Moreover, CF were better evaluated in the RT than SS. Interestingly, the contrary was observed for the control group: monolinguals were rated less native-like in the RT than in SS.

Results per speaker detailed in fig. 2 show that inter-speaker variation is higher in CF speakers than in C speakers. Ordering speakers by LOR suggests no linear relation between LOR and meanResponse.

4. DISCUSSION AND CONCLUSIONS

The present study showed that production in Czech of Czechs living in France is perceived by native speakers of Czech with experience in phonetics as deviating from the usual pronunciation of the Czech language. The study also demonstrated that in the bilinguals' groups, the influence of French is more perceptible in semi-spontaneous speech than in the reading task. These two main results confirm our initial hypothesis based on previous studies (de Leeuw, 2009; Major, 1992; Schmid & Hopp, 2014).

This study showed the important inter-speaker variation in both bilinguals' sub-group and no linear relation between LOR and perceived phonetic CLI. Indeed, to the best of our knowledge, no study in phonetic CLI and attrition proved any universal LOR since which phonetic CLI appears in L1 production. It was shown that phonetic drift may appear after 5

weeks of L2 intensive classes (cf. Chang, 2010, 2012) or one hour of intensive training of target vowels (cf. Kartushina, Hervais-Adelman, Frauenfelder & Golestani, 2016). Nevertheless, other studies did not confirm such a rapid shift (see "Study Abroad group" in Lang & Davidson, 2017). Likewise, Chang (2013) showed that phonetic drift occurs more in novice learners than in experienced learners. Conversely, other studies found more drift in graduate learners' (Herd, Walden, Knight & Alexander, 2015) or in attriters' production (Lang & Davidson, 2017) than in beginning learners, and that the better speakers' mastery of L2 is, the more L2-like their L1 production (Major, 1992). Because of the discrepancy of these results and missing linear relation between LOR and meanResponse in fig. 2, we suppose that perceived phonetic CLI is due to multiple sociolinguistic factors such as speaker's language use, proficiency, and others, and not only to LOR.

Finally, the results of PT merits to be verified by acoustic analysis. The listeners' observations invite us to explore, namely vowels' quantity and quality, intonation, accent, the pronunciation of /r/, /r̥/, and /r̥̃/ sound, and fricative consonants. In the oncoming study, we will compare the results of acoustic analysis with the results of PT and explore them in relation to sociolinguistic factors.

5. ACKNOWLEDGEMENTS

This research was supported by a doctoral grant of the French Research Ministry, a Charles University PhD scholarship, the Charles University project SVV – 2019, *Jazyk a nástroje pro jeho zkoumání* 2019 and the Charles University project Progres Q10, Language in the shiftings of time, space, and culture.

6. REFERENCES

- Aoyama, K. & Guion, S. (2007). Prosody in second language acquisition: an acoustic analysis on duration and F0 range. In O.-S. Bohn and M. J. Munro (eds.), *The role of language experience in second-language speech learning: in honor of James Emil Flege*, 281–97. Amsterdam: John Benjamins.
- Boersma, Paul & Weenink, David (2019). Praat: doing phonetics by computer [Computer program]. Version 6.1.08, retrieved 5 December 2019 from <http://www.praat.org/>
- Boril, T., & Skarnitzl, R. (2016). Tools rPraat and mPraat. In P. Sojka, A. Horák, I. Kopeček, & K. Pala (Eds.), *Text, Speech, and Dialogue* (pp. 367–374). Springer International Publishing.
- Bullock, B. E., Dalola, A. & Gerfen, C. (2006). Mapping the Patterns of Maintenance versus Merger in Bilingual Phonology: The Preservation of [a] vs. [ɑ] in Frenchville French. In J.-P. Y. Montreuil (Ed.), *New Perspectives on Romance Linguistics* (pp. 15–29). Amsterdam/Philadelphia: John Benjamins.
- Cancila, J., Celata, C., & Giannini, S. (2005). Phonological attrition and perceptual decay: The consonant length feature in the Lucchese community of San Francisco (CA). Paper presented at ICFLA 2005, Amsterdam.
- Chang, C. B. (2010). First language phonetic drift during second language acquisition. PhD thesis, University of California, Berkeley, CA.
- Chang, C. B. (2012). Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics*, 40 (2), 249–268. <https://doi.org/10.1016/j.wocn.2011.10.007>.
- Chang, C. B. (2013). A novelty effect in phonetic drift of the native language. *Journal of Phonetics*, 41 (6), 520–533.
- Chevrie-Muller C., Dodart F., Seguir-Dermier N. & Salomon D. (1971). Etude des paramètres acoustiques de la parole au cours de la schizophrénie de l'adolescent. *Fo/ia Phoniatria*, 23, 401–428
- Colantoni, L., & Steele, J. (2007). Acquiring /R/ in context. *Studies in Second Language Acquisition*, 27, 381–406.
- Curtin, S., Goad, H., & Pater, J. (1998). Phonological transfer and levels of representation: the perceptual acquisition of Thai voice and aspiration by English and French speakers. *Second Language Research* 14: 389–405.
- Čapek, K. (1991). *Od člověka k člověku III*. Praha: Československý spisovatel.
- de Leeuw, E. (2009). When your native language sounds foreign: A phonetic investigation into first language attrition. PhD thesis, Queen Margaret University, Edinburgh, United Kingdom.
- de Leeuw, E., Mennen, I., & Scobbie, J. M. (2012). Singing a different tune in your native language: First language attrition of prosody. *International Journal of Bilingualism*.
- Dmitrieva, O., Jongman, A., & Sereno, J. A. (2010). Phonological neutralization by native and non-native speakers: the case of Russian final devoicing. *Journal of Phonetics*, 38(3), 483–492.
- Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48. doi:10.18637/jss.v067.i01.
- Flege, J. E. (1984). The detection of French accent by American listeners. *The Journal of Acoustical Society of America*, 76, 692–707.
- Flege, J. E. (1987b). The production of “new” and “similar” phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15: 47–65.
- Flege, J.E., & Fletcher, K. L. (1992). Talker and listener effects on degree of perceived foreign accent. *Journal of the Acoustical Society of America*, 91, 370–389.
- Herd, W. J., Walden, R. L., Knight, W. L., & Alexander, S. N. (2015). Phonetic drift in a first language dominant environment. *Proceedings of Meetings on Acoustics*, 23, 060005.
- Hollien, H., & Shipp, T. (1972). Speaking fundamental frequency and chronologic age in males. *Journal of Speech and Hearing Research* 15, 155–159.
- Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2016). Mutual influences between native and non-native vowels in production: Evidence from short-term visual articulatory feedback training. *Journal of Phonetics*, 57, 21–39. <https://doi.org/10.1016/j.wocn.2016.05.001>
- Kijak, A. (2009). How stressful is L2 stress? A cross-linguistic study of L2 perception and production of metrical systems. Utrecht: Landelijke Onderzoekschool Taalwetenschap.
- Köpke, B. (2019). First language attrition: from bilingual to monolingual proficiency? In De Houwer, A. & Ortega, L. (Eds.), *The Cambridge Handbook of Bilingualism* (pp. 349–365). Cambridge: CUP.
- Lang, B. & Davidson, L. (2017). Effects of exposure and vowel space distribution on phonetic drift: Evidence from American English Learners of French. *Language and Speech*, 0023830917737111
- Lord, G. (2008). Second language acquisition and first language phonological modification. *Proceedings of the 10th Hispanic linguistics symposium*. ed. Joycece Bruhn de garavito and E. Valenzuela, 184–193. Somerville, MA: Cascadilla proceedings project.
- Major, R. C. (1986). The ontogeny model: Evidence from L2 acquisition of Spanish r. *Language Learning* 36: 453–503.
- Major, R. C. (1992). Losing English as a first language. *The Modern Language Journal* 76: 190–208.
- Major, R. C. & Baptista, B.O. (2009). First language attrition in foreign accent detection. In M. A. Watkins, A. S. Rauber, B. O. Baptista (eds.), *Recent research in Second Language Phonetics/Phonology: Perception and production*, 256–267. Newcastle Upon Tyne: Cambridge Scholars Publishing.
- Mayr, R., Price, S., & Mennen, I. (2012). First language attrition in the speech of Dutch–English bilinguals: The case of monozygotic twin sisters. *Bilingualism: Language and Cognition*, 15(04), 687–700.
- McGuire, M. (2010). A brief primer on experimental designs for speech perception research (Tech. Rep.). UC Santa Cruz. [CK]
- Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. *J. Phonetics*, 32, 543–563.
- Paillereau, N. (2015). Perception et production des voyelles orales du français par des futures enseignantes tchèques de Français Langue Etrangère (FLE). PhD Thesis. Université Sorbonne Paris Cité.
- R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, 25(4), 421–436. <https://doi.org/10.1006/jpho.1997.0051>
- Schmid, M. & Hopp, H. (2014). Comparing foreign accent in L1 attrition and L2 acquisition: Range and rater effects. *Language testing* 31(3), 367–388.
- Sharwood Smith, M. (1983). On explaining language loss. In R. Felix & H. Wode (Eds.), *Language development at the crossroads* (pp. 49–59). Tübingen: Gunther Narr Verlag.
- Skarnitzl, R., Šturm, P. & Volín, J. (2016). *Zvuková báze řečové komunikace: Fonetický a fonologický popis řeči*. Praha: Karolinum. 170 p.
- Stoehr, A., Benders, T., Van Hell, J., & Fikkert, P. (2018). Heritage language exposure impacts voice onset time of Dutch–German simultaneous bilingual preschoolers. *Bilingualism: Language and Cognition*, 21(3), 598–617. doi:10.1017/S1366728917000116
- Ventureyra, V. A. G., Pallier, C., & Yoo, H-Y. (2004). The loss of first language phonetic perception in adopted Koreans. *Journal of Psycholinguistics*, 17, 79–91.
- Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.
- Wickham, H., François, R., Henry, L. & Müller, K. (2019). dplyr: A Grammar of Data Manipulation. R package version 0.8.3. <https://CRAN.R-project.org/package=dplyr>

FRICATIVAS ALVEOLARES DEL EUSKERA: APROXIMACIÓN A SU CARACTERIZACIÓN ARTICULATORIA

Alexander Iribar, Rosa Miren Pagola, Itziar Túrrez

Equipo Fonetiker, Laboratorio de Fonética, Facultad de Ciencias Sociales y Humanas, Universidad de Deusto
alexiribar@deusto.es, rpagola@deusto.es, itziar.turrez@deusto.es

ABSTRACT

Se presenta una descripción articuladora experimental de las fricativas sibilantes /s/ y /ʃ/ del euskera, basada en el análisis de 15 colecciones de imágenes de MRI, en 2D y en 3D, correspondientes a 11 informantes. Se identifican dos modelos articulatorios para /s/ y tres para /ʃ/.

The articulation of the Basque sibilant fricatives /s/ and /ʃ/ is described by means of the analysis of 15 collections of magnetic resonance images, in 2D and 3D, corresponding to 11 informants. Two articulatory models for /s/ and three for /ʃ/ are identified.

Keywords: sibilantes, articulación, MRI-2D-3D, euskera. sibilants, articulation, MRI-2D-3D, Basque

1. INTRODUCCIÓN

El euskera posee un subsistema de seis sibilantes –en proceso de defonologización parcial–, compuesto por 3 fricativas y 3 africadas, definidas habitualmente como alveolares dorsales /s, ts/ (o /ʃ, tʃ/), alveolares apicales /s̺, t̺/ y prepalatales /ʃ, tʃ/ (Pagola et al. 2014). Estos elementos fueron estudiados a partir de imágenes mediosagitales de resonancia magnética (MRI) en Elejabeitia, Iribar y Pagola (2008a, 2008b). Estos trabajos iniciales mostraron la necesidad de profundizar especialmente en la descripción articuladora de la pareja alveolar, cuyos modelos de producción no quedaban suficientemente diferenciados entre sí.

Para avanzar en la descripción articuladora del par fricativo /s, ʃ/, se aumenta el número de los informantes, se incorporan las reconstrucciones en 3D y se aplica una diferente colección de parámetros articulatorios, ya contrastada en un trabajo anterior aplicado a las consonantes velares del español (Iribar, Pagola y Túrrez 2019).

2. METODOLOGÍA

2.1. Obtención del material

Las imágenes analizadas pertenecen a las siguientes colecciones:

- DAREMOSE, obtenida en el proyecto *Descripción articuladora experimental y modelizada de los sonidos del euskera* (ref. PI2004-1). Se trata de imágenes mediosagitales de 4 informantes repitiendo las secuencias [sa] y [ʃa]. La figura 1 muestra un ejemplo representativo de las dos articulaciones.

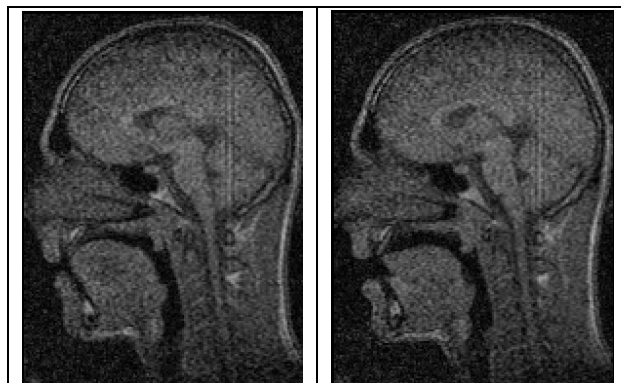


Figura 1: Realización de [s] y [ʃ] de un mismo informante (corpus DAREMOSE).

- DAELPACE, obtenida en el proyecto *Descripción articuladora experimental de la lengua: parámetros articulatorios del castellano y el euskera mediante cine-MRI* (ref. FFI2009-10706). Contiene imágenes mediosagitales de 7 informantes emitiendo aislada y sostenidamente los elementos en cuestión. La figura 2 muestra unos ejemplos representativos.

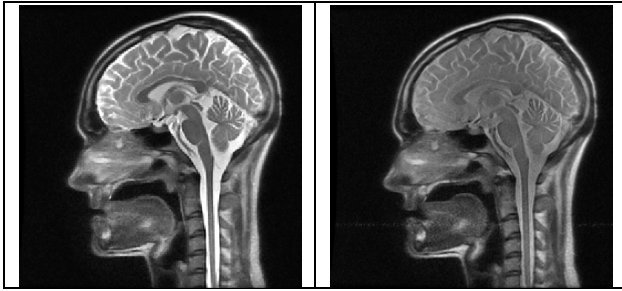


Figura 2: Realización de [s] y [ʃ] de un mismo informante (corpus DAELPACE).

- CASE-3D, obtenida en el proyecto *Caracterización articulatoria de la producción de los sonidos del euskara a partir de la modelización 3D del tracto vocal* (ref. PI2013-34). Consiste en la reconstrucción tridimensional de la lengua y la cavidad oral de 4 informantes, que emiten ambos elementos de la misma manera que en la colección anterior. La reconstrucción se realizó a partir de las imágenes de MRI desde la triple perspectiva sagital, axial y coronal. Los ejemplos correspondientes se muestran en la figura 3.

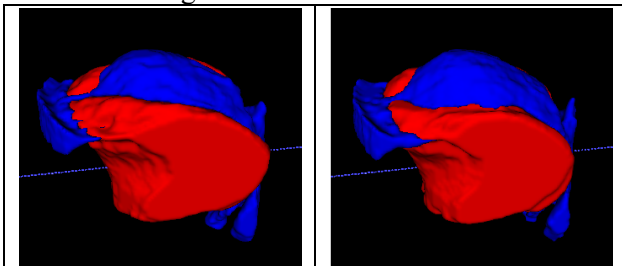


Figura 3: Realización de [s] y [ʃ] de un mismo informante (corpus CASE-3D).

Más información sobre la metodología utilizada para la obtención y el análisis de las tres colecciones puede encontrarse, respectivamente, en Elejabeitia, Iribar y Pagola (2008a), Pagola, Iribar y Túrrez (2012) e Iribar, Pagola y Túrrez (2013).

2.2. Informantes

La información relativa a los 15 informantes analizados, todos ellos universitarios, se resume en la Tabla 1:

Tabla 1: Información sobre los informantes analizados en el trabajo.

Inf	Corpus	Origen	Sexo	Edad
01	Daremore	G – Oiartzun	H	40
02		G – Zestoa	H	38
03		N – Leitza	M	18
04		G – Donostia	M	25
05	Daelpace	G – Mutriku	H	18

06	Case-3D	G – Azpeitia	H	48
07		G – Hondarribia	H	23
08		N – Leitza	M	23
09		G – Donostia	M	30
10		B – Gernika	M	20
11		N – Bera	M	22
12		G – Azpeitia	H	54
13		B – Igorre	M	23
14		B – Gernika	M	23
15		G – Donostia	M	34

Cuatro de los quince informantes han sido grabados para dos colecciones, de modo que son la misma persona los siguientes informantes: 03 y 08; 04 y 09; 06 y 12; 10 y 14.

Las realizaciones de todos los informantes fueron previamente analizadas en el Laboratorio para comprobar su adecuada distinción. Por otra parte, todos fueron oportunamente entrenados para las sesiones de grabación de las imágenes.

2.3. Análisis

Para el análisis de las imágenes se han tenido en cuenta los siguientes parámetros articulatorios, cualitativos y cuantitativos:

- Parámetros de constricción:
 - Articulador pasivo: dentoalveolar / alveolar.
 - Articulador activo: ápice / ápice-lámina / dorso.
 - Dimensiones de la constricción: longitud, largo, ancho y volumen.
 - Constricción secundaria: uvular / faríngea.
- Parámetros de lengua:
 - Contacto incisivos inferiores: sí / no.
 - Ápice: destacado / redondeado.
 - Dorso: recto / curvado.
 - Cuerpo: recto / convexo / cóncavo-convexo.
 - Depresión central: sí / no.
 - Plano: recto / ascendente / descendente.
 - Punto más alto del arco lingual: paladar / frontera velo-palatina / velo.
 - Raíz: recta / atrasada.
 - Volumen porción ápice.
 - Volumen porción ápice-dorso.
- Parámetro de labios:
 - Distancia interlabial.
 - Protrusión.
- Parámetros de cavidades (área de la sección midsagital y volumen):
 - Cavidad externa.
 - Cavidad bucal.
 - Cavidad faríngea.

Todos los parámetros se han obtenido mediante el software DicomPas (García Arroyo et al. 2013) para las imágenes en 2D y el programa ITKSnap (Yushkevich y Gerig 2017) para las de 3D. Las variables continuas se han convertido también en discretas (con los valores grande/mediano/pequeño) para facilitar un primer análisis conjunto.

3. PRINCIPALES RESULTADOS

Se observa una gran variación en los procedimientos de producción, pero también se encuentran algunas semejanzas y regularidades que permiten establecer dos modelos articulatorios para /s/ y tres para /ʃ/, cuyos parámetros se muestran en la tabla 2.

Tabla 2: Caracterización articulatoria de los cinco modelos identificados. (Las abreviaturas se corresponden con las iniciales de los parámetros expuestos en 2.3.)

	/s/		/ʃ/		
	Z1	Z2	S1	S2	S3
Art. Pas.	DA	DA	A	A(da)	DA
Art. Act.	D	AL	A	D	D
Long.	(G-M)p	M(p-g)	P(mg)	P(mg)	G
C. sec.	F(u-Ø)	U-F	Øf	Fu	Uf
C.i.i	S>N	N	N	N(s)	S
Ápice	R>D	D	D	D	R
Dorso	C>R	R	R	R-C	C
Cuerpo	(R-CX)>CCX	CCX-R	(R-CCX)cx	R(cx-ccx)	R-CX
Plano	A-D	R	R>A	A-D	A-R
Pmadal	(V-P)vp	V-VP	(VP>P)av	P(v-vp)	VP-P
Raíz	A-R	R	Ar	Ar	R
LL	M(p-g)	M-G	G>M>P	M	P-M
CE	M(p-g)	M(p-g)	Mp	P-M	P-M
CB	M(p-g)	M(p-g)	P>G>M	Pmg	P-M
CF	M(p-g)	Mg	Mpg	M	M
CB > CF	Ns	S	Sn	S	S-N

Las combinaciones de estos modelos establecen cuatro grupos de informantes (Z1-S1, Z1-S2, Z1-S3 y Z2-S1). Si la realización de /s/ es dorsal, entonces /ʃ/ puede ser dorsal o apical; si /s/ es apical, entonces /ʃ/ también lo es.

Por otra parte, ni los modelos, ni los informantes, ni, por supuesto, los parámetros distinguen los dos elementos con la misma potencia. Estas diferencias se examinarán con cierto detalle.

4. ALGUNAS CONCLUSIONES

La articulación de /s/ es ligeramente más adelantada que la de /ʃ/ y presenta un carácter dorsal con mayor frecuencia que /ʃ/. Pero ambos elementos pueden articularse con el ápice o con la lámina. Se han identificado 5 modelos articulatorios distintos, pero no es descartable que pueden encontrarse más. El modelo predominante de /s/ es dentoalveolar dorsal y el de /ʃ/, alveolar apical.

Los informantes no siempre distinguen los dos elementos del mismo modo. Los parámetros más potentes para la distinción son la cavidad externa, la abertura labial y las dimensiones del canal articulatorio. Algunas características se producen casi únicamente en los modelos dentoalveolares dorsales. Con esta información, sigue sin aclararse del todo la cuestión de la transcripción fonética de los elementos: <z> podría transcribirse como [s̺], [s̺̟] o incluso como [s̺̟̟]; para <s>, podrían utilizarse [s] y [s̺].

4. BIBLIOGRAFÍA

- Elejabeitia, A., Iribar, A. y Pagola, R.M. (2008a): Nuevos datos acerca de la articulación de las sibilantes vascas. *New trends in experimental phonetics: selected papers from the IVth International Conference on Experimental Phonetics*, A. Pamies y E. Melguizo (eds.), Language Design, Special Issue, 1, 55-62.
- Elejabeitia, A., Iribar, A. y Pagola, R.M. (2008b): El cine-MRI aplicado a la descripción de las sibilantes vascas. *Estudios de Fonética Experimental*, XVIII, 145-160.
- García Arroyo, J.L., García Zapirain, M.B., Oleagordia, I., Méndez, A. (2013). Marco de trabajo tecnológico para la realización de estudios de caracterización articulatoria sobre imágenes MRI. *Estudios de Fonética Experimental*, XXII, 367-404.
- Iribar, A., Pagola, R.M. y Túrrez, I. (2013). Caracterización articulatoria de ELE en español y euskara. *Estudios de Fonética experimental*, XXII, 129-169.
- Iribar, A., Pagola, R.M. y Túrrez, I. (2019). La articulación de las consonantes velares en español. *Estudios de Fonética Experimental*, XVIII, 125-160
- Pagola, R.M. et al. (2014-). *Euskal Hotsak Aholkatzen*. [http://euskalhotsakahoskatzen.deusto.eus/].
- Pagola, R.M., Iribar, A. y Túrrez, I. (2012): La descripción articulatoria de los sonidos en euskara y castellano: el proyecto DAELPACE. *Marcos interpretativos de la realidad social contemporánea*, Acillona, M. (ed.), Universidad de Deusto, Bilbao, 107-118.
- Yushkevich, P.A. y Gerig, G. (2017). *ITK-Snap v. 3.6.0*. [http://www.itksnap.org]

MODELOS MIXTOS Y CRITERIOS DE INFORMACIÓN EN FONÉTICA

Miguel Jiménez-Bravo

Universidad Complutense de Madrid
miguel.jimenez.bravo@ucm.es

ABSTRACT

La actual presentación, como parte de la sesión monográfica *Nuevas metodologías para el estudio fonético*, trata sobre cómo el análisis de los datos en fonética está cediendo al uso de modelos lineales mixtos (MLM) para solventar problemas asociados a ANOVA. Este cambio de paradigma está permitiendo abordar preguntas de investigación más ambiciosas y realizar análisis más complejos. Igualmente, el uso de criterios de información es cada vez más común en el proceso de validación de hipótesis y se ha propuesto que, junto al valor- p , también se comuniquen otros, como la tasa de falsos positivos.

Palabras clave: estadística, fonética, modelos lineales mixtos, criterio de información, valor- p

This presentation, belonging to the monographic session *New methodologies for the study of phonetics*, deals with how the analysis of data in phonetics is moving towards the use of linear mixed models (LMM) in order to overcome some limitations associated to ANOVAs. This paradigm shift is allowing to tackle more ambitious research questions and to enrich the possibilities of analysis. Similarly, the use of information criteria have become common in hypothesis testing, and it has been suggested that together with the p -value, also other values such as the false positive rate should be reported.

Keywords: statistics, phonetics, linear mixed models, information criteria, p -value

1. INTRODUCCIÓN

El uso de la estadística inferencial para estudiar un determinado fenómeno –como complemento a la estadística descriptiva– permite generalizar conclusiones a partir de unos datos tomados de una muestra representativa. Para ello es posible aplicar diferentes tests estadísticos dependiendo de cuál sea el tipo de análisis requerido y de la naturaleza de los datos. Entre los análisis más frecuentes están los que permiten correlacionar diferentes grupos de datos (p.ej. coeficiente de correlación de Pearson) o comparar sus medias aritméticas para establecer si dos grupos de datos son diferentes (p.ej. t -Student). En la mayoría de los análisis resultan clave tanto las medidas de tendencia central, es decir, las que dan cuenta de cómo se distribuyen los datos en torno a un determinado valor, como las medidas de dispersión, que muestran cómo es la variabilidad de su distribución.

La idea básica es que cualquier análisis que permita generalizar conclusiones a partir de una muestra representativa siempre requiere la creación de un modelo, a una escala reducida, del fenómeno estudiado; es decir, un modelo estadístico, sea éste un

modelo sencillo, como la media aritmética y su desviación estándar, o más complejo, como el análisis de la varianza de un conjunto de datos (ANOVA).

A su vez, el uso de estos modelos estadísticos para realizar inferencias estadísticas es fundamental para confirmar o rechazar, con un cierto grado de certeza, cualquiera que sea la hipótesis inicial de la investigación, la *hipótesis nula*. Este grado de certeza, el valor- p , equivale a la probabilidad de que un determinado resultado no sea fruto del azar y es un valor arbitrario y determinado de antemano.

Sin embargo, tanto el uso de ciertos modelos estadísticos para extrapolar los resultados obtenidos, como la búsqueda obstinada del valor- p con el fin de confirmar una cierta hipótesis no siempre son el modo más acertado de aproximarse a los datos. En lo que sigue se abordarán las razones que han dado lugar a un cambio de paradigma en numerosas ciencias, entre ellas la lingüística.

2. LIMITACIONES DE ANOVA

En lingüística, diversos ámbitos –entre ellos la fonética experimental– han presenciado la llegada de

un modo alternativo de analizar estadísticamente los datos (Quené y van den Bergh, 2008; Quené, 2013; Torreira et al., 2014; Gahl, 2015; Adamou et al., 2017; Kirby y Sonderegger, 2018; Speelman et al., 2018). Los modelos estadísticos empleados tradicionalmente, entre ellos a menudo diferentes tipos de análisis de la varianza (ANOVA), han ido dando paso al uso de modelos lineales mixtos (MLM), especialmente en casos de medidas repetidas, datos de tipo categórico y diseños incompletos (Jaeger, 2008).

Uno de los problemas con los que suelen encontrarse los investigadores al recabar datos es la necesidad de registrar más de una respuesta por cada participante (o por palabra, si es esta la unidad de observación). Por ejemplo, en un hipotético estudio sobre la realización de la prominencia acústica en función de las diferencias que puede haber en el timbre vocálico, se registra la duración de la vocal tónica de un conjunto de pseudopalabras (Tabla 1).

Tabla 1. Tabla ficticia de datos para un experimento sobre la realización de la prominencia acústica con medidas repetidas.

Pseudopalabra	Vocal	Duración (seg)
1	[a]	0.045
1	[o]	0.034
1	[i]	0.029
2	[a]	0.035
2	[o]	0.031
2	[i]	0.028
3	[a]	0.037
3	[o]	0.041
3	[i]	0.034

Siguiendo la lógica del análisis de la varianza (ANOVA), que permite comparar las medias de más de dos grupos de datos, las observaciones repetidas registradas para cada pseudopalabra violan el principio de independencia de las observaciones – más exactamente, de los errores de las observaciones–, el cual constituye un pilar sobre el que se asientan los modelos de regresión lineal y ANOVA. Este principio puede, sin embargo, ser respetado si se emplea un análisis de la varianza para medidas repetidas (RM-ANOVA) (p.ej. Field et al., 2012, cap. 13).

No obstante, los modelos RM-ANOVA tienen serias limitaciones para analizar datos donde faltan

observaciones (p.ej. Raaijmakers et al., 1999, Quené y van den Bergh, 2004). En estos casos, los datos ausentes pueden reemplazarse por la media del participante, lo cual distorsiona las estimaciones; o pueden descartarse junto a todas las demás observaciones registradas para ese participante. De modo similar, se desaconseja utilizar RM-ANOVA en diseños experimentales incompletos, como por ejemplo los diseños entregrupos, donde los participantes no reciben todos los estímulos, o en casos donde de haya una aleatorización de estímulos mediante la técnica de cuadrado latino para contrarrestar efectos de aprendizaje.

Supongamos, sin embargo, que quisiera aplicarse el conjunto de estímulos del experimento anterior a un grupo de participantes en un estudio sobre la percepción de la prominencia en función del tipo de vocal y su duración. En este caso la variable dependiente sería categórica y no continua (Tabla 2), y como se ha señalado en varias ocasiones, la transformación arco-seno necesaria para analizar datos categóricos en ANOVA a menudo incurre en errores de estimación (p.ej. Jaeger, 2008; Warton y Hui, 2011).

Tabla 2. Tabla ficticia para un experimento sobre percepción de la prominencia acústica. La variable dependiente *prominencia* es categórica.

Partic.	Pseudopal.	Vocal	Dur.	Prom.
1	1	[a]	0.045	SI
1	1	[o]	0.034	NO
1	1	[i]	0.029	SI
1	2	[a]	0.035	NO
1	2	[o]	0.031	SI
1	2	[i]	0.028	SI

En este caso, además, el diseño multifactorial del experimento presenta otro inconveniente. Con los participantes como unidad de observación, los estímulos muestran claramente un jerarquía en la que cada vocal cuenta con tres niveles, uno para cada palabra. Es decir, suponiendo que hubiese 12 participantes a los que se les administrasen 3 pseudopalabras para cada vocal, se registrarían 9 respuestas por participante, dando un total de 108 observaciones (12x3x3). Tradicionalmente se ha buscado la forma de modelar la correlación que existe en las observaciones para los niveles de un mismo factor –puesto que estos tienden a obtener resultados más similares entre sí–, lo cual incrementa los errores de falso positivo, o errores de tipo I (Coleman, 1964; Clark, 1973). La solución más común a este problema

ha sido analizar de forma separada la varianza que los estímulos aportan al modelo para que ésta no afecte a las estimaciones; y lo mismo se ha hecho para la correlación en las observaciones registradas para los participantes. Por ello, tanto estímulos como participantes a menudo han sido tratados en el modelo estadístico como efectos aleatorios, y por tanto diferenciados de los efectos fijos, es decir, los efectos reales que se pretenden analizar, causados por la variable independiente.

3. MODELOS LINEALES MIXTOS (MLM) Y GENERALIZADOS (MLMG)

Frente a lo expuesto anteriormente, los modelos lineales mixtos (MLM) –en ocasiones también llamados modelos de regresión jerárquica o de regresión multinivel– son simplemente modelos de regresión lineal que no solo solventan de manera sencilla la mayoría de las limitaciones enumeradas, sino que además permiten declarar fácilmente los efectos aleatorios de estímulos y participantes.

Por ejemplo, en un estudio real, similar al ejemplo mencionado, se investigó la influencia de la gesticulación de los hablantes y el peso de los correlatos acústicos en la percepción de la prominencia (Jiménez-Bravo, 2020). Los participantes de un experimento tenían que determinar qué palabras de una serie de frases les resultaban prominentes, y el diseño multifactorial incompleto incluía, como variables independientes: dos modalidades perceptivas (auditiva y audiovisual); los valores acústicos de la vocal/sílaba tónica (frecuencia fundamental, intensidad y duración) homogeneizados en tres condiciones experimentales; y las diferentes fases en las que se dividían los gestos manuales realizados por el hablante al emitir la frase (Fig. 1).

Con este diseño experimental, habría sido imposible abordar el análisis estadístico de los datos mediante cualquiera de los análisis ANOVA. Por un lado, las respuestas repetidas para cada participante violan el principio de independencia de las observaciones e introducen una correlación para cada participante. Además, puesto que estos no podían juzgar la misma frase en ambas modalidades, los participantes no recibieron todos los estímulos, sino que un grupo recibió la mitad de los estímulos en la modalidad auditiva, mientras que el otro grupo los recibió en la modalidad audiovisual. Además, los estímulos presentan una estructura jerárquica de palabras anidadas dentro de frases, las cuales habían sido a su vez emitidas por diferentes hablantes.

Sí es posible, sin embargo, analizar los datos utilizando MLMs con el software estadístico R

(2019), concretamente mediante el paquete *lme4* (Bates et al., 2015), el cual incluye la función *glmer* para modelar respuestas categóricas de tipo binario como la del experimento mencionado. Esta función se basa en una variante de los MLMs llamada modelos lineales mixtos generalizados (MLMG), capaz de analizar distribuciones no normales utilizando un modelo de regresión logística. Por otro lado, y de acuerdo con la lógica de los MLMs, la formulación del modelo estadístico en R permite incluir la variabilidad que los estímulos y los participantes introducen en el modelo con el fin de que ésta no afecte a la estimación de los efectos fijos (Fig. 1).

```
prominencia ~ modalidad + condición + f0 + intensidad + duración + fase-gesto +
+ (1 | participante) + (1 | hablante/frase/palabra)
```

Figura 1. Ejemplo simplificado de la formulación del modelo estadístico en R. La notación (1 | ...) estima la variabilidad de participantes y estímulos como efectos aleatorios.

4. INFERENCIA: CRITERIO DE INFORMACIÓN Y VALOR-P

Como en otros modelos de regresión múltiple, es posible incluir más de una variable independiente en los MLMs y por ello es necesario averiguar cuáles de ellas tienen un efecto real en el fenómeno estudiado; es decir, ha de ajustarse el modelo para que éste describa la realidad de la mejor manera posible.

En el caso de los MLMs no existen fórmulas perfectas para realizar esta estimación, por lo que a menudo se realiza una comparación entre modelos basada en la ‘razón de verosimilitud’ (*likelihood ratio*). En esta comparación, el modelo difiere solamente en un único efecto –fijo o aleatorio– respecto al modelo con el que se compara –el modelo *nulo*– y el valor-*p* resultante de esta comparación determina la contribución que tiene el efecto en cuestión sobre la variable dependiente.

Sin embargo, varias críticas han señalado debilidades asociadas a este procedimiento (p.ej. Bolker et al., 2009) y han censurado la búsqueda obstinada del valor-*p*, lo cual se ha relacionado con la dificultad de reproducir resultados que otros investigadores previamente calificaron como ‘estadísticamente significativos’ (Cumming, 2014; Halsey et al., 2015; Colquhoun, 2017).

Como alternativa se ha propuesto el uso de procedimientos basados en criterios de información, los cuales no se limitan a comparaciones únicamente de dos modelos que difieren en un solo efecto, sino que permiten comparar los efectos de varios de ellos

(Burnham y Anderson, 2002). Por ejemplo, el criterio de información de Akaike (AIC) (1973), basado en el concepto de ‘distancia’ entre dos modelos (Kullback y Leibler, 1951), permite comparar varios modelos a partir del número de efectos incluidos en cada uno. El valor AIC, por tanto, es solamente informativo: clasifica los modelos para seleccionar el que mejor describe la realidad a partir de los datos.

Las recomendaciones en el uso de este procedimiento aconsejan incluir, partiendo de un modelo inicial, sólo los efectos que estén justificados por las preguntas de investigación, para ir desestimando progresivamente aquellos que no resulten significativos. También es importante no declarar un conjunto excesivo de efectos en los modelos construidos a partir de este modelo inicial y se recomienda, además, que aquellos modelos a menos de 2 puntos AIC del mejor modelo sean también incluidos en los resultados (Arnold, 2010; Grueber et al., 2011). El paquete *AICcmodavg* (Mazerolle, 2017) permite comparar con facilidad en R los modelos a partir de su valor AIC.

Finalmente, como se ha señalado, muchas veces resulta insuficiente dar cuenta únicamente del valor-*p* al comunicar los resultados. Dada la probabilidad de incurrir en un error de tipo I y considerar como significativo un efecto que en realidad no lo es, se ha propuesto, entre otras cosas, incluir también la tasa de falsos positivos, el tamaño del efecto y los intervalos de confianza (Colquhoun, 2017; Cumming, 2014). En este sentido, por ejemplo, la tasa de falsos positivos de un modelo puede calcularse fácilmente en R mediante el paquete *caret* (Kuhn, 2008), el cual permite crear una tabla de confusión que cruza las predicciones hechas por el modelo con los datos reales, para así valorar su precisión –o poder predictivo– (Kuhn y Johnson, 2013) (Tabla 3). En este sentido, como ironiza Andy Field (2012, p. 339): “Running a regression without checking how well the model fits the data is like buying a new pair of trousers without trying them on – they might look fine on the hanger but get them home and you find you’re Johnny-tight-pants.”

Por tanto, la tasa de falsos positivos, como uno de los elementos asociados a la validación del modelo estadístico se convierte en un valor que permite contextualizar el valor-*p* necesario en la inferencia. No obstante, este proceso de validación del modelo, en el que se cruzan las predicciones con los datos reales, puede realizarse bien utilizando el conjunto completo de datos bien dividiéndolos en dos subconjuntos –a menudo 70/30–, en lo que se conoce como validación cruzada (*cross-validation*). Esta división de los datos permite ‘entrenar’ previamente el modelo con el conjunto menor de los datos y

realizar las predicciones sobre un conjunto de datos diferente; en este caso, a falta de un conjunto de datos independiente a la investigación, la predicción se hace sobre la partición mayor, obtenida a partir de los datos originales.

Tabla 3. Matriz de confusión para un modelo sobre la percepción multimodal de la prominencia (Jiménez-Bravo, 2020). La precisión corresponde a la suma de las predicciones correctas dividida por el gran total. El valor 117 muestra los falsos positivos, equivalente a una tasa del 19%.

		Datos		
		No prom	Prom	Total
Predicción	No prom	500	250	750
	Prom	117	367	484
	Total	617	617	1234
Precisión		70%		

5. CONCLUSIONES

Este cambio de paradigma en el análisis estadístico surge de la limitación que los diferentes análisis de la varianza (ANOVA) presentan al analizar los datos. Ello hace posible abordar preguntas de investigación más ambiciosas y realizar análisis más complejos. En el ámbito de la lingüística, numerosos estudios fonéticos están haciendo uso de modelos lineales mixtos (MLM) o de su variante para distribuciones no normales, MLMs generalizados (MLMG). De forma similar, el criterio de información de Akaike (AIC) se hace enormemente útil al permitir comparar modelos con múltiples variables independientes con el fin de seleccionar aquel que mejor dé cuenta de los datos.

Por otro lado, el proceso de inferencia está estrechamente asociado al valor-*p*. No obstante, es necesario complementarlo con otros valores que ayuden a replicar los resultados. Entre ellos, destacan el tamaño del efecto y sus intervalos de confianza, así como la tasa de falsos positivos, que no es más que uno de los elementos propios de la validación del modelo estadístico. Este nuevo paradigma ha ido de la mano del desarrollo de R, cuya utilización está haciendo que la disciplina se enriquezca. Todo ello, sin embargo, plantea la necesidad de establecer criterios de buenas prácticas que permitan llevar a cabo tanto un correcto análisis de los datos como una adecuada comunicación de los resultados.

6. BIBLIOGRAFÍA

Adamou, E., Gordon, M., Gries, S.T. (2018). Prosodic and morphological focus marking in Ixcatec (Otomanguean).

- In E. Adamou, K. Huade, & M. Vanhove (Eds.) *Information Structure in Lesser-described Languages*, (pp. 51–84). Amsterdam: John Benjamins.
- Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. In B. N. Petrov, & F. Csaki (Eds.) *Proceedings of the 2nd International Symposium on Information Theory*, (pp. 267–281). Budapest: Akadémiai Kiadó.
- Arnold, T.W. (2010). Uninformative parameters and model selection using Akaike's Information Criterion. *Journal of Wildlife Management*, 74(6), 1175–1178.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using *lme4*. *Journal of Statistical Software*, 67(1), 1–48.
- Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S. S. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology and Evolution*, 24(3), 127–135.
- Burnham, K.P., & Anderson, D.R. (2002). *Model Selection and Multimodel Inference: a Practice Information-Theoretic Approach*. Berlin/Heidelberg/New York: Springer-Verlag.
- Clark, H.H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 12, 335–359.
- Coleman, E.B. (1964). Generalizing to a language population. *Psychological Reports*, 14, 219–226.
- Colquhoun D. (2017). The reproducibility of research and the misinterpretation of *p*-values. *Royal Society Open Science*, 4.
- Cumming, G. (2014). The New Statistics: Why and How. *Psychological Science*, 25(1), 7–29.
- Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics using R*. London: SAGE.
- Gahl, S. (2015). Lexical competition in vowel articulation revisited: Vowel dispersion in the Easy/Hard database. *Journal of Phonetics*, 49, 96–116.
- Grueber, C.E., Nakagawa, S., Laws, R.J., & Jamieson, I.G. (2011). Multimodel inference in ecology and evolution: Challenges and solutions. *Journal of Evolutionary Biology*, 24(4), 699–711.
- Halsey, L.G., Curran-Everett, D., Vowler S.L., & Drummond G.B. (2015). The fickle P value generates irreproducible results. *Nature Methods*, 12(3), 179–85.
- Jaeger, T.F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Archives of General Psychiatry*, 59(4), 434–446.
- Jiménez-Bravo, M. (2019). *Multimodal perception of acoustic prominence in Spanish*. Tesis doctoral. UNED.
- Jiménez-Bravo, M. (2020). Multimodal Perception of Prominence in Spontaneous Speech: A Methodological Proposal using Mixed Models and AIC. *Speech Communication*, 124, 28–45.
- Kirby, J., & Sonderegger, M. (2018). Mixed-effects design analysis for experimental phonetics. *Journal of Phonetics*, 70, 70–85.
- Kuhn, M., 2008. Building Predictive Models in R using the caret Package. *Journal of Statistical Software*, 28(5), 1–26.
- Kuhn M., & Johnson K. (2013). Measuring Performance in Classification Models. In: *Applied Predictive Modeling*, (pp. 247–273). New York: Springer.
- Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, 22(1), 79–86.
- Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, 59(4), 413–425.
- Quené, H. (2013). Longitudinal trends in speech tempo. *Journal of the Acoustical Society of America*, 133 (6), EL453–EL457.
- R Development Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Computer program. <https://www.R-project.org/>.
- Raaijmakers, J.G., Schrijnemakers, J.M., & Gremmen, F. (1999). How to deal with “the language-as-fixed-effect fallacy”: Common misconceptions and alternative solutions. *Journal of Memory and Language*, 41, 416–426.
- Speelman, D., Heylen, K., & Geeraerts, D. (Eds.) (2018). *Mixed-Effects Regression Models in Linguistics*. Amsterdam: Springer.
- Torreira, F., Simonet, M., & Hualde, J.I. (2014). Quasi-neutralization of stress contrasts in Spanish. In *Proceedings of the 7th International Conference on Speech Prosody (SP2014)*, (pp. 197–201). Dublin, Ireland.
- Warton, D.I., & Hui, F.K.C. (2011). The arcsine is asinine: the analysis of proportions in ecology. *Ecology*, 92(1), 3–10.

RELEVANCIA DE LA DURACIÓN EN EL ESPAÑOL DE CANARIAS: ETIQUETAJE Y DIALECTOMETRÍA A PARTIR DE DATOS ACÚSTICOS

Carolina Jorge Trujillo, Josefa Dorta Luis, Chaxiraxi Díaz Cabrera

Universidad de La Laguna
cJORGETR@ULL.EDU.ES, JDORTA@ULL.EDU.ES, CHADIAZ@ULL.EDU.ES

ABSTRACT

El propósito del presente trabajo es estudiar las pautas de duración en el español hablado en Canarias. Abordamos una doble perspectiva para comprobar si las características analizadas acústicamente se corroboran desde el punto de vista del análisis dialectométrico. Nuestro corpus se compone de 756 oraciones declarativas e interrogativas con estructura SVO, emitidas por catorce informantes canarios: siete mujeres y siete hombres de las zonas capitalinas de las islas, sin estudios superiores.

Palabras clave: prosodia, duración, acento, dialectometría, variedades del español

The purpose of this paper is to study the duration patterns in the Spanish spoken in the Canary Islands. We take a double perspective to verify if the acoustically analyzed characteristics are corroborated from the point of view of the dialectometric analysis. Our corpus consists of 756 declarative and interrogative sentences with SVO structure, issued by fourteen Canarian informants: seven women and seven men from the islands' capital areas, without higher studies.

Keywords: prosody, duration, accent, dialectometry, Spanish varieties

1. INTRODUCCIÓN

La frecuencia fundamental (F0) es el parámetro más estudiado dentro de los estudios dedicados a la prosodia del español. A diferencia de esta, el análisis de la duración y de la intensidad se ha realizado de manera más esporádica, si bien la bibliografía precedente contiene trabajos de gran interés al respecto. Esta investigación se plantea como contribución al análisis de la duración en la variedad del español hablada en Canarias. Para ello, realizamos un análisis acústico de las vocales pretónicas, tónicas y postónicas con el objetivo de etiquetarlas según su mayor o menor duración (Muñetón Ayala, Díaz y Dorta, 2018) y, asimismo, establecer la proximidad entre los diferentes puntos a la luz del análisis dialectométrico con ProDis v2 (Elvira García *et al.* 2018).

2. ANÁLISIS ACÚSTICO DE LA DURACIÓN

2.1. Etiquetaje y umbrales psicoacústicos

En el marco del proyecto internacional AMPER, han sido varios los trabajos que han acometido el etiquetaje de las oscilaciones de duración entre las vocales pretónicas, tónicas y postónicas como propuesta de etiquetaje similar al de la F0 en el modelo Métrico Autosegmental (AM) (Pierrehumbert, 1987). De este modo, si para la F0 se consideran dos tonos básicos, L* (bajo) y H* (alto), para la duración se han empleado dos categorías temporales: G* (grande) y P* (pequeña) (Fernández Planas y Martínez Celdrán, 2003) o L* (larga) y B* (breve) (Muñetón Ayala *et al.*, 2018).

2.2. Datos precedentes

Fernández Planas y Martínez Celdrán (2003) analizan el castellano septentrional de España y apuntan que, en el 50% de los casos, las diferencias temporales no alcanzan el umbral de percepción, si bien es cierto

que la duración marca el tonema de forma notable. En trabajos acerca de otras variedades se ha llegado a conclusiones paralelas. En Dorta *et al.* (2009, 2011), Dorta y Mora (2011), Dorta (ed., 2013), Dorta y Jorge Trujillo (2015), Muñetón *et al.* (2018) y Dorta (2019) se aprecia, en líneas generales, que si bien las diferencias temporales no son muy numerosas, las vocales tónicas suelen destacar por su duración y, además, las vocales finales de oración (y de sintagma en algunos casos) registran valores muy relevantes.

3. ANÁLISIS DIALECTOMÉTRICO

3.1. La dialectometría en los estudios prosódicos

La introducción de la dialectometría ha constituido un importante avance, en tanto que permite procesar y agrupar o distanciar matemáticamente grandes cantidades de datos. En consecuencia, dentro de AMPER se ha tratado de adaptar las técnicas dialectométricas al estudio específico de la prosodia. Podemos mencionar, para comenzar, a Rilliard y Lai (2008), quienes han trabajado en el programa *Stat-Distances*. Seguidamente, contamos con los trabajos realizados por el equipo del Laboratorio de Fonética de la Universidad Autónoma de Barcelona, que comenzaron sus investigaciones con *Calcu-Dista* y con *DiaTech*, este último desarrollado en un principio por el equipo de la Universidad del País Vasco (Aurrekoetxea *et al.*, 2013; Fernández-Planas *et al.*, 2011, 2015). El grupo de AMPER-Galicia, por su parte, ha diseñado una herramienta que permite procesar los datos de manera estadística a través del programa R (Martínez Calvo and Rei, 2015). Romano *et al.* (2011) y De Castro Moutinho *et al.* (2011) también han empleado métodos de correlación en el estudio del italiano y del portugués, respectivamente. Entre los avances más recientes, debemos mencionar nuevamente al equipo de Barcelona, que ha diseñado un nuevo programa de análisis, denominado *ProDis* (Elvira-García *et al.*, 2015, 2016; Fernández-Planas 2016a, 2016b; Fernández-Planas *et al.*, 2017, Fernández-Planas *et al.*, 2019) que cuenta ya con dos versiones.

3.2. Datos precedentes

Como detallaremos en el apartado de “Metodología”, en el presente trabajo nos proponemos emplear *ProDis* v2 en el estudio de la duración en Canarias. Fernández-Planas *et al.* (2015) analizan con *Calcu-Dista* y *DiaTech* la F0 de catorce puntos de encuesta españoles e hispanoamericanos. En Dorta (ed., 2018) se aplica *ProDis* v1 a la F0 de veinte puntos de encuesta procedentes de cinco zonas del español, mientras que en Dorta y González Rodríguez (2019) se estudia la F0 de San Antonio de Texas. En general,

el procesamiento dialectométrico de los datos ha permitido corroborar la mayoría de los resultados obtenidos con el análisis acústico.

4. OBJETIVOS

El objetivo de esta investigación es realizar una aproximación al estudio de la duración en las Islas Canarias desde una doble perspectiva (acústica y dialectométrica) para contribuir al estudio de este parámetro en el ámbito del español.

5. METODOLOGÍA

5.1. Informantes y corpus de análisis

Se han analizado las oraciones emitidas por siete mujeres y siete hombres de procedencia urbana (zonas capitalinas de las Islas Canarias), sin estudios superiores y entre los 25 y 55 años, naturales y residentes en sus respectivos puntos de encuesta. El corpus de análisis está integrado por 756 oraciones SVO procedentes del corpus fijo o formal de AMPER (con igual número de declarativas e interrogativas) y que combinan inicios y finales oxítonos, paroxítonos y proparoxítonos.

5.2. Grabación del corpus

Las grabaciones se realizaron en lugares familiares para los informantes. El equipo utilizado fue un magnetofón Marantz PM222 Digital Compact Flash Portable Recorder y un micrófono Shure SM48 8700. Cada oración se repitió tres veces, lo que favorece una mayor fiabilidad de los datos.

5.3. Análisis de los datos

5.3.1. Análisis acústico

Una vez que se procedió a la digitalización y optimización (en su caso) de los ficheros con el programa *GoldWave Digital Audio Editor*, se etiquetó y analizó acústicamente cada frase según la metodología del proyecto con rutinas de *Matlab* creadas para AMPER, el programa *Amper-2006* (Romano, 1995; Brezmes, 2007; López Bobo *et al.*, 2007). Nos hemos centrado en las vocales pretónicas, tónicas y postónicas y, además, en las pospostónica de los proparoxítonos. Se aplicó el umbral diferencial de 33,33% establecido por Fernández Planas y Martínez Celdrán (2003) con el que pudieron hallarse las diferencias mínimas perceptibles (DMP) entre los datos obtenidos. Estas DMP servirán de base para realizar el etiquetaje de la duración según la propuesta detallada en Muñetón *et al.* (2018).

5.3.2. Análisis dialectométrico

Se ha realizado un análisis dialectométrico con ProDis v2 (Elvira-García *et al.*, 2018). El programa trabaja con los valores numéricos obtenidos con *Amper-2006*. ProDis utiliza la correlación de Pearson para calcular la similitud entre los contornos comparados solo a partir de los valores de F0 o ponderados por la intensidad, por la duración o por ambas. Esto permite expresar numéricamente qué grado de correlación existe entre estos contornos, y realizar a partir de aquí un análisis de clúster que los clasifica en grupos según su similitud.

6. RESULTADOS

6.1. Análisis acústico

Presentamos a modo de ilustración las medias de duración correspondientes a las vocales pretónicas, tónicas y postónicas de las dos modalidades analizadas (Figs. 1 y 2).

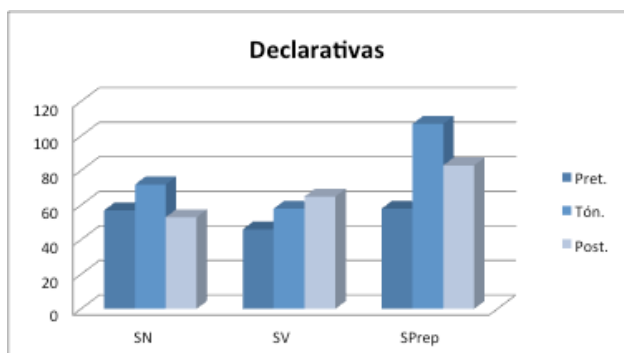


Figura 1: Medias de duración en los tres sintagmas de las declarativas canarias.

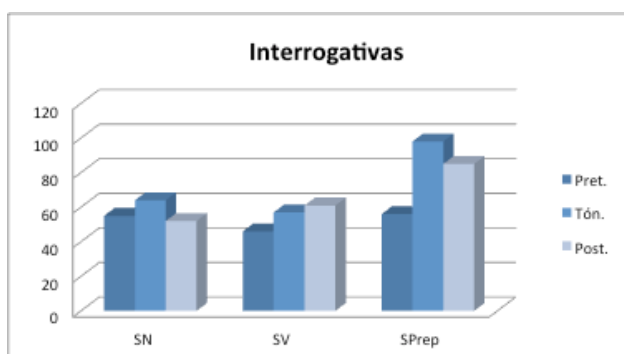


Figura 2: Medias de duración en los tres sintagmas de las interrogativas canarias.

Las figuras precedentes muestran que, tanto en declarativas como en interrogativas, la vocal acentuada sobresale respecto de las colindantes en el sintagma nominal (SN) y en el preposicional (SPrep). En este último se advierte la influencia de la posición final de oración, en tanto que la tónica se distancia de la pretónica de forma más

destacada que en el SN y, además, la postónica ofrece valores muy relevantes. En el sintagma verbal (SV), por el contrario, se produce un progresivo incremento de duración.

6.2. Análisis dialectométrico

De manera similar a lo expuesto en el análisis acústico, incluimos ahora los resultados para una y otra modalidad desde el punto de vista dialectométrico (Figs. 3 y 4).

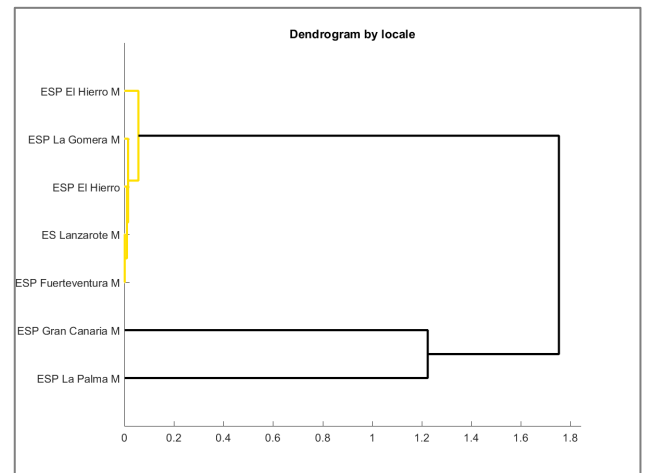


Figura 3: Dendrograma de los puntos de encuesta a partir de los datos de duración de las declarativas.

El dendrograma de la Fig. 3 señala tres grupos prosódicos claramente diferenciados (en una distancia no superior a 1.2): los formados por La Palma, Gran Canaria y el del resto de las islas.

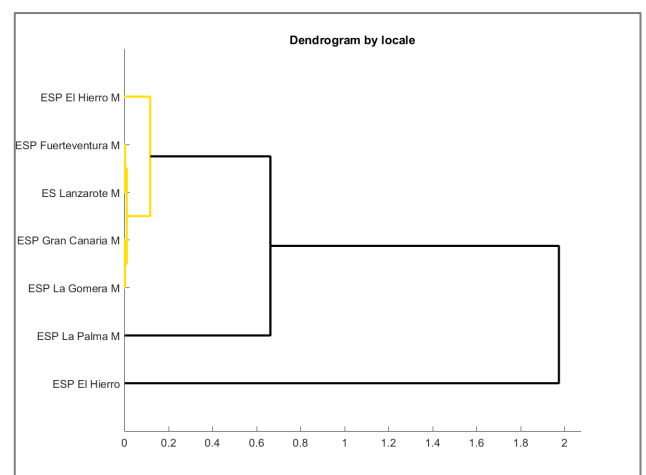


Figura 4: Dendrograma de los puntos de encuesta a partir de los datos de duración de las interrogativas.

El dendrograma de la Fig. 4 distancia El Hierro del resto de las islas y, a su vez, entre estas destaca la mayor distancia de La Palma frente al resto, por lo que podemos considerar el agrupamiento en tres clúster.

7. CONCLUSIONES

El análisis acústico ha permitido comprobar, en general, cómo las vocales acentuadas destacan enormemente por su duración y, además, las situadas en la frontera final ofrecen valores de gran interés. Por su parte, el análisis dialectométrico ha relacionado la mayoría de islas entre sí y apunta cierta distancia con respecto a otras, de modo que se realizará un etiquetaje de la duración para profundizar en estos resultados.

8. BIBLIOGRAFÍA

- Aurrekoetxea, G., Fernández-Aguirre, K., Rubio, J., Ruiz, B., y Sánchez, J. (2013). DiaTech: A new tool for dialectology. *Literary and Linguistic Computing*, 28, 23-30.
- Brezmes-Alonso, D. (2007). *Desarrollo de una aplicación software para el análisis de características fundamentales de la voz*. Proyecto de fin de carrera. Oviedo: Universidad de Oviedo.
- De Castro Moutinho, L., Coimbra, R. L., Rilliard, A., y Romano, A. (2011). Measure of the variation prosodique diatopique en portugais européen. *Estudios de Fonética Experimental*, 20, 33-55.
- Dorta, J. (2019). Estructuras tonales y de duración en la entonación del español de hablantes bilingües americanos con ascendencia mexicana. *Tonal and duration structures in the intonation of Spanish language by bilingual Americans of Mexican ancestry ONOMÁZEIN*, 45, 232-258.
- Dorta, J., y González Rodríguez, M^a J. (2019). Tonal proximity relationship in the Spanish of the Canary Islands in the light of dialectometry. *Languages, Special Issue Contemporary Perspectives in Geolinguistics and Dialectology* 4(2), 29.
- Dorta, J., y Jorge Trujillo, C. (2015). Estudios prosódicos en Canarias: análisis de la duración en habla formal de El Hierro y Fuerteventura. *Estudios de Fonética Experimental*, 24, 11-33.
- Dorta, J., y Mora, E. (2011). Patrones temporales en dos variedades del español hablado en Venezuela y Canarias. *Revista Internacional de Lingüística Iberoamericana*, 9, nº 1 (17), 91-100.
- Dorta, J., Hernández, B., y Díaz, Ch. (2009). Interrogativas absolutas: relación entre F0, duración e intensidad". *Estudios de Fonética Experimental*, 18, 123-144.
- Dorta, J., Hernández, B., y Díaz, Ch. (2011). Duración e intensidad en la entonación de las declarativas e interrogativas de Canarias. En F. Hernández González, M. Martínez Hernández y L. M. Pino Campos (eds.): *Sosalivm Mvnera. Homenaje a Francisco González Luis* (pp. 143-154). Madrid: Ediciones Clásica.
- Dorta, J. (ed.) (2013). *Estudio comparativo preliminar de la entonación de Canarias, Cuba y Venezuela*. Madrid-Santa Cruz de Tenerife: La Página ediciones S/L, Colección Universidad.
- Dorta, J. (ed.) (2018). *La entonación declarativa e interrogativa en cinco zonas fronterizas del español*. Studien zur romanischen Sprachwissenschaft und interkulturellen Kommunikation, herausgegeben von Gerd Wotjak, Peter Lang Edition.
- Elvira García, W., Balocco, S., Roseano, P., y Fernández-Planas, A. M. (2018). ProDis: A dialectometric tool for acoustic prosodic data. *Speech Communication*, 97, 9-18.
- Elvira-García, W., Balocco, S., Fernández-Planas, A. M^a, Roseano, P., y Martínez Celdrán, E. (2015). Presentació d'una aplicació informàtica per a l'anàlisi dialectomètrica de dades prosòdiques en el marc de l'Atlas Multimèdia de la Prosòdia de l'Espai Romànic. Ponencia presentada en *Workshop sobre la prosòdia del català*. Universitat de Barcelona, Barcelona, España, 22 de junio.
- Fernández-Planas, A. M^a. (2016a). Aspectos de ProDis, una nueva herramienta para el análisis dialectométrico prosódico. Ponencia presentada en *Workshop Approaches to Sociolinguistics Aspect of Romanian and Spanish Intonation*. Alexandru Ioan Cuza University of Iasi, Iasi, Rumanía, 21 de octubre.
- Fernández-Planas, A. M^a. (2016b). Características generales de ProDis (herramienta para analizar distancias prosódicas). Ponencia presentada en *Servei de Rtactament de la Parla i el So (STPS)*. UAB, Birmingham, AL, USA, 18 de noviembre.
- Fernández-Planas, A. M^a, y Martínez Celdrán, E. (2003). El tono fundamental y la duración: dos aspectos de la taxonomía prosódica en dos modalidades de habla (enunciativa e interrogativa) del español. *Estudios de Fonética Experimental*, 12, 166-200.
- Fernández-Planas, A. M^a, Elvira-García, W., Roseano, P., y Balocco, S. (2019). Análisis dialectométrico con ProDis: un paso más en los estudios prosódicos de AMPER. En J. Dorta (ed.), *Investigación geoprosódica. AMPER: análisis y retos* (pp. 119-134). Frankfurt: Iberoamericana Vervuert.
- Fernández-Planas, A. M^a, Roseano P., Elvira-García, W., y Balocco, S. (2017). Génesis y aspectos fundamentales de ProDis. Ponencia presentada en *Congreso Subsidia: Herramientas y recursos para las ciencias del habla*. Universidad de Málaga, Málaga, España, 21-23 de junio.
- Fernández-Planas, A. M^a, Roseano P., Martínez Celdrán, E., y Romera Barrios, L. (2011). Aproximación al análisis dialectométrico de la entonación en algunos puntos del dominio lingüístico catalán. *Estudios de Fonética Experimental*, 20, 141-78.
- Fernández-Planas, A. M^a, Dorta, J., Roseano, P., Díaz, Ch., Elvira-García, W., Martín Gómez, J. A., y Martínez Celdrán, E. (2015). Distancia y proximidad prosódica entre algunas variedades del español: un estudio dialectométrico a partir de datos acústicos. *Revista de Lingüística Teórica y Aplicada*, 53, 13-45.
- López Bobo, M^a J., Muñoz Cachón, C., Díaz Gómez, L., Corral Blanco, N., Brezmes Alonso, D., y Alvarellos Pedrero, M. (2007). Análisis y representación de la entonación. Replanteamiento metodológico en el marco del proyecto AMPER. En J. Dorta (ed.), *La prosodia en el ámbito lingüístico románico* (pp. 17-34). Santa Cruz de Tenerife: La Página.
- Martínez Calvo, A., y Fernández Rei, E. (2015). Unha ferramenta informática para a análise dialectométrica

- da prosodia. *Estudios de Fonética Experimental*, 24, 289-303.
- Muñetón Ayala, M., Díaz, Ch., y Dorta, J. (2018). La duración en oraciones sin expansión en la voz femenina de dos países fronterizos: Colombia (Bogotá-Medellín) y Venezuela (Caracas-Mérida). *Literatura y Lingüística*, 37, 401-423.
- Pierrehumbert, J. (1987) [1980]. *The Phonology and Phonetics of English Intonation*. Bloomington: Indiana University Linguistics Club.
- Rilliard, A., y Lai, J. P. (2008). Outils pour le calcul et la comparaison prosodique dans le cadre du projet AMPER-L'exemple des variétés occitane et sarde. En A. Turculeț (ed.), *La variation diatopique de l'intonation dans le domaine roumain et roman* (pp. 217-229). Iași: Presses de l'Université "Al. I. Cuza" de Iași.
- Romano, A. (1995). *Développement d'un environnement de travail pour l'étude des structures sonores et intonatives de la parole*. Mémoire de DEA en Sciences du Langage. ICP, Univ. Stendhal.

DISTANCIA DEL COSENO ENTRE VECTORES CEPSTRALES COMO MÉTODO PARA CUANTIFICAR LA DIFERENCIA ACÚSTICA ENTRE MARCAS EN CONFLICTO

José María Lahoz-Bengoechea, Marianela Fernández Trinidad

Universidad Complutense de Madrid
jmlahoz@ucm.es, marianela.fernandez@ucm.es

ABSTRACT

Este trabajo propone una metodología innovadora en el ámbito de la lingüística forense para poder estimar la distancia acústica entre nombres de marcas comerciales compuestas por similar número de segmentos. La iniciativa surge de un estudio lingüístico pericial en el que se pretendía esclarecer el grado de diferencia entre dos marcas comerciales en conflicto judicial. La medida de la distancia acústica que se propone es el ángulo formado por los vectores mel-cepstrales (MFCC), comparados para cada instante temporal entre los sonidos de los nombres pronunciados por un mismo locutor. Este método permite calcular de forma objetiva la distancia fonética global entre los nombres de las marcas, la distancia en la parte coincidente y en la discrepante o diferencial. El artículo detalla todos los pasos metodológicos seguidos en el estudio y discute sus ventajas y posibles limitaciones.

Palabras clave: distancia acústica, análisis cepstral, marcas comerciales, lingüística forense

This paper presents a new method, of interest to forensic linguistics, that estimates the acoustic distance between the names of two trademarks that consist of a similar number of segments. The proposal is based on an experts' linguistic report aimed at establishing the degree of differentiation between two trademarks, one of which had sued the other for alleged similarity. The measure of acoustic distance that is suggested here is the angle formed by the mel-cepstral (MFCC) vectors, compared for every time frame between the sounds of both names, as pronounced by the same speaker. This method provides an objective measure of the phonetic distance between the trademarks, averaged over the word, or restricted to either the differential part or that in common. The paper explains all the methodological steps and discusses the contributions and possible limitations.

Keywords: acoustic distance, cepstral analysis, trademarks, forensic linguistics

1. INTRODUCCIÓN

1.1. Nombres de marcas en contienda judicial

El examen comparativo de nombres de marcas comerciales constituye un ámbito de actuación reciente en lingüística judicial (Cicres y Turell, 2014 y Ramírez Salado, 2017).

La Ley 17/2001, de 7 de diciembre, de Marcas regula sobre el concepto de marca y las prohibiciones para su registro. En el art. 6.1 de la mencionada ley se establece que no podrá registrarse una nueva marca si su signo (nombre, dibujos, colores, etc.) es idéntico o semejante al de otro ya existente, máxime si coinciden los productos o servicios que se ofrecen.

La identidad de una marca viene determinada por varios factores y, entre ellos, asume un papel importante su nombre, especialmente si en ella prevalece el carácter denominativo sobre el gráfico. En caso de que una marca vea peligrar sus derechos y recurra por vía legal el registro de la nueva, una de las partes implicadas o ambas pueden solicitar a los especialistas un estudio lingüístico pericial que compare los nombres de las marcas en liza.

1.2. Descripción del presente caso de estudio

Este trabajo surge de un caso real de estudio lingüístico pericial en el que se buscaba establecer las posibles similitudes o diferencias entre los nombres de dos marcas comerciales en conflicto. Por

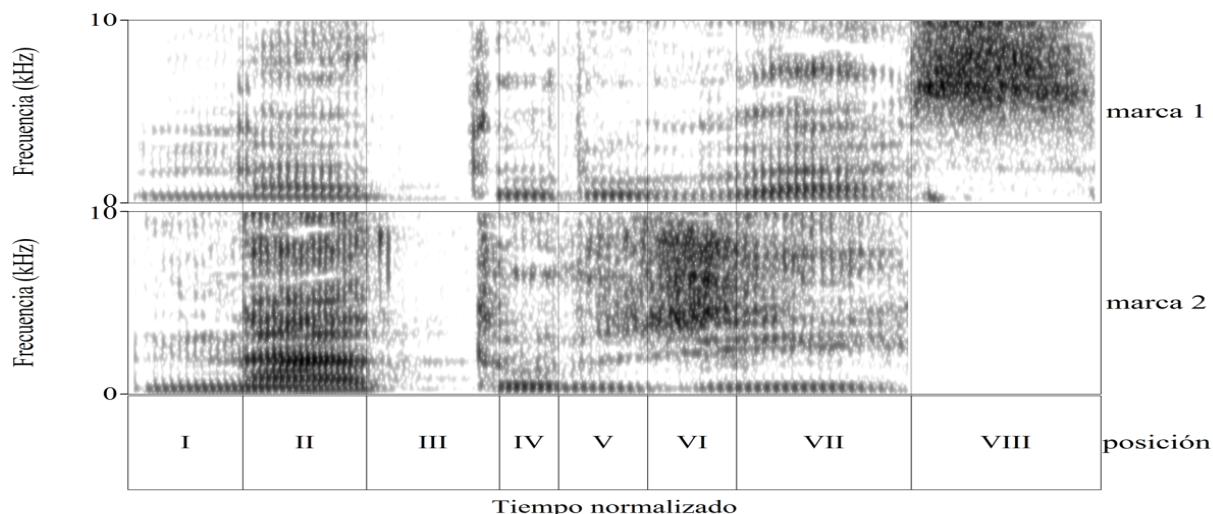


Figura 1. Espectrogramas de ambas marcas alineados temporalmente

motivos de confidencialidad, no se mencionarán aquí los nombres en cuestión.

La marca 1 consta de ocho segmentos y es una palabra aguda, mientras que la marca 2 está formada por siete segmentos y su acentuación es esdrújula. Los primeros cinco fonemas, con una estructura CVCVC, coinciden entre ambos nombres. La parte diferencial presenta en la posición VI un contraste entre una espirante velar sonora [ɣ] y una fricativa postalveolar sonora [ʒ] (parcialmente ensordecida según el modelo de pronunciación de los anuncios de la marca). En la posición VII, la marca 1 presenta una vocal [a] y la marca 2, una [i]. La posición VIII destaca por el contraste entre la presencia de un segmento adicional en la marca 1 frente a la ausencia de cualquier segmento en la marca 2.

Ahora bien, un método basado exclusivamente en contar los segmentos que cambian de una palabra a otra no sería suficiente para establecer una comparación sólida entre los nombres, pues no basta cuantificar los elementos diferentes, sino que es necesario esclarecer cuán diferentes son. Por ello, se decidió partir de la comparación acústica de las pronunciaciones de ambas marcas como método para establecer la distancia fonética real entre ellas.

2. METODOLOGÍA

La medida de la distancia acústica que aquí se propone es el ángulo formado por los vectores de coeficientes cepstrales, comparados para cada instante temporal entre las grabaciones de la pronunciación de cada palabra. A continuación, se detallan los sucesivos pasos de este cálculo.

2.1. Obtención de las muestras sonoras

Se grabó la pronunciación de los nombres de ambas marcas por un único hablante para minimizar las diferencias inter-locutor y poder aislar aquellas derivadas de las características de la palabra.

Para agilizar el posterior análisis cepstral, se redujo la frecuencia de muestreo a 16 kHz.

2.2. Alineamiento temporal de los segmentos

Las marcas que se compararon no constaban del mismo número de segmentos: una tenía siete y la otra, ocho. Esto impedía establecer comparaciones emparejadas segmento a segmento.

Una posibilidad habría sido añadir un fragmento de silencio al final del nombre más corto. Sin embargo, Ahocoder (Erro *et al.*, 2011), el programa utilizado para extraer los coeficientes cepstrales, rellena con ruido aleatorio los silencios que detecta en la señal. Mediante la adición matemática de un silencio no se consigue realmente el objetivo pretendido de calcular la distancia acústica entre un sonido y su ausencia. Por lo tanto, fue necesario excluir del análisis el último segmento del nombre más largo.

A su vez, la duración de los segmentos difería en una palabra y otra, incluso en aquellos segmentos que corresponden a un mismo fonema. Para permitir una comparación alineada entre ambas palabras, se modificaron con Praat (Boersma y Weenink, 2017) las duraciones de cada pareja de segmentos para equipararlas a la duración promediada entre ambos. La Figura 1 muestra los espectrogramas de los dos nombres de marcas ya alineados temporalmente. En ellos se puede observar también el contraste entre

la presencia o ausencia de segmento en la posición VIII, al que se ha aludido hace un momento (recuérdese que dicho segmento no se tuvo en cuenta para el cálculo de la distancia acústica).

2.3. Extracción de los coeficientes cepstrales

Se utilizó el programa Ahocoder (Erro *et al.*, 2011) para extraer, cada 0.005 s (frecuencia de muestreo de 200 Hz), un vector de 40 coeficientes MFCC.

2.4. Cálculo de la distancia angular

Se computó el coseno del ángulo formado por cada pareja de vectores (aquellos correspondientes a un mismo instante temporal), utilizando para ello la versión 4.0.0 del programa Octave (Eaton *et al.*, 2015), según la fórmula generalizada para vectores n-dimensionales, tal y como se muestra en (1):

$$(1) \quad \cos = \frac{A \cdot B}{|A||B|} = \frac{A \cdot B}{\sqrt{A_n^2} \sqrt{B_n^2}}$$

A continuación, mediante el arco coseno, se calculó el ángulo definido por dichos vectores expresado en grados sexagesimales. Este ángulo se tomó como medida de la distancia acústica entre los segmentos, con la siguiente interpretación. Un ángulo de 0° solo se obtendría si se comparara una determinada producción sonora consigo misma. Incluso el mismo tipo de sonido pronunciado dos veces por la misma persona presentaría una distancia mayor que 0°. En el extremo opuesto, 180° significaría una correlación perfecta pero de signo opuesto o inversamente proporcional. Por su parte, un ángulo de 90° supondría la falta absoluta de correlación, como sucede por ejemplo cuando se calculan los componentes principales de una serie de variables para eliminar su covarianza y obtener factores ortogonales (Baayen, 2008, p. 127 y ss.). El rango de valores de distancia calculados con este método puede oscilar entre 0° y 90°.

2.5. Obtención de magnitudes de referencia

En la práctica, los valores de la distancia acústica no llegan a ser tan bajos como 0° ni tan altos como 90°. Para tener una idea aproximada de los límites efectivos de la variación con los que poder comparar los resultados obtenidos se decidió realizar dos comparaciones adicionales. Por un lado, se calculó la distancia entre dos emisiones diferentes de cada una

de los nombres de marca pronunciados por el mismo locutor en la misma sesión de grabación. Esta distancia se tomó como orientativa del mínimo de distancia efectivo, teniendo en cuenta la variación intralocutor. Por otro lado, se comparó una producción de [u] con una de [s] y se tomó como referencia de un máximo razonable en la modalidad intralocutor, habida cuenta de que estos dos sonidos no coinciden ni en punto de articulación, ni en modo, ni en sonoridad, y sus resonancias características se hallan en zonas muy separadas del espectro.

3. RESULTADOS

3.1. Referencias de los valores mínimo y máximo

El promedio de la distancia obtenida entre las dos realizaciones de la marca 1 fue de 12° y el mínimo alcanzado en algún punto de dicha comparación fue de 6°. En el caso de la segunda marca, los valores respectivos fueron 11° de promedio y un mínimo de 6°. Como se constata de manera repetida con estos resultados, el mínimo efectivo ronda los 6°.

Por otra parte, del contraste [u] – [s] intralocutor se obtuvo una media de 42° y un máximo de 44°. Esta medida (44°) se tomó como techo razonable de diferencia entre dos sonidos altamente dispares en la condición intralocutor.

Estos valores mínimo y máximo aportaron un marco de referencia para interpretar los resultados de la distancia acústica entre los nombres de las marcas.

3.2. Distancia acústica entre las marcas

La Figura 2 muestra la evolución de la distancia acústica en los distintos instantes emparejados entre los nombres de ambas marcas, medida en términos de distancia angular entre los vectores cepstrales, tal y como se ha explicado.

3.2.1. Análisis de la distancia global entre las palabras

Los valores alcanzados se enmarcan entre el punto mínimo, obtenido durante la segunda vocal de la parte común, de 8° —un dato similar al mínimo de 6° previamente establecido como suelo— y el instante de mayor diferencia, que corresponde a los 33° de distancia entre la [ɣ] (aproximante espirante, velar, sonora) y la [ʒ] (fricativa, postalveolar, sonora parcialmente ensordecida).

Entre estos extremos, el valor promedio de la distancia acústica constatada entre la pronunciación de ambas marcas es de 18°. Es decir, 18° es la

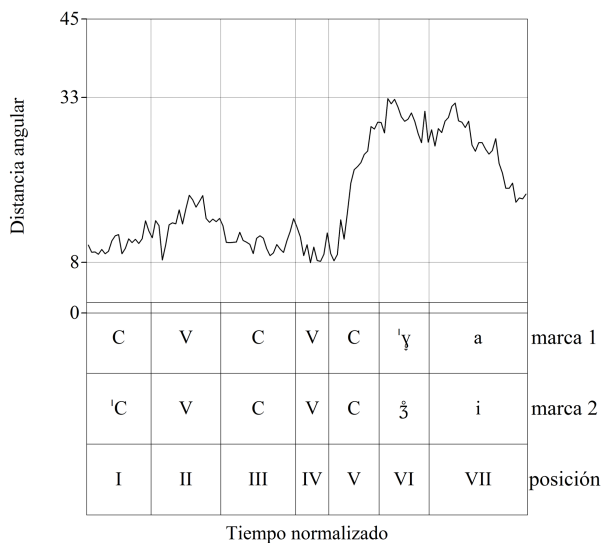


Figura 2. Distancia acústica entre ambas marcas

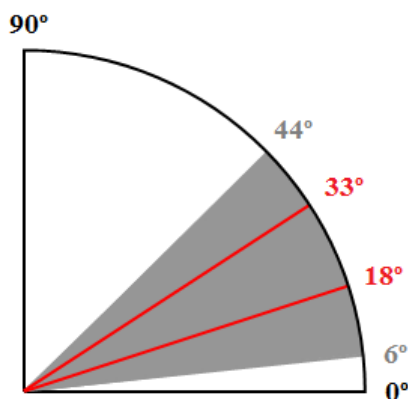


Figura 3. Interpretación de la distancia angular con respecto a su mínimo y a su máximo

distancia global considerando las palabras enteras, con excepción del último segmento del nombre más largo, que se descartó del análisis por las razones arriba mencionadas y que, en todo caso, aumentaría la distancia, debido a que la presencia de dicho sonido se opone a su ausencia en la otra marca.

3.2.2. Análisis de la distancia en la parte común

Como es lógico, la parte común de ambos nombres presenta una menor distancia (con un promedio de 13°), similar al promedio de 12° que se calculó anteriormente como atribuible a la variación intralocutor. No obstante, nótese que dentro de esta secuencia de cinco segmentos hay dos en los que la distancia se amplía. Uno de ellos es la primera vocal, que presenta un mínimo de 8° y un promedio de 13° pero llega en algún punto hasta los 18°. Esta diferencia se justifica por el hecho de que dicha vocal

se encuentra en una sílaba acentuada en un caso, pero no acentuada en el otro. El último segmento de la parte común muestra una distancia mayor entre ambas marcas, con un mínimo de 8° y un máximo de 30° (promedio de 19°). Esto se debe sin duda al efecto de la coarticulación entre dicho sonido y la consonante siguiente, que es muy distinta en ambos casos, de modo que la propia pronunciación se ve ya modificada hasta cierto punto por anticipación del sonido que viene a continuación. Esta idea se ve corroborada por el hecho de que la distancia acústica entre las dos realizaciones del quinto segmento aumenta a lo largo de su duración (es decir, cuanto más próximo se halla al sonido siguiente).

3.2.3. Análisis de la distancia en la parte distintiva

Por su parte, el intervalo distintivo entre las dos marcas ([^hɣa] y [ʒi]) se corresponde con una distancia media de 27° (30° de promedio en la consonante y 25° de promedio en la vocal), con el máximo de 33° en el contraste [ɣ] – [ʒ], como se ha mencionado. La Figura 3 ofrece una representación gráfica de lo que supone un ángulo de 18° (promedio de distancia acústica entre los dos nombres) y un ángulo de 33° (máxima distancia acústica entre dichas palabras). Como referencia, aparece sombreada la zona comprendida entre el umbral mínimo estimado en 6° y el máximo, que corresponde a los 44°. Teniendo en cuenta esta zona de variación efectiva, el promedio de 18° representa casi un tercio (32 %) del total del rango disponible, mientras que el punto máximo de distancia acústica entre los nombres de ambas marcas (33°) supera los dos tercios (71 %).

4. DISCUSIÓN

4.1. Interpretación de los coeficientes cepstrales

Un análisis basado en valores mel-cepstrales, como el que aquí se propone, ofrece la ventaja de que los distintos coeficientes permiten descomponer la señal acústica en piezas de información perceptivamente relevantes. Como contrapartida, puede suponer un problema para interpretar las distintas dimensiones de los vectores MFCC. Es razonable pensar que no todas esas dimensiones sean igual de importantes para caracterizar las diferencias en la identidad de los segmentos, que es lo primordial a la hora de estudiar el grado de parecido entre dos nombres de marcas. Por el contrario, algunos de los coeficientes se refieren a propiedades de la fuente, lo que puede incluir el valor de la f_0 o distintas características de la cualidad

de la voz. Otros, aun referidos al filtro, recogen información sobre frecuencias altas, que pueden no ser relevantes para el contenido segmental. Finalmente, resulta necesario deslindar qué coeficientes se ven más afectados por características individuales del locutor.

Por todo esto, resulta necesario estudiar con mayor profundidad los correlatos de los distintos MFCC, mediante experimentos muy controlados que aislen la variación en tan solo uno de los aspectos mencionados (propiedades de la fuente, identidad de los segmentos, características del locutor).

4.2. Interpretación de la distancia angular

La interpretación de una distancia medida en ángulos puede no resultar intuitiva y, en todo caso, depende de haber establecido previamente los límites inferior y superior del rango de variación posible. En este trabajo se ha propuesto, como método para determinar el mínimo, calcular la distancia entre distintas producciones de la misma palabra por parte de un locutor. Por su parte, la estimación del máximo se ha basado en la distancia, también intra-locutor, entre dos sonidos altamente dispares (aquí se ha escogido la [s] y la [u], por las razones expuestas en § 2.5).

Ahora bien, con el fin de consolidar unas referencias estables para el mínimo y el máximo, en el futuro se impone ampliar el corpus sobre el que se han basado estos cálculos. Por un lado, cabe cuestionarse cuánta más distancia podría existir entre los pares comparados si fueran pronunciados por distintas personas (es decir, incorporando la variación inter-locutor). Esto se relaciona con uno de los aspectos mencionados en el apartado anterior (§ 4.1). En todo caso, sería recomendable aumentar el número de locutores y grabar muchas más instancias de un número también mayor de palabras, para establecer sus mínimos de variación intra-locutor y poder calcular un promedio de dichos mínimos. Asimismo, conviene obtener el máximo a partir de un promedio de distancias que consideren más pares de sonidos.

4.3. Comparación de nombres de longitud desigual

Una de las limitaciones de este estudio aparece cuando se intenta comparar dos palabras con diferente número de segmentos. Como se ha explicado, no se puede insertar un silencio en la palabra más corta y calcular la distancia entre este y el sonido de la palabra más larga. En el proceso de cálculo de los MFCC, los silencios se sustituyen por

ruido blanco. Aunque los vectores se definieran manualmente como una secuencia de ceros, tampoco sería posible utilizarlos para calcular la distancia angular, puesto que en la fórmula (1) habría que dividir por cero.

En este trabajo, la longitud de los denominativos comerciales comparados tan solo difería en un segmento, y se ha decidido descartarlo del análisis acústico, con el razonamiento de que, en todo caso, la diferencia entre la presencia y la ausencia de cualquier sonido solo aumentaría la distancia acústica entre ellos. No obstante, podrían considerarse otras posibles soluciones ante este tipo de casos, si bien es cierto que no se anticipa que existan verdaderos casos de conflicto entre marcas si sus nombres difieren demasiado en longitud.

5. CONCLUSIONES

Este artículo presenta una propuesta metodológica para calcular la distancia fonética entre dos nombres de marcas en conflicto. En el ámbito de la comparación de marcas, el hecho de que con frecuencia los nombres carezcan de significado hace que el peso de la distinción entre dichas marcas recaiga fundamentalmente sobre el significante (frente a otros posibles planos de análisis, como el morfológico, el semántico, etc.). Por lo tanto, merece la pena seguir explorando la distancia entre vectores cepstrales como una posible manera de objetivar la diferencia entre las marcas.

6. REFERENCIAS BIBLIOGRÁFICAS

- Boersma, P., & Weenink, D. (2017). *Praat: Doing phonetics by computer* (Versión 6.0.35) [Programa informático].
- Cicres, J. y Turell, M. T. (2014). Investigación, docencia y práctica profesional en lingüística forense. En Garayzábal Heinze, E., Jiménez Bernal, M. y Reigosa Riveiros, M. (Coords.), *Lingüística forense: la lingüística en el ámbito legal y policial* (pp. 376-391), Euphonia Ediciones.
- Eaton, J. W., Bateman, D., Hauberg, S., & Wehbring, R. (2015). *GNU Octave version 4.0.0 manual: A high-level interactive language for numerical computations*.
- Erro, D., Sainz, I., Navas, E., & Hernández, I. (2011). HNM-based MFCC+F0 extractor applied to statistical speech synthesis. En *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 4728-4731).
- Ramírez Salado, M. (2017). Antecedentes de la lingüística forense: ¿desde cuándo se estudia el lenguaje como evidencia? *Pragmalingüística* (25), 525-539.

SPEECH PRODUCTION IN NOISE BY CHINESE LEARNERS OF ENGLISH

Bin Li, Qianwen Guan, Zixia Fan, Mosi He
City University of Hong Kong, China

ABSTRACT

This study examined effects of noise on speech production in first and second languages (L1 and L2) by examining correlates of segments and suprasegmentals produced by experienced Chinese learners of English. Vowels were significantly lengthened while consonants remained comparable under babble noise in both Chinese and English. Fundamental frequency was unchanged in Chinese, but raised in English in both quiet and noisy conditions. Phonetic realization of lexical stresses was partially maintained under noise to achieve contrast between strong and weak syllables. Our findings reveal Lombard effect in Chinese speakers' L1 and L2 speech production. Patterns of phonetic modification in Lombard speech appear language-specific, which may be attributed to L1 transfer or limited experience in L2 communication under adverse conditions.

Keywords: lombard effect, speech production in noise, Chinese learners of English, lexical stress

1. INTRODUCTION

Speakers tend to involuntarily increase vocal efforts when speaking in adverse conditions such as under noise, which is referred to as the Lombard effects (Lombard, 1911). Lombard speech has been examined and found in a few languages, such as English (Junqua, 1993), French (Garnier, 2014), Spanish (Castellanos et al., 1996), Japanese (Takizawa and Hamada, 1990), etc. Acoustically, Lombard effect is accompanied typically by an increase in vocal amplitude. However, previous studies found that modifications to other phonetic cues, such as duration and fundamental frequency (F0), were present in Lombard speech depending on phonetic contexts. For instance, acoustic adaptation in noisy conditions seemed to favour vowels rather than consonants. Junqua (1993) found that vowels are more likely to be elongated while consonants slightly shortened. Garnier (2014) reported background noise induced greater changes in intensity and spectral energy of vowels than in those of consonants. In addition to fine-grained modification of segmentals, Arciuli and colleagues (2014) examined English lexical stresses in noise and found that English speakers did not increase vocal intensity equally across syllables, but altered degree of contrast between strong and weak syllables in Lombard speech. In the production of lexical stresses, both intensity and fundamental frequency

showed greater differences between strong and weak syllables in noise than in quiet.

The phonetic modifications triggered by background noise are interpreted as attempts to facilitate speech communication in adverse conditions, where speakers employ measures available to enhance the intelligibility of speech in their native language. Much less is known, however, whether and how non-native speech shows such effect and modification, as second-language (L2) speakers who are limited by their linguistic experience may have more difficulty in identifying L2 speech targets in noisy conditions than native listeners (Li, Wayland, Zhang, 2012). For the production, Marcoux and Ernestus (2019) reported the non-native Lombard speech in terms of F0 range by examining native speakers of Dutch whose Lombard speech showed the influence of their native language. Previous studies warrant an investigation of the impact of noise on L2 speech production and remedial strategies if any by L2 learners.

2. STUDY DESIGN

This study aimed to investigate universality and embodiments of the Lombard effect by comparing native and non-native speech production under noise. To this end, we examined the Lombard effect in Chinese learners' English-as-an-L2 (ESL) speech production under noise, in terms of phonetic realizations of vowels, consonants, and lexical

stresses. Much research has been devoted on ESL acquisition of suprasegmentals by learners whose native rhythmic mechanisms differ. Chinese learners are known to have difficulties in mastering English lexical stresses. In this study, we intended to examine the presence of Lombard effect in Chinese speakers' ESL production in noise with a reference to their L1. In other words, if L2 phonetic patterns by Chinese speakers shows any variation in noise from that in quiet, and any difference from their L1 production. As L2 learners may lack experience in dealing with degraded speech signals or self-monitoring under adverse conditions, whichever patterns of interlanguage or L1 transfer they show in their L2 speech in a quiet condition may not be fully maintained under noise. L2 learners' phonetic performance, if different from that of native speakers in quiet, would remain different if not more distant in babble noise, because of the acoustic and verbal interference.

In order to elicit phonetic modification if any and for systematic comparison, we selected bisyllabic Chinese words varying in lexical tones and English words varying in the location of stress assignment. Altogether, we selected 18 Mandarin words, 30 Cantonese words, and 28 English words. The background noise in the noisy condition was synthesized from intensity-normalized babbles recorded in cafeteria and canteens. The babble noise was played out via loudspeakers at 76 dB while participants talked. Speech enhancement method using A Priori Signal-to-Noise Ratio (Scalart, 1996) was adopted before acoustic analysis to filter out noise from speech signals. However, this method may suppress harmonics in speech signals during noise reduction, and thus cause undesirable degradation or normalization of intensity levels. Therefore, we did not include intensity in acoustic analysis, though it is an important cue to lexical stresses. The parameters to measure include duration of consonant, duration and peak F0 of vowel, the latter two of which were then fed to the formula below, where d_k is duration or peak F0 of the k th syllable in a word, to calculate normalized pairwise-variability-index (nPVI) for stress realization. The nPVI reflects the degree and direction of contrastive stress across syllables in a word (Arciuli et al., 2014, Nolan and Asu, 2009). A higher absolute nPVI indicates greater contrast between strong and weak syllables.

$$nPVI = 100 \times |(d_k - d_{k+1}) / ((d_k + d_{k+1})/2)|$$

We recruited three native speakers of Chinese who are experienced ESL users: two Mandarin speakers and one Cantonese speaker. They produced the word list in their L1 and then that in English, in both quiet and noisy conditions, totaling 300 trials. All recordings were then transcribed for phonetic analysis in Praat (Boersma and Weenink, 2019).

For data analysis, we adopted linear mixed models in R (R Core Team, 2015) using the lme4 package (Bates, Maechler, Bolker, and Walker, 2015). A set of models was run with the tested acoustic parameters as dependent factors. The fixed factors were Condition (noise & quiet) and Language (L1 & L2) and the interaction between them. All models were specified for random intercept for Speaker and for Word. *P*-values were obtained by a likelihood-ratio-test in which the full model was compared to a model without the factor in question. If necessary, Tukey posthocs were conducted with the lsmeans package (Lenth and Hervao, 2008).

3. RESULTS AND DISCUSSION

3.1 Consonants

We measured duration of consonantal parts of all syllables. As presented in Table 1, consonants in noise were numerically shorter than their counterparts in L1 but sustained in L2.

Table 1. Mean duration of consonants. (Values in brackets are standard deviation)

	Language	Condition	Mean (SD)	<i>P</i> value
Duration (ms)	L1	Quiet	96.1 (47.3)	0.24
		Noise	90.2 (47.8)	
	L2	Quiet	113.2(51.4)	0.92
		Noise	113.2 (53)	

P-values were determined by least-squares means comparison using the lsmeans package in R.

3.2 Vowels

We measured duration and peak F0 of vowels. As listed in Table 2, vowels under babble noise were much longer than their counterparts in the quiet condition ($\chi^2(2) = 8.33, p = 0.01$) in both L1 and L2. Chinese speakers' F0 did not change in L1, but was raised in L2 in both quiet and noise. The spectral changes in F0 however, were not significant.

Table 2. Mean duration and peak F0 of vowels. (Values in brackets are standard deviation)

	Language	Condition	Mean (SD)	P value
Duration (ms)	L1	Quiet	158.4(61.4)	0.04
		Noise	167.6(62.5)	
	L2	Quiet	123.3(64.2)	0.03
		Noise	133.4(65.6)	
F0 (Hz)	L1	Quiet	217 (54.3)	0.88
		Noise	217.6(53.6)	
	L2	Quiet	228.6(55.1)	0.32
		Noise	232.3(54.6)	

P-values were determined by least-squares means comparison using the lsmeans package in R.

Furthermore, for both consonants and vowels, language (L1 and L2) was found to interact with speaking conditions (vowels: $\chi^2(2) = 30, p < 0.0001$; consonants: $\chi^2(2) = 8.49, p = 0.014$). In other words, Chinese speakers adjusted to noisy conditions well in their native tongues in a similar way, but not as successfully in their L2, showing more individual variations in phonetic modifications.

3.3 Lexical stresses

To further compare the Lombard effect in L1 and L2, we examined ESL production of lexical stresses by calculating nPVI using duration and peak F0 of vowels in strong and weak syllables. There were both trochaic and iambic words on the English word list, and thus SW and WS patterns.

In quiet condition, durational contrast in SW words is significantly smaller than that in WS ones, and the pattern is maintained under noise. The variation pattern between trochaic and iambic words is consistent with previous findings (Li, Cao, Yan 2014). But the mean range in SW words slightly shrinks under noise. Moreover, nPVI of vowel duration showed no general difference for either SW or WS syllables in quiet or noisy conditions, as shown in Figure 1.

3.4 General Discussion

Results from acoustic analysis reveal patterns of modification in Lombard speech by Chinese speakers. In general, consonants were not lengthened as much as vowels were in noise. The temporal increment in vowel production under noise is significant in both L1 and L2. The nPVI reveals that durational contrast of syllables in trochaic words was weakened, suggesting durational lengthening may have been applied to all syllables. Chinese speakers' F0 in L1 was the same in quiet and noise, which is expected in tonal languages. Interestingly, their peak

F0 was raised when producing English in quiet and further in the noise condition.

The tendency of modification in F0 and duration echoes previous findings that Chinese ESL learners may use a falling tone to signal a stress or apply a high tone overall while lengthening a stressed syllable (Lai, 2009; Zhang, Nissen, Francis, 2008). In our study, Chinese ESL users adopted both temporal and spectral methods in achieving the phonetic contrast between strong and weak syllables, which were sustained under noise.

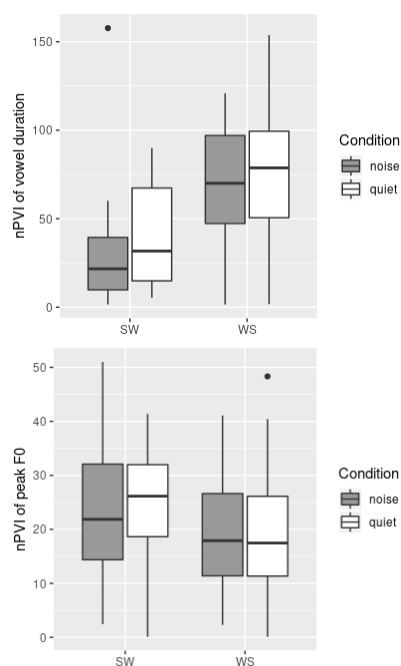


Figure 1: nPVI for vowel duration (upper panel), and peak F0 (lower panel) of English words produced by Chinese speakers in noise and in quiet.

The slight weakening of contrast in terms of vowel duration in SW words prompts certain impact of noise on temporal adjustment. Noise may have distracted the speakers' attention to phonetic details. They engaged much effort in filtering out interference from noise and that from linguistic cues in babbles. Cognitive load in maintaining clear speech may have been too demanding for non-native users to allot sufficient attention to phonetic accuracy at a higher-level in the target language. There is, however, another interpretation of the changes to phonetic contrast under noise. Chinese speakers could have realized the impact of noise on their speech and thus adopted the general strategy of cue enhancement, including F0 increment and vowel lengthening in L2 and in noise. Compensation to signal degradation may override that to phonetic cues, therefore their modification was implemented

in an overall manner on all syllables, softening the contrast between strong and weak syllables rather than maintaining or least enhancing it.

4. SUMMARY

This study investigated effects of noise on L2 production of segments and suprasegmentals by comparing phonetic correlates of consonants, vowels and lexical stresses by Chinese ESL users in quiet and in noise. Results confirmed Lombard effects in Chinese speakers' L1 and L2 production. Modification in amount and strength is observed more in vowels than consonants in noise, which is consistent with Junqua (1993) that reported greater Lombard effect in native speakers' production of vowels than in that of consonants. Moreover, in contrast to some previous studies (Junqua 1993, Garnier et al. 2010), fundamental frequency (F0) was maintained under the noise condition in L1 Chinese but slightly raised in L2 English. This finding confirms L1 influence in ESL speech production and language-specific strategies in phonetic modifications in Chinese and English. ESL data also suggest a potential trade-off relation between responses to general speech degradation and modification to fine phonetic details. But this hypothesis should be examined using well-controlled tonal contexts since Chinese (both Cantonese and Mandarin) is a tone language. Future research could test hypotheses of our study using more acoustic parameters and recruiting a larger pool of speakers.

5. ACKNOWLEDGEMENT

This study is supported by EDB(LE)/P&R/EL/175/12 from SCOLAR of Hong Kong and by Project 6391207 from the UGC T&L Scheme.

6. BIBLIOGRAPHY

- Arciuli, J., Simpson, B. S., Vogel, A. P., & Ballard, K. J. (2014). Acoustic changes in the production of lexical stress during Lombard speech. *Language and speech*, 57(2), 149-162.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 667,1-48.
- Boersma, P. & Weenink, D. (2019). Praat: Doing phonetics by computer (version 6.1.04). [Computer program], <<http://www.praat.org/>>.
- Castellanos, A., Benedí, J. M., & Casacuberta, F. (1996). An analysis of general acoustic-phonetic features for Spanish speech produced with the Lombard effect. *Speech Communication*, 20(1-2), 23-35.
- Garnier, M., Henrich, N., & Dubois, D. (2010). Influence of sound immersion and communicative interaction on the Lombard effect. *Journal of Speech, Language, and Hearing Research*, 53, 588-608.
- Garnier, M. & Henrich, N. (2014). Speaking in noise: How does the Lombard effect improve acoustic contrasts between speech and ambient noise? *Computer Speech & Language*, 28(2), 580-597.
- Junqua, J.C. (1993). The Lombard reflex and its role on human listeners and automatic speech recognizers. *The Journal of the Acoustical Society of America*, 93(1), 510-524.
- Lai, Y. W. (2009). Acoustic Realization and Perception of English Lexical Stress by Mandarin Learners. Ph.D. dissertation, University of Kansas.
- Lenth, R. & Hervao, M. (2018). Using the lsmeans package. <http://CRAN.R-project.org/package=lsmeans>.
- Lombard, E. (1911). Le signe de l'élévation de la voix [The sign of the rise of the voice]. *Maladies Oreille, Larynx, Nez, Pharynx*, 27, 101-119.
- Li, B., Wayland, R., & Zhang, C. (2010). Contextual information in (non)native speech perception in noise. *Proceedings of the Third ISCA Tutorial and Research Workshop on Experimental Linguistics*. 25-27 August, 2010: 101-104. Athens: International Speech Communication Association.
- Li, B., Cao, W., & Yan, X. (2014). Phonetic cues of lexical stress in Chinese learner English. *Proceedings of 2014 International Conference on Phonetic Research and Language Learning & English Phonetic Conference in China*. 17-19 October, 2014: 150-153. Changsha: Human University.
- Marcoux, K. P., & Ernestus, M. T. C. (2019). Differences between native and non-native Lombard speech in terms of pitch range. *Proceedings of the 23rd International Congress on Acoustics*. 9-13 September, 2019: 5713-5720.
- Nolan, F., & Asu, E. L. (2009). The pairwise variability index and coexisting rhythms in language. *Phonetica*, 66(1-2), 64-77.
- R Core Team (2015). R: A Language and Environment for Statistical Computing. In. Vienna, Austria: R Foundation for Statistical Computing.
- Scalart, P. (1996). Speech enhancement based on a priori signal to noise estimation. In *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*. Vol. 2, 629-632. IEEE.
- Takizawa, Y., & Hamada, M. (1990). Lombard speech recognition by formant-frequency-shifted LPC cepstrum. *Proceeding of First International Conference on Spoken Language*, 293-296.
- Zhang, Y. H., Nissen, S. L., & Francis, A. L. (2008). Acoustic characteristics of English lexical stress produced by native Mandarin speakers. *The Journal of the Acoustical Society of America*, 123(6), 4498-4513.

LA PRODUCCIÓN DE LAS OCLUSIVAS CHINAS POR HISPANOHABLANTES

Zhao Liu, María J. Machuca

East China Normal University, Universitat Autònoma de Barcelona
liuz_liuzhao@163.com, mariajesus.machuca@uab.cat

ABSTRACT

Este artículo trata sobre la producción de las oclusivas chinas por parte de hablantes españoles. Los informantes se dividen en dos grupos: uno de cuatro hablantes nativos del español, aprendices del chino, y otro de seis hablantes nativos del chino, que constituyen el grupo de control. Para la recogida de muestras, se utilizaron dos textos chinos, uno fonéticamente equilibrado en cuanto a los fonemas consonánticos y, otro, que servía para completar los diferentes contextos vocálicos. Se categorizan las realizaciones en función de las observaciones espectrográficas.

Los resultados muestran que los aprendices pueden producir correctamente las oclusivas del chino en un nivel de aprendizaje intermedio. La aspiración no es un obstáculo, el punto de articulación dentoalveolar es el que presenta problemas en su producción. La revisión detallada de los datos señala que las realizaciones pueden ser idénticas, pero debidas a contextos diferentes y que la influencia interlingüística en la adquisición de las lenguas extranjeras es un fenómeno complejo.

Palabras clave: fonemas oclusivos, producción, chino, inglés, español

The aim of this paper is to look at the production of Chinese stop by Spanish learners. 10 participants took part in this study. They were divided into two groups according to their native language: 4 native Spanish speakers and 6 native Mandarin Chinese speakers (control group). They were asked to read two Chinese stories that contained the target sounds. It is essential to notice that one of the two passages is phonetically balanced in terms of the consonant. The other one was used to obtain more vowel contexts. The data were categorized using spectrographic analysis.

The results suggest that Spanish learners seem to produce Mandarin stops correctly. Overall, the aspirated stops were produced successfully. The place of articulation which appears to be more challenging is the dentoalveolar. Despite these findings, the detailed revision of the data indicates that even though the production seems to be similar between the two groups, it may be motivated by different vowel contexts. Therefore, the crosslinguistic influence seems to be a complex phenomenon.

Keywords: stop production, Mandarin Chinese, English, Spanish

1. INTRODUCCIÓN

En primer lugar, es necesario hacer una descripción breve de las oclusivas, objeto de este estudio, y una comparación de estas consonantes en las dos lenguas tratadas: el chino y el español. Estas dos lenguas cuentan con seis fonemas oclusivos, a pesar de que hay una diferencia en cuanto a la sonoridad, a la aspiración y al punto de articulación.

Los fonetistas están de acuerdo en que los fonemas oclusivos del español se clasifican en tres

puntos de articulación: el bilabial, el dental y el velar. No obstante, en chino no se da ese acuerdo. Los estudios sobre los fonemas oclusivos del chino señalan que el chino y el español comparten los puntos de articulación bilabial y velar. Sin embargo, no ocurre lo mismo con el punto de articulación dental. Existe una discrepancia en cuanto al punto de articulación de los fonemas oclusivos /t, t^h/; hay autores que los clasifican como dentales o alveolares (Y. -H. Lin, 2007). En este estudio, los hemos considerado dentoalveolares siguiendo a T. Lin y Wang (2013).

Centrándonos en estas clasificaciones arriba mencionadas, podemos establecer una primera comparación entre los fonemas oclusivos en chino y en español, que se observa en la Tabla 1, donde podemos apreciar que el chino y el español comparten los fonemas oclusivos sordos /p, t, k/. Además, mientras que el chino cuenta con los fonemas oclusivos sordos aspirados /p^h, t^h, k^h/, inexistentes en español, el español tiene los fonemas sonoros, de los que carece el chino. Los fonemas oclusivos sonoros del español, además de la realización de una oclusiva, también posee la de una aproximante en determinados contextos.

Tabla 1: Fonemas oclusivos en chino y en español.

		Bilabial		Dentoalveolar		Velar	
		No	Sí	No	Sí	No	Sí
Aspiración	Sonoridad						
	Sí	/p ^h /		/t ^h /		/k ^h /	
No	Sonoridad						
	Sí	/p/		/t/		/k/	
Español			/b/		/d/		/g/

Existen muchos estudios sobre la adquisición de una lengua extranjera, pero la mayoría de ellos no consideran si esa lengua adquirida, en realidad, es segunda o tercera lengua; cada día hay un número mayor de hablantes que dominan más de una lengua extranjera, por lo que los estudios sobre la adquisición de lenguas no deberían limitarse a la segunda lengua extranjera aprendida (L2), sino también a la tercera (L3) o la cuarta (L4) (véase, por ejemplo, Cenoz, 2013, De Angelis, 2007). Los aspectos que se suelen estudiar sobre la adquisición de terceras lenguas varían en función del parámetro que se esté considerando: la fuente de transferencia, el efecto de la influencia de las lenguas adquiridas con anterioridad, entre otros.

En el nivel fonético y fonológico, la adquisición de la L3 es compleja (Wrembel, 2010), ya que todas las lenguas que domina un hablante pueden ser la fuente de influencia, tanto la L1 (Llisterri, 1987; Patience, 2018) como la L2 (Gut, 2010; Tremblay, 2007; Zhu y Mok, 2015), o una combinación de ambas (Llama y López-Morales, 2016; Wrembel, 2011). Además, la influencia proveniente de estas lenguas puede ser positiva (Bailey, 2013), facilitando la adquisición de la L3, o negativa (Chamot, 1973, citado en De Angelis, 2007), obstaculizándolo. También hay autores que señalan la inadecuación de generalizar la influencia, ya que puede variar de acuerdo con más factores (Bentahila, 1975). De esta manera, hay estudios que afirman que los aprendices pueden llegar a realizar producciones

similares a la de los nativos (Wrembel, 2014) y hay los que apuntan una dificultad de poseer una producción correcta en la lengua meta (Maldonado y Álvarez Piña, 2018).

2. HIPÓTESIS

Partiendo de lo que se ha comentado en el apartado de *Introducción* sobre los fonemas oclusivos de las dos lenguas, este estudio pretende contestar a la pregunta general de si los aprendices hispanohablantes del chino tienen alguna dificultad en la producción de los fonemas oclusivos del chino. Para ello, debemos contestar también a otras preguntas específicas relacionadas:

1. ¿Hay algún fonema que constituya un obstáculo mayor para los aprendices de chino en comparación con otros?
2. ¿La aspiración, por ser inexistente en los fonemas oclusivos del español, es un problema para los aprendices de chino?
3. A raíz de los resultados, ¿se podría observar alguna influencia interlingüística entre las lenguas dominadas por los hablantes?

3. METODOLOGÍA

3.1. Informantes

Para la selección de informantes, se repartió a los posibles candidatos un formulario de Google para obtener el historial lingüístico del hablante. En él se hacían preguntas detalladas sobre las lenguas que dominaban y otros aspectos relacionados con la dificultad de adquirir ciertos sonidos.

Los hablantes que conforman el objeto de este estudio se clasifican en dos grupos atendiendo a su lengua materna: un grupo de hablantes nativos del chino (grupo CH) y otro, de nativos del español (grupo ES). Ambos grupos comparten la similitud en el aprendizaje de una segunda lengua, el inglés, y el nivel de su L2 inglés es comparable, entre B2 y C1.

El grupo CH, que es considerado el grupo de control, consta de seis hablantes femeninos. El grupo ES, grupo de observación, debido a que no se han encontrado más hablantes con un historial lingüístico similar, está constituido por cuatro hablantes femeninos que tienen el español como lengua materna y que están aprendiendo chino como L3 y su nivel de conocimiento del chino es de B2. Todos los hablantes tienen una edad de entre 21 y 24 años.

3.2. Corpus

Los datos para el análisis se han extraído de un corpus más amplio empleado para un estudio sobre terceras lenguas (véase, para más detalle, Liu, 2019).

Se crearon dos textos *ad hoc* para las grabaciones de la tarea de lectura de textos. El primer texto, titulado *El año nuevo*, es un texto fonéticamente equilibrado, en cuanto a los fonemas consonánticos, que tuvo que ser creado a propósito, ya que no existía un texto fonéticamente equilibrado en chino.

Se creó un texto más para la grabación, con el fin de obtener más contextos vocálicos, aspecto que no se había podido tener en cuenta en el primer texto.

El uso de los dos textos nos permite recoger un número relativamente mayor de muestras para su posterior análisis.

3.3. Procedimiento

La grabación se llevó a cabo en una sala de grabación de insonorización, perteneciente al laboratorio del Servei de Tractament de la Parla y del So de la Universitat Autònoma de Barcelona.

Cabe apuntar que los textos fueron presentados en la caligrafía china, es decir, los caracteres, ya que el sistema romanizado (*pinyin*) podría haber perjudicado o afectado a la lectura de los informantes, causando una correlación equivocada entre los sonidos y las grafías.

Antes de proceder a la grabación, se mantuvo una conversación en lengua china con los informantes de una duración aproximada de cinco minutos. Pensamos que esto podría facilitar la recogida de muestras, ya que, por una parte, se introducía a los informantes en un contexto en el que se empleaba la lengua que iba a ser objeto del estudio y, por otra, el hablante iniciaba de una forma más natural y relajada la lectura del texto.

Los informantes rellenaron un formulario de consentimiento dando permiso a la exploración y al uso de estos datos.

3.4. Tratamiento de los datos

Las muestras de los hablantes han sido segmentadas y etiquetadas manualmente con la herramienta PRAAT (Boersma y Weenink, 2017). Se extrajo un total de 673 casos, 392 corresponden a las realizaciones del grupo de los hablantes nativos (Grupo CH) y 281, al de los aprendices (Grupo ES). El análisis estadístico de los datos se llevó a cabo con la herramienta software R (R CoreTeam, 2014). Se

analizaron los datos teniendo en cuenta el fonema, el punto de articulación y la aspiración.

4. RESULTADOS

En esta sección se presentarán los resultados del análisis de las muestras considerando el fonema, el punto de articulación y la aspiración.

4.1. Fonema

Las realizaciones del fonema bilabial sordo aspirado /p^h/ por parte de ambos grupos de informantes son categorizadas como canónicas, de ahí que no se ha procedido el análisis estadístico.

En cuanto a /t^h/, los hablantes aprendices produjeron una sola realización para el fonema dentoalveolar sordo aspirado correspondiente a una realización canónica, mientras que el grupo de nativos del chino produjeron dos realizaciones: una canónica (93.8 %) y otra sonorizada (6.2 %). Los resultados del análisis estadístico muestran que hay una diferencia significativa entre ambos grupos de informantes: $\chi(1, N=108) = 4.45, p < .05$.

En la realización del fonema velar sordo aspirado /k^h/, el grupo ES, aprendices del chino, presentaba realizaciones canónicas en todos los casos, mientras que los hablantes del grupo CH, nativos del chino, además de presentar la realización canónica (95.8 %), produjeron sonorizaciones (4.2 %), pero esta diferencia no resulta ser significativa: $\chi(1, N=39) = 2.46, p > .05$.

Con relación al fonema bilabial sordo no aspirado /p/, los hablantes nativos de esa lengua, los del grupo CH, presentaban cuatro realizaciones diferentes: un 56.68% correspondiente a una realización canónica, un 38.33% a una oclusiva bilabial sonora, un 3.33% a una aproximante y un 1.66% a una fricativa. El grupo ES, por su parte, solo mostraban las dos primeras realizaciones. El análisis estadístico indica que el grupo de los nativos (grupo CH) tiene más realizaciones sonorizadas (38.33 % vs. 4.35 %) que el grupo de los aprendices (grupo ES) y esa diferencia es significativa: $\chi(1, N=106) = 35.29, p < .01$.

Respecto a la realización del fonema dentoalveolar sordo no aspirado /t/, los hablantes nativos (grupo CH) presentan cuatro realizaciones: la de una oclusiva dentoalveolar sorda no aspirada (la canónica, 41.87%), la de una oclusiva dentoalveolar sonora (36.88%), la de una aproximante (20%) y casos de elisión de este fonema (1.25%). Los aprendices (grupo ES) presentan las primeras tres

realizaciones, pero con un porcentaje diferente: la canónica (59.61%), la realización sonorizada (35.58%) y la aproximantizada (4.81%). Las pruebas estadísticas sobre las tres realizaciones comunes para ambos grupos de hablantes revelan una diferencia significativa entre ellos: $\chi(2, N=262) = 12.46, p < .01$. Las comparaciones de dos en dos señalan que las diferencias se encuentran entre la realización canónica y la aproximantizada y entre la sonorizada y la aproximante.

Las realizaciones del fonema sordo no aspirado /k/ por parte de ambos grupos de hablantes se clasifican en cuatro tipos: la canónica (grupo CH, 69.87%; grupo ES, 76.36%), la de una oclusiva sonora (grupo CH, 21.69%; grupo ES, 20%), la de una aproximante (grupo CH, 7.23%; grupo ES, 1.82%) y la de una fricativa (grupo CH, 1.21%; grupo ES, 1.82%). Las pruebas estadísticas sobre las primeras tres realizaciones, ya que la fricativizada cuenta con un número limitado de casos, indican que no hay una diferencia significativa entre ambos grupos de hablantes en la realización de este fonema ($\chi(2, N=136) = 3.49, p > .05$).

4.2. Aspiración

Si se analizan los datos considerando la aspiración (véase Figura 1), se puede apreciar que las oclusivas no aspiradas cuentan con un mayor porcentaje de realizaciones no canónicas que las aspiradas en ambos grupos de hablantes.

Según los resultados de las pruebas estadísticas, el rasgo de aspiración juega un papel significativo en la realización de los fonemas oclusivos del chino, tanto para los nativos (grupo CH, $\chi(1, N=422) = 46.71, p < .01$) como para los aprendices de esa lengua (grupo ES, $\chi(1, N=281) = 30.01, p < .01$).

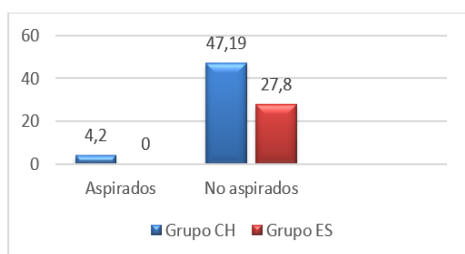


Figura 1: Porcentaje de realizaciones no canónicas de los fonemas oclusivos de ambos grupos en función del punto de articulación.

4.3. Punto de articulación

Si se lleva a cabo el análisis agrupando los fonemas por el punto de articulación (Figura 2), se puede

señalar que hay diferencias entre los dos grupos en el punto de articulación bilabial y en el velar. Las pruebas estadísticas arrojan diferencias significativas entre los tres puntos de articulación para ambos grupos de hablantes: grupo CH, $\chi(2, N=422) = 8.83, p < .05$; grupo ES, $\chi(2, N=281) = 25.58, p < .01$.

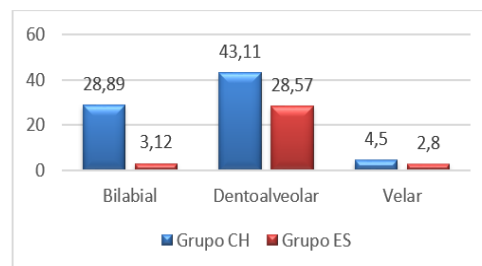


Figura 2: Porcentaje de realizaciones no canónicas de los fonemas oclusivos de ambos grupos en función del punto de articulación.

5. DISCUSIÓN

Los resultados señalan que en todos los fonemas se observa un porcentaje mayor de realizaciones no canónicas en el grupo de nativos, lo que puede indicar que el nativo no se fija en las realizaciones que está produciendo, puesto que es su lengua nativa. Los hablantes nativos del chino, al realizar estos fonemas en su lengua materna, pueden permitirse realizaciones no canónicas. Además, es necesario señalar que los fonemas /t/ y /k/ son los más difíciles de realizar para los aprendices. Es interesante observar que el fonema oclusivo dentoalveolar sordo no aspirado es el que presenta un porcentaje mayor de realizaciones no canónicas para ambos grupos de hablantes (58.13% y 40.39%). Sin embargo, tras una revisión detallada de los datos teniendo en cuenta el contexto precedente (nasal, vocal o pausa), se han observado tendencias diferentes en los dos grupos de hablantes; en cuanto a los nativos, es el contexto nasal que favorece la sonorización de este fonema, mientras que, para los aprendices, una realización sonorizada sucede, en la mayoría de los casos, en la posición postpausal. Esto muestra que los datos deben tratarse con cautela, ya que la misma realización no canónica puede proceder de diferentes motivos.

En cuanto a la aspiración, el hecho de que los aprendices del chino lo realicen correctamente podría deberse a que los aprendices ya hayan adquirido el rasgo de aspiración gracias a su L2 (inglés), y eso haya favorecido el dominio de ese rasgo en la adquisición de su L3 chino. Por tanto, podríamos atribuirlo a una influencia positiva de una L2.

6. CONCLUSIONES

De acuerdo con los resultados de este trabajo, los aprendices del chino pueden llegar a producir las realizaciones como los hablantes nativos chinos.

Siguiendo las preguntas formuladas en el apartado 2, el fonema que más realizaciones no canónicas presenta por parte de los aprendices del chino es el oclusivo no aspirado dental, ya que muestra muchas realizaciones no canónicas, al igual que los hablantes del grupo de control, pero las razones, como hemos visto, son diferentes.

El rasgo de la aspiración no supone un obstáculo porque ya lo tienen asimilado para su L2. En este sentido, se puede pensar que la influencia interlingüística ha favorecido positivamente la adquisición de un rasgo que no se encuentra en la lengua nativa de los aprendices.

Este estudio no dejar de ser un estudio preliminar que cuenta con algunas limitaciones, ya que el número de informantes es relativamente reducido. A pesar de los esfuerzos por conseguir un número mayor de hablantes, ha resultado difícil encontrar más informantes que reunieran los criterios de selección. Un estudio acústico de estas realizaciones nos señalará si las realizaciones que se han categorizado como canónicas son idénticas para los dos grupos de hablantes como se ha mostrado en otros estudios relacionados con los sonidos fricativos (véase Liu y Machuca, 2020).

6. BIBLIOGRAFÍA

- Bailey, A. A. (2013). Similar, yet different: acquisition of Brazilian Portuguese nasal vowels by Spanish-English bilinguals. En *Selected Proceedings of the 16th Hispanic Linguistics Symposium* (pp. 128–142). Cascadilla Proceedings Project Somerville, MA.
- Bentahila, A. (1975). *The influence of L2 on the learning of L3* (TFM). Bangor University, Reino Unido.
- Boersma, P. y Weenink, D. (2017). Praat: Doing phonetics by computer, versión 6.0.36. 5. Consultado el 31 de diciembre de 2019, en <http://www.praat.org>.
- Cenoz, J. (2003). The additive effect of bilingualism on third language acquisition: A review. *International Journal of Bilingualism*, 7(1), 71 – 87.
- Chamot, A. U. (1973). Phonological problems in learning English as a third language. *International Review of Applied Linguistics in Language Teaching*, 11(3), 243.
- De Angelis, G. (2007). *Third or additional language acquisition*. Clevedon: Multilingual Matters.
- Gut, U. (2010). Cross-linguistic influence in L3 phonological acquisition. *International Journal of Multilingualism*, 7(1), 19 – 38.
- Lin, T. y Wang, L. J. (2013). 语音学教程增订版 Yuyinxue Jiaocheng Zengdingban [Manual de fonética (edición revisada por Wang Yunjia y Wang Lijia)]. Peking University Press.
- Lin, Y.-H. (2007). *The Sounds of Chinese*. Cambridge: Cambridge University Press.
- Liu, Z. (2019). *Análisis de las obstruyentes en chino y en español como L3. Estudio acústico y perceptivo para la categorización de errores*. (Tesis doctoral). Universitat Autònoma de Barcelona, Departament de Filologia Espanyola, España.
- Liu, Z. & Machuca, M. (2020, enero). Similitud fonética en la comparación de los fonemas fricativos del chino y del español. Presentación de artículo en el XLIX Simposio Internacional de la Sociedad Española de Lingüística, Tarragona.
- Llama, R. y López-Morales, L. P. (2016). VOT production by Spanish heritage speakers in a trilingual context. *International Journal of Multilingualism*, 13(4), 444 – 458.
- Llisterri, J. (1987). Phonetic interference in bilingual's learning of a third language. En *Proceedings of the 11th International Congress of Phonetic Sciences*, 1– 7 de agosto de 1987 (Vol. 5, pp. 134–137). Tallin, Estonia: Academy of Sciences of the Estonian S. S. R.
- Maldonado, F. O. y Álvarez Piña, M. Á. (2018). *Realizaciones del fonema aproximante retroflejo sonoro en inglés por parte de estudiantes bilingües (árabe y español) en el CEVAC* (TFG). Universidad de Carabobo.
- Patience, M. (2018). Acquisition of the Tap-Trill Contrast by L1 Mandarin – L2 English – L3 Spanish Speakers. *Languages*, 3(4), 42.
- R CoreTeam. (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. 2013.
- Tremblay, M.-C. (2007). L2 influence on L3 pronunciation: Native-like VOT in the L3 Japanese of English-French bilinguals, 3 – 4 de agosto de 2007. *Satellite Workshop of ICPHS XVI*. Freiburg, Germany.
- Wrembel, M. (2010). L2-accented speech in L3 production. *International Journal of Multilingualism*, 7(1), 75–90.

- Wrembel, M. (2011). Cross-linguistic influence in third language acquisition of voice onset time. En W.-S. Lee y E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 2157–2160). City University of Hong Kong. Hong Kong.
- Wrembel, M. (2014). VOT patterns in the acquisition of third language phonology. *Concordia Papers in Applied Linguistics*, 5, 751 – 771.
- Zhu, Y. y Mok, P. P. K. (2015). The acquisition of third language German consonant clusters by Cantonese-English bilinguals. En *Proceedings of the 2015 International Conference on Speech Sciences* (pp. 173 – 174). Seoul.

LA FUNCIÓN DE LAS HESITACIONES EN LA IDENTIFICACIÓN DEL HABLANTE

Joaquim Llisterri, María J. Machuca, Antonio Ríos

Universitat Autònoma de Barcelona

joaquim.listerri@uab.cat, mariajesus.machuca@uab.cat, antonio.rios@uab.cat

ABSTRACT

El análisis de 10.146 hesitaciones, extraídas de un corpus de habla espontánea grabado por 50 hablantes femeninas de español y clasificadas en nueve categorías atendiendo a criterios tanto lingüísticos como estadísticos pone de manifiesto que la frecuencia de aparición, la duración mediana y el rango de duración de los distintos tipos de hesitaciones pueden discriminar a los hablantes en una media del 98 % de los casos. A pesar de que se recogieron dos sesiones de habla espontánea diferentes, no aparecen diferencias significativas entre sesiones de grabación en lo que respecta al empleo de esas hesitaciones en ninguno de los hablantes analizados. Estos datos permiten confirmar que el estudio de las hesitaciones puede ofrecer datos relevantes para la comparación entre muestras de habla en el contexto de la fonética judicial.

Palabras clave: hesitaciones, pausas, alargamientos, habla espontánea, fonética judicial, español

A total of 10.146 hesitations were obtained from a corpus of spontaneous speech in Spanish recorded by 50 female speakers and were classified into nine categories following linguistic and statistical criteria. The results of the analysis show that the frequency of occurrence, the median of duration and the range of duration of the different types of hesitations can discriminate speakers in an average of 98 % of cases. Although two different sessions were recorded, there are no significant differences in the use of hesitations between the two sessions in any of the speakers. Overall, the results confirm that the study of hesitations may offer relevant data for the comparison of speech samples in forensic phonetics.

Keywords: hesitations, fillers, pauses, lengthening, spontaneous speech, forensic phonetics, Spanish

1. INTRODUCCIÓN

El habla conversacional resulta de una interacción entre dos o más interlocutores que comparten la misma situación comunicativa y que contribuyen a su desarrollo sin una planificación previa del mensaje. Por tal razón, este estilo se define como espontáneo: el hablante elabora su discurso en tiempo real y debe planificarlo a medida que va emitiendo su mensaje, de ahí que introduzca alargamientos, pausas llenas, palabras comodín y otros tipos de vocalizaciones o de elementos de relleno (fenómenos definidos como ‘hesitaciones’ o ‘disfluencias’) que le ayudan a organizar lo que va a decir al mismo tiempo que va procesando el habla.

Estos fenómenos, que se han analizado desde diferentes enfoques lingüísticos, pueden ser idiosincrásicos del hablante, por lo que resultan útiles para su identificación en contextos judiciales. Fehringer y Fry (2007), por ejemplo, señalan que las

hesitaciones no solo se dan en todos los hablantes, sino que los caracterizan hasta tal punto que, en una lengua extranjera, un locutor tiende a emplear los mismos recursos relacionados con la planificación del habla que utiliza en su primera lengua.

Ishihara y Kinoshita (2010) estudiaron la elección de los elementos de relleno por parte de hablantes de japonés y la posibilidad de que el elemento elegido caracterice al locutor. Los resultados, basados en la frecuencia de aparición de las hesitaciones, sugieren que el uso de un determinado elemento de relleno identifica a un hablante concreto. No obstante, los autores constataron algunas diferencias que podrían relacionarse con los parámetros acústicos que caracterizan a esos elementos. En este sentido, McDougall y Duckworth (2017) analizaron la frecuencia de aparición y la duración de diferentes hesitaciones extraídas de entrevistas realizadas a 20 hablantes de inglés británico: pausas sonoras, silencios, repeticiones, prolongaciones e interrupciones. El análisis discriminante de los datos

puso de manifiesto que todos los fenómenos observados presentan algún grado de especificidad con respecto al hablante, pero que no es suficiente para distinguir a los 20 locutores. Las hesitaciones que proporcionaron un porcentaje más alto de discriminación fueron las pausas llenas, las repeticiones de palabras y los alargamientos de consonantes fricativas y oclusivas.

Sin embargo, en un estudio similar realizado para el español (Llisterri, Machuca y Ríos, 2019a), se halló que los recursos empleados más frecuentemente son los alargamientos vocálicos y las pausas silenciosas. No obstante, un análisis discriminante considerando estos dos tipos de hesitación solo permite identificar a un 56 % de los 21 hablantes considerados. La duración de las pausas silenciosas y la de los alargamientos vocálicos resulta relevante para la identificación del locutor, mientras que ni la duración de las pausas sonoras ni la de los alargamientos consonánticos contribuyen a mejorar esta identificación.

El presente trabajo se centra en los tipos de hesitación relacionados con el nivel fonético: los alargamientos segmentales, tanto vocálicos como consonánticos, la realización de la conjunción y con valor expletivo, las pausas silenciosas (vacías) y las pausas sonoras (llenas). Aunque estas hesitaciones, como ya se ha señalado, se han abordado desde diferentes perspectivas, el enfoque de esta investigación es el propio de la fonética judicial, con el fin de obtener datos, extraídos de muestras de habla espontánea, que resulten útiles para caracterizar la individualidad de un locutor. Por ello, se analizan, además de la frecuencia de aparición de los tipos de hesitación mencionados, los valores de los parámetros acústicos relacionados con su duración: media, mediana y rango.

2. METODOLOGÍA

Las muestras de habla analizadas proceden de las grabaciones de habla espontánea del corpus CIVIL (San Segundo, Alves y Fernández Trinidad, 2013). En ellas, una entrevistadora conversaba con los participantes durante tres o cuatro minutos sobre temas de trabajo, aficiones, etc., en dos sesiones separadas en el tiempo. La muestra seleccionada está formada por 50 hablantes femeninas.

Los archivos sonoros se segmentaron y etiquetaron mediante el programa Praat (Boersma y Weenink, 2019), siguiendo los criterios definidos en el proyecto VILE (Llisterri, Machuca y Ríos, 2019b); los más relevantes para el presente trabajo se resumen a continuación.

Las pausas silenciosas se analizaron únicamente en el interior de los turnos de palabra de los

participantes. Si después del silencio aparece una oclusiva sorda, se considera que la pausa incluye la oclusión cuando la ausencia de energía sonora supera los 100 ms.

Por otra parte, se ha diferenciado entre pausa sonora, alargamiento vocálico y elemento expletivo. Se etiquetaron como pausas sonoras aquellos segmentos sonoros que no forman parte de una palabra; como alargamientos vocálicos, los segmentos sonoros que se encuentran en una palabra y que duran más de 150 ms; como expletivo, el segmento sonoro correspondiente a una vocal [i], pero que no ejerce la función propia de una conjunción copulativa (Cortés, 1991).

Finalmente, se anotó si los alargamientos pertenecían a palabras funcionales o a palabras con contenido léxico y, en las pausas sonoras y en los alargamientos, también se tuvo en cuenta si se encontraban en posición prepausal o pospausal.

En total se han analizado 10.146 hesitaciones, distribuidas tal como se muestra en la Tabla 1.

Tabla 1: Tipos de hesitación y número de casos analizados (*N*).

Tipo de hesitación	<i>N</i>
Alargamiento consonántico en palabra con contenido léxico	411
Alargamiento consonántico en palabra funcional	237
Alargamiento vocálico en palabra con contenido léxico	3308
Alargamiento vocálico en palabra funcional	1655
Elemento expletivo	218
Pausa silenciosa	2932
Pausa silenciosa seguida de consonante oclusiva	1035
Pausa sonora	350

El análisis estadístico se realizó mediante el programa SAS (versión 9.4), en colaboración con el Servei d'Estadística de la Universitat Autònoma de Barcelona.

Para determinar si existen diferencias estadísticamente significativas entre las dos sesiones de grabación se ha ajustado un modelo lineal mixto, considerando el tipo de hesitación, la sesión y su interacción como factores fijos y el hablante como factor aleatorio.

Se ha realizado también un análisis predictivo de los datos que ha permitido detectar patrones característicos del comportamiento de los hablantes. Para cada uno de ellos, se han ajustado cuatro modelos de regresión logística: el primero a partir de la frecuencia de aparición de cada tipo de hesitación, el segundo, a partir de las medianas de las duraciones, el tercero, a partir del rango de las duraciones y el cuarto, a partir de los factores

obtenidos en un análisis de componentes principales. Este último análisis resulta útil para determinar la correlación entre las variables, ya que cuando estas están muy correlacionadas entre sí, la mayor parte de su variabilidad se puede explicar con muy pocos componentes. Para determinar el poder discriminante de cada uno de los modelos se ha calculado el porcentaje de individuos bien clasificados en cada modelo.

3. RESULTADOS

En primer lugar, se ha comprobado el efecto en la duración de las hesitaciones del tipo de palabra, de la posición con respecto a la pausa y de la presencia de una consonante oclusiva en las pausas silenciosas. Los resultados muestran que existen diferencias estadísticamente significativas entre los alargamientos en las palabras funcionales y los que aparecen en palabras con contenido léxico, tanto en vocales ($p < 0,0001$) como en consonantes ($p < 0,0001$), por lo que se han considerado como

categorías separadas. También se tratan como dos tipos diferentes las hesitaciones (pausas sonoras y alargamientos) que se encuentran simultáneamente precedidas y seguidas de pausa y las que aparecen en el resto de los contextos, ya que se dan diferencias estadísticamente significativas entre ambos tipos. Por el contrario, no se observan diferencias en la duración de la pausa silenciosa en función de si el segmento que le sigue es o no una oclusiva sorda ($p = 0,0844$), así que se toman estos silencios como pertenecientes al mismo grupo.

Mediante esta agrupación se obtienen los nueve tipos de hesitaciones recogidos en la Tabla 2, En conjunto, la frecuencia de aparición de cada tipo no presenta diferencias estadísticamente significativas si se comparan las dos sesiones de grabación realizadas por cada locutor ($p = 0,315$), y tampoco se observa una interacción significativa entre las hesitaciones y las sesiones ($p = 0,2028$). En cambio, se aprecian diferencias estadísticamente significativas entre las duraciones de cada tipo de hesitación ($p < 0,0001$).

Tabla 2: Porcentaje de hablantes que emplean cada tipo de hesitación, valor medio, mediana y desviación estándar de la duración y frecuencia de aparición de las hesitaciones en el corpus.

Tipo de hesitación	Porcentaje de hablantes que emplean el tipo de hesitación	Duración media (en ms)	Duración mediana (en ms)	Desviación típica	Frecuencia de aparición (en %)
Alargamiento consonántico en palabra con contenido	95	196,8	183,1	39,7	4,03
Alargamiento consonántico en palabra funcional	82	272,6	243,4	106,8	2,32
Alargamiento vocálico en palabra con contenido	100	225,3	219,5	35,3	32,45
Alargamiento vocálico en palabra funcional	100	280,1	265,6	69	16
Alargamiento vocálico en palabra funcional simultáneamente seguida y precedida de pausa silenciosa	17	467,8	456	193,9	0,24
Elemento expletivo	74	447,6	419,7	182,4	2,14
Pausa silenciosa	100	412,6	432,7	106,5	38,91
Pausa sonora	86	323,8	277,1	134,7	3,43
Pausa sonora simultáneamente seguida y precedida de pausa silenciosa	28	568,8	611,3	219,5	0,48

3.1 Análisis de componentes principales para la discriminación entre hablantes

Para llevar el cabo el análisis de componentes principales se han excluido dos de los nueve tipos de hesitación presentados en la Tabla 2, puesto que son utilizados por un porcentaje muy bajo de hablantes: las pausas sonoras simultáneamente seguidas y

precedidas de pausas silenciosas (28 % de los locutores) y los alargamientos vocálicos en palabras funcionales seguidas y precedidas simultáneamente de pausas silenciosas (17 % de los locutores). Así, en el análisis de componentes principales se consideran 21 variables: siete relacionadas con la duración mediana, siete con el rango de duración y siete con la frecuencia de aparición de cada tipo de hesitación. Se utiliza la mediana de la duración porque se

considera un valor estadísticamente más robusto que la media en distribuciones de valores como las obtenidas en este trabajo.

Existen nueve factores o componentes que explican un 70,33 % de la variabilidad total y que se correlacionan, de modo diferente cada factor, con los

valores de la mediana y del rango de duración de los diferentes tipos de hesitaciones, así como con su frecuencia de aparición, tal como se muestra en la Tabla 3. El grado de variabilidad total explicado disminuye del primer factor (17,17 %) hasta el noveno (4,92 %).

Tabla 3: Correlación entre parámetros relacionados con cada tipo de hesitación y los nueve factores del análisis de componentes principales.

Tipo de hesitación	Mediana de la duración	Rango de la duración	Frecuencia de aparición
Pausa silenciosa	Factor 4	Factores 1, 3 y 5	Factor 1
Pausa sonora	Factores 5 y 9	Factor 3	Factores 1 y 3
Alargamiento consonántico en palabra funcional	Factor 4	Factores 3 y 8	Factor 1
Alargamiento consonántico en palabra con contenido	Factor 2	Factor 6	Factores 1 y 9
Alargamiento vocálico en palabra funcional	Factor 2	Factor 2	Factor 1
Alargamiento vocálico en palabra con contenido	Factor 2	Factores 2, 3 y 8	Factor 1
Elemento expletivo		Factores 1 y 7	Factor 1

3.2. Modelos predictivos por hablante

En la Tabla 4 se recoge el porcentaje de discriminación de los 50 hablantes analizados en función de cuatro variables globales, referidas a los siete tipos de hesitación empleados en el análisis de

componentes principales. La variable relativa al análisis de componentes principales comprende las variables frecuencia de aparición, mediana de la duración y rango agrupadas en los nueve factores mencionados en el apartado 3.1.

Tabla 4: Resumen de los datos sobre discriminación de 50 hablantes en función de las variables consideradas.

Variable	Media de hablantes discriminados	Mediana de hablantes discriminados	Desviación típica	Porcentaje mínimo de hablantes discriminados	Porcentaje máximo de hablantes discriminados
Frecuencia de aparición del tipo de hesitación	96 %	99 %	6 %	76 %	100 %
Mediana de la duración de la hesitación	96 %	100 %	6 %	73 %	100 %
Rango de duración de la hesitación	93 %	96 %	8 %	67 %	100 %
Análisis de componentes principales (9 factores)	98 %	100 %	4 %	80 %	100 %

4. CONCLUSIONES

Los resultados muestran que con los nueve factores del análisis de componentes principales se discrimina una media del 98 % de los 50 hablantes considerados. En el corpus analizado, la frecuencia de aparición de cada tipo de hesitación combinada con el rango de duración constituye el elemento más importante para identificar a un locutor. Si se toman aisladamente las variables, la frecuencia de aparición del tipo de hesitación es la que discrimina entre

locutores en mayor medida, mientras que el rango de duración es la que discrimina en un porcentaje más bajo. Por tanto, en el presente trabajo se confirma lo que ya habían puesto de relieve Ishihara y Kinoshita (2010) y Braun y Rosin (2015) en el sentido de que la combinación de diversos elementos contribuye a establecer los rasgos que permiten determinar la semejanza entre muestras de habla. En el caso de los parámetros acústicos, parece que los valores relacionados con la dispersión, como el rango, tienen una mayor incidencia en la discriminación que los valores centrales como la mediana, siempre que se

combinen con la frecuencia con la que aparece cada tipo de hesitación.

Es interesante también señalar, desde una perspectiva contrastiva, que en el corpus analizado en este trabajo las hesitaciones más frecuentes son los alargamientos vocálicos y las pausas silenciosas (véase la Tabla 1), mientras que en el estudio sobre el inglés británico de McDougall y Duckworth (2017), las pausas sonoras y las silenciosas constituyen las hesitaciones más frecuentes y los alargamientos aparecen en menos ocasiones que en español.

5. BIBLIOGRAFÍA

- Boersma, P. y Weenink, D. (2019). *Praat: Doing phonetics by computer* (Versión 6.1.05) [Programa informático]. University of Amsterdam. <http://www.praat.org>
- Braun, A. y Rosin, A. (2015). On the speaker-specificity of hesitation markers. En The Scottish Consortium for ICPHS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. University of Glasgow. <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0731.pdf>
- Cortés, L. (1991). *Sobre conectores, expletivos y muletillas en el español hablado*. Ágora.
- Fehringer, C. y Fry, C. (2007). Hesitation phenomena in the language production of bilingual speakers: The role of working memory. *Folia Linguistica*, 41(1-2), 37–72. <https://doi.org/10.1515/flin.41.1-2.37>
- Ishihara, S. y Kinoshita, Y. (2010). Filler words as a speaker classification feature. En M. Tabain, J. Fletcher, D. Grayden, J. Hajek y A. Butcher (Eds.), *Proceedings of the 13th Australasian International Conference on Speech Science and Technology* (pp. 34–37). Australasian Speech Science and Technology Association.
- Llisterri, J., Machuca, M. J. y Ríos, A. (2019a). Caracterización del hablante con fines judiciales: fenómenos fónicos propios del habla espontánea. *e-AESLA. Revista digital de lingüística aplicada*, 5, 265–278. <https://cvc.cervantes.es/lengua/eaesla/pdf/05/26.pdf>
- Llisterri, J., Machuca, M. J. y Ríos, A. (2019b). VILE-P: un corpus para el estudio prosódico de la variación inter e intralocutor. En J. M. Lahoz-Bengoechea y R. Pérez Ramón (Eds.), *Subsidia: Tools and Resources for Speech Sciences / Subsidia: herramientas y recursos para las ciencias del habla* (pp. 117–123). Universidad de Málaga. <https://hdl.handle.net/10630/18177>
- McDougall, K. y Duckworth, M. (2017). Profiling fluency: An analysis of individual variation in disfluencies in adult males. *Speech Communication*, 95, 16–27. <https://doi.org/10.1016/j.specom.2017.10.001>
- San Segundo, E., Alves, H. y Fernández Trinidad, M. (2013). CIVIL Corpus: Voice quality for speaker forensic comparison. *Procedia – Social and Behavioral Sciences*, 95, 587–593. <https://doi.org/10.1016/j.sbspro.2013.10.686>

6. AGRADECIMIENTOS

El presente trabajo se ha realizado en el marco del proyecto ESTIVOZ, *Grado de espontaneidad y factores temporales en el disimulo de la voz* (FFI2014-59848-C2-2-P) financiado por el Ministerio de Economía y Competitividad del Gobierno de España.

LAS TÉCNICAS DE NEUROIMAGEN Y LA FONÉTICA EXPERIMENTAL

Victoria Marrero Aguiar

UNED
vmarrero@flog.uned.es

ABSTRACT

En esta presentación, que forma parte de la sesión monográfica “Nuevas metodologías para el estudio fonético: qué nos aportan y qué nos exigen”, se presentará el tipo de datos que nos ofrecen las principales técnicas de neuroimagen que se están utilizando para el estudio del habla y, en general, del nivel fónico del lenguaje. Nuestro foco de interés no es la descripción técnica de esas herramientas, sino las evidencias lingüísticas que proporcionan: cómo interpretarlas, qué esperar de ellas y en qué medida pueden enriquecer el estado actual de conocimientos en el área de la fonética experimental.

Palabras clave: Neuroimagen, EEG, EcoG, ERP, MEG, MRI, fMRI, PET, NIRS

In this presentation, which is part of the monographic session "New methodologies for phonetic study: what they provide us with and what they demand from us", the type of data offered by the main neuroimaging techniques that are being used for the study of speech and, in general, the phonic level of language, will be presented. Our focus is not on the technical description of these tools, but on the linguistic evidence they provide: how to interpret them, what to expect from them and to what extent they can enrich the current state of knowledge in the area of experimental phonetics.

Keywords: Neuroimaging, EEG, EcoG, ERP, MEG, MRI, fMRI, PET, NIRS

1. INTRODUCCIÓN

Las técnicas de neuroimagen utilizan un conjunto de instrumentos para obtener una representación gráfica de la estructura y el funcionamiento del cerebro en vivo.

Actualmente, esas imágenes se pueden generar estudiando dos procesos neuronales diferentes:

a) La actividad eléctrica: una medida directa del funcionamiento de las neuronas. Es la base de técnicas como la electroencefalografía (EEG), la electrocorticografía (ECoG) los potenciales evocados, o los potenciales relacionados con eventos (ERP); aunque no proporcionan imágenes sobre la estructura del cerebro, sí permiten conocer aspectos de su funcionamiento. También la magnetoencefalografía (MEG) toma como base el campo magnético asociado a la actividad eléctrica neuronal; permite identificar grandes áreas del cerebro activadas, y su resolución temporal es también muy buena.

b) El metabolismo: es una medida indirecta, se detecta la respuesta hemodinámica, las zonas de

mayor consumo metabólico; es la base de técnicas como la resonancia por emisión de positrones (PET), la resonancia magnética funcional (fMRI) o la espectroscopía del infrarrojo cercano (NIRS). Permiten conocer qué áreas del cerebro se han activado durante una determinada tarea. Todas ofrecen una buena resolución espacial, pero peor resolución temporal que las técnicas ligadas a la actividad eléctrica.

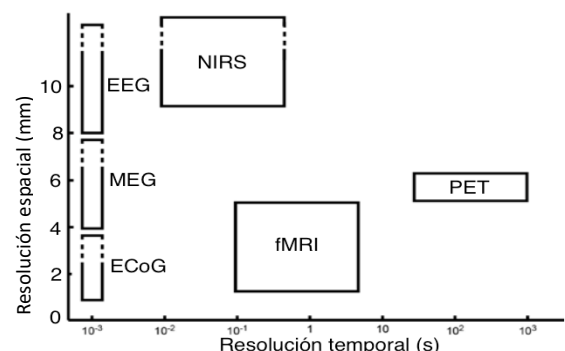


Figura 1: Resolución espacial y temporal de diferentes técnicas de neuroimagen. Adaptado de Lizarazu 2017, fig. 9.

2. OBJETIVO

En esta presentación, tras una breve descripción técnica de las metodologías mencionadas, se hará una sucinta revisión bibliográfica sobre la información que las principales técnicas de neuroimagen nos ofrecen en lo que concierne al estudio del habla y, en general, del nivel fónico del lenguaje. El objetivo es analizar en qué medida esos resultados contribuyen a mejorar la descripción del sistema de sonidos de las lenguas, por una parte, y como contrapartida, qué requisitos y bases de conocimiento son necesarios para que los datos se puedan integrar en el ámbito de la fonética experimental.

3. TÉCNICAS DE BASE ELECTROFISIOLÓGICA

3.1. EEG, EcoG y potenciales

Todas estas técnicas requieren colocar (con mayor o menor densidad por mm.) un conjunto de electrodos con el fin de captar el potencial postsináptico de grupos de neuronas, cuya pequeña potencia (entre 10 y 200 μ V) a veces consigue superar las barreras físicas que las separan del receptor (profundidad en la propia corteza, meninges, cráneo, cuero cabelludo...).

3.1.1. Electroencefalografía (EEG)

La primera de estas técnicas (fue descubierta por Hans Berger en la década de los 20 del siglo pasado), nos proporciona información en términos de bandas de frecuencia para todo el cerebro o, como mucho, por hemisferios. Especialmente relevantes para la decodificación del habla serían las bandas delta, theta y gamma, que se han relacionado con la segmentación temporal de la información prosódica, silábica y fonológica, respectivamente. La decodificación segmental y subsegmental se realizaría predominantemente en el hemisferio izquierdo, en la banda más rápida (*gamma*, por encima de 30 Hz \approx 33 ms. o menos); el hemisferio derecho, por su parte, extraería información en una ventana más amplia, adecuada para el nivel silábico (banda *theta*, 4-10 Hz \approx 100- 250 ms.) y prosódico (banda *delta*, 0.5-3.5 Hz \approx 290-2000 ms) (Ghitza y Greenberg 2009, Poeppel, Idsardi y Van Wassenhove 2008; Morillon et al. 2010; Ghitza 2011, Giraud y Poeppel 2012; Leong y Goswami 2014).

3.1.2. Potenciales relacionados con eventos (ERP)

Los ERP son variaciones en los potenciales de amplitud de la señal de EEG, comparada con la actividad de base de esta señal. Pueden ser

provocados en respuesta a un evento determinado, visual, auditivo o de cualquier otro tipo. En el nivel fónico es de sobra conocido el papel del *mismatch negativity* (MMN), un componente deflectivo que aparece entre 160-220 ms después de la presentación de un estímulo auditivo, y que se ha relacionado con la detección preatencional de cambios en *input* sonoro (una revisión en Beres 2017); solo en el presente año se han publicado varias decenas de artículos que utilizan este índice para estudiar elementos de naturaleza frecuencial (Gu et al. 2019, Chang, Lin y Kuo 2019, Shen y Froud 2019), de naturaleza temporal (Oron et al. 2019, Cheng et al. 2019), en ontogénesis (Lee 2019, Ferreira et al. 2019, Marklund, Schwarz y Lacerda 2019), o en la intersección entre percepción y producción, considerando el paso de lo variable –fonético– a lo invariable –fonológico– (Rodrigues Silva y Rothe-Neves 2019).

3.1.3. Electroencefalografía (EcoG)

Utiliza electrodos corticales, pero colocados directamente sobre la corteza cerebral, para lo cual es necesaria una intervención quirúrgica que hace este método muy invasivo, a diferencia de la EEG, y útil solo para personas cuya patología lo justifique. La información eléctrica neuronal se procesa mediante sistemas intracraneales de interfaz cerebro-ordenador (*Brain-Computer Interface, BCI systems*), con el fin de convertirla en habla (el primer caso fue presentado por Vansteensel et al. en 2016). Para ello, una hipótesis básica es que “la producción de fonemas implica la secuencia de patrones de actividad robustos y reproducibles en la superficie cortical”, muy dependientes de la decodificación temporal (Ramsey et al. 2018, pg. 301; Bouchard et al. 2013).

3.2. Magnetoencefalografía (MEG)

Toda corriente eléctrica genera un campo magnético a su alrededor. Mientras los métodos anteriores captan el potencial eléctrico neuronal, la MEG capta el campo magnético asociado, que también es mínimo (10^{-12} T aproximadamente, entre 0.1 y 1.0 picotesla, la billonésima parte del generado por un buen altavoz), por lo que es imprescindible un aislamiento del aparato que evite interferencias (para empezar, del propio magnetismo terrestre, que es de 10^{-5} T). Aunque la MEG no tiene tan buena resolución espacial como las técnicas de base hemodinámica, sus resultados se pueden representar anatómicamente, tras un proceso de reconstrucción de la fuente donde se originó la señal.

La localización anatómica del sistema auditivo en la corteza cerebral hace que la MEG, con su resolución temporal en ms., sea una técnica especialmente

adecuada para estudiar, como indican Lütkenhöner y Poeppel “cómo se transforman las señales acústicas para producir las representaciones auditivas que forman la base de la percepción del habla, la cognición musical y otros aspectos de la cognición auditiva.” (Lütkenhöner y Poeppel 2011, 605). En su revisión, estos autores proponen la siguiente taxonomía de áreas de estudio con MEG:

- (a) Claves elementales de la señal acústica (tono, sonía y timbre).
- (b) Estrategias básicas de procesamiento del habla: segmentación, integración, detección de cambios...
- (c) Procesamiento de sonidos, desde segmentos aislados hasta habla continuada.
- (d) Música.
- (e) Interacción e integración multisensorial y sensorio-motora.

4. TÉCNICAS DE BASE METABÓLICA

4.1. Tomografía por emisión de positrones (PET)

Requiere la ingesta, inhalación o inyección de un radiofármaco que actúa como “trazador”; se trata, por tanto, de una técnica algo invasiva. El tomógrafo detecta los fotones que se producen cuando se aniquila un positrón del radiofármaco y un electrón de la corteza cerebral. Su ventaja, frente a técnicas como la fMRI, es que no genera ruido (por tanto, puede usarse para pruebas de percepción auditiva) y no es incompatible con elementos metálicos como los que se usan en implantes cocleares.

En lo que respecta a sus aplicaciones lingüísticas o cognitivas, “los estudios de la función auditiva con PET pueden dividirse entre los que exploran la representación, dentro de la corteza auditiva, de la tonotopía y otros parámetros de estímulo como la sonía, el ancho de banda, la amplitud o la tasa de modulación de la frecuencia, y los estudios que tratan de caracterizar los circuitos neuronales a una escala más amplia, como los de la localización espacial auditiva, el procesamiento de la música o el habla y la modulación de la atención” (Johnsrude, Giraud y Frackowiak, 2002, 255).

4.2. Imagen por resonancia magnética funcional (fMRI)

Esta es, probablemente, la técnica estrella en neuroimagen cognitiva. Consiste en generar, mediante grandes imanes conectados a una elevada corriente eléctrica, un potentísimo campo magnético (1,5, 3 o 7 Teslas), que provoca la polarización de los átomos de hidrógeno del cuerpo (y del cerebro, cuyo estudio conlleva el uso de un escáner en forma de casco). Hay imágenes de resonancia magnética estáticas, en las que el sujeto no realiza ninguna tarea, pero nos interesan más las

funcionales, que miden las diferencias magnéticas entre la sangre oxigenada empleada por las neuronas activas y la desoxigenada de las inactivas (Glover 2012).

Hay tanta bibliografía sobre fMRI y lenguaje que la publicada en un único año dio lugar al artículo de revisión de Price (2010) con 135 referencias. En la imagen inferior se presentan los datos relacionados únicamente con el nivel fónico.

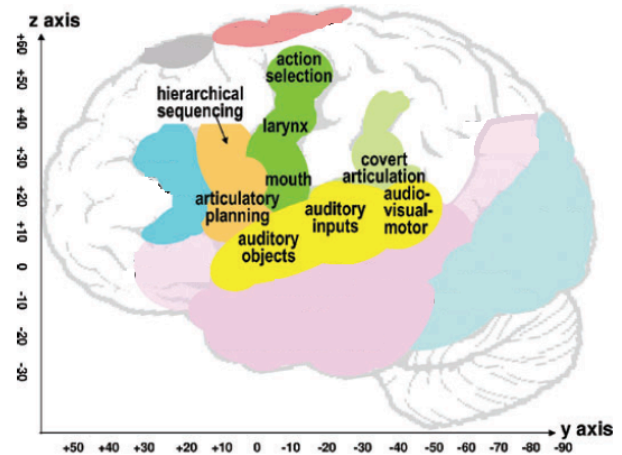


Figura 2: Resumen de las funciones atribuidas a las áreas del lenguaje relacionadas con el nivel fónico. Adaptado de Price 2010, pg. 76.

4.2.1. Imágenes de resonancia magnética por difusión (dMRI); imágenes con tensor de difusión (DTI) – tractografía

La particularidad de estas técnicas es que permiten visibilizar la sustancia blanca del cerebro, la que compone las fibras mielinizadas de los axones que conectan las neuronas; por lo tanto, proporciona representaciones de las redes, de las conexiones neuronales. Tanto la dMRI como la DTI o tractografía son sensibles a la difusión de los protones del agua en el cuerpo; como en los axones esa difusión no es libre, sino que se ve constreñida por la vaina de mielina, el escáner capta las diferencias en la dirección de los protones y, tras una compleja computación, se representan como líneas de colores.

Ambos métodos se han utilizado para investigar las conexiones entre las áreas implicadas en los procesos de articulación y los de percepción (Ocklenburg, Hugdahl y Westerhausen 2013, Powell et al. 2006) o la existencia de una doble vía para el procesamiento auditivo, una dorsal, implicada en la relación entre el sonido y los movimientos articulatorios, y otra ventral, que relacionaría secuencias sonoras con significados (Saur et al. 2008; Ford et al. 2013, Ries et al. 2019); para el aprendizaje de sonidos en una L2 (Golestani et al. 2007), y en múltiples aplicaciones clínicas.

4.2.3. Métodos ópticos: espectroscopía funcional del infrarrojo cercano (fNIRS) y tomografía óptica difusa de alta densidad (HD-DOT)

Los métodos ópticos de neuroimagen miden las variaciones en la absorción o reflexión de rayos de luz infrarroja en función del nivel de oxígeno en la sangre. A pesar de que su resolución espacial es peor que la de otras técnicas de base hemodinámica, y la temporal es peor que la de los electrofisiológicos, son técnicas no invasivas con un equipamiento relativamente portátil, permite movimientos e interacción social y es bastante silenciosa, lo cual les ha hecho ganar terreno en los estudios cognitivos.

Según Pelle, en su revisión de 2017, las técnicas con base óptica se han utilizado especialmente en tres áreas:

- (a) Desarrollo del lenguaje en bebés y niños.
- (b) Procesamiento del habla en poblaciones excluidas de la resonancia (implantados cocleares, por ejemplo).
- (c) Comprensión del habla sin interferencia de ruido.

También se han utilizado para la discriminación de fonemas en L1 y L2 (Dubins 2009, Minagawa-Kawai, Mori & Sato 2005), entre otros temas.

5. CONCLUSIONES

El conjunto de técnicas descritas, y las que se puedan sumar a esta sucinta exposición, nos permiten obtener una gran cantidad de datos de gran interés para un lingüista interesado en el estudio del nivel fónico: desde la realidad cognitiva de los rasgos distintivos o el fonema, al procesamiento y la producción y las relaciones entre producción y percepción de segmentos o suprasegmentos. Acceder a esos datos, entender su alcance e integrarlos en la descripción y análisis de la fonética y fonología de una lengua, o del lenguaje como capacidad general, sin duda enriquecerá este ámbito de estudio.

Nos planteábamos en la introducción qué exigencias se requieren para acceder a estas fuentes de información. Evidentemente, si nos planteamos como meta el realizar todo el diseño experimental y llevarlo a cabo, es necesaria una formación especializada, con distintos niveles de complejidad entre los diferentes métodos. Pero acceder a los datos generados por una bibliografía cada vez más numerosa y específica está sin duda al alcance de cualquier fonetista interesado.

Por otra parte, también la lingüística tendría elementos que aportar al ámbito de los estudios neurocientíficos del lenguaje, como un mejor control de las variables, considerando elementos que a

menudo no parecen tenerse en cuenta: diferencias de complejidad silábica entre los estímulos, de posición en la sílaba, de acento léxico o prosódico, de frecuencia de uso, diferencias geosociales, etc.

En definitiva, incrementar la interdisciplinariedad y el conocimiento mutuo entre la fonética experimental y la neurociencia del lenguaje probablemente es una magnífica vía de crecimiento para ambas disciplinas.

6. BIBLIOGRAFÍA

- Beres, A. M. (2017). Time is of the essence: A review of electroencephalography (EEG) and event-related brain potentials (ERPs) in language research. *Applied psychophysiology and biofeedback*, 42(4), 247-255.
- Bouchard KE, Mesgarani N, Johnson K, Chang EF. (2013) Functional organization of human sensorimotor cortex for speech articulation. *Nature*. 2013; 495:327-332.
- Chang, C. C., Lin, T. H., & Kuo, W. J. (2019). Does phonological rule of tone substitution modulate mismatch negativity?. *Journal of Neurolinguistics*, 51, 63-75.
- Cheng, B., Zhang, X., Fan, S., & Zhang, Y. (2019). The role of temporal acoustic exaggeration in high variability phonetic training: A behavioral and ERP study. *Frontiers in psychology*, 10, 1178.
- Dubins, M. (2009). *Phonetic Discrimination in the First and Second Half-year of Life: An Investigation of Monolingual and Bilingual Infants using Event-Related Functional Near-infrared Spectroscopy (fNIRS)* Tesis doctoral. Universidad de Toronto, Canadá.
- Ferreira, D. A., Bueno, C. D., de Costa, S. S., & Sleifer, P. (2019). Mismatch Negativity in Children: Reference Values. *International archives of otorhinolaryngology*, 23(02), 142-146
- Ford, A., et al. (2013). Broca's area and its striatal and thalamic connections: a diffusion-MRI tractography study. *Frontiers in neuroanatomy*, 7, 8.
- Ghitza, O. (2011). "Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm", *Frontiers in Psychology*, 2, p. 130
- Ghitza, O., & Greenberg, S. (2009). On the possible role of brain rhythms in speech perception: intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica*, 66(1-2), 113-126.
- Giraud, A. L. y Poeppel, D. (2012). "Cortical oscillations and speech processing: emerging computational principles and operations", *Nature Neuroscience*, 15, 511-517
- Glover G. H. (2011). Overview of functional magnetic resonance imaging. *Neurosurgery clinics of North America*, 22(2), 133-vii.
- Golestani, N., Molko, N., Dehaene, S., LeBihan, D., & Pallier, C. (2007). Brain structure predicts the learning of foreign speech sounds. *Cerebral cortex*, 17(3), 575-582.
- Gu, F., Wong, L., Hu, A., Zhang, X., & Tong, X. (2019). A Lateral Inhibition Mechanism Explains the Dissociation between Mismatch Negativity and

- Behavioral Pitch Discrimination. *Brain Research*, 146308
- Johnsrude, I. S., Giraud, A. L., & Frackowiak, R. S. (2002). Functional imaging of the auditory system: the use of positron emission tomography. *Audiology and Neurotology*, 7(5), 251-276.
- Lee, C. Y. (2019). Neural Underpinnings of Early Speech Perception and Emergent Literacy. *Folia Phoniatrica et Logopaedica*, 1-10
- Leong, V. y Goswami, U. (2014). "Assessment of rhythmic entrainment at multiple timescales in dyslexia: Evidence for disruption to syllable timing", *Hearing Research*, 308, pp. 141–161
- Lizarazu, M. (2017). *Speech-brain synchronization: a possible cause for developmental dyslexia*. Tesis doctoral. Consultado el 20/01/20 en https://www.researchgate.net/publication/315524493_Lizarazu_PhD_Thesis_Speech-brain_synchronization_a_possible_cause_for_developmental_dyslexia
- Lütkenhöner, B., y Poeppel, D. (2009). From tones to speech: magnetoencephalography studies. En *The Auditory Cortex*, eds J. A. Winer and C. E. Schreiner (New York, NY: Springer), 407–428.
- Marklund, E., Schwarz, I. C., & Lacerda, F. (2019). Amount of speech exposure predicts vowel perception in four-to eight-month-olds. *Developmental cognitive neuroscience*, 36, 100622.
- Minagawa-Kawai, Y., Mori, K., & Sato, Y. (2005). Different brain strategies underlie the categorical perception of foreign and native phonemes. *Journal of cognitive neuroscience*, 17(9), 1376-1385.
- Morillon et. al (2010). Neurophysiological origin of human brain asymmetry for speech and language. *PNAS*, 107, pp. 18688–18693.
- Ocklenburg, S., Hugdahl, K., & Westerhausen, R. (2013). Structural white matter asymmetries in relation to functional asymmetries during speech perception and production. *Neuroimage*, 83, 1088-1097.
- Oron, A., Szelag, E., Nowak, K., Dacewicz, A., & Szymaszek, A. (2019). Age-related differences in Voice-Onset-Time in Polish language users: An ERP study. *Acta psychologica*, 193, 18-29.
- Peelle, J. E. (2017). Optical neuroimaging of spoken language. *Language, cognition and neuroscience*, 32(7), 847-854.
- Poeppel, D., Idsardi, W. J., & Van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1493), 1071-1086.
- Powell et al. (2006). Hemispheric asymmetries in language-related pathways: a combined functional MRI and tractography study. *Neuroimage*, 32(1), 388-399.
- Price, C. J. (2010). The anatomy of language: a review of 100 fMRI studies published in 2009. *Annals of the new York Academy of Sciences*, 1191(1), 62-88.
- Ramsey, N. F., Salari, E., Aarnoutse, E. J., Vansteensel, M. J., Bleichner, M. G., & Freudenburg, Z. V. (2018). Decoding spoken phonemes from sensorimotor cortex with high-density ECoG grids. *Neuroimage*, 180, 301-311.
- Ries, S. K., et al. (2019). Roles of ventral versus dorsal pathways in language production: An awake language mapping study. *Brain and language*, 191, 17-27.
- Rodrigues Silva, D. M., & Rothe-Neves, R. (2019). Context-dependent categorisation of vowels: a mismatch negativity study of positional neutralisation. *Language, Cognition and Neuroscience*, 1-16
- Saur, D., et al. (2008). Ventral and dorsal pathways for language. *Proceedings of the national academy of Sciences*, 105(46), 18035-18040.
- Shen, G., & Froud, K. (2019). Electrophysiological correlates of categorical perception of lexical tones by English learners of Mandarin Chinese: an ERP study. *Bilingualism: Language and Cognition*, 22(2), 253-265.
- Vansteensel MJ, et al. (2016) Fully Implanted Brain-Computer Interface in a Locked-In Patient with ALS. *N Engl J Med*. 375:2060–2066.

DESEMPEÑO FONÉTICO-ACÚSTICO DE OCLUSIVAS ÁFONAS EN HABLANTES ADULTOS MAYORES DEL ESPAÑOL CHILENO, DE LA CIUDAD DE CHILLÁN, CHILE

Renato Martínez-Cifuentes¹, Jaime Soto-Barba¹, Mauricio Figueroa-Candia¹, Katia Sáez-Carrillo²

¹Departamento de Español, Universidad de Concepción, Chile

²Departamento de Estadística, Universidad de Concepción, Chile
remartínez@udec.cl

ABSTRACT

Los estudios fonético-acústicos de /p t̥ k/ en Chile principalmente se han efectuado con menores de 50 años. La única investigación que se realiza con mayores de 55 años analiza solo el V.O.T. OBJETIVO. Describir el desempeño fonético-acústico de oclusivas áfonas en hablantes adultos mayores del español chileno, de la ciudad de Chillán, Chile. MATERIALES Y MÉTODOS. Estudio cuantitativo y descriptivo. 16 hablantes con un promedio de 70,6 años (D.E. = 8,26) producen 15 enunciados que contienen /p t̥ k/ antecedendo a cada una de las cinco vocales. RESULTADOS. La fracción áfona media de las oclusivas es mayor al 91% y la intensidad relativa media es mayor a 40,92 dB. Hay diferencias estadísticamente significativas en la duración absoluta de /k/ cuando está antes de /i/. CONCLUSIONES. Se reportan características acústicas de las oclusivas áfonas en un rango etario que no se ha estudiado en Chile.

Palabras clave: oclusivas áfonas, intensidad, duración, español chileno

Acoustic-phonetic research on /p t̥ k/ in Chile have been chiefly conducted in a population of under 50. The only research conducted in a population of over 55 is just VOT analysis. OBJECTIVE. To assess the acoustic-phonetic performance on voiceless occlusives of Chilean Spanish-speaking senior citizens living in Chillán, Chile. MATERIALS AND METHODS. Quantitative and descriptive study. 16 speakers with an average age of 70.6 (SD = 8.26) produce 15 statements including /p t̥ k/ preceding each of the five Spanish vowels. RESULTS. The mean voiceless fraction is over 91% and the mean relative intensity is over 40.92 dB. There are statistically significant differences in the absolute duration of /k/ when followed by /i/. CONCLUSIONS. Acoustic features of voiceless occlusives are described in an age range that has not been studied in Chile.

Keywords: voiceless occlusives, intensity, duration, Chilean Spanish

1. INTRODUCCIÓN

Las oclusivas en el español de Chile, al igual que en el español peninsular, son: bilabial oclusivo sordo /p/, velar oclusivo sordo /k/, bilabial oclusivo sonoro /b/, y velar oclusivo sonoro /g/ (Hidalgo & Quilis, 2012; Hualde & Colina, 2014). Sin embargo, a distinción de España, el español chileno presenta la oclusiva postdental áfona /t̥/ y la oclusiva postdental sonora /d̥/ (Sadowsky & Salamanca, 2011).

Específicamente en la serie áfona, los alófonos de /p t̥ k/ que se producen con el modo de articulación oclusivo en el español de Chile son los que se ubican en todos los contextos, excluyendo los de coda silábica, donde se manifiesta el alófono

oclusivo no liberado [p̚], [t̥̚] y [k̚] (Sadowsky & Salamanca, 2011).

Desde un punto de vista acústico, las oclusivas áfonas se caracterizan por una zona de silencio y una barra de explosión en el momento de su resolución. Esta barra corresponde a una zona de ruido muy breve a lo largo de las frecuencias en el espectrograma (Hidalgo & Quilis, 2012; Martínez, 2003; Martínez & Fernández, 2013). La barra de explosión de las oclusivas se describe como “una línea vertical seguida de un corto intervalo de energía aperiódica (sin un patrón reiterado definido) o ruido, antes de comenzar los formantes de la vocal siguiente” (Hualde & Colina, 2014, p. 47). Puntualmente, [p] presenta una “barra de explosión muy breve, con un punto máximo de energía en bajas

frecuencias” (Hidalgo & Quilis, 2012, p. 185), [t] evidencia una “barra de explosión muy visible, con su punto máximo de energía alto”, y [k] tiene una “explosión bastante amplia, el punto de máxima energía es intermedio” (Hidalgo & Quilis, 2012, p. 186).

Asimismo, las oclusivas son caracterizadas en base al V.O.T. (voice onset time). Este índice acústico se refiere al “tiempo que se encuentra entre la relajación de la oclusión (o momento de la explosión) y el inicio de la vibración laríngea por la articulación siguiente” (Hidalgo & Quilis, 2012, p. 59). En el caso de las oclusivas áfonas, “el valor temporal del VOT es positivo” (Hidalgo & Quilis, 2012, p. 59).

En Chile, se cuenta con varios estudios fonético-acústicos que analizan la producción de oclusivas, áfonas y sonoras, en hablantes normotípicos del español chileno. Las investigaciones se orientan principalmente a la duración (Soto-Barba, 1994; Soto-Barba & Valdivieso, 1999), V.O.T. (Camacho et al., 2018; Roldán & Soto-Barba, 1997; Soto-Barba & Valdivieso, 1999), intensidad de la onda periódica (Soto-Barba & Valdivieso, 1999), velocidad de las transiciones vocálicas (Soto-Barba & Valdivieso, 1999), variantes aproximantes (Pérez, 2007) y lenición (Figueroa, 2016; Rogers, 2016; Rogers & Mirisis, 2018). De estas, la investigación de Camacho et al. (2018) es la única que se ha enfocado en la producción de hablantes urbanos mayores de 55 años (específicamente en el V.O.T.), y el reporte de Roger & Mirisis (2018) es el estudio acústico más reciente de las oclusivas áfonas con hablantes urbanos de entre 18 y 49 años.

A pesar de la tradición en estudios fonético-acústicos de oclusivas en hablantes normotípicos del español de Chile, estas investigaciones reportan parcialmente las características acústicas de las oclusivas áfonas en hablantes mayores de 55 años. Por ende, este estudio se propone describir el desempeño fonético-acústico de oclusivas áfonas en hablantes adultos mayores del español chileno, de la ciudad de Chillán, Chile.

2. MATERIALES Y MÉTODOS

El estudio se enmarca en un enfoque cuantitativo, un diseño cuasi-experimental y un alcance descriptivo. La muestra es de tipo no probabilística. Los sujetos corresponden a 16 personas, siete hombres y nueve mujeres, con una edad promedio de 70,6 años (D.E. = 8,26). Estas provienen de los participantes del Proyecto FONDECYT 1150336 “Representaciones espaciales en la comprensión del lenguaje en pacientes con

enfermedad de Parkinson”. Todos cuentan con una escolaridad básica completa, son hablantes del español de Chile como lengua materna y residen en zonas urbanas de las comunas de Chillán y Chillán Viejo.

El instrumento de elicitación de datos consiste en una lista de 15 enunciados que contienen las tres oclusivas áfonas y las cinco vocales del español de Chile. Cada enunciado se compone de dos palabras, tres sílabas y seis segmentos. Las oclusivas /p t k/ se encuentran en el ataque de la segunda sílaba, precedidas de la consonante nasal /n/ y seguidas de cada una de las vocales /i e a o u/. De esta forma, la oclusiva se sitúa en el tercer segmento de la frase, en una sílaba acentuada de estructura abierta o libre.

El procedimiento de recolección de datos, que consiste en la lectura de las frases, se efectúa luego del proceso de consentimiento informado. La grabación de la producción del hablante se realiza en una sala acústicamente aislada y con reverberación atenuada del laboratorio de Fonética de la Universidad de Concepción. Se utiliza una grabadora Tascam DR-40, conectada a un micrófono de cabeza Sennheiser EW-152-G3. La frecuencia de muestreo (sampling frequency) es de 44.1 kHz, la profundidad (bit rate) es de 24 bits, la grabación es monofónica y el formato del archivo es WAV.

La segmentación, el etiquetado y el análisis acústico de la señal se realiza con el software Praat versión 6.0.56 (Boersma & Weenink, 2019). La anotación de la señal se efectúa en un TextGrid, cuyos estratos se componen de intervalos. En el primer estrato se identifica el inicio y el término de cada frase elicitada por el hablante. En el segundo, se identifican dos segmentos en él: la oclusiva y la vocal que la sucede. Y, en el tercero, se identifican el punto de cierre de la oclusiva y su resolución.

El análisis acústico de las oclusivas se realiza en base a los índices acústicos: fracción áfona, intensidad relativa y duración absoluta. La fracción áfona del segmento se mide mediante la función Voice report de Praat. Esta función entrega el Fraction of locally unvoiced frames del intervalo seleccionado. La intensidad relativa se mide en el objeto Intensity. En este objeto se identifica el valle del segmento oclusivo estudiado y la cima de la vocal que lo sucede. Luego, se calcula la diferencia entre la cima y el valle, en el marco del procedimiento de normalización Diferencia de intensidad A (Figueroa, 2016). La duración absoluta se mide desde el punto de cierre de la oclusiva hasta su resolución (Rogers & Mirisis, 2018).

El análisis estadístico considera estadígrafos descriptivos, como medidas de tendencia central y de dispersión, así como estadística inferencial. Se contrasta la fracción áfona, la intensidad relativa y la

duración absoluta de las oclusivas áfonas fonológicas, y se identifican las interacciones de estos contrastes con la vocal fonológica que sucede a la oclusiva. Se utiliza para los análisis estadísticos el software R base versión 3.6.2 (The R Foundation for Statistical Computing, 2019).

3. RESULTADOS

Al analizar la fracción áfona (F.A.) de /p t̪ k/ los resultados muestran que, en conjunto, el promedio es de 0,92 (D.E. = 0,08). Además, se evidencia que la F.A. no presenta interacción estadística entre los factores: oclusiva y vocal (p-valor > 0,05). La media de la fracción áfona de /p/ es 0,92 (D.E. = 0,07); de /t̪/ es 0,92 (D.E. = 0,07); y de /k/ es 0,91 (D.E. = 0,09). Al contrastar la F.A. entre las oclusivas, se observa que no existen diferencias estadísticamente significativas entre ellas (p-valor > 0,05).

La intensidad relativa (I.R.) de las oclusivas áfonas se manifiesta con un promedio global de 41,52 dB. (D.E. = 5,48). Se evidencia que la I.R. no presenta interacción estadística entre los factores: oclusiva y vocal (p-valor > 0,05). La media de la intensidad relativa de /p/ es 41,61 dB. (D.E. = 5,41); de /t̪/ es 42,02 dB. (D.E. = 5,36); y de /k/ es 40,92 dB. (D.E. = 5,66). Al comparar la I.R. entre las oclusivas, se observa que entre ellas no existen diferencias estadísticamente significativas (p-valor > 0,05).

La duración absoluta (D.A.) de /p t̪ k/ se expresa con un promedio general de 0,13 s. (D.E. = 0,04). Se evidencia que la D.A. presenta interacción estadística entre los factores: oclusiva y vocal (p-valor < 0,05). La media de la duración absoluta de las oclusivas áfonas por vocal se presenta en la tabla 1. Al contrastar la D.A. de las oclusivas por vocal, se observa que existen diferencias estadísticamente significativas entre /p t̪ k/ cuando preceden a /i/ (p-valor < 0,05), tal como se muestra en la figura 1.

Tabla 1: Promedios y desviaciones estándar de la duración absoluta de las oclusivas áfonas por vocal

Oclusiva por vocal	Duración absoluta (s.)
/a/	
/p/	0,13 (0,04)
/t̪/	0,12 (0,03)
/k/	0,12 (0,03)
/e/	
/p/	0,14 (0,04)
/t̪/	0,13 (0,03)
/k/	0,12 (0,03)
/i/	
/p/	0,15 (0,05)
/t̪/	0,14 (0,04)

/k/	0,11 (0,03)
/o/	
/p/	0,13 (0,05)
/t̪/	0,14 (0,04)
/k/	0,13 (0,06)
/u/	
/p/	0,12 (0,03)
/t̪/	0,14 (0,05)
/k/	0,13 (0,04)

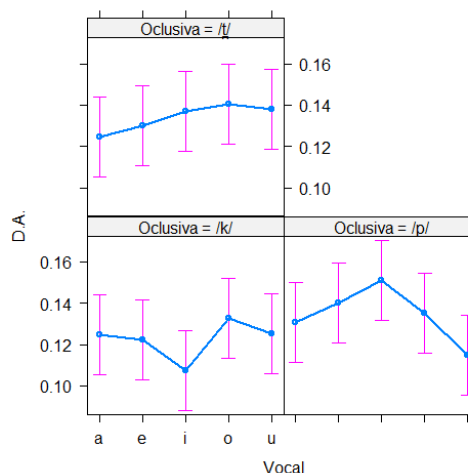


Figura 1: Duración absoluta de las oclusivas áfonas por vocal.

4. CONCLUSIONES

Frente a la pregunta de investigación ¿cómo son las oclusivas áfonas desde un punto de vista fonético-acústico en hablantes normotípicos del español chileno de la ciudad de Chillán, Chile?, este estudio evidencia que /p t̪ k/ tienen, en promedio, una fracción áfona mayor al 91% y una intensidad relativa media mayor al 40,92 dB. Estos índices tienden a ser más altos para /p/ y más bajos para /k/, sin embargo, estas diferencias no son estadísticamente significativas. Respecto de la duración absoluta de /p t̪ k/, este estudio provee resultados por cada una de las vocales vecinas, debido a la interacción estadística que se presenta entre los factores oclusiva y vocal. La duración absoluta de /k/, respecto de /p/ y /t̪/, tiende a ser baja cuando antecede a las vocales /a e i o/, no obstante, estas diferencias solo tienen significancia estadística cuando precede a la /i/.

Los aportes de este estudio se plantean en dos ámbitos. El primero, la presentación de evidencia sobre las características fonético-acústicas de las oclusivas áfonas en un rango etario que no se ha estudiado en Chile. Y el segundo, sobre el tratamiento estadístico de los datos. En este estudio se ha cautelado analizar las interacciones estadísticas de las oclusivas con las vocales adyacentes, antes de comprar los índices acústicos de las producciones de /p t̪ k/.

Finalmente, respecto de las proyecciones de esta investigación, se advierte la necesidad de nuevos estudios que describan acústicamente la serie sonora de las oclusivas en hablantes adultos mayores normotípicos del español chileno. De igual manera, se propone que los estudios acústicos de las oclusivas en adultos mayores: 1) amplíen los instrumentos y procedimientos de elicitación de datos, 2) incorporen factores sociofonéticos (geográficos y socioeconómicos), y 3) consideren hablantes con trastornos adquiridos del habla de origen neurológico.

5. CONFLICTO DE INTERESES Y FINANCIACIÓN

Conflicto de intereses: los autores declaran no tener ningún conflicto de intereses.

Financiación: la presente investigación ha recibido ayuda de la Agencia Nacional de Investigación y Desarrollo (A.N.I.D.) del Gobierno de Chile mediante la beca de doctorado nacional, y del Proyecto UCO 1866 “Plan de Fortalecimiento de la Internacionalización del Postgrado y la Investigación de la Universidad de Concepción”.

6. BIBLIOGRAFÍA

- Boersma, P., & Weenink, D. (2019). *Praat: Doing phonetics by computer* (Versión 6.0.56) [Computer software]. <http://www.praat.org/>
- Camacho, C., Véliz, M., & Soto-Barba, J. (2018). *Comportamiento del índice acústico voice onset time (VOT) en la serie [p-t-k] y [b-d-g] en personas con Parkinson vs personas sin esta enfermedad diagnosticada* [Tesis magíster, Universidad de Concepción]. repositorio.udec.cl/bitstream/handle/11594/3253/Tesis_Comportamiento_del_indice_acustico.Image.Marked.pdf
- Figuroa, M. (2016). *Lenition in the production and perception of Chilean Spanish approximant consonants: Implications for lexical access models—UCL Discovery* [Tesis doctorado, University College London]. https://discovery.ucl.ac.uk/id/eprint/1503903/1/phd_thesis_MFiguroa_2016.pdf
- Hidalgo, A., & Quilis, M. (2012). *La voz del lenguaje: Fonética y fonología del español*. Tirant Humanidades.
- Hualde, J., & Colina, S. (2014). *Los sonidos del español*. Cambridge University Press.
- Martínez, E. (2003). *El sonido en la comunicación humana: Introducción a la fonética* (2.^a ed.). Octaedro.
- Martínez, E., & Fernández, A. (2013). *Manual de fonética española: Articulaciones y sonidos del español* (2.^a ed.). Ariel.
- Pérez, H. (2007). Estudio de la variación estilística de la serie /b-d-g/ en posición intervocálica en el habla de los noticieros de la televisión chilena. *Estudios de fonética experimental*, 16, 228-259-259.
- Rogers, B. (2016). The influence of linguistic and social variables in the spirantization of intervocalic /b,d,g/ in Concepción, Chile. *Studies in Hispanic and Lusophone Linguistics*, 9(1), 207-237. <https://doi.org/10.1515/shll-2016-0008>
- Rogers, B., & Mirisis, C. (2018). Voiceless stop lenition and reduction as linguistic and social phenomena in Concepción, Chile. *Borealis – An International Journal of Hispanic Linguistics*, 7(2), 187-215. <https://doi.org/10.7557/1.7.2.4401>
- Roldán, Y., & Soto-Barba, J. (1997). El V.O.T. de /p-t-k/ y /b-d-g/ en el español de Valdivia: Un análisis acústico. *Estudios filológicos*, 32, 27-33. <https://doi.org/10.4067/S0071-17131997003200003>
- Sadowsky, S., & Salamanca, G. (2011). El inventario fonético del español de Chile: Principios orientadores, inventario provisorio de consonantes y sistema de representación (AFI-CL). *Onomázein*, 24(2), 61-84.
- Soto-Barba, J. (1994). ¿Los fonemas /b/ y /p/ se diferencian por la sonoridad? *Estudios filológicos*, 29, 33-38.
- Soto-Barba, J., & Valdivieso, H. (1999). Caracterización fonético-acústica de la serie de consonantes /p-t-k/ vs /b-d-g/. *Onomázein*, 4, 125-133.
- The R Foundation for Statistical Computing. (2019). *R* (Versión 3.6.2) [Computer software]. <https://www.r-project.org/>

LEXICAL EFFECTS IN THE USE OF WORD-LEVEL STRESS FOR WORD RECOGNITION BY ENGLISH AND KOREAN LEARNERS OF SPANISH

María Teresa Martínez-García¹, Dong-Jin Shin², Goun Lee³

¹Hankuk University of Foreign Studies, ²Jeonju University, ³Sungkyunkwan University
¹mtmg87@gmail.com, ²djshin@jj.ac.kr, ³cconni80@gmail.com

ABSTRACT

The reliance on suprasegmental information for word recognition varies among languages: While Spanish uses only suprasegmental cues to word-level stress (e.g., *Papa* ‘potato’ vs. *paPÁ* ‘father’), English uses both segmental and suprasegmental cues (e.g., *REcord* vs. *reCORD*), and Korean does not have lexical stress. Unclear is whether L2 learners of Spanish can make use of suprasegmental cues to recognize stress and how lexical information influences word recognition. Advanced English and Korean L2 learners of Spanish completed a cross-modal word-identification task. They heard semantically ambiguous sentences that ended with incomplete fragments, and they were asked to choose the word corresponding to the fragment heard. Results showed that all learners can make use of stress as a cue for word recognition, but Korean learners seem to be more influenced by their L1 prosodic structure. While other analyses are undergoing, L2 proficiency has shown to lead to more native-like results.

Keywords: second language acquisition, Spanish, suprasegmental information, word-level stress, word recognition

1. INTRODUCTION

The reliance on suprasegmental information for recognizing words in speech varies depending on the listeners’ language background. For example, since Spanish has word-level stress and does not have vowel reduction, suprasegmental information is used to distinguish among competing lexical words, in both minimal pairs (e.g., *Papa* ‘potato’ vs. *paPÁ* ‘father’), and temporarily overlapping pairs (e.g., *peLOta* ‘ball’ vs. *peLoTÓN* ‘platoon’) (e.g., Soto-Faraco, Sebastián-Gallés, & Cutler, 2001). Like in the case of Spanish, English has word-level stress, but it also has vowel reduction, with most minimal pairs not being segmentally ambiguous (e.g., *REcord* vs. *reCORD*) and with temporarily overlapping, segmentally similar pairs being less common (e.g., *MYStery* vs. *misTAKE*). Accordingly, while native speakers of Spanish heavily rely on suprasegmental cues for word recognition, native English listeners make limited use of these cues for recognizing English words (e.g., Cooper, Cutler, & Wales, 2002; Tremblay, 2008). Interestingly, Korean is a language without word-level stress. However, previous findings on their perception of English word-level stress indicated that Korean listeners pay more

attention to vowel reduction (not existing in Spanish) than suprasegmental cues (e.g., Lee, 2015).

The current study aims to investigate whether English and Korean listeners at advanced proficiency level in L2 Spanish would make use of stress for word recognition in Spanish, and whether they will make similar use of suprasegmental cues to stress as native Spanish listeners. Furthermore, the study also aimed to see whether the presence of stress (a ‘positive’ cue) and the absence of stress (a ‘negative’ cue) similarly constrain word recognition in all groups (Altenberg, 2005) and how the effect may be modified by lexical information (such as word frequency or word familiarity).

2. METHODOLOGY

2.1. Participants

Thirty-two ‘late’ English-speaking L2 learners of Spanish and 32 ‘late’ Korean-speaking “L2” learners of Spanish (most of which self-reported having an advanced level of proficiency in English) participated in the study. Both groups of learners had an intermediate-to-advanced proficiency in Spanish and were matched in most individual variables (such as age of acquisition or years of instructions).

Finally, 32 native Spanish speakers (bilingual in Valencian and Spanish) were included as a control group.

2.2. Procedure

The experiment was administered using Paradigm (Perception Research Systems, Inc., Tagliaferri, 2005). All the groups completed a cross-modal word-identification task (adapted from Cooper, Cutler, & Wales (2002); Soto-Faraco, Sebastián-Gallés, & Cutler (2001); Tremblay (2008)). Participants heard non-constraining auditory sentences that ended with two-syllable word fragments (e.g., *Elena dijo peLO...* ‘Elena said peLO...’) presented in two possible conditions. In the stressed condition, the fragments were stressed on the penultimate syllable (e.g., *peLO-*); in the unstressed condition, the fragments were unstressed (e.g., *pelo-*). Participants were asked to choose the word corresponding to the fragment on the screen (e.g., “*peLOta*” ‘ball’ vs. “*peloTON*” ‘squad’). All fragments belonged to words that follow regular stress patterns in Spanish and no visual information regarding the stress placement was used, other than the regular diacritics.

After the main test, participants also completed a background questionnaire, a word-familiarity test, and a proficiency test.

2.3. Stimuli

Twenty-four experimental stimuli were created and distributed into four lists in a counterbalanced way. All of the experimental items were trisyllabic words with ‘regular’ stress placement (Harris, 1967). Words with penultimate stress (stressed fragment) and final stress (unstressed fragments) were matched in terms of lexical frequency. An example of the two conditions can be seen in Table 1.

Table 1: Example of an experimental trial in the two stress conditions (underlined is the correct response).

AUDIO	WORD CHOICE	
Stressed Fragment		
▶ peLO-	<u>pelota</u>	pelotón
Unstressed Fragment		
▶ pelo-	pelota	<u>pelotón</u>

The experimental task also included 36 filler items, in which the difference between the two potential target words was segmental rather than suprasegmental (e.g., *balido* ‘bleat’ vs. *batido* ‘shake’).

2.4. Data Analysis

Participants’ accuracy was analyzed with logit mixed-effects models (cf. Baayen, 2008). A first model on participants’ results included the following fixed variables: group (Spanish vs. English L2 Learners vs. Korean L2 learners; baseline=Spanish), type of fragment (stressed vs. unstressed; baseline=stressed), and the interaction between the two variables. Subjects and trial were used as a random effect. As a post-hoc analyses, three additional models on the learner group’s accuracy results were conducted. In these analyses, either the proficiency scores, the log-transformed frequency of the competitor word, or the log-transformed word familiarity ratings of the L2 learners were entered as independent variables. Both models included participant and test item as crossed random variables.

2.5. Results

Figure 1 presents the percentage of correct responses of the three groups in each of the two fragment type conditions (unstressed and stressed).

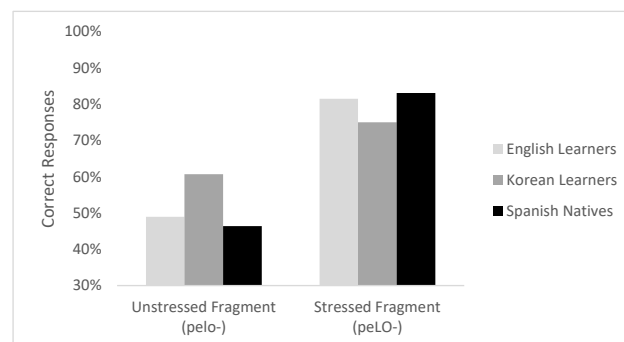


Figure 1: Accuracy results of the cross-modal word-identification task.

A first logit mixed-effects model revealed a main effect of fragment type, showing that the stressed fragment received, overall, more accurate responses; a main effect of group, indicating that the Korean learners patterned differently with respect to the other two groups (native speakers and English learners); and a stress by group interaction, showing that the group effect varied depending on the stress fragment type. While Korean learners were more accurate in the unstressed fragment condition, as compared with the other two groups, they showed the reverse pattern (being less accurate) in the stressed fragment condition.

A second logit mixed-effects model, on only the L2 learners’ data and including proficiency, revealed a main effect of fragment type and proficiency and a significant interaction between fragment type and group, as well as between fragment type and

proficiency. As can be seen in Figures 2 and 3, these results indicate that, while learners were better at using word-level stress for word recognition with increasing proficiency in Spanish, this effect was in general stronger for the Korean learners (similar effect for both fragments types) and, among the English learners, it influenced more positively the stressed fragment than the unstressed fragment.

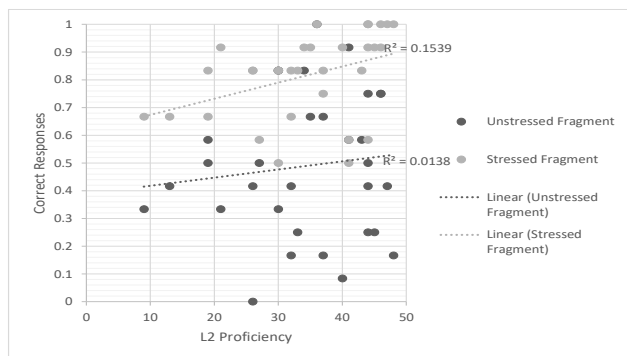


Figure 2: Correlation between accuracy results of English learners and L2 proficiency.

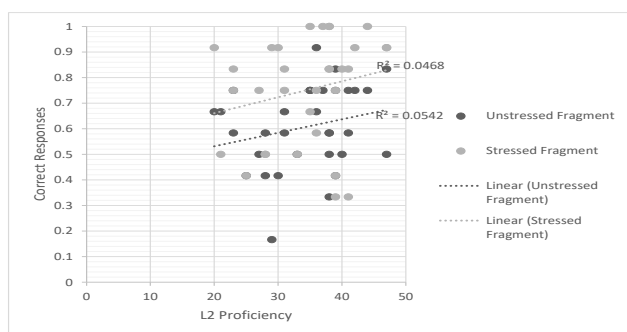


Figure 3: Correlation between accuracy results of Korean learners and L2 proficiency.

A third logit mixed-effects model, on only the L2 learners' data and including the log-transformed frequency of the competitor word, revealed a main effect of fragment type, a main effect of frequency and an interaction between fragment type and group. As can be seen in Figures 4 and 5, these results indicate that, the more frequent the competitor word is, the less accurate participants are (there is a tendency for them to choose the competitor word over the target word). However, this effect reverses in the stressed fragment condition (Korean learners), where the frequency of the competitor word increases the likelihood of selecting the correct target response. However, there is not an interaction with frequency. That is, the fragment type differs between the two groups, but this effect is not influenced by the frequency of the competitor word.

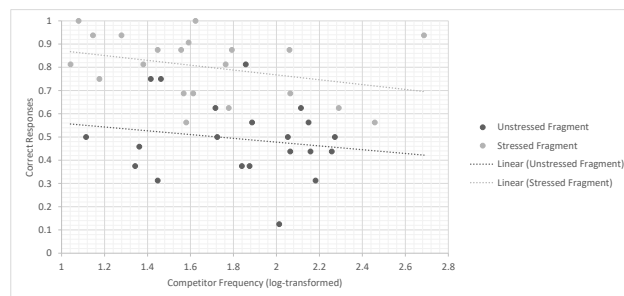


Figure 4: Correlation between accuracy results of English learners and competitor frequency (log-transformed).

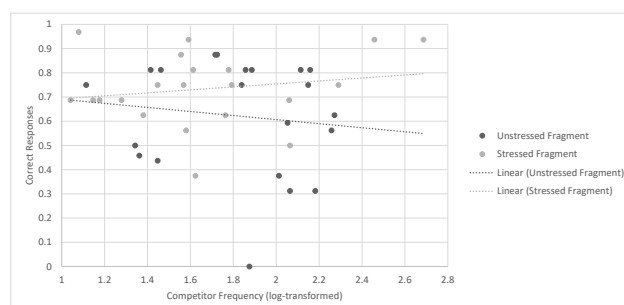


Figure 5: Correlation between accuracy results of Korean learners and competitor frequency (log-transformed).

3. GENERAL DISCUSSION AND CONCLUSION

The current study aimed to explore whether English and Korean listeners at an advanced proficiency level in L2 Spanish could make use of suprasegmental cues to stress for word recognition in Spanish and how lexical information (such as competitor frequency or word familiarity) may modify this effect. The results of this study indicate that the three groups showed a similar pattern in the word-identification task: The overall accuracy is overall higher in the stressed fragment condition than unstressed condition. That is, stressed fragments similarly constrain lexical access, while unstressed fragments showed a smaller effect for all three groups. Thus, these L2 learners appear to rely on similar mechanisms as native speakers for recognizing Spanish words.

These results are in line with previous findings showing that L2 learners can show sensitivity to stress cues for word recognition (for studies on English learners of Spanish, see: Martínez-García, Van Anne, Brown, R., & Tremblay, n.d.; for studies on Korean learners of English, see: Kwon, 2016; Lee, 2015). Moreover, the evidence for L2 learning is further reinforced by the proficiency effects reported in the study. Learners show increased sensitivity to stress as their proficiency in L2 Spanish increases. This suggests that English- and Korean-speaking L2

learners of Spanish can use Spanish stress when recognizing Spanish words. Thus, these L2 learners appear to learn to use suprasegmental cues for word recognition, at least as measured by the word identification task.

However, it is important to keep in mind that other possible interpretations of the results may exist. For example, while it is possible that learners show sensitivity to the greater acoustic saliency of the stressed fragments (that is, that learners have learned to use suprasegmental cues to stress for word recognition in a similar way as native speakers of Spanish), it is also possible that the results obtained are identifying learners' sensitivity to other types of information (or a combination of factors). As preliminary explored in the current study (considering that the word familiarity results have not yet been analyzed), learners seem to also be sensitive to lexical information (their pattern of results interacted with the frequency of the competitor word). Thus, further studies should explore how the frequency of the words and/or of that specific stress pattern or even other variables (such as the amount of lexical competition generated) explain the way in which learners use suprasegmental cues to word recognition in L2 Spanish.

Interestingly, while the three groups showed a similar pattern of results, the absence of stress cues (the unstressed fragment) affected the 3 groups differently, such that the Korean learners were more accurate in the unstressed condition than any of the other two groups, and that the proficiency effect was smaller for the English learners than Korean learners in this specific condition. One of the possible interpretations of this different pattern is related to the Korean prosodic structure (Kwon, 2016; Shin & Speer, 2012). Korean learners have a preference towards unstressed fragments, because they match the preferred High-Low (HL = trochaic) pitch pattern in Korean. Since Korean learners' L1 only allows for trisyllabic words to have HL or Low-Low (LL) pitch pattern, they seem to be transferring this preference towards the stress fragment that follows a similar structure.

While future studies still need to explore variables such as how lexical knowledge (e.g., word familiarity) influences online word recognition or how the specific properties of the L1 affect the use of suprasegmental cues, this is one of the first studies to investigate and compare how two groups of learners of Spanish (English-speaking and Korean-speaking learners) make use of suprasegmental cues to stress during word recognition in an almost native-like manner and how this effect seems to be partially modulated by lexical information.

4. BIBLIOGRAFÍA

- Altenberg, E. P. (2005). The perception of word boundaries in a second language. *Second Language Research*, 21, 325-358.
- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.
- Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and Speech*, 45, 207-228.
- Harris, J. W. (1967). *Spanish phonology* (Doctoral dissertation, Massachusetts Institute of Technology).
- Kwon, Na-Yeon. (2016). *Learning Stress Cues for Word Segmentation by Korean Learners of English* (Doctoral dissertation, 서울대학교 대학원).
- Lee, G. (2015). *Production and perception of Korean and English word-level prominence by Korean speakers*. Ph.D. Dissertation. University of Kansas.
- Martínez-García, M. T., Van Anne, K., Brown, R., & Tremblay, A. (n.d.). *English and Spanish Listeners' Use of "Positive" Stress cues in Spanish Word Recognition*.
- Shin, J., & Speer, S. R. (2012). English lexical stress and spoken word recognition in Korean learners of English. In *Speech Prosody 2012*.
- Soto-Faraco, S., Sebastián-Gallés, N., & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language*, 45, 412-432.
- Tagliaferri, B. (2005). *Paradigm*. Perception Research Systems, Inc. Available at www.perceptionresearchsystems.com
- Tremblay, A. (2008). Is second language lexical access prosodically constrained? Processing of word stress by French Canadian second language learners of English. *Applied Psycholinguistics*, 29, 553-584.

UN CAMBIO ENTONATIVO EN EL HABLA DE LOS JÓVENES DE MAÓ Y ALAIOR (MENORCA)

Ignasi Mascaró¹, Paolo Roseano^{2,3}

¹Institut Menorquí d'Estudis, ²Universitat de Barcelona, ³University of South Africa
ipmascaro@gmail.com, paolo.roseano@ub.edu

ABSTRACT

El cambio lingüístico a nivel suprasegmental ha sido objeto de un número limitado de estudios. Este artículo presenta nuevos datos sobre uno de los pocos cambios entonativos que se han descrito hasta ahora y que está teniendo lugar en el catalán de Menorca. Estudios previos (Mascaró y Roseano, 2015; Roseano et al., 2019) muestran que los adolescentes urbanos en la parte occidental de la isla utilizan un contorno de pregunta total que difiere del patrón utilizado por los adultos, mientras que los adolescentes en los centros rurales de la misma zona no lo hacen. En este artículo presentamos datos de la parte oriental de la isla, donde hemos grabado 78 hablantes que han producido 584 preguntas totales informativas. Los resultados muestran que los adolescentes en el este de Menorca también usan el patrón innovador, pero en menor medida que los jóvenes hablantes urbanos en el oeste de la isla.

Palabras clave: sociofonética, entonación, cambio lingüístico, variación, catalán

Sound change at the suprasegmental level has been the object of a limited number of studies. This paper brings about new data on one of the few intonational changes that has been described so far and which is taking place in Minorcan Catalan. Previous studies (Mascaró & Roseano, 2015; Roseano et al., 2019) show that urban teenagers in the Western part of the island use a yes-no question contour that differs from the tune used by adults, whereas teenagers in rural centres do not do so. In this paper we present data from the Eastern part of the island, where we have recorded 78 speakers who produced 584 yes-no questions. The results show that teenage speakers in Eastern Minorca use the innovative tune, too, but less than young urban speakers in Western Minorca.

Keywords: sociophonetics, intonation, language change, variation, Catalan

1. INTRODUCCIÓN

Si bien la literatura sobre el cambio de sonido a nivel segmental es abundante, el cambio a nivel suprasegmental ha sido objeto de un número mucho menor de estudios. Una de las razones detrás de esta escasez es el hecho de que los estudios sobre la prosodia se han desarrollado en especial modo solo en las últimas tres décadas, por lo que es una disciplina relativamente reciente en comparación con la fonética segmental.

Además de ser escasa, la literatura sobre el cambio a nivel suprasegmental muestra otra limitación: la mayoría de los estudios analizan casos en los que el cambio entonativo es el efecto del contacto lingüístico.

Los casos documentados de cambios entonativos que no pueden atribuirse al contacto entre lenguas son pocos. Los casos conocidos se limitan

esencialmente a la descripción de un cambio entonativo en el japonés de Tokio (Fumio, 1998) y algunos relatos de la propagación de los tonos de juntura final ascendentes en varios dialectos del inglés (ver, entre otros, Jarman & Cruttenden, 1976; Ching, 1982; Guy, Horvath, Vonwiller, Daisley & Rogers, 1986; Allan, 1990; Britain, 1992; Mayo, Aylett, & Ladd, 1997; Daly & Warren, 2001; Fletcher, Grabe, & Warren, 2005; Warren, 2005; Warren, 2015; Sullivan, 2011; Ritchard & Arvaniti, 2014; Armstrong, Piccinini, & Ritchart, 2015).

A los estudios que se acaban de enumerar, se añaden dos trabajos recientes (Mascaró & Roseano, 2015; Roseano, Mascaró, Elvira-García, & Fernández Planas, 2019) que describen un cambio en curso en la entonación del catalán de Menorca. Los adolescentes de la parte occidental de la isla están abandonando el patrón entonativo tradicional (Vanrell & Mascaró, 2013) de las preguntas totales absolutas neutras (que es de tipo descendente-

ascendente, v. Figura 1) y están adoptando un patrón innovador de tipo circunflejo (Figura 2). Esa innovación está casi generalizada en el área urbana de Ciutadella (en la que un 75.9% de los adolescentes utilizan el patrón innovador), mientras que en las zonas rurales de la parte centro-occidental de la isla, como por ejemplo Ferreries, el cambio está en un estadio inicial (solo el 5.1% de los jóvenes utilizan el patrón innovador).

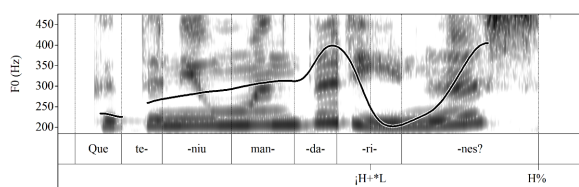


Figura 1: Contorno de F0, transcripción ortográfica y anotación entonativa de la pregunta total informativa *Que teniu mandarines?* '¿Tendrás mandarinas?' pronunciada por una hablante de catalán de Es Migjorn Gran (Menorca).

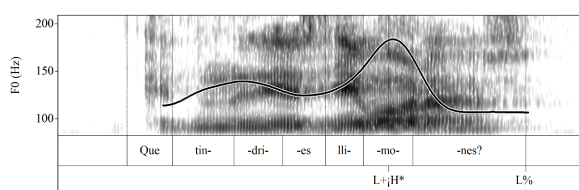


Figura 2: Contorno de F0, transcripción ortográfica y anotación entonativa de la pregunta total informativa *Que tindries llimones?* '¿Tendrás limones?' pronunciada por un hablante de catalán de Ferreries (Menorca).

Los trabajos existentes sobre el cambio entonativo en curso en Menorca plantean la exigencia de recoger datos también en la parte oriental de la isla, para comprobar si (y hasta qué punto) la innovación está presente en la otra área urbana de la isla (es decir, en la ciudad de Maó) y en las zonas menos urbanizadas del sector centro-occidental de Menorca.

En este artículo se presentan los datos que se han recogido en la parte oriental de la isla para completar el cuadro de la descripción del cambio suprasegmental en curso.

2. METODOLOGÍA

Se han grabado datos en dos puntos de encuesta: Maó y Alaior. Maó es un centro urbano costero de unos 30 000 habitantes, mientras que Alaior es una pequeña ciudad de unos 8 000 habitantes en el interior de la isla (Figura 3).

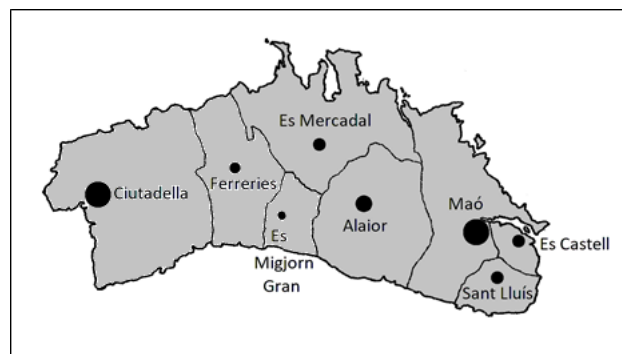


Figura 3: Mapa de Menorca con la indicación de los principales núcleos habitados.

Se han grabado 78 informantes en total, todos ellos hablantes nativos de la variedad menorquina de catalán y con una edad de entre 12 y 19 años. La distribución de los informantes por lugar y sexo se detalla en la Tabla 1.

Tabla 1: Informantes por punto de encuesta y sexo.

Lugar	Hombres	Mujeres	Total
Maó	20	28	48
Alaior	14	16	30

Las grabaciones se han llevado a cabo con una grabadora digital Marantz 600 conectada con un micrófono Shure SM58. Para la elicitación de las preguntas que constituyen el corpus se ha utilizado la técnica del *Discourse Completion Task* (Vanrell, Feldhausen, & Astruc, 2018).

En total, una vez descartados los enunciados que por diferentes razones técnicas no era posible o aconsejable analizar, el corpus así obtenido contiene 584 interrogativas absolutas informativas.

3. RESULTADOS

3.1. Alaior

Los hablantes de Alaior utilizan principalmente el patrón tradicional, que aparece en el 61,2% de los casos (es decir, en 150 de 245), mientras que el contorno innovador está presente en menos de la mitad de las preguntas (38,8%, es decir, 95/245). Los 14 hablantes masculinos y 16 femeninos de Alaior no muestran diferencias importantes (Figura 4), en la medida en que ambos prefieren el contorno tradicional y los porcentajes son similares (63.7% entre los hablantes masculinos, 58.7% en los hablantes femeninos). Aunque las mujeres son ligeramente más innovadoras, esta diferencia no es estadísticamente significativa ($\chi^2(1) = 0.6532$, $p =$

0.418975), lo que también se confirma con la V de Cramér ($\varphi_c = 0.052$).

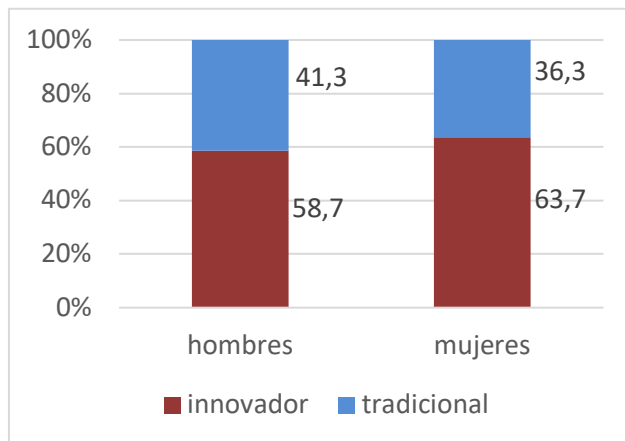


Figura 4: Porcentaje de interrogativas absolutas producidas con el patrón innovador y con el patrón tradicional por los informantes (hombres y mujeres) de Alaior.

3.2. Maó

Los resultados del análisis entonativo muestran que en Maó, en general, predomina el contorno innovador, que aparece en el 66.4% de los casos (es decir, 225/339).

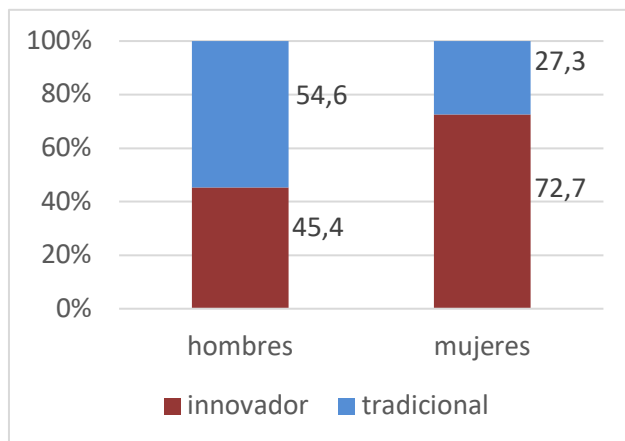


Figura 5: Porcentaje de interrogativas absolutas producidas con el patrón innovador y con el patrón tradicional por los informantes (hombres y mujeres) de Maó.

Si se analizan por separado las producciones de hombres y mujeres (Figura 5), se observa una diferencia entre los hablantes masculinos y femeninos. Estos últimos son más innovadores, ya que usan el contorno innovador en el 72.7% de los casos (es decir, 160/220), mientras que los hablantes masculinos lo hacen solo en el 45.4% de los casos (es decir, 65/119). Esta diferencia es estadísticamente significativa ($\chi^2 (1) = 11.3423, p < .01$), aunque la V de

Cramér sugiere que la asociación no es fuerte ($\varphi_c = 0.183$).

4. DISCUSIÓN Y CONCLUSIONES

Los resultados que se han resumido en la Sección 3 demuestran que en la parte oriental de la isla de Menorca está en marcha el mismo cambio entonativo que se ha descrito para la zona oriental (Mascaró & Roseano, 2015; Roseano et al., 2019): mientras los adultos utilizan el patrón tradicional descendente-ascendente en las interrogativas absolutas neutras, entre los adolescentes se está difundiendo otro patrón, de tipo ascendente-descendente, que está suplantando el primero.

Tal y como se ha descrito para la parte occidental de Menorca (Roseano et al., 2019), uno de los factores que contribuyen a explicar la incidencia del cambio es el tipo de asentamiento: mientras que en las zonas urbanas (Ciudadella y Maó) el cambio está en un estadio avanzado, en las zonas rurales o semi-rurales (Ferrerries y Alaior) se encuentra en una fase menos avanzada. La primera conclusión que se puede extraer de los datos es, por lo tanto, que existe una correlación entre el grado de innovación entonativa y el “grado de urbanidad” de una comunidad. Esto es similar a lo que Trudgill (1974) observó en la península de Brunlanes (Noruega) o Callary (1975) describe para el norte de Illinois: ambos autores encontraron niveles más altos de innovación en los centros urbanos que en las zonas rurales entre ellos.

Si consideramos cómo la variable sexo afecta la entonación entre los adolescentes, la imagen que obtenemos no está completamente clara. Mientras que en Alaior no hay una diferencia significativa entre los hablantes adolescentes masculinos y femeninos, en Maó sí existe, pero es poco significativa. Puesto que esta situación es la misma que se encontraba en Ferrerries y Ciudadella, parece sugerir que las mujeres que hablan, una vez que se inicia el cambio, fomentan o el cambio más que los hombres. En otros términos, en base a nuestros datos no podemos confirmar el principio de Labov (1990, pp. 213-215) según el cual las mujeres son a menudo las iniciadoras de la innovación.

6. BIBLIOGRAFÍA

Allan, S., (1990). The rise of New Zealand intonation. In A. Bell & J. Holmes (Eds.), *New Zealand ways of speaking English* (pp. 115-128). Clevedon: Multilingual Matters.

- Armstrong, M., Piccinini, P., & Ritchart, A. (2015). The phonetics and distribution of non-question rises in two varieties of American English. In M. Wolters, J. Livingstone, B. Beattie, R. Smith, M. MacMahon, J. Stuart-Smith & J. Scobbie (Eds.), *Proceedings of the 18th International Congress of Phonetic Sciences*. London: International Phonetic Association. Retrieved from <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPhS0927.pdf/>
- Britain, D. (1992). Linguistic change in intonation: The use of high rising terminals in New Zealand English. *Language Variation and Change*, 4 (1), 77-104.
- Callary, R., (1975). Phonological change and the development of an urban dialect in Illinois. *Language in Society*, 4, 155-170.
- Ching, M.K.L., (1982). The question intonation in assertions. *American Speech*, 57, 95-107.
- Daly, N., & Warren, P., (2001). Pitching it differently in New Zealand English: Speaker sex and intonation patterns. *Journal of Sociolinguistics*, 5 (1), 85-96.
- Fletcher, J., Grabe, E., & Warren, P., (2005). Intonational variation in four dialects of English: The high rising tune. In S-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 390-409). Oxford: Oxford University Press.
- Fumio, I., (1998). Sociolinguistics of intonation change in progress in Tokyo. *東京外国語大学論集[Area and Culture Studies]* 56, 11-24.
- Guy, G.R., Horvath, B., Vonwiller, J., Daisley, E., & Rogers, I., (1986). An intonation change in progress in Australian English. *Language in Society*, 15, 23-52.
- Jarman, E., & Cruttenden, A., (1976). Belfast intonation and the myth of the fall. *Journal of the International Phonetic Association*, 6 (1), 4-12.
- Labov, W., (1990). The intersection of sex and social class in the course of linguistic change. *Language Variation and Change*, 2, 205-254.
- Mascaró, I.; Roseano, P. (2015). Un canvi entonatiu en curs: Les interrogatives absolutes entre els adolescents menorquins. *Randa* 74, 139-153.
- Mayo, C., Aylett, M., & Ladd, D. R., (1997). Prosodic transcription of Glasgow English: An evaluation study of GlaTol. In Botinis, A., Kouroupetroglou, G. & Carayannis, G. (Eds.), *Proceedings of the ESCA Workshop on Intonation: Theory, Models and Applications*. Athens, pp. 231-234.
- Ritchart, A., & Arvaniti, A., (2014). The form and use of uptalk in Southern Californian English. N. Campbell, D. Gibbon, & D. Hirst (Eds.), *Proceedings of Speech Prosody 2014*. Dublin: Fondúireacht Eolaíochta Éireann/Science Foundation Ireland, 331-335.
- Roseano, P., Mascaró, I., Elvira-García, W., & Fernández Planas, A.M., (2019). La dimensión rural-urbana en un cambio lingüístico en curso en la entonación del catalán de Menorca. *Dialectologia*, 23, 173-192.
- Sullivan, J.N., (2011). Approaching intonational distance and change. Doctoral dissertation, University of Edinburgh.
- Trudgill, P., (1974). Linguistic change and diffusion: Description and explanation in sociolinguistic dialect geography. *Language in Society*, 3, 215-246.
- Vanrell, M.M., Feldhausen, I., & Astruc, L., (2018). The Discourse Completion Task in Romance prosody research: Status quo and outlook. In I. Feldhausen, J. Fliessbach & M. M. Vanrell (Eds.), *Methods in prosody: A Romance language perspective* (pp. 191-227). Berlin: Language Science Press.
- Vanrell, M.M., & Mascaró, I., (2013). Balear. In P. Prieto & T. Cabré (Eds.), *L'entonació dels dialectes Catalans* (pp.75-100). Barcelona: Publicacions de l'Abadia de Montserrat.
- Warren, P., (2005). Patterns of late rising in New Zealand English: Intonational variation or intonation change?. *Language Variation and Change*, 17, 209-230.
- Warren, P., (2015). *Uptalk*. Cambridge: Cambridge University Press.

LA ENTONACIÓN DE LAS PREGUNTAS ABSOLUTAS DEL ESPAÑOL DE SAN JUAN (ARGENTINA) EN HABLA ESPONTÁNEA

Miguel Mateo Ruiz

Universidade Federal do Rio de Janeiro (UFRJ)
miguelmateoruiz@letras.ufrj.br

ABSTRACT

Los estudios de la entonación del español de Argentina se centran, en general, en la variedad de Buenos Aires, la variedad dominante (Toledo y Gurlekian, 2009; Gomes, 2014) y apenas existen otros corpus ni estudios de otras variedades (Terán, V. y Ortega-Llebaria, 2017), de ningún tipo de registro, excepto de las zonas de contacto, por ejemplo, en la triple frontera Argentina-Paraguay-Brasil (Figueredo, 2018). En este trabajo, parte de un proyecto mayor -centrado en la descripción de las variedades dialectales del español latinoamericano-, presentamos las características de las preguntas absolutas en habla espontánea del español de San Juan, a partir de un corpus de 76 enunciados producidos por 18 informantes. Para el análisis, hemos seguido las pautas del protocolo del Análisis Melódico del Habla (AMH), un método fonético. Como era de esperar, los hablantes sanjuaninos han utilizado, principalmente, patrones interrogativos, pero también, patrones que, en otras variedades de español, son enfáticos.

Palabras clave: preguntas absolutas, entonación, habla espontánea, AMH

Generally, studies of Argentina Spanish intonation focus within the variety the Spanish spoken in Buenos Aires, which is the dominant variety (Toledo and Gurlekian, 2009; Gomes, 2014). There is no many other corpus or studies of other varieties or any type of linguistic register (Terán and Ortega-Llebaria, 2017), except for the studies of contact areas, i.e. the triple border Argentina-Paraguay-Brazil (Figueredo, 2018). This study, part of a major project - focused on the description of the dialectal varieties of Latin American Spanish. We present the characteristics of absolute questions in spontaneous speech of the variety of Spanish from San Juan, from a corpus of 76 statements produced by 18 participants. For the analysis, we have followed the guidelines of the phonetic method: Melodic Analysis of Speech (MAS) protocol. As expected, speakers from San Juan used mainly interrogative patterns, but also other patterns that in other Spanish varieties are emphatic.

Keywords: absolute questions, intonation, spontaneous speech, MAS

1. INTRODUCCIÓN

En este trabajo, vamos a describir las melodías /+interrogativas/ del español de San Juan (Argentina), obtenidas mediante el método Análisis Melódico del Habla (AMH) a partir de un corpus de habla espontánea extraído de un corpus más amplio. El objetivo es, por lo tanto, analizar una variedad de la que existen estudios prosódicos y contribuir, de esta forma, a la descripción de las variedades dialectales, no dominantes, del español, especialmente del área latinoamericana. El trabajo se encuadra en un proyecto mayor, de descripción tanto de dichas variedades como del español en contacto con otras lenguas, por ejemplo: con el portugués en Misiones

(Argentina), o, con el quechua y el aymara en Puno (Perú). Presentaremos, pues, sus rasgos melódicos, a partir de enunciados producidos por informantes anónimos en habla espontánea.

2. METODOLOGÍA

En este apartado, presentamos, de forma breve, el método de análisis utilizado (y la teoría que le da soporte); el corpus del que forma parte el corpus analizado y los criterios con los que hicimos la selección.

2.1. El método de análisis

Este trabajo se basa en los presupuestos de la teoría de Análisis Melódico del Habla (AMH) propuestos en

Cantero (2002) y Font-Rotchés (2007), que ofrece un método de análisis formal, *Análisis Melódico del Habla* (véase el protocolo de actuación en Cantero y Font, 2009), basado en el análisis acústico y perceptivo del discurso con medios instrumentales, en este caso, el programa de análisis y síntesis de voz *Praat* (Boersma y Weenink, 1992-2019), con la finalidad de describir el fenómeno de la entonación desde un punto de vista fonético y fonológico.

Esta teoría establece la existencia de tres tipos de entonación: la prelingüística, la paralingüística y la lingüística. En la primera, la entonación actúa como un contenedor que integra el discurso y sus principales manifestaciones son el acento dialectal y el acento extranjero. La entonación paralingüística, por su parte, abarca los amplios márgenes de dispersión de la entonación que veremos a continuación, la lingüística, y permite expresar emociones o características idiolectales, por ejemplo, y como atestiguan los cada vez más numerosos estudios de la entonación “afectiva”, de las emociones, puede estar parcialmente codificada.

La entonación lingüística, en la que se centra este estudio, trata de los rasgos melódicos que tienen un rendimiento fonológico y permiten caracterizar funcionalmente la entonación. En el español, se han identificado ocho tonemas (resultantes de la combinación de estos rasgos fonológicos: /±interrogativo, ± suspendido, ±enfático/). Ello permite hablar de melodías típicas de entonación interrogativa, suspendida, enfática y neutra (que sería la ausencia de todas las anteriores (Cantero 2002). En total, son trece patrones diferentes, con sus variantes y márgenes de dispersión (Cantero y Font-Rotchés, 2007; Font-Rotchés y Mateo, 2011).

El método de análisis, del que en este trabajo solo hemos llevado a cabo el análisis acústico, su primera fase, se puede ver de forma detallada en Cantero y Font-Rotchés (2009) y en Cantero y Mateo (2011). Esta fase acústica consiste, de manera sintética, en la extracción de los valores tonales de los segmentos que configuran el contorno (nucleados en torno a las vocales): anacrusis y primer pico (en qué vocal recae – primera, o bien está desplazado-), cuerpo del contorno, y, finalmente, la inflexión final, el elemento más importante: su dirección y el porcentaje de aumento/disminución.

Lo más destacable del método de análisis, como se podrá observar en los diferentes ejemplos que ilustran

el apartado de resultados, es el procedimiento de estandarización de la curva melódica. Esto permite analizar amplios corpus de habla (de cualquier tipo) y comparar, de forma homogénea y coherente las producciones de grupos heterogéneos de informantes.

2.2. Corpus

El corpus analizado se ha extraído del Corpus de Voces Sanjuaninas (Cuadros y Villavicencio 2013a, 2013b)¹. Este corpus, de más de 13 horas de grabaciones, con hablantes nativos de San Juan o que llevan, un mínimo de 10 años residiendo, está compuesto por distintos tipos de habla: entrevistas guiadas, narraciones espontáneas o conversaciones espontáneas. De este último tipo, del bloque que las autoras denominan “Hechos cotidianos”, 4 horas, hemos extraído los 76 enunciados analizados en este trabajo. Son preguntas absolutas emitidas por informantes anónimos en el contexto de conversaciones sobre cómo ha ido un examen, una visita médica o cuál va a ser la opción de ocio del siguiente fin de semana; son, pues, intervenciones que se pueden encuadrar en el concepto de habla dialogal, desde una perspectiva interactivo-funcional (Hidalgo, 2019). Como explican en su trabajo las autoras, los informantes son de ambos sexos, están clasificados en cuatro franjas etarias y tres niveles de instrucción. A su codificación, hemos añadido el código del audio y el número de enunciado, así, por ejemplo, el enunciado JA3F19-01, de la figura 1, es el enunciado 01 de la grabación 19, que es una persona de San Juan (J inicial), de nivel de instrucción alto (A), de la tercera franja de edad, de 31 a 60 años (3) y de sexo femenino (F).

3. RESULTADOS

A continuación, presentamos los resultados, en los que hemos diferenciado la caracterización de las preguntas totales emitidas con patrones caracterizados como /+ interrogativos/ en el español de España y los enunciados que presentan otros patrones de realización, /+enfáticos/, principalmente.

3.1. Patrones interrogativos

En la figura 1, podemos observar un ejemplo del patrón melódico III, con primer pico desplazado a la sílaba posterior (-te a) y una inflexión final entre los márgenes descritos para el español de España (40-60%), un caso, sin embargo, no habitual, pues 14 de los 21 contornos de este patrón interrogativo, un

¹ Agradecemos a las autoras que nos hayan permitido utilizar su Corpus para esta investigación.

66,6%, presentan inflexiones inferiores a las descritas para dicha variedad del español (Cantero y Font-Rotchés 2007).

Del patron II, hemos encontrado 6 casos, un 7,9% del total del corpus (12,24% de los patrones interrogativos). Al igual que en el caso del patron II, todos los casos presentan inflexiones inferiores a las definidas por Cantero y Font-Rotchés (2007): en nuestro corpus, entre un 8% y un 54% frente al más de un 70% que indican los mencionados autores.

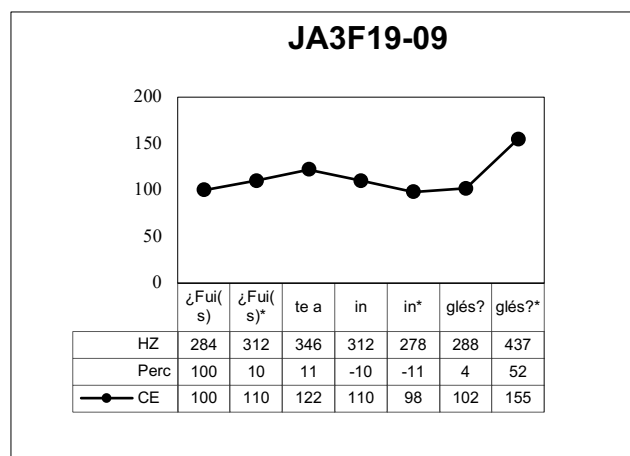


Figura 1: Contorno del enunciado: *¿Fuiste a inglés?*

De los patrones con inflexión final circunfleja (IVa y IVb), solo hemos encontrado casos del primero de ellos (7 casos, un 9,2% del corpus; un 14,3% de los enunciados que siguen un patrón interrogativo). Se puede ver un ejemplo en la figura 2, en el que se puede constatar un segmento nuclear, *-ti-*, con tres valores tonales que configuran el comentado final circunflejo, con un ascenso del 30% y un descenso del -25% (más el -3% de la sílaba final, *-vo*).

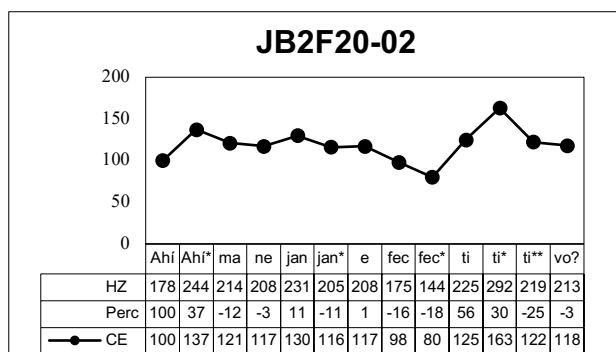


Figura 2: Contorno del enunciado, *¿Ahi manejan efectivo?*

Finalmente, en la figura 3, podemos observar un ejemplo de patrón XIII, descrito en Font-Rotchés y Mateo (2011); un patrón habitual en español, para realizar preguntas enfáticas (+interrogativas + enfáticas), con un ascenso total del 52%. En nuestro corpus hemos encontrado enunciados con ascensos totales entre el 19% y el 194%, similares a los encontrados en el español de España (Font-Rotchés y Mateo, 2017).

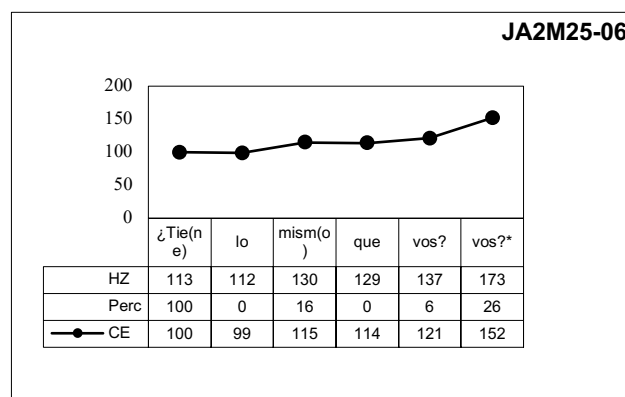


Figura 3: Contorno del enunciado *¿Tiene lo mismo que vos?*

Si comparamos estos resultados con los estudios realizados con el español de Buenos Aires, con metodologías y tipos de corpus diferentes, hallamos distintos grados de coincidencia, menor con lo que señalan Toledo y Gurlekian (2009), Gabriel *et alii* (2010), o Gomes da Silva (2014), en cuanto al uso mayoritario de patrones con inflexión final circunfleja, pues en nuestro corpus, estos usos no son los más utilizados. En nuestros resultados, destacan, por un lado, la presencia de los dos patrones de inflexión final ascendente (II y III), un 55,1% de los casos de patrones interrogativos y, por otro, los casos de patrón XIII, de cuerpo ascendente, un 30,6% de los enunciados con contorno interrogativo.

3.2. Patrones no interrogativos

En este apartado, presentamos los resultados y ejemplos de los 35 enunciados que, aunque en su contexto son emitidos como preguntas absolutas, con respuestas afirmativas o negativas explicitadas en las conversaciones del corpus, presentan contornos melódicos de patrones no interrogativos: neutro o enfáticos. Son 27 enunciados, un 35,5% del corpus, de los cuales, 25 son enfáticos, un 32,9%.

Cabe señalar que los dos casos de patrón I presentan inflexiones finales inferiores a las que describieron

Cantero y Font-Rotchés (2007), para el español de España, entre un 10% y un 15%, en nuestro corpus las podríamos considerar planas, pues son de un 7% y un 8%. Son, además, de la misma informante, por lo que tal vez podría ser una característica idiolectal.

Entre los patrones enfáticos utilizados, cabe destacar los patrones VII y VIII. En la figura 4, podemos ver un ejemplo del patrón VIII, un caso en el que el contorno presenta esta forma circunfleja característica y, por lo tanto, el primer pico y núcleo del enunciado coinciden.

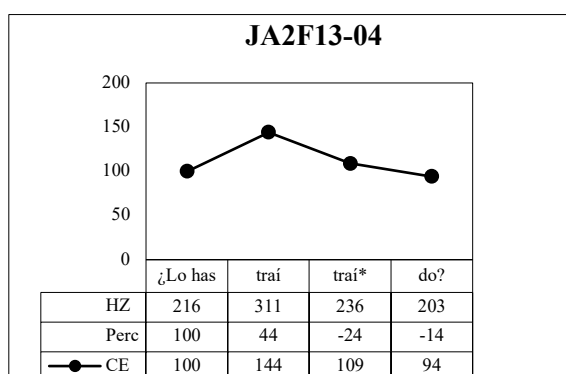


Figura 4: Contorno del enunciado *¿Lo has traído?*

De este patrón, hemos identificado un total de 5 casos en nuestro corpus, lo que supone un 20% del total de patrones enfáticos utilizados por los informantes sanjuaninos, y, sobre el total de los enunciados del corpus, un 6,6%.

Para cerrar este apartado, presentamos un cuadro resumen (tabla 1), con los datos de cada uno de los patrones encontrados en nuestro corpus de habla espontánea de San Juan.

Tabla 1: Resumen de patrones encontrados en el corpus analizado.

	Patrón	Casos	Porc.
+Inter.	II	6	64,47%
	III	21	
	IVa	7	
	XIII	15	
+Enfát.	VII	7	
	VIII	5	
	IX	2	
	Xa	1	
	Xb	2	
	XI	4	
	XIIa	2	32,90%
	XIIb	2	
Neutro	I	2	2,63%

4. CONCLUSIONES

Los hablantes tienen a su disposición numerosos patrones para comunicarse y los utilizan, en habla espontánea, con diferentes finalidades a las que serían esperables en habla más monitorada o intervenciones puntuales. Así, los hablantes de San Juan (Argentina), utilizan en sus preguntas absolutas, mayoritariamente, patrones interrogativos. Sin embargo, como era esperable (ver Font-Rotchés y Mateo, 2017), también han utilizado patrones enfáticos, con los que, además de formular una pregunta a su interlocutor, están expresando algún tipo de emotividad, de afectividad, en la línea de lo que apuntan otros estudios en curso de los grupos de investigación *Grup de Recerca en Entonació i Parla* (GREP) y *Fonoemoció* de Val.Es.CO.

Quedan pendientes algunas cuestiones, entre otras: (1) cuáles son los valores pragmáticos que se transmiten y qué patrones se utilizan en cada caso; (2) ¿coinciden con los descritos para el español europeo? (Font-Rotchés y Mateo, 2017); en los casos de patrones no interrogativos, (3) qué papel juegan las diferencias tonales respecto a enunciados anteriores de los mismos hablantes y qué emoción transmiten (Padilla, 2020) o (4) qué interrelaciones se establecen entre melodía y sus correlatos con la información dinámica –intensidad- o rítmica –duración- (Cantero, 2019).

5. BIBLIOGRAFÍA

- Boersma, P. & Weenink, D. (1992-2012): PRAAT. *Doing phonetics by computer*. Institute of Phonetic Sciences, University of Amsterdam. <http://www.praat.org>
- Cantero, F. J. 2002. *Teoría y análisis de la entonación*. Barcelona: Edicions de la Universitat de Barcelona
- Cantero, F. J. 2019: Análisis prosódico del habla: más allá de la melodía, en María Rosa Álvarez Silva; Alex Muñoz Alvarado & Leonel Ruiz Miyares (eds.): *Comunicación Social: Lingüística, Medios Masivos, Arte, Etnología, Folclor y otras ciencias afines*. Volumen II. Santiago de Cuba: Ediciones Centro de Lingüística Aplicada.
- Cantero, F. J. y Font-Rotchés, D. 2007. Entonación del español en habla espontánea: patrones melódicos y márgenes de dispersión, *Moenia* 13, 69-92.
- Cantero, F.J. y Font-Rotchés, D. 2009. Protocolo para el análisis melódico del habla, *Estudios de Fonética Experimental*, núm. XVIII, 17-32
- Cantero, F.J. y Mateo, M. 2011 Análisis Melódico del Habla: Complejidad y entonación en el discurso, *Oralia*, 14, 105-127.
- Cuadros, M. y Villavicencio, L. 2013a. Voces sanjuaninas: un corpus oral del español en Argentina, en Labastía, L. (ed). *Cuestiones de fonética, fonología y oralidad*. Mendoza: Editorial FFyT-UnCuyo y SAL, 21-32

- Cuadros, M. y Villavicencio, L. 2013b. *Voces Sanjuaninas: Un corpus oral del español en Argentina*. San Juan, Argentina: EFU.
- Figueiredo, N. (2018): *Variação pragmática e ecologia das línguas: análise multimodal de atos de fala no espanhol do paraguai e da argentina*. Tesis doctoral inédita. Universidade Federal do Rio de Janeiro (UFRJ).
- Font-Rotchés, D. 2007. *L'entonació del català*, Biblioteca Milà i Fontanals, 53. Barcelona :Publicacions de l'Abadia de Montserrat,
- Font-Rotchés, D.y Mateo, M. 2011. Absolute interrogatives in Spanish, a new melodic pattern. *Anais do VII Congresso Internacional ABRALIN*. Curitiba: Ed. ABRALIN. Associação Brasileira de Lingüística, 1111-1125.
- Font-Rotchés, D.y Mateo, M. 2017. Melodías para confirmar, preguntar, sugerir o pedir en español, *Phonica*, 13, 49-67.
- Gabriel, C., I. Feldhausen, A. Pešková, L. Colantoni, Su-Ar Lee, V. Arana & L. Labastía. 2010. Argentinian Spanish Intonation. En Prieto, P y Roseano, P. (eds.) *Transcription of Intonation of the Spanish Language*. Munich: Lincom Europa. 285-317
- Gomes da Silva, C. 2014. *Análise entonacional e pragmática de conversas telefônicas coloquiais: os enunciados interrogativos totais nas variedades de Buenos Aires e Santiago do Chile*. Disertação de Mestrado. UFRJ
- Padilla, X. 2020. Las emociones en la conversación coloquial: una aproximación pragmaprosódica. Comunicación presentada en el XLIX Simposio Internacional de la Sociedad Española de Lingüística. Tarragona, 21-24 de enero de 2020.
- Terán, V. y Ortega-Llebaria, M. (2017). A Description of Tucumán Spanish Intonation in Argentina. *Open Linguistics*, 3(1), pp. 456-490, doi:10.1515/opli-2017-0023
- Toledo, G. y Gurlekian, J. (2009): Amper-Argentina: tonemas en oraciones interrogativas absolutas. *EFE XVIII*, 401-415

UNA APROXIMACIÓN AL ANÁLISIS PROSÓDICO DE LA CONVERSACIÓN COLOQUIAL

Miguel Mateo Ruiz¹, Francisco José Cantero Serena²

¹Universidade Federal do Rio de Janeiro (UFRJ), ²Universitat de Barcelona (UB)
miguelmateoruiz@letras.ufrj.br, cantero@ub.edu

ABSTRACT

En este trabajo, presentamos los resultados de un estudio en curso de los diferentes rasgos prosódicos (melódicos, dinámicos y rítmicos) de una conversación coloquial del corpus Val.Es.Co. Se trata de un análisis multidimensional, de los datos que expresan las relaciones entre los parámetros físicos de primer orden (el tono y la intensidad) de los centros fónicos de los segmentos tonales y las distancias entre ellos (Cantero, 2019a). Asimismo, establecemos la relación de este análisis acústico con la propuesta interactivo-funcional (Hidalgo, 2019), que presenta puntos de convergencia con la propuesta de Cantero y supone un punto de partida imprescindible para el estudio de la entonación y de la prosodia de la interacción coloquial. A partir de nuestro análisis establecemos las correlaciones existentes entre los citados parámetros acústicos, para interpretar qué valor aportan al contorno entonativo, especialmente en los patrones enfáticos.

Palabras clave: prosodia, conversación coloquial, patrones suspendidos y enfáticos

In this paper, we present the results of an ongoing study of the different prosodic features (melodic, dynamic and rhythmic) of a colloquial conversation taken from the Val.Es.Co.corpus. This analysis is multidimensional: of the data that expresses the relationships among those parameters of first physical order (tone and intensity); of the tonal segments phonic canthers, and of the distances among them (Cantero, 2019a). Likewise, we establish the relationship between this acoustic analysis and the interactive-functional proposal (Hidalgo, 2019), which presents points of convergence with the Cantero proposal and represents an essential starting point for the study of intonation and prosody of the colloquial interaction. From our analysis, we establish the correlations between the aforementioned acoustic parameters, in order to interpret what value they contribute to the intonation contour, especially in emphatic patterns.

Keywords: prosody, colloquial conversation, suspended and emphatic patterns

1. INTRODUCCIÓN

En este trabajo, vamos a presentar los resultados de un estudio piloto sobre la prosodia de la conversación coloquial, a partir de una muestra del corpus Val.Es.Co. El estudio se encuadra en un proyecto que pretende analizar la conversación coloquial desde una perspectiva integradora entre los modelos del Análisis Prosódico del Habla (APH) (Cantero, 2019a), y el Enfoque Interactivo-Funcional (Hidalgo, 2019).

Como se puso de manifiesto en Mateo y Fonseca (2020), en la conversación coloquial raramente se producen entonaciones genuinamente “neutras” o no marcadas; por el contrario, la entonación coloquial es abrumadoramente enfática, a menudo muy emocional (Cantero, 2014), y no siempre basta con un análisis de

la melodía para caracterizarla. El análisis de la interacción conversacional debe afrontarse desde una perspectiva integradora, que incluya las tres dimensiones de la prosodia del habla: la melodía, la intensidad y las duraciones.

En Cantero (2002) y Font-Rotchés (2007), se ofrece un método de análisis formal de la entonación, el *Análisis Melódico del Habla* (AMH) (cuyo protocolo experimental, acústico y perceptivo, se formuló en Cantero y Font-Rotchés, 2009 y, más recientemente, 2020), que permite el análisis fonético objetivo de la melodía del habla y su caracterización fonológica. En Cantero (2019a), se añadieron al método el análisis de las intensidades (análisis dinámico) y el análisis de las duraciones (análisis rítmico) a lo largo del enunciado: lo que llamamos, en conjunto, Análisis Prosódico del Habla (APH).

El modelo establece tres niveles de análisis de la entonación (Cantero y Mateo, 2011): un nivel prelingüístico, un nivel lingüístico y un nivel paralingüístico. En nuestro estudio, nos centramos en estos dos últimos.

La entonación lingüística se refiere a los rasgos melódicos que tienen un rendimiento fonológico y que permiten caracterizar funcionalmente la entonación. En español peninsular, por ejemplo, se han identificado ocho tonemas (resultantes de la combinación de los rasgos fonológicos: /±interrogativo/ /±suspendido/ /±enfático/) y trece patrones diferentes, con sus variantes y sus márgenes de dispersión (Cantero y Font-Rotchés, 2007; y Font-Rotchés y Mateo, 2011).

La entonación paralingüística, por su parte, abarca los amplios márgenes de dispersión de la entonación lingüística (sobre todo, de las entonaciones enfáticas) y permite expresar la afectividad en la interacción (y, en primer lugar, la emoción, que puede estar parcialmente codificada, como atestiguan los cada vez más numerosos estudios, en el ámbito del español, por ejemplo: Cantero, 2014, 2019b; Hidalgo, 2020; Padilla, 2020; Planas, 2020).

La propuesta Interactivo-Funcional de Hidalgo (2019: 254-256) parte de la premisa de la existencia de 5 contornos básicos (ver figura 1) y presta atención a los contornos hipotéticos del cuerpo y de la inflexión final de los diversos subactos que se pueden dar tanto en el nivel monológico como en el nivel dialógico del habla, en el que dos o más hablantes interactúan.

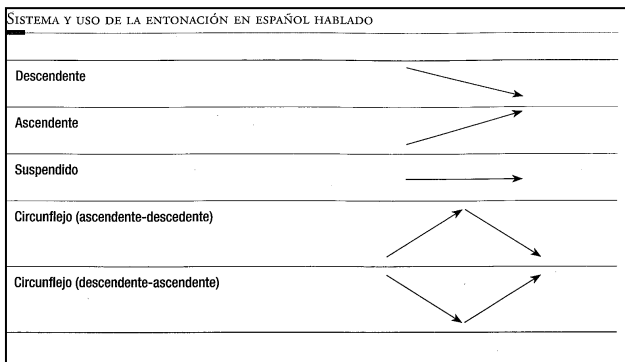


Figura 1: Distinción básica de contornos (Hidalgo, 2019:254)

Tras presentar estos contornos básicos, señala que en el habla, con su combinación (en cuerpo e inflexión final), se pueden generar hasta 20 posibles “patrones holísticos (teóricos, al menos)” (Hidalgo, 2019:254).

2. METODOLOGÍA

En este apartado, presentamos, de forma muy resumida, el método de análisis utilizado, así como el corpus que hemos analizado en nuestro trabajo.

2.1. El método de análisis

A partir de los procedimientos establecidos en AMH, en su fase de análisis acústico, en Cantero (2019a) se implementó el análisis melódico con un análisis dinámico (que da cuenta de las relaciones de intensidad entre los centros tonales del enunciado) y un análisis rítmico (que parte de una medición objetiva de la duración de los pies rítmicos). Esta fase acústica se compone, de manera resumida, de los siguientes pasos:

- Análisis melódico: extracción de los valores tonales de los segmentos que configuran el contorno, nucleados en torno a las vocales: anacrusis y primer pico (en qué vocal recae - primera, o bien está desplazado-), cuerpo del contorno, y, finalmente, la inflexión final, el elemento más importante: su dirección y el porcentaje de aumento o disminución.
- Análisis dinámico: obtención de los picos de intensidad de cada uno de los segmentos o centros tonales del enunciado.
- Análisis rítmico: cálculo de las distancias (cantidad de tiempo) entre los picos de intensidad de cada centro tonal, que será la duración de cada pie rítmico.
- Estandarización de los valores de cada una de las tres magnitudes, y representación gráfica.

En todos los casos, como se puede observar en las figuras 2 y 3, de un estudio en curso sobre la prosodia del portugués de Brasil, como se ha señalado, se estandarizan los valores tonales, de intensidad y de duración, para poder realizar comparaciones y estudiar las correlaciones entre los diversos valores de cada una de las magnitudes.

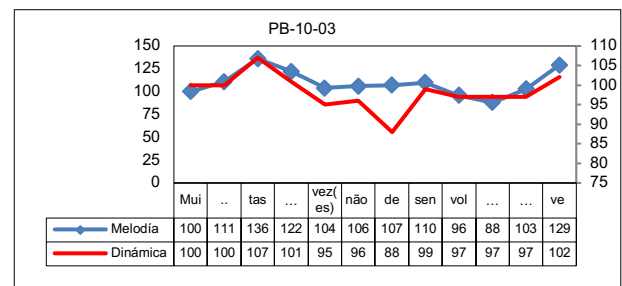


Figura 2: Correlación melódico (tono, color azul) – dinámica (intensidad, color rojo)

En los casos del tono y de la intensidad, hemos utilizado gráficos de líneas que permiten ver su evolución de forma paralela; en cambio, para la duración, hemos optado por un gráfico de barras, que creemos que da una idea más ajustada de la magnitud que realmente mide, la *cantidad* de tiempo entre los picos tonales de los segmentos vocálicos.

Así, por ejemplo, en la figura 3, nuestra propuesta de análisis y su representación gráfica nos permite observar de forma clara cómo el acento en portugués brasileño es un fenómeno de cantidad o duración: *mui-*, tónica, tiene un valor de F0 significativamente inferior a la sílaba átona, *-tas*, igual que en el caso de la sílaba *-vol-*, de *desenvolve*, pero sus duraciones, tanto de *mui-* como de *-vol-*, son significativamente mayores.

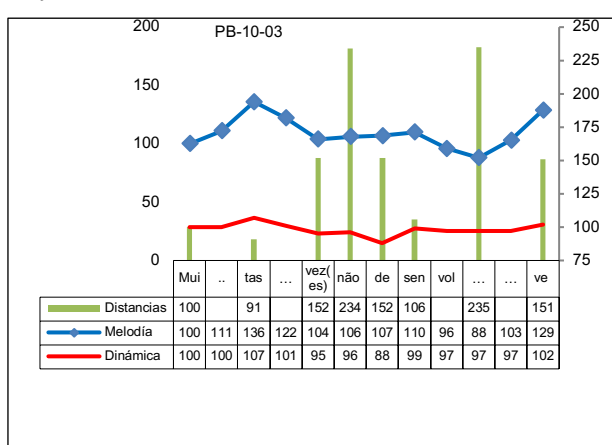


Figura 3: Correlación melódico (tono) – dinámica (intensidad) y rítmica (cantidad de tiempo)

En nuestro trabajo, nos centraremos en el análisis prosódico de dos grupos principales de intervenciones:

(1) los operadores discursivos, que han sido abordados principalmente desde aspectos sintácticos y pragmáticos en el seno del grupo Val.Es.Co (Pons, 2012, a modo de ejemplo), pero también con aproximaciones acústicas (Pereira, 2011); y

(2) los enunciados enfáticos, fundamentales en la conversación coloquial, para el mantenimiento del turno de habla y para comunicar emociones, como es habitual en nuestras conversaciones.

2.2. Corpus

El corpus analizado se ha extraído del Corpus Val.Es.Co. (Briz y grupo VAL.Es.Co, 2002; Cabedo y Pons, 2013). Son conversaciones coloquiales espontáneas, auténticas y genuinas. Los actos y subactos de habla analizados son intervenciones que podemos situar, por lo tanto, en el ámbito dialogal, desde la perspectiva Interactivo-Funcional propuesta por Hidalgo (2019). En concreto, se trata de la conversación *h25a1*, en la que intervienen cuatro interlocutores; con una duración de 10 minutos, contiene 179 actos de habla y 360 subactos.

Para nuestro análisis hemos excluido las intervenciones del investigador (hablante D) y no hemos considerado los siguientes subactos: las intervenciones no verbales (que podrían ser objeto de otro estudio, multimodal), los actos que solo contienen monosílabos (“sí”, “no” o similares, sin cambios prosódicos) ni los subactos con voces sobrepuestas, inevitables en una grabación espontánea. En total, hemos seleccionado y analizado 50 subactos.

3. RESULTADOS

Mostramos a continuación dos ejemplos del análisis del marcador discursivo más utilizado en la conversación objeto de análisis, *claro*, que aparece hasta en 19 ocasiones (excluyendo, como se ha indicado, las intervenciones del investigador).

En la figura 4, presentamos un ejemplo de análisis melódico (tono) y dinámico (intensidad), en el que se puede observar claramente la divergencia entre ambos parámetros físicos: un descenso inapreciable en la intensidad, pero un ascenso marcado (de un 46% en total) en el tono. Se trata de un uso enfático del ascenso, en el que se alarga la segunda sílaba (con dos valores).

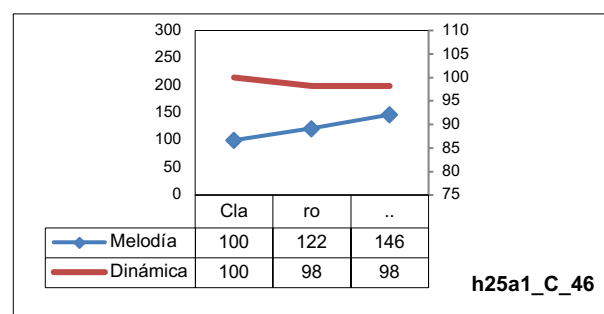


Figura 4: Análisis melódico y dinámico del subacto *CLARO*¹(C-46)

¹ Mantenemos la transcripción del original Val.Es.Co, en mayúsculas.

En la figura 5, podemos comprobar, al añadir la duración, que las dos sílabas tienen el mismo valor. Es decir, la vocal átona (con la misma intensidad y la misma duración que la vocal tónica, pero con dos centros tonales: con una inflexión tonal) se ha convertido en la vocal con mayor relieve. En este caso, el marcador *claro* está enfatizado, al convertir en “tónica” a la vocal átona.

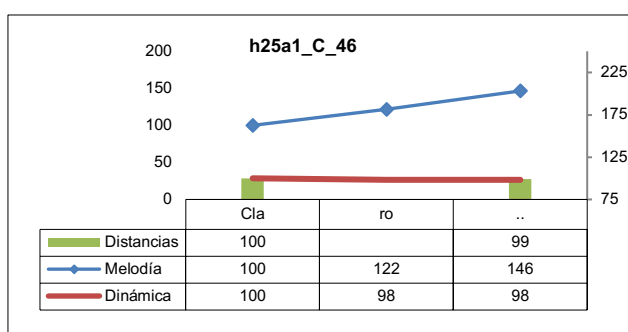


Figura 5: Análisis prosódico del subacto CLARO (C_46)

Si comparamos estos datos con otro ejemplo de la misma informante (v. figura 6), en el que el adverbio *claro* tiene un mero carácter confirmativo, las diferencias resultan evidentes: no hay variación significativa ni en tono (un descenso de un 5%), ni en intensidad (descenso de un 6%). En este caso, la vocal átona no constituye una inflexión tonal, y su duración es un 15% menor a la vocal tónica. Es decir, el marcador, en este caso, no está enfatizado.

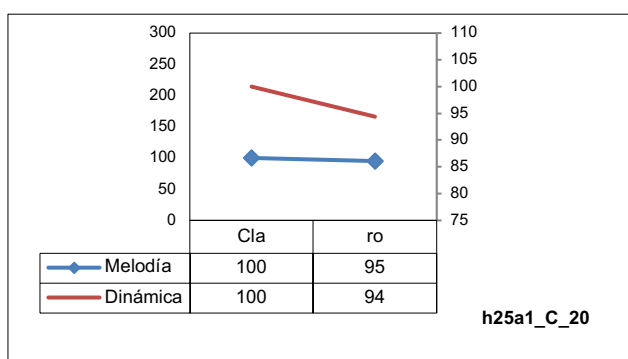


Figura 6: Análisis melódico y dinámico del subacto CLARO (C_20)

Padilla (2020) señala que una de las características de la entonación emocional es la diferencia tonal y de intensidad respecto a los subactos contextuales (anterior y posterior) del hablante que lo emite. En la figura 7, podemos observar otro enunciado producido por el mismo hablante del ejemplo anterior: se trata, justamente, de su intervención posterior a ese *claro* confirmativo, que también presentamos en el gráfico, para poder contrastar las diferencias que se producen.

Podemos observar cómo los picos tonales y los picos de intensidad, que marcan el carácter enfático del contorno (desde las dos perspectivas de análisis, APH e Interactivo-Funcional) coinciden excepto en la parte final y en el segmento *-por-*, un elemento átono, lo cual confirmaría el desplazamiento, en general, a los segmentos átonos del enunciado (*ga, to, por, fo*), tanto de los picos tonales como de los de intensidad. El énfasis en la expresión *por teléfono* se marca, precisamente, con ese desajuste entre la melodía y la intensidad, convirtiendo (como ocurría en el subacto C-46) un elemento átono *-por-* en el elemento relevante, (como si se convirtiera en la sílaba tónica). Finalmente, el ascenso final de intensidad marca el énfasis del enunciado: de nuevo, un desajuste entre la melodía (con un final plano) y la dinámica.

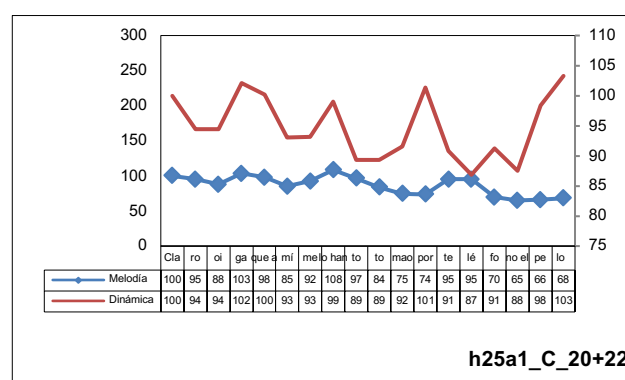


Figura 7: Análisis melódico y dinámico de los subactos CLARO y oiga que a mí me lo han tomao por teléfono el pelo

4. CONCLUSIONES

Estos desajustes entre el perfil melódico y el perfil dinámico del enunciado constituyen un rasgo de énfasis prosódico ampliamente documentado en nuestro estudio. Las diferencias de duración, por su parte, parecen correlacionarse, básicamente, con los desdoblamientos tonales, independientemente de si se trata de vocales tónicas o átonas: hemos comprobado que también es un rasgo de énfasis prosódico frecuente poner de relieve elementos que deberían ser átonos.

El mero análisis de la melodía no nos hubiera permitido, en ningún caso, identificar esos rasgos de énfasis con los que los hablantes ponen en valor elementos de su discurso con marcas afectivas que dependen, como hemos podido comprobar, de las otras dimensiones prosódicas: la intensidad, la duración y, sobre todo, el desajuste cruzado entre el perfil de cada dimensión.

Este trabajo solo es una primera aproximación a las características prosódicas de la conversación coloquial espontánea. En adelante, será necesario el estudio de un corpus mayor, para verificar los

resultados que el estudio de la conversación seleccionada nos permite apuntar. En concreto, profundizar en el papel que juegan las diferencias tonales, de intensidad y de duración con respecto a los enunciados adyacentes, y determinar qué funciones pueden correlacionarse con los desajustes prosódicos en la interacción, como marcas de énfasis emocional.

5. BIBLIOGRAFÍA

- Briz, A. y grupo Val.Es.Co. 2002. *Corpus de conversaciones coloquiales*, Madrid.: Arco Libros.
- Briz, A. y otros .2014: Las unidades del discurso oral. La propuesta Val.Es.Co. de segmentación de la conversación (coloquial), *Estudios de Lingüística del Español* 35, 13-73.
- Cabedo, A. y Pons, S. (eds.). 2013. *Corpus Val.Es.Co 2.0*. Valencia: Val.Es.Co. (Valencia, Español Coloquial), Departamento de Filología Española, Universidad de Valencia. Consultado en <http://www.valesco.es>
- Cantero, F. J. 2002. *Teoría y análisis de la entonación*. Barcelona: Edicions de la Universitat de Barcelona.
- Cantero, F. J. 2014. Códigos de la entonación y entonación emocional, en A. Díaz, M.C. Fumero, M.P. Lojendio, S. Burgess, E. Sosa y A. Cano Ginés (eds.): *Actas del 31 Congreso Internacional AESLA*. La Laguna: Universidad de La Laguna, 618-629.
- Cantero, F. J. 2019a. Análisis prosódico del habla: más allá de la melodía, en María Rosa Álvarez Silva; Alex Muñoz Alvarado & Leonel Ruiz Miyares (eds.): *Comunicación Social: Lingüística, Medios Masivos, Arte, Etnología, Folclor y otras ciencias afines*. Volumen II. Santiago de Cuba: Ediciones Centro de Lingüística Aplicada.
- Cantero, F. J. 2019b. Expresión y contacto: dimensiones de la afectividad en prosodia. *Moenia* 25, 521-537.
- Cantero, F. J. y Font-Rotchés, D. 2007. Entonación del español en habla espontánea: patrones melódicos y márgenes de dispersión, *Moenia* 13, 69-92.
- Cantero, F.J. y Font-Rotchés, D. 2009. Protocolo para el análisis melódico del habla, *Estudios de Fonética Experimental*, núm. XVIII, 17-32
- Cantero, F. J. y Font-Rotchés, D. 2020. Melodic Analysis of Speech (MAS). *Phonetics of Intonation*. En: Abasolo, J.; de Pablo, I.; Ensunza, A.: *Contributions on education*. Universidad del País Vasco, 20-47.
- Cantero, F.J. y Mateo, M. 2011 Análisis Melódico del Habla: Complejidad y entonación en el discurso, *Oralia* 14, 105-127.
- Font-Rotchés, D. 2007. *L'entonació del català*. Biblioteca Milà i Fontanals, 53. Barcelona :Publicacions de l'Abadia de Montserrat,
- Font-Rotchés, D. y Mateo, M. 2011. Absolute interrogatives in Spanish, a new melodic pattern. *Anais do VII Congresso Internacional ABRALIN*. Curitiba: Ed. ABRALIN. Associação Brasileira de Lingüística, 1111-1125.
- Font-Rotchés, D. y Mateo, M. 2017. Melodías para confirmar, preguntar, sugerir o pedir en español, *Phonica* 13, 49-67.
- Hidalgo, A.2019. *Sistema y uso de la entonación en español hablado Aproximación interactivo-funcional*. Santiago de Chile: UAH/Ediciones.
- Hidalgo, A. 2020. Rasgos prosódicos de la emoción: estudio de un corpus conversacional. *Phonica* 16, 36-53.
- Mateo, M. y Fonseca de Oliveira, A. 2020. Español meridional: más allá de las melodías de la conversación. Comunicación presentada en el *XLIX Simposio Internacional de la Sociedad Española de Lingüística*. Tarragona, 21-24 de enero de 2020.
- Padilla, X. 2020. Prosodia emocional y conversación espontánea: bases para el establecimiento de un protocolo de identificación perceptiva. *Phonica* 16, 4-35.
- Pereira, D. 2011. Análisis acústico de los marcadores discursivos *a ver, bueno, claro, vale, ¿cómo? y ya*. *Onomázein* 24/2, 221-234.
- Planas, S. 2020. Sobre los códigos de entonación emocional convencional de la preguntas periodísticas en las entrevistas radiofónicas. *Phonica* 16, 104-121.
- Pons, S. 2012. *Claro*. Una palabra sobre los apellidos de la sintaxis. En José Jesús de Bustos Tovar, et al (eds.), *Sintaxis y análisis del discurso hablado en español: homenaje a Antonio Narbona*. Universidad de Sevilla, 375-389

P-CENTRES AND NUCLEAR VOWEL DURATION IN THE CONTEXT OF IRISH (GAELIC) INITIAL MUTATION

Connor McCabe

Trinity College, Dublin
cmccabel@tcd.ie

ABSTRACT

A pilot study investigated the behaviour of one of two initial mutations exhibited by Irish (Gaelic) in a metronome synchronisation experiment. Two of three participants were found to treat mutated items almost identically to their lexical matches, rather than to their ostensibly identical homophones. Further investigation found shorter vowel durations in mutated conditions with earlier P-centres. Compensatory shortening of long vowels in order to maintain syllable weight furnishes a possible explanation. This connects to evidence of weight reckoning beginning with the P-centre rather than with the syllable rhyme, and to previous evidence of compensatory nucleus clipping conditioned by language-specific metrical structure. Further ongoing study is discussed.

Keywords: P-centre, initial mutation, vowel duration, compensation, weight

1. INTRODUCTION

Initial mutation (IM) refers to a typologically rare class of processes which alter (or ‘mutate’) the first segment of lexical items for grammatical purposes. Two such processes exist in Irish (Gaelic): so-called ‘eclipsis’ and ‘lenition’. They are used to indicate grammatical gender and number, amongst other features. The character of mutated items versus their unmutated ‘radical’ counterparts remains an open question (Welby *et al.* 2011, 2016; Ussishkin *et al.* 2017). Segmental work has found little evidence of distinction between mutated and radical/unmutated versions of a given consonant, e.g. unmutated vs mutated <bp>, both /b^v/.

A pilot study investigated the behaviour of initial consonant mutation in the context of a metronome synchronisation task. This paradigm is used to estimate the location of the *P-centre*, the perceptual ‘downbeat’ of a syllable (Morton *et al.* 1976; Fowler 1979; Šturm & Volín 2016; Franich 2018). A mutation-like process of initial prenasalisation in Medumba was investigated by Franich (2018) using metronome synchronisation. Derivation status of prenasalised onsets was found to influence the rhythmic treatment of segmentally identical targets. Irish eclipsis evolved from a process of coarticulatory nasalisation across word boundaries, notionally similar to Medumba’s morphological prenasalisation (McCone 1994). This encouraged a metronome investigation of the relationship between P-centres

and eclipsis as a conceptual extension of Franich’s work.

Results were compared to durational measures for onset consonants and nuclear vowels. Earlier P-centres appear to correspond to shorter vowel durations. Syllable weight is invoked as a possible factor, based on suggestions by Ryan (2014) that weight reckoning may begin at the P-centre rather than the nucleus.

The experiment and its results are described below, with discussion and plans for future work outlined in Sections 3 and 4, respectively.

2. METRONOME EXPERIMENT

Metronome synchronisation experiments estimate P-centre location as the point with which speakers align metronome beats in a target repetition task. This paradigm is well suited to investigating P-centre movement under tight contextual controls, but would not necessarily be suited to answer questions about the acoustic correlates of the P-centre itself.

2.1. Participants and materials

Three native speakers of Irish were recruited to take part in this experiment. All were originally from the village of An Rinn (Ring) in County Waterford and were residing in Dublin at the time of recording.

The present experiment centred on 5 basic targets comprising one of three bilabial consonants (/p^v b^v m^{v-j}/) followed by the long vowel /a:/. All targets are real-word monosyllables in Irish, and were preceded

by the possessive particle *a* /ə/. For /bʲaː/ and /mʲaː/ there was ambiguity between an unmutated form and an eclipsed one: *a bá* ‘her sympathy’ vs *a bpá* ‘their pay’, both /ə bʲaː/, and *a meá* ‘her/their scales’ vs. *a mbá* ‘their sympathy’, both /ə mʲaː/. The discrepancy in palatalization/velarization for the latter pair is set aside for the present discussion. The five target phrases are detailed in Table 1 below. Homophones align horizontally, lexical items align vertically.

		LEXICAL		
SEGMENTAL		<i>PÁ</i> ‘PAY’	<i>BÁ</i> ‘SYMPATHY’	<i>MEÁ</i> ‘SCALES’
	/ə pʲaː/	<i>a pá</i>		
	/ə bʲaː/	<i>a bpá</i>	<i>a bá</i>	
	/ə mʲaː/		<i>a mbá</i>	<i>a meá</i>

Table 1: Elicitation materials with three lexical categories and two eclipsis-derived homophone pairs.

Participants were presented with one target at a time on a computer monitor in the semi-anechoic chamber of the Trinity College Phonetics & Speech Laboratory. They were instructed to repeat each target ‘in time’ with a 60 beats-per-minute metronome track generated in Audacity being played over headphones. Two participants (DB and NÓM) aligned one beat per disyllabic phrase, as intended. One (CB) split phrases over two beats, producing substantially different results.

2.2. Analysis

Results were analysed in Praat (Boersma & Weenink 2019). P-centres were estimated by measuring participants’ placement of metronome beat (MB) relative to segmental landmarks, including beginning of onset closure and vowel onset. A relative measure of MB as a percentage of onset duration was selected. This allowed for a degree of normalisation relative to individual variation in raw duration of onsets.

2.3. Results

DB and NÓM appear to treat ambiguous *a bpá/a mbá* targets in line with their lexical correspondents, rather than with their (near) homophones *a bá/a meá*. CB’s results are distinct due to his idiosyncratic task treatment. Even so, treatment of mutated targets *a bpá* and *a mbá* differed from that of unmutated homophones.

A Mann-Whitney U Test, which evaluates the null hypothesis that two samples are taken from the same distribution was run using R (2019). There was significant differentiation between homophones (for *a bpá–a bá*, $U=75$, $p=0.04777$; for *a mba–a meá*, $U=68$, $p=0.0249$). No such differentiation emerges

within lexical categories (for *a pá–a bpá*, $U=134$, $p=0.8358$; for *a bá–a mbá*, $U=111$, $p=0.5339$). This is consistent with the visually striking pairwise grouping of lexical counterparts, in addition to stark differentiation within homophone pairings.

Medians for each category roughly correspond for lexical pairs *a pá–a bpá* (‘her/their pay’, median MB locations of 30% and 25% through duration of onset consonant, respectively) and *a bá–a mbá* (‘her/their sympathy’, respective medians 52% and 58%). For unmutated items, P-centres are located as expected for different consonants: earliest for voiceless stop /pʲ/ and latest for sonorant /mʲaː/, with voiced /bʲ/ in the middle (Pompino-Marschall 1989; Patel *et al.* 1999). This makes the treatment of mutated items even more striking, as it defies the location predicted by their acoustic and/or durational profiles.

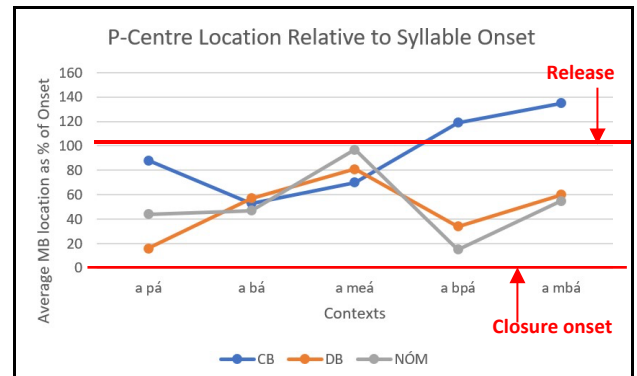


Figure 1: Average MB location within onset consonants for all three participants. DB and NÓM closely correspond.

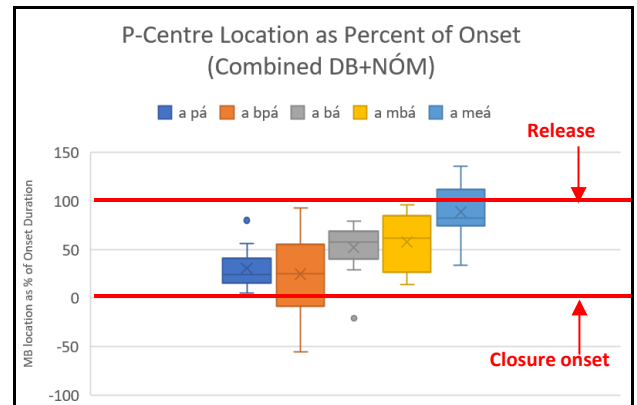


Figure 2: Pooled results for DB and NÓM, $N=16$ per condition., 8 from each participant.

2.4. Durational correlates

Duration is the most direct acoustic correlate of the temporal organisation under observation in this experiment. Additionally, there is ample evidence that onset duration influences P-centre location, with longer onsets corresponding to earlier P-centres (Marcus 1981; Pompino-Marschall 1989; Patel *et al.*

1999; Franich 2017). For this reason, duration of onset consonants in the stimuli under consideration was initially suspected to be behind the apparent P-centre grouping effect. Average durations of onset consonants for all participants are presented in Figure 3 below.

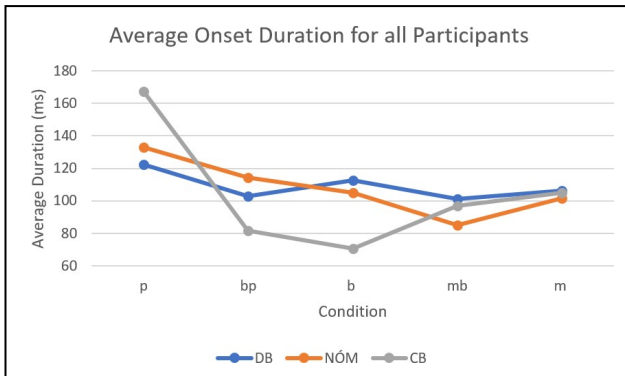


Figure 3: Raw consonant durations for participants in milliseconds; N=10 to 13 per participant per condition.

Mutated consonants showed no grouping with their lexical counterparts in terms of duration. Substantial overlap with homophones was the norm. For stops, this is particularly evident if closure and burst portions are measured individually, in which case homophones are virtually identical with respect to duration for all three participants. This is consistent with previous phonetic studies of mutated consonants in Irish, which found no durational correspondence between mutated onsets and their radical counterparts (Welby *et al.* 2011, 2016).

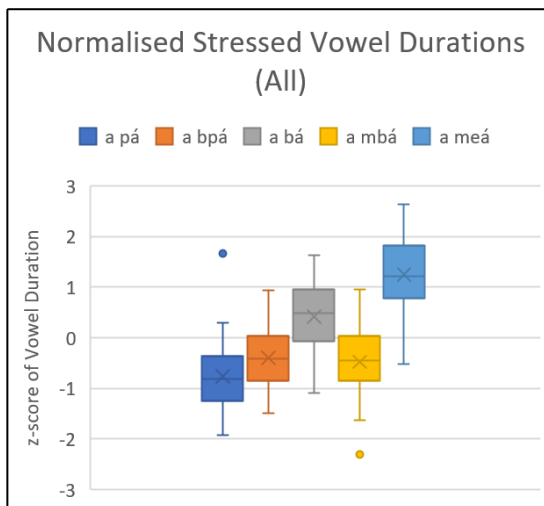


Figure 4: Relative /a:/ durations across conditions. Raw durations normalised as z-scores for each participant and then pooled together; N=34 / condition.

Duration of stressed /a:/ is unexpectedly more promising. Effects of onset variation on nuclear duration are typically negligible, especially when the vowel in question across conditions is controlled for

quality and stress. For all three participants, mutated targets *a bpá* and *a mbá* exhibited strikingly shorter /a:/ duration than in unmutated homophones *a bá* and *a meá*, with no overlap in interquartile range. Pooled data for the three participants, normalised as z-scores relative to individual ranges in raw duration across all conditions, are displayed in Figure 4 above. All participants showed roughly equivalent patterns of vowel durations for raw measurements, range aside.

3. DISCUSSION

The combined patterns of P-centre movement and stressed vowel duration as an apparent function of lexical identity and mutation status are conceptualised in Figure 5 below.

The mutated targets, shown in red, pattern together in two distinct ways. Most directly, for all three participants they have nearly identical vowel durations to the exclusion of the three non-mutated targets. In both cases, the phonemically long vowel is notably shorter than in a corresponding non-mutated homophone. Additionally, for participants DB and NÓM, MB-estimated P-centres are (i) earlier than those of non-mutated homophones, and (ii) consistent with those of non-mutated lexical counterparts. The close correspondence of results between the two participants who treated the task in the same (intended) way, to the exclusion of DB, provides a useful contrast.

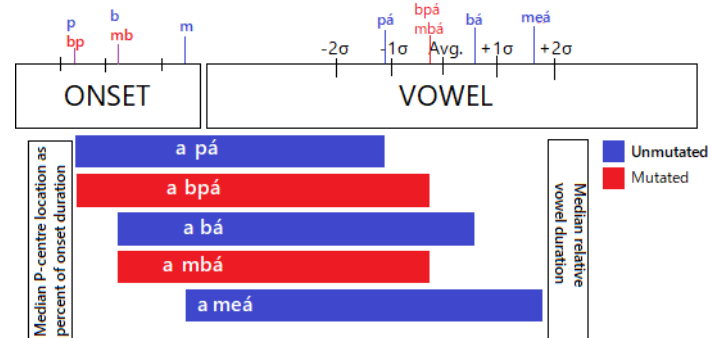


Figure 5: Conceptual visualisation (not to scale) juxtaposing P-centre location as a proportion of onset duration (left) and stressed vowel duration (right).

The correspondence for each of these features is both strikingly clear relative to the small amount of data, and is unlikely to be under conscious control to this degree of accuracy (e.g. for exaggerated disambiguation from homophones). Informal pre-testing of perception with native speakers of Irish further suggests that the homophones produced are indistinguishable above chance; this is a productive phenomenon with no direct basis in perceptual considerations. See Section 4 for further discussion of perception testing, currently underway.

Syllable weight furnishes a possible explanation for the observed pattern, and also has precedent for linking P-centres and syllable rhymes. Ryan (2014) suggests that weight reckoning begins not with the rhyme, but with the P-centre, which is often coincidentally located roughly at the onset-nucleus transition. Independently attested leftward P-centre movement is significantly correlated with traditionally weight-couched phenomena such as stress assignment. This can be accounted for if slightly earlier P-centres result in the partial inclusion of the onset for weight-reckoning, potentially yielding a slightly heavier syllable. Cross-linguistic evidence of compensatory clipping of nuclear vowels in the case of ‘mora sharing’ between nuclei and coda consonants is also attested (Broselow *et al.* 1997).

‘Heavy’ syllables in Irish comprise those containing long vowels or diphthongs and the sequence [ax] (Green 1997: 65; Bennett 2012: 208-219). Weight is also of particular importance for this dialect of Irish (Munster), as stress assignment in the variety is conditioned by weight. A weight-based explanation for the pilot results would predict that earlier P-centres in mutated targets are driving a slight compensatory clipping of long vowel duration. This would satisfyingly tie later P-centres to longer, unclipped vowels. Consider, for example, that the earliest P-centres in *a pá* corresponding to the shortest vowels, and the latest P-centres in *a meá* corresponding to the longest vowels.

Further work is required to support or refute what is currently little more than conjecture. In particular, the behaviour of light syllables (i.e. those containing short vowels) under the same conditions is of interest. Larger scale data collection is underway, outlined briefly below.

4. ONGOING WORK

At the time of writing, the metronome experiment described above is being run on a larger scale, with minor modifications based on considerations from the pilot study. Pilot *meá* has been replaced by *má* ‘plain’, originally avoided due to very low frequency and homophony with *má* ‘if’. Light syllables have also been included in order to, first, test the treatment of these items in the metronome synchronisation paradigm, and second to investigate duration of phonemically short vowels if the lexical grouping effect found for heavy syllables holds. The light syllable stimuli comprise the homophone gradation: *a* (*b*)*paca* ‘her/their pack’, *a* (*m*)*bacadh* ‘her/their bother’, *a* (*m*)*bata* ‘her/their stick’, and *a mata* ‘her/their mat’. All 5 items are initially stressed, with the short vowel /a/ in a light open syllable. Lexical

availability precluded the use of open light monosyllables (e.g. non-existent *pa/ba/ma*).

Participants are also being given more explicit instruction to avoid discrepancies in task treatment. Presentation of individual targets across three separate trials within each session has been implemented to allow a much larger number of tokens per participant to be gathered (10 repetitions per trial x 3 trials, for 30 tokens per target per participant).

Additionally, participants are taking part in a perception task using extracts from the pilot study. This will explore whether apparent rhythmic distinctions produced in the pilot study are perceptible to listeners, and if so, under what conditions this occurs. Listeners are presented with 3 tokens of either *a pá* or *a b(p)á*, both with and without the metronome track included, and assessed in (i) a pairwise AX (same/different) judgement task, (ii) a free-response identification task, and (iii) a forced-choice identification task.

Participants are being recruited in the field, across the three sub-varieties of Munster Irish in Counties Waterford, Cork, and Kerry. Seven Waterford Irish speakers have been recorded thus far, with the ultimate goal of 10 participants per sub-variety.

5. CONCLUSIONS

In a pilot metronome synchronisation task, three native Irish speakers were presented with five segmentally controlled disyllables. These items covered all grades of the ‘eclipsing’ mutation for the bilabial place of articulation (/p^v b^v m^vi/), with two crucial overlaps between basic/unmutated and eclipsis-derived instances of /b^v m^vi/. Two of three participants treated mutated items in line with their underlying lexical match for timing purposes, rather than to their homophonous unmutated counterpart. This pattern emerged as strikingly crisp relative to both the microscopic nature of the variation and the small amount of data involved. Segmental controls suggest that this effect is not the result of contextual acoustic effects on P-centre location. Measurements ruled out onset duration as the source of this.

P-centre groupings for these two participants broadly correspond to subphonemic groupings of nuclear vowel duration for all three participants across these same conditions. An explanation couched in syllable weight would link subtle onset effects to rhyme phenomena. Ongoing work seeks to replicate these results in a larger group of speakers before further exploring the syllable weight hypothesis. This has implications for our understanding of the implementation of initial mutation systems, and of the relationship between onset phenomena and syllable rhymes.

6. REFERENCES

- Audacity Team. (2019). Audacity®: Free Audio Editor and Recorder [Computer application]. Version 2.3.3.
- Bennett, R. (2012). *Foot-conditioned phonotactics and prosodic constituency*. PhD dissertation, University of California, Santa Cruz.
- Boersma, P. and D. Weenink. (2019). Praat: doing phonetics by computer (Computer programme). Version 6.1.08.
- Broselow, E., S-I. Chen & M. Huffman. (1997). "Syllable weight: convergence of phonology and phonetics." *Phonology* 14, 47-82.
- Fowler, C.A. (1979). "'Perceptual centers' in speech production and perception." *Perception & Psychophysics* 25, 375-388.
- Franich, K. (2018). "Tonal and morphophonological effects on the location of perceptual centers (p-centers): Evidence from a Bantu language." *Journal of Phonetics* 67, 21-33.
- Green, A.D. (1997). *The Prosodic Structure of Irish, Scots Gaelic, and Manx*. PhD dissertation, Cornell University.
- McCone, K. (1994). "An tSean-Ghaeilge agus a Réamhstair." in K. McCone, D. McManus, C. Ó hAinle, N. Williams & L. Breathnach (eds) *Stair na Gaeilge: in ómós do Pádraig Ó Fiannachta*, 61-220. Naas, Co. Kildare: Leinster Leader Ltd.
- Patel, A.D., A. Löfqvist & W. Naito. (1999). "The Acoustics and Kinematics of Regularly Timed Speech: A Database and Method for the Study of the P-Center Problem" in *ICPhS-14*, 405-408.
- Pompino-Marschall, B. (1989). "On the psychoacoustic nature of the P-center phenomenon." *Journal of Phonetics* 17, 175-192.
- R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- Ryan, K. (2014). "Onsets contribute to syllable weight: Statistical evidence from stress and meter." *Language* 90(2), 309-341.
- Šturm, P. & J. Volín. (2016). "P-centres in natural disyllabic Czech words in a large-scale speech-metronome synchronization experiment." *Journal of Phonetics* 55, 38-52.
- Ussishkin, A., N. Warner, I. Clayton, D. Brenner, A. Carnie, M. Hammond & M. Fisher. (2017). "Lexical representation and processing of word-initial morphological alternations: Scottish Gaelic mutation." *Laboratory Phonology* 8(1), 1-34.
- Villing, R.C., B.H. Repp, T.E. Ward & J.M. Timoney. (2011). "Measuring perceptual centers using the phase correction response." *Attention, Perception & Psychophysics* 73, 1614-1629.
- Welby, P., M. Ní Chiosáin & B. Ó Raghallaigh. (2011). "A phonetic investigation of Irish eclipsis: preliminary results and challenges." *ICPhS-17*, 2122-2125.
- Welby, P., M. Ní Chiosáin & B. Ó Raghallaigh. (2016). "Total eclipse of the heart? The production of eclipsis in two speaking styles of Irish." *JIPA* 47(2), 125-152.

“JUST COMPARE THAT!” THE USE OF INSTANT AUDIO MESSAGES IN FORENSIC PHONETICS

Chiara Meluzzi¹, Sonia Cenceschi², Nicholas Nese¹

¹University of Pavia, ²University of Applied Sciences and Arts of Southern Switzerland (SUPSI)
chiara.meluzzi@unipv.it, sonia.cenceschi@supsi.ch, nicholas.nese01@universitadipavia.it

ABSTRACT

Recordings available in forensics usually come in different varieties of quality and formats, and specialists are often required to identify or analyse voices recorded in very bad environments and/or with different devices. Moreover, these materials do not last long enough to achieve a scientifically reliable result. Recently, WhatsApp and other instant messaging app have increased the number of available audio data extractable during investigations and criminal trials: this study is part of a larger work focused on the use of this type of data can be used in forensics.

An Italian speech corpus has been recorded considering both controlled and spontaneous speech. These tasks have been repeated with high quality mics, WhatsApp audios, and phone-call recording Apps in order to investigate the variability of intra- and inter-speaker parameters within technology. Results show the problems in normalizing small size corpora and comparing normalized formant values of low-quality recordings.

Keywords: forensic phonetics, speech variability, speaker comparison, forensic linguistics, WhatsApp audio messages, instant messaging

1. INTRODUCTION

Can WhatsApp audio messages be used in solving forensic issues? We think that social media apps (Kaplan, 2015) could represent a very large digital pool from which to draw (Cenceschi & Al. 2019) that could be useful in fill the lack of spontaneous speech corpora available for research purposes.

However, these audio messages belong to a new speech communication style (Cenceschi & Al. 2018 and Nencioni, 1983) introducing new interaction behaviours such as different speakers' expectations, rhythm, pauses, etc., providing audio files with different qualities if compared to laboratory recordings. It is so necessary to investigate similarities and variations with respect to phone-calls and live speech. For these reasons, we made up a controlled experiment partially built with a perfectly balanced corpora and controlled speech, in order to focus this work on intra-speaker and inter-device variations for a limited number of speakers. A large variety of spontaneous Italian speech corpora have

been collected among years (Cresti 2002, Leoni 2006) but none with social media-style speech.

In this scenario, this paper aims at investigating the possibilities offered by instant messaging and environmental audio recordings for forensic purposes. In particular, we will address the following research questions:

- 1) how much does audio quality affect vowel formants values (F1 and F2) at the quantitative and quality levels?
- 2) Are recordings made with different devices directly comparable, or it is better to normalize formant values according to standardized sociophonetic procedures?
- 3) To what extent is possible to compare/identify the voice of the same speaker from different recording devices in a forensic analysis?

In order to answer these questions, we set up a laboratory experiment, in order to compare the voices of the same speakers recorded through different devices and digital applications. Contrary to what usually happens in forensics, in this case the identity of the speakers is known, and the only external (i.e.,

non-linguistic) variable to be addressed will be the recording devices.

Moreover, in the analysis we will address the issue of the automatic quantitative analysis of these data, in order to discuss possible methods for speaker recognition in forensic reports.

2. THE CORPUS

2.1. Materials

For the purpose of the experiment, we selected the three cardinal vowels of Italian /a/ - /i/ and /u/; middle vowels were excluded since their opening degree strongly depends on the phonological context (open vs. closed syllable) and the geographic origin of the speaker (Canepari 1980). We prepare a list of 12 target disyllabic real words with the target vowel surrounded by bilabial or dental voiceless occlusives /p/ - /t/, with a phonological structure of CV.CV or CVC.CV (e.g., *Papa* ‘pope’, *pappa* ‘food’). Words were presented within a sentence balanced for prosodic contour, with the 12 target words and 12 fillers in an automatic random order.

Speakers were asked to read the sentence twice, without interpreting the sentence, and assuming they are just providing a series of neutral statements. After readings, speakers were also asked to describe their bedrooms and their favourite food and/or what they usually cook; this topic were aimed at eliciting a more informal speech.

2.2. The Speakers

12 speakers (6 males and 6 females) have been selected among students of the University of Pavia living at the boarding school “Giasone del Maino”. All speakers were aged between 19 and 25 y.o., with no previously known hearing or speaking problems, and all speaking North-Western Regional Italian as their L1.

2.3. Method and Technology

Each speaker read the list of sentences and provide the descriptions of room and favourite food twice for each recording modality. The 3 recording modalities were performed shortly after one another in the following order:

1. Professional recording (**MIC**), in the soundproofed room in the boarding school “Giasone del Maino”, with a Tascam DR20 (44.1kHz, 16 bit) placed at 30 cm from the speaker;
2. WhatsApp message recordings (**WA**) performed by each participant in his/her own private room at the same boarding school;

3. Voice Recorder APP (**VR**), installed on the Samsung J7 device of the first researcher, recording the incoming phone calls made by each speaker in a quiet environment.

Each device, thus, differ for format of the incoming audio and its digitalization, as summarized in Table 1. WhatsApp is based on the SILK VoIP (Voice over Internet Protocol) codec.

Table 1: Audio quality

Type	Format	Quality
MIC	Wav	44,1 kHz – 16 bit
WA	Ogg/M4a	48.0 kHz – 16 bit
VR	Mp3	44.1 kHz – 16 bit

The corpus composed of 6 sentence-list readings (Task 1) and 6 spontaneous thematic monologues for each speaker (Task 2). In this paper, we will focus on the first task, since it allows comparison at the segmental level without interferences of the prosody, since they are all neutral statement provided during a reading task. Our corpus includes 864 tokens balanced per speakers’ sex, phonological context (singleton - geminate; consonantal environmental), target vowel and recording format.

All data were manually segmented and annotated on PRAAT (Boersma & Weenick 2008), on three different tiers: sentence, target word, target vowel, the latter being identified through the presence/absence of F2 .

3. ANALYSIS

A first pilot study on the corpus was conducted as a target analysis on the midpoint of the target vowels, by automatically extracting the values of the first two formants. A random manual check was conducted on the dataset to verify the data (in particular, for vowel /u/).

The values were entered in a matrix on the software IBM SPSS 20, together with extra-linguistic variables; in different columns we entered non-normalized data in Hertz, as extracted by the script, normalized data in Bark (Schroeder et al. 1979). An Excel matrix with the same values was also extracted, in order to visually inspect data through the web application Visible Vowels (Heeringa & Van de Velde 2018).

For a first qualitative analysis of the vowel space, subjects were divided in two groups according to sex (male vs. female), and we focus on the comparison of recording devices (MIC - WA - VR). Fig. 1 shows the areal variation in the vowel space of the 3 target vowels in unnormalized values: it could be noted how the vowel /u/ shows a great degree of variation among

the 3 recorded modalities, whereas the vowel /a/ seems more stable with respect to speakers' sex.

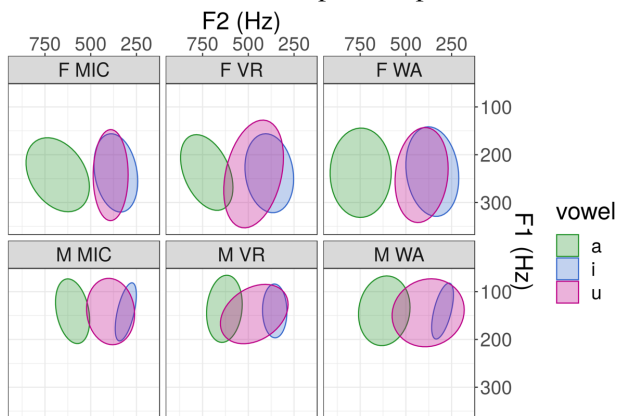


Figure 1: Vowels' areal diffusion according to unnormalized values of F1 and F2 measured on the midpoint, divided by speakers' sex (F vs. M) and recording device (MIC - VR- WA).

We then run a series of univariate ANOVAs with Tukey post-hoc test: our null hypothesis was that speakers recorded in the 3 modalities were the same ones, so in this case we expected that the comparison between formant values show no statistical significance, since we previously know that speakers are the same ones.

Table 2: Results of Anova - Tukey Post-hoc for F1 in M/F groups with the 3 recording devices.

Sex	Vowel	Normalization	df	F	p	partial η-squared	R-squared
Female	/a/	None	2	16,721	0,0001	0,192	0,18
		Bark	2	418,058	0,0001	0,856	0,854
	/i/	None	2	5,435	0,005	0,72	0,058
		Bark	2	301,98	0,0001	0,811	0,808
	/u/	None	2	17,303	0,0001	0,197	0,186
		Bark	2	522,793	0,0001	0,881	0,879
Male	/a/	None	2	7,111	0,001	0,092	0,079
		Bark	2	254,511	0,0001	0,783	0,78
	/i/	None	2	17,831	0,0001	0,202	0,191
		Bark	2	392,752	0,0001	0,848	0,846
	/u/	None	2	22,618	0,0001	0,243	0,232
		Bark	2	169,252	0,0001	0,883	0,881

Table 3. Results of Anova - Tukey Post-hoc for F2 in M/F groups with the 3 recording devices.

Sex	Vowel	Normalization	df	F	p	partial η-squared	R-squared
Female	/a/	None	2	2,827	0,063	0,039	0,025
		Bark	2	131,826	0,0001	0,652	0,647
	/i/	None	2	20,533	0,0001	0,226	0,215
		Bark	2	70,719	0,0001	0,501	0,494
	/u/	None	2	11,014	0,0001	0,135	0,123
		Bark	2	72,726	0,0001	0,508	0,501
Male	/a/	None	2	12,506	0,0001	0,151	0,139
		Bark	2	131,826	0,0001	0,64	0,634
	/i/	None	2	6,703	0,002	0,087	0,074
		Bark	2	60,691	0,0001	0,463	0,455
	/u/	None	2	4,624	0,011	0,062	0,048
		Bark	2	78,129	0,0001	0,526	0,519

However, as shown in Table 2 and 3, a significant difference has been found for normalized (Bark) data for both F1 and F2, with high R-squared values.

The unnormalized Hertz values, obviously divided between male and female subjects, resulted more different for comparison of high vowels /i/ and /u/ in the MIC vs. WA for both F1 and F2, while the

formants of /a/ differ the most between WA and VR in both subgroups.

5. DISCUSSION

These first results confirm what has been widely discussed in sociophonetic research (Van der Harst 2011): standardization procedures have to be applied with cautious since the reduction of variability could also result in a drastic reduction of socio-indexical information about the speaker. In forensic research, this could lead to extreme (and disastrous) consequences, failing at identifying the two samples as coming from the same speakers, as in the experiment presented here. In this respect, if different recording devices have been used in the data collection, un-normalized data seems to be better indicated for speakers' comparison and identification.

Finally, the typical forensic case of comparing environmental interceptions (VR) with professional recordings in court (MIC) will be better not be addressed through semi-automatic analysis. Given the fact that forensic reports on speakers' identification are usually based on very small samples, results from this experiment strongly encouraged not to rely solely on quantitative parameters, but to combine them with qualitative evaluation (data inspection and visualization, speech perception tests).

6. BIBLIOGRAPHY

- Boersma, P., & Weenink, D. (2008). Praat, A System for Doing Phonetics by Computer. Web site: www.praat.org.
- Canepari, L. (1980). *Italiano standard e pronunce regionali*. Cooperativa libreria editrice degli studenti dell'università di Padova.
- Cenceschi, S. & Al. (2019). Collecting Italian spontaneous social media speech: the WAsp2 project. In *XV AISV conference*.
- Cenceschi, & Al. (2018). Verso il riconoscimento automatico della prosodia. *STUDI AISV*, 433-440.
- Cresti, E., Moneglia, M., do Nascimento, F. B., Moreno-Sandoval, A., Véronis, J., Martin, P. & Blum, C. (2002). The C- ORAL-ROM Project. New methods for spoken language archives in a multilingual romance corpus. In *LREC*.
- Heeringa, W., & Van de Velde, H. (2018). Visible Vowels: a Tool for the Visualization of Vowel Variation. In *Proceedings CLARIN Annual Conference* (pp. 8-10).
- Kaplan, A. M. (2015). Social Media, the Digital Revolution, and the Business of Media. *International Journal on Media Management*, 17(4), 197-199.

- Leoni F. A. (2006). Il corpus clips. presentazione del progetto.
- Nencioni, G. (1983). Di scritto e di parlato. Zanichelli.
- Schroeder, M. R., Atal, B. S., & Hall, J. L. (1979). Optimizing digital speech coders by exploiting masking properties of the human ear. *The Journal of the Acoustical Society of America*, 66(6), 1647-1652.
- Van der Harst, S. (2011). *The vowel space paradox: A sociophonetic study on Dutch*. Netherlands Graduate School of Linguistics.

THE ROLE OF ATTENTION CONTROL AND STIMULI LEXICAL STATUS IN HIGH VARIABILITY PHONETIC TRAINING

Joan C. Mora¹, Mireia Ortega¹, Ingrid Mora-Plaza¹, Jonás Fouz-González²

¹Universitat de Barcelona, ²Universidad de Murcia
mora@ub.edu, m.ortega@ub.edu, imoraplaza@ub.edu, jfouz@um.es

ABSTRACT

Cognitive attention control guides auditory processes during speech perception and production, but its contribution to L2 speech learning remains under-researched. We examined the role of attention control in L2 learners' ability to benefit from high-variability phonetic training (HVPT) administered under different conditions (word vs. nonword stimuli). L1-Spanish/Catalan participants were trained on the perception and production of English /æ/-/ʌ/ through AX discrimination, identification and immediate repetition tasks. Training gains were assessed through categorical ABX discrimination and lexical decision tasks, and delayed word and sentence repetition tasks. Individual measures of auditory attention control (selective attention and attention switching) were used to predict perception and production test scores and gains. Findings showed that phonetically-oriented (nonword) HVPT was significantly more effective than lexically-oriented (word) HVPT. Trainees with stronger selective attention were more accurate at perceptually discriminating /æ/ from /ʌ/ and attention switching skill predicted variance in production training gains for the nonword training group.

Keywords: auditory attention control, phonetic training, L2 vowel perception and production

1. INTRODUCTION

When acquiring a second language (L2), learners frequently struggle with the pronunciation of L2 words. Learning pronunciation is a challenging task because learners' perceptual and production systems have become attuned to the phonetic features and distributional properties of the speech sounds of their native language (L1; Kuhl et al., 2008), developing general speech processing skills biased by previous L1 experience. At the phonetic level, L2 speech processing is hindered by L1-based perception mechanisms, often making L2 sound contrasts confusable. This is the case of the English contrast /æ/-/ʌ/ (*cap* vs. *cup*), which poses perceptual difficulties for Spanish learners of English, as they perceptually map both vowels onto their vowel category /a/ (Cebrian, 2019). In addition, the encoding mechanisms responsible for the execution of articulatory gestures in speech production may also reflect the way phonological contrasts have been encoded confusably in the mental lexicon, resulting in a failure to maintain L2 phonological contrasts at the lexical level, leading to the inaccurate pronunciation of words.

Previous research has shown large individual variation in the acquisition of L2 pronunciation, suggesting that learners may differ in their ability to overcome perception and production difficulties originating in L1-based processing mechanisms in order to develop a more target-like pronunciation (Moyer, 2014). Sources of individual differences in L2 phonological acquisition beyond age- and experience-related factors (Piske, MacKay, & Flege, 2001) include speech processing skills (Lengeris & Hazan, 2010) and cognitive skills such as phonological memory (Aliaga-Garcia, Mora, & Cerviño-Povedano, 2011), inhibition (Darcy, Mora, & Daidone, 2016; Lev-Ari & Peperkamp, 2013) and attention (Mora & Mora-Plaza, 2019). However, research on cognitive individual differences in the context of high-variability phonetic training (HVPT) studies is largely under-researched (Perrachione et al., 2011), and the existing research has often produced mixed results (Mokari & Werner, 2019). The current study extends this line of research by exploring the role of auditory attention control in HVPT for an L2 vowel contrast L1-Spanish/Catalan learners of English have difficulty with (/æ/-/ʌ/).

2. ATTENTION AND HVPT CONDITIONS

HVPT has been found to be effective at improving L2 learners' perception and production of difficult L2 phonological contrasts. It facilitates L2 phonological acquisition (Iverson, Pinet, & Evans, 2012) and enhances efficient phonological processing in the recognition and production of L2 words (Bradlow, 2008). Gains after HVPT have been shown to be robust, generalizing to new items and speakers (Thomson, 2012; Wang & Munro, 2004) and across perception and production modalities (Carlet, 2017; Kartushina et al., 2015). However, little research to date has examined the potential differential effects of using word vs. nonword stimuli in training L2 sound contrasts. It is known that phonetically-oriented training with nonwords is more effective than lexically-based training with words, a finding explained by nonwords possibly being more effective at directing trainees' attention to phonetic detail (Thomson & Derwing, 2016), thus resulting in larger training benefits. Additionally, the extent to which individual differences in attention control interact with the lexical status of training stimuli to effect gains in the perception and production of an L2 vowel contrast has not been explored yet. The present study therefore addressed the following research questions (RQs):

RQ1. Is there a differential effect of training stimuli (words vs. nonwords) on HVPT benefits?

RQ2. Are individual differences in auditory attention control related to trainees' HVPT performance and gains?

3. METHODOLOGY

This study followed a pre-test post-test design with a 4-session phonetic training intervention. The stimuli we used for testing and training targeted the English vowels /æ/ and /ʌ/ and consisted of monosyllabic CVC minimal pairs produced by native English male and female speakers.

3.1. Participants

Participants were advanced Catalan/Spanish L2-English ($N=41$) learners randomly assigned to one of two experimental training groups differing in the type of HVPT stimuli used, either words ($N=12$), nonwords ($N=14$), or to a control group ($N=15$). Participant groups were comparable in terms of L2 learning history and proficiency, which we assessed through a language background questionnaire and an elicited imitation task (Ortega et al., 2002).

3.2. Training

The experimental groups were trained in four 45-minute sessions on the perception and production of the English vowel contrast /æ/-/ʌ/ through AX discrimination, identification, and word repetition tasks. The AX discrimination training task consisted of 96 trials per session ($4 \times 96 = 384$ trials) with feedback for error and response latency. AX trials consisted of an equal number of same (AA, BB) and different (AB, BA) trials, and each one of the 4 voices (2 female) used appeared equally frequently in A and B positions. The identification training task consisted of 32 trials per session ($4 \times 32 = 128$ trials) with feedback for error and response latency. Response labels combined orthographic (*cap* vs. *cup*), phonetic (/æ/ vs. /ʌ/) and visual semantic (picture of a *cap* vs. picture of a *cup*) representations. Participants identified 2 minimal pairs (4 words/nonwords) in 4 different voices twice in every session. The word repetition task consisted of the same 32 trials per session used in the identification task, which they were asked to repeat twice (listen-repeat + listen-repeat) within a 2000ms time frame. Participants repeated 2 minimal pairs (4 words/nonwords), uttered by 4 voices, twice in every session.

3.3. Testing

The testing stimuli included words and nonwords, half of which were untrained and half trained. Untrained voices were used for all testing stimuli to ensure training gains involved generalization. Trainees' gains in perception were assessed through categorical ABX discrimination and lexical decision (LD) tests, and gains in production through delayed word (DWR) and sentence repetition (DSR) tests. Attention control was measured through an auditory selective attention test (ASA) and an attention switching test (AttSw).

The ABX discrimination test provided a measure of sensitivity to the target phonological contrast /æ/-/ʌ/ and consisted of 136 trials: 120 (A)/æ/-(B)/ʌ/ test trials (12 nonwords and 18 words \times 4 orders: ABA, ABB, BAB, BAA) and 16 control trials (/æ/-/i:/ or /ʌ/-/ɪ/) with 500ms ISI. All three items in ABX triads were produced by different voices and X was an untrained voice. We obtained accuracy (% correct) and response latency (RT in milliseconds) scores separately by target (X) vowel.

The LD test was used to obtain a measure of perceptual sensitivity to /æ/-/ʌ/ in a lexical context. Participants were asked to decide whether a sound

sequence was an English word or not. The test consisted of 56 trials: 14 English filler words (*jumping*) and nonwords (*sagreem*), 14 English test words containing the target vowels (*map, sun*) and 14 English test nonwords created by substituting the target test vowels by their contrasting counterparts (*mup, san*). We computed an average accuracy rate by participant based on nonword test trials.

The DWR test consisted of 76 items participants were asked to repeat after a sound signal occurring 1500ms after stimulus onset: 12 nonwords (half untrained) and 26 words (20 untrained) per target vowel (12+26=38x2=76). We assessed vowel accuracy gains through normalized Bark-converted spectral distance scores (i.e. Euclidean distances) between learners' and native speakers' vowels in the same words. Vowel measurements (f0, F1, F2) were extracted from a 10-millisecond window by manually placing a cursor at the midpoint of the steady-state portion of the target vowels. To minimize age, gender and vocal tract size effects, frequency values were converted from Hertz (Hz) to Bark (B), and then a Bark-distance normalization procedure was used to provide speaker-independent estimates of vowel quality. The difference in Bark between F1 and f0 (B1-B0) estimated degree of vowel height, whereas the difference between F2 and F1 (B2-B1) estimated degree of vowel frontness (Bohn & Flege, 1990).

The DSR test consisted of 16 sentences participants had to produce twice (32 trials) containing 16 target words (8 /æ/-/ʌ/ minimal pairs) also included in the DWR test. In one DSR trial the participants first read the sentence appearing in standard orthography on the computer screen for 4000ms, then they heard the sentence over headphones and were asked to repeat it from memory after a sound signal occurring 1500ms after the offset of the sentence stimulus. The same measures as those used in the DWR were obtained for the target words in the DSR test.

Individual measures of auditory attention control were obtained from the ASA and AttSw tests. The ASA test was based on the single-talker competition paradigm (Humes, Lee, & Coughlin, 2006). This task consisted of 64 trials of pairs of different sentences presented simultaneously, target vs. competitor, spoken by a male voice and a female voice respectively (e.g. male: *Ready CHARLIE go to BLUE SIX now*; female: *Ready TIGER go to RED EIGHT now*). The sentences were normalized for duration (1700ms). In every trial, a call signal (e.g. *TIGER*) cued the voice participants had to attend to

before the sentence for correctly identifying 1 of 4 colours and 1 of 8 digits visually presented on the screen. Individual ASA scores were computed by adding up all correctly identified colours and digits up to a maximum score of 128.

The AttSw test was an auditory version of the task-switching paradigm (Monsell, 2003) where participants were asked to switch attention between the duration (500ms long or 200ms short) or the voice (male or female) of isolated vowels (/i e ε a ɔ o u/) produced by a male and a female speaker (Safronova & Mora, 2013): 7 vowels x 2 speakers x 2 durations = 28 stimuli x 8 presentations = 224 trials). Trials were randomly presented for categorization (long, short, female, male) in duration (2 trials) + voice (2 trials) alternating runs. Speaker icons appeared predictably in clockwise fashion inside a 4 square framework at the onset of the auditory stimuli, alternating between duration (D) and voice quality dimensions (V) predictably. Participants were expected to obtain lower accuracy and speed scores on switch than on repeat trials. The switching cost (the difference between switch and repeat RTs) was used as a measure of attention control, so that the smaller the switching costs, the stronger the attention control.

4. RESULTS

Generalized Linear Mixed Models were used to statistically assess the effectiveness of HVPT. *Training Type* (nonwords, words), *Testing Time* (T1, T2) and *Vowel* (/æ/, /ʌ/) were entered as fixed effects and *Subject* and *Trial Number* as random factors. The results were then aggregated by *Subject* and Pearson correlations were used to assess the relationship between attention control and L2 vowel perception and production. In general the control group did not improve significantly and overall RTs supported accuracy scores in the perception tests.

The results indicate that both training groups improved significantly and did not differ from one another in discrimination (ABX) accuracy (all $p < .001$) for both vowels (Figure 1), but identification accuracy was lower for /ʌ/ than for /æ/ at both testing times. In the lexical decision test both groups improved, but gains reached significance for the nonword training group only, and did so for /æ/ ($p = .036$) and /ʌ/ ($p = .020$), suggesting that phonetically-oriented training was more effective at improving the perceptual lexical encoding of the /æ/-/ʌ/ contrast (Figure 2).

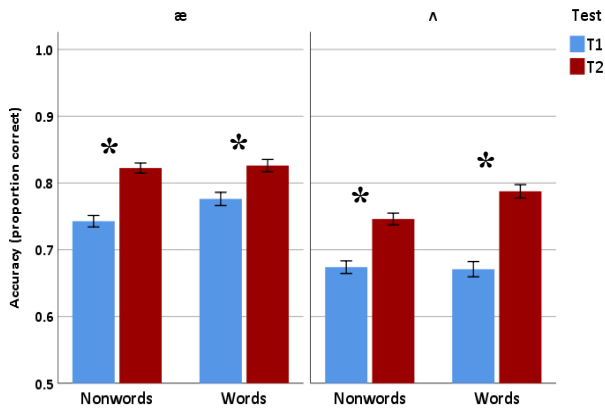


Figure 1: ABX accuracy as a function of *Training Group, Testing Time and Vowel*.

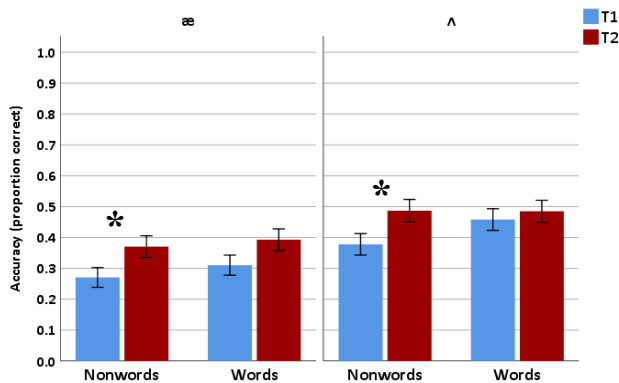


Figure 2: Nonword identification accuracy rate as a function of training group, testing time and vowel.

Training gains in production were much more limited for both groups. However, only the nonword training group improved significantly on /æ/ ($p=.004$) in the DWR test (Figure 3). In general participants were less accurate on /ʌ/ than on /æ/, producing /ʌ/ with larger spectral distance scores than /æ/. In the DSR test, improvement was visible for both vowels in both groups (Figure 4), but it only reached significance for the nonword training group for /ʌ/ ($p=.027$). These results suggest that phonetically-oriented HVPT was more effective than lexically-based HVPT (RQ1).

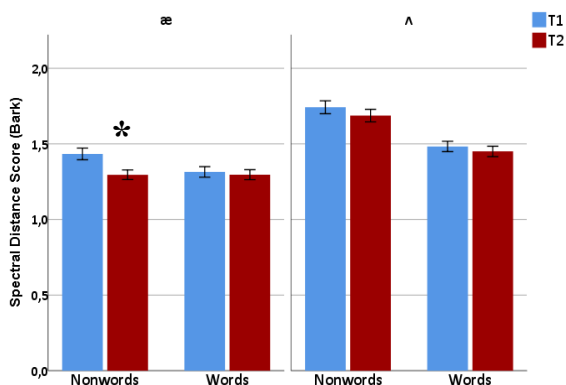


Figure 3: Spectral distance scores as a function of *Training Group, Testing Time and Vowel*.

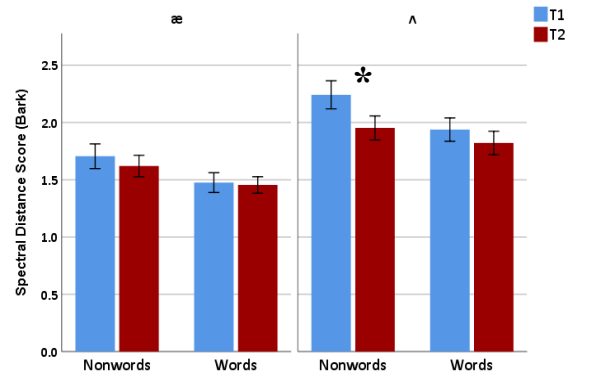


Figure 4: Spectral distance scores as a function of *Training Group, Testing Time and Vowel*.

In the ASA test, the word training group obtained scores ($M=95.58$, $SD=8.93$) that did not differ significantly ($p=.934$) from those of the nonword training group ($M=95.14$, $SD=16.16$). Similarly, in the AttSw test, the word training group ($M=127$ ms, $SD=56$) did not differ significantly ($p=.298$) from the nonword training group ($M=163$ ms, $SD=104$) in the switching cost measure, suggesting that both groups were comparable in cognitive attention control.

Spearman's rank correlation coefficients showed that learners in the nonword training group with stronger selective attention (ASA) obtained significantly higher discrimination scores at T1 ($r=.783$, $p<.001$) and at T2 ($r=.523$, $p=.031$) for the perceptually more difficult /ʌ/. Also for this group and the same vowel, stronger attention switching skill (AttSw) was significantly associated with gains in spectral distance scores in the DWR ($r=-.523$, $p=.029$) and the DSR ($r=-.577$, $p=.039$) tests. Attention control was unrelated to perception gains for this group and to perception and production gains in the word training group.

7. DISCUSSION AND CONCLUSION

Phonetically-oriented training with nonwords was found to be superior to training with words at improving the perception and production of /æ/ and /ʌ/, thus supporting Thomson and Derwing's (2016) findings. We found a modest role of attention control in predicting training gains. Selective attention was related to performance in discriminating the target contrast, but did not predict variance in trainees' gains; however, attention switching skills helped improve the production of /ʌ/ in the nonword training group.

Analyses currently under way will relate attention to individual learning curves during the 4-session HVPT for perception and production.

8. REFERENCES

- Aliaga-García, C., Mora, J., Cerviño-Povedano, E. (2011). L2 speech learning in adulthood and phonological short-term memory. *Poznań Studies in Contemporary Linguistics*, 47(1), 1–14.
- Bradlow, A. R. (2008). Training non-native language sound patterns: Lessons from training Japanese adults on the English /r-/l/ contrast. In Hansen Edwards, J. G. and Zampini, M. L. (Eds.), *Phonology and Second Language Acquisition*. John Benjamins Publishing Company. 287–308
- Bohn, O.-S. & Flege, J. E. (1990) Interlingual identification and the role of foreign language experience in L2 vowel perception, *Applied Psycholinguistics*, 11, 303–328
- Carlet, A. (2017). Different high variability procedure for training L2 Vowels and consonants. In Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (eds.) *Proceedings of the 19th International Congress of Phonetic Sciences*, Melbourne, Australia 2019, 944–948.
- Cebrian, J. (2019) Perceptual assimilation of British English vowels to Spanish monophthongs and diphthongs. *Journal of the Acoustical Society of America*, 145 (1), EL52–EL58.
- Darcy, I., Mora, J. C. & Daidone, D. (2016). The role of inhibitory control in second language phonological processing. *Language Learning*, 66 (4), 741–773.
- Humes, L.E., Lee, J. H., & Coughlin, M.P. (2006). Auditory measures of selective and divided attention in young and older adults using single-talker competition. *The Journal of the Acoustical Society of America*, 120(5), 2926–2937.
- Iverson, P., Pinet, M., & Evans, B. G. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics*, 33(1), 145–160.
- Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *The Journal of the Acoustical Society of America*, 138(2), 817–832.
- Kuhl, P., Conboy, B., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society*, 363, 979–1000.
- Lengeris, A., & Hazan, V. (2010). The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek. *The Journal of the Acoustical Society of America*, 128(6), 3757–3768.
- Lev-Ari, S., & Peperkamp, S. (2013). Low inhibitory skill leads to non-native perception and production in bilinguals' native language. *Journal of Phonetics*, 41(5), 320–331.
- Mokari, P. G., & Werner, S. (2019). On the role of cognitive abilities in second language vowel learning. *Language and Speech*, 62(2), 260–280.
- Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, 7(3), 134–140.
- Mora, J. C. & Mora-Plaza, I. (2019) Contributions of cognitive attention control to L2 speech learning. In Nyvad, A. M., Hejná, M., Højen, A., Jespersen, A. B., & Sørensen, M. H. (eds.) *A Sound Approach to Language Matters - In Honor of Ocke-Schwen Bohn*. Dept. of English, School of Communication & Culture, Aarhus University, Denmark. 477–499.
- Moyer, A. (2014). Exceptional outcomes in L2 phonology: The critical factors of learner engagement and self-regulation. *Applied Linguistics*, 35, 418–440.
- Ortega, L., Iwashita, N., Norris, J.M., & Rabie, S. (2002). An investigation of elicited imitation tasks in crosslinguistic SLA research. In *Second Language Research Forum*, Toronto.
- Perrachione, T. K., Lee, J., Ha, L. Y., & Wong, P. C. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, 130(1), 461–472.
- Piske, T., MacKay, I. R., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, 29(2), 191–215.
- Safronova, E. and Mora, J. C. (2013). Attention control in L2 phonological acquisition. In Llanes Baró, A., Astrid Ciro, L., Gallego Balsà, L. and Mateus Serra, R. M. (Eds.) *Applied Linguistics in the Age of Globalization*. Lleida: Edicions de la Universitat de Lleida. 109–114.
- Thomson, R. I. (2012). Improving L2 listeners' perception of English vowels: A computer-mediated approach. *Language Learning*, 62, 1231–1258.
- Thomson, R.I., & Derwing, T.M. (2016). Is phonemic training using nonsense or real words more effective? In J. Levis, H. Le, I. Lucic, E. Simpson, & S. Vo (Eds.), *Proceedings of the 7th Pronunciation in Second Language Learning and Teaching Conference*. Iowa State University. 88–97.
- Wang, X., & Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System*, 32(4), 539–552.

SECUENCIAS VOCÁLICAS EN FRONTERA DE PALABRAS EN DATOS DEL ESPAÑOL MEXICANO

João Paulo Moraes Lima dos Santos

Universidad de Salamanca
joaopaulomls@hotmail.com, joao.paulo@usal.es

ABSTRACT

El presente estudio es un análisis de secuencias de vocales en datos del español mexicano, dentro de los tres procesos en frontera de palabra: formación de monoptongos, reducción y mantenimiento de hiatos. Con base en el análisis de la duración de las secuencias, de los valores de F1, F2 y de la tonicidad, los resultados muestran que: i) hubo una mayor producción de monoptongos, seguida por hiatos y un número menor de reducciones; ii) los monoptongos tienen una duración menor que las reducciones y estos, a su vez, tienen menor duración que los hiatos, coincidiendo con los resultados de algunos estudios anteriores (Aguilar, 2003; Hualde et al. 2008; Gubian et al. 2015); iii) en secuencias en que por lo menos una de las vocales es átona, la preferencia es por la monoptongación.

Palabras clave: secuencias vocálicas, frontera de palabra, español mexicano

This study is an analysis of vowel sequences in Mexican Spanish data, with respect to the three processes across word boundaries: monophthong formation, reduction, and hiatus maintenance. Based on the analysis of the duration in the sequences, F1 and F2 values, and stress, the results show that (i) there was a high number of monophthongs, followed by hiatuses and, lastly, reductions (ii) monophthongs are shorter in duration than reductions, and these in turn are shorter than hiatuses, corroborating previous studies (Aguilar, 2003; Hualde et al. 2008; Gubian et al. 2015) and (iii) in sequences with at least one unstressed vowel, the tendency is monophthongization.

Keywords: vowel sequences, word boundary, Mexican Spanish

1. INTRODUCCIÓN

En los estudios sobre secuencias de vocales en español se comprueba una tendencia a la tautosilabación de las vocales en frontera de palabras (Jenkins, 1999, Alba, 2006; Bacovik, 2006; Hualde et al., 2008; Smith et al., 2008; Barberia, 2006, 2012; Alcoholado Feltstrom, 2017). La contracción puede llevar a: la monoptongación, a partir de la elisión de una de las vocales (como por ejemplo en m[a]cuerdo para *me acuerdo*); la reducción de las vocales altas o medias, caracterizando la formación de un diptongo (m[ja]cuerdo) o lo que algunos autores denominan de sinalefa (m[ea]cuerdo).

Los trabajos observan las secuencias de vocales desde diferentes perspectivas y en distintas variedades. Alba (2006), por ejemplo, analiza en datos /aV/ del español hablado en Nuevo México la variabilidad entre hiatos/no hiatos a partir de la calidad vocálica y de la tonicidad, pero también añade otros elementos como el tipo de sílaba y el

tipo de palabra. Sus resultados muestran un mayor número de resoluciones de hiatos frente al mantenimiento, en todos los aspectos llevados a cabo. Hualde et al. (2008), en un análisis de la contracción en secuencias de vocales no altas en datos del español peninsular, señalan la altura de las vocales y la duración como características imprescindibles de la contracción en las secuencias. Barberia (2012) observa las posibles situaciones de resolución de los hiatos para el español peninsular, entre ellos el de frontera de palabras, y concluye que la contracción siempre requiere la reducción de la duración de la secuencia.

Smith et al. (2008) realiza un análisis en datos del español mexicano hablado en Veracruz, y sus resultados presentan la preferencia por no mantener el hiato en 94% de los datos. Los autores también observan que la duración de hiatos es más larga en comparación con las contracciones, pero que el acento parece no afectar las combinaciones vocálicas.

A partir de estas observaciones, el presente estudio analiza las secuencias en frontera a partir de la calidad vocálica, duración y tonicidad. El objetivo es verificar si hay preferencia por alguno de los tres procesos en los datos del español mexicano y examinar cuáles son las condiciones para los resultados obtenidos. Se parte de la hipótesis de que, así como en los trabajos anteriores de algunas variedades del español, los datos siguen la tendencia en el idioma a la contracción en los encuentros vocálicos en el límite de las palabras.

2. METODOLOGÍA

El experimento fonético realizado para ese estudio se basa en datos recopilados de cuatro estudiantes mexicanos de las ciudades de León y Oaxaca (informantes del sexo femenino), Puebla y Toluca (del sexo masculino), todos con edad entre 25 y 30 años. Se etiquetó a los informantes como LGua1, Oaxa1, Puebla1 y Tolu1. La recopilación de los datos se hizo por medio de entrevistas con los informantes, y a partir de algunos temas específicos: a) hablar sobre la ciudad donde nacieron; b) hablar sobre viajes; c) estudios. Los temas propuestos tuvieron como objetivo crear un ambiente relajado para responder a las preguntas, con la finalidad de recopilar los datos de habla lo más espontáneo. Para cada tema hubo un conjunto de 5 preguntas, y en general las entrevistas tuvieron una duración de 20 a 25 minutos.

Para las grabaciones se utilizó un micrófono unidireccional y un interfaz Scarlett 2.0. Los datos fueron grabados en el software Audacity, en una frecuencia de muestreo de alta calidad (44100Hz). Todas las entrevistas se hicieron en un despacho de la Universidad de Salamanca, en España. Considerando posibles influencias del español salmantino en el habla de los mexicanos, hay que resaltar que todos los estudiantes seleccionados habían llegado unos dos meses antes de las entrevistas.

El análisis acústico se hizo por medio del software libre Praat, versión 6.0.53. Con el programa se han analizado los valores de F1 y F2 en 15%, 50% y 85% del espacio temporal de los encuentros vocálicos y la duración de las secuencias. El entorno RStudio (interfaz para R) fue utilizado para las pruebas y análisis de estadística inferencial y las construcciones de los gráficos. Los paquetes de R utilizados fueron el ggplot, dplyr, MASS y phonTools.

La normalización de las variables continuas se hizo por medio del método Z-score/Lobanov. Las diferencias entre los diptongos y los hiatos se hicieron mediante las pruebas ANOVA y post-hoc Fisher y la

distancia euclidiana, utilizados por Hualde et al. (2008).

3. ANÁLISIS DE LOS DATOS

Los datos analizados del español mexicano presentan un mayor número de monoptongos en las producciones ($f_i = 0.56$) que los otros dos procesos, corroborando la tendencia a la contracción de las secuencias en el idioma (Alba, 2006; Hualde et al., 2008; Smith et al., 2008; Barberia, 2012, entre otros). Los diptongos/sinafeas tuvieron una frecuencia mucho menor que los hiatos ($f_i = 0.31$ para hiatos y $f_i = 0.13$ para reducciones). La figura 1 presenta la distribución de las secuencias que fueron significativas dentro de los tres procesos:

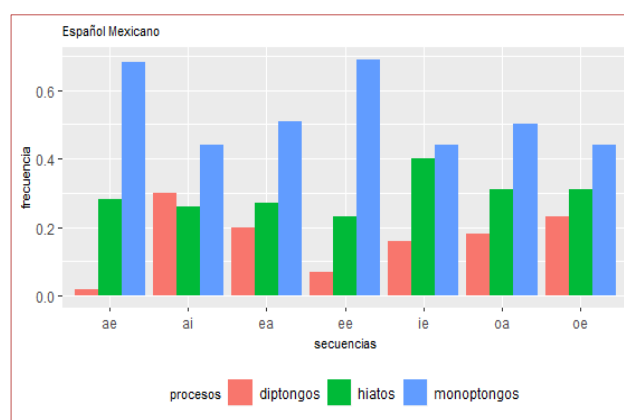


Figura 1: Distribución de secuencias con un número de casos significativos ($p < 0.05$).

3.1. Calidad de la vocal y posición en la secuencia

En los procesos de elisión de las secuencias /ae/, /ai/ y /oa/, las vocales /e, i, o/ se eliden frente a la vocal baja /a/.

La secuencia /ai/ monoptongada aún tuvo un promedio de 0.5 de coalescencia, resultando [e]. Esta conclusión se debe al hecho de que los valores formánticos en 15%, 50% y 85% de /ai/ en coalescencia se asimilan a los valores de los monoptongos [e] (con un rango de variabilidad de F1 entre 435Hz y 512Hz y de F2 entre 1765Hz y 2114Hz). En /ae/, /ea/, /oa/ y /oe/ monoptongados también hubo una coalescencia en los formantes, pero con una frecuencia de casos mucho menor. En ese sentido, los resultados coinciden con los hallazgos de coalescencia en Smith et al. (2008), también en el español mexicano.

En cambio, la elisión de una de las vocales en las secuencias /ea/, /oe/ y /ei/ es variable: para /ea/, la elisión de la primera vocal tuvo una frecuencia de $f_i = 0.53$ (IC = 0.37, 0.68) y para la segunda vocal $f_i = 0.46$ (IC = 0.31, 0.62); para /oe/, la elisión de /o/ $f_i = 0.63$ (IC = 0.48, 0.76) y de /e/ $f_i = 0.37$ (IC = 0.23,

0.51); para /ei/, la elisión de /e/ $f_i = 0.28$ (IC = 0.08, 0.58) y de /i/ $f_i = 0.71$ (IC = 0.41, 0.91). Se puede observar que los intervalos de confianza muestran la posibilidad de frecuencia de 0.5 para esas combinaciones, es decir, hay probabilidad de que la elisión de la segunda vocal ocurra en 50% de los casos. Así pues, el contexto /ea/ se contrapone a los resultados encontrados en Hualde et al. (2008) para el español peninsular, cuyo estudio describe la posibilidad de eliminar solamente la primera vocal. Las figuras 2 y 3 ejemplifican la elisión de la primera (2) y segunda (3) vocal a partir de la producción de uno de los informantes del español mexicano:

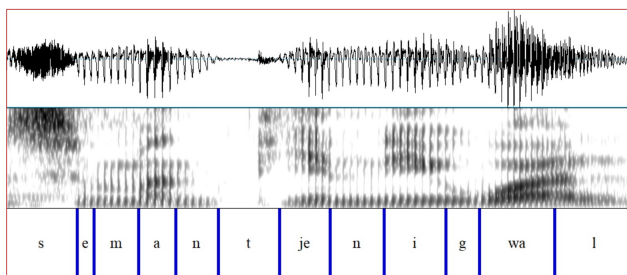


Figura 2: Onda sonora, espectrograma y TextGrid en Praat de ‘se mantiene igual’, con elisión de la primera vocal (informante Pueb1, valores en 15%, 50% y 85% fueron, en Hertz: $f_{1.1} = 348$; $f_{2.1} = 347$; $f_{1.3} = 343$; $f_{2.1} = 2430$; $f_{2.2} = 2367$; $f_{2.3} = 2321$).

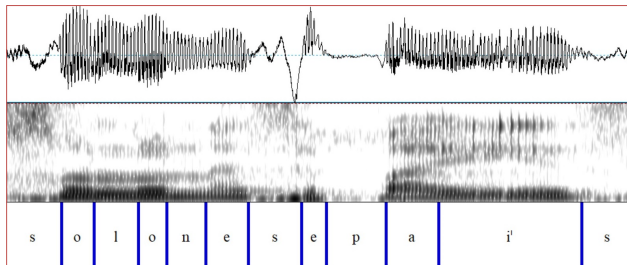


Figura 3: Onda sonora, espectrograma y TextGrid en Praat de ‘sólo en ese país’, con elisión de la segunda vocal (informante LGua1, valores en 15%, 50% y 85% fueron, en Hertz: $f_{1.1} = 411$; $f_{2.1} = 426$; $f_{1.3} = 408$; $f_{2.1} = 1200$; $f_{2.2} = 1210$; $f_{2.3} = 1237$).

Respecto a las reducciones de vocales medias, hubo muy pocos contextos en que las vocales /e, o/ se convirtieron en deslizadas, resultando en sinalefas ($f_i = 0.09$). En la mayoría de los datos en contexto de vocales medias hubo contracción mediante la monoptongación ($f_i = 0.65$).

3.2. Duración de las secuencias

Las pruebas ANOVA y Tukey presentan diferencias significativas entre monoptongos, reducciones y hiatos ($p < 0.01$), constatando la importancia de la duración como pista prosódica en la caracterización de los procesos (monoptongos: $m = 57.9$, $sd = 18.0$); diptongos: $m = 77.9$, $sd = 15.5$);

hiatos: $m = 182.2$, $sd = 49.7$, valores en milisegundos). Los valores de diptongos y monoptongos se encuentran más distantes de los valores de hiatos, como observado en gráfico de caja de bigotes:

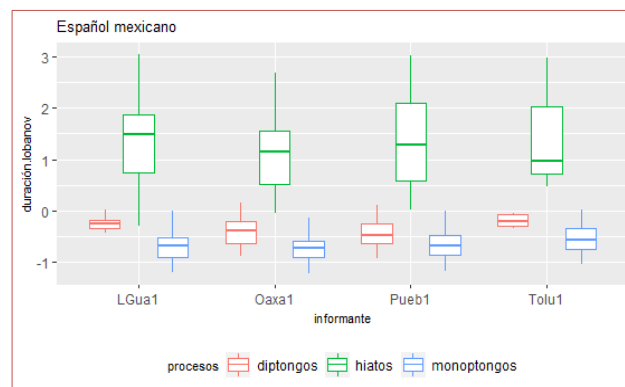


Figura 4: Datos normalizados (Z-score/Lobanov) de la duración de las secuencias dentro de los tres procesos.

Así como se describe en Aguilar (1999; 2003), Hualde et al. (2008), Gubian et al. (2015), la duración también favorece la distinción diptongos/hiatos, aunque no sea el único parámetro para el contraste. Los datos reducidos señalan una menor duración de las vocales /i/, /e/ y /o/ formando de esta manera un diptongo o una sinalefa con la vocal adyacente: con /i/ se realizan fonéticamente [aj] y [je] (éste último derivado no solo de /ie/ sino también de /ee/), con /e/: [æ], [ɛ̃], y con /o/: [õ], [ɔ̃].

3.3. Tonicidad

En las secuencias de dos vocales átonas, todos los informantes produjeron un mayor número de monoptongos ($f_i = 0.87$). Los contextos V'V y VV' también ocasionaron más monoptongación que hiatos o reducciones, con frecuencias de $f_i = 0.73$ y $f_i = 0.52$, respectivamente. Los resultados divergen de Alba (2006), en su análisis con datos del español de Nuevo México, pero están de acuerdo con la investigación de Smith et al. (2008) sobre el español de Veracruz. La figura 4 presenta la distribución de cada combinación por tonicidad:

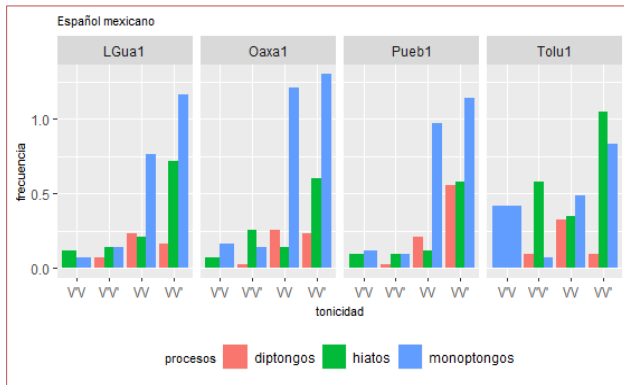


Figura 5: Distribución de secuencias del español mexicano por tonicidad ($p < 0.05$).

4. CONCLUSIONES

Las secuencias de vocales analizadas en frontera siguen un proceso gradual de contraerlas en una única sílaba. No obstante, en la resilabificación los datos del español mexicano son más susceptibles a la monoptongación (elisión y algunos casos de coalescencia), diferente de la posibilidad de producción de diptongos o sinalefas observados para otras variedades del español (Aguilar, 2003; Hualde et al., 2008).

Duración y calidad de la vocal son los dos factores principales para la caracterización de los procesos. La tonicidad muestra resistencia a la contracción solamente cuando ambas vocales son tónicas. Los resultados se aproximan a los de Smith et al. (2008), que también analizó una variedad mexicana del español. Por fin, los datos reafirman el antihiatismo del español en contextos de frontera.

5. BIBLIOGRAFÍA

- Aguilar, L. (1999). Hiatus and diphthong: Acoustic cues and speech situation differences. *Speech Communication*, 28(1), 57-74. [https://doi.org/10.1016/S0167-6393\(99\)00003-5](https://doi.org/10.1016/S0167-6393(99)00003-5)
- Aguilar, L. (2003). Effects of prosodic and segmental variables on vowel sequences pronunciation in Spanish. *Proceedings of 15th International Congress of Phonetic Sciences*, editado por Solé, M, Recasens D. y Romero. J. Barcelona.
- Alba, M. C. (2002). ¿Cómo se llega a l'escuela? : a study of usage effects on hiatus resolution in New Mexican Spanish. *Paper presented at NWAV-XXXI*. Stanford University.
- Alba, M. C. (2006). Accounting for Variability in the Production of Spanish Vowel Sequences. *Proceedings of the 9th Hispanic Linguistics Symposium*, 273-285. Somerville, MA: Cascadilla Proceedings Project.
- Alcoholado Feltstrom, A. (2017). *Fenómenos fonéticos y antihiatismo en hablantes cultos de español*. (Tesis doctoral). Universitat Jaume I, España.
- Bacovik, E. (2006). Hiatus resolution and incomplete identity. In: Martínez-Gil, F. & Colina, S. (Orgs) *Optimality-Theoretic Studies in Spanish Phonology*, *Linguistics Today*, 99, 62-73.
- Barberia, I. (2006). Effects of VV-sequence deletion across word boundaries in Spanish. *Proceedings of ISCA Tutorial and Research Workshop on Experimental Linguistics*, 28-30 August 2006, Athens, Greece. Recuperado de https://www.isca-speech.org/archive/exling_2006/papers/ex16_085.pdf
- Barberia, I. (2012). *The gradient nature of hiatus resolution in Spanish*. (Tesis doctoral). Universidad de Deusto, España. Recuperado de <https://dialnet.unirioja.es/servlet/tesis?codigo=118847>
- Gubian, M., Torreira, F., Boves, L. (2015). Using Functional Data Analysis for investigating multidimensional dynamic phonetic contrasts. *Journal of Phonetics*, 49, 16 - 40. <https://doi.org/10.1016/j.wocn.2014.10.001>
- Hualde, I., Simonet, M., Torreira, F. (2008). Postlexical contraction of nonhigh vowels in Spanish, *Lingua*, 118, 1906-1925. <https://doi.org/10.1016/j.lingua.2007.10.004>
- Jenkins, D. L. (1999). Hiatus resolution in Spanish: phonetic aspects and phonological implications from Northern New Mexico. (Tesis doctoral). University of New Mexico, USA.
- Smith, J. M., Flores, T., & Gradoville, M.S. (2008). *An analysis of vowels across word boundaries in Veracruz, Mexican Spanish*. Indiana University Bloomington, Volume 8. Recuperado de <https://www.semanticscholar.org/paper/An-analysis-of-vowels-across-word-boundaries-in-Smith-Flores/406571f38f8de7a0e11d556521ce89bb8ea13251>

PROSODIA DEL ASTURIANO CENTROMERIDIONAL: RASGOS CARACTERÍSTICOS Y ANOTACIÓN EN ASTUR_ToBI

Carmen Muñiz Cachón¹, Paolo Roseano^{2,3}

¹Universidad de Oviedo, ²Universitat de Barcelona, ³University of South Africa
carmu@uniovi.es, paolo.roseano@ub.edu

ABSTRACT

En los últimos veinte años, la prosodia de los distintos dialectos de la lengua asturiana ha sido el objeto de diferentes estudios que se han centrado sobre todo en dos modalidades oracionales: la declarativa de foco ancho y la interrogativa total neutra. Este trabajo presenta los primeros resultados de un análisis de los rasgos prosódicos de un conjunto más amplio de modalidades oracionales. Los resultados del estudio permiten describir de forma más completa los rasgos prosódicos del asturiano y, al mismo tiempo, permiten esbozar una primera versión del Astur_ToBI, un sistema de transcripción de la prosodia del asturiano basado en las premisas del modelo métrico-autosegmental.

Palabras clave: asturiano, prosodia, entonación, Astur_ToBI

In the last twenty years, the prosody of Asturian has been the object of different studies that have focused mainly on two sentence-types: broad-focus statements and information-seeking yes-no questions. This paper presents the first results of an analysis of the prosodic features of a broader set of sentence-types. The results of the study permit a more complete description of the prosodic features of Asturian and, at the same time, allow sketching the first version of Astur_ToBI, an annotation system of Asturian prosody based on the autosegmental-metrical model.

Keywords: Asturian, prosody, intonation, Astur_ToBI

1. INTRODUCCIÓN

Este trabajo tiene dos objetivos principales. El primero de ellos es el de describir los rasgos prosódicos —tanto entonativos como durativos— de la variedad centromeridional asturiana de Mieres, que en trabajos anteriores (v. Sección 1.1) ha sido parcialmente analizada. El segundo es el de proponer una primera versión de un sistema de transcripción de la prosodia del asturiano, llamado Astur_ToBI, desarrollado en el marco teórico métrico-autosegmental (v. Secciones 1.2 y 2.5).

1.1. La prosodia del asturiano

Son relativamente recientes los estudios sobre la prosodia de las lenguas de Asturias, salvando una nota muy interesante de Canellada (1984) y algunos datos muy generales presentados en ciertas monografías dialectales. En los estudios actuales, realizados generalmente en el marco del proyecto

AMPER, se ha analizado un conjunto importante de enunciados de las cinco zonas dialectales, emitidos por cinco informantes de cada zona, y se ha medido la F0, intensidad y duración de cada una de las sílabas de los enunciados. Se han analizado estos valores en dos modalidades oracionales: declarativas neutras y en oraciones interrogativas totales neutras.

El cómputo de la duración confirmó un comportamiento muy distinto en ambas modalidades, en asturiano, particularmente destacado en el habla de Mieres. Paralelamente se aprecia en esta variedad centromeridional un contorno entonativo (F0) con una configuración semejante —o muy próxima— en enunciados declarativos neutros y en interrogativos totales neutros, con un acento nuclear H+L* seguido de idéntico tono de frontera bajo L%. Desde los primeros trabajos descriptivos y comparativos de esta zona (López Bobo et al., 2005; Díaz Gómez et al., 2007; Congosto et al., 2010) se destacó el papel relevante que juega la duración en la discriminación

de modalidades oracionales. Corroborar estas afirmaciones un estudio específico de la duración en relación con los límites sintagmáticos (Muñiz Cachón, 2017) que cuantifica el incremento de la duración en la frontera final ligado a las diferencias entre modalidades.

Son varias las descripciones fonéticas realizadas de las distintas variedades lingüísticas de Asturias, sin embargo, apenas se han realizado sistematizaciones fonológicas de sus resultados (Alvarellos et al., 2011).

1.2. Los sistemas ToBI para la anotación de la prosodia

En la actualidad, el marco teórico más influyente para los estudios prosódicos es posiblemente el métrico-autosegmental (de aquí en adelante AM). Desarrollado inicialmente en la década de los ochenta por Pierrehumbert (1980) a partir de las propuestas de Goldsmith (1979), el modelo AM se ocupa principalmente de fenómenos entonativos y afirma que los contornos pueden entenderse como secuencias de dos tipos principales de unidades: los acentos tonales (que se asocian con las sílabas tónicas) y los tonos de frontera (que se asocian con los límites de las unidades sintácticas).

Sobre los fundamentos teóricos del modelo AM, se ha desarrollado un conjunto de convenciones de anotación prosódica, de carácter más bien fonológico, conocidas como *Tones and Break Indices* (abreviado como *ToBI*). A pesar de la reconocida necesidad de utilizar un sistema de transcripción de la entonación que sea lo más universal y compartido posible (Roseano y Fernández Planas, 2013; Hualde y Prieto, 2016), los sistemas de transcripción actualmente en uso son específicos para cada lengua (para una reseña, v. Jun, 2005, 2014; v. Hualde y Prieto, 2016).

Para las lenguas románicas, en la última década se ha elaborado un conjunto de sistemas de transcripción tendentes a converger en sus anotaciones (v. Frota y Prieto, 2015).

En este trabajo se pretende establecer la base para el sistema de transcripción de la prosodia del asturiano, llamado *Astur_ToBI* (v. Sección 2.5). Los antecedentes directos de este sistema de transcripción se encuentran en el trabajo de Alvarellos et al. (2011).

Un aspecto crucial que cabe destacar es que los sistemas ToBI suelen prever solo un sistema de transcripción para uno de los elementos que constituyen la prosodia, es decir la entonación,

mientras que obvian la duración, ya que esa última no suele desempeñar un papel destacado. Puesto que en el asturiano de Mieres la duración tiene una función distintiva, en cuanto que permite diferenciar modalidades oracionales (v. Sección 1.1), el sistema *Astur_ToBI* tiene que incluir los aspectos durativos. Para la elaboración de un sistema de anotación de la duración, los antecedentes más cercanos en un marco AM son las propuestas de Roseano y Fernández Planas (2013) así como de Muñetón, Díaz y Dorta (2018).

2. MÉTODO

2.1. Informantes

Se ha entrevistado a cuatro hablantes nativos de la variedad asturiana de Mieres: dos hombres y dos mujeres de edad comprendida entre los 23 y los 58 años. Todos ellos han nacido y residen en esta localidad asturiana.

2.2. Elicitación de datos

Los datos se han obtenido mediante la técnica del *Discourse Completion Task* (Vanrell, Feldhausen y Astruc, 2018), que consiste en plantear a los entrevistados una situación comunicativa en la que tienen que interactuar con un interlocutor y dirigirse a él según las instrucciones, como la que se ejemplifica en (1).

- (1) ENTREVISTADOR: *Entres nuna tienda y preguntes al tenderu si tien mermelada.*
RESPUESTA ESPERADA: *¿Tien mermelada?*

Para este trabajo se ha adaptado al asturiano el *Discourse Completion Task* que se ha utilizado en el marco del *Atlas interactivo de la entonación del español* (Prieto y Roseano, 2009-2013). La tarea prevé 32 situaciones comunicativas en total, que corresponden a otros tantos tipos de oración. En este trabajo se van a presentar los resultados relativos a siete modalidades oracionales básicas (v. Sección 3).

2.3. Grabaciones

Las grabaciones se han llevado a cabo en un despacho silencioso de la Universidad de Oviedo utilizando una grabadora Marantz PMD 671 y un micrófono de diadema AKG C520.

La persona que ha realizado las entrevistas es hablante nativa de la variedad asturiana de Mieres y

conocida por los entrevistados. Estaba presente en las grabaciones la primera autora de este trabajo, que también es hablante nativa de la misma variedad.

Antes de suministrar el *Discourse Completion Task*, la entrevistadora ha mantenido una conversación de algunos minutos sobre temas variados con cada uno de los entrevistados, a fin de que se familiarizaran con el entorno de grabación.

Cada grabación ha tenido una duración aproximada de 45 minutos (incluyendo la fase de familiarización).

2.4. Corpus

En conjunto, el corpus analizado para este trabajo incluye 128 enunciados (4 informantes * 32 situaciones conversacionales).

2.5. Anotación prosódica: Astur_ToBI

Tal y como se puede apreciar en la Figura 1, para la anotación de la prosodia de la variedad de Mieres se utilizan, además de un *tier* con transcripción segmental, un *tier* donde aparecen los *Break Indices*, también un *tier* –el tercero– con la anotación de la duración y, por último, un *tier* con la transcripción entonativa. En los apartados siguientes se van a resumir las características más destacadas de los sistemas para la transcripción de la entonación (2.5.1) y de la duración (2.5.2).

2.5.1. Anotación de la entonación

Tal y como es común en los sistemas ToBI, para la anotación de la entonación se han utilizado tres símbolos para representar las dianas tonales: H para los tonos altos (del inglés *high*), L para los tonos bajos (del inglés *low*), y !H que indica un tono medio. El símbolo % indica que un tono está asociado con la frontera inicial o final de un enunciado (por ejemplo, entre otras posibilidades, L% indica un tono bajo asociado con las sílabas finales, mientras que %LH indica un tono ascendente asociado con la frontera inicial de un enunciado). A su vez, el símbolo - indica que un tono es asociado con los límites de la frase intermedia (como, por ejemplo, en ¡H- o L-). Para los acentos tonales, un asterisco * señala el tono asociado con la sílaba tónica (por ejemplo, L+H*).

2.5.2. Anotación de la duración

Los conocimientos de los que disponemos actualmente sobre la prosodia de la variedad

asturiana de Mieres (v. Sección 1.1) permiten afirmar que una de las diferencias prosódicas entre las oraciones declarativas de foco ancho y las interrogativas absolutas es que las segundas presentan un alargamiento considerable de la sílaba átona final, mientras que en las primeras ese alargamiento no se observa. Puesto que este contraste durativo se realiza en la sílaba final, se trata de un fenómeno relacionado con las fronteras prosódicas. Por esa razón, se ha decidido anotar con G% el fenómeno de alargamiento de la sílaba final, y con P% su ausencia. El símbolo % es el que se utiliza de manera habitual en los sistemas ToBI para los fenómenos que se asocian con las fronteras prosódicas. Las letras G y P, tal y como se explica en Roseano y Fernández Planas (2013), hacen referencia a las iniciales de palabras que en distintas lenguas románicas indican tamaño grande (esp. *grande*, cat. *gran*, ita. *grande*, fri. *grant*, fr. *gros*, ast. *grande*) o pequeño (esp. *pequeño*, cat. *petit*, ita. *piccolo*, fri. *piçul*, fr. *petit*, ast. *pequeñu*).

3. RESULTADOS

5.1. Declarativas de foco ancho

Las declarativas de foco ancho (Figura 1) están caracterizadas por una configuración nuclear que incluye una bajada de F0 entre la sílaba pretónica y la tónica (H+L*), seguida por un tono de frontera bajo (L%). Desde el punto de vista de la duración, no se observa alargamiento de las sílabas átonas finales (P%).

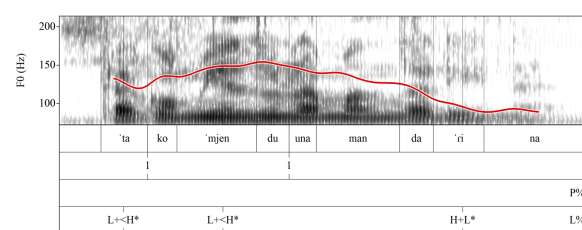


Figura 1: Declarativa de foco ancho *Ta comiendo una mandarina* 'Está comiendo una mandarina'.

5.2. Enumeraciones

La característica más destacada de las enumeraciones declarativas (Figura 2) es la de presentar una configuración L+H* H!H- en los primeros elementos de la serie hasta el penúltimo.

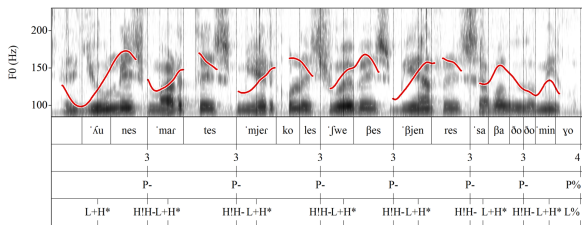


Figura 2: Enumeración declarativa *Lunes, martes, miércoles, jueves, viernes, sábado, domingo* 'Lunes, martes, miércoles, jueves, viernes, sábado, domingo'.

5.3. Declarativas de obviedad

Las oraciones declarativas de obviedad (Figura 3) presentan un acento nuclear bajo (L*) seguido por un movimiento complejo de tipo ascendente-descendente en correspondencia con la frontera final (HL%). La sílaba final presenta alargamiento (G%).

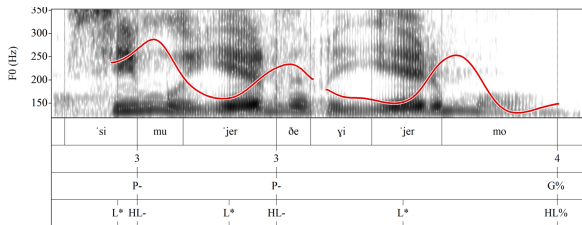


Figura 3: Declarativa de obviedad *Sí, mujer, de Guillermo* 'Sí, mujer, de Guillermo'.

5.4. Interrogativas totales neutras

Las interrogativas totales neutras (Figura 4) presentan un acento inicial del tipo L*+H, mientras que en la configuración nuclear aparece un acento descendente (H+L*) y un tono final bajo (L%). A diferencia de lo que se observa en otras modalidades oracionales, en las interrogativas totales la sílaba final es alargada (G%).

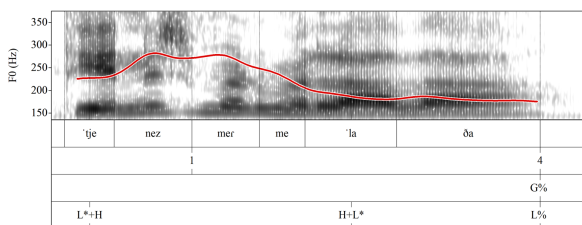


Figura 4: Interrogativa total neutra *¿Tienes mermelada?* '¿Tienes mermelada?'.

5.5. Órdenes

Las órdenes (Figura 5) están caracterizadas, desde el punto de vista entonativo, por una configuración nuclear descendente del tipo H+L* L%.

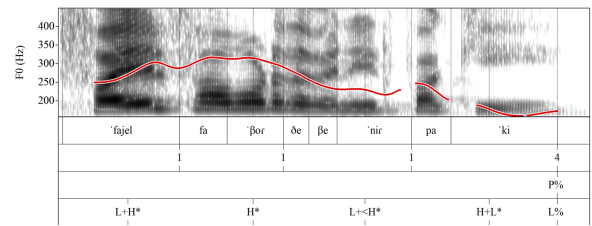


Figura 5: Orden *¡Haz el favor de venir para aquí!* '¡Haz el favor de venir para aquí!'.

5.6. Ruegos

En los ruegos como el de la Figura 6, además de marcas morfológicas (modo imperativo) y léxicas (partícula *ho*, v. Muñiz Cachón, 2016), se ha observado la presencia del contorno entonativo L+H* L%. En los casos en que está presente la partícula *ho*, se observa alargamiento final (G%).

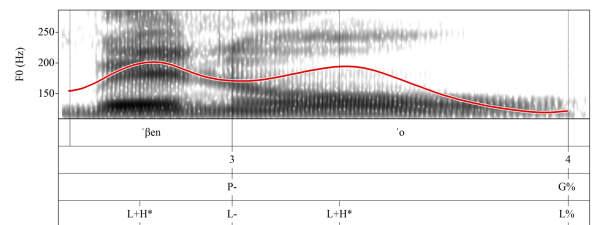


Figura 6: Ruego *¡Ven, ho! ¡Ven!*

5.7. Vocativos

Los vocativos (Figura 6) presentan un patrón caracterizado por una subida inicial en las primeras sílabas del nombre. Como primera aproximación y en espera de futuros análisis más detallados, transcribimos ese movimiento como [%LH]. En correspondencia con la sílaba tónica se observa una bajada de F0 (que interpretamos como un acento H+L*) y, a continuación, una bajada menos marcada (que proponemos relacionar con la presencia de un tono final bajo L%). Destaca la presencia de alargamiento final (G%) y, en algunos casos, de la partícula *ho* con función atenuativa.

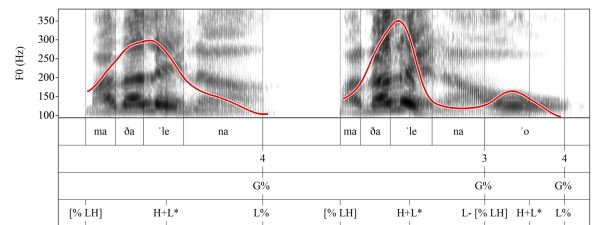


Figura 6: Serie de vocativos *¡Madalena... Madalena, oh!*.

4. DISCUSIÓN Y CONCLUSIONES

Este trabajo ha descrito los primeros resultados de un análisis de los rasgos prosódicos (tanto

entonativos como durativos) de la variedad asturiana de Mieres a partir de datos recogidos con un *Discourse Completion Task* diseñado para elicitar enunciados diferentes modalidades oracionales. Ha destacado la relevancia de la duración en la discriminación de esos modos oracionales. Además, ha formulado una primera propuesta de un sistema de transcripción de la prosodia del asturiano llamado Astur_ToBI.

5. BIBLIOGRAFÍA

- Alvarellos Pedrero, M., Muñiz Cachón, C., Díaz Gómez, L., y González Rodríguez, R. (2011). La entonación en las variedades lingüísticas de Asturias: estudio contrastivo. *RILL*, 17, 111-120.
- Canellada, M.J. (1984). Notas de la entonación asturiana. *Lletres Asturianes*, 10, 23-26.
- Congosto Martín, Y., Díaz Gómez, L., Viejo Lucio-Villegas, M., y Fernández Rodríguez, R. (2010). Estudio contrastivo de la entonación del castellano de Don Benito y del asturiano de Mieres en el marco del Proyecto AMPER. En Iliescu, M., Siller-Runggaldier, H., y Danler, P. (Eds.), *Actes du XXVe CILPR Congrès International de Linguistique et de Philologie Romanes. Innsbruck 3 à 7 de septembre de 2007* (vol. 2, p. 53-66). Berlin / New York: De Gruyter.
- Díaz Gómez, L., Muñiz Cachón, C., López Bobo, M.J., González Rodríguez, R., y Cuevas Alonso, M. (2007). La entonación en dos zonas rurales del centro de Asturias. *Archivum*, LVI, 263-289.
- Frota, S. y Prieto, P. (Eds.) (2015). *Intonation in Romance*. Oxford: Oxford University Press.
- Goldsmith, John. (1979). *Autosegmental Phonology*. New York: Garland.
- Hualde, J.I. y Prieto, P. (2016). Towards an International Prosodic Alphabet (IPrA). *Laboratory Phonology: Journal of the Association for Laboratory Phonology*. Disponible en: <http://doi.org/10.5334/labphon.11>
- Jun, S.A. (Ed.) (2005). *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press.
- Jun, S.A. (Ed.) (2014). *Prosodic Typology II: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press.
- López Bobo, M.J., González Rodríguez, R., Cuevas Alonso, M., Díaz Gómez, L., y Muñiz Cachón, C. (2005). Rasgos prosódicos del centro de Asturias: comparación Oviedo-Mieres. *Estudios de Fonética Experimental*, 14, 167-199.
- Muñetón Ayala, M., Díaz, Ch., y Dorta, J. (2018). La duración en oraciones sin expansión en la voz femenina de dos países fronterizos: Colombia (Bogotá-Medellín) y Venezuela (Caracas-Mérida). *Literatura y lingüística*, 37, 401- 423.
- Muñiz Cachón, C. (2016). Aspectos prosódicos y pragmáticos de la partícula *ho*. *Lletres Asturianes*, 115, 73-88.
- Muñiz Cachón, C. (2017). Implicaciones de la duración en la prosodia: asturiano y castellano del centro de Asturias. *Estudios de Fonética Experimental*, 26, 223-243.
- Pierrehumbert, J.B., (1980). *The Phonetics and Phonology of English Intonation*. (Tesis doctoral, Massachusetts Institute of Technology)
- Prieto, P., y Roseano, P. (Coords). (2009-2013). *Atlas interactivo de la entonación del español*. <http://prosodia.upf.edu/atlasentonacion/>
- Roseano, P. y Fernández Planas, AM. (2013). Transcripció fonètica i fonològica de l'entonació: una proposta d'etiquetatge automàtic. *Estudios de Fonética Experimental*, 22, 275-332.
- Vanrell, M.M., Feldhausen, I., y Astruc, L., (2018). The Discourse Completion Task in Romance Prosody Research: Status quo and Outlook. En: Feldhausen, I., Fliessbach, J., y Vanrell, M.M. (Eds.), *Methods in Prosody: A Romance Language Perspective* (p. 191-227). Berlin: Language Science Press.

VARIACIÓN SOCIAL EN LA REALIZACIÓN DE LA FRICATIVA /S/ EN EL ESPAÑOL DE ANTIOQUIA, COLOMBIA

Diana Muñoz-Builes, Wendy Elvira-García

Pontificia Universidad Católica de Chile, Universidad Nacional de Educación a Distancia
dmmunoz2@uc.cl, welvira@flog.uned.es

ABSTRACT

Este trabajo presenta las características acústicas del sonido [s] en el departamento de Antioquia, Colombia. Específicamente, se analizan las diferencias encontradas según la variable estrato social. En total, se han analizado 943 realizaciones de /s/, producidas por 32 colaboradores. Los resultados preliminares sugieren que las /s/ en posición de ataque se realizan como más apuntadas en el estrato socioeconómico bajo en Medellín y como más sonorizadas en el estrato socioeconómico alto en Apartadó.

Palabras claves: fonética, fricativas, español

This work examines the acoustics of the of the sound [s] in two cities of Antioquia, Colombia. More specifically we analyze social differences. The study analyzes 343 [s] productions produced by 32 speakers. Preliminary results suggest that in onset position [s] is produced with more kurtosis in Medellín by low class speakers and as more voiced in Apartadó by high social class.

Keywords: phonetics, fricatives, Spanish

1. INTRODUCCIÓN

Son numerosos los estudios que apuntan que, para el español, la realización de /s/ es un marcador social. Así, características como la aspiración de la /s/, ya sea en ataque o en coda, la sonorización de /s/ intervocálica o la producción de /s/ a ciertas frecuencias, sirven como indicadores sociolingüísticos a los hablantes (Chappel, 2019a).

Tradicionalmente se considera que la aspiración de /s/ es propia de las clases más bajas, mientras que el mantenimiento de /s/ se asocia con un nivel socioeconómico más alto (Chappell, 2019b; Walker, García, Cortés y Campbell-Kibler, 2014).

Otros estudios plantean que la sonorización de las /s/ intervocálicas se interpreta en Costa Rica como indicador de menor status social aunque, a la vez, sus hablantes son considerados como más confiables, amables y más locales (Chappel, 2016). En Ecuador, por su parte, esta sonorización es un rasgo de mayor

status social y edad, en el grupo de las mujeres (García, 2019).

En Chile, la producción de /s/ con un centro de gravedad (CoG) y un apuntamiento del espectro especialmente altos es un marcador sociolingüístico: el CoG aumenta con el género femenino y el estrato socioeconómico medio, mientras que disminuye en el estrato socioeconómico más alto (Perdomo-Pinto y Sadowsky, 2019).

En el caso de Colombia, y más específicamente de Medellín, se han documentado diferencias en diversos niveles lingüísticos de acuerdo al estrato socioeconómico de los hablantes. Por ejemplo, en el nivel léxico se ha estudiado el “parlache”, léxico usado mayoritariamente por jóvenes de barrios populares y que ha permeado otras clases sociales (Castañeda, 2005) y, además, se han analizado las diferencias en el uso de las formas de tratamiento pronominales (Jang, 2010) y nominales (Castellano Ascencio, 2008) en razón al estrato sociocultural.

Pasando al campo de la fonética se han documentado diferencias diastráticas en la entonación de mujeres de Medellín, específicamente en la configuración nuclear, el rango tonal y los valores máximos (Muñoz-Builes, en evaluación).

Para el caso específico de /s/ en Colombia, los autores clásicos defendían que la aspiración era un marcador de clase. Así, por ejemplo, planteaban que “aunque /s/ final de sílaba se resiste al debilitamiento en las tierras altas colombianas (...) a veces /s/ se aspira en posición inicial de palabra intervocálica. La pronunciación de *nosotros* como *nojotros*, propia del español popular, está muy extendida en el centro de Colombia, incluso entre hablantes con un nivel cultural relativamente alto. (Flórez, 1951b: 29; Flórez, 1973: 82-3)” Citado en Lipski (2005: 232-233) [[subrayado nuestro].

Otra alusión a la producción de /s/, también realizada por Flórez (1951: 193) plantea que la “/s/ prevocálica inicial de palabra se aspira a menudo. Esta pronunciación no es habitual, salvo en estratos sociolingüísticos muy bajos (Flórez, 1951: 193)”

El objetivo de este trabajo es doble, por una parte, explorar la realización fonética de /s/ en dos ciudades de la región de Antioquia en Colombia (Medellín y Apartadó) y por otra, explorar si la variable de estrato socioeconómico tiene algún efecto en los datos.

2. MÉTODO

2.1. Informantes, puntos de encuesta y grabaciones

Se grabaron a 32 hablantes antioqueños, preestratificados en las variables: municipio, sexo (femenino, masculino), grupo etario (20-35; 36-65 años) y estrato socioeconómico (alto, bajo).

El estrato socioeconómico fue determinado mediante la ponderación de tres variables: 1) nivel educacional propio y de la pareja, 2) profesión u oficio propio y de la pareja y, 3) estrato socioeconómico del DANE (Departamento Administrativo Nacional de Estadística), que clasifica a las viviendas colombianas en seis estratos, donde 1 es el estrato más bajo y 6 es el estrato más alto. La inclusión de los datos de la pareja para la determinación del estrato sociocultural utilizada en

este estudio es una adaptación, para el contexto colombiano, de la propuesta de estratificación EMIS (Metodología ESOMAR Modificada para la Investigación Sociolingüística) realizada por Sadowsky (2012).

Los municipios puntos de encuesta fueron Medellín y Apartadó, cada uno de los cuales tiene características geográficas, sociales y dialectales particulares, entre las que destacan, la pertenencia de Medellín al superdialecto andino y de Apartadó al superdialecto costeño (Montes Giraldo, 1982).

Las entrevistas fueron realizadas en cada localidad, utilizando para la captura de datos una grabadora Zoom H4nPro y un micrófono de diadema, Audix Ht7. Las grabaciones se digitalizaron en PCM con una frecuencia de muestreo de 48 KHz.

2.2. Elicitación de datos

El corpus analizado hace parte del Mapa prosódico de Colombia (Muñoz Builes, en preparación). Específicamente, los datos de este estudio se han elicitado mediante un *Discourse Completion Task* (Blum-Kulka, 1980; Vanrell, Feldhausen y Astruc, 2018) que contenía veintitrés situaciones conversacionales, basadas en Prieto y Roseano (2010), las cuales respondían a diferentes actos de habla.

2.3. Análisis

Los datos fueron anotados a partir del análisis acústico en Praat (2019). Específicamente, se anotaron todas las /s/ producidas fonéticamente como [s], ya fuera en posición de ataque o de coda. Los datos acústicos de esas /s/ fueron extraídos a través del script Elvira-García (2019). Se analizaron los parámetros de duración, intensidad, pasos por cero normalizados, frecuencia de máxima intensidad del espectro, centro de gravedad del espectro, desviación estándar del espectro, coeficiente de apuntamiento del espectro (curtosis) y coeficiente de asimetría del espectro.

El análisis estadístico fue realizado a través de R (R Core Team). Específicamente, se realizó un análisis de modelos mixtos usando como factores fijos la ciudad, el Estrato, la Edad y el Sexo y como factores aleatorios el Hablante y el Ítem. Los modelos creados se testaron para buscar la significatividad del estrato a través de Likelihood Ratios usando el modelo creado contra el modelo sin la variable Estrato.

3. RESULTADOS

El siguiente apartado explora las diferencias en la producción de /s/ según el estrato de los informantes (3.1) y describe dichas producciones (3.2).

En los datos analizados en este estudio llaman la atención algunas producciones de [s], como por ejemplo la representada en la Figura 1. Una [s] apical, que presenta la mayor concentración de energía entre los 3600 Hz y los 4350 Hz, con una gran intensidad de ruido, con el inicio de frecuencia de ruido en 2700 Hz y con el centro de gravedad en 3810,70 Hz. Además, tiene una desviación estándar de 1053,86 Hz, una curtosis de 60,279 y un coeficiente de asimetría de 5,5. Estas producciones son, en opinión de las autoras, propias de hablantes de estrato bajo, sin embargo, esta percepción subjetiva debería corroborarse con estudios perceptivos.

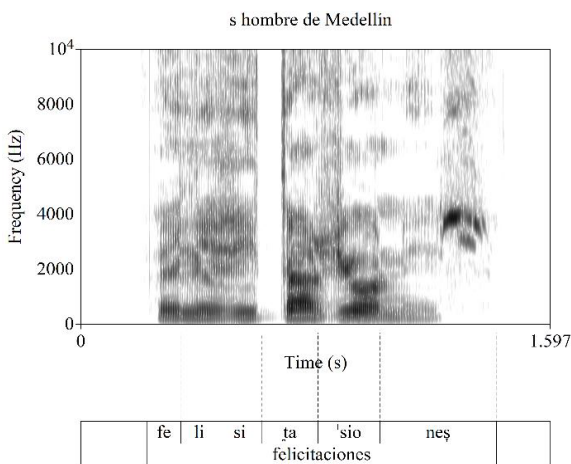


Figura 1: Espectrograma de /s/ producida en la palabra “felicitaciones”. Hablante masculino, 55 años, estrato bajo, Medellín.

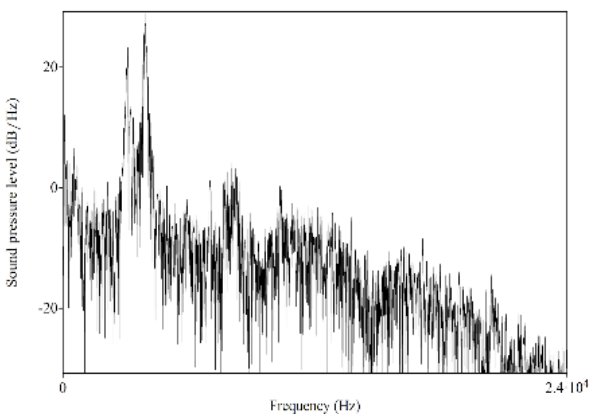


Figura 2. Espectro de [s] codal. Hablante masculino, 55 años, estrato bajo, Medellín.

3.1. Descripción cualitativa de las diferencias de la realización de /s/ en posición de ataque

Tal y como demuestran estudios anteriores, (Muñoz-Builes y Elvira-García, en prensa) la producción de /s/ difiere en Medellín y Apartadó, por ello los datos se presentan en esta sección separados por ciudad (Medellín, Apartadó), sexo y posición en la sílaba (solo para ataque por razones de espacio).

En las tablas 1 y 2 se presentan los valores acústicos descriptivos en ataque para hablantes femeninas (Tabla 1) y masculinos (Tabla 2) en Medellín y Apartadó.

En general, los valores acústicos de las [s] producidas son parecidos, tal y como probaban los resultados de la sección anterior. Sin embargo, en Medellín es reseñable la diferencia en el parámetro de la curtosis (apuntamiento del espectro) entre los dos estratos tanto en hombres como en mujeres. El valor alto de este parámetro en el estrato bajo, es un reflejo de la realización de la fricativa que se puede observar en la Figura 2, donde la concentración de energía es muy alta en un punto del espectro mientras que más baja en el resto.

Tabla 1: Valores acústicos de las producciones de /s/ en posición de ataque por parte de las hablantes femeninas del corpus.

		Mujeres (valores medios para /s/ en posición de ataque)			
		Estrato A		Estrato B	
Parámetro		Media	Desviación estándar	Media	Desviación estándar
A p a r t a d ó	Pico espectral (Hz)	6343	3862	7087	3355
	CoG (Hz)	6562	2540	6734	2199
	Desviación (Hz)	2329	745	2292	628
	Asimetría	-0,38	1,87	-0,37	1,65
	Kurtosis	3,99	6,29	3,64	9,44
	Ruido relativo	0,74	4,18	0,03	4,73
	Pasos por cero	98,39	62	98,59	62,43
duración (ms)	55,24	21	55,88	20,69	

Medellín	intensidad (dB)	59,92	6,06	59,33	5,52
	Pico espectral (Hz)	4254	2863	4590	3125
	CoG (Hz)	4947	2116	5158	2477
	Desviación (Hz)	2048	583	1918	736
	Asimetría	0,86	1,73	1,09	2,77
	Kurtosis	5,55	14,74	15,26	40,95
	Ruido relativo	1,64	4,87	0,99	3,91
	Pasos por cero	65,07	52,2	72,52	58,6
	duración (ms)	52,84	25,93	50,75	20,79
	intensidad (dB)	61,67	5,77	61	5,22

Tabla 2: Valores acústicos de las producciones de /s/ en posición de ataque por parte de los hablantes masculinos del corpus.

		Hombres (valores medios para /s/ en posición de ataque)			
		Estrato A		Estrato B	
Parámetro		Media	Desviación estándar	Media	Desviación estándar
Apartadó	Pico espectral (Hz)	2482	1853	5566	3321
	CoG (Hz)	3302	1614	6078	1626
	Desviación (Hz)	1672	775	2665	439
	Asimetría	2,84	2,94	0,01	0,94
	Kurtosis	29,9	58,2	0,15	2,49
	Ruido relativo	1,25	3,72	-1,78	5,7
	Pasos por cero	38,89	1853	96,39	51,63
	duración (ms)	47,88	21,11	55,67	16,08
	intensidad (dB)	59,51	5,9	61,1	4,12
	Medellín	Pico	2692	1741	2482

Medellín	espectral (Hz)				
	CoG (Hz)	3880	1470	3302	1674
	Desviación (Hz)	2084	541	1672	775
	Asimetría	1,38	1,59	2,84	2,94
	Kurtosis	6,52	18,37	29,9	58,20
	Ruido relativo	0,11	3,5	1,25	3,72
	Pasos por cero	35,05	28,96	38,89	46,94
	duración (ms)	44,6	13,58	47,88	21,11
	intensidad (dB)	58,62	5,52	59,51	5,9

3.1. Análisis estadístico de la variable estrato sociolingüístico

El estudio de la variable de estrato sociolingüístico ha arrojado datos significativos para pocos de los parámetros estudiados.

En el caso de las /s/ producidas en coda, ningún parámetro resulta significativo, es decir, no hay diferencias entre las producciones de /s/ en posición codal según el estrato de los hablantes.

En el caso de las /s/ producidas en posición de ataque, en cambio, algunos de los parámetros sirven para distinguir el estrato.

Tal y como se podía predecir a partir de los datos de la sección anterior, en el caso de las /s/ producidas en Medellín, los espectros del estrato bajo son significativamente más apuntados (efecto + 10.158, error 5.196, $p < 0.06$). Además, las /s/ producidas por hablantes de estrato bajo duran significativamente más que las del estrato alto (efecto +0.38786, error 0.54955 $p < 0.06$).

En el caso de las /s/ producidas en Apartadó, sin embargo, no hay diferencias de apuntamiento ni duración entre estratos. En cambio, los parámetros que sí son significativos son: pasos por cero que, en el estrato bajo aumentan (+22.09, error 12.19, $p > 0.1$.) y el parámetro de ruido relativo, que en el estrato bajo se reduce (efecto: -1.29407, error

0.70863, $p > 0.1$). En los dos casos, el cambio en los parámetros es compatible con un menor grado de sonorización en el estrato bajo.

4. CONCLUSIONES

La /s/ en posición de ataque analizada en el departamento de Antioquia, Colombia, muestra que las realizaciones en Medellín tienen un espectro más apuntado en el estrato bajo, en comparación con el estrato alto. Por su parte, en Apartadó las /s/ están menos sonorizadas en el estrato bajo, que en el estrato alto.

5. REFERENCIAS

- Blum-Kulka, S. (1980). Learning to Say What You Mean in a Second Language; a Study of the Speech Act Performance of Learners of Hebrew as a Second Language. *Applied Linguistics*, 3(1), 29–59.
- Castañeda, L. S. (2005). El parlache: resultados de una investigación lexicográfica. *Forma y Función*, 18, 74–101.
- Castellano Ascencio, M. (2008). Fórmulas de tratamiento nominales para la pareja en el habla juvenil medellinense. *Íkala, Revista de Lenguaje y Cultura*, 13(20), 163–181.
- Chappell, W. (2016a). On the social perception of intervocalic /s/ voicing in Costa Rican Spanish. *Language Variation and Change* 28, 357–378.
- Chappell, W. (2016b). Sociophonetic perception of intervocalic [z] in Costa Rican Spanish. Paper presented at New Ways of Analyzing Variation 45, Simon Fraser University, November 3–6.
- Chappell, W. (2019a). Introduction: Spanish sociophonetic perception: The state of the field. In *Recent Advances in the Study of Spanish Sociophonetic Perception* (pp. 2–12). John Benjamins.
- Chappell, W. (2019b). Caribeño or mexicano, profesionista or albañil?: Mexican listeners' evaluations of /s/ aspiration and maintenance in Mexican and Puerto Rican voices. *Sociolinguistic Studies* 12(3–4), 367–393.
- Elvira-García, Wendy. *Zero-crossings-and-spectral-moments*, 2019 [Praat script]
- Flórez, L. (1951). *La pronunciación del español en Bogotá*. Bogotá, Instituto Caro y Cuervo.
- García, C. (2019). Regional identity in Highland Ecuador: Social evaluation of intervocalic/s/voicing. In *Recent Advances in the Study of Spanish Sociophonetic Perception* (pp. 126–152). John Benjamins.
- Jang, J. S. (2010). Fórmulas de tratamiento pronominales en los jóvenes universitarios de Medellín (Colombia) desde la óptica sociopragmática: estrato socioeconómico y sexo. *Íkala: Revista de Lenguaje y Cultura*, 15(26), 43–116.
- Lipski, J. M. (2005). *El español de América*. Cátedra.
- Montes Giraldo, J. J. (1982). El español de Colombia: propuesta de clasificación dialectal. *Thesaurus*, 37(1), 23–92.
- Muñoz-Builes, D. (en evaluación). Variación diastrática de la entonación femenina en el español colombiano de Medellín.
- Muñoz-Builes, D. (en preparación). *Mapa prosódico de Colombia*.
- Muñoz-Builes, D. y Elvira-García, W. (En prensa). Descripción acústica de la realización de [s] en Antioquia (Colombia). *Rilce Revista de Filología Hispánica*.
- Perdomo-Pinto, L., & Sadowsky, S. (2019). The Ultra-High-Frequency Whistled /s/ of Southern Chilean Spanish: Socioeconomic and Gender Stratification of its Spectral Moments and Prevalence. In *19th International Congress of Phonetic Sciences* (pp. 48–52). Melbourne.
- Prieto, P., & Roseano, P. (2010). *Transcription of Intonation of the Spanish Language*. Lincom Europa.
- Quilis, A. (1988). Estudio comparativo entre la entonación portuguesa (de Brasil) y la española. *Revista de Filología Española*, 68(1-2), 33–65.
- Sadowsky, S. (2012). *Naturaleza fonética y estratificación sociolingüística de los alófonos vocálicos del castellano de Concepción (Chile)*. Universidad de Concepción.
- Vanrell, M. D. M., Feldhausen, I., & Astruc, L. (2018). The Discourse Completion Task in Romance prosody research: Status quo and outlook. In I. Feldhausen, J. Fliessbach, & M. D. M. Vanrell (Eds.), *Methods in prosody: A Romance language perspective* (pp. 191–228). Berlín: Language Science Press.
- Walker, A., García, C., Cortés, Y., & Campbell-Kibler, K. (2014). Comparing social meanings across listeners and speaker groups: The indexical field of Spanish /s/. *Language Variation and Change* 26, 169–189.

ESTUDIO PRELIMINAR DE LA VOCAL ROTIZADA EN EL HABLA ESPONTÁNEA DE ANDALUCÍA Y ASTURIAS

Núria Ortiz-de-Pinedo Sánchez

Universidad Internacional de la Rioja
n.ortizdepinedo@gmail.com

ABSTRACT

El fenómeno de la vocal rotizada o *r-coloring*, registrado en las consonantes róticas del inglés, afecta directamente a las vocales y no está descrito en los manuales de pronunciación del español.

En la comunicación presentada en el VII Congreso Internacional de Fonética Experimental (Cantero y Ortiz-de-Pinedo, 2017) se introdujeron algunas características acústicas generales del fenómeno de la vocal rotizada en el habla espontánea del español peninsular. En esta ocasión, se exponen los resultados obtenidos tras el análisis acústico detallado de todos los sonidos vibrantes hallados en posición implosiva de las comunidades autónomas de Asturias y Andalucía, representación de la variedad septentrional y la variedad meridional respectivamente. El total de sonidos vibrantes analizados es de 176, 60 pertenecientes a Asturias y 116 propios de Andalucía.

Palabras clave: pronunciación, análisis acústico, vibrantes, vocales rotizadas

The rhotacized vowel or r-coloring, recorded in rhotic consonants of English, directly affects the vowels and is not described in the Spanish pronunciation manuals.

In the paper presented at the *VII Congreso Internacional de Fonética Experimental* (Cantero y Ortiz-de-Pinedo, 2017) some general acoustic characteristics of the rhotacized vowel in spontaneous speech in peninsular Spanish language were introduced. In this occasion, the results obtained after the detailed acoustic analysis of all the vibrating sounds found in implosive position in Asturias and Andalusia, representing the northern and southern variety respectively, are presented. The total number of vibrating sounds analysed is 176, 60 belonging to Asturias and 116 to Andalusia.

Keywords: pronunciation, acoustic analysis, rhotics, rhotacized vowels

1. INTRODUCCIÓN

El fenómeno de la vocal rotizada o *r-coloring* se produce al estar en contacto la /r/ con la vocal y el cambio más significativo se observa en las vocales. En efecto, según Knowles (1987), el cambio que se ha producido en estos últimos siglos ha sido mayormente vocálico, pero también con efectos sobre la rótica. Lo que ha ocurrido es que la realización vibrante se ha visto reemplazada por la creación de estas vocales largas que surgen a raíz de estar en contacto la vocal corta con la vibrante /r/.

Ladefoged (1975) realiza una descripción de las *rhotacized vowels* (vocales rotizadas) y explica la rotización como una cualidad auditiva de las vocales adyacentes a la rótica. También se refiere a esta característica como la *r-coloring*. Tanto Malmberg

(1963) como O'Connor (1973) coinciden que es un fenómeno retroflejo, se produce gracias a la pronunciación retrofleja de la rótica. Cuando articulamos la vibrante retrofleja se crea un hueco en la parte trasera del arco del paladar y por esto, la vocal siguiente está caracterizada por un timbre especial.

Esta característica de la vocal no es un fenómeno usual en las lenguas del mundo, tan solo ocurre en un 1% de estas (Maddieson 1984). De todas formas, sí que se observa un alargamiento de la vocal posterior a la rótica en sueco y se observan algunas modificaciones en las vocales anteriores a la vibrante en los sonidos uvulares del francés o del danés (Ladefoged and Maddieson, 1996).

En lengua inglesa, este rasgo de las vocales, lo podemos encontrar en pronunciaciones del inglés americano (Ladefoged, 1975; Malmberg, 1963;

Knowles, 1987; Catford, 1988), en el inglés de Irlanda (O'Connor, 1973; Knowles, 1987) y en algunas zonas dialectales de Inglaterra (Malmberg, 1963; O'Connor, 1973; Knowles, 1987). En estos dialectos del inglés se distinguen entre ellos como *rhotics* o *non-rhotics*. En el primer caso, se llaman dialectos róticos cuando permanece en la vocal algún rasgo característico del sonido /r/ y, por el contrario, si desaparece completamente la /r/ de la vocal, se agruparán con los dialectos no-róticos (Knowles, 1987).

La vocal rotizada en español

En Cantero y Ortiz-de-Pinedo (2017) se presentó el fenómeno de la vocal rotizada como un hecho no esperado que surgió durante el análisis acústico de las vibrantes realizado con motivo de los estudios doctorales de Ortiz-de-Pinedo (2017). Por este motivo, se presentaron algunas aproximaciones al fenómeno, pero no se detallaron sus características.

Se observó que no se trataba de un fenómeno aislado y característico de una realización de la vibrante en concreto ni tampoco de un contexto preciso. En la investigación anterior (Ortiz-de-Pinedo, 2017) se pudo ver esta característica de las vocales en un total de nueve realizaciones distintas de la vibrante y en cuatro de los seis contextos registrados: posición inicial, margen silábico, posición intervocálica y posición implosiva.

Las características acústicas de las vocales rotizadas (vocales parcial o totalmente rotizadas) que se habían observado en el espectrograma del sonido de forma repetida son las siguientes:

1. La baja frecuencia del F3
2. La pérdida de intensidad a partir del F4
3. El acercamiento entre el F3 y el F2.

Para comprobar que se estaba ante un sonido rotizado, se tuvieron en cuenta dichas características y también la manipulación del sonido: segmentando las partes críticas, eliminándolas, manteniendo la duración sin sonido, etc.

2. METODOLOGÍA

En esta investigación se ha realizado nuevamente un análisis acústico de todas las vibrantes en posición implosiva de las comunidades autónomas de Asturias y Andalucía para observar en qué ejemplos se encuentran rasgos de vocales rotizadas.

Así como se hizo durante el estudio previo a esta investigación (Cantero y Ortiz de Pinedo, 2017) y para poder determinar cuál es el comportamiento de las vibrantes en posición implosiva del español peninsular, se realizaron pruebas de manipulación del sonido. Estas pruebas consisten en eliminar la vibrante para comprobar que si se seguía percibiendo; segmentar el sonido vocálico para comprobar si la rotización de la vocal era total o parcial; eliminar por completo la vibrante y toda la parte rotizada de las vocales contiguas, manteniendo la duración para comprobar si este rasgo era también característico de las vocales rotizadas, etc.

Todas estas comprobaciones nos han permitido rellenar nuestra hoja de toma de datos con los siguientes datos:

- ¿La vocal está rotizada?
- ¿Qué tipo de rotización nos encontramos?
¿Total o parcial?
- ¿Qué vocal está rotizada? (anterior, posterior o ambas)
- ¿Cuál es la duración de dicha vocal?
- ¿Cómo es el comportamiento de los F2 y F3?
- ¿Hay disminución de intensidad? (con respecto a los sonidos colindantes)

2.1. Corpus

El corpus con el que se ha trabajado en esta investigación fue creado por Ballesteros, Mateo y Cantero (2010) con la finalidad de analizar melódicamente las variedades del español peninsular. Se divide en dos variedades dialectales: la variedad septentrional y la variedad meridional. Es un corpus completamente de habla espontánea o habla no-planificada.

Este corpus cumple con las siguientes características:

1. Es un corpus formado por enunciados orales de habla no planificada. Todos los enunciados han sido extraídos de programas de televisión (desestimando aquellos informantes considerados locutores profesionales y habituados al medio en el que trabajan).
2. Es una representación muy amplia de las variedades habladas en el territorio español peninsular.
3. Contiene un número amplio, representativo, válido y equilibrado de informantes.
4. Tiene buena calidad acústica.

Para su elaboración se desestimaron, por un lado, las primeras intervenciones de los informantes en los programas que eran de debates o de tertulias porque se sobreentendió que quizás eran las partes más preparadas de sus intervenciones; también se descartaron las intervenciones de personas que hacían visible su nerviosismo o su emoción por estar delante de una cámara. Por otro lado, se eliminaron las intervenciones en las que se detectaron ruidos o interferencias para que estas no modificaran el posterior análisis.

2.2. Muestra analizada en la investigación

Para esta investigación, se ha realizado un análisis pormenorizado de todas las vibrantes en posición implosiva de las comunidades autónomas de Asturias y Andalucía. Un total de 176 sonidos repartidos de la siguiente manera:

Tabla 1: Muestra de sonidos analizados para esta investigación

Comunidad	Nº de sonidos
Andalucía	116
- Andalucía oriental	45
- Andalucía occidental	71
Asturias	60

El corpus total de Asturias está formado por 200 enunciados de 13 horas de grabación que proceden de tres programas que se emiten en la Televisión del Principado de Asturias (TPA):

- De tarde
- Asturias en 25
- Conexión Asturias

En estos programas se dan tertulias, reportajes y concursos. Los informantes que se han seleccionado son un total de 67.

El corpus total de Andalucía está formado por 520 enunciados de 53 horas, aproximadamente, de grabaciones de distintos programas de la televisión autonómica Canal Sur:

- Alégame la vida
- Andalucía directo
- Apasiona2
- Aquí estamos
- El club de las ideas
- La tarde aquí y ahora
- La trastienda
- Mira la vida
- Parlamento Andaluz

- Tesis
- Coloquio
- 75 minutos
- TV local Almería.

Son magazines, programas informativos, de debate, entrevistas y reportajes. Los informantes que se han seleccionado son un total de 139.

3. OBJETIVOS

Los objetivos que se han planteado en este estudio son los siguientes:

- Determinar el número total de sonidos rotizados de la muestra analizada.
- Clasificar los sonidos rotizados según el tipo de rotización que presenten: parcial o total.
- Caracterizar los sonidos rotizados.

4. RESULTADOS

En el análisis general de todo el corpus, encontramos dos tipos de *rotización*: la *rotización total* de la vocal anterior o posterior al sonido y la *rotización parcial* de la vocal anterior o posterior al sonido. Como sus nombres indican, en el primer caso nos encontramos con una vocal completamente rotizada, de inicio a fin del sonido; y en el segundo caso registramos la rotización tan solo en una parte de la vocal.

En el ejemplo siguiente observamos que la vocal siguiente de la vibrante está parcialmente rotizada.

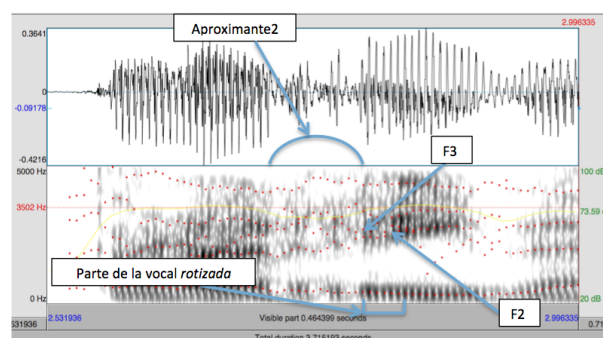


Figura 1: Ejemplo de vocal posterior rotizada. Sonido AN-63-02-05/25: arriba.

Se observan en este ejemplo las características siguientes:

- En la parte seleccionada de la vocal que está rotizada vemos cómo el F3 y el F2 están tan próximos que en algún momento es imposible separarlos.
- Vemos también cómo a partir del F4 se pierde intensidad con respecto a la parte que no está rotizada.

En este ejemplo en concreto consideramos que la vocal está “coloreada” parcialmente por la vibrante porque cuando eliminamos los segmentos de la vocal coloreados, llega un momento que dejamos de percibir acústicamente la vibrante.

En la figura siguiente, se puede observar un sonido vibrante que “colorea” completamente a su vocal anterior.

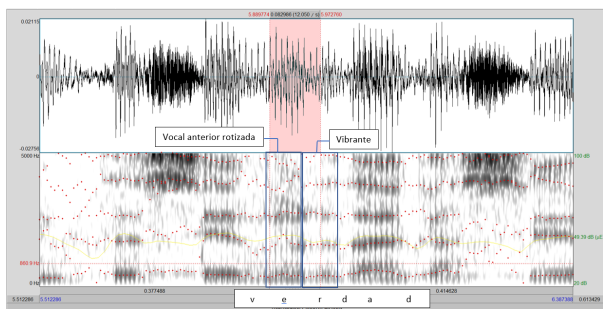


Figura 2: Ejemplo de la vocal anterior rotizada. Sonido AN-01-02-04/32: *verdad*.

En este ejemplo, la vocal anterior al sonido vibrante está completamente rotizada. Por el contrario, en este caso, tan solo observamos en la mitad del sonido vocálico previo a la rótica que hay una elevación del F2 que coincide con un pequeño descenso de intensidad, pero que no se mantiene durante toda la vocal.

5. CONCLUSIONES

El análisis se ha realizado mediante un método instrumental y descriptivo con la finalidad de caracterizar acústicamente los rasgos de las vocales rotizadas en habla no controlada.

De la muestra analizada de 176 vibrantes en posición implosiva de las comunidades de Asturias y Andalucía, 55 sonidos presentan vocales rotizadas. Esto supone un 31,25% del total.

El contexto en el que se encuentran estos sonidos, posición implosiva, determina que la vocal rotizada siempre sea la anterior al sonido. La vocal, como ya

hemos comentado anteriormente, puede estar rotizada de forma total o parcial. En este estudio, hemos observado que el 30,90% de las vocales rotizadas presentan una rotización total, mientras que el 69,09% tienen solo una parte rotizada.

El comportamiento de los formantes es bastante dispar. En la gran mayoría de sonidos observamos el acercamiento del F2 y el F3 debido al ascenso del primero y al descenso del segundo. También observamos en muchos otros ejemplos que ambos formantes son paralelos durante todo el sonido vocálico. Después podemos observar de forma más esporádica distintos comportamientos: ambos formantes aparecen tan juntos que es imposible dividirlos, el acercamiento de los formantes se produce hacia el final del sonido, hay un descenso de ambos formantes de forma paralela, etc.

Referente a la intensidad, el 52,72% de las vocales rotizadas presentan una disminución de intensidad, mientras que en el 47,27% no se observa.

En esta comunicación observaremos, mediante ejemplos de nuestro corpus analizado, todos estos datos que han descrito el fenómeno de las vocales rotizadas del habla espontánea en posición implosiva de las comunidades autónomas de Andalucía y Asturias.

6. FUTURAS INVESTIGACIONES

Este es el primer paso de una investigación más extensa sobre el comportamiento de las vocales en contacto con la vibrante en posición implosiva en todas las comunidades autónomas de nuestro corpus.

Por un lado, sería interesante también ver si el comportamiento de estas vocales cambia según el contexto en el que se encuentren en contacto con la vibrante.

Por otro lado, sería conveniente poder realizar una serie de pruebas perceptivas para poder observar si existe, o no, un proceso de mediación entre los hablantes que puede influir entre lo que realmente se pronuncia y lo que percibimos.

7. BIBLIOGRAFÍA

- Ballesteros, M. P., Mateo, M., & Cantero, F. J. (2010). Corpus oral para el análisis melódico de las variedades del español. *Actas del XXIX Simposio Internacional de la Sociedad Española de Lingüística*. Santiago de Compostela: Universidad de Santiago de Compostela.
- Cantero, F.J. & Ortiz-de-Pinedo, N. (2017). La vocal rotizada o r-coloring en el español peninsular. En V. Marrero & E. Estebas (coord.): *Tendencias actuales en fonética experimental* (pp.49-52). Madrid: UNED.
- Catford, J. (1988). *A practical introduction to phonetics* (2001, 2a ed.). Oxford: Oxford University Press.
- Knowles, G. (1987). *Patterns of Spoken English. An Introduction to English Phonetics* (1995, 6a ed.). New York: Longman Group Limited.
- Ladefoged, P. (1975). *A course in phonetics* (1993, 3a ed.). New York: Harcourt Brace Jovanovic.
- Ladefoged, P. & Maddieson, I. (1996) *The sounds of the world's languages*. Blackwells,
- Maddieson, Ian. 1984a. *Patterns of Sounds*. Cambridge University Press, Cambridge.
- Malmberg, B. (1963). *Phonetics*. New York: Dover Publications.
- O'Connor, J. (1973). *Phonetics* (1991, 2a ed.). England: Penguin Books.
- Ortiz-de-Pinedo, N. (2017). *Análisis acústico de las vibrantes del español en habla espontánea*. Tesis Doctoral. Universidad de Barcelona.

L2 PHONETIC ACCURACY DEVELOPMENT IN A FOREIGN LANGUAGE SETTING: A LONGITUDINAL STUDY

Alejandra Pesantez, Volker Dellwo

University of Zurich, Switzerland
alejandra.pesantezpesantez@uzh.ch, volker.dellwo@uzh.ch

ABSTRACT

This study investigates second language phonetic development in the accuracy of the high front vowels produced by 21 Ecuadorian learners of English as a foreign language. The participants were enrolled in the First and Foreign Language Teaching program at the University of Cuenca-Ecuador. At the outset of the study, they were in the third semester (T1) and at the completion of it, they were in the fifth semester (T3) of the program.

The acoustic analysis revealed that Ecuadorian learners of English started forming new categories of the vowel contrast /i/-/ɪ/ for F1 and F2 formant values at T2 and T3. They did not produce any difference between the English and Spanish vowel /i/ tokens in the F1 and F2 formant values in T3. Comparing to teachers and American English speakers, Ecuadorian learners of English differ in formant values production.

Keywords: accuracy, development, exposure, Ecuadorian EFL learners

1. INTRODUCTION

The goal of a student who learns a second language (L2) is to communicate effectively with native and non-native speakers of English after considerable exposure to the target language. During the learning process of the L2 language, adult learners struggle to produce the sounds of the L2 language in a “particular setting or long term- configuration” (Esling & Wong, 1983). For example, the L1 Spanish speakers have difficulties producing differences between the tense and lax vowel sounds because they rely more on duration cues to produce the English vowel contrasts than on spectral information as native speakers do (Escudero, 2005).

The main characteristics of English and Spanish vowels relate to the position of the tongue body on a front-back axis, the relative height of the tongue in the mouth, and the shape of lips, rounded or unrounded during the production of vowel sounds (Ladefoged, 1971). The pure Spanish vowels are tense and short,

with minor variations in duration from the stress and unstress vowels (Celdrán & Elvira-García, 2019). In contrast, English vowels are both tense and lax, and comparing the American English tense vowels /i u e o/ with their Spanish counterparts, which are all pronounced as true monophthongs, they are comparatively less tense than Spanish vowels (Stevens, 2011). For producing the English vowels /i/ and /u/ the tongue position is higher compared to the Spanish vowels (Flege, 1989).

In trying to explain why L2 adults fail to produce differences between the vowel sounds of the first language (L1) and target language, researchers have paid attention to the role of input (Flege & Wayland, 2019), the L2 language experience (Flege, Bohn & Jang 1997; Baker & Trofimovich 2006; Colantoni, Steele & Escudero, 2015), the age of learning (Munro, Flege & Mackay, 1996; Abrahamsson & Hyltenstam, 2009), and the L2 aptitude (Huensch & Thompson, Saito, 2017; Saito, Suzukida & Sun, 2019). Apart from these factors affecting L2 phonetic acquisition, the developing phonological awareness

(the study of patterns of sounds) and phonetics aspects (production of the sounds) will enhance learners to decode information from L2 input and form new phonetics categories (Saito & Lyster, 2012).

Most of the cross-sectional studies of segmental acquisition comparing learners with a different kind of experience and exposure, quantity, and quality of input, in a naturalistic setting have not yield solid conclusions about adult phonetic learning development. The necessity of longitudinal studies within the field of L2 phonetic learning in an FL context is crucial to understand how L2 adult learners develop new phonetic categories over time as the number of English and Linguistics classes increases.

Taking into account the phonological differences between English and Spanish vowels and the exposure to English in an FL classroom setting, we formulated the following research questions based on the acoustic analysis of the English and Spanish vowels:

- As exposure increases, will the Ecuadorian EFL learners show improvement in the production of the vowel contrast /i/-/ɪ/ in terms of quality over T1, T2, and T3?
- What influence does the Spanish /i/ have on Ecuadorian EFL learners' production of the English /i/ in T3?
- To what extent the English formant values produce by Ecuadorian EFL learners in T3 will resemble those of their teachers or native American English speakers?

2. METHOD

In order to investigate the Ecuadorian English phonetic development over time, a production experiment was replicated three times during their studies: T1 (third semester), T2 (fourth semester), and T3 (fifth semester). The experiment consisted of a production task that was applied at the end of every semester providing information about the first two formant values of the English vowels. Another production task was administered to have information about the Spanish first two formant values.

2.1. Participants

Twenty-one undergraduate Ecuadorian L1 Spanish speakers who study EFL participated in our study. They were 16 female and 5 male adults, students of the First and Foreign Language Teaching program at

the University of Cuenca - Ecuador. Their ages ranged from 18 to 29 years, (M = 21.31). All the participants were enrolled in the phonetics class in the fourth semester and at the time of the second part of the data collection. Thus, it was expected that their exposure to English in the classroom and explicit phonetic instruction could have already contributed to help learners to form phonetic categories for the L2 /i - ɪ/ vowel contrasts.

Apart from the learner's data, we also collected data related to native American English speakers (NEs), 3 females, and 2 males. Their ages ranged from 19 to 43 years old. The students' teachers (T), 6 females and 1 male, also were recorded with a mean age of 37.14. Most of the teachers were not L1 English speakers, except for one bilingual simultaneous Spanish-English teacher.

2.1. Experiment 1: L1 and L2 production

2.1.1. Material

Ecuadorian EFL learners, T and NEs were asked to produce 40 monosyllabic isolated words containing the English vowel contrasts /i/-/ɪ/, /u/-/ʊ/, /ɛ/-/æ/, /ʌ/-/ɑ/ in a CVC and CVCC context, and 5 disyllabic words were added as a control group. For the purpose of this study, we only analyzed the tense and lax vowels /i:/-/ɪ/ which are ones of the most difficult to perceive and produce by L1 Spanish speakers (Flege, Bohn & Jang, 1997). Learners also said 40 isolated Spanish (Sp) words containing the five vowels /a /, /e/, /i/, /o/, /u/ in a CVC, CVCV, CVCVCV context. Only the vowel /i/ first CV (containing stress) was examined.

2.1.2. Procedure

A picture-naming task was used to elicit target vowels and their relative label in Spanish was presented together on a computer screen. The recording section was done in the radio station of the University of Cuenca, and their productions were recorded by using the Zoom H2n handy recorder at 44.1 kHz sampling, 16-bit quantization.

2.1.3. Data Analysis

To compare the spectral characteristics between the group of learners, NEs, and T we manually annotated the vowels, measured the first two formants F1 and F2. The researcher used waveform, spectrogram, and auditory discrimination cues in determining vowel boundaries. For extracting the mean of the first and second formant frequency (F1 and F2) and plotting the vowel space for both languages, two different scripts

were created using Praat software (2019). Vowels were Lobanov's z score normalized.

2.1.4. Statistical Analysis

To examine whether the same participants can start producing differences between /i-/ as exposure increased in the T1, T2, and T3, two mixed-effects models (one for F1 and F2) with formants as dependent variables were used, speaker as an independent variable, and for random effects, we had intercepts for speaker and word. The same model was used for the following questions with variation in the number of participants for question two (T3, Spanish) and for question three (T3, NEs, and T). The Post-hoc Tukey test was also used during the analysis.

3. RESULTS

To answer the first question, the mean of the F1 and F2 formants values of the /i-/ /ɪ/ vowels produced by Ecuadorian EFL learners were plotted indicating that learners made improvements over time (Fig.1). Two mixed-effect models were performed to evaluate the effect of exposure on the production of the English vowel contrasts /i-/ /ɪ/ during three semesters. The results demonstrated that there was a statistically significant interaction between time and vowel on F1 values, ($F(2, 552.56) = 10.240, p < .001$) and F2 values, ($F(2, 582.45) = 6.0168, p < .01$). Post-hoc tests (with Tukey correction) indicated that the effect of the amount of exposure for F1 and F2 values was not significant at T1. They became significant for F1 at T2 ($p = 0.0123$) and T3 ($p = 0.0024$), and for F2 at T2 ($p = 0.0067$) and T3 ($p = 0.0007$).

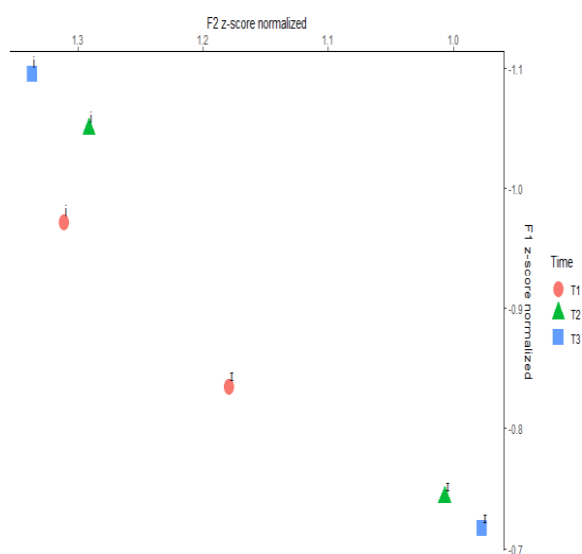


Figure 1: Vowel chart mean values of the /i-/ /ɪ/ vowel contrasts produced by Ecuadorian EFL learners at T1, T2, T3.

To know if the Ecuadorian EFL learners produce differences between the F1 and F2 formant values of the Spanish vowel sound /i/ compared to the F1 and F2 formant values of the English vowel sound /i/, a mixed model using formants as the dependent variable was carried out. The results revealed no significant difference in vowel production for F1 values, ($F(1, 13.29) = 0.0396, p > 1$) and F2 values, ($F(1, 38.703) = 0.0062, p > 1$).

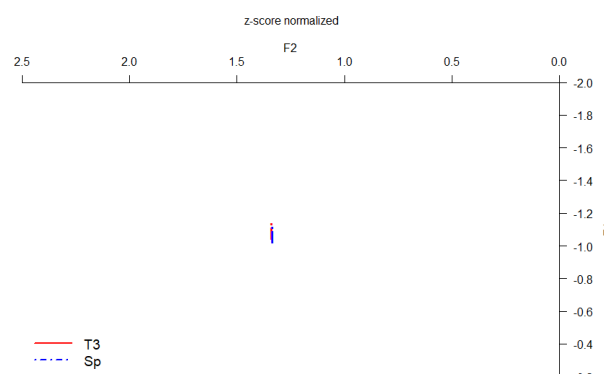


Figure 2: The average position of the native Spanish vowel /i/ and English vowel /i/ in the acoustic F1/F2 space produced by Ecuadorian participants.

A visual inspection of the boxplots indicates us that F1 and F2 formants had different acoustic realization between T3, NEs and T (Fig 3). The result of the fixed model built illustrated that the main effect of group for the English vowel /i/ was significant on F1 values, ($F(2, 30.026) = 9.3922, p < .001$) and F2 ($F(2, 29.666) = 8.4595, p < .01$). Post-hoc test showed that F1 values for T3 differ significantly compared to NEs ($p = 0.0010$) and T ($p = 0.0510$); and for F2 also differ between T3 and NEs (0.0037) and T3 and T ($p = 0.0222$). There was also a significant difference in the production of /i/ between T3, NEs and T for F1, ($F(2, 30) = 6.7499, p < .01$) and F2 ($F(2, 30) = 6.1476, p < .01$). Post-hoc tests indicated that the effect of group for F1 values was significant between T3 and NEs ($p = 0.0137$), and T3 and T ($p = .0296$). However, for the F2 values, it was significant only for T3 and NEs ($p = .0063$).

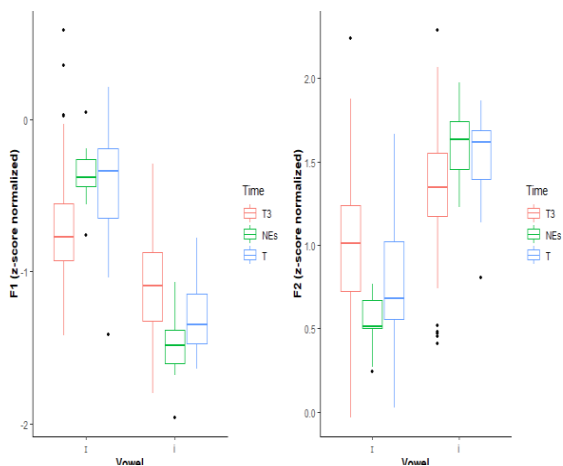


Figure 3: Boxplot of the /i- ɪ/ F1 and F2 formant values produced by the Ecuadorian EFL learners at T3 (fifth level), NEs (native English speakers), and T (teachers).

5. CONCLUSION

This study examined the L2 phonetic accuracy between the vowel contrast /i-/ɪ/ in a foreign language setting. The findings indicated that Ecuadorian EFL learners improved in F1 and F2 formants as exposure increased. The biggest improvement was found in T2 and T3 when learners were in the fourth and fifth semesters. One reason could be attributed to phonetics classes which helped L2 English learners to start forming clear-cut categories for these vocalic phonemes of the English language. Another explanation may be the experience and exposure that learners had gained during these three semesters. In terms of Spanish and English /i/ tokens, Ecuadorian learners of English did not produce any statistical difference for F1 and F2 values. A possible explanation could be the influence of the Spanish vowel system creating a tendency to produce the same F1 and F2 formant values for both languages. Comparing the spectral measurements between T3, T, and NEs, learners produced instances of /i/ tokens that were closer to their T for F2 formants. Even though learners started applying rules to produce phonetic differences in the two sounds contrast, these contrasts still differ from their T and from NEs even more.

6. REFERENCES

Abrahamsson, N., & Hyltenstam, K. (2009). Age of acquisition and nativelikeness in a second language-listener perception vs. linguistic scrutiny. *Language Learning*, 59, 249-306.

- Baker, W., & Trofimovich, P. (2006). Perceptual paths to accurate production of L2 vowels: The role of individual differences. *IRAL-International Review of Applied Linguistics in Language Teaching*, 44(3), 231-250.
- Celdrán, E. M., & Elvira-García, W. (2019). Description of Spanish vowels and guidelines for teaching them. *Key Issues in the Teaching of Spanish Pronunciation: From Description to Pedagogy*.
- Colantoni, L., Steele, J., Escudero, P., & Neyra, P. R. E. (2015). *Second language speech*. Cambridge University Press.
- Esling, J. H., & Wong, R. F. (1983). Voice quality settings and the teaching of pronunciation. *Tesol Quarterly*, 17(1), 89-95.
- Escudero, P. (2005). Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization. *Netherlands Graduate School of Linguistics. Quilis*.
- Flege, J. E. (1989). Differences in inventory size affect the location but not the precision of tongue positioning in vowel production. *Language and Speech*, 32(2), 123-147.
- Flege, J. E., Bohn, O. S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of phonetics*, 25(4), 437-470.
- Flege, J. E., & Wayland, R. (2019). The role of input in native Spanish Late learners' production and perception of English phonetic segments. *Journal of Second Language Studies*, 2(1), 1-44.
- Huensch, A., & Thompson, A. S. (2017). Contextualizing attitudes toward pronunciation: Foreign language learners in the United States. *Foreign language annals*, 50(2), 410-432.
- Ladefoged, P. (1971). *Preliminaries to linguistic phonetics*. University of Chicago Press.
- Munro, M. J., Flege, J. E., & MacKay, I. R. (1996). The effects of age of second language learning on the production of English vowels. *Applied psycholinguistics*, 17(3), 313-334.
- Saito, K. (2017). Effects of sound, vocabulary, and grammar learning aptitude on adult second language speech attainment in foreign language classrooms. *Language Learning*, 67(3), 665-693.
- Saito, K., & Lyster, R. (2012). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of /ɪ/ by Japanese learners of English. *Language Learning*, 62(2), 595-633.
- Stevens, John. (2011). Vowel duration in second language Spanish vowels: Study abroad versus at-home learners. *Arizona Working Papers in SLA & Teaching*, 18, 77-104. (1988).
- Saito, K., Suzukida, Y., & Sun, H. (2019). Aptitude, Experience, and Second Language Pronunciation Proficiency Development in Classroom Settings: A Longitudinal Study. *Studies in Second Language Acquisition*, 41(1), 201-225.

PERCEPTUAL ASSIMILATION OF TASHLHIYT CONSONANTS BY SPANISH-DOMINANT BILINGUALS

Lucrecia Rallo Fabra¹, Mustapha Achichaou¹, Michael D. Tyler²

¹Universitat de les Illes Balears, ²Western Sydney University
lucrecia.rallo@uib.es, m.tyler@westernsydney.edu.au

ABSTRACT

Tashlhiyt is a variety of Tamazight, an Afro-Asiatic Language spoken mostly in Southern Morocco. In the present study, we investigate categorization of Tashlhiyt consonants by a group of Spanish-dominant bilinguals with no experience in that language. The task consisted of a perceptual assimilation task followed by goodness ratings. Overall, listeners categorized quite consistently the Tashlhiyt consonants with a Spanish counterpart. Pharyngeal and aryepiglottal consonants were mostly uncategorized, with two or more Spanish categories selected above chance, which led to perceptual overlap between various Tashlhiyt consonants. The results allow predictions of discriminability to be made, based on the Perceptual Assimilation Model (Best, 1994, 1995), and add to the existing research on categorization of nonnative phones.

Keywords: categorization, perceptual assimilation, language experience, Spanish, Tashlhiyt

1. INTRODUCTION

Humans perceive many linguistic and non-linguistic events and objects categorically. For instance, in vision, observers can label the chromatic spectrum using four basic color terms – blue, green, yellow and red, or their combinations (Bornstein, 1987, among others). In the case of speech, there is evidence that pre-linguistic infants are born with sensitivity to categorical boundaries between speech sounds. These have a facilitative role for language learning and can be updated and fine-tuned on the basis of experience with specific languages (Eimas et al., 1987).

The *Perceptual Assimilation Model-PAM* (Best, 1994, 1995) proposes that native language experience shapes the perception of phonological categories in a non-native language. Naïve listeners perceive individual non-native phones as: 1) assimilated to a native phonological category (ranging from a good to a poor exemplar); 2) assimilated as speech, but as unlike any native phonological category (uncategorized-dispersed, see Faris, Best, & Tyler, 2016); 3) assimilated as speech, and weakly consistent with one (uncategorized-focalized) or more (uncategorized-clustered) native phonological categories, or; 4) not assimilated as speech. Predictions for pairs of contrasting non-native phones are derived by combining these perceptual assimilation patterns. For example, if each non-native phone is assimilated to a different native category (a *two-category* assimilation) then discrimination is

predicted to be excellent. If both non-native phones are assimilated as equally good exemplars of the same native category (a *single-category* assimilation) then discrimination should be poor.

Perceptual assimilation studies are common in cross-language L2 perception. Understanding how L2 learners categorize L1 sounds is a prerequisite for L2 discrimination and it allows researchers to make predictions about L2 learnability. At present, there is an extensive body of research investigating perceptual assimilation of L2-English sounds to L1-Spanish (Cebrian, 2019; Baigorri et al., 2018; Escudero & Chládková, 2012) and L1-Catalan (Cebrian, 2006; Rallo Fabra, 2005; Safronova, 2015). However, studies investigating perceptual assimilation of non-native sounds from less common languages are scarce (see Tyler et al., 2014; Faris, Best & Tyler, 2016, 2018), which limits the generalizability of existing theories such as PAM.

The present study was designed with two major goals in mind. On the one hand, we intend to fill a gap in the literature, testing L1-Spanish listeners' categorization of non-native consonants from a minority language, Tamazight. On the other hand, we intended to provide preliminary evidence on how listeners would use their diverse language experience in more than one language to categorize the consonants of an unfamiliar language.

Tashlhiyt is a variety of Tamazight spoken mostly in Southern Morocco (Ridouane, 2014). From the phonological perspective, one of the striking aspects for the Spanish listener is the existence of the

pharyngeal coronal consonants that have phonemic status /tʰ dʰ sʰ zʰ ʒʰ rʰ lʰ/, thus distinguishing pairs of words such as /arʰa/ *give me!* /ara/ *write!*, /askʰa/ *comb*, /aska/ *tomorrow*. At the surface level, pharyngealization is also used to express emphasis (exclamation). Articulatory differences between the pharyngeal and plain stops involve the supra-laryngeal and laryngeal levels. In the case of /tʰ/, the tongue is shifted backwards towards the posterior laryngeal wall and the glottal opening is smaller than for /t/. These gestures account for the acoustic differences in VOT duration, which is shorter for /tʰ/ than for /t/.

2. METHOD

2.1. Participants

A group of 24 Spanish-Catalan bilinguals living in Majorca, with a mean age of 18 years, volunteered to participate in the study. The sample included first-year undergraduates, faculty, and administrative staff. The participants' L1 dominance in either Spanish or Catalan was controlled via an exhaustive language background questionnaire. Seventy-five percent of the participants were raised in predominantly Spanish-speaking environments since both parents spoke Spanish at home. A few of the participants reported speaking both languages at home but they interacted mostly in Spanish.

2.2. Stimuli

The stimuli used in the experiment were produced in a sound-treated booth by a young-adult male native speaker of Tashlhiyt. Twenty-three consonants were used in the experiment (see Tables 1-3), each produced in /Ca/, /Ci/, and /Cu/ frames. There were 6 tokens per frame for /k, g, q, ɣ, ʁ, h, ɦ, ʕ/ and 1 token per frame for the remaining consonants.

2.3. Procedure

The participants were tested in groups of up to two in a quiet room. On each trial, the participants listened to a CV syllable and clicked an on-screen response. Immediately after, the same sound was played a second time, and they rated the goodness of fit using a 7-point scale (1 = a poor fit, 7 = a perfect fit). The response categories were the 20 Spanish consonant phonemes printed in orthography and arranged alphabetically (b, c/qu, ch, d, f, g, j, l, ll, m, n, ñ, p, r, rr, s, t, w, y, z), plus three extra response options: (1) *Not a speech sound* (e.g., door slamming, twig snapping, hands clapping). Participants then typed a description of what they heard. (2) *Other speech sound* sounded like a combination of sounds, or like one or more speech sounds in a language other than Spanish, or in a different Spanish dialect. Participants then typed the language and sound(s) that they heard. (3) *Unknown speech sound* was for any sound that was heard as speech but was unlike any consonant in

Table 1: Mean percent categorization of Tashlhiyt stop and nasal consonants onto Spanish categories. Categorized consonants are in bold.

Tashlhiyt Stimuli	Spanish Response Categories							
	k	g	p	T	b	d	m	n
k	93.8 (6.0)							
g		99.1 (6.0)						
q	47.1 (5.0)		25.7 (5.3)	13.8 (5.1)				
t				93.1 (5.6)				
tʰ			55.6 (5.6)	19.4 (5.6)				
d						97.2 (5.9)		
dʰ					56.9 (4.8)	20.8 (5.6)		
b					70.8 (5.6)	23.3 (5.7)		
m							75.7 (6.4)	20.1 (5.9)
n								98.6 (6.2)

Table 2: Mean percent categorization of Tashlhiyt fricative consonants onto Spanish categories. Categorized consonants are in bold.

Tashlhiyt Stimuli	Spanish Response Categories								
	g	x	r	f	s	ʃ	tʃ	θ	Uncat
f				96.2 (6.3)					
s					82.6 (6.2)				
z					30.6 (4.2)			66.7 (4.9)	
s^ɸ				16.3 (4.7)	63.2 (5.2)			14.9 (4.4)	
z^ɸ					25.3 (4.3)			56.6 (4.8)	
ʃ					40.6 (4.4)	21.9 (6.2)	29.2 (4.5)		
ʒ		30.6 (5.1)						17.4 (4.9)	
ʒ^ɸ		20.8 (4.5)							
ʁ	33.2 (4.3)		25.6 (3.8)						
χ	19.3 (5.9)	79.6 (6.1)							
ħ		52.3 (4.3)							
ʕ			28.0 (3.3)						23.0
h	15.0 (3.9)	13.9 (3.2)	21.2 (3.2)						16.2

any language known by the listener. No rating was required for unknown speech sounds.

The experiment was run using *Psyscope X B77* on MacBook Pro laptops. The stimuli were presented through headphones and in a random order. Each testing session lasted around 45 minutes. The participants filled a language background questionnaire at the end of the session.

3. RESULTS

The mean percent categorization for the Tashlhiyt consonants is shown in Tables 1 to 3. Please note that response categories not chosen above chance by the listeners have been omitted from confusion matrixes. (1/23, as tested by a one-sample *t*-test, see Faris et al., 2016, 2018). A 70% assimilation was considered to be the benchmark for determining whether a given consonant was categorized to a L1 consonant (see Tyler et al., 2014). Each percentage corresponds to the average responses given by the 24 listeners to each Tashlhiyt stimulus, collapsed across vowel.

3.1. Stops and nasals

Overall, the Spanish listeners perceptually assimilated Tashlhiyt /k/, /g/, /t/, /b/, /d/ to their Spanish counterparts (see Table 1). The high goodness ratings indicate a very good fit to their native categories. These four consonants can be considered categorized in terms of PAM and the high assimilation percentages suggest that participants understood the task and their performance was reliable. The uvular /q/ and pharyngeal stops /t^ɸ/ and /d^ɸ/ were uncategorized-clustered, that is, listeners heard them as a mixture of two or more L1 consonants. In the case of /q/, listeners perceived it predominantly as moderate-to-good instances of Spanish /k/ and also as /p/ and /t/. Strangely, the pharyngeal stops /t^ɸ/ and /d^ɸ/ were mostly categorized as /p/ and /b/, respectively.

3.2. Fricatives

Only the three fricatives that have an equivalent in Spanish, namely /f/, /s/ and /χ/, were categorized (see Table 2). The rest of the fricatives were not clearly

assimilated to a single native category, but there was at least one label selected above chance for each (i.e., they were uncategorized-clustered or -focalized patterns of assimilation). The modal response for the Tashlhiyt fricatives /s/, /sʰ/ and /ʃ/ was Spanish /s/, but only /s/ was categorized. There appear to be different degrees of goodness-of-fit, /s/ being the most similar followed by /sʰ/ and then /ʃ/. A more complex case is that of the voiced post-alveolar /z/ and /zʰ/, voiceless uvular /χ/, and voiceless aryepiglottal /ħ/ fricatives. Spanish /x/ was selected for all four consonants, but only /χ/ was categorized. /ħ/ and /zʰ/ were uncategorized-focalized because /x/ was the only above-chance label. It is interesting to note that /x/ was selected only 20% of the time for /zʰ/, suggesting that it has little similarity to any Spanish category. Responses for /z/ were divided between /x/ and /θ/ (uncategorized-clustered). Where /x/ was selected among the four consonants, goodness of fit ranged from moderate (4.5/7) to very good (6.1/7). Responses for the glottal fricative /h/ were shared between /g/, /x/, and /r/, and a significant proportion of responses were heard as unknown sounds. For the voiced aryepiglottal /ʕ/, responses were shared between /r/ and unknown sound.

3.3. Laterals and Approximants

As shown in Table 3, the pairs /l-/lʰ/ and /r-/rʰ/ show similar patterns of assimilation, although only /l/ and /lʰ/ were categorized. In both cases the plain consonants appear to be perceived as better fits to the Spanish categories than their pharyngealized counterparts. It is interesting to note that the modal response category for the palatal approximant /j/ was /ʎ/ rather than /j/.

Table 3. Mean percent categorization of Tashlhiyt lateral and approximant consonants onto Spanish categories. Categorized consonants are in bold.

Tash Cat.	Spanish Response Categories					
	r	r	j	l	ʎ	w
l				92.0		
				(6.3)		
lʰ				83.3		
				(5.2)		
r	67.7					
	(4.8)					
rʰ	55.2	21.5				
	(4.2)	(4.7)				
w						58.3
						(5.3)
j			25.7		53.8	
			(5.5)		(5.3)	

4. DISCUSSION

In general, only Tashlhiyt consonants that have an equivalent in Spanish, namely /b, t, d, k, g, b, m, n, f, s, χ, l/ were consistently categorized. Following PAM, discrimination of phonological contrasts involving any two consonants in that group should be accurate. The only exception is the pharyngealized /lʰ/, which was also assimilated to Spanish /l/. In PAM terms, the /l-/lʰ/ contrast would be an instance of category goodness, that is, one member of the pair is a better fit to the native category than the other.

We speculate that the perceptual overlap in many of the fricative consonants is not likely to hinder discrimination ability. For instance, the varying degrees of assimilation percentages and goodness ratings of /s/, /sʰ/, and /ʃ/ suggest that there will be differences in perceived phonological overlap between those sounds and their native Spanish categories (see Faris et al., 2018). The /b-/dʰ/ is completely overlapping phonologically, so discrimination should be poor.

Two assimilation patterns are surprising. One of them involves the modal response /ʎ/, which was preferred over /j/ to categorize Tashlhiyt /j/. This could be due to *yeísmo*, a sound change in progress in which the /j-/ʎ/ contrast is lost in Castilian Spanish (Hualde, 2005) and Majorcan Spanish (Rost Bagudanch, 2019). The assimilation patterns of Tashlhiyt /z/ and /zʰ/ onto Spanish /θ/ are odd. We would expect a trend towards Spanish /s/. Again, the fact that the listeners were Spanish-Catalan bilinguals might have confused them. The letter “z” is pronounced /θ/ in Spanish but /z/ in Catalan. Aside from that, /s-/z/ have phonemic status in this language. Future studies should include a Spanish keyword to ensure that participants only select “z” when they intend to indicate /θ/.

PAM was conceived from the perspective of a monolingual perceptually assimilating the sounds of a non-native language to the native language. In this study, participants were Spanish-dominant bilinguals who had been exposed to Catalan at school since childhood. Most of them had learnt English and other foreign languages outside their home such as German, French or Japanese. In fact, many listeners resorted to their knowledge of additional languages to categorize unfamiliar Tashlhiyt phones, although not to the extent that they resulted in above-chance responding. If, as PAM suggests, language experience shapes our perception of non-native speech sounds, future research testing perceptual assimilation in bilingual contexts should consider ways to allow participants to indicate categories from either or both of their languages that are similar to non-native stimuli.

5. ACKNOWLEDGEMENTS

This research was funded by grant FFI2017-84479-P AEI/FEDER from the Spanish Ministry of Economy and Competitiveness and a Visiting Scholars grant awarded to the third author by the UIB Research Office.

6. REFERENCES

- Baigorri, M., Campanelli, L. & Levy, E. S. (2018). Perception of American–English vowels by early and late Spanish–English bilinguals. *Language and Speech*, 62, 681-700. doi.org/10.1177/0023830918806933.
- Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167-244). Cambridge, MA: MIT Press.
- Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171-204). Baltimore: York Press.
- Bornstein, M. H. (1987). Perceptual categories in vision and audition. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 287–300). Cambridge: CUP.
- Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics*, 34(3), 372–387.
- Cebrian, J. (2019). Perceptual assimilation of British English vowels to Spanish monophthongs and diphthongs. *Journal of the Acoustical Society of America*, 145, EL52–EL58.
- Eimas, P. D., Miller, J. L. & Jusczyk, P. W. (1987). On infant speech perception and the acquisition of language. In S. Harnad (Ed.), *Categorical perception. The groundwork of cognition* (pp. 287–300). Cambridge: CUP.
- Escudero, P., & Chládková, K. (2010). Spanish listeners' perception of American and Southern British English vowels. *Journal of the Acoustical Society of America*, 128, EL254–EL259.
- Faris, M. M., Best, C. T. & Tyler, M. D. (2016). An examination of the different ways that non-native phones may be perceptually assimilated as uncategorized. *Journal of the Acoustical Society of America*, 139, EL1–EL5.
- Faris, M. M., Best, C. T. & Tyler, M. D. (2018). Discrimination of uncategorized non-native vowel contrasts is modulated by perceived overlap with native phonological categories. *Journal of Phonetics*, 70, 1–19.
- Hualde, J. I. (2005). *The sounds of Spanish*. Cambridge: CUP.
- Rallo Fabra, L. (2005). Predicting ease of acquisition of L2 speech sound sounds: A perceived dissimilarity test. *Vigo International Journal of Applied Linguistics*, 2, 75–92.
- Ridouane, R. (2014). Tashlhiyt Berber. *Journal of the International Phonetic Association*, 44, 207–221.
- Rost Bagudanch, A. (2019). Yeísmo in Majorcan Spanish: Phonetic variation in a bilingual context. *Zeitschrift für Romanische Philologie*, 135, 426–441.
- Safronova, E. (2016). *The Role of Cognitive Ability in the Acquisition of Second Language Perceptual Phonological Competence*. Unpublished PhD Dissertation. Universitat de Barcelona.
- Tyler, M. D., Best, C. T., Faber, A. & Levitt, A. G. (2014). Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica*, 71, 4–21.

FUNDAMENTACIÓN ARTICULATORIA Y AERODINÁMICA DE LOS PROCESOS FONOLÓGICOS. LAS ASIMILACIONES

Daniel Recasens

Departament de Filologia Catalana, Universitat Autònoma de Barcelona
daniel.recasens@uab.cat

ABSTRACT

Experimental data on heterorganic C#C sequences with /n/ and /t/ in coda position in Catalan indicate that regressive assimilations of place of articulation exerted by labials and velars are generally optional and more frequent in the case of the nasal than in that of the stop. Regressive manner assimilations in /t#C/ sequences before a nasal, lateral and a trill are favored by the homorganicity condition between the two consonants of the group. On the other hand, Catalan C#C sequences composed of (quasi)-homorganic consonants may be realized through two places of articulation, regressive assimilation or gestural blending depending on the degree of articulatory constraint and the order of occurrence of the two consecutive consonants. Also in Catalan, there is regressive voicing assimilation in sequences with one coda obstruent (C#C) but less clearly so in sequences with two obstruents in syllable-final position (CC#C).

Datos experimentales de grupos C#C con /n/ y /t/ en posición de coda del catalán indican que las asimilaciones regresivas de lugar de articulación ejercidas por labial y velar son generalmente optativas y resultan más frecuentes en el caso de la nasal que en el de la oclusiva. Las asimilaciones regresivas de modo de articulación en secuencias /t#C/ con C2 nasal, lateral o vibrante se ven favorecidas por la condición de homorganicidad entre las dos consonantes del grupo. Por su parte, los grupos C#C integrados por consonantes (casi)-homorgánicas muestran dos lugares de articulación, asimilación regresiva o mezcla gestual en función del grado de constreñimiento y el orden de aparición de las consonantes consecutivas. La asimilación regresiva de sonoridad actúa en grupos con una obstruyente en posición final de sílaba (C#C) y resulta incierta en secuencias de obstruyentes en posición de coda silábica (CC#C).

Keywords: Asimilaciones de lugar y modo de articulación, asimilación de sonoridad, catalán, constreñimiento articulatorio, electropalatografía, electroglotografía.

1. INTRODUCCIÓN

Las asimilaciones en secuencias de consonantes pueden ser vistas como fonologizaciones de efectos coarticulatorios. Se produce asimilación cuando una consonante adquiere las mismas propiedades articulatorias que la consonante adyacente ya desde su inicio. Este trabajo se ocupa de las asimilaciones de lugar de articulación (sección 2), modo de articulación (sección 3) y sonoridad (sección 4) en grupos de consonantes heterosilábicas del catalán. Los datos provienen especialmente de la publicación Recasens (2014), que no citamos nuevamente en las secciones siguientes.

2. ASIMILACIÓN DE LUGAR DE ARTICULACIÓN

Para adquirir una comprensión adecuada de los procesos asimilatorios de lugar de articulación en grupos de consonantes heterosilábicas conviene distinguir si las dos consonantes del grupo son heterorgánicas o más o menos homorgánicas. En el primer caso, los articuladores son independientes entre sí y, por lo tanto, pueden superponerse en el tiempo (p. ej., /tp/, /kt/, /pk/); en el segundo, las dos consonantes se articulan mediante el mismo articulador o articuladores contiguos, como el ápice y la lámina de la lengua o bien la lámina y el predorso (p. ej., /lt/, /sp/).

2.1. Grupos de consonantes heterorgánicas

La probabilidad de que tenga lugar una asimilación regresiva de C1 a C2 en grupos como /np/ y /nk/ varía en progresión directa con el grado de superposición de los gestos articulatorios de las dos consonantes y también con el grado de reducción articulatoria de C1, i.e., a mayor brevedad y menos contacto linguopalatal durante C1, más coarticulación y mayor probabilidad de asimilación.

Las asimilaciones en consideración no son siempre completas sino que pueden resultar también parciales cuando comportan un cambio discreto de lugar de articulación a la vez que se produce algún residuo gestual de la consonante asimilada. Así, por ejemplo, en inglés, /n/ puede asimilarse a /k/ y por lo tanto exhibir una oclusión dorsovelar desde su inicio, y a su vez, puede presentar contacto lingual a ambos lados del paladar hasta la zona alveolar lo cual indica que el predorso y quizás la lámina adoptan una posición relativamente elevada (Ellis & Hardcastle, 2002). Dicha elevación predorsal y laminal es propia de /n/ pero no de /k/ por cuanto la nasal alveolar se articula mediante la región lingual frontal y la producción de la oclusiva dorsovelar comporta descenso predorsal y laminal.

Estudios electropalatográficos (EPG) del catalán, italiano y español indican que la asimilación de lugar de articulación en secuencias /n#C/ con C2 labial o velar es generalmente completa puesto que C1 no presenta ningún residuo gestual de la alveolar nasal (Celata et al., 2013, Martínez Celdrán & Fernández Planas, 2007: 131-134). No obstante, en el caso de la secuencia /n#k/ del francés, no hay asimilación de ningún tipo sino superposición de los gestos linguales anterior de /n/ y dorsal de /k/ durante la segunda mitad del periodo oclusivo de la consonante nasal (Steele et al., 2016). Por otro lado, según datos del catalán, el proceso asimilatorio en secuencias como *són pobres* ‘son pobres’ o *són catalans* ‘son catalanes’ no es obligatorio sino que su implementación está condicionada por el hablante y la clase de palabra (la asimilación es especialmente factible en vocablos de uso frecuente como *un* ‘un’ y *són* ‘son’).

En catalán, datos de EPG muestran que la asimilación de lugar de articulación ocurre en menor medida en secuencias con /t/ (*set cases* ‘siete casas’, *set bosses* ‘siete bolsas’) que en grupos con /n/. La frecuencia de asimilaciones en grupos /t#C/ depende de los mismos factores citados para /n#C/ (i.e., el proceso asimilatorio afecta especialmente la consonante final de vocablos como *set* ‘siete’), así como de las características articulatorias de C2 (i.e., hay más asimilaciones ante consonante velar sonora que ante velar sorda o ante labial sorda o sonora).

Una mayor propensión a la asimilación regresiva de lugar de articulación en el caso de /n/ que en el de /t/ guarda relación con diferencias de presión intraoral motivadas por el hecho de que el aire procedente de los pulmones se almacena en la cavidad oral durante /t/ pero durante /n/ y, en consecuencia, en comparación con la primera consonante, la segunda presenta una explosión más breve y menos intensa, así como menos contacto linguopalatal y más coarticulación vocálica. Factores aerodinámicos relacionados con las diferencias de duración y de contacto linguopalatal entre obstruyentes sordas y sonoras pueden explicar la mayor frecuencia de asimilaciones en el caso de /t#g/ que en el de /t#k/.

2.2 Grupos de consonantes (casi)-homorgánicas

Las consonantes (casi)-homorgánicas en contacto se interfieren entre sí, y pueden resolverse mediante dos lugares de articulación, asimilación regresiva de lugar de articulación o mezcla de gestos (véase la definición de este concepto en el apartado 2.2 (c)). La solución fonética final de estos grupos de consonantes está condicionada por el grado de constreñimiento articulatorio y el orden de aparición de las dos consonantes consecutivas. Por lo que respecta al primer factor, las fricativas /s, ʃ/ y la rótica /r/ (sobretudo si es vibrante múltiple) son más constreñidas y a la vez más resistentes a la coarticulación vocálica que las oclusivas /t, d/ y las nasales /n, ñ/. Por su parte, la formación de uno o dos canales laterales hace que /l, ʎ/ sean bastante constreñidas, si bien menos que las fricativas y la vibrante. Es interesante constatar además que estas diferencias de grado de constreñimiento pueden estar en consonancia con diferencias de lugar de articulación: en catalán, /s, r/ son invariablemente centroalveolares o postalveolares, mientras que /n/ y en menor medida /l/ se articulan en la zona alveolar anterior, central o posterior según el contexto vocálico.

(a) C1 no constreñida, C2 constreñida

Secuencias como /ts, tʃ, tr, ns, nʃ, nr/ (*set sous* ‘siete salarios’, *un xai* ‘un cordero’, *són reis* ‘son reyes’) presentan asimilación regresiva completa y de aplicación obligatoria, con realización alveolar central o posterior de /t/ y /n/, que en otros contextos y posiciones se articulan en un lugar más anterior, ya sea de forma invariable (/t/) o no (/n/). Los datos de EPG de la Figura 1 muestran la asimilación de /n/ a /s, ʃ, r/ y, por lo tanto, la identificación del lugar de la oclusión de la nasal con el lugar de la constricción de las dos fricativas y la vibrante desde el inicio de C1. Así pues, en secuencias como /nr/, hay que admitir la

existencia de asimilaciones dentro de una misma zona articuladora tradicional (i.e., la zona alveolar en este caso).

(b) *C1 constreñida, C2 no constreñida*

En combinaciones como /st, ft, rt/ (*està tort* ‘está torcido’, *eix tort* ‘eje torcido’, *mar Tirrè* ‘mar Tirreno’), las dos consonantes del grupo mantienen su lugar de articulación propio si bien puede ocurrir coarticulación retardatoria y realización exclusivamente alveolar de C2. El hecho de que, a diferencia de grupos como /ts, tʃ, tr/, la consonante constreñida no prevalezca de forma habitual sobre la no constreñida guarda relación con la naturaleza regresiva de las asimilaciones en general. Por su parte, en los grupos /sn, fn, rn/, /n/ puede realizarse centroalveolar o postalveolar en virtud de la aplicación de un mecanismo coarticulatorio y, en consecuencia, de forma no sistemática.

(c) *C1 y C2 no constreñidas*

Los grupos de consonantes no constreñidas se realizan mediante mezcla de gestos. En grupos como /nʎ, nt, nɲ/ (*un llamp* ‘un relámpago’, *any terrible* ‘año terrible’, *any nou* ‘año nuevo’), este proceso se manifiesta a través de la formación de una única oclusión que engloba las áreas de oclusión de C1 (alveolopalatal) y de C2 (dentoalveolar, o alveolar más o menos anterior). Una particularidad relevante del mecanismo de mezcla gestual es que no actúa desde el inicio de C1. Así, según la Figura 2, /n/ del grupo /nʎ/ empieza siendo alveolar y, más tarde, presenta oclusión simultánea en las zonas alveolar y palatal. Asimismo, secuencias como /nt, lt/, que, según las descripciones tradicionales, registran cambio de C1 alveolar a dental, muestran a menudo mezcla gestual con oclusión lingual simultánea en los incisivos de la mandíbula superior y en la zona alveolar anterior.

(d) *C1 y C2 constreñidas*

El resultado fonético de las secuencias /rs, rʃ, sr, sʃ, fr, fʃ/ (con realización vibrante múltiple de C2= /r/) permite asegurar que la consonante rótica es más constreñida que /ʃ/, y que /s/ es menos constreñida que las otras dos consonantes. Efectivamente, /r/ no se asimila ni a /s/ ni a /ʃ/ ([rs] *mar salada* ‘mar salada’, [rʃ] *llar xinesa* ‘hogar chino’); /s/ suele asimilarse a /r/ y a /ʃ/ ([r(:)] *dos russos* ‘dos rusos’, [f(:)] *dos xais* ‘dos corderos’), mientras que /ʃr/ presenta generalmente asimilación regresiva (*peix rodó* ‘pez redondo’) y /ʃs/, mezcla de gestos (*peix salat* ‘pescado salado’). Una mayor propensión a la asimilación en el caso de /sʃ/ que en el de /ʃs/ también se da en francés y en inglés (Holst & Nolan, 1995, Nolan et al., 1996, Niebuhr et al., 2011, Niebuhr & Meunier, 2011). El

resultado fonético de /ʃs/ en catalán es una realización fricativa con lugar de la constricción intermedio entre los lugares de /s/ y /ʃ/. Además, esta realización se observa bien avanzado el grupo (véase el gráfico derecho de la Figura 3) pero no al inicio del mismo donde la posición lingual es muy parecida a la de /ʃ/ (gráfico izquierdo de la misma figura).

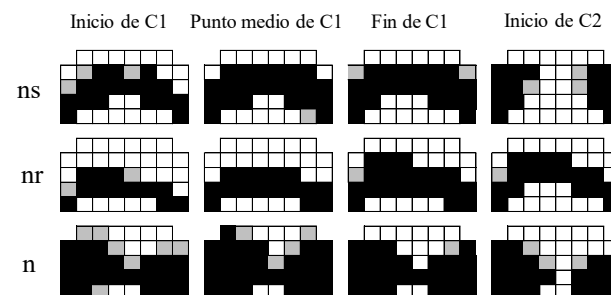


Figura 1: Configuraciones de EPG correspondientes a cuatro momentos temporales de /ns, nr, nʃ/ según un hablante de catalán central. Las cuatro filas de electrodos superiores corresponden a la zona alveolar (la primera empezando por arriba es la más anterior) y la quinta a la zona prepalatal.

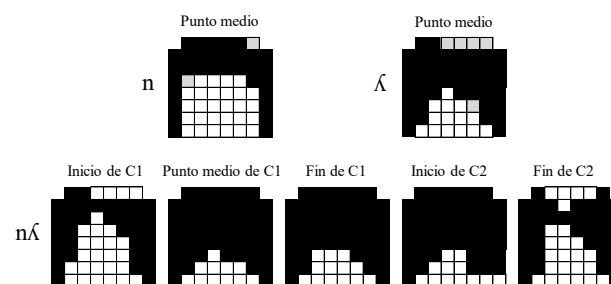


Figura 2: Configuraciones de EPG correspondientes al punto medio de /n/ y /nʎ/ en las secuencias /ana/ y /aʎa/ (parte superior) y a cinco momentos consecutivos durante la secuencia /nʎ/ (parte inferior) según un hablante de catalán central. Las cuatro filas de electrodos superiores corresponden a la zona alveolar y la cuatro inferiores a la zona palatal.

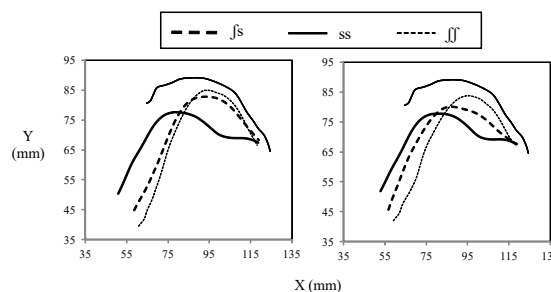


Figura 3: Configuraciones lingüales correspondientes a un 25% (gráfico izquierdo) y un 75% (gráfico derecho) del periodo de fricción de /ʃ#s/, /s#s/ y /ʃ#ʃ/ según un hablante de catalán central. Los ejes X e Y indican las distancias en mm en relación con el origen del campo de ultrasonidos.

3. ASIMILACIÓN DE MODO DE ARTICULACIÓN

En catalán, se puede producir asimilación regresiva de modo de articulación en grupos con oclusiva seguida de nasal, lateral o vibrante.

Datos de EPG correspondientes a secuencias oclusiva + nasal y oclusiva + lateral indican que esta asimilación es opcional, y se aplica más a menudo a los grupos /pm/, /tn/, /tl/, /tʎ/ donde C1 y C2 comparten el mismo lugar bilabial ([m:], *cap mort* ‘ningún muerto’), dentoalveolar ([n:, l:], *set nois* ‘siete chicos’, *set làmines* ‘siete láminas’) o dento-alveolopalatal ([ʎ:], *set llamps* ‘siete relámpagos’) que a los grupos heterorgánicos /pn/, /tm/, /km/, /kn/ (*cap noi* ‘ningún chico’, *set morts* ‘siete muertos’, *sac marró* ‘saco marrón’, *sac nou* ‘saco nuevo’). La condición de homorganicidad es una característica subyacente en el caso de /pm/, mientras que en /tn/, /tl/, /tʎ/ resulta de la aplicación de un proceso de mezcla de gestos. Las asimilaciones de modo se ven favorecidas por la condición de homorganicidad, de manera que resulta especialmente factible crear pasajes laterales (/tl/, /tʎ/) y promover el descenso del velo del paladar (/tn/) cuando el articulador lingual primario se mantiene inalterado a lo largo de la secuencia de dos consonantes.

Por otro lado, la oclusiva /k/ es más refractaria que /p/ y /t/ a convertirse en nasal delante de nasal por dos motivos. En primer lugar, un nivel elevado de presión intraoral condicionado por la formación de una cavidad posterior al lugar de articulación de dimensiones especialmente reducidas dificulta el descenso del velo. En segundo lugar, por lo que respecta a /kn/, un control estricto de la actividad del cuerpo lingual durante la consonante dorsal puede dificultar la anticipación del gesto lingual anterior de /n/.

También en el caso de /s#r/ (*dos russos* ‘dos rusos’, *vas rus* ‘vaso ruso’) resulta válida la hipótesis según la cual la asimilación de modo de articulación con resultado [r(:)] es facilitada por la condición de homorganicidad entre las dos consonantes centroalveolares o postalveolares (véase la sección 2.2). En este caso la aplicación del proceso depende del ritmo de habla, de la frecuencia de uso de la palabra que contiene la consonante fricativa y de si la frontera de palabra es también frontera de frase o no (Solé, 1999).

4. ASIMILACIÓN DE SONORIDAD

En catalán, la fonética descriptiva y la fonología contemplan la aplicación de un proceso de sonorización regresiva en virtud del que el valor del rasgo de sonoridad de la última consonante de un

grupo determina si las obstruyentes (una o más) precedentes finales de sílaba se realizan sonoras o sordas: [bd] *sap dir mentides* ‘sabe mentir’, [pk] *sap coses* ‘sabe cosas’, [ps] *en saps* ‘sabes de ello’, [bzð] *saps dibuixar* ‘sabes dibujar’, [psk] *saps coses* ‘sabes cosas’. Conviene indicar que, en el caso de la sonoridad, el grado de constreñimiento de las consonantes está condicionado por los requisitos aerodinámicos en la medida que pueden facilitar o dificultar la vibración de las cuerdas vocales durante un periodo más o menos largo. Así, las consonantes fricativas resultan especialmente constreñidas dado que la vibración glotal durante su producción se encuentra en conflicto con el paso de un volumen considerable de aire proveniente de los pulmones a través de la glotis.

Con el fin de investigar la validez de esta caracterización descriptiva del proceso de asimilación de sonoridad grabamos datos electroglotográficos de grupos C#C y CC#C con /p, t, k, f, s, ʃ/ y /ps, ts, ks, tʃ, sp, sk/ en posición de coda silábica, y una consonante sonora de modo oclusivo, fricativo, nasal, lateral, rótico y aproximante en posición inicial de la sílaba siguiente. Ocho catalanohablantes produjeron estas secuencias en frases con sentido.

4.1 Grupos C#C

El gráfico superior de la Figura 4 muestra que el porcentaje de sonoridad correspondiente a C1 de los grupos C#C suele ser inferior al 100 %, y que mientras que algunos hablantes se aproximan a esta cifra (VA, DR), otros sonorizan mucho menos (EV, SO). Además y de acuerdo con las expectativas iniciales, se observa un porcentaje de sonoridad inferior en grupos con C1 fricativa, que corresponde a la media de los porcentajes de /f/, /s/ y /ʃ/ (línea discontinua), que en grupos con C1 oclusiva, que corresponde a la media de los porcentajes de /p/, /t/ y /k/ (línea continua).

En principio, estos datos no nos informan sobre si hay asimilación de sonoridad de C1 a C2 o no. Consideramos que para que haya asimilación regresiva, C1 debe presentar vibración glotal desde su inicio con independencia de que pueda haber interrupción de la sonoridad más adelante durante esta misma consonante. Además, puesto que parte del periodo de sonoridad existente al inicio de C1 puede deberse a la vocal precedente, será también necesario que, para poder asegurar que hay asimilación, este periodo sea más largo que el periodo de sonoridad atribuible a la vocal, que suele denominarse VTT (‘voice termination time’) en inglés. La duración del VTT puede establecerse a partir del análisis de la sonoridad en secuencias C#C con C2 sorda (p. ej.,

[pk] *sap coses* ‘sabe cosas’), dado que en este caso C1 se realiza sorda y el periodo de sonoridad existente al inicio de esta consonante está asociado necesariamente y de forma exclusiva con la vocal precedente. En suma, solo en el caso de que la duración del periodo de sonoridad existente al inicio de C1 sea mayor que la correspondiente al VTT podremos asegurar que nos encontramos ante un proceso asimilatorio con independencia de si el resto de C1 presenta o no vibraciones glotales.

El gráfico inferior de la Figura 4 aporta datos sobre esta cuestión. Muestra, separadamente para los grupos C#C con C1 oclusiva oral y con C1 fricativa en posición final de sílaba, el porcentaje de casos a través de grupos de consonantes y repeticiones de cada grupo en el que cada uno de los hablantes analizados presenta un periodo de sonoridad al inicio de C1 cuya duración excede la duración del VTT. Si asumimos que aquellos hablantes con más de un 80% de casos programan C1 como sonora, entonces podemos concluir que prácticamente todos los catalanohablantes analizados exceptuando SO presentan asimilación regresiva de sonoridad.

4.2 Grupos CC#C

Los gráficos superior y medio de la Figura 5 presentan los porcentajes de sonoridad durante C1 y C2 en grupos CC#C, y ello separadamente para C1 oclusiva (línea continua) y C1 fricativa (línea discontinua). Una comparación con los datos de la Figura 4 muestra que, por regla general, estos porcentajes son inferiores a los registrados en los grupos C#C por cuanto un incremento del número de obstruyentes promueve un ascenso del nivel de presión intraoral y dificulta la vibración de las cuerdas vocales. También se observa aquí un periodo de vibración glotal inferior durante las consonantes fricativas que durante las oclusivas.

Por su parte, el gráfico inferior de la Figura 5 muestra para cada hablante el porcentaje de casos en que los grupos CC#C presentan un periodo de sonoridad al inicio de C1 más largo que el VTT. Asumimos también aquí que los hablantes con porcentajes de un 80% o superior programan C1 como sonora y en consecuencia asimilan C1 (y C2) a C3. Los resultados indican un escenario menos claro que el expuesto en la Figura 1 dado que muestran la presencia de tres hablantes con porcentajes superiores al 80% (MA, LO, MO) y cinco con porcentajes bajos especialmente si C1 es fricativa (EV, PE, SO, DR, VA). A tenor de estos datos no podemos asegurar que se dé una asimilación regresiva mayoritaria al menos de C1 a C3 en los grupos CC#C que nos ocupan.

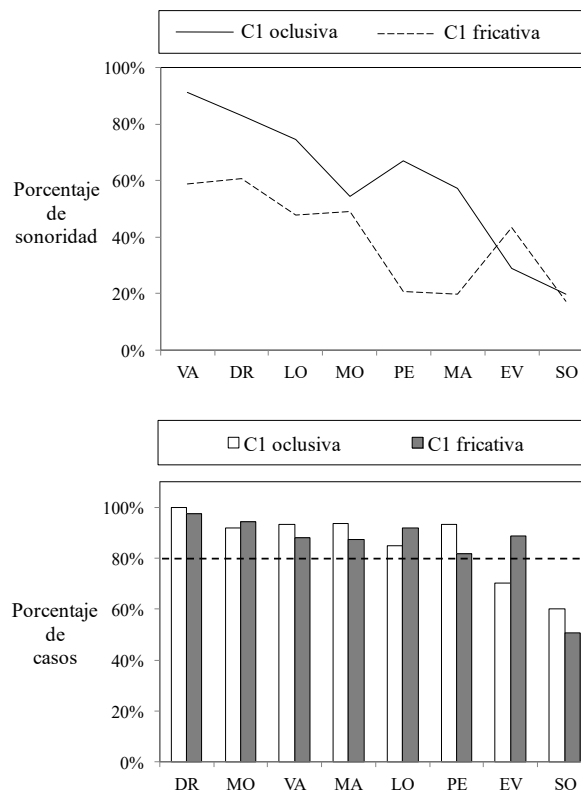


Figura 4: Sonorización regresiva en grupos C#C por parte de los hablantes VA, DR, LO, MO, PE, MA, EV y SO. (Gráfico superior) Porcentaje de sonoridad durante C1 oclusiva y C1 fricativa. (Gráfico inferior) Porcentaje de casos con un periodo de sonoridad al inicio de C1 cuya duración excede la duración del periodo de sonoridad atribuible a la vocal precedente.

Pueden sugerirse diversas hipótesis para explicar esta situación. Una posibilidad es que todos los hablantes programen C1 y C2 como sonoras y que los elevados requisitos aerodinámicos de ambas obstruyentes en posición de coda impidan la vibración de las cuerdas vocales de forma considerable. También podría darse un proceso de cambio en virtud del cual los hablantes más conservadores sonorizan todo el grupo CC#C, mientras que los menos conservadores tienden a no sonorizar las dos consonantes en posición de coda silábica. En este caso la regla de sonorización regresiva se aplicaría a todos o la mayoría de hablantes en grupos con una obstruyente final de sílaba (C#C) y solo a un subgrupo de hablantes en grupos con dos obstruyentes finales (CC#C).

5. CONCLUSIONES

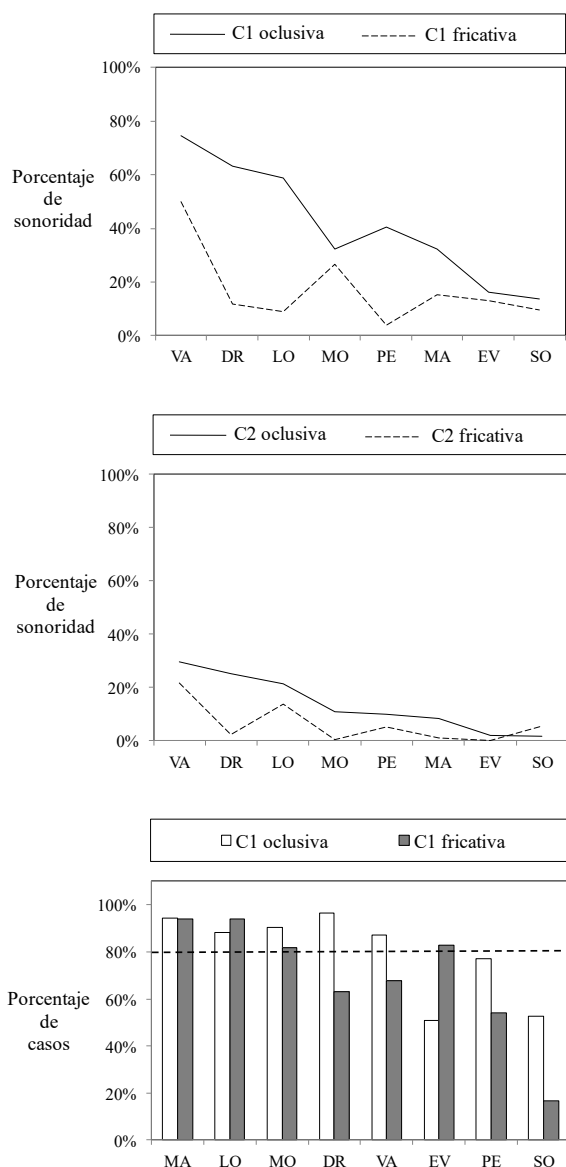
Los datos de producción fonética de grupos de dos consonantes a través de frontera de palabra del catalán analizados en este estudio confirman que, tal como han constatado las descripciones tradicionales y la fonología generativa, en secuencias C#C con C1 oclusiva oral dentoalveolar y nasal alveolar, las asimilaciones regresivas de lugar de articulación son

generalmente optativas, y las asimilaciones de modo, también optativas, dependen de la condición de homorganicidad entre C1 y C2. Se concluye asimismo que la asimilación de lugar de articulación es más frecuente si C1 es /n/ que si es /t/. Por su parte, los grupos C#C integrados por consonantes (casi)homorgánicas muestran dos lugares de articulación,

las consonantes en contacto, y las asimilaciones de lugar de articulación pueden operar en una misma zona articulatoria tradicional (p. ej., /n/ se realiza obligatoriamente alveolar central o posterior ante /s, r/ en catalán). Finalmente, los datos electroglotográficos indican que, mientras que la asimilación regresiva de sonoridad es esencialmente obligatoria en secuencias C#C, su aplicación en secuencias CC#C resulta incierta en más de la mitad de los hablantes analizados.

6. BIBLIOGRAFÍA

- Celata, C., Calamai, S., Ricci, I. & Bertini, C. (2013). Nasal place assimilation between phonetics and phonology: An EPG study of Italian nasal-to-velar clusters. *Journal of Phonetics*, 41, 88-100.
- Ellis, L. & Hardcastle, W.J. (2002). Categorical and gradient properties of assimilation in alveolar to velar sequences: Evidence from EPG and EMA data. *Journal of Phonetics*, 30, 373-396.
- Holst, T. & Nolan, F. (1995). The influence of syntactic structure on [s] to [ʒ] assimilation, en Connell, B. & Arvaniti, A. (eds.). *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*. Cambridge: Cambridge University Press, 315-333.
- Martínez Celdrán, E. & Fernández Planas, A.M. (2007). *Manual de fonética española*. Barcelona: Ariel.
- Niebuhr, O. & Meunier, C. (2011). The phonetic manifestation of French /s#f/ and /ʃ#s/ sequences in different vowel contexts: On the occurrence and the domain of sibilant assimilation. *Phonetica*, 68, 1-28.
- Niebuhr, O., Clayards, M., Meunier, C. & Lancia, L. (2011). On place assimilation in sibilant sequences - comparing French and English. *Journal of Phonetics*, 39, 429-445.
- Nolan, F., Holst, T. & Kühnert, B. (1996) Modelling [s] to [ʒ] accommodation in English. *Journal of Phonetics*, 24, 113-137.
- Recasens, D. (2014) *Fonètica i fonologia experimentals del català*. Barcelona: Institut d'Estudis Catalans.
- Recasens, D. & Mira, M. (2014) Gradient phonetic implementation of regressive voicing assimilation in Catalan heterosyllabic two- and three-consonant clusters. *Phonetica*, 71, 128-156.
- Recasens, D. & Mira, M. (2015) Place and manner assimilation in Catalan consonant clusters. *Journal of the International Phonetic Association*, 45, 115-147
- Solé, M..J. (1999). Production requirements of apical trills and assimilatory behavior, en Ohala, J. J., Hasegawa, Y., Ohala, M., Granville, D. & Bailey, A. C. (eds.). *Proceedings of the 14th ICPHS*, vol. 1, San Francisco, 77-94.
- Steele, J., Colantoni, L. & Kochetov, A. (2016). Gradient assimilation in French cross-word /n/+velar stop sequences. *International Phonetic Association*, 49, 87-108.



asimilación regresiva o mezcla gestual en función del grado de constreñimiento y el orden de aparición de

LOS RASGOS PROSÓDICOS DE LOS NEOLECTORES: UN ESTUDIO MULTIDIMENSIONAL CON UN ENUNCIADO INTERROGATIVO

Sara Recio-Pineda¹, Alicia Sola², Francisco José Cantero Serena³

Universitat de Barcelona (UB)
srecio@ub.edu, aliciasola@ub.edu, cantero@ub.edu

ABSTRACT

En este trabajo, mostramos los resultados de un estudio que realiza el análisis de los diferentes rasgos prosódicos (melódicos, dinámicos y rítmicos) de un enunciado interrogativo parcial leído por 24 neolectores de 2º curso de Primaria en el transcurso de un ejercicio de lectura. Se trata de un análisis multidimensional en el que se reúnen datos que expresan las relaciones entre los parámetros físicos de primer orden (el tono y la intensidad) de los centros fónicos y sus distancias (Cantero, 2019). A partir de este análisis establecemos las correlaciones existentes entre estos parámetros para interpretar el valor o significado que aportan al contorno entonativo.

Palabras clave: prosodia, melódico, dinámico, rítmico, neolectores

In this paper, we display the results of a study that performs a complete analysis of all the different prosodic features (melodic, dynamic and rhythmic) of a partial interrogative statement read by 24 students in the course of a reading aloud text exercise. It is a multidimensional analysis in which data expresses the relationships between the first-order physical parameters (tone and intensity) of phonic centres and their distances (Cantero, 2019). From this analysis, we establish the existing correlations among these parameters in order to interpret their value or significance into the intonation contour.

Keywords: prosody, melodic, dynamic, rhythmic, young readers

1. INTRODUCCIÓN

El análisis de la prosodia se ha centrado tradicionalmente en el análisis de la entonación. Esto ha hecho que la variable central de estudio haya sido la secuencia de los intervalos de tono más relevantes que se dan en un enunciado.

Con el método de Análisis Melódico del Habla (véase Cantero y Font-Rotchés, 2020) y utilizando un instrumento de análisis acústico como *Praat*, se identifican y aíslan los centros tonales (vocales) de las unidades de habla que se quieren analizar (grupos fónicos); se determina su valor central de F_0 en Hz; se establece que el primer valor de la primera vocal es un punto relativo del 100%; se mide el intervalo entre este primer punto relativo y el que sigue (el siguiente valor central de F_0 en Hz de la siguiente vocal) en porcentajes y, por último, se hace la representación gráfica del contorno.

Con la aplicación de esta primera fase del método (fase acústica) se obtiene la estandarización de los contornos melódicos de cada unidad de análisis. De

este modo, se pueden describir las melodías del habla y compararse independientemente de la edad, el sexo o cualquier otra característica del informante. Después de la fase acústica, se comprueba la validez de los resultados mediante una serie de pruebas perceptivas.

Para este trabajo, merece la pena destacar que en el contexto del *Grup de Recerca en Entonació i Parla* (GREP), y gracias a este método, se han podido establecer los patrones del catalán y sus variedades dialectales (Font-Rotchés, 2007, 2009). También se han podido describir diversas interlenguas fónicas (véase, por ejemplo, Liu, 2005; Cantero y Devís, 2011; Devís, 2011; Baditzné, 2012; Martorell, 2014 o Fonseca, 2013) y se ha podido establecer un marco para el análisis de la entonación emocional (Cantero, 2014) o la entonación de cortesía (Devís, 2020).

Recientemente, Recio-Pineda, Sola & Cantero (2020), utilizando este método, han mostrado los patrones melódicos y los rasgos de énfasis de un enunciado interrogativo parcial en catalán leído por 24 participantes en edad de adquisición y de aprendizaje inicial de la lectura. Este trabajo ha

demostrado que la interrogativa de los neolectores está afectada por rasgos muy enfáticos. El ascenso tonal que se detectó al final del 41,67% de los casos no servía para añadir un valor modal, y tampoco las entonaciones enfáticas de núcleo elevado (20,83%) tenían un valor intencionado.

Los énfasis de palabra detectados estaban provocados por dificultades en la descodificación: por ende, los rasgos de énfasis podían considerarse un indicio de poca competencia lectora. Por una parte, el fenómeno se interpretó como efecto del contagio de modelos pedagógicos de entonación que trabajan la entonación interrogativa desde ejemplos, aislados y marcados pedagógicamente, de interrogativa absoluta; por otra parte, y al tratarse de una interrogativa parcial, se consideró que la lectura mostraba un cierto grado de afectación. De hecho, la frase incluye un vocativo final en el que no se observa el efecto de eco esperable (Bonet, 1986): en la práctica totalidad de los casos, el vocativo se comporta como un elemento más del enunciado, sin un esquema melódico propio (véase un ejemplo en la figura 1):

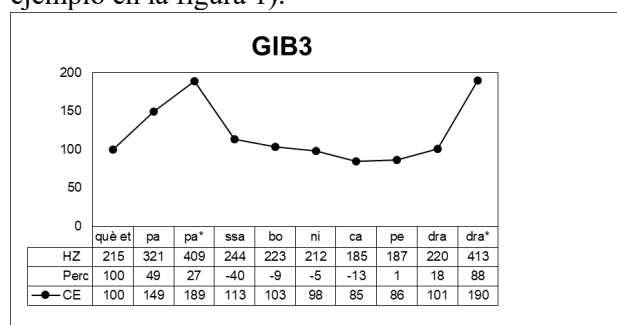


Figura 1: El vocativo “bonica pedra” no tiene un esquema melódico propio.

Para obtener un mayor entendimiento de los fenómenos paralingüísticos de afectación y contagio de la entonación de los neolectores, entonces, se advirtió que, más allá de la melodía, era interesante indagar en otros rasgos prosódicos que intervenían con fuerza, como la intensidad y el ritmo. Por ello, en este trabajo pondremos a prueba el análisis dinámico y el análisis rítmico de esta interrogativa parcial en catalán leída por 24 informantes de 2º curso de Primaria.

2. METODOLOGÍA

De acuerdo con Cantero (2019, p.491-494), para el análisis de la intensidad (o análisis dinámico) se identifican primero, y del mismo modo que en el análisis melódico, los segmentos tonales del enunciado. En estos segmentos, se toma el valor de intensidad máximo. Estos valores (dB) se convierten en porcentajes relativos, para establecer intervalos de intensidad desde el primer punto hasta cada uno de los centros tonales. La relación de intervalos constituye una curva estándar de intensidades.

Para el análisis rítmico se parte también de la identificación de los valores relevantes de tono e intensidad del enunciado: se toma el valor de tiempo absoluto en el momento en que ocurre el valor de máxima intensidad y se mide su distancia respecto al siguiente. A estas distancias se las denomina “pie rítmico”. A continuación, se expresan las distancias en porcentajes y se estandarizan tomando como valor de referencia (100) el primer centro tonal. El gráfico para expresar las relaciones rítmicas es de barras.

Para completar el análisis, se realiza la correlación de estos parámetros prosódicos (melodía, ritmo y entonación) y se obtiene una visión dinámica del valor que cada una de estas dimensiones añade a la entonación.

2.1. Corpus

El corpus (Recio-Pineda, 2017) contiene más de 36 horas de grabación de niños entre los 7 y los 12 años leyendo un relato adecuado a su edad y contestando, después, a preguntas de comprensión en una entrevista oral. De este corpus, se ha extraído un enunciado que aparece leído por los 24 informantes de Ciclo Inicial (24) y que corresponde a una interrogativa parcial: *què et passa, bonica pedra?* [¿qué te pasa, bonita piedra?].

La muestra de lectores es equilibrada y heterogénea en niveles de lectura, pues contiene 8 lectores con un nivel alto de lectura, 8 lectores con un nivel medio y 8 lectores con un nivel bajo, según indicaciones de los maestros. Excluye a lectores con dificultades de aprendizaje debidas a algún diagnóstico como dislexia, hipoacusia o TDHA.

3. RESULTADOS

Mostramos, a continuación, algún ejemplo de los resultados.

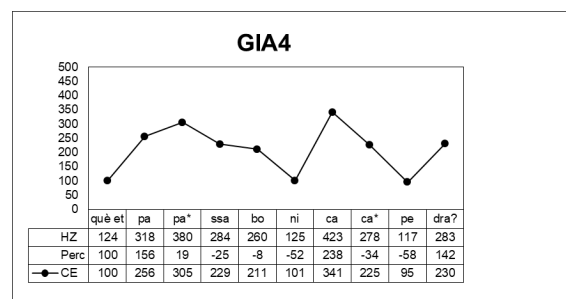


Figura 2: Análisis melódico del enunciado leído por un informante con competencia lectora alta.

De acuerdo con lo que muestra la figura 2, este informante lee utilizando lo que Font-Rotchés (2009) identifica como patrón entonativo 3 de las interrogativas parciales del catalán. Este patrón es muy poco común en su corpus de habla espontánea

adulta, ya que se localiza solo en un 3,7% de los casos¹. Sin embargo, se trata de un patrón que aparece en el 25% de los casos del corpus de este trabajo (lectura de niños en una etapa de adquisición y aprendizaje inicial de las habilidades de lectoescritura).

El patrón 3, con el que se corresponden los ejemplos de las figuras 1 y 2, se caracteriza por mostrar un ascenso en la inflexión final de más de un 80% (en estos ejemplos, son del 88% y del 142%, respectivamente).

En el caso de la figura 2, el informante lee el enunciado con un énfasis de primer pico desplazado y un énfasis de palabra en postónica. Este último énfasis, lejos de tener un valor modal intencionado, puede responder al hecho de que el informante cree que la frase acaba antes (en *bonica*); después se da cuenta de que la frase sigue y duplica el patrón entonativo hasta el final (*pedra*).

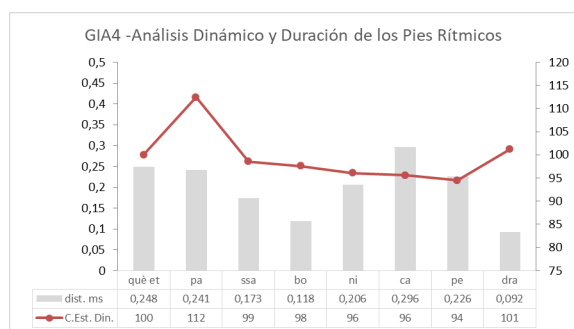


Figura 3: Análisis dinámico y rítmico del mismo enunciado e informante de la figura 2 (GIA4).

En el gráfico de la figura 3, puede observarse que, a diferencia del tono, la intensidad se mantiene estable, pero que experimenta un pico muy alto que coincide con el énfasis de primer pico desplazado en el tono. Llama la atención que allí donde la figura 2 mostraba una inflexión de tono muy acusada (énfasis de palabra en postónica), aquí no hay variación de la intensidad relevante. Por último, la intensidad asciende moderadamente en la inflexión final.

En la correlación de los tres rasgos prosódicos (tono, intensidad y duración), se comprueba que ni el primer pico tonal desplazado, ni su correspondiente énfasis dinámico se corresponden con una mayor duración de la vocal. El énfasis tonal de palabra en postónica no lleva asociado un ascenso en la intensidad, pero sí lleva asociada una mayor duración de la vocal que experimenta un intervalo de tono. Esto no ocurre así en la inflexión final, que experimenta un ascenso tonal muy acusado acompañado de un ascenso

moderado de la intensidad, pero que no va asociado a una mayor duración de la vocal.

En la figura 4 observamos los análisis dinámico y rítmico correspondientes al ejemplo de la figura 1:

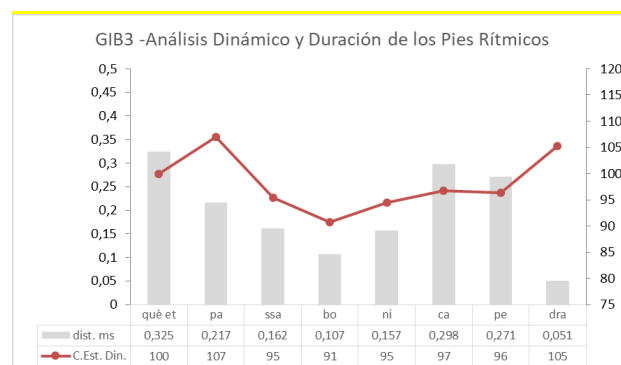


Figura 4: Análisis dinámico y rítmico del mismo enunciado e informante de la figura 1 (GIB3).

En este caso, el perfil dinámico reproduce muy fielmente el perfil melódico del enunciado, con un pico de intensidad en *passa* (paralelo al primer pico tonal) y un final ascendente. En cuanto a las duraciones, en cambio, el desajuste es claro: la sílaba con pico tonal y pico dinámico (*passa*) es más breve que la sílaba anterior (*què et*) y, como en el ejemplo anterior, la sílaba final de *bonica* es más larga que las vocales tónicas adyacentes.

4. CONCLUSIONES

El estudio de la prosodia comprende fenómenos tales como el acento, el ritmo y la entonación de las unidades del habla. Para estudiarla en su conjunto, es necesario tener en cuenta las tres variables fonéticas que tienen alguna funcionalidad lingüística: el tono, la intensidad y la duración.

En este trabajo, mostramos los patrones dinámicos y los patrones rítmicos de los enunciados analizados, los relacionamos con los patrones melódicos y trazamos los fenómenos de correlación entre las tres dimensiones prosódicas, con especial atención a los puntos de sincronía entre las tres dimensiones (no necesariamente ubicados en las vocales tónicas). Entre las conclusiones de la investigación, es particularmente relevante la constatación de que los desajustes entre el perfil dinámico y la melodía del enunciado deben interpretarse como un rasgo claro de énfasis prosódico.

Este trabajo es solo una primera aproximación a las características prosódicas de un grupo de neolectores leyendo un enunciado interrogativo parcial, en catalán. Es necesario seguir trabajando en este

¹ En el corpus Font-Rotchés (2007), los patrones que más ocurrencia tienen en las interrogativas parciales del catalán son

acercamiento para profundizar en el papel que juegan las diferencias tonales, de intensidad y de duración en cada punto del enunciado, especialmente en sus núcleos; y en las interrelaciones que se establecen entre el perfil melódico del enunciado y sus correlatos dinámico y rítmico, como índices objetivos de su competencia lectora.

5. BIBLIOGRAFÍA

- Baditzné, k. 2012. Spanish Intonation of Hungarian Learners of Spanish; Yes or No Questions, *Biblioteca Phonica*, 15.
- Bonet, E. 1986. L'entonació de les formes interrogatives en barceloní, *Els Marges*, 33, 103-117.
- Cantero, F. J. 2014. Códigos de la entonación y entonación emocional, en A.Díaz, M.C. Fumero, M.P. Lojendio, S. Burgess, E. Sosa y A.Cano Ginés (eds.): *Actas del 31 Congreso Internacional AESLA*. La Laguna: Universidad de La Laguna, 618-629.
- Cantero, F. J. 2019. Análisis prosódico del habla: más allá de la melodía, en María Rosa Álvarez Silva; Alex Muñoz Alvarado & Leonel Ruiz Miyares (eds.): *Comunicación Social: Lingüística, Medios Masivos, Arte, Etnología, Folclor y otras ciencias afines*. Volumen II. Santiago de Cuba: Ediciones Centro de Lingüística Aplicada.
- Cantero, F.J. y Devís, E. 2011. Análisis melódico de la interlengua, en A. Hidalgo; Y. Congosto y M. Quilis (eds.): *El estudio de la prosodia en España en el siglo XXI: perspectivas y ámbitos*. Anejo nº 75 de la Revista Quaderns de Filologia. Universitat de Valencia. 285-299.
- Cantero, F. J. y Font-Rotchés, D. (2020). Melodic Analysis of Speech (MAS). *Phonetics of Intonation*. En: Abasolo, J.; de Pablo, I.; Ensunza, A.: *Contributions on education*. Universidad del País Vasco, 20-47.
- Devís, E. 2011. La entonación del español hablado por italianos, *Didáctica (Lengua y Literatura)*, vol.23. 35-58.
- Devís, E. 2020. La entonación de (des)cortesía en catalán, *Phonica*, vol. 16, 54-82.
- Fonseca, A. 2013. *Caracterización de la entonación del español hablado por brasileños*. Tesis doctoral, Universitat de Barcelona.
- Font-Rotchés, D. 2007. *L'entonació del català*, Biblioteca Milà i Fontanals, 53. Barcelona: Publicacions de l'Abadia de Montserrat.
- Font-Rotchés, D. 2009. Les interrogatives pronominals del català central. Anàlisi melòdica i patrons entonatiu, *Els Marges*, 87, 41-64.
- Liu, Y.H. 2005. La entonación del español hablado por taiwaneses. *Biblioteca Phonica*, 2.
- Martorell, L. 2014. Aproximació als trets melòdics de les interrogatives de l'espanyol parlat per suecs, *Phonica*, vol.13-14. 99-106.
- Recio-Pineda, S. 2017. Prosodia y comprensión lectora en Educación Primaria. Tesis doctoral, Universitat de Barcelona. En <http://diposit.ub.edu/dspace/handle/2445/117325>
- Recio-Pineda, S.; Sola, A.& Cantero, F.J. 2020. La interrogativa leída por escolares de Ciclo Inicial: afectación y contagio, *Libro de Resúmenes. XLIX Simposio de la Sociedad Española de Lingüística (SEL) 2020*.

EMERGÈNCIA FONÈTICA: CAL APRENDRE i ENSENYAR A PRONUNCIAR LA [ʎ]

A. Rius-Escudé, F. Torras

Facultat d'Educació. Universitat de Barcelona
agnesrius@ub.edu, ftorras@ub.edu

ABSTRACT

Presentarem els resultats d'una experiència d'ensenyament-aprenentatge amb els alumnes que cursen l'assignatura Llengua Catalana per a l'Ensenyament, de primer curs del Grau de Mestre d'Educació Primària de la Facultat d'Educació de la UB, la majoria dels quals no tenen el català com a segona llengua i presenten dificultats en la pronúncia del so lateral palatal sonor del català [ʎ].

Es tracta d'un pla d'acció inspirat en els principis del mètode verbotonol, en què es treballa a partir de l'audició, la percepció, la imitació i la producció. Per a portar-lo a la pràctica i per afavorir l'autonomia de l'estudiant, comptem amb la plataforma Moodle, les aplicacions del telèfon mòbil (Kahoot i codis QR) i pàgines web (els sons del català i les guies de pronunciació). La finalitat és afavorir la comunicació i aproximar les produccions orals a l'accent nadiu.

Paraules clau: pronunciació, so lateral palatal sonor del català [ʎ], català com segona llengua, aprenentatge mòbil

In this paper, we offer the results obtained by a teaching-learning experience tested on 81 students who attend the Catalan language subject on the first degree of primary teacher training (Faculty of Education. University of Barcelona). Most of the students have Catalan as a second language and have some difficulties to pronounce the Catalan sound [ʎ].

The experiment is based on the verbotonol method according to which hearing, perception, imitation and production is practised. To put into practise and promote learner empowerment Moodle platform, Kahoot and QR codes apps and *The Catalan sounds* and *Guides of pronunciation* websites will be used. The results show most of the students improve significantly their pronunciation and therefore they feel more confident communicating in Catalan an enhance their self-esteem.

Keywords: pronunciation, sound [ʎ], Catalan as a second language, mlearning

1. INTRODUCCIÓ

En aquesta comunicació, presentarem els resultats d'una experiència d'ensenyament-aprenentatge amb els alumnes que cursen l'assignatura *Llengua Catalana per a l'Ensenyament*, de primer curs dels Grau de Mestre d'Educació Primària a la Facultat d'Educació de la UB, la majoria dels quals no tenen el català com a llengua primera i presenten dificultats en la pronúncia del so lateral palatal sonor en català. Quan un alumne ha d'aprendre una segona llengua, entra en contacte amb un nou sistema fonològic. Generalment, accepta com a hàbit la solució més fàcil: sent i pronuncia els sons de la llengua estrangera sobre la base del sistema de sons de la

seva llengua materna. Sovint, doncs, els errors fonètics que comet depenen de la confrontació dels sistemes fonològics de les dues llengües.

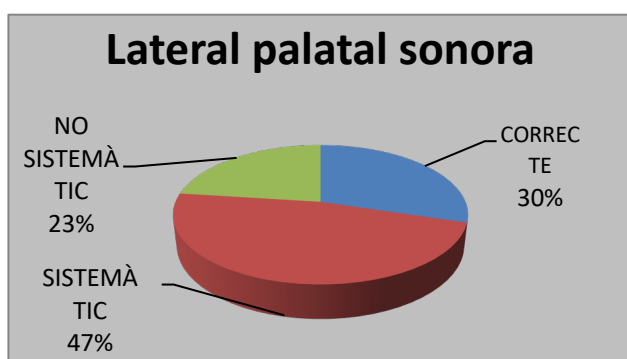
Actualment, sobretot a zones urbanes, que és el lloc de procedència de la majoria de l'alumnat que estudia aquest grau, s'observa que els trets definitoris de la pronúncia del català, com és la pronúncia del so lateral palatal sonor, se substitueixen per formes d'articular d'altres llengües, sobretot de l'espanyol, de manera que la fonètica, que és la cara de la llengua, cada vegada té un rostre més desfigurat.

Tot i que l'alumnat que accedeix a la facultat d'educació de la UB presenta dificultats en la realització de diversos sons (la e i la o oberta, la vocal neutra, la essa sonora i africada, la prepalatal sonora

i africada i la lateral palatal sonora), en aquest estudi, ens centrarem en el darrer so perquè està experimentant una recessió important en castellà i també en català. Com podem veure al gràfic de la figura 1, pràcticament la meitat dels nostres estudiants no el saben articular en cap context, i un 23% el pronuncia correctament només en alguns contextos.

Figura 1. Resultat de la producció del so lateral palatal sonor.

Actualment, els joves que accedeixen a la universitat pertanyen a la generació de nadius digitals i el mòbil forma part del seu dia a dia. Utilitzar-lo a l'aula és una motivació més. Si el mòbil és a la butxaca de l'alumnat i el professorat, per què no el podem fer servir com a recurs educatiu?



Els objectius, doncs, que ens plantejem en aquest estudi són, d'una banda, aportar els resultats d'una experiència d'ensenyament-aprenentatge en l'adquisició de la pronúncia del so lateral palatal sonor, amb la finalitat que adquireixin una pronúncia adequada en català perquè tinguin una bona competència lingüística quan exerceixin la seva professió.

I, d'altra banda, avaluar si l'aportació de les noves tecnologies, com la plataforma Moodle, la web de los sons del català, la web de les Guies de pronunciació i el telèfon mòbil, afavoreix aconseguir un aprenentatge més eficaç.

2. MARC TEÒRIC

La pronunciació, com a destresa lingüística derivada de la fonologia, sol ser un aspecte molt secundari en l'ensenyament d'una llengua. En el marc de les metodologies comunicatives, no es dona importància a la pràctica de la pronunciació, tanmateix, quan una persona parla amb gran precisió gramatical, però comet errors de pronúncia, pot perdre, a vegades, el seu objectiu comunicatiu o tenir malentesos amb el seu interlocutor.

Sovint, pels mateixos fonetistes, la pronunciació és una matèria secundària. Kelly (1969) afirmava que la pronunciació era la ventafocs de l'ensenyament de la LE, Renard (1971) parlava del parent pobre de la didàctica de la llengua, Göbel i Graffmann (1977) s'hi referien com la fillastra i Stanley (2013) comentava que els professors no donen importància a aquesta àrea.

Segurament, aquest menyspreu s'ha anat perpetuant al llarg del temps. A més, podem afegir-hi que la majoria de professors en tenen un cert desconeixement perquè, d'una banda, tenen poca formació fonètica i, de l'altra, perquè hi ha la creença que l'adquisició de la pronunciació és un procés automàtic del qual no cal preocupar-se.

Actualment, la part teòrica no ha canviat gaire, l'ensenyament de la pronúncia, si ens fixem en la majoria de manuals, es treballa molt poc, segurament per la creença que és un procés automàtic i per la poca formació del professorat. Ens els darrers anys, han aparegut pàgines web per proporcionar recursos al professorat i per afavorir que l'alumnat pugui fer un aprenentatge autònom d'aquesta disciplina. En aquesta mateixa línia, tot i la modernització amb les pàgines webs, creiem que cal fer un pas endavant, adaptar-se als nous temps i crear noves tasques amb l'ajuda de les aplicacions que ens ofereix el telèfon mòbil.

El que sí que ha canviat és la tendència a la desfonologització del so lateral palatal sonor a favor d'un element aproximant palatal en algunes zones del domini lingüístic català (Rost, 2016). Recasens (1991:322) creu que la confusió perceptiva, atesa la similitud acústica entre ambdues articulacions, la lateral i l'aproximant, afavoreix aquest canvi.

Les investigacions i els estudis sobre l'aprenentatge d'una segona llengua a través del mòbil (*mlearning*) constaten l'impacte positiu en la motivació dels alumnes que aprenen una L2 (Burston, 2013; Godwin-Jones, 2017), en el canvi en el rol de joc i en el concepte tradicional d'aula (Drexler, 2010; Moreno i Traxler, 2016). Trucar, doncs, és la característica menys important del mòbil perquè en destaca la connectivitat, la multimèdia, la ubiqüitat i la capacitat de comunicació.

3. METODOLOGIA

Per portar a terme els objectius que ens proposem, ens hem basat en un pla d'acció inspirat en els principis del mètode verbotal, en què es treballa a partir de l'audició, la percepció, la imitació i la producció, amb la finalitat d'afavorir la comunicació

i d'aproximar-ne les produccions orals a l'accent nadiu (veg. Dalmau et. al., 1985; Renard, 1971; Billières et al., 2013). El mètode verbotonal vol que l'estudiant integri el fonema en el sistema fonològic de manera inconscient perquè, quan un so es percep correctament, és quan es pot (re)produir. Per produir, doncs, primer s'ha de percebre.

Quant a la metodologia, es parteix de la gravació que cada estudiant fa de 20 mots, que corresponen a imatges que contenen el so lateral palatal sonor en diferents posicions: inicial (CVC/CVV: llit/lleó), intervocàlica (VCV: orella), final (CVC/VC: cavall, ull) o en la combinació VCCV (bitllet). De les 1360 paraules obtingudes, es va analitzar acústicament el so de la consonant amb Praat (Boersma & Weenink, 1992-2019) i es va comprovar que només el 30% d'estudiants produeixen aquest so, mentre que la resta hi tenen dificultats.

Inspirats en el manual de Bau, Pujol i Rius-Escudé, 2007, i en el mètode verbotonal, vam preparar unes activitats per treballar el so a partir de 4 nivells de tasques, cadascun dels quals s'ha de superar per poder continuar: discriminar, percebre, imitar i produir.

A més a més, per a portar-les a la pràctica i per afavorir l'autonomia de l'estudiant, comptem amb la plataforma Moodle, la pàgina web "Els sons del català" (<http://www.ub.edu/sonscatala/#.html>), la pàgina web "Guies de pronunciació del català" (<http://www.ub.edu/guiesdepronunciacio>) i aplicacions de mòbil, com Kahoot i els codis QR.

L'experiència es va portar a terme durant el primer semestre del curs 2019-2020 a 81 estudiants de dos grups de primer curs del grau de Mestre d'Educació Primària (UB). Els informants són noies, un 65%, i nois, un 35%, d'entre 18 i 20 anys dels quals un 37% és catalanoparlant, un 39% castellanoparlant, tot i que el 95% és nascut a Catalunya, un 20% parla ambdues llengües i un 4% parla altres llengües. (Vegeu figura 2).

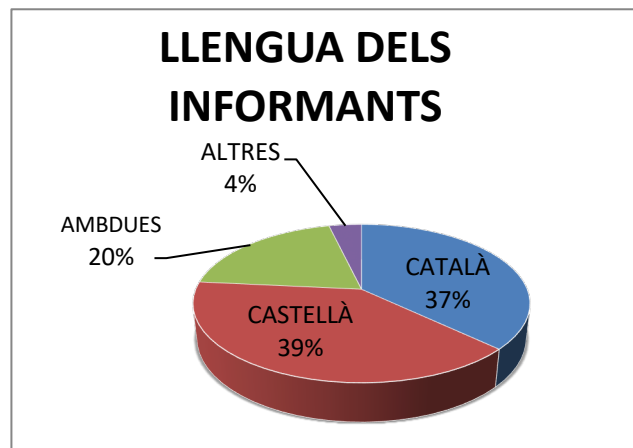


Figura 2. La llengua dels informants.

4. LES ACTIVITATS

Les activitats que proposem per treballar el so lateral palatal sonor són, en primer lloc, començar amb activitats d'escalfament per preparar el cos abans de començar, igual com fan els músics. Per realitzar-les, ens hem de posar dempeus i fer exercicis d'inspiració/expiració, d'obertura i tancament de la mandíbula, repetició de mots i frases curtes tenint en compte la velocitat, l'entonació, el ritme i la intensitat. Cal mencionar que, sempre que sigui possible, quan es practica, l'alumne no ha de tenir cap referent ortogràfic, i és millor partir d'una imatge o d'un referent auditiu (el professor o un arxiu d'àudio). És important destacar, també, que, perquè l'alumnat no s'acostumi a una única veu i a un únic to, cal que les mostres auditives que escolti siguin variades.

En segon lloc, presentar el so i, a continuació, fer exercicis de discriminació auditiva, mitjançant l'aplicació Kahoot perquè, un so que no se sent, no es pot produir. Les activitats consisteixen a escoltar dos mots i saber quin conté el so que s'està treballant, el primer o el segon; després escoltar una frase interrogativa curta i dir si conté el so o no.

En tercer lloc, articular el so a partir d'activitats i tasques diverses: les que poden trobar al campus virtual, les de les webs <http://www.ub.edu/sonscatala/#.html> o <http://www.ub.edu/guiesdepronunciacio> i de les aplicacions a través del mòbil (codi QR i Kahoot). A mesura que les activitats de producció són més complexes, s'introdueixen també aspectes locutius com ara la velocitat, la intensitat, el ritme, etc. Per exemple, per treballar la locució i el so lateral palatal sonor, el professor tria dos embarbussaments que tinguin aquest so, grava 5 versions de cadascun a diferent velocitat, to i ritme, i els penja a l'aula amb un codi QR.



L'alumnat, per parelles, ha d'escollir-los i ha d'escollir quins sons els dos embarbussaments que tenen la pronúncia i la locució adequada. Han de compartir la tria amb el grup classe. Finalment, a partir del model adequat, es demana a l'alumnat que en gravi la seva versió i, amb un codi QR la faci arribar al professor. Per aconseguir un progrés òptim, però, cal que totes les tasques siguin tutoritzades per un professor.

5. RESULTATS

Després de realitzar les activitats, l'alumnat ha tornat a gravar els mateixos mots de les imatges que contenen el so de la lateral palatal sonora i ha produït un discurs. S'ha analitzat acústicament el so de la consonant a Praat i hem constatat que gairebé tots els estudiants que no sabien produir el so han millorat de manera considerable la seva pronúncia. Tots han aconseguit articular-lo, però no tothom l'ha integrat en la seva parla diària. És a dir, un 40% el pronuncia correctament en tots els mots i en totes les posicions (inicial: CVC/CVV; intervocàlica: VCV, final: CVC/VC; i la combinació VCCV), mentre que un 60% el pronuncia correctament, tot i que, puntualment, en alguns mots o en alguna posició el canvia per un altre so. No obstant això, ara són conscients de les seves mancances i saben que han de continuar treballant per sistematitzar-lo.

Les activitats de discriminació auditiva realitzades amb l'aplicació del telèfon mòbil Kahoot són les que els motiven més, per la qual cosa, en el bloc de producció, se'ls anima a crear-ne més i compartir-los per poder continuar discriminant, produint i, en conseqüència, millorant.

6. CONCLUSIONS

Des del punt de vista de l'aprenentatge, l'alumnat millora la pronúncia del so, és conscient dels seus errors i s'anima els alumnes més avantatjats a crear-ne més per poder-los compartir. Així mateix, l'ús de les noves tecnologies ha permès que l'alumnat estigui motivat, treballi i continuï practicant de

manera autònoma per solucionar les seves mancances.

I, finalment, des del punt de vista psicològic, els estudiants augmenten l'autoestima, perden la vergonya i la por de parlar en català i se senten més segurs i satisfets quan el parlen.

7. BIBLIOGRAFIA

- Bau, M.; Pujol, M.; Rius, A. (2007). *Curs de pronunciació. Exercicis de correcció fonètica*. Barcelona: PAMSA.
- Boersma, P.; Weenink, D. (1992-2019). PRAAT. *Doing phonetics by computer*. Institute of Phonetic Sciences, University of Amsterdam. Disponible en <http://www.praat.org>
- Billières, M., Alazard, C., Astésano, C.; Nocaudie, O. (2013). Phonétique corrective en FLE. Méthode verbo-tonale. Retrieved from <http://w3.uohprod.univ-tlse2.fr/UOH-PHONETIQUE-FLE/>
- Burston, J. (2013). Mobile Assisted Language Learning: A selected Annotated Bibliography of implementatios Studies 1994-2012. *Language Learning & Technology*, vol. 17, n.3.
- Carrera Sabaté, J. i Prieto, P. (coords.) (2012-2019). *Guies de pronunciació del català*. <http://www.ub.edu/guiesdepronunciacio> [Consultada el 25 de septiembre de 2019]
- Dalmau, M.; Miró, M; Molina, D. A. (1985). *Correcció fonètica. Mètode Verbo-Tonal*. Vic: EUMO.
- Drexler, W. (2010). The networked student model for construction of personal learning environments: Balancing teacher control and student autonomy. *Australasian Journal of Educational Technology*, 26(3), 369–385.
- Els sons del català (2009). <http://www.ub.edu/sonscatala/#.html> Universitat de Barcelona. [Consultada el 25 de septiembre de 2019].
- Göbel, H. i Graffmann, H. (1977). "Ein Stiefkind des Unterrichts DaF: Ausprachschulung", en *Zielsprache Deutsch*, 3.
- Godwin-Jones, R. (2017) Smartphones and language learning. *Language Learning & Technology*, 21 (2), 3-17.
- Kelly, L. (1969). *25 centuries of language teaching*. Rowley, MA: Newbury House.
- Moreno. A. I. i Traxler, J. (2016). MALL-based MOOCs for language teachers: Challenges and opportunities. *Porta Linguarum Monograph*, 1, 73-85.

- Renard, R. (1971). *Introduction à la méthode verbo-tonale de correction phonétique*. París: Didier.
- Rost Bagudanch, A. (2016). La percepción de /ʎ/ i /j/ en catalán y en español. Implicaciones en la explicación del yeísmo. *Estudios de Fonética Experimental*, XXV, 2016, p.39-80.
- Recasens, D. (1991). *Fonética descriptiva del català (Assaig de caracterització de la pronúncia del vocalisme i consonantisme del català al segle XX)*, Barcelona: IEC.
- Stanley, G. (2013). *Language Learning with Technology: ideas for integrating technology in the classroom*. Cambridge: Cambridge University Press.

RHYTHMIC TRANSFER IN GALICIAN LEARNERS OF ENGLISH

Rosalía Rodríguez Vázquez¹, Paolo Roseano²

¹Universidade de Vigo; ²Universitat de Barcelona and ³University of South Africa
rosalia@uvigo.es; paolo.roseano@ub.edu

ABSTRACT

Languages around the world have been classified according to their rhythm as stressed timed, syllable timed, and mora timed. Conforming to this classification, a limited number of studies have sought to find out whether prosodic transfer may occur between languages belonging in different rhythmic classes. The aim of this study is to verify the existence of rhythmic transfer from L1 to L2 in a foreign language learning context, where a group of native Galician (L1) speakers utter a text in English (L2). Bearing in mind the above-mentioned classification of languages, the rhythmic productions of six speakers of Central Galician, Central Peninsular Spanish, and Southern British English, and four speakers of English L2 are analysed and labelled using *Praat*. Subsequently, five rhythm metrics (V%, ΔC , ΔV , CrPVI and VnPVI) are calculated for each language by means of *Correlatore*. The resulting data show that there is rhythmic transfer between Galician L1 and English L2.

Keywords: English L2, rhythm, acquisition, rhythmic transfer

1. INTRODUCTION

1.1. Rhythmic typologies

Linguistic rhythm has been traditionally defined as isochrony of speech intervals (Abercrombie, 1967: 97-98). The classification of languages regarding their rhythmic makeup relies on the speech intervals (stress feet, syllables, morae) that are considered to be isochronous. From the initial twofold classification into stress- and syllable-timed languages established by Pike (1945) and Abercrombie (1967) to the stress-based continuum subsequently proposed by Dauer (1983) and Bertinetto (1989), a considerable amount of phonological research has delved into the rhythmic characterisation of languages as (predominantly) stress-timed, where stressed syllables are perceived to fall at roughly isochronous intervals, syllable-timed languages, where all syllables are perceived to be isochronous, and mora-timed languages, where morae are taken to be isochronous. The most quoted examples of each one of these types of language are English, Spanish and Japanese, respectively. In recent years, phonologists have come to accept the view that linguistic rhythm is a percept; in other words, isochrony is perceptually imposed on a largely non-isochronous speech signal (Ordin & Polyanskaya, 2015: 535). Consequently, efforts have been made to find objectively quantifiable measures

to prove the phonetic reality of the aforementioned classification. With this purpose in mind, data-driven studies by Ramus, Nespore & Mehler (1999), Grabe & Low (2002) and Dellwo (2006), among others, have made use of a number of metrics to confirm the existence of objective, acoustically measurable differences between stress-timed and syllable-timed languages. Such differences are arguably linked to contrasts regarding syllable structure, and the presence (or not) of vowel reduction, accent-related lengthening, and final lengthening. In spite of the problems observed when interpreting some metric scores to try and classify languages rhythmically (see Arvanity, 2009; Arvanity & Ross, 2010; Arvanity, 2012), no other reliable classification method has been put to the test.

1.2. Rhythmic transfer

As much as the research on the rhythmic differences between languages has evolved, the production of L2 in contrast to L1 rhythmic patterns has been paid little attention in the literature on L2 acquisition and linguistic transfer (see Ordin & Polyanskaya, 2015). Linguistic transfer may be defined as the phenomenon that occurs when the learners of an L2 use one or several features that characterise their L1 when they speak the L2. As Roseano et al. (2015) indicate, linguistic transfer affects not only the lexical and syntactic levels of a language, but also its

prosodic level, i.e. intonation and rhythm. Only a scarce number of studies have analysed phenomena related to prosodic transfer in contexts of language contact connected to migrations—examples of which are the studies on the production of American English by Latin American Spanish speakers in North Carolina (Carter, 2005); *Porteño* and Castilian Spanish by Italian speakers in Buenos Aires (Gabriel & Kireva, 2014); Argentinian Spanish by Afrikaans speakers in Patagonia (Coetzee et al., 2015)—and in contexts of foreign language learning—the production of English by Chinese (Ding & Xu, 2016), Chinese and German (Li & Post, 2014), Japanese (Grenon & White, 2008) and French (Tortel & Hirst, 2008) learners. A relevant percentage of those studies note that the speakers’ rhythmic productions in L2 show certain rhythmic features that characterise the prosody of their L1. The existing literature on prosodic acquisition proves that “the acquisition of second language speech rhythm is demonstrably a challenge for language learners” (Kinoshita & Sheppard, 2011: 1086). Moreover, the transfer of certain prosodic features from L1 to L2 is implicitly or explicitly supported by the view that “the phonic elements making up the [...] phonetic subsystems” of the languages in question “exist in a common phonological space, and so will necessarily influence one another” (Flege, Schirru, & MacKay, 2003: 469). As a natural consequence of such view, “it is not surprising that [...] L1 transfer is an important factor in L2 prosody learning.” (Rasier & Hiligsmann, 2007: 44). Although some scholars have pointed to the possibility that certain rhythm acquisition processes might be universal, i.e. they always appear in the process of L2 language prosody learning regardless of the characteristics of the learner’s L1 (see Rasier & Hiligsmann, 2007; Li & Post 2014; Ordin & Polyanslaya, 2015), this does by no means entail that the rhythm of L2 will not get affected by the rhythmic properties of L1, particularly when the two have very different rhythmic structures (see Tortel & Hirst, 2008). In Wenk’s (1985: 157) view, when advanced learners of a given L2 produce sentences or texts in that L2, “[...] they] pass through a transitional stage characterised by the production of rhythmic groups which, while displaying features of both the L1 and L2 systems, are unique to the learner’s ‘interlanguage’”. Whether or not one endorses the existence of the interlanguage, the analysis of the potential influence of the rhythm of L1 on L2 must be carefully explored so as to come to a sound conclusion regarding prosodic transfer.

1.3. Objectives

The present study analyses the production of speech rhythm in a foreign language (ENG_L2) by bilingual learners who speak Spanish, which is rhythmically contrastive to the target language, and Galician, a Romance language that has not yet been classified with regards to its rhythm. The aim of the study is to verify whether the production of ENG_L2 is affected by rhythmic transfer from Galician to English.

2. METHODS

In order to conduct this investigation, speakers of four different varieties were recorded:

- 6 speakers of Central Galician (henceforth GAL), the native language of the group of students recorded for this experiment;
- 6 speakers of Central Peninsular Spanish (henceforth CPS), a prototypical syllable-timed language included in our dataset as a rhythm reference point;
- 6 speakers of Southern British English (henceforth SBE), a prototypical stress-timed language and the students’ target language;
- 4 university students, native speakers of Central Galician and Peninsular Spanish currently studying English (ENG_L2) at the University of Vigo who, at the moment of recording, certified a B2/C1 level of English and had never lived abroad. Regarding the production task, speakers read the tale *The North Wind and the Sun* in their respective language. This is a text commonly employed in phonetics studies, which has been used in research works about rhythm (see Grabe & Low, 2002). The resulting corpus contained 22 recordings. The acoustic analysis was carried out with *Praat* (Boersma & Weenink, 2019). For each recording, the vocalic and consonantal intervals were annotated in a textgrid. The statistical analysis was carried out by means of *Correlatore 2.3.4* (Mairano & Romano, 2010), a software used to calculate the global metrics V%, ΔC and ΔV (Ramus, Nespors, & Mehler, 1999), and the local metrics CrPVI and VnPVI (Grabe & Low, 2002), all of which have been put to the test in previous studies about linguistic rhythm. As is well known, ΔC and CrPVI relate to the variability of consonantal intervals, while ΔV and VnPVI have to do with the variability of vocalic intervals. Syllable-timed languages like CPS have been reported to show low values for all of these metrics (i.e. they have simple consonant clusters and do not show vowel reduction). Stress-timed languages like SBE,

on the other hand, have been observed to display high values for the same metrics (i.e. they have complex consonant clusters and show vowel reduction). Similarly, the metric V% reflects both vocalic and consonantal characteristics of languages. Syllable-timed languages such as CPS have been proven to deploy a high value of V%, while stress-timed languages such as SBE have been noted to show a low value of V%.

3. RESULTS

3.1. V%, ΔC and ΔV (Ramus et al., 1999)

Following Ramus et al. (1999), two combinations of metrics, namely ΔC and ΔV (Figure 1), and ΔC and V% (Figure 2) were used. In both cases, ENG_L2 emerges in an intermediate position between GAL (which turns out to be, like most Romance languages, syllable-timed) and SBE on both axes, which means that the variety of English used by the students shows features that appear between the source (GAL) and target (SBE) languages regarding consonantal and vocalic features.

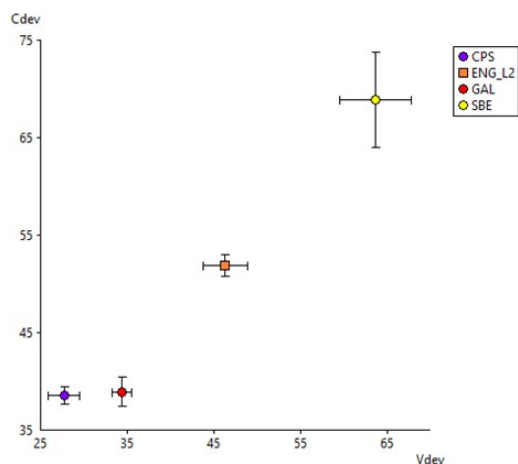


Figure 1: Results for $\Delta C/\Delta V$.

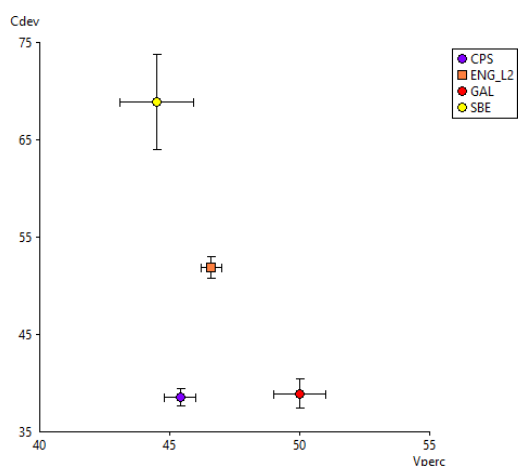


Figure 2: Results for $\Delta C/V\%$.

3.2. CrPVI and VnPVI (Grabe & Low, 2002)

Following Grabe & Low (2002), we analysed CrPVI and VnPVI (Figure 3). Again, ENG_L2 stands in an intermediate position between GAL and SBE on both axes, which indicates that it shows intermediate characteristics between the source and the target languages as far as consonantal and vocalic features are concerned.

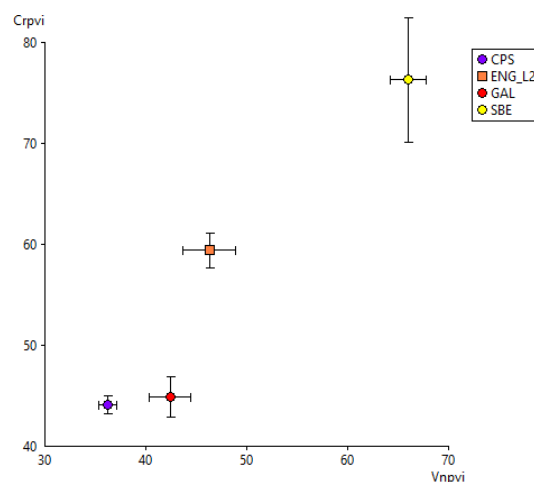


Figure 3: Results for CrPVI/VnPVI.

4. DISCUSSION

The graphic representations provided in the analysis support the view that the acoustic parameters best suited to grasp the rhythmic differences between SBE, CPS GAL, and ENG_L2 are, on the one hand, the parameters that measure the consonant material—Ramus et al.'s (1999) ΔC and Grabe and Low's (2002) CrPVI—and, on the other hand, the parameters that measure the vowel material—Grabe and Low's (2002) VnPVI and Ramus et al.'s (1999) ΔV . The graphics clearly reveal that SBE and CPS stand at opposite poles of the representation axes, which enables us to locate two benchmarks that frame the rhythmic area occupied by English spoken by Galician learners. The rhythmic area where ENG_L2 stands is situated in the space between GAL and SBE, and slightly closer to GAL. This points to the existence of a certain degree of transfer from Galician to English, as ENGL_L2 is characterised by the production of vowel, consonants and rhythmic groups which are different from those of SBE and, therefore, unique to the Galician learners' language.

5. CONCLUSIONS

The purpose of this work was to verify whether the production of English by Galician speakers would be affected by rhythmic transfer from Galician to English. The results prove that there is rhythmic transfer from L1 (in this case, Galician) to L2 (in this case, English spoken by Galician speakers) in both the consonant and the vowel intervals. In light of this, language transfer may be considered as an inescapable learning stage to make it possible for foreign language learners to eventually acquire L2 speech rhythm.

6. REFERENCES

- Abercrombie, D. (1967). *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.
- Arvanity, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics*, 40, 351-373.
- Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica*, 66, 46-63.
- Arvanity, A., & Ross, T. (2010). Rhythm classes and speech perception, *Proceedings of Speech Prosody 2010*.
- Bertinetto P. M. (1989). Reflections of the dichotomy "stress" vs. "syllable-timing". *Revue de Phonétique Appliquée*, 91-93.
- Boersma, P., & Weenink, D. (2019). *Praat: Doing phonetics by computer [Computer program]. Version 6.1.07*, retrieved from <http://www.praat.org/>
- Carter, P. M. (2005). Prosodic Variation in SLA: Rhythm in an Urban North Carolina Hispanic Community. *University of Pennsylvania Working Papers in Linguistics*, 11(2), 59-71.
- Coetzee, A. W., García-Amaya, L., Henrikisen N., & Wissing, D. (2015). Bilingual Speech Rhythm: Spanish-Afrikaans in Patagonia, *Proceedings of the 18th International Congress of Phonetic Sciences 2015*.
- Dauer, R. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, 11, 51-62.
- Dellwo, V. (2006). Rhythm and Speech Rate: A Variation Coefficient for deltaC. In P. Karnowski, and I. Sziget (eds.) *Language and language-processing*. Frankfurt: Peter Lang, 231-241.
- Ding, H., & Xu, X. (2016). L2 English Rhythm in read Speech by Chinese Students. *Proceedings of Interspeech 2016*, 2696-2700.
- Flege, J. E., Schirru, C., & MacKay, I. R. A. (2003). Interaction between the native and second phonetic subsystems. *Speech Communication*, 40, 467-491.
- Gabriel, C. & Kireva, E. (2014). Prosodic transfer in learner and contact varieties. *Studies in Second Language Acquisition*, 36, 257-281.
- Grabe, E., & Low E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven, & N. Warner (Eds.), *Papers in Laboratory Phonology 7*, (pp. 515-546). Berlin: de Gruyter.
- Grenon, I., & White, L. (2008) Acquiring Rhythm: A Comparison of L1 and L2 Speakers of Canadian English and Japanese. *Proceedings of the 32nd Boston University Conference on Language Development*, 155-166.
- Kinoshita, K., & Sheppard, C. (2011). Validating acoustic measures of speech rhythm for second language acquisition. In W. S. Lee, & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences*, (pp. 1086-1089). Hong Kong: City University of Hong Kong.
- Li, A., & Post, B. (2014). L1 Acquisition of Prosodic Properties of Speech Rhythm. *Studies in Second Language Acquisition*, 36, 223-255.
- Mairano, P., & Romano, A. (2010). Un confronto tra diverse metriche ritmiche usando Correlatore. In S. Schmid, M. Schwarzenbach, & D. Studer (Eds.), *La dimensione temporale del parlato*, Proc. of the V National AISV Congress, University of Zurich, Collegiengebäude, 4-6 February 2009, (pp. 79-100), Torriana (RN): EDK.
- Ordin, M. & Polyanskaya, L. (2015). Acquisition of speech rhythm in a second language by learners with rhythmically different native languages. *The Journal of the Acoustic Society of America*, 138(2), 533-545.
- Pike, K. (1945). *The intonation of American English*. Ann Arbor: University of Michigan Press.
- Ramus, F., Nespors, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73, 265-292.
- Rasier, L. & Hiligsmann, Ph. (2007). Prosodic transfer. Theoretical and Methodological Issues. *Nouveaux cahiers de linguistique Française*, 28, 41-66.
- Roseano, P., Fernández Planas, A. M., Elvira-García, W., & Martínez Celdrán, E. (2015). Contacto lingüístico y transferencia prosódica bajo una perspectiva. *Dialectologia et Geolinguistica: Journal of the International Society for Dialectology and Geolinguistics*, 23(1), 95-123.
- Tortel, A. & Hirst, D. (2008). Rhythm and rhythmic variation in British English: subjective and objective evaluation of French and native speakers. *Proceedings of Speech Prosody 2008*, 359-362.
- Wenk, B. J. (1985). Speech Rhythmic in Second Language Acquisition. *Language and Speech*, 28(2): 157-175.

TraFo: AN ONLINE G2P TOOL FOR SPANISH

Francesco Rodriguez^{1,2}

¹Universitat de Barcelona, ²Albert-Ludwigs-Universität Freiburg
francesco.rodriguez@ub.edu

ABSTRACT

This paper presents *TraFo*, an online Grapheme-to-Phoneme (G2P) transcriber of dialectal varieties of Spanish. While at present there is a wide range of tools for automatic phonetic transcription of a variety of languages (including Spanish), only very few tackle the issue of automatic transcription on the phonemic level, especially when it comes to considering dialectal phonological variation. *TraFo* is an online tool that generates automatic phonological transcriptions of Spanish and which has been designed to take into account dialectal features known as *lleísmo*, *yeísmo*, *zheísmo*, *sheísmo*, and *seseo*, *ceceo*, *distinguidor*. This set of parameters is sufficient to allow for the automatic phonological description of a wide range of Spanish varieties and can hence be used for teaching, (self-)learning or research.

Keywords: phonological transcription, dialectal variation, speech technology, Grapheme-to-Phoneme (G2P)

1. INTRODUCTION

This paper presents *TraFo*, a Grapheme-to-Phoneme (G2P) transcriber of Spanish that has five main characteristics: 1) it is available for public use, 2) it operates online (<http://stel.ub.edu/trafo>), 3) it uses IPA conventions, 4) it performs a phonological transcription, and 5) it takes into consideration dialectal variation. The reasons why a transcriber with these characteristics was needed are explained in Section 2, where we present a short overview of the technical features of the existing tools that carry out automatic transcriptions of Spanish. In Section 3 we describe the methodological aspects of *TraFo*. Section 4 summarizes the validation of the tool and, finally, Section 5 draws some conclusions and envisages a few future developments.

2. STATE OF THE ART AND MOTIVATION

Currently, a few tools carry out automatic transcriptions of Spanish. The features they present are not homogeneous.

An important characteristics of existing transcribers for Spanish, as pointed out by Garrido, Codina, & Fodge (2019), is that the majority of them have been developed specifically for a system or for a concrete application, usually commercial, and are not available for public use (e.g. Bonaventura, Giuliani, Garrido, & Ortín, 1998).

Among the transcribers that are freely available, the majority carry out a phonetic transcription, like *Fonemolabs* (Molino de Ideas, 2012), *Transcriptor fonético automático* (López, 2004), *SAGA* (TALP-UPC, 2017), and *TransText* (Garrido et al., 2019). In addition, most of them (with the exception of *TransText* and *SAGA*) only provide a transcription in Central Peninsular Spanish, while other dialects are not taken into account.

The only transcriber that carries out a phonological transcription is, to the best of our knowledge, the *Silabeador y transcriptor fonético y fonológico* (Armario, 2003-2008). This notwithstanding, this transcriber is based on some theoretical and methodological assumptions that can discourage its use. From a theoretical point of view, the *Silabeador y transcriptor fonético y fonológico* is clearly structuralist, insofar as the transcriptions it produces include the archiphonemes /L/, /N/, /D/ and /R/ (see example in Section 4.2). From a methodological point of view, a shortcoming of the *Silabeador y transcriptor fonético y fonológico* is that it does not take into consideration dialectal variation: in its present state, it only offers transcriptions in one sociolect, that is the variety spoken in Castille by older generations, who do not show either the /θ-/s/ merger nor the /ʎ-/j/ merger (Chapman, Dubra, Martínez-Gil, & Tritica, 1983). On the other hand, this transcriber has the advantage of using not only the traditional RFE phonetic alphabet, but also the more widely used IPA.

This overview of the existing transcribers of Spanish reveals that there is no tool that has the features we advanced in Section 1. Specifically, no online free transcriber performs an IPA phonological transcription that takes into account dialectal differences and is not bound to a specific theoretical model. Such tool is desirable not only because it contributes to the development of speech sound studies, but also because it has practical applications in (i) clinical linguistics (e.g. creation of sets of phonemically balanced texts for logaudiometries and training in speech therapy), (ii) teaching of Spanish phonology and phonetics to university students, (iii) teaching/learning of the pronunciation of Spanish as a foreign language.

3. METHODOLOGY

The web tool is based a *Python* script (version 3.6) whose input (Section 3.1) consists of a text and of some dialectal variation options. By means of some G2P formulas (Section 3.2), the script produces the final output (Section 3.3), i.e. the phonological transcription of the text.

3.1. Input

The text to be transcribed has to be written (or pasted) in a dialogue box. It may contain words and punctuation marks, numbers and symbols; nevertheless, punctuation marks, numbers and symbols will not be transcribed and will appear as such in the output.

The dialectal features that the user will be required to choose have to do with two mergers that appear in several Spanish dialects, where they show different outcomes.

If the /ʎ/-/j/ merger happens, its outcome can be /j/ (in this case the literature about Spanish talks about *yeísmo*), /ɜ/ (called *zheísmo*), or /j/ (called *sheísmo*). If the merger does not take place, the literature refers to this situation as *lleísmo*.

If the /θ/-/s/ merger happens, its outcome is usually /s/ (in this case the literature about Spanish talks about *seseo*), but in some varieties of Southern Spain it can be /θ/ (called *ceceo*). If the merger does not take place, the variety/speaker is traditionally called *distinguidor*.

The user will need to tick one of the four possible situations related to the first merger (i.e. *lleísmo*, *yeísmo*, *zheísmo*, or *sheísmo*), and one of the three possible situation related to the second merger (labelled as *distinguidor*, *ceceo*, and *seseo*).

3.2. G2P formulas

The central part of the transcriber consists of the phonetic transcription rules (summarized in Table 1). These rules have been encoded in *Python* by means of *if-then* conditions, so *TraFo* is a rule-based transcriber.

Table 1: Summary of the transcription rules.

Transcription rules
<a>, <á> ► /a/
 ► /b/
<c> ► /k/ if the following grapheme = <o>, <ó>, <a>, <á>, <u>, <ú>
<c> ► /θ/ if the following grapheme = <e>, <é>, <i>, <í> and if the user has chosen “ <i>distinguidor</i> ” or “ <i>ceceo</i> ”.
<c> ► /s/ if the following grapheme = <e>, <é>, <i>, <í> and if the user has chosen “ <i>seseo</i> ”.
<c> ► /tʃ/ if the following grapheme = <h>
<d> ► /d/
<e>, <é> ► /e/
<f> ► /f/
<g> ► /x/ before <e>, <é>, <i> e <í>.
<g> ► /g/ if the following grapheme ≠ <e>, <é>, <i>, <í>
<h> ► ∅
<i>, <í> ► /i/
<j> ► /x/
<k> ► /k/
<l> ► /l/ if the following grapheme ≠ <l> and if the preceding grapheme ≠ <l>
<l> ► /j/ if the following grapheme = <l> and if the preceding grapheme ≠ <l> and if the user has chosen “ <i>yeísmo</i> ”.
<l> ► /j/ if the following grapheme = <l> and if the preceding grapheme ≠ <l> and if the user has chosen “ <i>sheísmo</i> ”.
<l> ► /ɜ/ if the following grapheme = <l> and if the preceding grapheme ≠ <l> and if the user has chosen “ <i>zheísmo</i> ”.
<l> ► /ʎ/ if the following grapheme = <l> and if the preceding grapheme ≠ <l> and if the user has chosen “ <i>lleísmo</i> ”.
<l> ► ∅ if the preceding grapheme = <l> and if the following grapheme ≠ <l>
<m> ► /m/
<n> ► /n/
<ñ> ► /ɲ/
<o>, <ó> ► /o/
<p> ► /p/

<q> ► /k/
<r> ► /r/ if the preceding grapheme ∉ V={<a>,<á>,<e>,<é>,<i>,<í>,<o>,<ó>,<u>,<ú>,<ü>} and if the preceding grapheme ∉ C={,<c>,<d>,<f>,<g>,<h>,<j>,<k>,<l>,<m>,<n>,<p>,<q>,<r>,<s>,<t>,<v>,<w>,<x>,<y>,<z>}
<r> ► /r/ if the following grapheme = <r>
<r> ► /r/ if the preceding grapheme ∈ V={<a>,<á>,<e>,<é>,<i>,<í>,<o>,<ó>,<u>,<ú>,<ü>} and the following grapheme ∈ V={<a>,<á>,<e>,<é>,<i>,<í>,<o>,<ó>,<u>,<ú>,<ü>}
<r> ► ∅ if the preceding grapheme = <r> and if the following grapheme ≠ <r>
<s> ► /s/ if the user has chosen “seseo” or “distinguidor”
<s> ► /θ/ if the user has chosen “ceceo”.
<t> ► /t/
<u>,<ú> ► /u/ if the preceding grapheme ≠ <g> or <q>
<u>,<ú> ► /u/ if the preceding grapheme = <g> or <q> and the following grapheme ≠ <e>,<é>,<i>,<í>
<u>,<ú> ► ∅ if the preceding grapheme = <g> or <q> and if the following grapheme = <e>,<é>,<i>,<í>
<ü> ► /u/
<v> ► /b/
<w> ► /gu/
<x> ► /ks/
<y> ► /i/ if the preceding grapheme ∈ V={<a>,<á>,<e>,<é>,<i>,<í>,<o>,<ó>,<u>,<ú>,<ü>}
<y> ► /i/ if the preceding grapheme ∉ V={<a>,<á>,<e>,<é>,<i>,<í>,<o>,<ó>,<u>,<ú>,<ü>} y ∉ C={,<c>,<d>,<f>,<g>,<h>,<j>,<k>,<l>,<m>,<n>,<p>,<q>,<r>,<s>,<t>,<v>,<w>,<x>,<y>,<z>}
<y> ► /j/ if the following grapheme ∈ V={<a>,<á>,<e>,<é>,<i>,<í>,<o>,<ó>,<u>,<ú>,<ü>} and if the user has chosen “yeísmo” or “lleísmo”.
<y> ► /ʝ/ if the following grapheme ∈ V={<a>,<á>,<e>,<é>,<i>,<í>,<o>,<ó>,<u>,<ú>,<ü>} and if the user has chosen “sheísmo”.
<y> ► /z/ if the following grapheme ∈ V={<a>,<á>,<e>,<é>,<i>,<í>,<o>,<ó>,<u>,<ú>,<ü>} and if the user has chosen “zheísmo”.

<z> ► /θ/ if the user has chosen “ceceo” or “distinguidor”.

<z> ► /s/ if the user has chosen “seseo”.

3.3. Output

Once the transcription rules have been applied, the program offers the phonological transcription using IPA symbols (see examples in Section 4.2). It should be noted that the transcriber preserves punctuation symbols and, in addition, the spaces between words are not deleted either. The transcription can be copied and pasted in Office documents.

4. VALIDATION AND ASSESSMENT

4.1. Error rate

In order to validate the operation of the tool, the text *El viento norte y el sol*, which is used for IPA illustrations of the varieties of Spanish (Martínez Celdrán, Fernández Planas, & Carrera Sabaté, 2003), has been transcribed using different combinations of dialectal features: a) *lleísmo* + *distinguidor* that appears in (1) below (which corresponds to the sociolect of conservative speakers of Castilian Spanish, see Chapman et al., 1983), b) *sheísmo* + *seseo* that appear in (2) below (which corresponds to the sociolect of innovative speakers of Buenos Aires Spanish, see Wolf & Jiménez, 1979). In both cases, the transcriptions have been checked manually and no error was found.

4.2. Comparison with other transcribers

The comparison between the *lleísmo* + *distinguidor* transcription carried out with *TraFo* (1) and with the *Silabeador y transcriptor fonético y fonológico* (3) points out that the transcription made by *TraFo* is more easily readable and does not imply the assumption of the existence of archiphonemes.

- (1) el biento norte i el sol diskutian kual de los dos era el mas fuerte, quando paso un biaxero enbuelto en una kapa. se pusieron de akuerdo en ke akel ke primero pudiera obligar al biaxero a ke se le kajera la kapa seria considerado el mas fuerte. entonθes el biento norte komenθo a soplar kon mutʃa furia, pero kuinto mas soplabo, mas el biaxero agaraba su kapa, asta ke el biento norte desistio. entonθes el sol briño kon todo su esplendor, e inmediatamente el biaxero se kito la kapa. asi ke el biento norte tubo ke rekonoθer la superioridad del sol.

- (2) el biento norte i el sol diskutian kual de los dos era el mas fuerte, kuando paso un biaxero enbuelto en una kapa. se pusieron de akuerdo en ke akel ke primero pudiera obligar al biaxero a ke se le kafera la kapa seria considerado el mas fuerte. entonses el biento norte komenso a soplar kon mutfa furia, pero kuinto mas soplaba, mas el biaxero agaraba su kapa, asta ke el biento norte desistio. entonses el sol brijo kon todo su esplendor, e inmediatamente el biaxero se kito la kapa. asi ke el biento norte tubo ke rekonoser la superioridad del sol.
- (3) / | eL 'bieNto 'noRte i eL soL disku'tiaN kual de los dos 'era eL mas 'fueRte | 'kuaNdo pa'so uN bia'xero eN'bueLto eN 'una 'kapa | se pu'sieron de a'kueRdo eN ke a'kel ke pri'mero pu'diera obli'gaR aL bia'xero a ke se le ka'yera la 'kapa se'ria koNside'rado eL mas 'fueRte | eN'toNθes eL 'bieNto 'noRte komeN'θo a so'plaR koN 'muça 'furia | 'pero 'kuaNto mas so'plaba | mas eL bia'xero aga'raaba su 'kapa | 'asta ke eL 'bieNto 'noRte desis'tio | eN'toNθes eL soL bri'jo koN 'todo su espleN'doR | e iNmediata'meNte eL bia'xero se ki'to la 'kapa | a'si ke eL 'bieNto 'noRte 'tubo ke rekonos'eR la superioridaD deL soL /

4.3. Limitations

As stated before, *TraFo* does not offer any suprasegmental information. That is, it does not offer syllabification and does not mark the tonic syllable either. In addition, there are some specific problems with the transcription of certain words. Specifically, orthographically non-adapted loanwords (like in <halloween>, which *TraFO* transcribes as /ajogween/ instead of /xaloguin/) and the very few Spanish toponyms and anthroponyms where the grapheme <x> does not represent the cluster /ks/ but a fricative (like in <México>, which *TraFO* transcribes as /mekisiko/ instead of /mexiko/).

5. CONCLUSIONS

In this paper we have presented *TraFo*, a tool that carries our phonological transcriptions of Spanish. We have argued that a utility of this kind was necessary because, at present, there is no online tool that can generate a phonological transcription in several dialects of Spanish and without any embedded preference for a specific theoretical approach.

In the future, an updated version of *TraFo* should include a syllabification system, stress marks and a reliable method of transcribing loanwords and words with other spelling anomalies.

6. REFERENCES

- Armario, J. (2003-2008): *Silabeador y Transcriptor Fonético y Fonológico*. Retrieved from <http://www.respublicae.net/lengua/silabas/index.php>
- Bonaventura, P., F. Giuliani, J. M. Garrido, & I. Ortín (1998). Grapheme-to-phoneme transcription rules for Spanish, with application to automatic speech recognition and synthesis. In *Proceedings of the Workshop 'Partially Automated Techniques Transcribing Naturally Occurring Continuous Speech', 16th August 1998, Université de Montréal, Montreal, Quebec, Canada, Coling-ACL'98*, pp. 33-39.
- Chapman, P., A. Dubra, F. Martínez-Gil, & D. Tritica (1983). *El yeísmo en Covarrubias*. Unpublished manuscript, University of Southern California, Los Angeles.
- Garrido, J. M., M. Codina, M., & K. Fodge (2019). TransText, un transcriptor fonético automático de libre distribución para español y catalán. In J. M. Lahoz, & R. Pérez (Eds.), *Subsidia: Tools and Resources for Speech Sciences* (pp. 27-31). Málaga, Universidad de Málaga.
- López Morràs, X. (2004). *Transcriptor fonético automático del español*. Retrieved from <http://aucel.com/pln/transbase.html>
- Martínez Celdrán, E., A. M. Fernández Planas, & J. Carrera Sabaté (2003). Castilian Spanish. *Journal of the International Phonetic Association*, 33(2), 255–259.
- Molino de Ideas (2012). *Transcriptor fonético*. Retrieved from <http://www.fonemolabs.com/transcriptor.html>
- TALP-UPC (2017). *SAGA – Phonetic transcription software for all Spanish variants*. Retrieved from <https://github.com/TALP-UPC/saga>
- Wolf, Clara, & E. Jiménez (1979). El ensordecimiento de yeísmo porteño, un cambio fonológico en marcha. In A. M. Barrenechea (ed.), *Estudios lingüísticos y dialectológicos: Temas hispánicos* (pp. 115-135). París: Hachette.

EL ESTUDIO DE LOS VOCATIVOS DEL ESPAÑOL EN EL MARCO DEL PROYECTO AMPER-CAT

Lourdes Romera Barrios¹, Paolo Roseano^{1,2}, Ana Ma. Fernández Planas¹, Francesco Rodriguez^{1,3}, Diana Muñoz Builes⁴, Carolina Fabra Navarro¹, Eugenio Martínez Celdrán¹

¹Universitat de Barcelona, ²University of South Africa, ³Albert-Ludwigs-Universität Freiburg,

⁴Pontificia Universidad Católica de Chile

paolo.roseano@ub.edu, lromera@ub.edu, anamariafernandez@ub.edu, francesco.rodriguez@ub.edu, dmmunoz2@uc.cl, martinezceldran@ub.edu

ABSTRACT

Este trabajo presenta los resultados del análisis fonético y fonológico de vocativos en tres variedades de español peninsular (Soria, Albacete, Granada) en el marco del proyecto AMPER. Los vocativos fueron pronunciados por dos hablantes de cada localidad con tres funciones pragmáticas diferentes (saludo, confirmación, reprimenda). Con el fin de obtener datos, se utilizó un *Discourse Completion Task*. En total, se han analizado 162 enunciados.

Palabras clave: entonación, pragmática, vocativos, español, AMPER

This paper presents the results of the phonetic and phonological analysis of vocatives in three varieties of Peninsular Spanish (Soria, Albacete, Granada) within the framework of the AMPER project. Vocatives were uttered by two speakers of each locality with three different pragmatic functions (greeting, confirmation, reprimand). In order to obtain data, a *Discourse Completion Task* was carried out. On the whole, 162 tokens were analysed.

Keywords: intonation, pragmatics, calls, Spanish, AMPER

1. INTRODUCCIÓN

En las últimas décadas el estudio de la entonación de las variedades europeas peninsulares de español ha conocido un desarrollo notable. Desde los trabajos de Navarro Tomás (1944) centrados en el dialecto central, el espectro de las variedades estudiadas se ha ido ampliando, gracias sobre todo a dos proyectos: AMPER (Contini, 1992; Martínez Celdrán y Fernández Planas, 2003-2019) e AIEE (Prieto y Roseano, 2009-2013). Ambos proyectos prevén la recogida de datos acerca de la prosodia de algunas modalidades oracionales básicas. Entre ellas, el AIEE incluye dos tipos de vocativos (que se pueden definir *de llamada* e *insistente*, véase Estebas-Vilaplana y Prieto, 2010, pp. 41-43).

Sin embargo, el análisis de esta modalidad oracional no puede darse por concluido, puesto que quedan por explorar las características de los vocativos en la mayoría de puntos de la geografía peninsular y con funciones pragmáticas distintas a las dos que se han citado anteriormente.

Este trabajo tiene como objetivo principal describir las características entonativas de diferentes tipos de vocativos en tres puntos del dominio lingüístico español. El método empleado con ese fin se detalla en el apartado 2, mientras que el epígrafe 3 resume los resultados principales. A continuación, la sección 4 contiene información sobre la implementación fonética de los tonos fonológicos. Las conclusiones se presentan en el apartado 5.

2. MÉTODO

2.1. Puntos de encuesta, informantes y corpus

Se han grabado dos informantes (un hombre y una mujer) en tres puntos de encuesta: Soria, Albacete y Granada. Cada informante ha grabado 27 vocativos: 3 repeticiones de 3 tipos de vocativos (véase 2.2) con nombre caracterizados por posiciones acentuales distintas (el antropónimo esdrújulo *Bárbara*, el llano *Manolo*, y el agudo *Magalí*). En total, el corpus incluye 162 vocativos.

2.2. Elicitación de datos

Los datos se han elicitado mediante un *Discourse Completion Task* (Blum-Kulka, House y Kasper, 1989) que contenía tres situaciones conversacionales planteadas en trabajos anteriores (Huttenlauch, Feldhausen y Braun, 2018) como aptas para obtener vocativos de saludo (1), de confirmación (2) y de reprimenda (3).

- (1) Estás en el parque de tu ciudad, donde crees que no hay nadie conocido. De repente, ves a tu mejor amiga, que se llama Bárbara, a lo lejos. Te alegras mucho de verla. Llámala (por su nombre) para que se dé la vuelta, te vea y os saludéis.
- (2) Estás en el parque de tu ciudad, donde crees que no hay nadie conocido. De repente, ves a una chica que se parece mucho a tu mejor amiga que se llama Bárbara. Te preguntas si es ella porque no lo tienes muy claro. Llámala (por su nombre) para que se dé la vuelta y puedas comprobar si es ella o no.
- (3) Estás con tu mejor amiga Bárbara y unos amigos en un restaurante. Bárbara está contando una anécdota tuya que te da mucha vergüenza y que no quieres que sepa nadie. Te molesta y quieres que se calle. Llámala (por su nombre) para que se dé cuenta de que te molesta y no diga nada más.

2.3. Anotación prosódica

Una vez grabados, los datos se han anotado prosódicamente en *Praat*. La anotación se ha llevado a cabo en dos niveles: en el nivel fonológico se ha utilizado el sistema *Sp_ToBI* (Hualde y Prieto, 2015), mientras que para la anotación fonética se ha usado el *IPrA* (Hualde y Prieto, 2016).

Desde el punto de vista fonológico, se ha observado la presencia de siete configuraciones nucleares cuya frecuencia de uso varía en función del contexto pragmático y del punto de encuesta, tal y como se detalla en la sección 3: L* H% (Figura 1), L* HL% (Figura 2), L+H* !H% (Figura 3), L+H* H% (Figura 4), L+H* HL% (Figura 5), L+H* L% (Figura 6), y L+H* LH% (Figura 7). Los ejemplos de anotación fonética se proporcionan en el epígrafe 4.

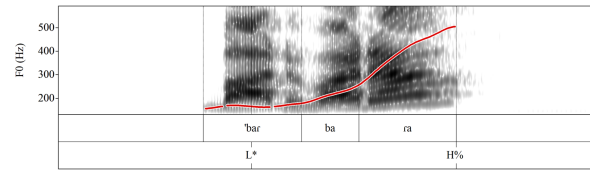


Figura 1: Vocativo L* H% de Albacete.

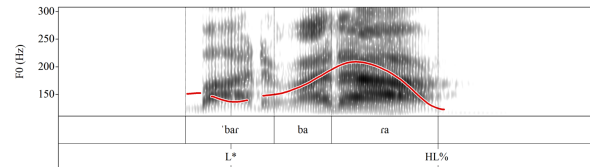


Figura 2: Vocativo L* HL% de Albacete.

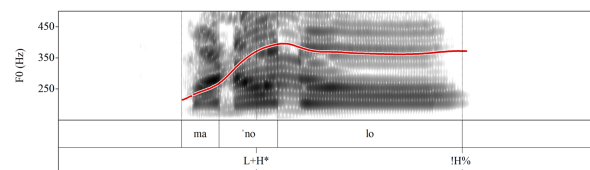


Figura 3: Vocativo L+H* !H% de Soria.

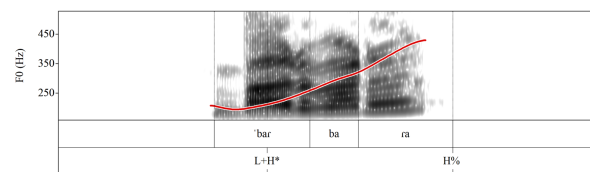


Figura 4: Vocativo L+H* H% de Granada.

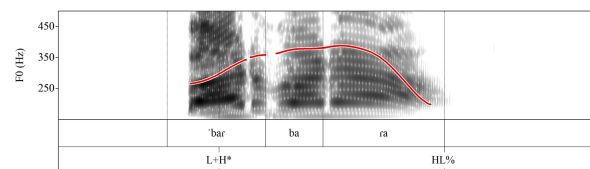


Figura 5: Vocativo L+H* HL% de Soria.

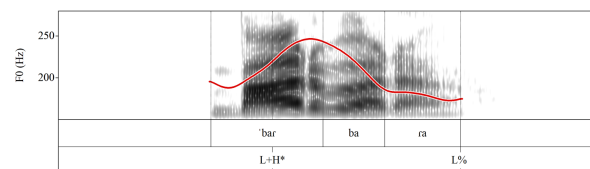


Figura 6: Vocativo L+H* L% de Granada.

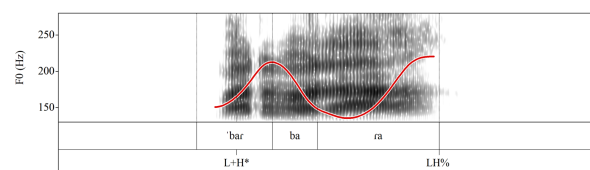


Figura 7: Vocativo L+H* LH% de Soria.

3. RESULTADOS

3.1. Soria

En los datos grabados en Soria aparecen 5 configuraciones nucleares: L+H* !H%, L+H* H%, L+H* HL%, L+H* L% y L+H* LH%. La última de ellas, en nuestro corpus, es exclusiva de esta localidad. Nuestros datos no permiten documentar la presencia de L* H% y L* HL% en la localidad castellana.

3.1.1. Vocativos de saludo

Los vocativos de saludo en Soria presentan el patrón L+H* HL% en un 83% de los casos y L+H* !H% en los demás (Figura 8).

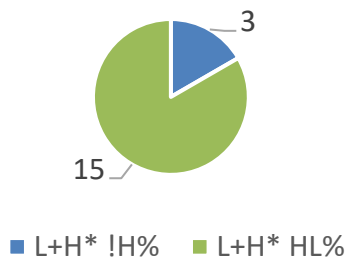


Figura 8: Vocativos de saludo en Soria.

3.1.2. Vocativos de confirmación

Los vocativos de confirmación en Soria presentan el patrón L+H* HL% en la mitad de los casos, el esquema entonativo L+H* H% en un 17% de los casos y L+H* !H% en un 33% de los casos (Figura 9).

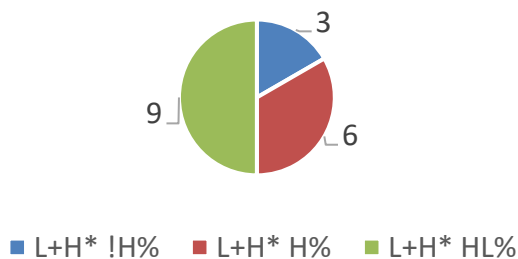


Figura 9: Vocativos de confirmación en Soria.

3.1.3. Vocativos de reprimenda

En el caso de los vocativos de reprimenda, la configuración nuclear predominante es L+H* HL% (50%), mientras son minoritarias (L+H* LH% (33%) y L+H* L% (17%) (Figura 10).

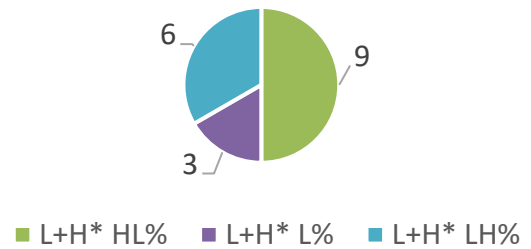


Figura 10: Vocativos de reprimenda en Soria.

3.2. Albacete

En los datos grabados en Albacete aparecen 6 configuraciones nucleares: L* H%, L* HL%, L+H* !H%, L+H* H%, L+H* HL% y L+H* L%.

3.2.1. Vocativos de saludo

En el punto de encuesta manchego los vocativos de saludo presentan los mismos patrones de Soria y con las mismas frecuencias: L+H* HL% en un 83% de los casos y L+H* !H% en los demás (Figura 11).

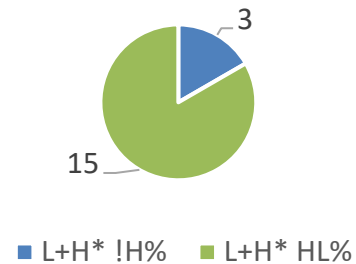


Figura 11: Vocativos de saludo en Albacete.

3.2.2. Vocativos de confirmación

En Albacete los vocativos de confirmación presentan dos configuraciones nucleares que se reparten equitativamente: L+H* H% y L* H% (Figura 12).

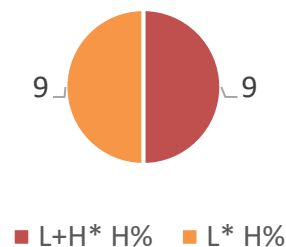


Figura 12: Vocativos de confirmación en Albacete.

3.2.3. Vocativos de reprimenda

Tal y como se aprecia en la Figura 13, en los vocativos de reprimenda de Albacete encontramos tres

esquemas entonativos: L+H* H% (5%), L+H* L% (28%) y L* HL% (67%).

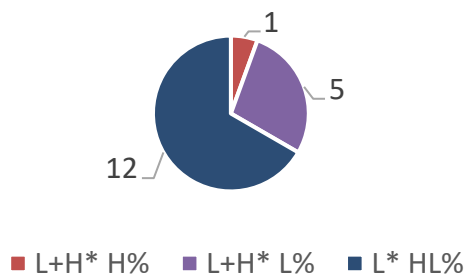


Figura 13: Vocativos de reprimenda en Albacete.

3.3. Granada

En los datos grabados en Granada aparecen las mismas 6 configuraciones nucleares que hemos descrito para Albacete.

3.3.1. Vocativos de saludo

En el punto de encuesta andaluz los vocativos de saludo (Figura 14) se reparten entre las dos configuraciones ya descritas con esta función en Soria y Albacete (L+H* !H% en un 50% y L+H* HL% en un 33%), mientras que el 17% restante utiliza L+H* L%.

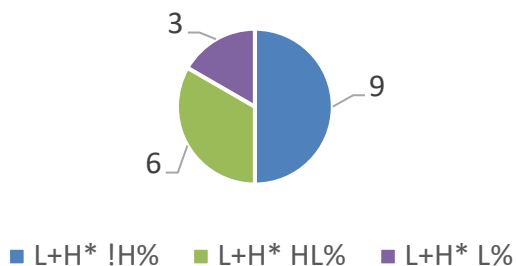


Figura 14: Vocativos de saludo en Granada.

3.3.2. Vocativos de confirmación

Los vocativos de confirmación granadinos (Figura 15) presentan la misma situación que en Albacete: dos configuraciones nucleares se reparten equitativamente el corpus: L+H* H% y L* H%.

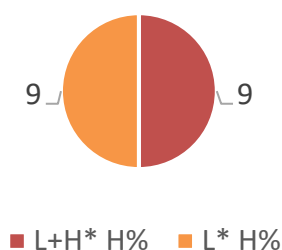


Figura 15: Vocativos de confirmación en Granada.

3.3.3. Vocativos de reprimenda

Tal y como se aprecia en la Figura 16, en los vocativos de reprimenda de Granada encontramos tres configuraciones nucleares: L+H* L% (50%), L+H* HL% (33%) y L* HL% (17%).

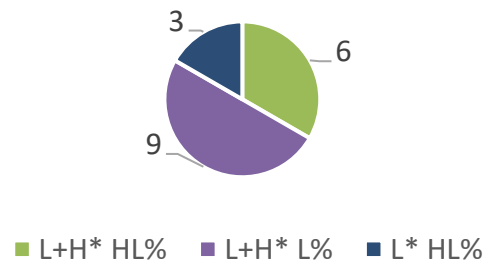


Figura 16: Vocativos de reprimenda en Granada.

3.4. Tendencias generales

En los vocativos de saludo, en todas las localidades aparecen las configuraciones L+H* !H% (que a menudo se ha descrito como la más común en los vocativos, Frota y Prieto, 2015) y L+H* HL% (para la cual se ha propuesto un significado de insistencia o énfasis, Estebas-Vilaplana y Prieto, 2010). Solo en Granada con esta función aparece L+H* L%.

Para los vocativos de confirmación, destaca en los tres puntos de encuesta la presencia de L+H* H%, que aparece de forma casi exclusiva con esta función pragmática. Eso lleva a hipotetizar que tenga una semántica más específica que otros patrones (como L+H* !H% o L+H* HL%) que aparecen en una pluralidad de situaciones.

En nuestros datos relativos a los vocativos de reprimenda, en cada una de las localidades aparecen varios patrones. Nos parece relevante destacar que algunos de ellos no parecen poder utilizarse con otras situaciones del DCT: es el caso de L+H* LH% en Soria y L* HL% en Granada. Estos patrones, por lo tanto, parecen tener una semántica muy específica.

4. IMPLEMENTACIÓN FONÉTICA

En el caso de nombres agudos (*Magali*), se puede producir una situación de *tonal crowding* que da lugar a fenómenos de *text-to-tune alignment* que merecen ser estudiados en profundidad en futuras investigaciones.

En concreto, se han detectado casos de alargamiento de la vocal tónica (Figura 17), como los que se ya se han descrito para el español y otras lenguas (Roseano y Fernández Planas, 2018; Rodríguez, Roseano y Elvira-García, 2020, entre otros).

Además, se han observado casos en que el alargamiento va acompañado de una hiporrealización de una de las dianas tonales. En la Figura 18, por ejemplo, se puede apreciar que la diana tonal L del tono de frontera HL% se realiza como [!H], puesto que el F0 no desciende hasta la franja baja de la tesitura del hablante como hace en las realizaciones canónicas (v. Figuras 7 y 17).

Finalmente, también se han detectado casos de truncamiento tonal parcial, que se caracterizan por no realizar por completo el último movimiento tonal. En la Figura 19, por ejemplo, se observa que la diana tonal L del tono de frontera HL% se realiza como [!H] porque el F0, a pesar de experimentar un descenso, no alcanza su nivel mínimo al final de la emisión como hace en las realizaciones canónicas (v. Figura 5).

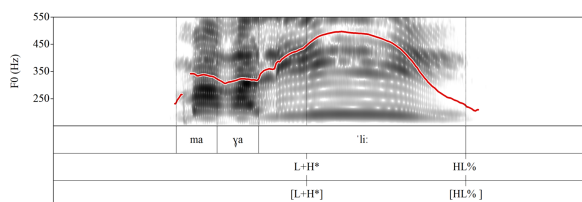


Figura 17: Vocativo L+H* HL% realizado con alargamiento de la vocal tónica (Soria).

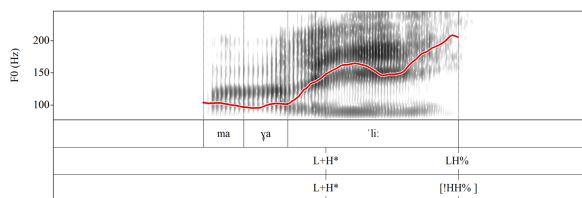


Figura 18: Vocativo L+H* LH% realizado con alargamiento de la vocal tónica e hiporrealización de la diana tonal L del tono de frontera LH% (Soria).

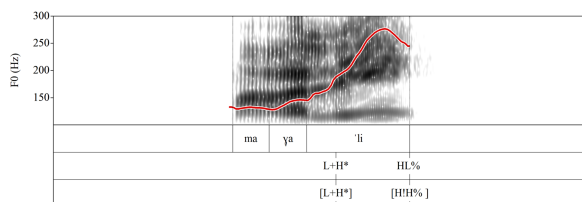


Figura 19: Vocativo L+H* HL% realizado con truncamiento parcial del tono de frontera final (Soria).

5. CONCLUSIONES

Este trabajo ha documentado la presencia de diferentes patrones entonativos en los vocativos de tres puntos de la península ibérica (Soria, Albacete, Granada). La distribución de los patrones por una parte indica la existencia de diferencias dialectales, y

por otra parte depende claramente del contexto pragmático y de la función comunicativa con la que se emiten los enunciados.

6. AGRADECIMIENTOS

Este estudio se ha llevado a cabo en el marco del proyecto de investigación *Tecnologías derivadas de AMPER-CAT y análisis de corpus complementarios*, FFI2015-64859-P (MINECO/FEDER).

7. BIBLIOGRAFÍA

- Blum-Kulka, S., House, J., & Kasper, G. (1989). Investigating cross-cultural pragmatics: An introductory overview. In S. Blum-Kulka, J. House, & G. Kasper (Eds.), *Cross-cultural pragmatics: Requests and apologies* (pp. 1-34). Norwood, NJ: Ablex.
- Contini, M. (1992). Vers une géoprosodie romane. In G. Aurrekoetxea & X. Videgain (Eds.), *Nazioarteko Dialektologia Biltzarra Agiriak* (pp. 83-109). Bilbo: Euskaltzaindia.
- Estebas-Vilaplana, E., & Prieto, P. (2010). Castilian Spanish intonation. In P. Prieto, & P. Roseano (Eds.), *Transcription of the Intonation of the Spanish Language* (pp. 17-48). Lincom: München.
- Frota, S., & Prieto, P. (2015). Intonation in Romance: Similarities and differences. In S. Frota, & P. Prieto (Eds.), *Intonation in Romance* (pp. 392-418). Oxford: Oxford University Press.
- Hualde, J.I., & Prieto, P. (2016). Towards an International Prosodic Alphabet (IPrA). *Laboratory Phonology*, 7(1):25, DOI: 10.5334/labphon.11
- Hualde, J.I., & Prieto, P. (2015). Intonational variation in Spanish: European and American varieties. In S. Frota, & P. Prieto (Eds.), *Intonation in Romance* (pp. 350-391). Oxford: Oxford University Press.
- Huttenlauch, C., Feldhausen, I., & Braun, B. (2018). The purpose shapes the vocative: Prosodic realisation of Colombian Spanish vocatives. *Journal of the International Phonetic Association*, 48(1), 33-56.
- Martínez Celdrán, E., & Fernández Planas, A.M. (coords.) (2003-2019). *Atlas Multimèdia de la Prosòdia de l'Espai Romànic*. <http://stel.ub.edu/labfon/ampere/>
- Navarro Tomás, T. (1944). *Manual de entonación española*. New York: Hispanic Institute in the United States.
- Prieto, P., & Roseano, P. (coords.) (2009-2013). *Atlas interactivo de la entonación del español*. <http://prosodia.upf.edu/atlasentonacion/>
- Rodriguez, F., Roseano, P., & Elvira-García, W. (2020, accepted). Grundzüge der sizilianischen Prosodie. *Dialectologia et Geolinguistica*, 28, 27-50.
- Roseano, P., & Fernández Planas, A.M. (2018). L'amuntegament tonal en castellà, català i friülà en la Teoria de l'Optimitat. *Sintagma*, 30, 23-37.

THE LINGUISTIC RHYTHM OF FRIULIAN: FIRST DATA

Paolo Roseano^{1,2}, Wendy Elvira-García³, Francesco Rodriguez^{1,4}

¹Universitat de Barcelona, ²University of South Africa, ³Universidad Nacional de Educación a Distancia,

⁴Albert-Ludwigs-Universität Freiburg

paolo.roseano@ub.edu, welvira@flog.uned.es, francesco.rodriquez@ub.edu

ABSTRACT

This paper aims at determining the position of Friulian, a Rhaeto-Romance language, in the rhythmic continuum. In order to do so, we collected data from 18 speakers of three languages: 6 speakers of Central Peninsular Spanish (a prototypical syllable-timed language), 6 speakers of Southern British English (a stress-timed language), and 6 speakers of Northern Friulian. The data were analysed acoustically by means of *Praat*. Statistical analyses were carried out with *Correlatore*. Global rhythm metrics (ΔC , ΔV , %V, varco ΔC and varco ΔV) suggest that Friulian is definitely not a syllable-timed language. While some metrics (ΔC , ΔV , varco ΔC , varco ΔV) seem to suggest that it is a stress-timed language, %V raises the question whether it could be a mora-timed language.

Keywords: suprasegmental phonetics, prosody, rhythm, Friulian

1. INTRODUCTION

In the supra-segmental phonetic field, languages are classified according to their rhythmic properties in a continuum ranging from a pole in which stress-timed languages (such as British English and other Germanic languages) are located to a pole in which the syllable-timed languages are found (including the Spanish variety spoken in the central part of the Iberian peninsula, but also several other Romance languages) (Lloyd James, 1940; Pike, 1945; Abercrombie, 1967). The position of each language in this continuum seems to be determined by its phonological characteristics. In this sense, stress-timed languages show complex consonant clusters and vocalic reduction processes, while syllable-timed languages are characterized by simpler consonant clusters and by the absence of vowel reduction processes (Dauer, 1987).

The general purpose of this study is to determine the position of Friulian in the aforementioned continuum, while the specific objective is to verify experimentally the hypothesis that can be formulated on the basis of the phonological features of the language in question.

In the following pages, we shall first offer a short overview of the phonological features of Friulian that can determine its position in the rhythmic continuum (Section 2). In section 3 we shall summarize the most important methodological

aspects. Section 4 contains a description of the results, while Section 5 puts forward some conclusions.

2. CHARACTERISTICS OF FRIULIAN

Friulian is the Easternmost Rhaeto-Romance language and it is spoken in Friuli (Figure 1), a region in Northern Italy, by roughly half a million people.

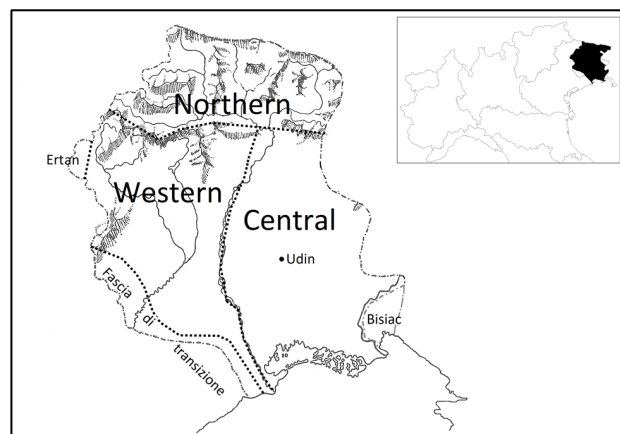


Figure 1: Map of the Friulian-speaking area, with the indication of the main dialects.

2.1. Vowels

Unlike most Romance languages, Friulian displays a contrast between long and short stressed vowels (e.g. ['bru:t] 'daughter in law' ~ ['brut] 'ugly') (Hualde, 1990; Prieto, 1992; losad, 2012; Torres

Tamarit, 2012, 2014). Figure 2 contains an illustration of the stressed vowels of Friulian.

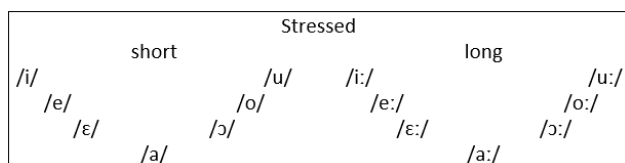


Figure 2: Stressed vowels of Friulian.

This feature, like Dankovičová and Dellwo (2007) put forward for Czech, is likely to have effects on rhythm. Since a contrast between long and short vowels implies a higher variability among vocalic intervals, we expect Friulian to be possibly closer to stress-timed languages than to syllable-timed languages. In addition, Friulian shows vowel reduction, insofar as unstressed vowels are shorter than stressed vowels, and the ratio is 5:9 (Roseano, 2012, p. 107). Since vowel reduction is believed to be a characteristic of stress-timed languages, Friulian seems to be close to such languages from the point of view of vocalic intervals.

2.2. Consonants

Previous studies (Finco, 2015) show that Friulian has a variety of syllable structures (Table 1).

Table 1: Frequency (in 5) of the different syllable structures in Friulian.

Syllable type	Frequency
CV	53.973 %
CVC	21.872 %
V	10.372 %
VC	4.977 %
CCV	3.892 %
CCVC	2.146 %
CVCC	1.323 %
CCVCC	0.818 %
CCCVC	0.370 %
CVCCC	0.074 %
VCC	0.052 %
CCCVC	0.043 %
CCVCCC	0.030 %
CCCVCC	0.021 %
VCCC	0.017 %
CCCVCCC	0.013 %

The presence of complex syllable structures (like CCCVCCC, e.g. in [ˈtrents] ‘tightened.pl’) has led some authors (Roseano, Vanrell, & Prieto, 2015) to suggest that Friulian could be rhythmically closer to

stress-timed languages like English (and in general of Germanic languages) than to Romance languages such as Spanish.

2. METHOD

In order to determine the position of Friulian in the rhythmic continuum, we collected data from 18 speakers of 3 languages (6 of Central Peninsular Spanish, 6 of Southern British English, 6 of Northern Friulian). Central Peninsular Spanish and Southern British English are the prototypical syllable-timed and stress-timed language, respectively, and will be taken as reference marks to establish the position of Friulian in the continuum.

The texts used are the respective versions of *The North Wind and the Sun*. The reading of the text took approximately 30 seconds for each informant. Speakers were recorded by means of a Marantz PMD671 digital recorder connected to a Shure SH58 microphone.

Praat (Boersma & Weenink, 2019) was used to annotate the vocalic and consonantal intervals of the recordings.

The duration of these intervals is the starting point for calculating some of the classical rhythmic parameters: ΔC and %V, ΔC and ΔV (proposed by Ramus, Nespors, & Mehler, 1999), $\text{varco}\Delta C$ and $\text{varco}\Delta V$ (proposed by Dellwo & Wagner, 2003). The metrics in question were calculated and plotted using *Correlatore* (Mairano & Romano, 2010; Mairano, 2011), a statistical software specifically designed for the analysis of rhythm.

3. RESULTS

3.1. ΔC and %V

The first combination of metrics suggested by Ramus et al. (1999) is ΔC and %V. It is worth reminding that ΔC is the standard deviation of the duration of consonant intervals. Linguistically, ΔC is related to the presence of complex syllabic structures (onsets and codas with several C). Typologically, stress-timed languages such as English have a high value of ΔC , while syllable-timed languages such as Spanish have a low value of ΔC .

On the other hand, %V is the percentage of vocalic intervals in the recording. Linguistically, according to Ramus et al. (1999), V is related to both phonological characteristics: with the presence of complex syllabic structures and with the presence of vowel reduction. Typologically, stress-timed languages such as English

have a low value of %V, while syllable-timed languages such as Spanish have a high value of %V. The graph in Figure 3 (where bars show the standard deviation among the speakers of the same language) shows that Friulian is halfway between Spanish and English when we consider ΔC .

The interpretation of %V is more problematic, because our data suggest that Friulian displays a rather high value that, according to Ramus et al. (1999), is typical of mora-timed languages like Japanese.

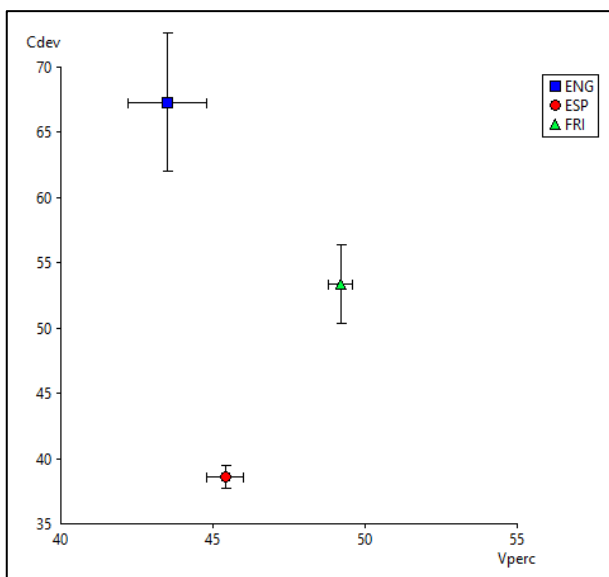


Figure 3: Mean values and standard deviation of ΔC and %V for English (ENG), Spanish (ESP) and Friulian (FRI).

3.2. ΔC and ΔV

The second combination of metrics suggested by Ramus et al. (1999) is ΔC and ΔV . The meaning of ΔC has already been explained in Section 3.1. On the other hand, ΔV is the standard deviation of vocalic intervals. Linguistically, ΔV is related to the phenomenon of vowel reduction. Typologically, stress-timed languages such as English have a high value of ΔV , while syllable-timed languages such as Spanish have a low value of ΔV .

The graph in Figure 4 shows that Friulian, in addition to having an intermediate ΔC value (as already seen in Section 3.1), displays a very high ΔV value, which is higher than English. According to this metric, thus, Friulian would then be definitely stress-timed.

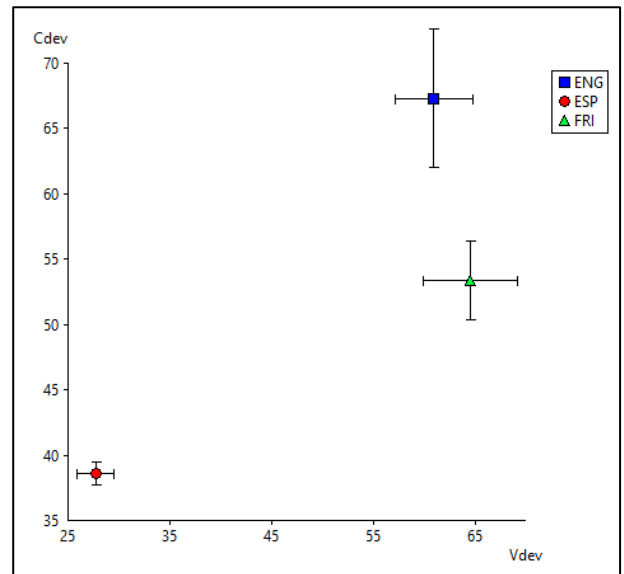


Figure 4: Mean values and standard deviation of ΔC and ΔV for English (ENG), Spanish (ESP) and Friulian (FRI).

3.3. varco ΔC and varco ΔV

The metric that was first used by Dellwo and Wagner (2003) is a version of ΔC and ΔV normalised by speech rate. In our data, the differences between $\Delta C/\Delta V$ and varco ΔC /varco ΔV are not noteworthy. The data plotted in Figure 5, in fact, confirm what we have already observed in Figure 4: Friulian shows varco ΔC values that are intermediate, i.e. between Spanish and English, while varco ΔV values are even higher than in the case of English.

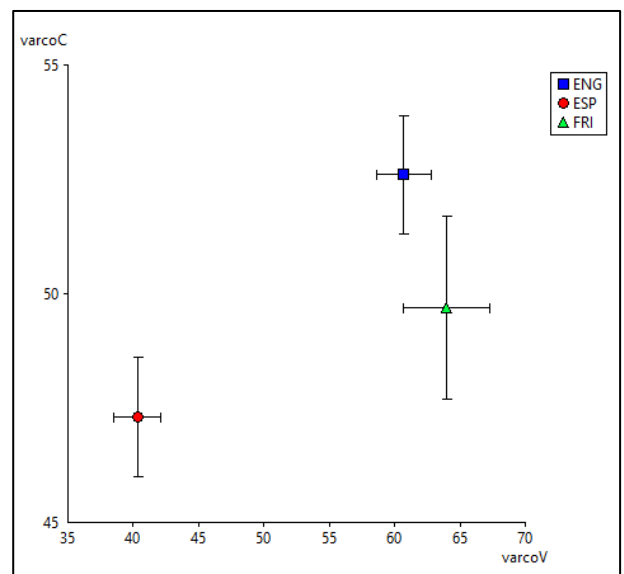


Figure 5: Mean values and standard deviation of varco ΔC and varco ΔV for English (ENG), Spanish (ESP) and Friulian (FRI).

4. DISCUSSION AND CONCLUSIONS

The general objective of this paper was determining the position of Friulian in the rhythmic continuum, while the specific objective was testing empirically the tentative hypothesis by Roseano et al. (2015), according to which Friulian was closer to stress-timed languages than to syllable-timed languages.

The analysis we carried out (which is based on different kinds of so-called *global* metrics of rhythm, like ΔC , ΔV , %V, $\text{varco}\Delta C$ and $\text{varco}\Delta V$) does not give a clear-cut answer to these questions.

As far as vocalic intervals are concerned, Friulian shows %V values that differ from both stress-timed and syllable-timed languages, and are closer to mora-timed languages. If we consider the other two vocalic metrics (ΔV and $\text{varco}\Delta V$), Friulian seems to be definitely stress-timed, like English.

Both ΔC and $\text{varco}\Delta C$ (which measure the variability of consonants) suggest that Friulian lies in an intermediate position between Spanish and English (although slightly closer to the second).

In sum, we can conclude that 1) Friulian is actually rhythmically distant from syllable-timed languages like Spanish, but 2) it is not clear whether it is closer to stress-timed languages like English, to mora-timed languages like Japanese, or if it lies in an intermediate position.

To solve these doubts, future research is needed in two directions. On one side, one will need to include in the analysis data of Japanese, in order to check whether this language and Friulian share any rhythmic features. On the other hand, one will have to explore the results provided by the so-called *local* metrics (like rPVI and nPVI suggested by Grabe & Low, 2002).

5. REFERENCES

- Abercrombie, D. (1967). *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.
- Boersma, P., Weenink, D. (2019). Praat: doing phonetics by computer [Computer program]. Version 6.0.55, Retrieved from <http://www.praat.org/>
- Dankovičová, J. & Dellwo, V. (2007). Czech speech rhythm and the rhythm class hypothesis. In *Proceedings of 16th ICPHS*, pp. 1241–1244.
- Dauer, R. M. (1987). Phonetic and phonological components of language rhythm. En T. Gamkrelidze (Ed.), *Proceedings XIth ICPHS: The eleventh International Congress of Phonetic Sciences* (pp. 447–450). Talin, Estònia: Academy of Sciences of the Estonian S.R.R.
- Dellwo, V., Wagner, P. (2003). Relations between language rhythm and speech rate. In M. J. Solé, D. Recasens, J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 461–474), Barcelona, España: Casual Productions.
- Finco, F. (2015). Phonetics and phonology. In F. Fabbro & C. Crescentini (Eds.), *Handbook of Friulian linguistics* (pp. 31–64). Udine: Forum.
- Grabe, E., & E. L. Low. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven & N. Warner (Eds.), *Papers in Laboratory Phonology* (Vol. 7, pp. 515–546). Berlin: de Gruyter.
- Hualde J.I. (1990). Compensatory lengthening in Friulian. *Probus*, 2, pp. 31–46.
- Iosad P. (2012). A Representational Approach to Final Devoicing and Vowel Lengthening in Friulian. *Lingua*, 122 (8), pp. 922–951.
- Lloyd James, A. (1940). *Speech Signals in Telephony*. London: Sir Isaan Pitman & Sons.
- Mairano, P., & Romano, A. (2010). Un confronto tra diverse metriche ritmiche usando Correlatore. In S. Schmid, M. Schwarzenbach, & D. Studer (Eds.), *La dimensione temporale del parlato*, (Proc. of the V National AISV Congress, University of Zurich, Collegiengebäude, 4–6 February 2009, pp. 79–100), Torriana (RN): EDK.
- Mairano, P. (2011). *Rhythm Typology: Acoustic and Perceptive Studies* (PhD Thesis, Università degli Studi di Torino).
- Pike, K. L. (1945). *The intonation of American English*. Ann Arbor: University of Michigan Press.
- Prieto P. (1992). Compensatory Lengthening by Vowel and Consonant Loss in Early Friulian. In *Catalan Working Papers in Linguistics*, 2, pp. 205–244.
- Ramus, F., Nespors, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73, 265–292.
- Roseano, P. (2012). *La prosòdia del friülà en el marc de l'Atles Multimèdia de Prosòdia de l'Espai Romànic* (PhD thesis, Universitat de Barcelona). Retrieved from <https://www.tdx.cat/handle/10803/96714> (2018-08-28).
- Roseano, P. (2015). Suddivisione dialettale del friulano. In: S. Heinemann, & L. Melchior (Eds.), *Manuale di linguistica friulana* (pp. 155–186). Berlin/New York: De Gruyter.
- Roseano, P., Vanrell, M.M., & Prieto, P. (2015). Intonational phonology of Friulian and its dialects. In S. Frota, & P. Prieto (Eds.), *Intonation in Romance* (pp. 101–139). Oxford: Oxford University Press.
- Torres-Tamarit, F. (2014). Length and voicing in Friulian and Milanese: Or why rule-free derivations are needed. *Natural Language and Linguistic Theory*, 2014. doi: <https://doi.org/10.1007/s11049-014-9271-7>
- Torres-Tamarit, F. (2012) *Syllabification and Opacity in Harmonic Serialism* (PhD thesis, Universitat Autònoma de Barcelona).

A PHONETICALLY BASED VALIDATION OF THE CLASSIFICATION OF FRIULIAN DIALECTS

Paolo Roseano^{1,2}, Franco Finco³

¹Universitat de Barcelona, ²University of South Africa, ³Pädagogische Hochschule Kärnten
 paolo.roseano@ub.edu, franco.finco@ph-kaernten.ac.at

ABSTRACT

This paper aims at validating the traditional classification of Friulian dialects by means of cluster analyses and cluster determinant analyses carried out with *Gabmap* (Nerbonne et al., 2011), a dialectometric online tool. The dataset used for the analyses includes 30 segmental phonetic features and 19 morphological features for each of the 23 local varieties of Friulian (Francescato, 1966; Frau, 1984). The results show that the clustering based on phonetic features largely coincides with the traditional classification of Friulian dialects, whereas the clustering based on morphological features does not. Overall, thus, one can say that the traditional classification of Friulian dialects is mostly –although not completely– based on segmental phonetics.

Keywords: segmental phonetics, cluster analysis, dialectometry, Friulian

1. INTRODUCTION

The beginnings of the study of dialectal variation of Friulian –the easternmost Rhaeto-Romance language– can be traced back to Graziadio Isaia Ascoli’s (1873) *Saggi Ladini*. However, it is only in the 20th century that this discipline experienced a greater development, which resulted in the publication of a considerable number of studies including dialectal data for Friulian, but very few attempts to classify the dialects of that language. The most important classifications were put forward by Francescato (1966) and Frau (1984), who present a concordant description of the dialects of Friulian (Figure 1). All works by other authors (Iliescu, 1972; Marcato, 2001; Vicario, 2005; Heinemann, 2007; Roseano, 2015, i.a.) agree with such classification. Both Francescato (1966) and Frau (1984) base their classification of Friulian dialects on a limited number of isoglosses. Such isoglosses correspond to about fifty linguistic features to which the authors in question attribute particular importance. Such features have to do mainly with the phonetic implementation of phonological categories and, to a lesser extent, with morphology and lexicon (while other aspects of the language, such as syntax, are not taken into consideration).

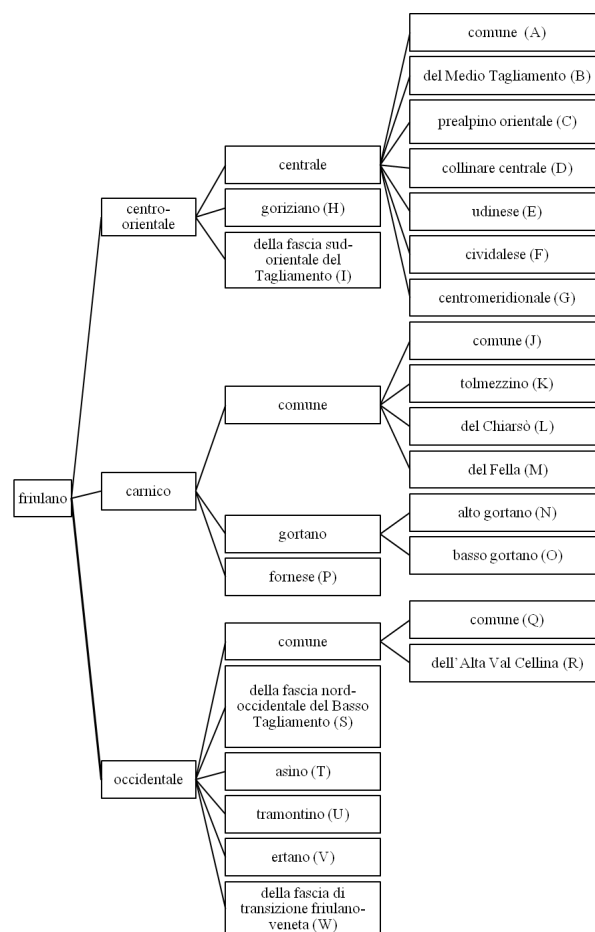


Figure 1: The classification of Friulian dialects according to Francescato (1966) and Frau (1984), adapted from Roseano (2015).

The 23 local varieties of Friulian (indicated with letters from A to W in Figures 1 and 2) are traditionally grouped into 3 dialectal clusters: Northern Friulian (marked in blue on the map in Figure 3), Western Friulian (marked in red), and Central Friulian (in yellow).

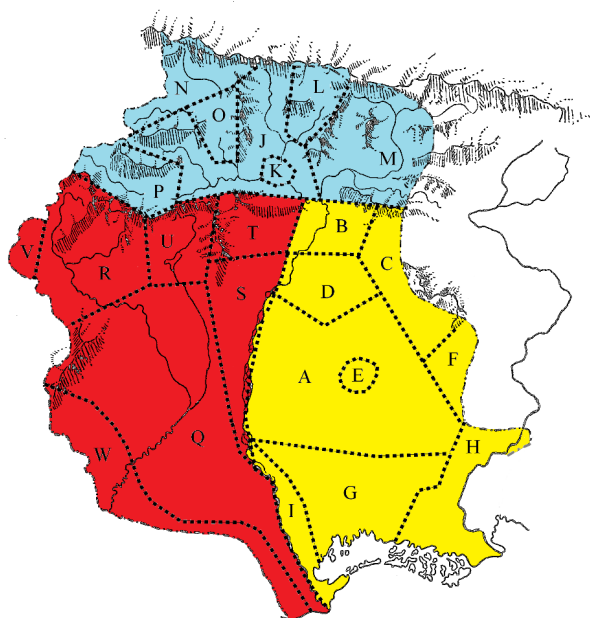


Figure 2: Map of the 23 local varieties and the 3 dialectal clusters of Friulian according to Francescato (1966) and Frau (1984).

A critical aspect in the studies about the classifications of Friulian dialects is the limited number of attempts to use dialectometric techniques (Lazard, 1985; Goebel, 1988; Bauer, 2010, i.a.). In addition to their limited number, all such studies aim at classifying Friulian within the Romance-speaking area (using data from few Friulian localities and other languages), while no attempt has been made to classify the dialects of Friulian themselves. The lack of a proper dialectometric analysis is due to the fact that the most important linguistic atlas for Friulian, the *Atlante storico-linguistico-etnografico friulano* (Pellegrini, 1972-1986), has not been transformed into a computer database yet. On the other hand, since linguistic atlases containing data about prosodic features of Friulian varieties are digitalized, there are several dialectometric analyses of this language based on suprasegmental data (Roseano, 2012, i.a.).

The goal of this paper is checking whether (and, if so, to what extent) the classification of Friulian varieties put forward by previous studies can be justified

statistically on the base of the phonetic and/or morphological features of such varieties.

2. METHODOLOGY

To check whether the classification of Friulian dialects can be validated on the base of the phonetic features of the local varieties of this language, we carried out cluster analyses by means of *Gabmap* (Nerbonne et al., 2011), an online dialectometric tool. Our analyses may not be defined dialectometric *stricto sensu* because in dialectometry the variables are lexical items, whereas in our database we have phonetic and morphological features. The dataset used for the dialectometric analyses, in fact, includes 30 phonetic features and 19 morphological features for each of the 23 local varieties of Friulian. The features are described in detail by Francescato (1966) and Frau (1984), and listing them would exceed the limits of this paper. As an example of a phonetic features we can cite the lengthening of stressed vowel before a rhotic (e.g. /'ar/ is realized as [a:r]); as an example of morphological features we can mention the suffix of 1pp of imperfect indicative (which can be [-'in], [-'on] or [-'en], depending on the variety).

The analyses in *Gabmap* were run using Ward's method, and the software was programmed so that it created 3 clusters (which is the number of dialectal areas described in the previous section). The results were plotted by means of dendrograms created by *Gabmap* and then transposed on dialectal maps adapted from Roseano (2015) by means of *Giphy*.

3. RESULTS

3.1. Analysis with phonetic and morphological features

The analysis carried out with both phonetic and morphological features produces three clearly identifiable clusters (see dendrogram in Figure 3).

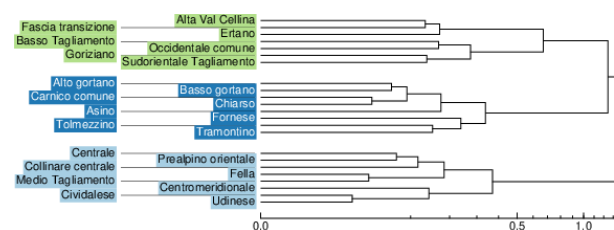


Figure 3: Dendrogram of the 23 local varieties of Friulian created with *Gabmap* using both morphological and phonetic features.

The three clusters roughly correspond to the three dialectal areas illustrated in Figure 2. Nevertheless, if we compare the clusters in Figure 2 with those in Figure 3, we detect that they do not match perfectly. In fact, 4 varieties out of 23 end up in a cluster different from the cluster where Francescato (1966) and Frau (1984) classified them. The differences can be plotted on two parallel maps (Figure 4): the one on the left is the same as in Figure 2, while the one on the right corresponds to the dendrogram in Figure 3. As one can observe, varieties labelled as U, T, M and H are clustered differently. The case of variety H is especially puzzling because, according to the dialectometric analysis we carried out, one cluster is split in two far-away geographic areas (in red in the panel on the right).

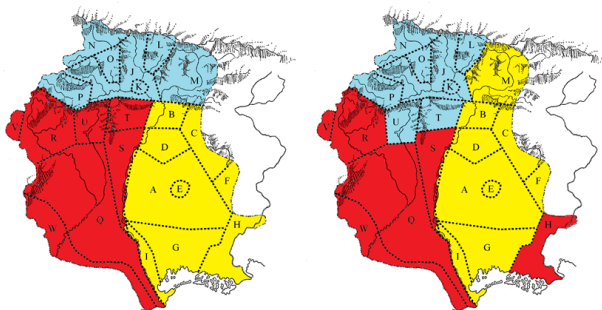


Figure 4: Maps corresponding to the traditional clusters of Friulian dialects (left panel) and to the dialectometric analysis carried out using phonetic and morphological features (right panel).

3.2. Analysis with morphological features only

The analysis carried out with morphological features generates the result illustrated in the dendrogram contained in Figure 5.

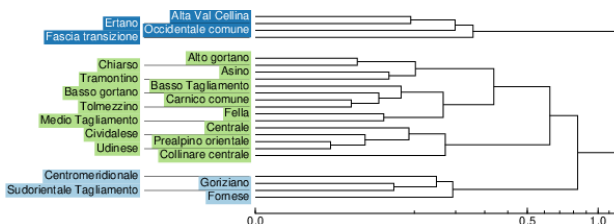


Figure 5: Dendrogram of the 23 local varieties of Friulian created with *Gabmap* using morphological features only.

If we compare the result with the traditional classification of Friulian dialects (by means of the parallel maps in Figure 6), we can conclude that using morphological data gives a much worse result: 10 varieties out of 23 end up in a cluster different from the cluster where Francescato (1966) and Frau (1984) classified them. In short, morphological

features are not a good criterion to validate the traditional classification of Friulian dialects.

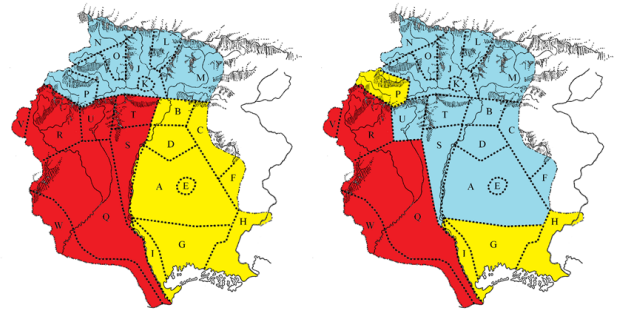


Figure 6: Maps corresponding to the traditional clusters of Friulian dialects (left panel) and to the dialectometric analysis carried out using morphological features only (right panel).

3.3. Analysis with phonetic features only

The analysis carried out with phonetic features only generates the result illustrated in the dendrogram contained in Figure 7.

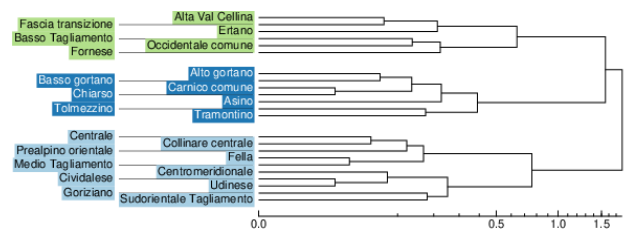


Figure 7: Dendrogram of the 23 local varieties of Friulian created with *Gabmap* using phonetic features only.

The comparison of this clustering with the traditional classification of Friulian dialects (see parallel maps in Figure 8) shows that phonetic features are more effective for validating the classification of Friulian dialect which is accepted by all dialectologists.

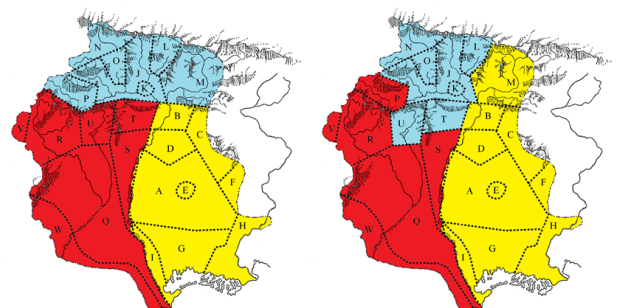


Figure 8: Maps corresponding to the traditional clusters of Friulian dialects (left panel) and to the dialectometric analysis carried out using phonetic features only (right panel).

In fact, 4 out of 23 varieties end up in a cluster different from the one that dialectologists expect. This is the same number of misplaced varieties as in Figure 4 (right panel); nevertheless, differently from what happens in Figure 4, no dialectal cluster is split in two. In this sense, the results of Figure 8 are better from a geolinguistic point of view.

3.3.1. Cluster determinant analysis

In order to find which phonetic features are characteristic of each of the three clusters, a cluster determinant analysis has been carried out. The method used for this analysis is the distance-based one (see Prokić et al., 2012, for an account of the method). The phonetics features that best characterize each dialectal cluster are the following:

- Northern Friulian: presence of word-final long vowels (e.g. Northern Friulian ['fa:] 'to do', instead of the form ['fa] commonly found in the other two areas);
- Western Friulian: presence of the diphthong ['ou] instead of the long vowel ['u:] commonly found in the other two areas (e.g. Western Friulian ['kouɾ] 'heart' instead of ['ku:r] predominant in other areas);
- Central Friulian: presence of [ɲ] instead of [n] in the context known as "*ɲ illegittima*" (Francescato 1966, 62-64) (e.g. Central Friulian ['ɲot] 'night' instead of ['not] found outside the Central Friulian area).

4. CONCLUSIONS

In the previous sections we have summarized the main results of an attempt to validate the traditional classification of Friulian dialects (Francescato, 1966; Frau, 1984) by means of a cluster analysis carried out with *Gabmap* (Nerbonne et al., 2011), a dialectometric online tool. We have shown that the cluster analysis carried out with the morphological features does not give a satisfactory result, insofar as the clusters are pretty different from the three dialectal areas accepted by dialectologists: only 56.5% (13 out of 23) of the local varieties end up in the expected cluster.

On the other hand, the cluster analysis carried out with the segmental phonetic features described by Francescato (1966) gives a better result, in the sense that the local varieties of Friulian end up in the expected cluster in 82.6% (19 out of 23) of the cases. Overall, thus, one can say that the traditional

classification of Friulian dialects is largely based on segmental phonetics.

5. REFERENCES

- Ascoli, G. I. (1873). Saggi ladini. *Archivio Glottologico Italiano* 1, 1-556.
- Bauer, R. (2010). Verifica dialettrometrica della Ladinia di Graziadio Isaia Ascoli (a 100 anni dalla sua morte). In M. Iliescu, H. Siller-Runggaldier, & P. Danler (Eds.), *XXVe CILPR Congrès International de Linguistique et de Philologie Romanes. Innsbruck, 3-8 septembre 2007* (vol. 7 pp. 3-10). Tübingen: Niemeyer.
- Francescato, G. (1966). *Dialettologia friulana*. Udine: Società Filologica Friulana.
- Frau, G. (1984). *I dialetti del Friuli*. Udine: Società Filologica Friulana.
- Goebel, H. (1988). Il posto dialettometrico che spetta ai punti-AIS 338 (Adornano, Friuli), 398 (Dignano/Vodnjan, Istria) e 367 (Grado, Friuli). Presentazione di tre carte di similarità. *Linguistica* 28, 75-103.
- Heinemann, S. (2007). *Studi di linguistica friulana*. Udine: Società Filologica Friulana.
- Iliescu, M. (1972). *Le frioulan à partir des dialectes parlés en Roumanie*. Den Haag/Paris: Mouton.
- Lazard, S. (1985). Les frontières du dialecte frioulan: étude dialectométrique à la lumière de la méthode globale d'Henri Guiter. *Revue de Linguistique Romane* 49, 27-70.
- Marcato, C. (2001). *Friuli Venezia Giulia*. Roma/Bari: Laterza.
- Nerbonne, J., R. Colen, C. Gooskens, P. Kleiweg, & T. Leinonen (2011). Gabmap: A Web Application for Dialectology. *Dialectologia* Special Issue II, 65-89.
- Pellegrini, G. B. (ed.) (1972-1986). *Atlante storico-linguistico-etnografico friulano*. Padova: Istituto di glottologia dell'Università di Padova.
- Prokić, J., Ç. Çöltekin, & J. Nerbonne (2012). Detecting Shibboleths. In M. Butt, & J. Prokić (Eds.), *Visualization of Language Patterns and Uncovering Language History from Multilingual Resources. Workshop at the 13th Conference of the European Chapter of the Association for Computational Linguistics* (pp.72-80). Avignon, France.
- Roseano, P. (2015). Suddivisione dialettale del friulano. In S. Heinemann, & L. Melchior (Eds.), *Manuale di linguistica friulana* (.pp. 155-186). Berlin: De Gruyter Mouton.
- Roseano, P. (2012). *La prosòdia del friülà en el marc de l'Atles Multimèdia de Prosòdia de l'Espai Romànic* (PhD dissertation), Universitat de Barcelona, Spain.
- Vicario, F. (2005). *Lezioni di Linguistica Friulana*. Udine: Forum.

IS CATALAN SYLLABLE-TIMED? AN ANSWER BASED ON CLUSTER ANALYSIS

Paolo Roseano^{1,2}, Patricia Marsà Morales³, Laura Alañá Vilas¹

¹Universitat de Barcelona, ²University of South Africa, ³Universitat Oberta de Catalunya
paolo.roseano@ub.edu

ABSTRACT

Studies about rhythm often try to determine which rhythm class a language belongs to. In this paper, we argue that cluster analysis is a method that can provide a statistically motivated clear-cut answer to this kind of question. In order to test this method, we first used *Correlatore* (Mairano & Romano, 2010) to compute nine rhythm metrics (namely V%, ΔC , ΔV , $\text{varco}\Delta C$, $\text{varco}\Delta V$, CrPVI, VrPVI, CnPVI, and VnPVI) for Catalan, Spanish, Italian, Southern British English and Mainstream American English. Such metrics were then used as variables in a cluster analysis that was carried out with Ward's method by means of *Gabmap* (Nerbonne, Colen, Gooskens, Kleiweg, & Leinonen, 2011). The outcome of the cluster analysis was plotted on a dendrogram. The results suggest that Catalan is syllable-timed, since it clusters with Spanish and Italian.

Keywords: rhythm, rhythm class, cluster analysis, Catalan

1. INTRODUCTION

The aim of this paper is twofold. On the one hand, it aims at contributing to the discussion about the rhythmic properties of Catalan, a language whose classification has been an object of discussion since the '90s (Nespor, 1990). On the other hand, it aims at bringing about a methodological contribution, insofar as it suggests that cluster analysis can be a useful tool for the classification of languages according to their rhythm.

The paper is organized as follows. Section 2 summarizes a few concepts about the rhythm that will be employed to motivate the use of cluster analysis. Section 3 presents the method, while Section 4 shows the most important results. Some conclusions are put forward in Section 5.

2. CLASSIFICATION METHODS IN RHYTHM STUDIES

In order to understand why cluster analysis can be an appropriate statistical tool in rhythm studies, one has to focus on some commonly accepted ideas in this field. Although studies about rhythm have undergone important paradigm shifts since the concept of linguistic rhythm appeared in the first half of the 20th Century (for an overview see Mairano, 2011), the majority of authors nowadays would agree on some basic concepts:

- Languages have different rhythmic patterns, and such differences can be measured.
- There is a linguistic continuum between (at least) two ideal poles of the continuum: stress-timed languages and syllable-timed languages.
- Some languages are commonly acknowledged to be stress-timed (e.g. Southern British English, German or Dutch); other languages are generally considered syllable-timed (e.g. Central Peninsular Spanish or Mandarin Chinese).

In spite of the fact that authors agree that there is a continuum, when researchers study the rhythmic properties of a specific language, they often ask themselves –in a more or less overt manner– which rhythmic *class* such language belongs to: is it syllable-timed or stress-timed? Trying to answer this question means trying to split the *continuum* in (at least) two *classes* (stress-timed and syllable-timed languages) or –which is the same– trying to *classify* rhythmic variation.

The methods used to classify languages rhythmically (i.e. to assign them to a rhythm class) are of two kinds: *graphic* methods and *statistical* methods. Both types of methods have advantages and disadvantages.

2.1. Graphic methods

Graphic methods have been used since the very first studies that employed rhythm metrics (Ramus, Nespors, & Mehler, 1999; Grabe & Low, 2002). These graphics are representations on a 2D Cartesian coordinate system where the axes show the values of the metrics. 2D representations are chosen for obvious practical reasons, but authors are aware that the best solution would be a representation in a multi-dimensional space (Ramus et al., 1999, p. 272). In the case of Catalan, graphic representations can be found in different studies, e.g. in Gavalda-Ferré (2007) (Figure 1) or in Marsà and Roseano (2019) (Figure 2).

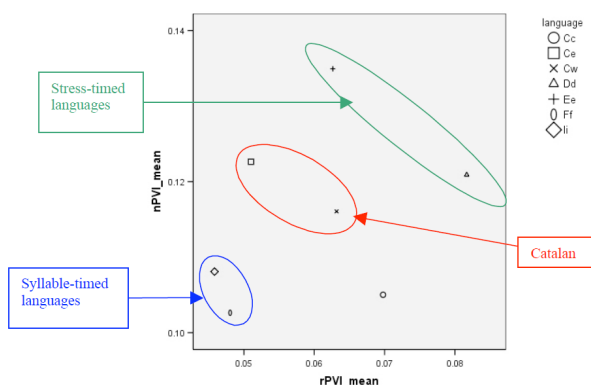


Figure 1: Distribution of languages in the nPVI/rPVI plane according to Gavalda-Ferré (2007); Catalan marked in red.

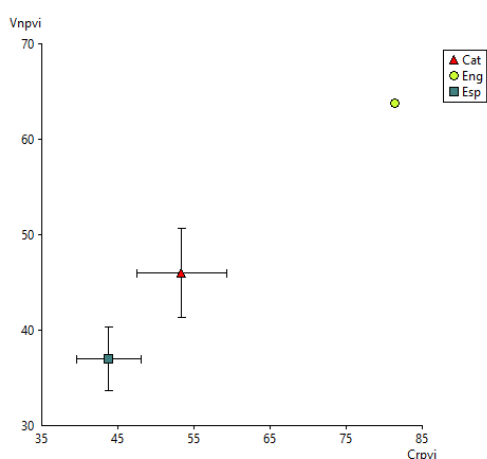


Figure 2: Distribution of languages in the nPVI/rPVI plane according to Marsà and Roseano (2019); Catalan marked in red.

Such representations usually include languages that represent the *poles* of the continuum (e.g. English or Spanish) and the language whose rhythmic properties are analysed. The representations are accompanied by non-quantified comments on the distance between the language studies and the poles

of the continuum, e.g. “both dialects of Catalan are grouped together with the syllable-timed languages –Italian and French–, while the stress-timed languages are far away” (Gavalda-Ferré, 2007, p. 28) or “our data allow us to conclude that Catalan is much closer to Spanish than to English” (Marsà & Roseano, 2019, p. 69).

The main advantage of graphic methods consists in the fact that 2D representations are persuasive and easy to interpret. Their possible disadvantages are two: 1) the clusters of languages are defined on the base of visual inspection of the graph and are not validated quantitatively; 2) graphs have only 2 dimensions corresponding to 2 metrics, even in the studies where three or more rhythm metrics are considered (e.g. Ramus et al., 1999).

2.2. Statistical methods

In order to overcome the limitations of the results provided by graphic analyses, different kinds of statistical methods have been used, though with different goals.

For example, Ramus et al. (1999, p. 272) use ANOVA to test to what extent the classification they had previously obtained using %V and ΔC was reliable. Tortel and Hirst (2010) use discriminant analysis to determine which rhythm metrics allow separating three predefined groups of native and non-native speakers of English. Loukina, Kochanski, Rosner, Keane, and Shih (2010) use machine classification to identify different languages from their rhythm measures. Crucially, in the studies we have mentioned so far, the objective of the statistical analysis was not exactly classifying languages in rhythm classes (at most, they aimed at validating an *existing* classification).

The studies that do use statistics to classify a language usually carry out metric-by-metric analyses. For example, Gavalda-Ferré (2007) uses ANOVA and post hoc tests to check whether the language to be classified (Catalan) differs from the reference languages in 5 rhythm metrics taken individually. Similarly, Prieto, Vanrell, Astruc, Payne and Post (2012) use a Generalized Linear Mixed Model to test whether Catalan differs from English and Spanish in 7 rhythm metrics taken one by one.

To the best of our knowledge, there has been only one attempt to carry out a multi-parameter statistical analysis to classify a language rhythmically: Pukevičiūtė and Kazlauskienė (2013) use cluster analysis to determine which rhythm class Lithuanian and Latvian belong to. In order to do so,

they take five metrics (ΔV , ΔC , $\%V$, $\text{varco}\Delta V$, $\text{varco}\Delta C$) as variables, and use data from other languages considered to be the prototypical examples for the two rhythmic classes.

The use of cluster analysis has several advantages in comparison to the previous approaches. The most noteworthy of them is that it provides a clear-cut answer to the research question ‘Is language X stress-timed or syllable-timed?’, and such answer is based on a statistical analysis.

3. METHODOLOGY

3.1. Speakers, recordings and acoustic analysis

We used recordings of the language to be classified (6 speakers of Central Catalan), in addition to the recordings of 2 syllable-timed varieties (6 speakers of Southern British English, 6 speakers of Mainstream American English) and 2 syllable-timed languages (6 speakers of Central Peninsular Spanish and 6 speakers of Italian). All speakers are native and were asked to read *The North Wind and the Sun* in their respective languages.

As usual in studies on rhythm, vocalic and consonantal intervals were annotated in a *Praat* textgrid (Boersma & Weenink, 2019) for each recording.

3.2. Statistical analysis

The statistical analysis was carried out in two stages. In the first stage, *Correlatore* (Mairano & Romano, 2010) was used to calculate nine rhythm metrics that had been put forward in classical studies about rhythm: $V\%$, ΔC , ΔV , $\text{varco}\Delta C$, $\text{varco}\Delta V$, CrPVI , VrPVI , CnPVI , VnPVI (Ramus et al., 1999; Grabe & Low, 2002; Dellwo & Wagner, 2003).

In the second stage, *Gabmap* (Nerbonne, Colen, Gooskens, Kleiweg, & Leinonen, 2011) was used. *Gabmap* is a free web-based application which has been designed for dialectometric studies. In addition to being able to process alphabetic data (like phonetic transcriptions of words), it also allows the analysis of numeric data (like those of rhythm metrics provided by *Correlatore*). The distance calculated for numeric data is the Euclidean one, while the linkage method several options are offered: Ward’s minimum variance method, group average, weighted average, single.

The two basic kinds of statistical analyses that *Gabmap* carries out are hierarchical cluster analysis and multidimensional scaling. The two techniques are complementary. Multidimensional scaling *per se*

does not group the varieties, but represents them on a continuum, in a way which is very similar to the plots generally used in rhythmic studies (like the ones in Figures 1 and 2). On the other hand, cluster analysis does classify varieties by creating groups or classes. In addition, *Gabmap* also provides graphic representations of the results of the statistical analyses it carries out.

By means of *Gabmap*, we carried out two different analyses:

- Like Pukevičiūtė and Kazlauskienė (2013), we run a cluster analysis. Nevertheless, differently from the above-mentioned authors, we plotted the results of the cluster analysis on dendrograms, that are easy to interpret. From a methodological point of view, one has to point out that for our analysis we have chosen Ward’s linkage method, which minimizes the total within-cluster variance and, therefore, favours the creation of homogeneous clusters.
- In addition to the cluster analysis used by Pukevičiūtė and Kazlauskienė (2013), we also used MDS plots, which are similar to 2D Cartesian representations typically used in rhythmic studies. The main difference between 2D plots (like the ones in Figure 1 and 2) and MDS plots is that the latter are based on the analysis of more than two metrics. To put it in intuitive terms, an MDS plot is the simplification in 2D of a N-dimensional space. In our case, since we have 9 metrics, our MDS plot will be the reduction on a 2D plot of a 9-dimensional space.

4. RESULTS

The values of the metrics obtained by means of *Correlatore* (Table 1) were uploaded in *Gabmap*, which generated the MDS plot in Figure 3 and the dendrogram in Figure 4.

Table 1: Values of the rhythm metrics for Central Catalan (CAT), Central Peninsular Spanish (CPS), Italian (ITA), Mainstream American English (MAE) and Southern British English (SBE).

	CAT	CPS	ITA	MAE	SBE
$V\%$	44.51	45.36	51.15	39.67	44.48
ΔV	36.81	27.68	44.27	51.01	36.55
ΔC	46.80	38.61	41.09	68.67	68.85
$\text{varco}\Delta V$	53.30	40.32	50.53	58.70	60.44
$\text{varco}\Delta C$	55.54	47.29	47.58	53.23	53.74
VnPVI	45.99	36.22	44.57	62.63	65.99
CnPVI	63.54	54.76	56.28	60.63	60.80
VrPVI	34.42	26.06	41.35	55.89	71.26
CrPVI	53.39	44.06	48.87	76.48	76.33

This MDS allows us to get an approximate idea of the distribution of the five varieties in the 9-dimensional space. The MDS plot gives a result that is very close to Gavalda-Ferré (2007) and Marsà and Roseano (2019) (see Figures 1 and 2): graphically speaking, Catalan ends up between stress-timed and syllable-timed languages, although it seems to be closer to stress-timed languages.

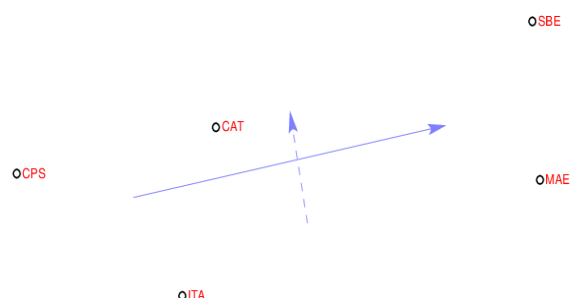


Figure 3: MDS plot based on the values of rhythm metrics for Central Catalan (CAT), Central Peninsular Spanish (CPS), Italian (ITA), Mainstream American English (MAE) and Southern British English (SBE) ($r = 1.0$).

Nevertheless, the MDS plot does not give a clear-cut answer to the research question “Is Catalan stress-timed or syllable-timed?”. On the contrary, the dendrogram corresponding to the cluster analysis (Figure 4) allows the researcher to easily find an unequivocal answer. In fact, Catalan appears grouped up with syllable-timed languages (Central Peninsular Spanish and Italian).

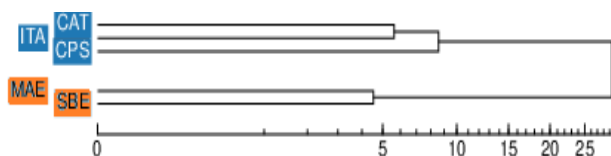


Figure 4: Dendrogram for Central Catalan (CAT), Central Peninsular Spanish (CPS), Italian (ITA), Mainstream American English (MAE) and Southern British English (SBE).

5. DISCUSSION

As we stated in the introduction, the aim of this paper was twofold. On the one hand, it aimed at contributing to characterize from a rhythmic point of view Catalan, a language whose classification has been an object of discussion since the '90s (Nespor, 1990). On the other hand, it intended to bring about a methodological contribution.

As far as the second objective is concerned, we argue that cluster analysis is a useful method in studies about rhythm because it can provide an answer to

the research question “What rhythm class does language X belong to?” and such answer is 1) statistically motivated and 2) clear-cut. MDS plots, on the other hand, do not display these advantages, insofar as they are statistically-based, but do not offer a clear-cut answer to the research question.

As far as the first objective is concerned, our analysis shows that Catalan is clustered with the two syllable-timed varieties analysed (Central Peninsular Spanish and Italian), and not with the two stress-timed varieties (Southern British English and Mainstream American English). This suggests that our data confirm the claim by Ramus, Dupoux and Mehler (2003) and Marsà and Roseano (2019) that Catalan is a syllable-timed language, and not an intermediate language or a language with mixed syllable/stress-timed features like other studies mentioned above argue.

6. FINAL REMARKS

Although the results obtained by means of cluster analysis are encouraging, further work is needed in order to solve some issues raised by this technique.

The first important issue has to do with the amount of gathered data: cluster analysis makes sense and performs well only if the number of objects to be classified is high. This implies that the results of the cluster analysis will be more solid when we add several more languages to the database.

The second point that needs to be solved is to find out which metrics should better be used in the cluster analysis and which should be avoided (and why).

Finally, one will need to consider the possible advantages of running a cluster discriminant analysis to determine which metrics define better each of the two clusters. This could also help shed some light on the capacity of the different rhythm metrics to capture the rhythmic properties of languages.

6. REFERENCES

(Ed.), *Proceedings of Speech Prosody, May 10–14 2010*, Chicago, IL.

- Boersma, P., & Weenink, D. (2019). *Praat: Doing phonetics by computer [Computer program]. Version 6.1.07*, retrieved from <http://www.praat.org/>
- Dellwo, V., & Wagner, P. (2003): Relations between language rhythm and speech rate. In D. Recasens, M. J. Solé, & J. Romero (Eds.), *Proceedings of the 15th ICPHS* (pp. 471-474). Barcelona: Universitat Autònoma de Barcelona.
- Gavaldà-Ferré, N. (2007). *Vowel reduction and Catalan speech rhythm* (MA dissertation, University College London, London, United Kingdom).
- Grabe, E., & Low E. L. (2002): Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven, & N. Warner (Eds.), *Papers in Laboratory Phonology 7*, (pp. 515-546). Berlin: de Gruyter.
- Loukina, A., Kochanski, G., Rosner, B., Keane, E., & Shih, C. (2010). Rhythm measures and dimensions of durational variation in speech. *Journal of the Acoustical Society of America*, 129, 3258-3270.
- Mairano, P. (2010). *Rhythm Typology: Acoustic and perceptive studies* (Doctoral dissertation, Università degli Studi di Torino, Torino, Italy).
- Mairano, P., & Romano, A. (2010). Un confronto tra diverse metriche ritmiche usando Correlatore. In Schmid, S., Schwarzenbach, M., & Studer, D. (Eds.), *La dimensione temporale del parlato, (Proc. of the V National AISV Congress, University of Zurich, Collegiengebäude, 4-6 February 2009)* (pp. 79-100). Torriana (RN): EDK.
- Marsà, P., & Roseano, P. (2019). El ritme del català: anàlisi a partir de textos fonèticament equilibrats. *Estudios de Fonética Experimental*, 28, 47-79.
- Nerbonne, J., Colen, R., Gooskens, C., Kleiweg, P., & Leinonen, T. (2011). Gabmap: A web application for dialectology. *Dialectologia, Special Issue II*, 65-89.
- Nespor, M. (1990). On the rhythm parameter in phonology. In I. M. Roca (Ed.), *Logical issues in language acquisition* (pp. 157-175). Dordrecht: Foris.
- Prieto, P., Vanrell, M.M., Astruc, L., Payne, E., & Post, B. (2012). Phonotactic and phrasal properties of speech rhythm: Evidence from Catalan, English, and Spanish. *Speech Communication* 54(6), pp. 681-702.
- Pukevičiūtė, A., & Kazlauskienė, A. (2013). Klasterinė tipologinių kalbos ritmo grupių analizė. *Lietuvių Kalbotyros Klausimai*, 69, 168-184.
- Ramus, F., Nespor, M., & Mehler, J. (1999): Correlates of linguistic rhythm in the speech signal. *Cognition*, 73(3), 263-292.
- Ramus, F., Dupoux, E., & Mehler, J. (2003): The psychological reality of rhythm classes: Perceptual studies. Paper presented at the *15th International Congress of Phonetic Sciences*, Barcelona, Spain, 3rd-9th August 2003.
- Tortel, A., & Hirst, D. (2010). Rhythm metrics and the production of English L1/L2. In Hasegawa Johnson, M.

ACTOS DE HABLA INDIRECTOS Y SU INTERPRETACIÓN EN EL MARCO PROSÓDICO Y GESTUAL (EN LA BASE DE INTERACCIONES EN ESPAÑOL DE APRENDICES RUSOS)

Albina Sarymsakova

Universidad de A Coruña
albina.sarymsakova@udc.es

ABSTRACT

El objetivo de esta comunicación es hacer una aproximación a la interpretación de los actos de habla indirectos producidos por los hablantes rusos en interacciones comunicativas en español. Dicha interpretación se realizará en función del significado semántico-pragmático que se construye a través de la interacción de los recursos discursivos, la prosodia y el gesto. Se trata de la identificación de dichos actos de habla en la interacción mediante los rasgos característicos de gestos y prosodia, que definimos a lo largo del presente trabajo. La interpretación de los actos de habla indirectos pretende resolver varios problemas en el marco de la comunicación intercultural: A) evitar los malentendidos comunicativos; B) alcanzar los objetivos comunicativos; C) resolver conflictos comunicativos.

Palabras clave: gestos, prosodia, ruso, español, actos de habla indirectos

The aim of this paper is to approach to the interpretation of indirect speech acts produced by Russian speakers in communicative interactions in Spanish. This interpretation will be based on the semantic-pragmatic meaning which is constructed through the interaction of discursive resources, prosody and gesture. This involves the identification of such speech acts in interaction by means of the characteristic features of gesture and prosody, which we define throughout this paper. The interpretation of indirect speech acts aims to solve several problems in intercultural communication: A) to avoid communicative misunderstandings; B) to achieve communicative goals; C) to solve communicative conflicts.

Keywords: gestures, prosody, Russian, Spanish, indirect speech acts

1. INTRODUCCIÓN

La presente investigación se enmarca en el análisis contrastivo del español y ruso; concretamente, nos interesa el estudio del fenómeno pragmático-discursivo del funcionamiento del acto de habla indirecto y su interpretación en relación con la prosodia y el gesto. Por tanto, se trata de una investigación que implica el estudio complejo de interacción de varios fenómenos, entre ellos prosodia, gesto y discurso.

El estado de la cuestión sobre los actos de habla indirectos atrae el interés de los investigadores desde el siglo pasado. En los trabajos de Searle (1975) y posteriormente en los estudios de Asher y Lascarides (2001) estos se definen como actos de habla en los que la intención comunicativa del hablante y el contenido proporcional de su mensaje no coinciden. Según Paul Grice (1975), algunos de los actos de habla indirectos poseen unos marcadores que activan un tipo de

significado implícito y se denominan las implicaturas conversacionales. Dichas implicaturas se transmiten a través de los numerosos indicios de contextualización¹ señalados por Gumperz (1982). En lo que se refiere a aproximación en la descripción de patrones prosódicos que actúan como indicios de contextualización, las propuestas de Hidalgo Navarro (2019) y García Riverón (1996, 1996a, 1998) muestran que el patrón se cambia en función del significado pragmático implícito que transmite el hablante. Finalmente, se han publicado numerosos estudios sobre el papel de gesto en la comunicación no verbal en las obras de Cestero Mancera (2006, 2018).

Sin embargo, existen escasas aproximaciones, entre las cuales aparecen las propuestas de Hidalgo (2019) y Kalan et al. (2019), en la descripción de la interacción de la prosodia y el gesto en los actos de habla indirectos producidos por los aprendices de español como L2; asimismo, encontramos unos estudios limitados, realizados por Иссакова (2003), Ivanova (2013) y

¹ Son pistas discursivas que operan en distintos niveles del discurso (prosodia, gestos, nociones léxicas, estructuras gramaticales, etc.), que transmiten valores comunicativos y conocimiento socio-cultural dentro de

un contexto.

Tereshchuk (2017), referidos a los actos de habla indirectos producidos por los aprendices cuya lengua materna es ruso. Así, estos autores afirman que tanto el gesto como la prosodia son recursos discursivos primordiales para aumentar la fuerza ilocutiva de los actos de habla directos de los hablantes rusos. No obstante, quedan algunas cuestiones que no han sido planteadas por los autores mencionados, y que intentamos abordar en el presente trabajo:

- ¿cuál es el rol de la interferencia de los recursos discursivos (prosodia y gesto) de la lengua rusa en los actos de habla indirectos a lo largo de las interacciones comunicativas en español?
- ¿esta interferencia puede provocar un malentendido comunicativo?

Este trabajo debería, en primer lugar, identificar los obstáculos de los hablantes rusos que surgen en el proceso comunicativo intercultural (en este caso analizamos las interacciones entre los hablantes de español y ruso). En segundo lugar, pretendemos explicar el origen de dichos obstáculos partiendo del análisis de los actos de habla indirectos empleados por los alumnos rusos hablando en español.

En relación con esto, planteamos los siguientes objetivos:

1. Destacar las características de los actos de habla indirectos en el marco prosódico.
2. Destacar las características de los actos de habla indirectos desde lo gestual y quinésico.
3. Analizar el significado semántico-pragmático de los actos de habla indirectos que se construye a través de la interacción de los recursos léxicos, prosódicos y gestuales.

2. METODOLOGÍA

En el presente estudio colaboraron las estudiantes de la Facultad de Filología de la Universidad Federal de Kazán (U FK), Rusia. Para la elaboración de esta investigación, hemos diseñado una actividad didáctica semi-espontánea (*role-play*) en tres situaciones diferentes de conflicto.

El grupo de nuestras informantes consta de tres mujeres rusas. Hicimos las grabaciones en parejas que han sido realizadas en la U FK. Su edad es de 22-28 años. Todas están cursando o han acabado los estudios de formación superior de Filología Extranjera (Filología Hispánica e Inglesa) en la U FK. Las alumnas rusas poseen nivel Avanzado de español (B2-C1). Las grabaciones duraron aproximadamente tres minutos, realizadas con una cámara de vídeo con el objetivo de analizar los gestos, la postura corporal y la expresión facial. Las grabaciones fueron llevadas a cabo siguiendo los tres escenarios de *role-play* mencionados anteriormente.

Hemos utilizado dos programas para procesar los datos grabados. Con el fin de realizar un análisis acústico de vídeos hemos usado Praat, herramienta de análisis acústico, y ELAN para anotar los gestos producidos por

las hablantes.

En cuanto a la base teórica aplicada en la presente investigación, partimos de las siguientes categorías de análisis:

- El paradigma de análisis prosódico incluye el concepto de acto de habla interaccional de García Riverón (1996, 1998), equivalente a la noción de “growth point” de McNeill y McNeill y Duncan (1992, 2000), y la noción de entonema que debe comprenderse como “una unidad de entonación compleja construida por un conjunto de rasgos distintivos (acústicos) y que se realiza en el marco sintagmático”, definido por García Riverón (1996, t. I, p. 99). Asimismo, hemos utilizado la propuesta de Cantero (2002) para analizar los rasgos melódicos a través de los elementos funcionales del contorno: el anacrusis, el primer pico, el cuerpo (o declinación), el núcleo y la inflexión final.
- El análisis de gestos ha sido llevado a cabo a través de la siguiente clasificación: la del grupo de los gestos faciales de González-Fuente, Escandell-Vidal y Prieto (2015); de ellos seleccionamos el movimiento de cejas, boca, cabeza, ojos, cambio de mirada; y la clasificación de los gestos manuales, introducidos por McNeill (1992): icónicos rítmicos deícticos y metafóricos.
- Finalmente, el análisis y la interpretación de los actos de habla indirectos se efectúa a partir de la noción de indicios de contextualización de Gumperz (1982) y de la propuesta de comunicación no verbal en el entorno intercultural de Kendon (2004).

3. EL ANÁLISIS DE LOS DATOS

A continuación, presentamos los datos recopilados y resultados de nuestro análisis.

De 187 actos de habla producidos por las hablantes rusas a lo largo de las interacciones semi-espontáneas, solo 13 actos de habla son indirectos. Así, el porcentaje de los actos indirectos asciende a 6,95%.

A lo largo del análisis de las interacciones de las hablantes no nativas, hemos intentado abordar distintas situaciones de conflicto (relaciones profesionales empleada-clientes (2 situaciones) y empleadas-jefe (1 situación)). Hemos identificado que la mayoría de los actos de habla indirectos aparecen al final de las estructuras argumentativas que despliegan las participantes para defender su imagen. Además, estos actos de habla forman parte de las estrategias comunicativas que emplean las informantes para conseguir sus objetivos: rechazar las estructuras contraargumentativas de su interlocutora, minimizar la imposición y

llegar a la mitigación del conflicto. Asimismo, hemos averiguado cuáles son las funciones pragmáticas de cada uno de los actos de habla indirectos y cuáles de ellos son más eficaces en resolución del conflicto comunicativo. Finalmente, hemos observado que las hablantes usan los rasgos prosódicos y gestuales como indicios de contextualización que activan implicaturas conversacionales y contribuyen a la realización de una u otra función pragmática.

Según se muestra en la Tabla 1, relacionamos las funciones pragmáticas con realización de distintos entonemas y producción de los gestos manuales, la expresión facial y la postura corporal:

Tabla 1: Funciones pragmáticas usadas por las hablantes rusas en relación con entonemas y rasgos gestuales

Función pragmática	Algunos rasgos melódicos relevantes	Entonemas destacados	Rasgos gestuales	Rasgos de expresión facial
Ironía	Movimiento ascendente hasta 73,67% en el cuerpo y el descenso leve en el núcleo hasta 6,61%.	VE-3a.	Rítmicos, deícticos	Sonrisa, subida de cejas, entrecierre de ojos, mirada fija en interlocutor, cabeza inclinada hacia delante
Petición implícita	Movimiento ascendente en el cuerpo hasta 8,84% y un ascenso en el núcleo hasta 55,29%.	VE-1c.	Rítmicos	Boca cerrada, ceño de cejas, ojos abiertos, mirada fija en el interlocutor, cabeza recta
Rechazo implícito	Movimiento ascendente hasta 74,57% en el primer pico y el descenso en la inflexión final hasta 70%.	VE-5a.	Rítmicos, deícticos	Boca abierta, subida de cejas, entrecierre de ojos, cabeza inclinada hacia izquierda, mirada fija en interlocutor
Cortesía positiva	Movimiento ascendente en el cuerpo hasta 31,8% y un descenso en la inflexión final hasta 50,6%.	VE-5b	Ícónicos	Boca cerrada, cejas bajadas, ojos abiertos, mirada fija en el interlocutor, cabeza inclinada hacia delante
Amenaza implícita	Movimiento ascendente en el cuerpo hasta 56,21% y un descenso mínimo en el núcleo hasta 2,48%.	VE-1a	Rítmicos	Boca abierta, cejas bajadas, abertura excesiva de ojos, cabeza recta, mirada fija en interlocutor

4. CONCLUSIONES

Con el presente estudio se ha realizado una aproximación a las cuestiones planteadas al principio de este trabajo:

- La interferencia de los recursos de la lengua rusa en los actos de habla indirectos se

transmite, sobre todo a nivel prosódico ya que el énfasis entonativo se desplaza en función de la intención de las hablantes, es decir con el cambio de la orden de las palabras (lo que observamos en realización de algunas IK (construcciones entonativas), según Svetozarova (1998)).

- Es posible que esta interferencia pueda provocar un malentendido comunicativo si las hablantes, en su habla espontánea, invierten la orden de las palabras en sus enunciados.

En lo que se refiere a los objetivos establecidos en este estudio, hemos conseguido:

- Destacar las características de los actos de habla indirectos en el marco prosódico y en el marco gestual y quinésico, lo que mostramos en la Tabla 1.
- Los recursos no verbales, tales como la entonación, la expresión facial, la postura corporal y el gesto actúan como indicios de contextualización y añaden un valor complementario; de esta manera la intención comunicativa de las no coincide con su significado literal de uno u otro acto de habla.

5. BIBLIOGRAFÍA

- Asher, N. y Lascarides, A. (2001). Indirect speech acts. *Synthese*, 128(1-2), 183-228.
- Cantero, F. J. (2002). *Teoría y análisis de la entonación* (Vol. 54). Edicions Universitat Barcelona.
- Cestero Mancera A. M. (2018). Nonverbal Communication in L2 Spanish Teaching. *L2 Spanish pragmatics: from research to teaching*. Routledge, 90-107.
- Cestero Mancera A. M. (2006). La comunicación no verbal y el estudio de su incidencia en fenómenos discursivos como la ironía. *Estudios de Lingüística*, N. 20, 57-77.
- García Riverón, R. (1996). *Aspectos de la entonación hispánica* (Vol. 22). Universidad de Extremadura.
- García Riverón, R. (1996a). *Aspectos de la entonación hispánica*, 2 tt., Universidad de Extremadura, Cáceres.
- García Riverón, R. M. (1998). *Aspectos de la entonación hispánica*, t. III, Universidad de Extremadura, Cáceres.
- González-Fuente, S., Escandell-Vidal, V., y Prieto, P. (2015). Gestural codas pave the way to the understanding of verbal irony. *Journal of Pragmatics*, 90, 26-47.
- Grice, Paul H. (1975), *Logic and Conversation. Studies in the Way of Words*. Harvard University Press, 22-40.
- Gumperz, John J. (1982). *Discourse strategies*.

- Cambridge: Cambridge University Press.
- Hidalgo Navarro, A. (2019). *Sistema y uso de la entonación en español hablado: Aproximación interactivo-funcional*. Ediciones Universidad Alberto Hurtado
- Ivanova, O. (2013). La competencia pragmática en español L2: un caso de inmigrantes rusohablantes en inmersión lingüística continua. *Lengua y migración*, 5(1), 57-81.
- Kalan, M. Š. (2019). L2 Spanish Pragmatics: From Research to Teaching. *Language and Dialogue*, 9(2), 316-323.
- Kendon, Adam (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.
- McNeill, D. y Duncan, S. (2000). Growth points in thinking-for-speaking. *Language and gesture*, 141-161.
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.
- Searle, J. R. (1975). Indirect speech acts. *Speech acts*, 59-82.
- Svetozarova, N. (1998): Intonation in Russian. *Intonation systems. A survey of twenty Languages*, Cambridge University Press, 261-274.
- Tereshchuk, A. (2017). *La asimilación lingüística de los inmigrantes rusos en Barcelona*. Universitat de Barcelona
- Иссакова, О. П. (2003). *Элементы жестово-мимической системы как значимые компоненты анализа интонационно-стилистических вариантов высказывания (на материале русского языка)*. Москва

WORD-FINAL [Z] IN ENGLISH PLURALIA-TANTUM AND REGULAR-PLURAL NOUNS: A COMPARISON OF DURATION

Marcel Schlechtweg¹, Greville G. Corbett²

Carl von Ossietzky Universität Oldenburg¹, University of Surrey²
marcelschlechtweg@gmail.com¹, G.Corbett@surrey.ac.uk²

ABSTRACT

The word-final [z] of the plurale-tantum noun *goggles* and that of the regular-plural noun *toggles* have traditionally been considered to be identical in acoustic terms. However, in recent years, it has been shown that presumed homophonous word-final sibilants in English actually differ in duration. Research has so far focused on, for instance, the plural, genitive, clitic, and non-affixal *s* in English. Pluralia-tantum nouns have been neglected in previous studies; the present experiment aimed to fill this gap by contrasting the word-final segment of these nouns to the equivalent segment in regular-plural nouns. Concentrating on the reading data of 40 native speakers of British English, we show that the final segment in words like *goggles* and *toggles* does not differ in duration, and we interpret the results against the background of the morpho-syntactic properties of the two.

Keywords: Pluralia-tantum nouns, regular-plural nouns, English, acoustics, word-final [z]

1. INTRODUCTION

It is well known that one and the same form can express different functions and meanings in language. In English, for instance, one finds non-affixal (e.g., *lapse*) and affixal (e.g., *laps*) word-final [s, z]. Further, affixal [s, z] can signal plurality (e.g., *the cars*), genitive case (e.g., *the car's*), both plurality and genitive case (e.g., *the cars'*), be a cliticized (e.g., *it's*), or a third-person form (e.g., *she uses*). Traditionally, it has been assumed that different types of [s, z] are identical in their pronunciation; hence, we should not expect acoustic differences between *cars* and *car's*. In recent years, however, several authors have shown that the different types of English [s, z] can indeed differ in their duration (see, e.g., Plag, Homann, & Kunter, 2017; Schwarzlose & Bradlow, 2001; Seyfarth, Garellek, Gillingham, Ackerman, & Malouf, 2018). The present contribution expands this line of research by investigating yet another type, namely the word-final segment of pluralia-tantum nouns. The work presented here is discussed in detail in Schlechtweg & Corbett (2021), which has been accepted for publication in *Morphology*.

2. THE STUDY

2.1. Type and objective

The study presented below was a reading experiment conducted with Praat (Boersma & Weenink, 2019). It aimed at examining whether the word-final segment of English pluralia-tantum nouns (= PTN) (e.g., *goggles*) and regular-plural nouns (= RPN) (e.g., *toggles*) differs in duration.

2.2. Method

2.2.1. Subjects

40 monolingual native speakers of British English without any speech disorder and with normal or normal-to-corrected vision participated in the study (29 females, 11 males).¹ They were students at the University of Surrey (UK) and their average age was 20.85 years (SD: 2.32).

¹ The data of bilinguals and a person with a speech problem was not considered in the analysis.

2.2.2. Materials

The nine pairs given in Table 1 were used as the test materials, each pair consisting of a RPN and a PTN.

Table 1: Test items.

RPN	PTN
<i>beers</i>	<i>shears</i>
<i>browsers</i>	<i>trousers</i>
<i>yearnings</i>	<i>earnings</i>
<i>fires</i>	<i>pliers</i>
<i>freezers</i>	<i>tweezers</i>
<i>toggles</i>	<i>goggles</i>
<i>gongs</i>	<i>tongs</i>
<i>screens</i>	<i>jeans</i>
<i>Pods</i>	<i>odds</i>

Each test item was embedded in a sentence (see 1).

- (1) a. *The toggles appear to be broken and they're useless.*
b. *The goggles appear to be broken and they're useless.*

All of the 18 test items ended in [z] and were inanimate. Within each pair, the same sentence was used, the only difference being the target word, which was either the RPN or the PTN. The RPN and the respective PTN had the same number of syllables, the same stress pattern, the same rhyme in the ultimate syllable, and at least two identical segments before the target fricative [z]. Further, the frequencies of the RPN and the PTN did not differ from each other (Mean RPN = 4.2 per million words (pmw); SD RPN = 4.7 pmw; Mean PTN = 4.2 pmw; SD PTN = 5.4 pmw; independent t test: $t = 0.02$, $p = .988$). The frequencies are based on the ukWaC corpus², a two-billion-words corpus derived from UK-based web pages. Finally, the sequence “target noun + following word” showed a frequency of 0 pmw in the ukWaC for all 18 test items.

2.2.3. Procedure

The experiment was conducted in a silent room. Subjects sat about 30 cm (12 inch) in front of a large-diaphragm condenser microphone, which, in turn, was placed approximately 30 cm (12 inch) in front of an elevated computer screen. They read the

sentences, which they saw on the screen, silently first and then aloud while being recorded with Praat.

Participants read both items of a pair, that is, 18 test sentences in total. In this way, everyone served as her or his own control and the issue of inter-subject variation was controlled for. 36 filler sentences were interspersed as well. In order to keep a distance between a RPN and the equivalent PTN and to minimize the influence of repetitions, 26 other sentences appeared between one RPN sentence (e.g., the sentences with *toggles*) and the respective PTN sentence (e.g., the sentence with *goggles*). The order of the two conditions “RPN” and “PTN” was counterbalanced both across and within subjects. The item order varied across participants.

The formatting of the computer slides was identical across the two conditions. All sentences occurred in a single line, left aligned, in the middle of the slide (vertical orientation), and in the same font type and size.

2.3. Data analysis

Out of the 720 sound files (40 subjects x 18 sound files per subject), 56 (8%) were excluded due to technical problems or slips of the tongue. The remaining 664 sound files were phonetically segmented with Praat. A RPN (e.g., *toggles*) and the PTN counterpart (e.g., *goggles*) uttered by the same person were analyzed together in order to maximize consistency during the segmentation process. The spectrogram and the waveform were used in combination to identify the beginning and end of the target fricative [z]. For this purpose, the spectrum settings were set to 5,000 to 11,000 Hz. To determine the beginning and end of the target fricative and in accordance with their acoustic characteristics and segmentation strategies documented in the literature (see, e.g., Ladefoged, 2003; Ladefoged & Maddieson, 1996; Machač & Skarnitzl, 2009; Schlechtweg & Härtl, 2020; Turk et al., 2006), we relied on the visibility of increased energy in the higher frequencies in the spectrogram (Criterion 1) and on the fricative noise visible in the waveform (Criterion 2). If the two did not coincide, priority was given to Criterion 1. An example is given in Figure 1.

² <http://corpus.leeds.ac.uk/itweb/htdocs/Query.html#> (Hartley, Sharoff, Stephenson, Wilson, Babych & Thomas 2011).

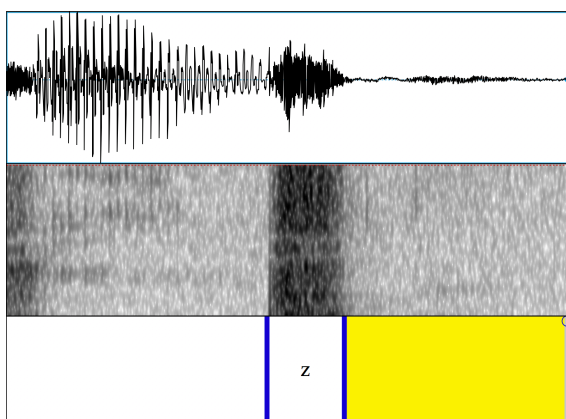


Figure 1: Segmentation of [z] using the information in the spectrogram and the waveform.

2.4. Results

After statistical outliers (= mean plus/minus 2.5 standard deviations) had been excluded, the descriptive statistics given in Table 2 were documented. We clearly see that the two groups of interest do not differ in their duration. In Schlechtweg & Corbett (2021), we present the complete statistical analysis, including linear mixed effects models, and confirm this null effect.

Table 2: Descriptive statistics of RPN and PTN. N = Number of observations. SD = Standard deviation. In seconds (s).

	RPN	PTN
N	321	327
Mean	0.067	0.067
SD	0.017	0.017
Median	0.066	0.066
Minimum	0.028	0.022
Maximum	0.118	0.114

3. CONCLUSION

Our experiment shows clearly that there is no duration difference between the word-final [z] in RPN and PTN. In previous research, it was found that non-affixal and affixal sibilants can be distinguished in their duration. The lack of a duration difference in the present study suggests that the [z] in RPN and PTN are morphologically similar, which is indeed the case as both control plural agreement.

4. REFERENCES

- Boersma, P., & Weenink, D. (2019). Praat: Doing phonetics by computer (Version 6.1.06). [Computer program]. Retrieved from <http://www.praat.org>.
- Hartley, T., Sharoff, S., Stephenson, P., Wilson, J., Babych, B., & Thomas, M. (2011). *IntelliText*. <http://corpus.leeds.ac.uk/itweb/htdocs/Query.html#>.
- Ladefoged, P. (2003). *Phonetic data analysis: An introduction to fieldwork and instrumental techniques*. Malden, MA: Blackwell.
- Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford, UK: Blackwell.
- Machač, P., & Skarnitzl, R. (2009). *Principles of phonetic segmentation*. Prague, Czech Republic: Epocha Publishing House.
- Plag, I., Homann, J., & Kunter, G. (2017). Homophony and morphology: The acoustics of word-final S in English. *Journal of Linguistics*, 53, 181–216.
- Schlechtweg, M., & Corbett, G.G. (2021). The duration of word-final s in English: A comparison of regular-plural and pluralia-tantum nouns. To appear in *Morphology*.
- Schlechtweg, M., & Härtl, H. (2020). Do we pronounce quotation? An analysis of name-informing and non-name-informing contexts. *Language and Speech* 63(4), 769–798.
- Schwarzlose, R., & Bradlow, A.R. (2001). What happens to segment durations at the end of a word? *The Journal of the Acoustical Society of America*, 109, 2292.
- Seyfarth, S., Garellek, M., Gillingham, G., Ackerman, F., & Malouf, R. (2018). Acoustic differences in morphologically-distinct homophones. *Language, Cognition and Neuroscience*, 33(1), 32–49.
- Turk, A., Nakai, S., & Sugahara, M. (2006). Acoustic segment durations in prosodic research: A practical guide. In S. Sudhoff, D. Lenertová, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, & J. Schließer (Eds.), *Methods in empirical prosody research* (Language, Context, and Cognition 3) (pp. 1–27). Berlin, Germany: Walter de Gruyter.

APPENDIX – TEST MATERIALS

His yearnings appear to worry his therapist more and more.

His earnings appear to worry his therapist more and more.

The pods eventually dropped.

The odds eventually dropped.

These new browsers unexpectedly caused a stir.

These new trousers unexpectedly caused a stir.

These old screens obviously need replacing.

These old jeans obviously need replacing.

These small freezers always cause problems.

These small tweezers always cause problems.

The extra beers ended up in the shed.

The extra shears ended up in the shed.

Our brass gongs amazingly sold for several pounds at auction.

Our brass tongs amazingly sold for several pounds at auction.

The fires eased the job of getting rid of the rubbish.

The pliers eased the job of getting rid of the rubbish.

The toggles appear to be broken and they're useless.

The goggles appear to be broken and they're useless.

MELODÍAS ENFÁTICAS EN PORTUGUÉS DE BRASIL

Raquel Sena Mendes¹ y Dolors Font-Rotchés²

¹Instituto Federal de Educación, Ciencia y Tecnología de Brasília, ²Universitat de Barcelona
raquel.mendes@ifb.edu.br, dolorsfont@ub.edu

ABSTRACT

En este trabajo, nos proponemos describir los patrones enfáticos del portugués de Brasil y sus rasgos melódicos. Para llevarlo a cabo, partimos de un corpus de habla espontánea constituido por 69 enunciados enfáticos emitidos por 48 informantes, 21 hombres y 27 mujeres, oriundos de los estados de Rio de Janeiro, Minas Gerais, São Paulo y Goiás. El método utilizado para esta investigación es el Análisis Melódico de Habla (AMH), descrito en Cantero (2002) y establecido en forma de protocolo en Cantero y Font-Rotchés (2009, 2020), que consiste en un análisis acústico-perceptivo de los contornos de las lenguas.

A partir del análisis de los enunciados, hemos obtenido y descrito los rasgos melódicos de cuatro patrones enfáticos en portugués de Brasil: a) el patrón con inflexiones tonales ascendentes en el cuerpo e Inflexión final descendente; b) el patrón Inflexión final prenuclear; c) el patrón Inflexión final descendente-ascendente; y d) el patrón Inflexión final ascendente-descendente

Palabras clave: entonación, enunciados enfáticos, portugués de Brasil, Análisis Melódico del Habla

This paper describes the emphatic patterns for Brazilian Portuguese and their melodic features. The study was based on a spontaneous corpus consisting of 69 emphatic statements issued by 48 informants, 21 men and 27 women, from the states of Rio de Janeiro, Minas Gerais, São Paulo and Goiás. The Melodic Analysis of Speech (MAS) method was used for this research, as described in Cantero (2002) and established as a protocol in Cantero and Font-Rotchés (2009, 2020), which consists of an acoustic-perceptual analysis of the language contours. From the analysis of the statements, the melodic features of four emphatic patterns in Brazilian Portuguese have been obtained and described: a) rising emphasis on the body and a falling final inflection pattern; b) prenuclear final inflection pattern; c) falling-rising final inflection pattern, and d) rising-falling final inflection pattern.

Keywords: intonation, emphatic statements, Brazilian Portuguese, Melodic Analysis of Speech

1. INTRODUCCIÓN

Los estudios sobre la entonación de una lengua y, en este caso, portugués de Brasil, son necesarios porque los resultados obtenidos servirán para establecer métodos de aprendizaje de la entonación y crear materiales de soporte para los profesores de portugués como primera lengua y como lengua extranjera. Con todo ello, los docentes podrán enseñar este aspecto de la oralidad casi siempre desconocido e ignorado en los manuales y gramáticas, pero tan crucial para obtener una buena competencia comunicativa.

El portugués de Brasil cuenta con investigaciones en entonación desde finales del siglo pasado (Moraes, 1993, 1998; Madureira, 1994, entre otros), cuyo número ha ido creciendo a medida que los instrumentos de análisis han sido cada vez más

eficaces. Hoy en día encontramos numerosos trabajos realizados siguiendo la metodología AMPER, cuyos investigadores se interesan principalmente por el análisis de la entonación de las declarativas y las interrogativas de distintas regiones de Brasil. En este sentido, hallamos las investigaciones de Reis et al. (2011), Seara et al., (2011), Antunes y Bodolay (2019) o el reciente volumen editado por Moutinho y Madureira (2019) en el que se presentan trabajos sobre la entonación de distintos estados de Brasil (São Paulo, Maranhão, Minas Gerais, Santa Catarina o Sergipe).

Basadas en el método métrico autosegmental, se han llevado a cabo investigaciones para caracterizar las interrogativas (Cunha y Castelo, 2015; Castelo y Frota, 2017) y las órdenes (Carnaval et al., 2016), entre otras.

Y utilizando el método *Análisis Melódico del Habla* (AMH) se ha descrito la entonación de las interrogativas de Goiás (Cantero y Font-Rotchés, 2013), de Río de Janeiro (Paixão, 2011, 2014), de São Paulo (Mendes, 2013, 2017) y de Minas Gerais (Araújo y Font-Rotchés, 2016), y la entonación de las declarativas de Minas Gerais (Araújo, 2017).

Todas estas investigaciones tratan, en general, la descripción de la entonación de los enunciados declarativos, interrogativos u órdenes del portugués de Brasil, pero, en ningún caso, la de las melodías que contienen énfasis y sus rasgos melódicos. Caracterizar esta tipología de contornos es relevante para poder tener descripciones de aspectos de la oralidad de esta lengua, tales como la (des)cortesía, las emociones o la focalización de una parte o de un enunciado entero. A partir de estas descripciones, se podrán desarrollar actividades didácticas en manuales de enseñanza y aprendizaje del portugués de Brasil, ya sea como primera lengua o como lengua extranjera.

Por todo ello, en el presente trabajo nos proponemos describir los patrones enfáticos del portugués de Brasil, obtenidos en un corpus de habla espontánea y siguiendo el método Análisis Melódico del habla (AMH).

2. METODOLOGIA

Tal como hemos afirmado, se ha utilizado para esta investigación el método Análisis Melódico de Habla (AMH), descrito en Cantero (2002), revisado y ampliado en Font-Rotchés (2007) y establecido en forma de protocolo en Cantero y Font-Rotchés (2009, 2020). Esta metodología consiste en un análisis acústico-perceptivo de los enunciados y ha sido utilizada en diferentes trabajos de investigación, que tratan sobre la entonación de lenguas, como el español (Cantero y Font-Rotchés, 2007; Ballesteros, 2011; Font-Rotchés y Mateo, 2013, 2017; Mateo, 2014), el catalán (Font-Rotchés, 2007), el alemán (Torregrosa y Font-Rotchés, 2017) o el portugués de Brasil (Mendes, 2013, 2017; Araújo, 2014, 2017; Paixão, 2014); sobre los dialectos del español (Mateo, 2021; Ballesteros y Font-Rotchés, 2019; Ballesteros, 2021) o la interlengua español hablado por brasileños (Fonseca, 2013, 2021) por húngaros (Baditzné, 2021) o por sino-hablantes (Cortés, 2004); o sobre los enunciados que expresan (des)cortesía en español (Devís, 2011). En todos estos trabajos de investigación, este método ha sido descrito y verificado, garantizando su eficacia y eficiencia y demostrando ser capaz de ofrecer al investigador soporte suficiente para alcanzar sus objetivos de investigación.

Otra característica de este método es la participación de múltiples informantes sin tener en cuenta características de edad, sexo o escolaridad, por ejemplo. Con esta particularidad, es posible obtener una infinidad de contornos, lo que nos da condiciones para determinar con precisión los patrones melódicos de determinada lengua. Además, este método está basado en habla totalmente espontánea, sin preparación, y en un contexto de habla real, lo que garantiza la proximidad de los análisis de los fenómenos genuinos de la lengua estudiada.

La metodología consiste así en dos fases: la fase acústica y la fase perceptiva. La fase acústica consiste en grabar, segmentar y clasificar enunciados; luego se extraen los valores en Hertz de cada vocal utilizando el *software Praat* (Boersma y Weenink, 1992-2019), que se relativizan contabilizando en porcentajes la distancia tonal entre un segmento tonal y el anterior. Con todo este proceso, se obtiene la curva estándar de la melodía de cada enunciado. La fase perceptiva se refiere a la manipulación de las curvas estándares obtenidas en la primera fase para formular hipótesis que serán comprobadas por medio de pruebas perceptivas para determinar los rasgos melódicos y los márgenes de dispersión de la lengua estudiada.

En este estudio nos centramos en la primera fase, la fase acústica, y describimos los patrones encontrados a partir de este análisis acústico. Para interpretar estos contornos y siguiendo a Navarro Tomás (1944), Cantero (2002) utiliza una estructura de tres partes y dos acentos para los contornos:

- *Anacrusis*: sílabas átonas anteriores que ascienden hasta a la primera vocal tónica o primer pico.
- *Primer pico*: la primera vocal tónica.
- *Cuerpo*: sílabas que van desde el primer pico hasta la última vocal tónica, o núcleo.
- *Núcleo*: la última vocal tónica del contorno.
- *Inflexión final*: las sílabas posteriores al núcleo, que puede presentar distintas direcciones y que se trata de la parte más significativa del contorno.

Analizando estos elementos, es posible definir la melodía del contorno y describir los patrones melódicos de la lengua estudiada.

2.1 El Corpus

El corpus de las enfáticas está compuesto por 69 enunciados extraídos de un corpus más amplio de habla espontánea de portugués de Brasil. Fueron producidos por 48 informantes nativos (21 hombres y 27 mujeres), de diferentes edades y nivel educacional, oriundos principalmente de los estados de Río de Janeiro, São Paulo, Minas Gerais y Goiás.

Los audios fueron extraídos de programas de TV con participación del público, como *Encontro com Fátima Bernardes*; entrevistas en la calle, como en *Jonal Hoje*; y reality shows, como *Big Brother Brasil*. Además, en el corpus de Goiás, fue posible hacer grabaciones de charlas, por ejemplo, una reunión en la que los profesores discutían acciones pedagógicas.

3. RESULTADOS

A continuación, presentaremos los cuatro patrones melódicos enfáticos del portugués de Brasil que hemos obtenido hasta el momento en nuestro corpus y los caracterizaremos a partir de los rasgos que presentan las melodías en las distintas partes de los contornos: presencia (o ausencia) y caracterización del primer pico, movimientos tonales en el cuerpo, posición del núcleo y la dirección de la inflexión final.

3.1 Patrón con inflexiones tonales ascendentes en el cuerpo e inflexión final descendente

En este grupo, hemos clasificado 28 enunciados del corpus que comparten rasgos similares: presentar un cuerpo con dos o más énfasis, que puede ser plano, descendente o ascendente. Con relación al primer pico, la mayor parte de los enunciados no presenta primer pico o tiene un falso primer pico, es decir, hay una intención de formación de primer pico, pero el ascenso o el descenso no superan el 10%. Con relación a la inflexión final, el descenso se sitúa entre un 10% y un 50%.

En la figura 1 presentamos un ejemplo de enunciado con este patrón, procedente de Goiás, aunque también observamos este tipo de contornos en los otros estados. El contorno no tiene primer pico, presenta dos énfasis en el cuerpo que tienen lugar en la palabra *compromiso*, con un ascenso de un 29,7% y un 55,6%, y una inflexión final en *uai*, con un 32% de descenso.

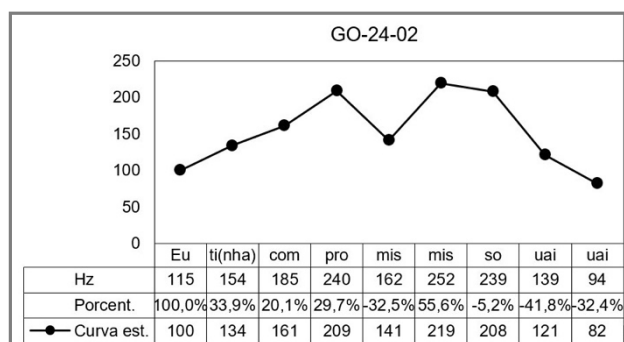


Figura 1: Representación del enunciado GO-24-02, "Eu tinha compromisso, uai!" (¡Yo tenía cita!)

3.2 Patrón con inflexión final prenuclear

El patrón, que se caracteriza por iniciar la inflexión final en la sílaba átona prenuclear, se encuentra en 16 enunciados del corpus analizado. Hay enunciados con la presencia de primer pico, falso primer pico y sin primer pico. Con relación al cuerpo, hay ejemplos de cuerpo descendente, ascendente y elevados. Lo que hace peculiares estos contornos es la posición del núcleo del contorno, que tiene lugar en la sílaba átona previa al núcleo, es decir, en la sílaba prenuclear. Las inflexiones finales de los ejemplos que hemos obtenido son descendentes, entre un 22% y un 71%.

En la figura 2, aportamos un ejemplo de un enunciado que contiene un primer pico en la sílaba *É*, que culmina en un ascenso de un 30%, y una inflexión final, en la que el núcleo empieza en la sílaba átona pretónica *di-* de la palabra *difícil* en lugar de tener lugar en la sílaba tónica *-fi-*. Este rasgo difiere de las tendencias que hemos observado en otro tipo de contornos del portugués de Brasil y de otras lenguas, como español (Cantero y Font-Rotchés, 2007) y catalán (Font-Rotchés, 2007).

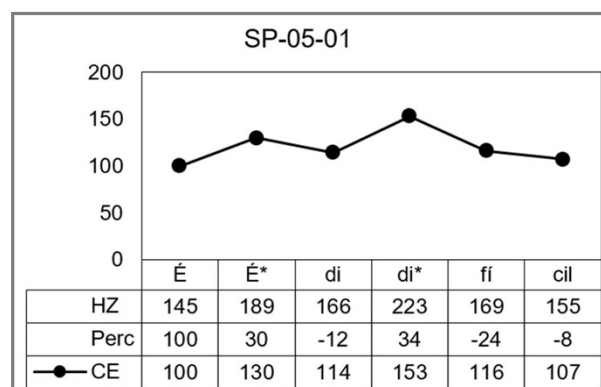


Figura 2: Representación del enunciado SP-05-01, "É difícil!" (¡Es difícil!).

3.3 Patrón con inflexión final descendente-ascendente

El patrón Inflexión final descendente-ascendente aparece en 5 enunciados del corpus y suele presentar un cuerpo ascendente. Los enunciados caracterizados por este patrón no presentan primer pico y la inflexión final contiene tres valores y dos direcciones, una descendente y otra ascendente, a partir del núcleo o de la sílaba tónica. El descenso en las melodías del corpus tiene lugar entre un 10% y un 40%, y el ascenso entre un 12% y un 118%.

En la Figura 3, tenemos un ejemplo de este patrón. La melodía se inicia con un énfasis en la sílaba *Po-* de *Pode*, que tiene tres valores y una inflexión descendente-ascendente. El descenso es leve, un 13%; sin embargo, el ascenso es muy marcado porque

continua hasta la sílaba postónica *-de*, de *Pode*, y suma un 82% de ascenso (un 100% equivale a una octava de la escala musical). La inflexión final, que tiene lugar en *claro*, presenta a partir del primer valor de *cla-*, un 19% de descenso, y, a continuación, asciende un 18% hasta la culminar en la sílaba postónica *-ro*.

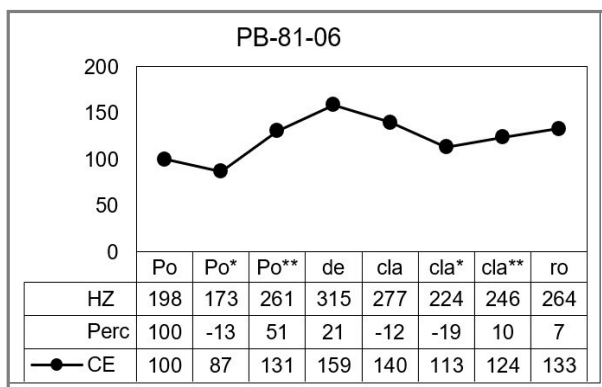


Figura 3: Representación del enunciado PB-81-06, "Pode, claro!" (¡Puedes, claro!)

3.4 Patrón con inflexión final ascendente-descendente

En el corpus, 20 enunciados fueron caracterizados con el patrón inflexión final ascendente-descendente. Gran parte de estos enunciados no presenta primer pico, 70%, aunque algunos contienen primer pico y falso primer pico. Además, con relación al cuerpo, fue posible observar cuerpo plano, elevado, descendente y ascendente. En la inflexión final, los ascensos se sitúan entre un 11% y un 86% y los descensos entre 19% y 73%.

En la Figura 4, aportamos un ejemplo de enunciado que sigue este patrón. Se trata de un contorno constituido solo por inflexión final y que tiene lugar en la palabra *Lógico*. En realidad, la sílaba *Ló-*, el núcleo, contiene la inflexión tonal circunfleja constituida por un ascenso de un 59% seguido de un descenso de un 29%, que culmina en la sílaba átona postónica, *-co*.

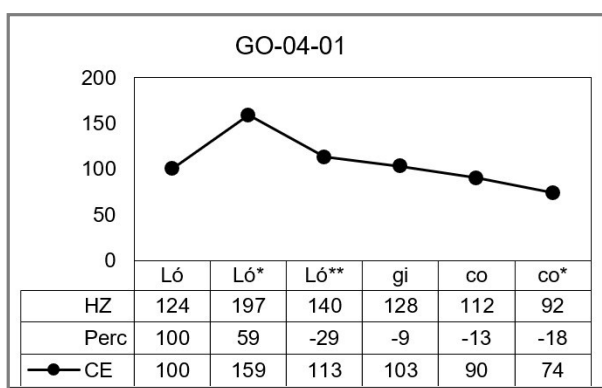


Figura 4: Representación del enunciado GO-04-01, "Lógico!" (¡Lógico!)

4. CONCLUSIONES

A modo de conclusión, constatamos que existen cuatro patrones enfáticos utilizados en la zona centro y sureste de Brasil (Rio de Janeiro, São Paulo, Minas Gerais y Goiás): a) el patrón con inflexiones ascendentes en el cuerpo e Inflexión final descendente; b) el patrón Inflexión final prenuclear; c) el patrón Inflexión final descendente-ascendente; y d) el patrón Inflexión final ascendente-descendente, patrones enfáticos que hasta ahora no habían sido descritos por los investigadores de la entonación. Se trata de una primera fase de una investigación mucho más amplia y, por ello, debemos continuar analizando más corpus para poder obtener más ejemplos y poder concretar, considerando el contexto en el que se dan, cuales son sus usos pragmáticos, y además someter estos rasgos melódicos a pruebas perceptivas para que puedan ser validados por hablantes nativos de portugués de Brasil.

5. BIBLIOGRAFÍA

- Antunes, L. B. y Bodolay, A. N. (2019). Variação prosódica mineira no âmbito do projeto amper-por. *Intercâmbio*, 39, 162-179.
- Araújo, M. (2014). *Entonação das interrogativas e das declarativas do português brasileiro falado em Minas Gerais: Modelos para o Ensino de Línguas*. [Masters Dissert.]. Universidade de Brasília.
- Araújo, M. (2017). Os contornos melódicos das declarativas do português brasileiro falado em Minas Gerais. *Phonica*, 13, 3-17. <https://doi.org/10.1344/phonica.2017.13.3-17>
- Araújo, M. y Font-Rotchés, D. (2016). As interrogativas do português brasileiro falado em Minas Gerais e sua entonação. *CHIMERA: Romance Corpora and Linguistic Studies*, 3(2), 359-374.
- Baditzné Pálvölgyi. (2021). Rasgos del perfil melódico del español meridional. En Cantero F.J.; Font-Rotchés, D.(eds.): *entonaciones del español. Acentos dialectales y acentos extranjeros* (pp. 149-166). Ed. Octaedro.
- Ballesteros, M. P. (2011). *La entonación del español del norte. PhD thesis*. [Universitat de Barcelona]. <https://www.tdx.cat/handle/10803/109963>
- Ballesteros, M. (2021). Rasgos del perfil melódico del español meridional. En Cantero F.J.; Font-Rotchés, D.(eds.): *entonaciones del español. Acentos dialectales y acentos extranjeros* (pp. 25-48). Ed. Octaedro.
- Ballesteros, M. y Font-Rotchés, D. (2019). Acoustic analysis of intonation: Comparison between two dialects of Spanish from the north of the peninsula. *Lingua*, 221, 49-71. <https://doi.org/10.1016/j.lingua.2019.01.006>.
- Boersma, P. y Weenink, D. (1992-2019). *Praat: Doing Phonetics by Computer* (Computer program). Version 6.1.02. <http://www.praat.org>
- Cantero, F. J. (2002). *Teoría y análisis de la entonación*. Edicions Universitat Barcelona.

- Cantero, F. J. y Font-Rotchés, D. (2007). Entonación del español peninsular en habla espontánea: Patrones melódicos y márgenes de dispersión. *Moenia. Revista Lucense de lingüística y literatura*, 13, 69-92.
- Cantero, F. J. y Font-Rotchés, D. (2009). Protocolo para el análisis melódico del habla. *Estudios de fonética experimental*, 18, 17-32.
- Cantero, F. J. y Font-Rotchés, D. (2013). The intonation of absolute questions of Brazilian Portuguese. *Linguistics and Literature studies*, 3, 142-149.
- Cantero, F. J. y Font-Rotchés, D. (2020). Melodic Analysis of Speech (MAS). Phonetics of Intonation. En Abasolo, J.; de Pablo, I.; Ensunza, A.: *Contributions on education* (pp. 20-47). Universidad del País Vasco.
- Carnaval, M., Miranda, L. S., Gomes da Silva, C. y Cunha, C. (2016). A entonação da ordem no português do Brasil: Uma descrição dialetal a partir do corpus ALiB. *JOSS. Journal of Speech Science*, 5, 29-45.
- Castelo, J. y Frota, S. (2017). The yes-no question contour in Brazilian Portuguese: A geographical continuum. En P. Barbosa, C. Paiva, C. Rodrigues(ed.): *Studies on Variation and Change in Varieties of Portuguese*. John Benjamins.
- Cortés, M. (2004). Análisis acústico de la producción de la entonación española por parte de sinohablantes. *Estudios de Fonética Experimental*, XIII, 142-149.
- Cunha, C. y Castelo, J. (2015). Caracterização prosódica das orações interrogativas. En M. Martins & J. Abraçado (Eds.): *Mapeamento sociolinguístico do Português Brasileiro: Teoria, descrição e análise*. Editora Contexto.
- Devís, E. (2011). La entonación de (des)cortesía en el español coloquial. *Phonica*, 7, 36-79.
- Fonseca de Oliveira, A. (2013). *Caracterización de la entonación del español hablado por brasileños* [Universitat de Barcelona]. <http://diposit.ub.edu/dspace/handle/2445/54688>
- Fonseca de Oliveira, A. (2021). Rasgos del perfil melódico del español meridional. En Cantero F.J.; Font-Rotchés, D.(eds.): *entonaciones del español. Acentos dialectales y acentos extranjeros* (pp. 93-110). Ed. Octaedro.
- Font-Rotchés, D. (2007). *L'entonació del català*. Publicacions de l'Abadia de Montserrat.
- Font-Rotchés, D. y Mateo, M. (2013). Entonación de las interrogativas absolutas del español peninsular del sur en habla espontánea. *Onomázein*, 28, 256-275.
- Font-Rotchés, D. y Mateo, M. (2017). Melodías para confirmar, preguntar, sugerir o pedir en español. *Phonica*, 13, 49-67.
- Madureira, S. (1994). Pitch patterns in Brazilian Portuguese: An acoustic phonetic analysis. En *Vth Australian International Conference on Speech Science and Technology*, 156-161.
- Mateo, M. (2014). *La entonación del español meridional*. Tesis Doctoral. Universitat de Barcelona. <http://diposit.ub.edu/dspace/handle/2445/53156>.
- Mateo, M. (2021). Rasgos del perfil melódico del español meridional. En Cantero F.J.; Font-Rotchés, D.(eds.): *entonaciones del español. Acentos dialectales y acentos extranjeros* (pp. 49-72). Ed. Octaedro.
- Mendes, R. S. (2013). *A entonação no processo de ensino aprendizagem de PLE. Proposta didáctica para o ensino de modelos de entonação interrogativa do português do Brasil-Estado de São Paulo. (Masters Dissert.)*. Universidade de Brasília.
- Mendes, R. S. (2017). Análise melódica de padrões interrogativos da fala espontânea do português do Brasil – Estado de São Paulo. *Phonica*, 13, 68-81.
- Moraes, J. A. (1993). *A entonação modal brasileira: fonética e fonologia*. Cadernos de estudos linguísticos, v. 25.
- Moraes, J. (1998). Intonation in Brazilian Portuguese. En Hirst, D.; Di Cristo, A. (eds.): *Intonation Systems. A Survey of Twenty Languages* (pp. 179-194). Cambridge University Press.
- Moutinho, L. y Madureira, S. (ed.). (2019). A prosódia das línguas românicas. *Intercâmbio*, 39.
- Paixão, V. (2011). A entonação das interrogativas absolutas neutras no português do Rio de Janeiro. En Reis, C. (ed.): *Anais do III Colóquio Brasileiro de Prosódia da Fala. Vol.1 (1)*, UFMG, 66-68.
- Paixão, V. (2014). *A prosódia das interrogativas absolutas na fala carioca – leitura versus fala espontânea*. Universidade Federal do Rio de Janeiro.
- Reis, C., Antunes, L. y Pinha, V. (2011). Prosódia de declarativas e interrogativas totais no falar marianense e belorizontino no âmbito do Projeto AMPER. En *Anais do III Colóquio Brasileiro de Prosódia da Fala*. Faculdade de Letras da UFMG.
- Seara, I. C., Figueiredo, M. C. y Berri, A. R. (2011). A entonação do SN-Sujeito no PB falado em Florianópolis sentenças declarativas e interrogativas totais. *RILL*, 17, 157-167.
- Torregrosa, J. y Font-Rotchés, D. (2017). Melodic patterns of absolute interrogative utterances in northern German in spontaneous speech. *Linguistics*, 55(6), 1311-1335. <https://doi.org/10.1515/ling-2017-0029>.

– UN ESTUDIO PILOTO DE EJERCICIOS DE LA PRONUNCIACIÓN DE LAS PALABRAS ÁTONAS Y TÓNICAS CON AYUDA DE MÚSICA

SENSUI Hirotaka¹, KIMURA Takuya², TAKASAWA, Miyuki³

Universidad Nanzan¹, Universidad Seisen², Universidad Asia³
hsensui@nanzan-u.ac.jp¹, kimura@seisen-u.ac.jp², mtakasa9@asia-u.ac.jp³

ABSTRACT

Este artículo es un estudio piloto para practicar la pronunciación de las palabras átonas y tónicas del español con ayuda de música en las clases destinadas a estudiantes japoneses del nivel elemental.

Aunque se utiliza la música en las aulas de español, hay escasos materiales con esta técnica destinados a los estudiantes japoneses. Bajo esta situación Kimura (2019) desarrolló un libro de texto que contiene piezas musicales para aprender este idioma, con atención especial a los aspectos fónicos. Presentamos un experimento piloto para observar su contribución a la didáctica de la pronunciación del español en Japón.

This article is a pilot study to practice pronunciation of stressed and unstressed words in Spanish with music in class targeted at Japanese students in elementary level.

Though music is used in language classes, there are little number of materials of Spanish with this technique for Japanese students. Under this situation Kimura (2019) developed a textbook which contains music pieces to learn this language, with special attention to phonetic aspects. A pilot experiment to observe its contribution to didactics of pronunciation of Spanish in Japan will be presented.

Keywords: palabras átonas, palabras tónicas, música, didáctica, estudiantes japoneses de ELE

1. INTRODUCCIÓN

Este artículo es un estudio piloto para practicar la pronunciación de las palabras átonas y tónicas del español con ayuda de música en las clases destinadas a estudiantes japoneses del nivel elemental.

Se cree que el sistema fónico del japonés es parecido al español, por ejemplo, cinco vocales en ambas lenguas o semejanza en el sistema consonántico. Esta proximidad da una impresión familiar a los estudiantes japoneses hacia este idioma y constituye uno de los elementos motivadores de estudio. Ura (2010: 25) indica que la pronunciación del español no es tan complicada como la del inglés y se habla, se escucha y se entiende durante la clase a través de diálogos aprendidos de memoria. Hamamatsu y Onoda (2014: 162) presentan el grado de dificultad percibida en una escala Likert de 5 puntos por parte de estudiantes japoneses de primer curso no especializados en español, sino que lo estudian como segunda lengua extranjera (n=184), donde se muestra que el promedio marca 2,85, cifra significativamente baja comparada con gramática (3,96), vocabulario (3,84), o comprensión auditiva (3,65).

Sin embargo, esta impresión no siempre refleja la realidad de los estudiantes japoneses; por ejemplo, en Fernández Lázaro *et al.* (2017: 65) es observada como *falso mito*. Los estudios de Carranza (2008, 2012) tratan de algunas dificultades típicas encontradas en la pronunciación de los estudiantes japoneses como /u/ sin redondeamiento o ensordecimiento de las vocales entre dos consonantes sordas. Los autores del presente estudio también han indicado problemas observados en los aspectos fónicos de estudiantes japoneses del español (Sensui (2014, 2017), Kimura (2018)). Sensui (2014) analiza la pronunciación de algunas oraciones de cuatro estudiantes japoneses e indica algunas características entonativas: movimiento ascendente más lento en la parte final de oraciones interrogativas, registro más compacto en preguntas, etc. Por otra parte, Sensui (2017) observa que los estudiantes japoneses tienden a pronunciar sílabas átonas concatenadas con más movimiento tonal que los hablantes nativos. En el estudio de Kimura (2018), se destaca especialmente la baja evaluación a la vocal /u/ no redondeada por parte de los hablantes nativos que participaron en su experimento. De estos estudios se deduce que hay problemas específicos de aspectos fónicos de los estudiantes japoneses de español, que

requieren algún tratamiento para resolverlos en su enseñanza.

En el presente estudio piloto elegimos uno de estos problemas, es decir, la distinción entre las palabras átonas y tónicas, una de las dificultades por parte de estudiantes japoneses (Hara 1990), y presentaremos una propuesta didáctica para entrenar este elemento con ayuda de la música.

2. ESTUDIOS PRELIMINARES Y NUEVOS ENFOQUES

En cuanto a la técnica musical aplicada a la enseñanza de lenguas, aunque ya hay numerosas sugerencias y posibilidades presentadas, quizá el estudio más conocido sea *Jazz Chants* (Graham 1978), que se utiliza en las aulas de inglés. Murphy (1992) también presenta una recopilación de usos de la música en clase de idiomas.

Entre los estudios relacionados con la aplicación de música a la enseñanza del español se encuentran Asensi (1996), Rodríguez López (2006), Cristóbal Hornillos y Villanueva Roa (2015), entre otros.

En cuanto al análisis musical relacionado con el español con atención especial a la relación con los hablantes de japonés, se puede consultar el estudio de Hattori (2010) desde el punto de vista más musical o el de Takasawa (2019), que se refiere más a los aspectos lingüísticos.

Sin embargo, hay escasos trabajos, que sepamos nosotros, sobre la aplicación de música para las aulas del español destinadas a los estudiantes japoneses, aunque algunos profesores pueden usar música en su clase esporádicamente. Una de las excepciones bajo esta situación es la reciente publicación de Kimura (2019), que está destinada a los estudiantes principiantes del idioma. Una de las características más destacadas y originales de este libro es el uso de canciones de corta duración para aprender elementos básicos de español con atención especial a los aspectos fónicos. La colección de todas las piezas del libro compuestas por el mismo autor está colgada en la página web de la editorial (<https://text.asahipress.com/free/spanish/hibikune/index.html>).

El compositor manifiesta que, a la hora de componer la canción titulada “El hombre, un hombre”, que utilizamos en el experimento al que se referirá más abajo (Véase el Apéndice 6.1), se dio más importancia a que la melodía reflejara las sílabas y los acentos. Más concretamente:

(1) Se intentó que cada nota correspondiera a una sílaba. Esto puede que parezca una cosa normal que no necesita mención, pero con este intento se procura ayudar a los estudiantes japoneses que se acostumbran a las sílabas españolas, porque en las canciones japonesas normalmente cada nota no corresponde a

una sílaba sino a una mora, que es la unidad rítmica básica del japonés que no siempre coincide con una sílaba. Por ejemplo, los japoneses tienden a pronunciar la palabra española de tres sílabas *fla-men-co* como *fu-ra-me-n-ko* en cinco moras.

(2) Se intentó que la sílaba tónica correspondiera al primer o tercer pulso, posiciones que llevan acento musical. Por ejemplo, mientras que el artículo *el* del sintagma *el hombre* de la canción arriba mencionada está puesto en la posición de anacrusa (*Auftakt*) y la sílaba tónica *hom-* está situado en el primer pulso, en el caso del sintagma *la mujer*, es la sílaba *-jer* que está puesta en el primer pulso, con las dos sílabas átonas precedentes *la* y *mu-* empujadas a la posición de anacrusa.

(3) Se pretendió también que la presencia y la ausencia de acento se reflejaran en el tono. Por ejemplo, a los artículos definidos *el*, *la*, *los*, y *las* de la canción arriba mencionada está asignado un tono más grave que los del sustantivo siguiente porque aquellos son palabras átonas. A la primera vocal *u-* de los artículos indefinidos *un*, *una*, *unos*, y *unas*, en cambio, se le asignó el tono más agudo de toda la melodía, ya que es una vocal tónica.

Al cantar, se cantó con resilabificaciones, un fenómeno comúnmente observado tanto en conversaciones como en canciones en español. Por ejemplo, aunque el sintagma *el hombre* aparece como *el hombre* en la partitura, está cantado con tres sílabas *e-lom-bre*. Igualmente, *los hombres* está cantado como *lo-som-bres*. La resilabificación es un importante fenómeno fonético que necesitan aprender los estudiantes japoneses de ELE.

3. EXPERIMENTO

3.1. Informantes

En este experimento colaboraron dos estudiantes voluntarios especializados en Estudios Hispánicos de una universidad japonesa (una mujer NF01 y un hombre NM03). Los estudiantes estaban en el segundo semestre del año académico 2019 de Japón y habían recibido aproximadamente un año y medio de instrucción en el aprendizaje del español en el momento del experimento. No tenían un previo conocimiento especial de fonética. Tampoco tenía experiencia de vivir largo tiempo en algún país hispanohablante, excepto un mes de estudios en México.

3.2. Procedimiento

3.2.1. Primera grabación

Los estudiantes grabaron las palabras incluidas en la lista (Apéndice 6.2), que se les presentaron aleatoriamente escritas en tarjetas. En este momento los

colaboradores no habían recibido ninguna instrucción particular con música.

La grabación fue realizada en el estudio de dicha universidad en octubre de 2019 con una grabadora PCM linear (SONY PCM-D10) (frecuencia de muestreo: 44.10 kHz, número de bit de cuantificación: 16bit) y un micrófono (SONY C-357).

3.2.2. Instrucciones en clase

Los estudiantes recibieron instrucciones con la música del Apéndice 6.1 en una parte (10 o 15 minutos) de la clase de gramática en noviembre de 2019. Tuvimos esta sesión en dos clases junto con otros estudiantes. En cada sesión, escucharon primero la música algunas veces, seguido a continuación por la sesión de práctica para cantarla.

3.2.3. Segunda grabación

Después de recibir las instrucciones mencionadas en 3.2.2., se realizó de nuevo la grabación de las mismas palabras. Seguimos el mismo procedimiento y con el mismo equipo de aparatos para esta grabación.

3.3. Análisis y resultados

Según las indicaciones del apartado 2., en este estudio piloto nos enfocamos en los siguientes elementos para analizar el resultado de las grabaciones.

- resilabificación: si no existe ninguna pausa ni oclusiva glotal entre la última consonante del artículo (*el*, *los*, *un*, *unos*) y el comienzo de la palabra *hombre* (En cuanto a la oclusiva glotal, véase Hara (1990: 373), que indica que “nosotros los japoneses ponemos con frecuencia un golpe de glotis ante una vocal que se encuentra en posición inicial de palabra”).
- altura de la vocal acentuada del sustantivo: si la altura de la vocal acentuada del sustantivo es más alta que la del artículo determinado y es más baja (o más o menos igual) que a la del artículo indeterminado
- duración de la vocal en el artículo: si la duración de la vocal del artículo determinado es breve y la del artículo indeterminado es larga

3.3.1. Resilabificación

Observamos gracias al software *Praat* (Boersma & Weenink 2019) si existe alguna pausa u oclusiva glotal entre la última consonante de los artículos y la palabra que sigue (*hombre*) para ver la resilabificación. Las casillas en la Tabla 1 muestran los casos donde apa-

rece estos elementos no favorables. Las tablas de la izquierda corresponden a la primera grabación (antes del entrenamiento) y las de la derecha, a la segunda (después del entrenamiento).

Tabla 1: Existencia de pausa u oclusiva glotal (o.g.) entre la última consonante de los artículos y la palabra que sigue (*hombre*) (duración en msec)

Pre-test			Post-test		
	1aNF01			2aNF01	
	pausa			pausa	
	Sí o No	duración (msec)		Sí o No	duración (msec)
<i>el hombre</i>	o.g.		<i>el hombre</i>	No	
<i>un hombre</i>	o.g.		<i>un hombre</i>	o.g.	
<i>los hombres</i>	No		<i>los hombres</i>	o.g.	100
<i>unos hombres</i>	o.g.		<i>unos hombres</i>	o.g.	

Pre-test			Post-test		
	1aNMD3			2aNMD3	
	pausa			pausa	
	Sí o No	duración (msec)		Sí o No	duración (msec)
<i>el hombre</i>	No		<i>el hombre</i>	No	
<i>un hombre</i>	No		<i>un hombre</i>	o.g.	
<i>los hombres</i>	o.g.		<i>los hombres</i>	No	
<i>unos hombres</i>	No		<i>unos hombres</i>	No	

o.g. = oclusiva glotal

En este elemento, no se observa una clara diferencia entre la primera grabación y la segunda.

3.3.2. Los artículos indeterminado / determinado y el sustantivo

Aquí analizamos la altura de la vocal *u* en los artículos indeterminados y la vocal de los artículos determinados en semitonos. Si la altura del artículo determinado está más alta o la del artículo indeterminado está más baja comparada con el sustantivo que sigue, es decir, en los casos problemáticos, las casillas correspondientes a la columna *diferencia (art-pal)* están sombreadas.

Tabla 2: Diferencias de altura entre la frecuencia del artículo indeterminado / determinado y la sílaba acentuada del sustantivo (en semitonos)

artículo determinado			
Pre-test		Post-test	
	1aNF01		
	frecuencia (st)		
	artículo	palabra	diferencia (art-pal)
<i>el hombre</i>	13.27	14.23	-0.95
<i>la mujer</i>	14.21	11.96	2.25
<i>los hombres</i>	13.89	15.94	-2.05
<i>las mujeres</i>	14.55	13.65	0.90
<i>el libro</i>	12.53	13.95	-1.43
<i>la casa</i>	12.79	13.55	-0.77
<i>los libros</i>	14.12	15.06	-0.93
<i>las casas</i>	12.62	15.04	-2.42

artículo determinado			
Pre-test		Post-test	
	1aNMD3		
	frecuencia (st)		
	artículo	palabra	diferencia (art-pal)
<i>el hombre</i>	1.35	0.69	0.66
<i>la mujer</i>	1.25	0.90	0.35
<i>los hombres</i>	2.68	0.86	1.82
<i>las mujeres</i>	2.03	2.72	-0.68
<i>el libro</i>	1.29	1.75	-0.46
<i>la casa</i>	1.41	2.43	-1.02
<i>los libros</i>	2.91	2.09	0.82
<i>las casas</i>	2.05	1.58	0.47

artículo determinado			
Pre-test		Post-test	
	1aNF01		
	frecuencia (st)		
	artículo	palabra	diferencia (art-pal)
<i>el hombre</i>	12.92	13.41	-0.49
<i>la mujer</i>	14.29	11.49	2.80
<i>los hombres</i>	14.98	13.84	1.14
<i>las mujeres</i>	12.82	13.70	-0.89
<i>el libro</i>	12.77	14.06	-1.29
<i>la casa</i>	11.50	13.35	-1.85
<i>los libros</i>	13.23	15.18	-1.95
<i>las casas</i>	13.25	14.37	-1.13

artículo determinado			
Pre-test		Post-test	
	1aNMD3		
	frecuencia (st)		
	artículo	palabra	diferencia (art-pal)
<i>el hombre</i>	2.89	2.38	0.52
<i>la mujer</i>	1.36	2.39	-1.03
<i>los hombres</i>	2.51	1.02	1.48
<i>las mujeres</i>	2.61	2.28	0.33
<i>el libro</i>	2.08	0.97	1.11
<i>la casa</i>	2.72	3.38	-0.65
<i>los libros</i>	2.67	2.51	0.17
<i>las casas</i>	2.26	1.67	0.60

Pre-test

	1aNF01		
	frecuencia (st)		
	artículo	palabra	diferencia (art-pal)
un hombre	15.60	13.89	1.70
una mujer	14.13	11.80	2.33
unos hombres	15.34	13.14	2.21
unas mujeres	15.78	12.13	3.65
un libro	14.42	12.07	2.35
una casa	15.06	12.04	3.02
unos libros	16.01	14.76	1.25
unas casas	15.56	11.70	3.86

Post-test

	1aNF01		
	frecuencia (st)		
	artículo	palabra	diferencia (art-pal)
un hombre	15.50	13.71	1.80
una mujer	13.19	10.90	2.29
unos hombres	14.86	13.89	0.98
unas mujeres	14.46	11.92	2.54
un libro	15.50	13.18	2.32
una casa	13.82	13.27	0.54
unos libros	14.51	14.07	0.44
unas casas	14.86	11.83	3.03

Pre-test

	1aNMO3		
	frecuencia (st)		
	artículo	palabra	diferencia (art-pal)
un hombre	1.44	1.42	0.03
una mujer	2.86	3.57	-0.71
unos hombres	2.98	1.32	1.66
unas mujeres	3.89	2.52	1.38
un libro	2.82	1.57	1.25
una casa	1.30	0.53	0.76
unos libros	4.63	1.99	2.64
unas casas	3.51	1.98	1.54

Post-test

	1aNMO3		
	frecuencia (st)		
	artículo	palabra	diferencia (art-pal)
un hombre	2.09	1.25	0.85
una mujer	2.27	2.53	-0.26
unos hombres	3.14	0.82	2.32
unas mujeres	3.79	-0.35	4.14
un libro	3.39	2.03	1.36
una casa	3.51	1.21	2.30
unos libros	2.90	1.25	1.65
unas casas	2.45	1.71	0.74

En el caso del artículo indeterminado, los dos estudiantes no muestran tantos problemas, ya que consiguen pronunciar más alto el artículo indeterminado que el sustantivo en mayoría de los casos antes y después del entrenamiento. Sin embargo, el estudiante NM03 empieza a pronunciarlo más alto después del entrenamiento, como está indicado en las casillas cuya cifra está en cursiva/negrita.

En cambio, aunque la estudiante NF01 básicamente no tiene problemas en la pronunciación del artículo determinado, el estudiante NM03 sí tiende a pronunciarlo más alto que el sustantivo, igualmente antes y después del entrenamiento.

3.3.3 Duración de la vocal en el artículo

En este apartado veremos la duración de la vocal y analizaremos la duración de la vocal *u* en los artículos determinados y la de los artículos determinados.

La columna que está más a la derecha significa la diferencia de la duración de la *u* en el artículo indeterminado y la de la vocal en el artículo determinado. Se observa en la Tabla 3 que hay un aumento de la diferencia de la duración en la segunda grabación (en las casillas sombreadas).

Tabla 3: Duración de la vocal *u* del artículo indeterminado, la vocal del artículo determinado y las vocales de las sílabas de cada sustantivo

Pre-test

	1aNF01			
	sílabas (duración msec)			
	artículo	1a sílaba	2a sílaba	3a sílaba
el hombre	120	293	281	21
un hombre	141	261	299	
la mujer	136	189	364	-39
una mujer	97	190	406	
los hombres	101	229	395	8
unos hombres	109	270	441	
las mujeres	132	199	183	408
unas mujeres	107	177	166	398
el libro	134	183	365	36
un libro	170	87	230	
la casa	105	278	201	31
una casa	136	275	298	
los libros	148	263	468	-64
unos libros	84	198	407	
las casas	151	252	457	-15
unas casas	136	232	387	

Post-test

	2aNF01			
	sílabas (duración msec)			
	artículo	1a sílaba	2a sílaba	3a sílaba
el hombre	93	308	292	7
un hombre	100	242	246	
la mujer	104	203	373	1
una mujer	105	170	328	
los hombres	108	300	416	-22
unos hombres	86	274	339	
las mujeres	97	163	206	359
unas mujeres	106	185	175	380
el libro	112	174	251	49
un libro	161	144	255	
la casa	89	280	187	3
una casa	92	263	289	
los libros	101	211	448	-24
unos libros	77	176	392	
las casas	116	250	449	13
unas casas	129	264	377	

	1aNMO3				diferencia (ind-det)
	sílabas (duración msec)				
	artículo	1a sílaba	2a sílaba	3a sílaba	
el hombre	92	180	106		-49
un hombre	43	188	135		
la mujer	41	117	207		
una mujer	40	110	199		-1
los hombres	44	172	275		-13
unos hombres	31	136	269		
las mujeres	56	92	132	241	-26
unas mujeres	30	152	137	252	
el libro	74	7	156		-22
un libro	52	163	177		
la casa	54	141	162		16
una casa	70	116	168		
los libros	43	82	277		13
unos libros	56	140	271		
las casas	58	138	317		-39
unas casas	19	129	327		

	2aNMO3				diferencia (ind-det)
	sílabas (duración msec)				
	artículo	1a sílaba	2a sílaba	3a sílaba	
el hombre	68	200	103		-44
un hombre	24	176	161		
la mujer	47	93	218		10
una mujer	57	142	238		
los hombres	54	134	286		-24
unos hombres	30	131	297		
las mujeres	71	106	166	258	-26
unas mujeres	45	127	144	243	
el libro	58	7	187		-26
un libro	32	132	223		
la casa	44	148	128		-13
una casa	31	139	145		
los libros	62	120	333		-23
unos libros	39	113	343		
las casas	79	137	289		-7
unas casas	72	134	308		

4. OBSERVACIONES

En cuanto a la resilabificación, al menos con el resultado de este experimento, no hemos podido confirmar un efecto concreto tras utilizar esta técnica.

Sobre la pronunciación de los artículos indeterminados, se observa que los dos estudiantes ya habían logrado controlar bastante bien la acentuación de este elemento. Además, como el estudiante NM03 empieza a pronunciar los artículos indeterminados más altos después del tratamiento, se espera la posibilidad de efecto positivo del uso de la canción. En cambio, sobre todo el estudiante NM03 no alcanza un nivel suficiente del manejo del movimiento tonal del artículo definido, pronunciando más alto los artículos que los sustantivos. Esto puede ser por una característica personal y también por una cuestión de hiper-corrección influida por el artículo indeterminado.

En cuanto a la duración, el aumento de la diferencia de la duración en la segunda grabación (en las casillas sombreadas de la Tabla 3) quiere decir que los dos estudiantes, especialmente la estudiante NF01, empiezan a pronunciar las vocales de los artículos indeterminados más largamente que los determinados, fenómeno en el que se espera que esté reflejado el intento de acentuar los indeterminados, después de recibir el tratamiento musical.

5. CONCLUSIONES

Como hemos podido analizar los datos recogidos pertenecen solo a dos personas y todavía no se puede decir algo definitivo. Sin embargo, en los datos de los artículos determinados e indeterminados o la duración de la vocal, hemos visto la posibilidad de que sirva para mejorar la pronunciación el uso de la música. Seguiremos llevando a cabo este tipo de experimentos y estamos en preparación de nuevos trabajos con más número de estudiantes e intentaremos recopilar sus datos.

6. APÉNDICE

6.1 Partitura de la música utilizada

El hombre, un hombre

© KIMURA, Takuya (2019)

6.2 Lista de las palabras grabadas

- | | | |
|------------------|------------------|------------------|
| (1) el hombre | (7) las mujeres | (13) los libros |
| (2) un hombre | (8) unas mujeres | (14) unos libros |
| (3) la mujer | (9) el libro | (15) las casas |
| (4) una mujer | (10) un libro | (16) unas casas |
| (5) los hombres | (11) la casa | |
| (6) unos hombres | (12) una casa | |

7. BIBLIOGRAFÍA

- Asensi, J. S. (1996). Música española contemporánea en el aula de español. *Actuales tendencias en la enseñanza del español como lengua extranjera II: actas del VI Congreso Internacional de ASELE*, 367-378.
- Boersma, P. & Weenink, D. (2019). Praat: doing phonetics by computer [software] (Version 6.0.48). <http://www.praat.org/> [último acceso 18-02-2019]
- Cristóbal Hornillos, R. y Villanueva Roa, J. D. D. (2015). La música, un instrumento en la enseñanza del español como lengua extranjera mediante la aplicación de las nuevas tecnologías. *Porta Linguarum* (Universidad de Granada), 23, 139-151.
- Carranza, M. (2008). Fenómenos de interferencia fónica relacionados con la vocal /u/ en la interlengua de estudiantes japoneses de español como lengua extranjera. *Estudios Lingüísticos Hispánicos* (Círculo de Estudios Lingüísticos Hispánicos de Tokio), 23, 1-22.
- Carranza, M. (2012). Errores y dificultades específicas en la adquisición de la pronunciación del español LE por hablantes de japonés y propuestas de corrección. Recogido en GIDE (Grupo de Investigación de la Didáctica del Español), *Nuevos enfoques en la enseñanza del español en Japón – Concha Moreno y GIDE* -, Editorial Asahi, 51-78.
- Fernández Lázaro, G.; Fernández Alonso, M.; Kimura, T. (2017). Corrección de errores de pronunciación para estudiantes japoneses de español como lengua extranjera. *Cuadernos CANELA* (Confederación Académica Nipona, Española y Latinoamericana) 27, 65-86.
- Graham, C. (1978). *Jazz chants: rhythms of American English for students of English as a second language*. New York: Oxford University Press.
- Hamamatsu, N., Onoda, R. (2014). Percepciones estudiantiles sobre el aprendizaje del español como lengua extranjera: Estudio particular sobre las actitudes y motivaciones de un grupo de universitarios japoneses respect [sic] examen del diploma del español. *Takushoku Language Studies* (Universidad Takushoku), 130, 149-174.
- Hara, M. (1990). Método de enseñanza de la pronunciación española a los alumnos japoneses. *Actas del segundo congreso nacional de ASELE*, 371-379.
- Hattori, Yoichi (2010). Supein kakyoku shidoho: Castillago no diction to enso kaishaku no kiso toshite ryuisubeki tokuchouteki sokumen [en japonés]. *Departmental Bulletin Paper* (Tokyo College of Music), 73-95. (descargable de <http://id.nii.ac.jp/1300/00000881/>) [último acceso: 13-02-2020]
- Kimura, T. (2018). Evaluación subjetiva de la pronunciación con "acento japonés" segmental por hablantes de español L1. *Estudios sobre enseñanza de lenguas* (Universidad Seisen), 10, 17-30. (descargable de <http://doi.org/10.24743/00001086>) [último acceso: 13-02-2020]
- Kimura, T. (2019). *¡Así suena!* Tokio: Editorial Asahi.
- Murphy, T (1992). *Music & Song (Resource books for Teachers)*, Oxford: Oxford University Press.
- Rodríguez López, B. (2006). Las canciones en la clase de español como lengua extranjera. *La competencia pragmática y la enseñanza del español como lengua extranjera. Actas del XVI Congreso Internacional de ASELE*, 806-816.
- Sensui, H. (2014). A Case Study on Some Prosodic Characteristics of Japanese Students of Spanish [en japonés]. *Academia. Literature and language* (Universidad Nanzan), 96, 21-55.
- Sensui, H. (2017). Pronunciation of Unstressed Words in Spanish by Japanese Learners [en japonés]. *Academia. Literature and language* (Universidad Nanzan), 102, 41-67. (descargable de <http://doi.org/10.15119/00001194>) [último acceso: 11-02-2020]
- Takasawa, M. (2019). On Syllables in Spanish Songs. [en japonés] *Journal of the Society for General Academic and Cultural Research* (Universidad Asia), 36, 39-52.
- Ura, M. (2010). A Study of Spanish Language Class for Increasing Motivation for Learning. [en japonés] *Journal of the Faculty of International Studies Bunkyo University*, 20(2), 17-35.

* Este estudio se ha llevado a cabo gracias a JSPS KAKENHI Grant Number JP19K0086 y al fondo para investigaciones académicas Pache I-A-2 para el año académico 2018 y 2019 de la Universidad Nanzan. / Funding for this study was provided by JSPS KAKENHI Grant Number JP19K0086 and Nanzan University Pache Research Subsidy I-A-2 for the 2018 and 2019 academic year.

ANÁLISIS COMPARATIVO DEL F0 RANGO Y DE LA FLUIDEZ ORAL ENTRE LOS HABLANTES CHINOS Y NATIVOS DEL ESPAÑOL

Peizhu Shang

Universitat de Barcelona
pshangsh7@alumnes.ub.edu

ABSTRACT

Este trabajo presenta una comparación sistemática de varias medidas del rango tonal y de la fluidez oral en los hablantes chinos y nativos del español peninsular. Los datos de F0 rango han sido analizado en dos dimensiones mediante el método LTD (*Long-term distributional*): nivel tonal y rango tonal, mientras la fluidez oral se cuantifica por medio de diferentes parámetros temporales. El objetivo de este estudio es examinar si (o cómo) los hablantes chinos y españoles difieren en la variación tonal y en la expresión oral de las preguntas. Para tal fin, se ha usado un corpus de habla inducida mediante contextos de un total de 37 informantes. Los resultados indican que, en comparación con los hablantes nativos de español, los aprendices chinos tienden a producir un F0 mínimo de nivel más alto, un rango tonal reducido y unas pausas de duración más larga en la L2.

Palabras clave: nivel tonal, rango tonal, fluidez oral, transferencia prosódica

This paper presents a systematic comparison of various measures of pitch range and speech fluency between Chinese and native speakers of Peninsular Spanish. The data of F0 range was analyzed along two dimensions by using the LTD methods (*Long-term distributional*): pitch level and pitch span, whereas speech fluency was quantified by various temporal features. The aim of this work is to examine whether (or how) Chinese and Spanish speakers differ in the tonal variation and oral fluency in questions. To this end, a corpus of 37 informants was analyzed using the *Discourse completion task*. Results indicate that compared to native Spanish speakers, Chinese learners tend to produce higher minimum F0, reduced pitch span and greater pause length in their L2.

Keywords: pitch level, pitch span, oral fluency, prosodic transfer

1. INTRODUCCIÓN

Uno de los objetivos en los estudios de aprendizaje de segundas lenguas consiste en describir y aclarar cómo el sistema fonético y fonológico de la primera lengua (L1) afecta el habla de la segunda (L2). Estudios previos han evidenciado que las lenguas humanas pueden diferir en el uso e interpretación del registro, el rango tonal y los patrones entonativos dado que la percepción de los eventos tonales tiene rasgos universales y otros específicos de cada lengua. El trabajo de Shi, Zhang y Xie (2014) ha comprobado, por ejemplo, que existe una diferencia del registro y del rango tonal entre los hablantes nativos del chino y del japonés. Este resultado ha sido apoyado por los datos propuestos por Keating y Kuo (2012), Mennen et al. (2012), Zimmerer et al. (2014) quienes han encontrado diferencias de nivel

tonal y de rango entre lenguas como el chino, el inglés, el francés, y el alemán.

Dado que la adquisición de la L2 suele implicar cierto grado de interacción, las diferencias entre la lengua materna y la lengua estudiada en lo referente al registro y el rango pueden aparecer también en el habla de la L2, como resultado de la transferencia prosódica. En general, según los estudios de adquisición, los hablantes de L2 tienden a usar un registro más alto y un rango tonal más comprimido que los hablantes nativos (Backman, 1979 para aprendices españoles del inglés; Busà y Urbani, 2011 para aprendices italianos del inglés; Shi et al., 2014 para aprendices chinos del japonés, entre otros).

Sin embargo, en el estudio de Ding et al. (2016), no se han encontrado diferencias del rango tonal entre los aprendices chinos y los nativos del alemán. Una posible explicación a la disparidad de resultados sería que el rango en este último trabajo está

cuantificado como la diferencia absoluta entre el F0 máximo y el F0 mínimo de la oración, lo cual se ve muy afectado por los valores extremos de cada individuo. Para identificar mejor las diferencias entre lenguas, en este estudio se va a usar el método LTD (*Long-term distributional*) desarrollado por Mennen et al. (2012), en el que para el rango tonal, además de la diferencia entre máximo y mínimo de F0, el 80% rango, la desviación estándar y la curtosis también han sido analizado.

Aparte de las diferencias tonales, la adquisición de la L2 se caracteriza a menudo por una reducción de la fluidez. La disminución de la fluidez oral suele relacionarse con parámetros temporales como el número y la duración total de pausa por minuto y la duración media de todas las pausas en el habla. Las medidas de fluidez son especialmente interesantes porque se relacionan directamente con el nivel lingüístico de los hablantes de L2 (Kahng, 2014).

El objetivo de este estudio, por lo tanto, es explorar las tres propiedades discutidas antes sobre la adquisición del lenguaje. En concreto, se pretende investigar si los aprendices chinos del español tienen un nivel tonal más alto y un rango más reducido en comparación con los hablantes nativos. Además, se examina la fluidez oral en la L2 en función del nivel de competencia lingüística de los informantes chinos.

2. METODOLOGÍA

2.1. Participantes

Los participantes del presente estudio son cinco hablantes nativos de español peninsular y 32 aprendices chinos con el español peninsular como lengua extranjera. La edad de todos los sujetos estaba comprendida entre los 18 y los 31 años, con la edad media de 23.97 años en el momento en el que se realizaron las grabaciones (v. Tabla 1). Ningún individuo en la investigación ha informado de ningún trastorno de habla, de audición, o de la voz.

Los informantes estaban divididos en tres grupos lingüísticos según la competencia del español: intermedio, avanzado y nativo. Todos los hablantes chinos eran estudiantes universitarios en Barcelona, con el dialecto peninsular del español como la variedad a la que se había expuesto predominantemente y como el objetivo de su L2.

Tabla 1: Datos de los tres grupos lingüísticos.

	Chino		Español	Total	
	Intermedio	Avanzado	Nativo	F	M
Sexo	F	M	F	F	M
Número	10	6	16	5	6
Edad media	23.44	24.75	23.2	23.97	
SD	2.34	2.62	4.87	2.87	

2.2. Corpus y grabación

Para garantizar la naturalidad del habla, se ha recurrido a la técnica de *Discourse Completion Task* (Billmyer y Varghese, 2000; Félix-Brasdefer, 2010) para elicitar las frases. En total, se ha presentado 15 situaciones conversacionales a los informantes, repartidas por cinco tipos de pregunta en español: pregunta total ('Tt'), pregunta parcial ('Pc'), pregunta disyuntiva ('Dy'), pregunta confirmatoria débil ('Cd') y pregunta confirmatoria con apéndice ('Ca'). Todas las grabaciones se realizaron en una cabina anecoica. Los archivos de sonido se digitalizaron en PCM con una frecuencia de muestreo de 44.1 kHz.

2.3. Extracción de los datos prosódicos

El análisis acústico de F0 se realizó en *Praat* (Boersma y Weenink, 2019) usando el método de autocorrelación, con un valor mínimo de F0 (*pitch floor*) de 75Hz y un valor máximo (*pitch ceiling*) de 500 Hz. Como la extracción automática de F0 fue insatisfactoria cuando hubieron glotalizaciones o habla crepitante (*creaky voice*) en la grabación, se ha llevado a cabo después una corrección manual para mejorar la exactitud de F0. En dicha corrección se ensordecían los fragmentos de F0 con errores a partir del objeto *Pitch*.

Mientras la fluidez oral ha sido evaluada mediante los tres parámetros temporales: el número de pausa por minuto, la duración total de pausa por minuto, y la duración media de pausa, las variaciones tonales se han analizado en dos dimensiones a través de una serie de medidas basadas en la distribución de F0. En concreto, se han calculado tres parámetros para el nivel tonal y cuatro parámetros para el rango:

- Nivel tonal: *F0 max*, *F0 min*, *F0 media*;
- Rango tonal: *100% rango* ($F0_{max}-F0_{min}$), *80% rango*, *desviación estándar*, *curtosis*.

Los valores de F0 en Hz se han convertido en la unidad de St e ERB para normalizar la frecuencia entre los españoles y chinos, los femeninos y masculinos. La transformación a St e ERB se aplicó exclusivamente a las medidas del rango, dado que la escala de St está diseñada especialmente para la diferencia de frecuencias y la transformación en ERB

es monotóna para el nivel tonal (Mennen et al., 2012).

3. ANÁLISIS ESTADÍSTICO

En este estudio, el análisis estadístico de los datos se ha realizado con el software R (R Core Team, 2016). En concreto, se ha escogido ejecutar modelos de efectos mixtos mediante el paquete *lmerTest* (Kuznetsova, Brockhoff y Christensen, 2017) para evaluar el efecto de cada variable. Como efectos fijos, se han incluido los factores de Grupo lingüístico, Sexo, y Tipo de pregunta. Como efectos aleatorios, el Ítem y el Hablante han sido incorporado con todas las posibles intercepciones aleatorias y pendientes aleatorias. Además, se ha utilizado la función de *Anova* (Faraway, 2002) para comparar en R los modelos ajustados. Los valores significativos de *p* se han obtenido por pruebas de razón de verosimilitud testando el modelo inicial contra el modelo sin el efecto de Grupo lingüístico, Sexo y Tipo de pregunta.

4. RESULTADOS

4.1. Resultados para el nivel tonal

Para el F0 máximo y el F0 media, no se han encontrado efectos principales significativos entre los tres grupos, aunque los hablantes chinos tenían un F0 media de nivel más alto que los españoles. Para el F0 mínimo, el modelo ha detectado diferencias significativas en función del grupo lingüístico ($\chi^2=37.105, p<0.0001$). Concretamente, el resultado indica que los dos grupos chinos tienden a producir un F0 mínimo más alto que los nativos (v. Figura 1). Sin embargo, el análisis post-hoc muestra que esa diferencia fue significativa solamente entre el grupo intermedio y el nativo ($\chi^2=4.2402, p<0.05$).

Para el factor Sexo, se han encontrado efectos significativos tanto del F0 media como del F0 máximo y mínimo (media: $\chi^2=47.499, p<0.001$; F0 máximo: $\chi^2=26.649, p<0.001$; F0 mínimo: $\chi^2=37.105, p<0.001$). Como es previsible, las mujeres de nuestro estudio tienden a usar una distribución de F0 más alta que los hombres (v. figura 1).

Respecto al Tipo de pregunta, diferencias significativas de F0 máximo, F0 mínimo y F0 media han sido detectado tanto para los aprendices chinos (máximo: $\chi^2=55.57, p<0.0001$; mínimo: $\chi^2=134.38, p<0.0001$; media: $\chi^2=74.457, p<0.0001$) como para los nativos españoles (máximo: $\chi^2=48.447,$

$p<0.0001$; mínimo: $\chi^2=11.262, p<0.05$; media: $\chi^2=45.898, p<0.0001$).

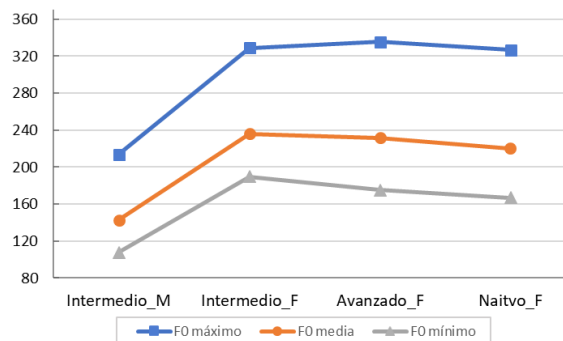


Figura 1: F0 máximo, F0 media, y F0 mínimo (Hz) producido por los informantes femeninos y masculinos de los tres grupos lingüísticos.

4.2. Resultados para el rango tonal

Entre las cuatro medidas del rango, solo el 80% rango calculado en St ha sido estadísticamente significativo en los tres grupos ($\chi^2=6.3734, p<0.05$). Más adelante, el análisis post-hoc demuestra que el grupo intermedio chino tenía un 80% rango tonal significativamente más reducido que el grupo nativo ($p<0.05$ en Hz, St y ERB) y el avanzado ($p<0.05$ en Hz, St y ERB). Sin embargo, no se han encontrado diferencias significativas de desviación estándar y de curtosis en función del grupo lingüístico.

Por otra parte, los resultados del Sexo indican que las mujeres tienden a usar un rango tonal más expandido que los hombres. No obstante, el efecto estadísticamente fue insignificante, probablemente porque la muestra tomada en este estudio no está equilibrada entre hombres y mujeres. En cambio, se han encontrado diferencias significativas de 100% y 80% rango en función del tipo de pregunta ($p<0.0001$ para el 100% y el 80% rango medido Hz, St y ERB). Además, el factor Tipo de pregunta tenía también un efecto significativo para las variables dependientes de desviación estándar ($p<0.001$ en Hz, St y ERB) y curtosis (Hz: $\chi^2=45.787, p<0.0001$).

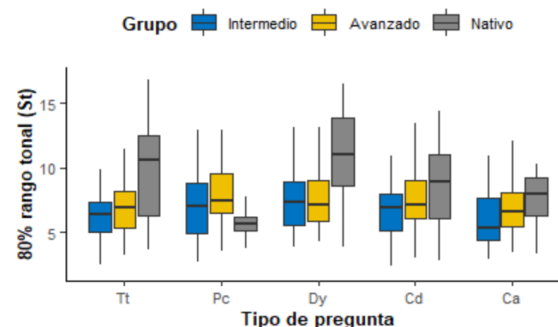


Figura 2: Boxplot del 80% rango de los cinco tipos de pregunta producidos por los tres grupos lingüísticos.

4.3. Resultados para la fluidez oral

Los datos en la tabla 2 muestran que los aprendices chinos hicieron pausas más frecuentemente y más largas en comparación con los hablantes nativos, resultando en que el habla de L2 queda dividido en muchos fragmentos que son menos inteligibles y esto disminuye la fluidez del habla.

Tabla 2: Estadística del número y la duración de pausa por minuto y la duración media de pausa de tres grupos.

Grupo	Nº_P/min	Duración_P/min	Duración media_P (SD)
Intermedio	49.67	3.2319s	0.0616s (±0.0307)
Avanzado	53.92	3.2667s	0.0573s (±0.0236)
Nativo	41.40	1.8936s	0.0449s (±0.0169)

Para la duración media de pausa, el modelo ha encontrado diferencias significativas entre los tres grupos ($\chi^2=7.5047, p<0.05$), así como entre los cinco tipos de pregunta ($\chi^2=95.409, p<0.0001$). Más adelante, las comparaciones múltiples indican que de estos tres grupos el nativo era el más distintivo: específicamente, la duración media de pausa de los españoles era significativamente más corta que en el grupo intermedio y el avanzado (v. Tabla 3). Además, se encuentra que las preguntas que se han realizado con pausas más largas son aquellas que tienen una construcción sintáctica más complicada, por ejemplo, preguntas disyuntivas y preguntas de apéndice (v. Figura 3). Respecto al factor Sexo, los hombres y mujeres no muestran diferencias significativas en la variable dependiente de duración media de pausa ($\chi^2=2.2142, p=0.137$).

Tabla 3: Comparaciones múltiples de la duración media de pausa (s) en los tres grupos lingüísticos (* $p<0.05$).

Contraste	Estimate	Std.Error	z value	Pr(> z)
Intermedio-Avanzado	0.001137	0.003856	0.295	0.7682
Nativo-Avanzado	-0.01382	0.005189	-2.663	0.0198 *
Nativo-Intermedio	-0.01496	0.005508	-2.716	0.0198*

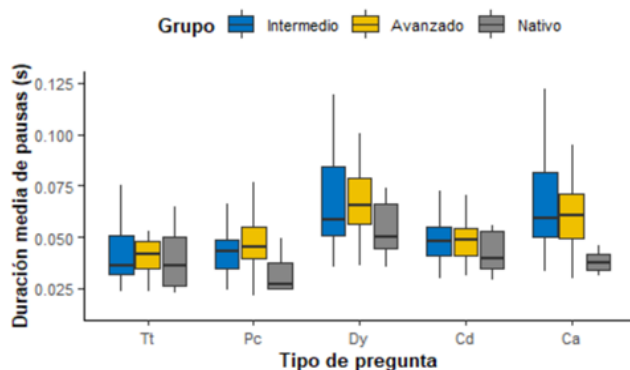


Figura 3: Boxplot de la duración media de pausa de los cinco tipos de pregunta producida por los tres grupos.

5. CONCLUSIONES

El objetivo de este estudio es comparar el nivel tonal, el rango, y la fluidez oral entre los aprendices chinos de diferentes niveles y los nativos españoles.

Los resultados obtenidos de F0 han revelado diferencias en dos aspectos. Por un lado, el rango comprimido encontrado para los hablantes chinos de L2 español es coherente con los estudios previos sobre la adquisición de segundas y terceras lenguas. Por otro, en vez de un nivel tonal más alto registrado para otros pares de lengua, los aprendices chinos en este estudio, han usado un F0 media similar al de los nativos españoles. Sin embargo, han presentado un valor significativamente más alto que los nativos del español en su F0 mínimo, lo cual se podría explicar por una transferencia negativa de la L1. Esta transferencia sería debida a la falta de un tono bajo constante (monotonal) en el inventario tonal del chino. Esta diferencia haría que los hablantes chinos asimilaran el tono monotonal bajo del español al único tono monotonal de su inventario, el tono 1, producido mediante un tono alto. El uso de un tono alto en lugar del bajo del español da como resultado, una distribución de F0 mínimo más alto en la lengua meta. Este error también es frecuente cuando el chino está en contacto con otras lenguas como el inglés y el alemán (Barto, 2015; Ding et al., 2012; Ji, Wang y Li, 2009; Yuan et al., 2019).

Las diferencias encontradas para el nivel tonal y el rango entre hombres y mujeres coinciden con el resultado de múltiples estudios previos (Martina, 2012; Shi et al., 2014). Estas se podrían explicar además de por las diferencias anatómicas, por cuestiones relacionadas con el comportamiento sociocultural (Van Bezooijen, 1995). Además, para los cinco tipos de pregunta, los hablantes chinos y españoles han empleado diferentes niveles tonales y distintos rangos para la distinción prosódica.

En cuanto a los resultados de fluidez oral, estos indican que los hablantes de L2 tienden a hacer pausas más largas y más frecuentes en comparación con los nativos. Este fenómeno lingüístico se puede atribuir, según Peters (2019), a un mayor esfuerzo cognitivo al hablar la L2. Es decir, los hablantes de L2 requieren más tiempo para la planificación verbal debido a la falta de competencia lingüística. Esta hipótesis también va en la línea de nuestro estudio ya que los datos muestran que la fluidez de los sinohablantes progresa a medida que el nivel de competencia lingüística en la L2 aumenta.

5. BIBLIOGRAFÍA

- Barto, K. A. (2015). *Mandarin Speakers' Intonation in their L2 English*. The University of Arizona.
- Billmyer, K., & Varghese, M. (2000). Investigating instrument-based pragmatic variability: Effects of enhancing discourse completion tests. *Applied Linguistics*, 21(4), 517–552.
- Boersma, P., & Weenink, D. (2019). Praat: Doing phonetics by computer [Computer program]. Retrieved September 2019 from www.Praat.Org/.
- Ding, H., Jokisch, O., & Hoffmann, R. (2012). A phonetic investigation of intonational foreign accent in Mandarin Chinese learners of German. In *Speech Prosody 2012*.
- Faraway, J. J. (2002). Practical regression and ANOVA using R. University of Bath Bath.
- Félix-Brasdefer, J. C. (2010). Data collection methods in speech act performance. *Speech Act Performance: Theoretical, Empirical and Methodological Issues*, 26, 41.
- Ji, X., Wang, X., & Li, A. (2009). Intonation patterns of yes-no questions for Chinese EFL learners. In *2009 Oriental COCOSDA International Conference on Speech Database and Assessments* (pp. 88–93). IEEE.
- Kahng, J. (2014). Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning*, 64(4), 809–854.
- Keating, P., & Kuo, G. (2012). Comparison of speaking fundamental frequency in English and Mandarin. *The Journal of the Acoustical Society of America*, 132(2), 1050–1060.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, 82(13).
- Martina, U. (2012). Pitch range in L1/L2 English. An analysis of f0 using LTD and linguistic measures. *Maria G. Bus Ya & Antonio Stella (Eds.), Methodological Perspectives on L2 Prosody: Papers from ML2P*, 79–83.
- Mennen, I., Schaeffler, F., & Docherty, G. (2012). Cross-language differences in fundamental frequency range: A comparison of English and German. *The Journal of the Acoustical Society of America*, 131(3), 2249–2260.
- Peters, J. (2019). Fluency and fundamental frequency in biligual speakers of high and low german. In *Proceedings of the 19th International Congress of Phonetic Sciences*. Melbourne, Australia.
- Shi, S., Zhang, J., & Xie, Y. (2014). Cross-language comparison of F0 range in speakers of native Chinese, native Japanese and Chinese L2 of Japanese: Preliminary results of a corpus-based analysis. In *The 9th International Symposium on Chinese Spoken Language Processing* (pp. 241–244). IEEE.
- Team, R. C. (2016). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria.
- Van Bezooijen, R. (1995). Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and Speech*, 38(3), 253–265.
- Yuan, C., González-Fuente, S., Baills, F., & Prieto, P. (2019). Observing pitch gestures favors the learning of Spanish intonation by Mandarin speakers. *Studies in Second Language Acquisition*, 41(1), 5–32.
- Zimmerer, F., Jügler, J., Andreeva, B., Möbius, B., & Trouvain, J. (2014). Too cautious to vary more? A comparison of pitch variation in native and non-native productions of French and German speakers.

PROSODIC ORGANIZATION OF ENGLISH FOLK RIDDLES AND THEIR EMOTIONAL AND PRAGMATIC POTENTIALS

Larysa Taranenko

National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”
larysataranenko@gmail.com

ABSTRACT

The paper outlines the results of an interdisciplinary experimental phonetic research aimed at defining emotional and pragmatic potentials of spoken English folk riddles, reflected in prosodic manifestation of the structural elements of their texts. The auditory and acoustic analyses were performed following the principal ideas and methodological guidelines of speech energetics theory and by applying a quantitative *K*-criterion as a new method of phonetic research. The obtained quantitative and qualitative data allowed the author to form the energy-grams which graphically represent the changes of the riddle emotional and pragmatic potentials throughout its plot unfolding. The carried out research revealed that a lapidary structure of the riddle is compensated by a set of prosodic means and their specific interaction, which being the result of the riddle emotional and pragmatic potentials variations, trigger creative and cognitive processes in the recipient’s mind aimed at searching for the riddle solution.

Keywords: English folk riddle, prosodic means, emotional and pragmatic potentials, experimental phonetic research

1. INTRODUCTION

The present-day development of phonetic research, performed at the intersection of cognitive, sociocultural and psychological studies, calls for new insights into the experimental research of oral speech with the use of methods of its quantitative analysis.

In view of this, the aim of the present paper is to experimentally verify the energetic potential of spoken English riddles, reflected in prosodic manifestation of emotional and pragmatic potentials inherent in their texts.

2. METHODOLOGICAL STIPULATIONS OF THE RESEARCH

The riddle serves as a material of our study since it consists of a minimal composition (“the topic”, or description of the object to be decoded, and “the commentary” (Taranenko, 2013)) and is by its nature a metaphoric world representation to be interpreted by the recipient. We believe that such a concise structure of the riddle text is to be compensated by prosodic means’ interaction, evoking in the

recipient’s mind creative and cognitive processes aimed at searching for the riddle answer.

Thus, in the course of analysis we registered typical prosodic patterns common for the outlined structural elements of the riddle. The prosodic patterns under study included length of pauses at the block juncture, distribution of utterance stress, types of terminal tone and head, pitch and voice range fluctuations within each compositional element. In order to specify the prosodic prominence of the riddle attention was also given to its loudness, tempo as well as voice quality modifications.

The research was carried out following the maxims of speech energetics theory put forward by Alla Kalyta (Kalyta, 2016) and by applying a new method of phonetic research known as a quantitative *K*-criterion (Kalyta & Taranenko, 2012, pp. 476-484). This criterion was calculated using the formula:

$$K = \frac{F_0 \times t \times I_0}{1000 \times I_3}$$

Legend: *K* – criterion of the level of the utterance emotional-and-pragmatic potential; *F*₀ – fundamental frequency (Hz); *t* – syllable duration

(ms.); I_0 – amplitude of F_0 (dB); I_3 – amplitude of F_3 (dB); 1000 – milliseconds to seconds conversion factor.

The experimentally defined values of the quantitative K -criterion were differentiated into: low (up to 30 points), medium ($K = 30-105$ points) and high (higher than >105 points) levels of the utterance emotional-and-pragmatic potential.

The use of this criterion makes it possible to define the riddles' emotional-and-pragmatic potential as well as graphically present its fluctuations in the form of the so-called energy-grams. These energy-grams being made up of three constituents (intono-, emotio- and pragma-grams) illustrate the changes of the text energetic potential throughout its plot unfolding.

Considering the outlined ideas, we completed an experimental phonetic study of 227 English folk riddles (Bryant, 2007, pp. 92-109) recited by two male speakers to the audience of schoolchildren. The total duration of the texts' oral actualization amounts to 42 minutes. The auditory and acoustic analyses were performed with a view to defining a correlation between prosodic organization of the riddle structural elements ("the topic" and "the commentary") and the level of their emotional and pragmatic potentials, being capable of activating the recipient's cognitive-and-creative processes while decoding the riddle solution.

2. RESULTS

2.1. Auditory analysis of emotional and pragmatic potentials of the structural element "the topic"

General assessment of *emotional and pragmatic potentials* of the riddle oral actualization allows us to qualify them as those of low or mid levels. It can be substantiated by the fact that while pronouncing the riddle out loud to its recipient, the speaker tries to conceal his/her emotional state and evaluative attitude to the object described, hiding his/her understanding of the riddle metaphorical meaning. At the prosodic level, the structural element "topic", whose content indirectly points to the object described in the riddle is characterized by the prevalence of neutral prosodic parameters:

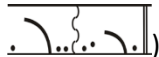
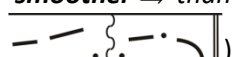
- moderate loudness and tempo,
- mid voice range,
- a rising kinetic tone in the non-terminal syntagm and a falling tone in the final one,

- wavy pitch contour (e.g.: *Lives with^oout a /body, | /hears with^oout /ears...* (Bryant, 2007, p. 102).

In case of enumeration of several features of the object described in the riddle, there is the alternation of different pitch levels of high-rising tones, e.g.: *It was /neither /fish, ⇒ /flesh, ⇒ nor /bone...* (ibid., 97). This type of intonation pattern is typical of emotionally neutral speech and serves as a marker of the low level of emotional potential of the riddle structural element "the topic".

It was also found out that the combination of two low-falling terminal tones in the adjoining intonation groups (e.g.: *East and /west ⇒ and /north and /south* (ibid., p. 92) is perceived as an indicator of a gradual increase in the emotional potential of "the topic" up to a low zone of its mid level.

The following prosodic means and their combinations signify a slight increase in the emotional potential of "the topic":

- the broken descending stepping head (e.g.: *When /I was /young and /beautiful | /wore a ↑blue /crown...* (ibid., 105),
- a high-falling terminal tone of a wide range and convex configuration within the riddle initial syntagm (e.g.: *My /fatherland ⇒ is A/rabia...* (ibid., p. 92) ,
- a falling tone having an ascending movement on the word stressed element and a low level tone of the unstressed tail (e.g.: *Runs /smoother ⇒ than /any /rhyme...* (ibid., p. 94) .

The outlined prosodic means, typical of the structural element "the topic", allow the listener's attention to be focused on those characteristics of a described object, which will contrast to those, given in the structural element "the commentary" and will serve for the creation of a metaphorical image in the listener's mind, necessary to find out the riddle solution.

2.2. Auditory analysis of emotional and pragmatic potentials of the structural element "the commentary"

Prosodic organization of "the commentary" differs from that of "the topic" since it contains the rheme of a riddle and, consequently, is marked by emphatic prosodic patterns whose role is to make the rheme prominent. Besides, the use of emphatic prosodic parameters is conditioned by the function of "the

commentary” to highlight specific features of the described object or give additional information about it. In other words, prosodic organization of “the commentary” serves to intensify those lexical units which are to excite in the recipient’s psychic sphere definite images and associations as well as trigger the search for analogies in the process of riddle decoding. All these factors explain the increase of the riddle emotional potential to its mid level.

The typical features of “the commentary” prosodic organization include:

- the combination of two falling tones having different pitch within one and same intonation group (e.g.: ...and \wasn’t a \beast... (ibid., p. 97); And I’m the \torment of \man (ibid., p. 106), the higher tone serving to highlight a specific feature of a described object, marked by the low falling tone,
- the use of emphatic sliding and stepping broken heads (e.g.: [^]Couldn’t \put [^]Humpty \Dumpty ⇒ to[^]gether a\gain || (ibid., p. 104), [^]One [^]hundred \teeth | and [^]never a \mouth || (ibid., p. 104), whose function in riddles, unlike other narrative folk texts, is to intensify the key lexical units which are to launch the image search in the recipient’s psychic sphere,
- prevalence of high falling tones in adjoining intonation groups, the latter having a slower rate of its movement (e.g.: [^]Love to \fall ⇒ but [^]cannot \climb || (ibid., p. 94),
- a slowed down tempo of the final syntagm and occurrence of short pauses (≡) that emphasize the rhematic element of a riddle (e.g.: [^]Lives in \winter, ⇒ [^]dies in \summer, | and [^]grows with its \roots ⇒ \upwards || (ibid., p. 99),
- highlighting of the riddle rhematic element by a preceding pitch interval (e.g.: ...and [^]when he *finds \water, | [^]he \perishes || (ibid., p. 102),
- the “enigmatic” voice coloring of the final intonation group of “the commentary”, qualified as a pragmatically most important component of the text of a riddle.

The analysis of “the commentary” prosodic organization gives us every reason to conclude that the riddle’s *pragmatic potential* is constantly growing, reaching its maximal mid-high level within the final intonation group.

2.3. The energy-gram of the riddle oral realization

In the course of acoustic analysis of riddle’s prosodic patterns we have calculated the quantitative *K*-criterion that defines the dynamics of the changes of the riddle emotional and pragmatic potentials within its structural elements. The values of the *K*-criterion have been calculated by the mentioned above formula by way of employing the computer-based sound processing software *Praat*.

Using the experimentally obtained data we have built a typical energy-gram of the English riddle oral actualization presented on figure 1.

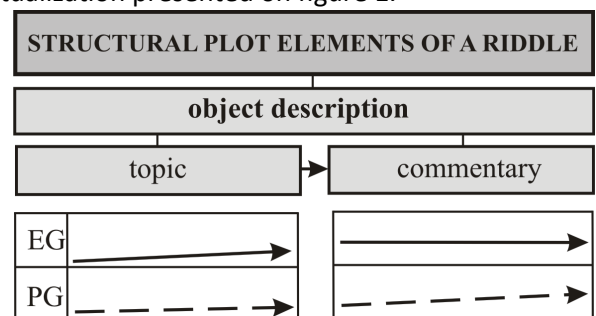


Figure 1: A typical energy-gram of the English riddle oral actualization (EG stands for emotio-gram, while PG – pragma-gram).

The figure graphically illustrates the changes of emotional and pragmatic potentials of the riddle oral actualization. It shows that the riddle emotional potential gradually increases from its low to mid levels within the structural element “the topic”, having the *K*-criterion fluctuations from $K = 22$ to $K = 32$. Within “the commentary” the emotional potential is characterized by a low zone of its mid level with the average value of $K = 56.5$. At the same time, the pragmatic potential of the riddle oral realization steadily increases from a low level in “the topic” to a high zone of the mid level within a rhyme of “the commentary”. We can state that the lower level of the riddle emotional potential as compared to the level of its pragmatic potential can be explained by the speaker’s desire to conceal his/her understanding of the riddle metaphorical meaning while uttering the riddle to its recipient.

3. DISCUSSION

The auditory and acoustic analyses showed that the emotional potential of the riddle either remains unchangeable at its low level, or gradually increases to the mid level in the direction from “the topic” to “the commentary” of the riddle text. Such a low emotional potential of the riddle is justified by the fact that its increase can bring additional meanings

to the riddle content, confusing the recipient or misleading him/her in search for the right answer to the riddle.

At the same time, since the riddles belong to a folklore genre with a creative-and-instructive pragmatic orientation, they are characterized by the higher level of pragmatic potential as compared to their emotional one.

It was experimentally defined that the riddle pragmatic potential tends to gradually increase, reaching its peak on the rheme within the last intonation group of "the commentary". The steady rise in the riddle pragmatic potential from low to mid levels is related to entertaining and training functional role of the riddle, which presupposes to get the recipient involved into the search for a riddle answer by means of launching his/her associative-and-creative mechanism that allows reconsidering of the known information.

The analysis also revealed the inverse relation between the level of the riddle emotional potential and the size of its text: the more lapidary the riddle content is, the higher is the level of its emotional potential, and vice versa – the increase of the text size leads to lowering of its emotional potential and growth of the pragmatic one.

4. CONCLUSIONS

Prosodic organization of the riddle serves to realize its pragmatic potential by emphasizing or contrasting the lexical units that affect the recipient's subconscious sphere, stimulating him/her to search for images, associations and analogies of the riddle solution.

Interaction of prosodic parameters outlined in the paper can be viewed as subliminal means of involving the listener into the process of riddle decoding.

5. BIBLIOGRAPHY

- Bryant, M. (2007). *Riddles Ancient and Modern*. London: Hutchinson.
- Kalyta, A. (2016). *Speech energetics*. Kyiv: Kafedra publishing house.
- Kalyta, A. & Taranenko L. (2012). Kriterij urovnja aktualizacii jemocional'no-pragmaticheskogo potenciala vyskazyvanija [The criterion of the level of the utterance emotional-and-pragmatic potential]. *Scientific notes*. Kirovograd: RVV KDPU im. V. Vinnichenka, 105(1), 476-484. (in Russian)
- Taranenko, L. (2013). Fabul'na structura folklornoyi zagadky ta kognityvno-kreativnyi mekhanizm yiyi decoduvan'y [The plot structure of the folk riddle and

the cognitive-and-creative mechanism of its decoding]. *Science and Education a New Dimension: Philology*, Budapest, 1(3), 167-174. (in Ukrainian).

SENSIBILIDAD PROSÓDICA ACENTUAL EN NIÑOS CON TRASTORNO DEL DESARROLLO DEL LENGUAJE DE SEGUNDO GRADO DE PRIMARIA

Valeska Torres Bustos¹, Bernardo Riffo Ocares¹, Katia Sáez Carrillo²

¹Departamento de Español, Universidad de Concepción, Chile;

²Departamento de Estadística, Universidad de Concepción, Chile
valtorres@udec.cl

ABSTRACT

La presente investigación buscó determinar la Sensibilidad prosódica (SP) acentual en niños con Trastorno del desarrollo del lenguaje (TDL) que cursan segundo grado de primaria. Se efectuó un estudio cuantitativo descriptivo en el que participaron 76 niños chilenos, 50 con un desarrollo normotípico del lenguaje (DNL) (\bar{X} = 7,74 años de edad; D.E.= 0,35) y 26 con TDL (\bar{X} = 7,69 años de edad; D.E.= 0,45). Se les administró el “*Test de acento léxico*” y el “*Test de nombres compuestos*” (ambos instrumentos adaptados lingüísticamente al español chileno). Los resultados muestran que en el test acento léxico los niños con DNL presentaron un desempeño superior en el reconocimiento de las palabras proparoxítonas y el en tiempo de las respuestas correctas de las palabras paroxítonas (p -valor < 0,05). En el análisis por ítem de ambas pruebas se observó que en algunos estímulos los niños con TDL presentan un mejor rendimiento que los con DNL.

Palabras clave: sensibilidad prosódica, acento léxico, acento métrico

The aim of this study was to determine the prosodic stress sensitivity of second graders with developmental language disorder (DLD). A quantitative descriptive study was carried out in which 76 Chilean children participated, 50 of them had a normal language development (\bar{X} = 7,74 years, SD= 0,35) and 26 had DLD (\bar{X} = 7,69 years, SD= 0,45). Both lexical stress and compound nouns tests (linguistically adapted to Chilean Spanish) were administered. The results reveal that, in the lexical stress test, children with normal language development had a better performance recognizing proparoxytone words and in correct response time of paroxytone words (p -value < 0,05). The analysis per items of both tests showed that, in response to some stimuli, children with DLD perform better than those with a normal language development.

Keywords: prosodic sensitivity, lexical stress, metric stress

1. INTRODUCCIÓN

La sensibilidad prosódica (SP) o también llamada conciencia prosódica (CP), se define como la habilidad de percibir, interpretar y manipular los cambios de duración, intensidad y tono (A Holliman et al., 2017).

Esta habilidad ayuda a segmentar la señal oral continua en unidades menores discretas (palabras, por ejemplo), lo que favorece la adquisición de las palabras léxicas (Shi & Werker, 2003) y la comprensión (Calet et al., 2016). Al igual que la conciencia fonológica (CF), la SP no solo está vinculada a la adquisición del lenguaje sino que también se relaciona con el aprendizaje de la lectura y la escritura tanto en lenguas con ortografía

transparente como no transparentes (Jordán et al., 2019).

Un tipo de SP es la acentual, esta a su vez se clasifica en SP acentual léxica y SP acentual métrica. La primera, consiste en reconocer el acento léxico de las palabras (Gutiérrez, 2011) y se adquiere alrededor de los cinco meses de vida (Yeung et al., 2013). La presencia de este tipo de acento está estrechamente relacionada con la función sintáctica de la palabra (Piñeros, 2008). En el español este tipo de acento aparece casi siempre en las tres últimas sílabas. De acuerdo con esta regla, es posible distinguir tres tipos de *palabras prosódicas*: las *oxítonas*, cuyo acento se ubica en la última sílaba; las *paroxítonas*, cuyo acento se ubica en la penúltima sílaba; y las *proparoxítonas*, cuyo acento se ubica en la antepenúltima sílaba (Hualde, 2013; Piñeros, 2008).

El segunda, es la SP acentual métrica que corresponde a la sílaba acentuada de la palabra prominente en el discurso (Hidalgo & Quilis, 2012), es decir, a la secuencia de sílabas fuertes y débiles en la oración (Calet et al., 2015). Esta constituye la base de la percepción del ritmo lingüístico.

Existen diversos estudios sobre la SP en inglés, español, entre otras lenguas (Calet et al., 2015, 2016; A Holliman et al., 2014; Kim & Petscher, 2016). En el español, se ha estudiado la vinculación de la SP con el desarrollo de la lecto-escritura tanto en niños con lenguaje normotípico como en niños con dislexia y trastorno específico del lenguaje (Calet et al., 2015, 2016, 2019; Riffo et al., 2018).

El Trastorno del desarrollo del lenguaje (TDL) (Bishop et al., 2017) o también llamado Trastorno específico del lenguaje (TEL), es una patología heterogénea cuya prevalencia es de alrededor del 7 % en niños preescolares (Tomblin et al., 1997) que afecta a todos los niveles del lenguaje en distinto grado. El TDL no está asociado a una etiología biomédica conocida. La existencia de factores de riesgo neurobiológicos o ambientales no impide su diagnóstico, y no requiere un desajuste entre la capacidad verbal y no verbal. Además, puede coexistir con otros trastornos del desarrollo neurológico, por ejemplo, el trastorno de Hiperactividad con Déficit de Atención. No obstante, la relación causal con los problemas del lenguaje no está clara (Bishop et al., 2017).

Cabe mencionar que las características lingüísticas de este trastorno van a variar dependiendo de la lengua. Sin embargo, existen especificadores que indican las dimensiones más afectadas. Estos especificadores son: la fonología, la sintaxis, búsqueda de palabras y la semántica, la pragmática, el discurso, aprendizaje verbal y memoria (Bishop et al., 2017).

En el nivel fonético-fonológico se ha descubierto que los niños con TDL presentan dificultades en: la discriminación auditiva de los fonemas, la conciencia fonológica (Coloma, Rodríguez, & De Barbieri, 2005; De Barbieri & Coloma, 2004), la inteligibilidad del discurso, el inventario consonántico, las configuraciones silábicas (Rescorla & Ratner, 1996), los segmentos fonológicos y la construcción silábica (Torres & Soto-Barba, 2016). A su vez, en el nivel suprasegmental se ha descubierto que los niños TDL presentan dificultades en la percepción, la conciencia prosódica y en la vertiente expresiva de la prosodia (Jordán et al., 2019).

En el inglés, Wells y Peppé (2003) estudiaron las habilidades entonación en los niños con TEL. Los autores les aplicaron el test *Profiling Elements of Prosodic Systems-Child (PEPS-C)* a 64 niños, de los

cuales 18 tenían TEL, 28 tenían desarrollo normotípico del lenguaje (DNL), ambos grupos tenían la misma edad cronológica, y 18 con DNL y el mismo rendimiento en comprensión lectora que los niños con TEL. Los resultados arrojaron que no existe diferencia significativa entre el rendimiento de los niños con TEL y el grupo de DNL con similar comprensión lectora. A su vez, en 5 de 16 tareas, los niños con TEL obtuvieron puntajes significativamente más bajos que los niños con DNL. Los autores afirman que las dificultades en la entonación se pueden deber a la memoria auditiva que se requiere en los enunciados prosódicos más largos, y al uso de prosodia con fines pragmáticos e interactivos.

From, Sundström y Samuelsson (2016) describieron la producción de palabras y pseudopalabras con estructura silábica variada y con diferentes acentos, en niños suecos con trastorno fonológico (TF) y trastorno específico del lenguaje expresivo (TEL-E). El estudio mostró que, a nivel prosódico, existe diferencia significativa entre los grupos respecto a la producción de sílabas sin acento y la producción de patrones de acento correctos en pseudopalabras. Los niños con trastorno fonológico presentaron altas puntuaciones en el acento correcto y la repetición de las sílabas correctas de las pseudopalabras en comparación con las palabras. Mientras, los niños con TEL-E presentaron mejores resultados para los acentos correctos en la repetición de palabras en comparación con las pseudopalabras. Los autores indican que las características prosódicas están vinculadas con las diferencias en la memoria de largo plazo y/o de la percepción de los elementos acentuados. También, suponen que el deterioro gramatical puede comprender problemas tanto con la percepción y la memoria de trabajo, como con la prosodia.

En el español, Jordán, Cuetos y Suárez-Coalla (2019) investigaron la prosodia en la lectura de niños con TEL españoles. Para ello, evaluaron la lectura en voz alta de 44 niños (22 TEL y 22 controles) entre siete y 13 años de edad. El texto contenía frases declarativas, interrogativas y exclamativas. Los autores midieron los parámetros prosódicos referentes a duración, tono e intensidad tanto en los párrafos como en las frases del texto. Los resultados mostraron que los niños con TEL leen más despacio, comenten más errores y hacen más pausas innecesarias, ya sea entre palabras (interléxicas), como dentro de las palabras (intra léxicas). Además, se observan menos variaciones tonales en la frase interrogativa y en rasgos finales de la frase.

Tal como se evidencia, la mayor parte de los estudios de SP en niños con TDL se han enfocado en la prosodia. En el español se ha investigado la fluidez

de la lectura en voz alta, la prosodia relacionada con la lectura, pero no la percepción de la SP acentual léxica ni la SP acentual métrica en niños con TDL. Por consiguiente, el presente estudio tuvo como propósito determinar la SP acentual de niños con TDL que cursan segundo grado de primaria.

2. METODOLOGÍA

La muestra estuvo conformada por 76 niños chilenos, que cursaban segundo grado de educación primaria. De los cuales, 50 tenían un DNL, cuya edad promedio fue de 7,74 años (D.E.= 0,35), y 26 niños con TDL, cuya edad promedio fue de 7,69 años (D.E.= 0,45). Tanto los niños con TDL como con DNL presentaban audición dentro de rangos normales. Todos los participantes con TDL contaban con el diagnóstico correspondiente. Los niños con DNL fueron seleccionados teniendo en cuenta la ausencia de antecedentes de trastornos de lenguaje o de aprendizaje. Toda la información necesaria para estos fines fue proporcionada por la escuela a la que los niños asistían.

Para evaluar la conciencia prosódica acentual se utilizaron dos test: *Test de acento léxico* y *Test de nombres compuestos*. Estos instrumentos se basaron en los empleados por Calet et al. (2015), y fueron adaptados lingüísticamente al español de Chile.

El *Test de acento léxico* consiste en la evaluación de la sensibilidad del acento léxico y fue implementado en el software *E-prime 3*. El niño escucha a través de auriculares una palabra trisilábica, al mismo tiempo que en la pantalla del computador ve tres imágenes. Cada imagen está conformada por tres “montañas”, dos del mismo tamaño y una más alta que representa la sílaba acentuada (ya sea, la primera, la segunda o la última sílaba, respectivamente). El niño debe elegir la imagen que corresponde a la palabra que acaba de escuchar. Por cada respuesta correcta, se le otorga un punto. El tiempo de respuesta fue registrado en milisegundos.

El estímulo auditivo corresponde a un tipo de palabra prosódica. Los estímulos fueron grabados en una cabina insonorizada con un micrófono Audio-Technica AT2020 y con una tarjeta Lexicon Omeg. La locución fue realizada por una mujer, hablante nativa del español de Chile, cuya profesión es profesor de lengua española. La grabación se efectuó con una frecuencia de muestreo de 44.1 kHz y una profundidad de 16 bits.

El *Test de nombres compuestos* se basó en la prueba de Calet et al. (2015). El objetivo de la prueba es medir la SP acentual a nivel métrico. Para ello, los niños deben distinguir entre un nombre compuesto y dos nombres, por ejemplo: “*abrelatas*” v/s “*Abre*” “*Latas*”. Cabe mencionar que el test presenta dos

estímulos, uno visual y uno auditivo. El estímulo visual se compone de dos imágenes, una correspondiente al nombre compuesto y la otra a los dos nombres. Mientras que el estímulo acústico es la palabra objetivo. El instrumento tiene dos ítems de práctica y 20 evaluados. Los niños deben elegir la imagen que corresponde al estímulo que escuchan. Por cada respuesta correcta, se le otorga un punto. El tiempo de respuesta fue registrado en milisegundos.

La grabación de los estímulos auditivos del test de nombres compuestos fue realizada de la misma forma que los estímulos de la prueba acento léxico. Cuando el estímulo correspondía a dos nombres se agregó una pausa de 200 ms, para ello se utilizó un script ad hoc en el software *Praat*.

3. RESULTADOS

El registro de los datos y la creación de las tablas se realizaron con el software Microsoft Excel 2019, mientras que los análisis estadísticos se efectuaron con el software IBM SPSS Statistics versión 24. Se utilizó el estadígrafo UNIANOVA para realizar las comparaciones entre los grupos.

Tal como se observa en la tabla 1, los niños con DNL en la prueba de acento léxico obtuvieron una media de 44,43 (D.E.= 14,84), mientras que la prueba de acento métrico la media fue de 79,58 (D.E.= 15,09). Los tiempos de respuestas correctas en la prueba de sensibilidad prosódica acentual léxica alcanzaron una media de 2415 ms (D.E.= 867), en cambio, en la prueba de acento métrico se obtuvo una media de 1540 ms (D.E.= 811).

Los niños con TDL en la prueba acento léxico obtuvieron una media de 37,46 (D.E.= 10,15) y en la prueba de acento métrico un promedio de 62,66 (D.E.= 17,18). Los tiempos de respuestas correctas en la prueba de sensibilidad prosódica acentual léxica alcanzaron una media de 2382 ms (D.E.= 1115), en cambio, en la prueba de acento métrico se obtuvo una media de 1442 ms (D.E.= 720).

Al comparar los datos de ambos grupos, se observó que los resultados de los niños con TDL están por debajo de los niños con DNL en ambas pruebas de SP acentual. Sin embargo, estas diferencias no son estadísticamente significativas (p -valor > 0,05).

Por lo otro lado, al analizar el desempeño del test acento léxico se constató que existen diferencias estadísticamente significativas en el reconocimiento de las palabras proparoxítonas y el en tiempo de reconocimiento de las respuestas correctas de las palabras paroxítonas (p -valor < 0,05). Por otra parte, se observó que en el desempeño del test acento métrico no existen diferencias estadísticamente significativas en la precisión de la respuesta, ni en el tiempo que la tarea demandó (p -valor > 0,05).

Además, se realizaron análisis estadísticos al desempeño por ítem en la prueba de acento léxico.

Para ello, se utilizaron los estadígrafos chi cuadrado (χ^2) y Prueba exacta de Fisher.

Tabla 1: Sensibilidad prosódica acentual en niños con desarrollo normotípico del lenguaje (DNL) versus niños Trastorno del desarrollo del lenguaje (TDL) de segundo grado de educación primaria.

Variable	DNL		TDL		Estadígrafo	valor p
	Media	D.E.	Media	D.E.		
Léxico	44,43	14,84	37,46	10,15	3,4	0,0671
Oxítonas	48,57	27,38	48,35	28,02	0	0,9748
Paroxítonas	44	25,98	33,76	23,2	1,9	0,1708
Proparoxítonas	40,86	30,27	31,32	21	6	0,016
T. Léxico	2415	867	2382	1115	0	0,9088
T. Oxítonas	2656	1198	2493	1124	3,1	0,0822
T. Paroxítonas	2566	1057	2071	1177	20,6	<0,0001
T. Proparoxítonas	2214	1125	2701	1598	0,3	0,5836
A. Métrico	79,58	15,09	62,66	17,18	20,1	0,9088
Nombre compuesto	83,85	19,43	63,31	19,74	1,9	0,1668
Dos nombres	74,55	22,87	61,89	21,67	0,1	0,757
T. Métrico	1540	811	1442	720	0,3	0,575
Nombre compuesto	1492	796	1548	792	1,3	0,2498
Dos nombres	1676	1098	1320	702	2,8	0,0965

Los resultados arrojaron que la distribución de la frecuencia para grupos TDL y DNL presentan diferencias significativas en las palabras oxítonas: *popular* (p -valor= 0,022) y *jabalí* (p -valor= 0,007); paroxítonas: *fantasma* (p -valor= 0,028), *linterna* (p -valor= 0,030), *mañana* (p -valor= 0,036) y *marido* (p -valor= 0,018); y en las proparoxítonas: *círculo* (p -valor= 0,039). Cabe mencionar que el desempeño de los niños con TDL no siempre es menor que en los con DNL. Los resultados muestran que en las palabras oxítonas *popular* (73,1%), *hospital* (65,4%) y *pantalón* (42,3%); y en las palabras paroxítonas *corona* (46,2%) y *garganta* (34,6%), los niños con TDL obtuvieron un mejor desempeño que lo niños con DNL.

También, se realizó el análisis estadístico por ítem en la prueba de acento métrico. Para ello, se utilizó chi cuadrado (χ^2) y Prueba exacta de Fisher.

Los resultados muestran que la distribución de la frecuencia tanto para TDL como para DNL presentan diferencias significativas en los ítems de nombre compuestos: *espuma* (p -valor= 0,000), *estrés* (p -valor= 0,004), *lloviendo* (p -valor= 0,043), *pezuñas* (p -valor= 0,006), *quitasol* (p -valor= 0,003), *saltamontes* (p -valor= 0,007) y una leve significancia en el ítem de “dos nombres” *pintauñas* (p -valor= 0,083) y en los ítems “nombre compuestos” *paraguas* (p -valor= 0,050), *soldado* (p -valor= 0,065).

4. CONCLUSIONES

El desempeño de los niños con TDL en las pruebas de sensibilidad prosódica acentual se encuentra por debajo de los niños normales. Sin embargo, estas diferencias no son estadísticamente significativas.

En el test acento léxico se constató que los niños con DNL presentaron un desempeño superior en el reconocimiento de las palabras proparoxítonas y el en tiempo de reconocimiento de las respuestas correctas de las palabras paroxítonas. Este hallazgo es estadísticamente significativo.

Al analizar por ítem, en ambas pruebas, se puede observar que en algunos estímulos, los niños con TDL presentan un mejor rendimiento que los niños con DNL. Estos hallazgos coinciden con la descripción de la definición de trastorno, ya que el mismo TDL se define como un trastorno heterogéneo, es decir presentan algunas habilidades conservadas, otras retrasadas y otras alteradas.

5. BIBLIOGRAFÍA

Bishop, D., Snowling, M., Thompson, P., & Greenhalgh, T. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi

- consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*, 58(10), 1068-1080. <https://doi.org/10.1111/jcpp.12721>
- Calet, N., Flores, M., Jiménez-Fernández, G., & Defior, S. (2016). Habilidades fonológicas suprasegmentales y desarrollo lector en niños de Educación Primaria. *Anales de Psicología*, 32(1), 72-79. <https://doi.org/10.6018/analesps.32.1.216221>
- Calet, N., Gutiérrez-Palma, N., Defior, S., & Jiménez-Fernández, G. (2019). Linguistic and non-linguistic prosodic skills in Spanish children with developmental dyslexia. *Research in Developmental Disabilities*, 90, 92-100. <https://doi.org/10.1016/j.ridd.2019.04.013>
- Calet, N., Gutiérrez-Palma, N., Simpson, I., González-Trujillo, M., & Defior, S. (2015). Suprasegmental Phonology Development and Reading Acquisition: A Longitudinal Study. *Scientific Studies of Reading*, 19(1), 51-71. <https://doi.org/10.1080/10888438.2014.976342>
- Coloma, C., Rodríguez, L., & De Barbieri, Z. (2005). Conciencia Fonológica Y Lengua Escrita En Niños Con Trastorno Específico Del Lenguaje Expresivo. *Revista CEFAC*, 7(4), 419-425.
- De Barbieri, Z., & Coloma, C. (2004). La conciencia fonológica en niños con trastorno específico de lenguaje. *Revista de Logopedia, Foniatría y Audiología*, 24(4), 156-163. [https://doi.org/10.1016/S0214-4603\(04\)75798-8](https://doi.org/10.1016/S0214-4603(04)75798-8)
- From, Å., Sundström, S., & Samuelsson, C. (2016). Differences in phonologic and prosodic abilities in children with phonological language impairment and phonological-grammatical language impairment assessed with non-word repetition. *Logopedics Phoniatrics Vocology*, 41(2), 66-76. <https://doi.org/10.3109/14015439.2014.982171>
- Gutiérrez, N. (2011). Procesamiento silábico y acento en español. *Revista de Logopedia, Foniatría y Audiología*, 31(1), 39-43. [https://doi.org/10.1016/S0214-4603\(11\)70170-X](https://doi.org/10.1016/S0214-4603(11)70170-X)
- Hidalgo, A., & Quilis, M. (2012). *La voz del lenguaje: Fonética y fonología del español*. Tirant Humanidades.
- Holliman, A., Critten, S., Lawrence, T., Harrison, E., Wood, C., & Hughes, D. (2014). Modeling the Relationship Between Prosodic Sensitivity and Early Literacy. *Reading Research Quarterly*, 49(4), 469-482. <https://doi.org/10.1002/rrq.82>
- Holliman, A., Mundy, I., Wade-Woolley, L., Wood, C., & Bird, C. (2017). Prosodic awareness and children's multisyllabic word reading. *Educational Psychology*, 37(10), 1222-1241. <https://doi.org/10.1080/01443410.2017.1330948>
- Hualde, J. (2013). *Los sonidos del español: Spanish Language edition*. Cambridge University Press.
- Jordán, N., Cuetos, F., & Suárez-Coalla, P. (2019). La prosodia en la lectura de niños con trastorno específico del lenguaje. *Infancia y Aprendizaje*, 42(1), 87-127. <https://doi.org/10.1080/02103702.2018.1550161>
- Kim, Y., & Petscher, Y. (2016). Prosodic Sensitivity and Reading: An Investigation of Pathways and Relations Using a Latent Variable Approach. *Journal of Educational Psychology*, 108(5), 630-645. <https://doi.org/10.1037/edu0000078>
- Piñeros, C. (2008). *Estructura de los sonidos del español*. Pearson Prentice Hall.
- Rescorla, L., & Ratner, N. (1996). Phonetic profiles of toddlers with specific expressive language impairment (SLI-E). *Journal of Speech and Hearing Research*, 39(1), 153-165.
- Riffo, B., Caro, N., & Sáez, K. (2018). Linguistic Awareness, Reading Aloud and Reading Comprehension. *Revista de Linguística Teórica y Aplicada*, 56(2), 175-198. <https://doi.org/10.4067/S0718-48832018000200175>
- Shi, R., & Werker, J. (2003). The basis of preference for lexical words in 6-month-old infants. *Developmental Science*, 6(5), 484-488. <https://doi.org/10.1111/1467-7687.00305>
- Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of Specific Language Impairment in Kindergarten Children. *Journal of speech, language, and hearing research*, 40(6), 1245-1260.
- Torres, V., & Soto-Barba, J. (2016). Ajustes fonético-fonológicos en niños con trastornos específicos del lenguaje mixto (TEL Mixto). *Onomázein*, 69-87. <https://doi.org/10.7764/onomazein.33.7>
- Wells, B., & Peppé, S. (2003). Intonation Abilities of Children With Speech and Language Impairments. *Journal of Speech, Language, and Hearing Research; Rockville*, 46(1), 5-20.
- Yeung, H., Chen, K., & Werker, J. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language*, 68(2), 123-139. <https://doi.org/10.1016/j.jml.2012.09.004>

¿ES MI PREGUNTA UNA INTERROGATIVA EN ESPAÑOL? LA ENTONACIÓN DEL ESPAÑOL HABLADO POR POLACOS

Weronika Urbanik-Pęk

Universidad Pedagógica de Cracovia, Polonia
weronika.urbanik-pek@up.krakow.pl

ABSTRACT

En el presente trabajo, describimos los rasgos melódicos de las interrogativas del español hablado por polacos. La investigación aquí presentada sigue el método Análisis Melódico del Habla (AMH) y forma parte de un trabajo más amplio y completo dedicado a esta interlengua. Para llevar a cabo el estudio, nos hemos basado en 36 preguntas absolutas, producidas por 16 informantes, que han sido analizadas con el software Praat, se han extraído los valores de F_0 , estandarizado y presentado en gráficos.

El estudio nos ha permitido constatar que la mayoría de interrogativas que producen los polacos hablando español no se sitúan dentro de los márgenes de los cuatro patrones interrogativos definidos para el español peninsular. Solo algunos siguen el *patrón interrogativo II. Inflexión Final +70%* y el *patrón XIII. Cuerpo e inflexión final ascendentes*. El resto de contornos se caracterizan, en su mayoría, por la ausencia del primer pico, el cuerpo muy plano y una inflexión final ascendente insuficientemente marcada o descendente, y, unos pocos, presentan rasgos enfáticos. En los tres casos, pueden tener lugar malentendidos en la comunicación porque podrían interpretarse como suspendidos, neutros y enfáticos, respectivamente.

Palabras clave: entonación, Análisis Melódico del Habla, oraciones interrogativas, interlengua, español

In this work, we describe the melodic features of the interrogative sentences in Spanish spoken by Poles. The research presented here follows the Melodic Analysis of Speech (MAS) method and it forms a part of a larger and more complete work dedicated to this interlanguage. To carry out the study, it was based on 36 absolute questions, produced by 16 informants, which have been analyzed with the Praat software. We have extracted the F_0 values, which have been standardized and presented in graphs.

The study has allowed us to verify that the majority of interrogative sentences produced by Poles speaking Spanish are not within the margins of the four interrogative patterns defined for the peninsular Spanish. Only some of them follow the interrogative *pattern II* with the final inflection +70% and others follow the *pattern XIII*, with the ascending body and final inflection. The rest of the contours are characterized by the absence of the first peak, a very flat body and an insufficiently marked ascending or descending final inflection, and some of them have emphatic features. In the cases mentioned above, there may occur some misunderstandings in the communication because they could be interpreted as suspended, neutral or emphatic.

Keywords: intonation, Melodic Analysis of Speech, interrogative sentences, interlanguage, Spanish

1. INTRODUCCIÓN

Hoy en día, el español es una de las lenguas extranjeras más populares en Polonia. Según los datos del Ministerio de Educación y Formación Profesional de España, el número de estudiantes de la educación reglada en el actual curso académico llega a unos 200.000. El hecho de que en las carreras universitarias el afán por estudiar el

castellano también sigue aumentando nos permite hablar del español como de la tercera lengua más estudiada en Polonia, tras el inglés y el alemán.

Hasta hoy en día la investigación sobre la entonación y la pronunciación en Polonia no ha provocado mucho interés y son escasos los trabajos que tratan este tema de manera detallada. Tras la publicación del libro de Dłuska sobre la prosodia del polaco en 1976 se han publicado muy pocos textos. En la última década aparecieron dos libros en cuyas

páginas aparecen fragmentos dedicados a la entonación, pero son más bien textos sobre todos los elementos suprasegmentales del habla analizados desde la perspectiva del discurso, turnos de habla, etc.

Aunque la adquisición de los rasgos prosódicos del español por los estudiantes extranjeros (de distintas nacionalidades) de ELE cuenta con mucha bibliografía, en Polonia no se ha trabajado todavía este tema, excepto Sobański (2012), el cual, utilizando la metodología de AMPER-España, estudia la adquisición de los patrones melódicos de las declarativas y de las interrogativas absolutas del español por parte de estudiantes polacos. A partir de este trabajo, no hay ningún estudio más que se centre en la prosodia y en la entonación de esta interlengua, o sea, del español hablado por polacos.

El tema que aquí trataremos con más detalle forma parte de la tesis doctoral sobre la entonación del español hablado por polacos analizada según la metodología AMH. La tesis doctoral junto con los artículos publicados previa y posteriormente rellenarán el hueco en la lingüística del español en el ámbito polaco. Desde este punto de vista es un trabajo pionero e innovador que permitirá a los estudiantes trabajar la entonación, practicarla y de este modo, mejorar sus competencias orales en la lengua.

2. LA METODOLOGÍA Y EL CORPUS

2.1. El método *Análisis Melódico del Habla (AMH)*

La metodología utilizada para realizar el presente estudio es el *Análisis Melódico del Habla (AMH)*, presentada en Cantero (2002) y revisada en Font-Rotchés (2007). Ambos investigadores anteriormente mencionados completaron el método con un protocolo de análisis (Cantero y Font-Rotchés, 2009), que ha sido revisado y ampliado en 2020.

El *Análisis Melódico del Habla* ofrece una interpretación fonológica, no dependiente de otros niveles de análisis lingüístico, de los fenómenos tonales de la lengua castellana (acento, ritmo y entonación). Gracias al criterio de la segmentación descrito en el AMH se puede analizar los valores relativos de las melodías de distintos enunciados de habla espontánea de los informantes. La visualización de la curva melódica y la posibilidad de analizarla de manera muy precisa permite aplicar el método no solo en el ámbito didáctico sino también en el científico o el clínico.

El Análisis Melódico del Habla distingue tres tipos de funciones o niveles de la entonación: lingüística, prelingüística y paralingüística. La entonación lingüística, o sea, la interpretación fonológica de la melodía está definida mediante la oposición de tres rasgos: interrogativo, enfático y suspendido. A base de combinar estos tres rasgos, se han distinguido 8 tonemas en español (Cantero, 2002):

1. /+interrog., +enfát., +susp./
2. /+interrog., +enfát., -susp./
3. /+interrog., -enfát., +susp./
4. /+interrog., -enfát., -susp./
5. /-interrog., +enfát., +susp./
6. /-interrog., +enfát., -susp./
7. /-interrog., -enfát., +susp./
8. /-interrog., -enfát., -susp./

La entonación interrogativa, analizada en el presente trabajo, corresponderá a los tonemas 4 y 8 que pueden realizarse con distintas melodías, que son típicas del castellano y que llamamos *patrones melódicos*. Aplicando el AMH se trabaja con el *software* Praat (Boersma, Weenink, 2018) en el que podemos identificar y segmentar las unidades melódicas de los enunciados y obtener el valor de la frecuencia fundamental F_0 de cada segmento tonal. Dichos valores absolutos, extraídos de este software y pasados al Excel, son estandarizados calculando la distancia tonal en porcentajes entre un segmento tonal y el siguiente. Los valores relativos resultantes permiten visualizar la curva melódica mediante un gráfico en el cual observamos ascensos o descensos del tono independientemente del sexo o de la edad de los informantes.

Cada unidad melódica se la analiza teniendo en cuenta las tres partes de cada contorno, anacrusis, cuerpo e inflexión final, y dos acentos, el primer pico y el núcleo. La anacrusis está formada por los primeros segmentos tonales ascendentes que culminan en la primera vocal tónica o primer pico. El cuerpo es la parte de contorno que va desde el primer pico hasta la última vocal tónica, llamada núcleo; y la inflexión final empieza en el núcleo y continua hasta el final del contorno.

2.2. El corpus

Las unidades melódicas elegidas para el presente estudio forman parte de un corpus de habla espontánea mucho más amplio del español hablado por polacos. Se han analizado 36 preguntas absolutas en el contexto, producidas por 16

informantes (3 hombres y 13 mujeres), de edad entre 19 y 47 años. Los informantes, en la mayoría de los casos eran estudiantes de español (o de filología hispánica o de cursos de lengua en el Instituto Cervantes de Cracovia). La mayoría de ellos ha declarado tener un dominio del español de nivel B2 o superior.

3. RESULTADOS

En este apartado, se presenta la descripción de los contornos obtenidos a partir de las interrogativas absolutas de nuestro corpus, teniendo en cuenta los rasgos significativos en cada parte del contorno: anacrusis, primer pico, cuerpo, núcleo e inflexión final. Después, se los compara con los cuatro patrones melódicos interrogativos del español peninsular descritos según la metodología AMH (Cantero y Font-Rotchés, 2007; Font-Rotchés y Mateo, 2013, 2017). Se trata del Patrón II Inflexión Final ascendente +70%; Patrón III Inflexión Final ascendente 40%-70%; Patrón IV Inflexión Final ascendente-descendente; y Patrón XIII Cuerpo e Inflexión Final ascendentes.

3.1. Inflexión final ascendente (+70%)

Los contornos del corpus que presentan una inflexión final ascendente (+70%) son los que siguen el patrón II del español /+interrog., -enfát., -susp./, que se caracteriza por un primer pico, que puede llegar a un 40% (puede estar en la primera sílaba tónica o estar desplazado a una sílaba átona posterior o también puede no presentar primer pico), un cuerpo con un descenso tonal suave y una inflexión final de un 70% o superior.

En nuestro corpus aparecen diez preguntas que siguen este patrón melódico. En todos los contornos encontrados la inflexión final supera el 70% y en ninguno de ellos encontramos el primer pico. El cuerpo suele ser muy plano con unos descensos muy suaves. No obstante, en caso de este patrón, solo basta analizar el ascenso muy notable de la inflexión final para poder clasificar un contorno como una interrogativa. Teniéndolo en cuenta, los ejemplos encontrados no nos ofrecen problema ninguno, ya que su inflexión final tan marcada nos permite interpretarlos directamente como preguntas.

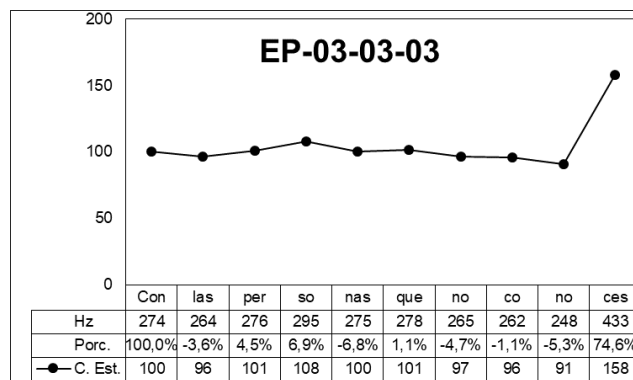


Figura 1: Contorno típico del patrón II: ¿Con las personas que no conoces?

Un ejemplo de este tipo de contorno es el que se encuentra en la Figura 1. Se trata de un contorno sin primer pico y cuerpo muy plano; los cambios del tono allí presentes no son significativos. En cuanto a la inflexión final, vemos una subida notable del tono (+75%), lo que nos permite constatar que este enunciado responde al patrón melódico II.

3.2. Inflexión final ascendente (40%-70%)

Los contornos del corpus con este ascenso final son similares al patrón III, caracterizado por los tonemas /+interrog., -enfát., -susp./ y complementario del patrón II. La única diferencia entre ellos consiste en que en el patrón II solo la inflexión final (+70%) permite clasificar a un enunciado como interrogativo, mientras que en el patrón III, se suele observar un ascenso de entre un 40% y un 70% en la inflexión final y el primer pico está desplazado a una sílaba átona posterior a la primera tónica.

Hemos obtenido en nuestro corpus diez interrogativas con un rasgo coincidente con el patrón melódico III: en la inflexión final observamos una subida tonal situada entre un 40% y un 69%. La mayoría de ellas, excepto dos (EP-10-02-03, ¿Queréis continuar los estudios? y EP-02-05-04 ¿Creéis que es necesario dormir por lo menos ocho horas al día?) carecen de primer pico, lo que es uno de los rasgos característicos de la interlenguaspañol hablado por polacos y también de otras, como el español hablado por italianos (Devís, 2011) o brasileños (Fonseca, 2013). En este sentido, los enunciados de este grupo es muy probable que, al no presentar primer pico, sean interpretados como enunciados no acabados (patrón suspenso VIb del español, Cantero y Font-Rotchés, 2007), en lugar de interrogativos.

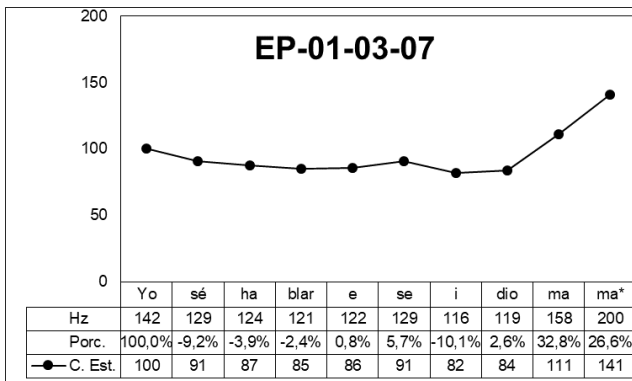


Figura 2: Contorno típico del patrón III: ¿Yo sé hablar ese idioma?

En la Figura 2, presentamos un contorno en el que la inflexión final coincide con el patrón interrogativo III y el patrón suspendido VIb. En este caso el cuerpo es muy plano y se nota en él una bajada muy suave pero regular. Este contorno, como los demás de este grupo, carecen de primer pico, lo cual implica que no se pueda garantizar que sea entendido como una pregunta.

Cabe mencionar tres contornos (EP-05-01-05 ¿Tus mujeres necesitan el dinero?, EP-01-03-07 ¿Yo sé hablar ese idioma?, EP-01-02-03 ¿Y yo tengo el pelo amarillo?) cuya inflexión supera el 65%. Estos enunciados, que presentan una inflexión final muy cercana al 70%, podrían ser interpretados como preguntas en diversos contextos.

3.3. Inflexión final 15%-40%

Entre los ejemplos obtenidos para ese estudio, se encuentran ocho cuya inflexión final no está tan marcada y oscila entre el 15%-40%. Estos contornos, dependiendo del contexto, podrían interpretarse como una suspendida representada por el patrón melódico VI. En éste, la inflexión final es de un 15%-40% (patrón melódico VIb) o de 15%-70% (patrón melódico VIa) (Cantero y Font-Rotchés, 2007).

Los contornos del corpus son similares al de la Figura 3: ausencia de primer pico o, si aparece, está tan poco marcado que apenas se lo considera como tal; cuerpo muy plano e inflexión final con un ascenso leve. Los rasgos presentes en estos enunciados asimilan el contorno al patrón VI lo que, sin situarse en el contexto, nos permitiría interpretarlo como una suspendida en lugar de una interrogativa. En estos casos, que son relativamente frecuentes en el español hablado por polacos, el emisor no consigue su intención comunicativa.

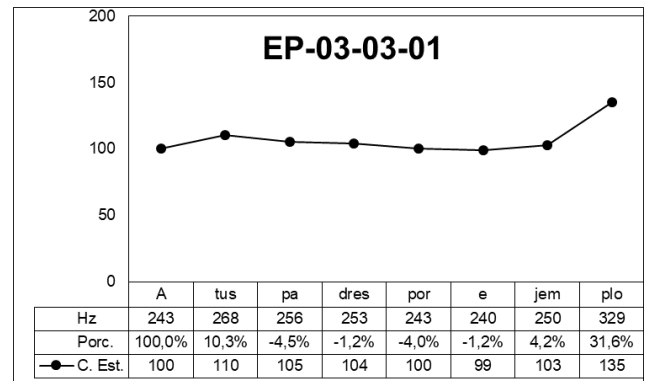


Figura 3: Contorno similar al patrón VI: ¿A tus padres por ejemplo?

3.4. Inflexión final inferior a un 15%

Se han registrado cuatro casos en los que podemos observar una inflexión final inferior a un 15%. Tres de ellos son contornos muy planos, sin primer pico y con la inflexión final poco marcada. Estas características permitirían interpretarlos como declarativas. En la Figura 4, podemos observar uno de ellos, en el que la inflexión final no llega al 10% y todo el cuerpo desciende suavemente.

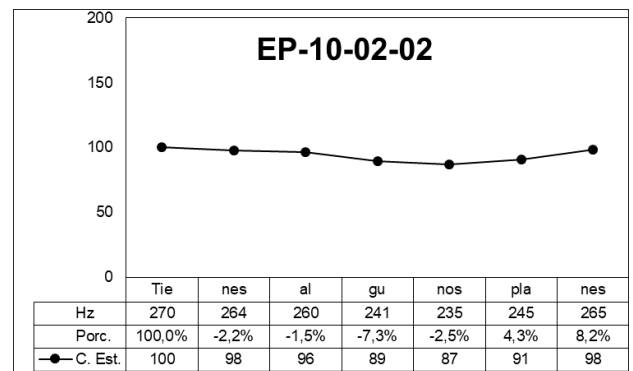


Figura 4: Contorno similar al patrón declarativo: ¿Tienes algunos planes?

3.5. Cuerpo e inflexión final ascendentes

Estos rasgos melódicos tienen lugar en solo dos enunciados del corpus y son coincidentes con el patrón XIII /+interrog.,+enfát., -susp./ del español peninsular (Font-Rotchés y Mateo, 2013, 2017). Este patrón se caracteriza por un cuerpo ascendente, a menudo sin primer pico y una inflexión también ascendente (de un 15% o más). El ascenso total en el cuerpo y en la inflexión final debe superar un 60%.

Los ejemplos de nuestro corpus no presentan el primer pico. En la Figura 5, vemos un contorno con

un cuerpo que asciende constantemente y culmina en la inflexión final con un ascenso de un 40,4%. El ascenso total es de 154% y en el otro contorno que hemos obtenido, de un 86%. Estas dos preguntas podrían ser interpretadas claramente como interrogativas ya que siguen el patrón melódico XIII, no obstante, como se demuestra aquí, no es una tendencia muy frecuente.

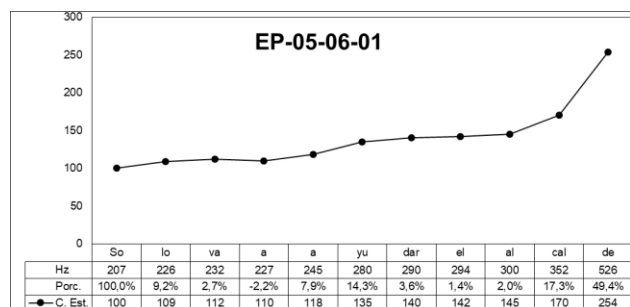


Figura 5: Contorno típico del patrón XIII: ¿Solo va a ayudar al alcalde?

3.6. Otros contornos

Un total de dos contornos presentan rasgos melódicos peculiares y los trataremos en este apartado.

En la Figura 6, la melodía es similar a la del patrón melódico XIIa, un patrón enfático cuyo cuerpo es muy plano y elevado, y la inflexión final, tras un énfasis, desciende (Cantero y Font-Rotchés, 2007).

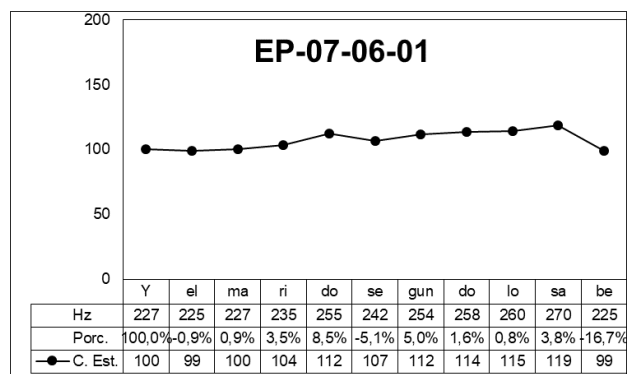


Figura 6: Contorno similar al patrón enfático: ¿Y el segundo marido lo sabe?

En la Figura 7, presentamos el único ejemplo del contorno sin primer pico, un cuerpo muy plano con un ascenso de un 48% iniciado en la sílaba prenuclear -ca- y que culmina en el núcleo del contorno -cio- de vacaciones; después del núcleo, la melodía desciende un 22% (no retorna al punto inicial) hasta el final del contorno. Este tipo de

esquema melódico se aproxima al patrón VII Inflexión final de núcleo elevado, que se caracteriza por ser enfático.

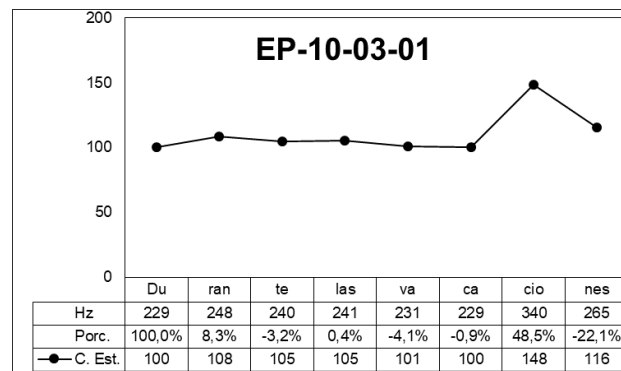


Figura 7: Contorno similar al patrón IV: ¿Durante las vacaciones?

4. CONCLUSIONES

A modo de conclusión, podemos constatar que de los cuatro patrones melódicos interrogativos descritos para el español peninsular, en nuestro corpus de la interlengua español hablado por polacos, solo se utilizan dos: el patrón melódico II. IF ascendente +70% y el patrón XIII. Cuerpo e inflexión final ascendentes. En cuanto al patrón III. Inflexión final ascendente 40-70% son diversos los contornos del cuerpo que siguen esta inflexión final, pero sin primer pico, lo que puede provocar que en la mayoría de ocasiones no se interpreten como preguntas. Finalmente, referente al patrón IV Inflexión final ascendente-descendente, que es interrogativo y enfático, no tenemos ningún ejemplo en el corpus. Otro rasgo melódico que cabe subrayar en el español hablado por polacos es la ausencia de primer pico en casi todos los contornos. La investigación llevada a cabo para el presente trabajo demuestra que los polacos cuando hablan español frecuentemente no aciertan a seguir el patrón melódico interrogativo y se podrían dar malentendidos en la comunicación, ya que sus preguntas podrían interpretarse como enunciados suspendidos y, en menor grado, neutros o enfáticos. Trabajar bien los patrones melódicos en la clase de ELE debería convertirse en uno de los objetivos porque aunque un estudiante polaco domine bien la gramática y el vocabulario, puede presentar dificultades comunicativas provocadas por el uso de una entonación inadecuada en diversos contextos.

6. BIBLIOGRAFÍA

- Boersma, P. y Weenink, D. (2018). *Praat: Doing Phonetics by Computer (Computer program). Version 6.1.02.* <http://www.praat.org>.
- Cantero Serena, F. J. (2002). *Teoría y análisis de la entonación.* Barcelona: Edicions Universitat de Barcelona.
- Cantero Serena, F. J. y Font-Rotchés, D. (2007). *Entonación del español peninsular en habla espontánea: Patrones melódicos y márgenes de dispersión.* Moenia. Revista Lucense de lingüística y literatura, 13, 69-92.
- Cantero Serena, F. J. y Font-Rotchés, D. (2009). Protocolo para el análisis melódico del habla. Estudios de fonética experimental, 18, 17-32.
- Cantero Serena, F. J. y Font-Rotchés, D. (2020). Melodic Analysis of Speech (MAS). Phonetics of Intonation. En Abasolo, J.; de Pablo, I.; Ensunza, A.: *Contributions on education* (pp. 20-47). Universidad del País Vasco.
- Devís Herráiz, E. (2011). *La entonación del español hablado por italianos.* en Didáctica. Lengua y Literatura, Madrid. 35-58
- Dłuska, M. (1976). *Prozodia języka polskiego.* Kraków.
- Fonseca de Oliveira, A. (2013). *Caracterización de la entonación del español hablado por brasileños* [Universitat de Barcelona]. <http://diposit.ub.edu/dspace/handle/2445/54688>
- Font-Rotchés, D. (2007). L'entonació del català. Publicacions de l'Abadia de Montserrat.
- Font-Rotchés, D. y Mateo, M. (2013). Entonación de las interrogativas absolutas del español peninsular del sur en habla espontánea. Onomázein, 28, 256-275.
- Font-Rotchés, D. y Mateo, M. (2017). Melodías para confirmar, preguntar, sugerir o pedir en español. Phonica, 13, 49-67.
- Sobański, T. (2012). *Adquisición de los patrones entonativos españoles por los alumnos polacos de ELE: una aproximación,* en Banyś, W. et all (2012). *Neophilologica.* Katowice.

ADQUISICIÓN DE LA ENTONACIÓN DE ELE: LA PRODUCCIÓN DEL FOCO ESTRECHO POR LOS SINOHABLANTES

Weiqi Li

Universitat Autònoma de Barcelona
weiqi.li@e-campu.uab.cat

ABSTRACT

En el aprendizaje de lenguas extranjeras, la entonación lanza un desafío particular a los estudiantes. El presente trabajo da cuenta de la producción prosódica del foco estrecho en el español peninsular, y tiene como objetivo analizar las dificultades mostradas por los sinohablantes. Los resultados indican que los sinohablantes son generalmente capaces de emplear variados recursos prosódicos para marcar la mayor prominencia en contextos experimentales. Sin embargo, el análisis enfocado en los correlatos acústicos (en particular, la elevación de F0, la alineación temporal del acento tonal focalizado y el grado de compresión tonal post-foco) sugiere que la implementación fonética tiende a causar mayores dificultades de adquisición que los patrones fonológicos categóricos. Además, entre las dificultades halladas, mientras que algunas se superan con el desarrollo general de la competencia de la lengua meta, algunas inciden ampliamente en el aprendizaje de todas las etapas, por lo que requieren mayor atención en la didáctica.

Palabra clave: entonación, adquisición de lengua extranjera, sinohablante, foco estrecho

Intonation poses a particular challenge for foreign language learners. This study takes into account the prosodic production of narrow focus in Peninsular Spanish, and aims to analyse the acquisition difficulties shown by Chinese native speakers. The results show that the Chinese learners are generally able to use various prosodic resources in marking narrow focus in experimental contexts. However, the analysis focused on acoustic correlates (in particular, the scaling and alignment of the focused pitch accent, also the degree of post-focal compression) suggest that phonetic implementation tends to cause greater acquisition difficulties than categorical phonological patterns. In addition, among the difficulties that have been found, while some of them are overcome along with the general development of the target language proficiency, others have broader impacts at all learning stages, hence it requires more efforts in teaching practice.

Keywords: intonation, foreign language acquisition, Chinese speaker, narrow focus

1. INTRODUCCIÓN

En el aprendizaje de lenguas extranjeras, la entonación lanza un desafío particular a los aprendices por las múltiples dimensiones en que incide. Se sabe que la delimitación de un movimiento entonativo implica descripciones de tres aspectos estrechamente relacionados: uno fonológico, que configura el patrón de modo categórico; uno fonético, que expone la forma de que los elementos estructurales de la entonación se implementan o se realizan acústicamente; y otro semántico, que relaciona el uso del patrón específico con un determinado significado. Desde el punto de vista de

la adquisición de lengua extranjera, una producción entonativa adecuada no solo supone emitir patrones melódicos apropiados según el contexto comunicativo, sino que también requiere realizaciones fonológicamente y fonéticamente precisas.

La descripción del contorno melódico a lo largo de este estudio se basa en el Modelo Autosegmental-Métrico (el modelo AM en lo que sigue) (Pierrehumbert, 1980; Ladd, 2008). Uno de los planteamientos cruciales del modelo consiste en que la categórica representación fonológica debe discriminarse de la realización fonética en la descripción entonativa. Fonológicamente, el modelo identifica dos tipos de eventos tonales importantes

en el contorno y los marca con tonos altos (H) y bajos (L): un tipo de evento asociado con las sílabas métricamente prominentes, que se denomina de acento tonal (etiquetado con el asterisco, p. ej. H* o L*); otro tipo, llamado tono de frontera, que se encuentra en las fronteras de determinadas unidades melódicas (en el caso de las frases intermedias, se etiqueta como H- o L-; en el caso de las frases de entonación, se marca como H% o L%). Fonéticamente, la implementación de los eventos fonológicos se refleja por dos parámetros: uno vertical, de la altura de F0, y otro horizontal, de la alineación temporal con relación a la secuencia segmental. Las lenguas no solo difieren en cuanto a la distribución y el número de los eventos fonológicos, sino también en los detalles acústicos.

El enfoque de este estudio es la producción del foco estrecho en el español peninsular central. En los trabajos de orientación fonológica, el concepto de foco se define en términos de la prominencia prosódica, en el sentido de que este fenómeno se realiza fonológicamente por medio de una serie de eventos suprasegmentales, con sus correlatos acústicos correspondientes. Se suele distinguir entre el foco amplio y foco estrecho: mientras que el primero hace referencia a que toda la frase se considera centro de interés comunicativo, el segundo, a aquellos casos donde solo determinados elementos reciben énfasis especial.

Centrándose en la realización prosódica del foco estrecho, este estudio tiene como objetivo revelar y analizar las dificultades de adquisición entonativa mostradas por los estudiantes de ELE con el chino estándar como lengua materna. En particular, por medio de establecer comparaciones entre la producción de hablantes nativos y estudiantes, así como entre la realización de estudiantes con diferentes niveles de competencia de la lengua meta, este estudio pretende contestar las siguientes preguntas: ¿Cómo se caracteriza fonológicamente y fonéticamente la producción de los sinohablantes? ¿Muestran avances en el dominio de la entonación con el desarrollo de la competencia general de la lengua meta? Por último, en vista de las características intrínsecas de la entonación, ¿los detalles con respecto a la implementación fonética del contorno melódico se adquieren de modo paralelo a los rasgos fonológicos categóricos?

2. LA REALIZACIÓN PROSÓDICA DEL FOCO ESTRECHO EN EL ESPAÑOL

En español peninsular central, varios recursos prosódicos se emplean de modo sistemático para contribuir a la realización del foco estrecho, entre los cuales se destacan principalmente tres tipos: el realce fonológico del elemento focalizado, las inflexiones tonales en las fronteras inmediatamente precedente y siguiente del foco; y la compresión tonal durante la parte posterior del foco.

Numerosos estudios previos (de la Mota, 1995; Nibert, 2000; Face 2001, 2002; Hualde, 2002; Frota & Prieto, 2015; entre otros) demuestran el realce fonológico del elemento focalizado con el cambio de la F0 y la alineación temporal. En concreto, en las posiciones prenucleares, se ha demostrado que la ubicación temporal del pico asociado con el foco estrecho se diferencia significativamente del caso de foco amplio (con localización típica en la mitad de la sílaba tónica en vez de la sílaba postónica). En cambio, la elevación de F0, a pesar de la presencia frecuente, no resulta constante en todos los estudios previos. En la posición nuclear, se comprueba que el foco estrecho solo causa aumentos en la altura tonal sin modificaciones en la alineación, dado que, en dicha posición, el pico del enunciado neutro se ubica igualmente dentro de la sílaba acentuada.

Además de la estrategia más común de enfatizar el acento tonal asociado con el elemento focalizado, otra posibilidad consiste en realizar, de modo opcional, inflexiones tonales en las fronteras inmediatamente precedente y siguiente del foco. En cuanto al patrón de las inflexiones, la presencia de H- resulta típica en la frontera pre-foco, con la función pragmática de separar las informaciones conocidas de las nuevas (Hualde, 2002); en la frontera después del foco, se encuentran ambos patrones H- y L- en los datos previos de producción (Face, 2002). Además, las inflexiones tonales suelen estar acompañadas por una pausa de cierta duración con el fin de reforzar el fraseo prosódico (de la Mota, 1995; Face 2002).

Por último, cabe señalar que el foco estrecho también tiene un efecto en el movimiento global del contorno melódico (Navarro Tomás, 1944; de la Mota, 1995; Face 2002; entre otros). El fenómeno, conocido como compresión tonal, se caracteriza por la reducción del rango tonal a lo largo de la parte posterior del elemento focalizado. Como consecuencia, la curva melódica se convierte más plana en comparación con la curva en el contexto neutro, y se sitúa en un nivel relativamente bajo del registro individual.

3. MÉTODO EXPERIMENTAL

En el presente trabajo, se analiza la producción prosódica del foco por parte de los sinohablantes. El corpus utilizado se compone de diez enunciados que se pronuncian como respuesta a distintos contextos. Como muestra el ejemplo dado en (1), los enunciados tienen una estructura sintáctica de SVO y llevan tres palabras prosódicas. Además, se controla el número de sílabas átonas entre cada dos tónicas para evitar el posible efecto causado por *tonal crowding* (cfr. Face 2002).

(1) a. ¿Qué pasa?

El hermano encuentra dificultades.

b. ¿La hermana encuentra dificultades?

No. EL HERMANO encuentra dificultades.

c. ¿El hermano encuentra soluciones?

No. El hermano encuentra DIFICULTADES.

En (1) se ilustran también los tres contextos manipulados. En (1a), el enunciado se pronuncia como respuesta de la pregunta *¿Qué pasa?*, de modo que corresponde al foco amplio. En (1b) y (1c), las preguntas de orientación son respectivamente una interrogativa absoluta con la primera y la última palabra prosódica diferente de la original del enunciado, con el fin de obtener un mayor énfasis puesto en la palabra en cuestión.

Colaboraron en total cinco hablantes nativos como grupo de control ($Media_{edad}=21$, $DE=0$; 20% hombres) y veinte sinohablantes con diferentes niveles de competencia de ELE ($Media_{edad}=27.3$, $DE=3.26$; 30% hombres). Las grabaciones se realizaron en la sala insonorizada y con equipamientos profesionales del *Servei de Tractament de la Parla i del So* de la Universitat Autònoma de Barcelona. El programa de Praat (Boersma & Weenink 2019) se emplea para el tratamiento de sonidos, y STATA (StataCorp. 2019), para el procesamiento estadístico.

4. RESULTADOS

4.1. Acento tonal focalizado

En la Figura 1 se compara el nivel de F0 que alcanza el acento tonal en el contexto de foco amplio y de foco estrecho en el corpus analizado. En primer lugar, se observa que los aprendices aumentan un promedio de 20-30Hz para subrayar la prominencia, independientemente de la posición en que tiene lugar el elemento focalizado (con $p < 0.001$ en todas las posiciones). Al contrario, los datos extraídos de la producción de los nativos no sugieren una

preferencia hacia el uso de la F0, dado que la significativa elevación tonal se halla únicamente en la posición nuclear. Este resultado por parte de los nativos concuerda en general con lo que informan las investigaciones previas (de la Mota, 1995; Nibert, 2000; Face, 2002; Hualde, 2002; Vanrell et al., 2013).

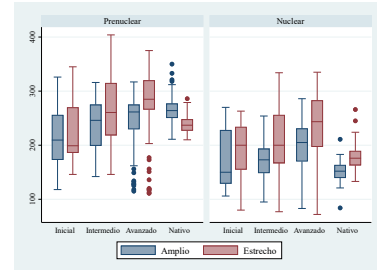


Figura 1: La comparación de la F0 de los picos tonales entre el contexto de foco amplio y de foco estrecho.

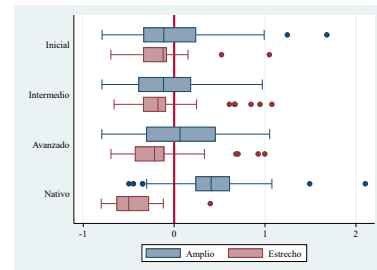


Figura 2: El tiempo relativo del pico tonal asociado con la primera palabra prosódica en contextos de foco amplio y de foco estrecho. El valor negativo denota un pico localizado dentro de la sílaba tónica, mientras que el valor positivo, un pico tardío encontrado en la sílaba postónica.

Además del cambio tonal causado por el foco estrecho, en el eje temporal, se ha observado el cambio de la alineación en la posición prenuclear. A fin de neutralizar las variaciones individuales, el análisis se ha basado en el tiempo relativo de los picos. Según lo ilustrado en la Figura 2, por un lado, la producción de los nativos coincide con las referencias en que el desplazamiento tonal típico en la declarativa neutra está inhibido en gran medida por la presencia del foco estrecho. En consecuencia, el pico tonal se encuentra regularmente dentro del dominio de la sílaba tónica y tiende a localizarse en la mitad de dicha sílaba. Por otro lado, las realizaciones de los aprendices, aunque carezcan de desplazamientos evidentes en las declarativas neutras, presentan una tendencia semejante de adelantamiento de los picos focalizados. Sin embargo, aunque el patrón fonológico coincide entre ambos grupos de informantes, la distribución de datos expone contrastes de carácter fonético: a diferencia de los hablantes nativos, que realizan el pico de F0 generalmente en la mitad de la sílaba tónica, la

alineación realizada por los sinohablantes, independientemente de su nivel lingüístico, tiende a aproximarse más a la frontera derecha de la sílaba tónica.

4.2. Tono de frontera intermedio precedente y siguiente del foco estrecho

La Figura 3 resume los diferentes patrones de tono de frontera intermedio respectivamente después y antes del elemento focalizado. Por lo general, el resultado ha sido bastante comparable entre los nativos y los estudiantes, con la mayor diferencia observada en la producción de los estudiantes de nivel inicial e intermedio. En concreto, cuando la primera palabra prosódica está focalizada, se hallan obviamente más separaciones prosódicas. En cuanto a los patrones emitidos, los estudiantes principiantes muestran más movimientos descendentes que los estudiantes con nivel intermedio, patrón que interrumpe por completo la pronunciación coherente.

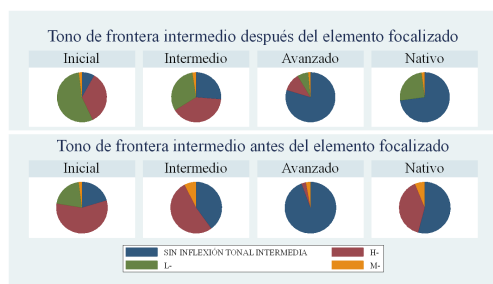


Figura 3: La presencia del tono de frontera intermedio después y antes del elemento focalizado.

4.3. Compresión tonal post-foco

Para evaluar la realización de la compresión tonal post-foco, se ha contrastado, entre el contexto de foco amplio y el de foco estrecho recaído en la primera palabra prosódica, los grados de cambio tonal entre cada dos de los picos tonales a lo largo del enunciado. Nótese que el cambio tonal analizado en este apartado también se ha experimentado una medición de relativización, que conlleva dividir la diferencia de F0 entre los dos picos por el rango tonal del enunciado neutro.

Como se puede observar en la Figura 4, el resultado ha sido equivalente entre ambos grupos de informantes. En particular, la notable diferencia de nivel tonal entre los primeros dos picos señala que el contorno melódico se caracteriza por un drástico descenso al finalizar el elemento focalizado. Además, el reducido grado de cambio entre los dos picos posteriores ilustra una configuración relativamente

plana de la curva post-foco. En conjunto, se demuestra que los estudiantes no han presentado dificultades relevantes en la producción de la compresión tonal, tanto desde la perspectiva fonológica como desde la perspectiva fonética.

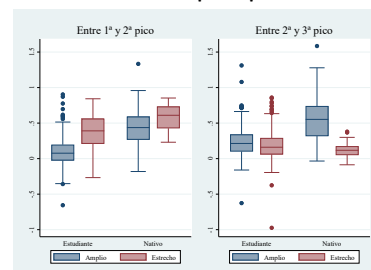


Figura 4: El cambio tonal entre cada dos de los picos tonales a lo largo del enunciado.

5. DISCUSIÓN Y CONCLUSIÓN

A lo largo del presente trabajo, por medio de un experimento de producción, hemos abordado la realización prosódica del foco estrecho en el español por parte de los estudiantes sinohablantes. En general, los resultados muestran que los estudiantes sinohablantes son capaces de emplear varios movimientos en el contorno melódico a la hora de poner énfasis especial en un determinado elemento del enunciado. No obstante, se hallan al mismo tiempo una serie de dificultades concretas en su expresión oral. A partir de la comparación entre la realización de patrones fonológicos categóricos y la precisión de implementación acústica, se concluye que los detalles fonéticos (en particular, la mayor dependencia de la elevación de F0, el modo inadecuado de alineación y el menor grado de compresión tonal post-foco) tienden a causar mayores dificultades a los sinohablantes y no se adquieren de modo paralelo que los patrones fonológicos. Desde otro punto de vista basado en la comparación entre estudiantes con distintos niveles de ELE, además de los problemas que se superan con el desarrollo de la competencia general, se destacan algunas dificultades, como el fallo con respecto a la alineación temporal de los picos, por su incidencia en el aprendizaje de todas las etapas. En futuras investigaciones, los hallazgos presentados nos motivan a reflexionar sobre el mejoramiento del proceso enseñanza-aprendizaje de la entonación de ELE orientado a sinohablantes, por ejemplo, el enfoque y método adoptado, la adecuación curricular de los conocimientos entonativos, así como la construcción de ejercicios y prácticas que favorezcan la adquisición de la entonación.

6. AGRADECIMIENTO

El estudio de doctorado en que se encuadra el presente trabajo está apoyado por la beca convocada por *China Scholarship Council* (Núm. 201708390050).

7. BIBLIOGRAFÍA

- Boersma, P., & Weenink, D. (2019). Praat: Doing phonetics by computer (Versión 6.0.53). Recuperado de <http://www.praat.org/>
- de la Mota, C. (1995). *La representación gramatical de la información nueva en el discurso* (Tesis doctoral). Universitat Autònoma de Barcelona.
- Face, T. L. (2001). Focus and early peak alignment in Spanish intonation. *Probus*, 13(2), 223-246.
- Face, T. L. (2002). *Intonational Marking of Contrastive Focus in Madrid Spanish*. Múnich: Lincom Europa.
- Frota, S., & Prieto, P. (2015). *Intonation in romance*. Nueva York: Oxford University Press.
- Hualde, J. I. (2002). Intonation in Spanish and the other Ibero-Romance languages: Overview and status quaestionis. En C. Wiltshire & J. Camps (Eds.), *Romance phonology and variation* (pp. 101–116). Ámsterdam: Benjamins.
- Ladd, D. R. (2008). *Intonational phonology* (2.^a ed.). Cambridge: Cambridge University Press.
- Navarro Tomás, T. (1944). *Manual de entonación española*. Nueva York: Hispanic Institute.
- Nibert, H. (2000). *Phonetic and phonological evidence for intermediate phrasing in Spanish intonation* (Tesis doctoral). University of Illinois at Urbana-Champaign.
- Pierrehumbert, J. (1980). *The phonetics and phonology of English intonation* (Tesis doctoral). Massachusetts Institute of Technology.
- StataCorp. (2019). *Stata Statistical Software: Release 16*. College Station, TX: StataCorp LLC.
- Vanrell, M. M., Stella, A., Gili Fivela, B., & Prieto, P. (2013). Prosodic manifestations of the Effort Code in Catalan, Italian and Spanish contrastive focus. *Journal of the International Phonetic Association*, 43(2), 195-220.

INTERROGATIVAS Y ENTONACIÓN EN EL ESPAÑOL HABLADO POR CHINOS

Tianshu Zhao, Dolors Font-Rotchés

Depart. Educació Lingüística i Literària. Facultat d'Educació. Universitat de Barcelona
ztsshengqing@hotmail.com, dolorsfont@ub.edu

ABSTRACT

Este estudio tiene como objetivo hacer una descripción de los rasgos melódicos de la entonación de las preguntas absolutas de la interlengua español hablado por chinos en habla espontánea.

Para tal finalidad, se ha establecido un corpus de 82 enunciados emitidos por cinco chinos estudiantes de Máster que hablan español. Después de analizar todos los enunciados con el método Análisis Melódico del Habla (AMH), hemos constatado que, dentro del corpus, solo hay 20 enunciados (un 24,4%) que siguen los patrones /+interrogativos/ del español peninsular, mientras que la mayoría, un 75,6%, se caracterizan por presentar melodías propias de los enunciados suspensivos y, en menor grado, declarativos y enfáticos. Estos resultados pueden dar explicación de algunas de las dificultades que tienen los sinohablantes al hablar español y de los malentendidos que pueden tener lugar en su comunicación con los nativos.

Palabras clave: análisis melódico, entonación, fonética, español hablado por chinos, habla espontánea

This paper aims to describe the melodic features of the intonation of absolute interrogatives in the Spanish interlanguage spoken by Chinese people in spontaneous speech.

A corpus of 82 statements was established, obtained from five Chinese Master's students who speak Spanish. After analysing all the statements using the Melodic Analysis of Speech (MAS), it was found that within the corpus, there are only 20 statements (24.4%) that follow the /+ interrogative / patterns of Peninsular Spanish, while the majority, 75.6%, is characterised by the typical melodies of the suspended statements and, to a lesser degree, declarative and emphatic ones. These results explain the difficulties that Chinese people have when speaking Spanish and the misunderstandings which they can cause when talking to native speakers.

Keywords: melodic analysis, intonation, phonetics, spanish spoken by Chinese, spontaneous speech

1. INTRODUCCIÓN

Debido al alto interés por aprender español en China, cada vez más colegios y universidades han comenzado a ofrecer cursos y grados de español. En estos momentos, en los que las relaciones comerciales entre China y los países de habla hispana han crecido exponencialmente, es crucial para los chinos aprender español, para poder comunicarse con los nativos hablantes de castellano con fluidez y eficacia, dado que el aprendiente chino de español presenta un marcado "acento extranjero" (Cantero, 2002; Liu, 2005) en su pronunciación que, en muchas ocasiones, provoca malentendidos en la comunicación con los nativos de español.

En cuanto a las investigaciones llevadas a cabo sobre la pronunciación del español hablado por chinos, existen algunos trabajos que se han centrado generalmente en los sonidos (Lu, 1991; Shen y Li, 2014; Yue, 2016; Chen, 2017; Cao y Rius-Escudé, 2019) o en la entonación de los hablantes de chino taiwaneses (Liu, 2005; Cortés, 2006). Existen, también, algunas investigaciones sobre la enseñanza de las interrogativas (Planas, 2012), o la percepción de los enunciados interrogativos de duda, enfado (Planas, 2015) o peticiones (Herrero et al. 2020).

Ante el interés por aprender español y la escasez de estudios sobre los rasgos melódicos de la interlengua español hablado por chinos, de China Continental, que son de gran utilidad para abordar el desarrollo de propuestas didácticas en manuales de aprendizaje, en el presente trabajo nos proponemos

describir los rasgos de la entonación de las preguntas absolutas del español hablado por chinos en habla espontánea.

2. METODOLOGÍA

2.1. Establecimiento del corpus

Para nuestro estudio, hemos seleccionado a cinco informantes, cuatro chicas y un chico en edades comprendidas entre 22 y 27 años de China Continental, que estudian un Máster en Barcelona. En cuanto a su procedencia, tres son del norte (provincias de Hebei y Hei Longjiang) y solo hablan chino mandarín; y dos del sur (provincias de Sichuan y Jiangsu), que son bilingües (hablan chino mandarín y el dialecto de su zona).

Para llevar a cabo esta investigación se realizaron tres grabaciones. En total, se obtuvieron un total de más de 300 minutos, de los que se extrajeron y segmentaron 82 preguntas absolutas con el programa Praat (Boersma y Weenink, 1992-2019).

2.2. El método Análisis Melódico del Habla (AMH)

Hemos analizado los enunciados siguiendo los criterios del método *Análisis Melódico del Habla* (AMH), expuesto de forma detallada en Cantero (2002), revisado y ampliado en Font-Rotchés (2007), y establecido en forma de protocolo en Cantero y Font-Rotchés (2009, 2020). Según el protocolo de este método de análisis acústico, el procedimiento se divide en tres etapas: la fase de identificación y segmentación, la fase acústica y la fase perceptiva. Sin embargo, para esta investigación, solo se necesitan las dos primeras.

Para hacer una comparación entre el español hablado por chinos y el español peninsular, nos basamos en los doce patrones melódicos del español peninsular establecidos por Cantero y Font-Rotchés (2007) y, más específicamente, en los cuatro interrogativos (Font-Rotchés y Mateo, 2013, 2017).

3. ANÁLISIS DE RESULTADOS

A continuación, exponemos los resultados que hemos obtenidos en la investigación en dos subapartados: el de los contornos melódicos que siguen la estructura de patrones interrogativos y el de los que contienen estructuras no interrogativas.

3.1. Patrones interrogativos

En el corpus, solo 20 enunciados, un 24,4%, presentan los rasgos melódicos de los patrones interrogativos. De todos ellos, 7 corresponden al patrón melódico II; 3, al patrón III; 3, al patrón IV y 7, al XIII.

3.1.1 Patrón melódico II. IF ascendente (+70%)

El patrón melódico II es el patrón típico de la entonación interrogativa, cuyo primer pico (1r P) tiene lugar en la primera vocal tónica del contorno, situada en la culminación de un ascenso de entre un 20% y un 40%; el cuerpo se caracteriza por una declinación suave y constante hasta llegar a la última sílaba tónica o núcleo, desde donde empieza la inflexión final (IF), que debe presentar un ascenso igual o superior al 70%. De este grupo, tenemos 7 enunciados, un 8,5% del corpus, de los cuales, 6, un 85,7% no tienen primer pico (SP), y el único enunciado que sí lo tiene se encuentra desplazado a la vocal átona posterior (AP). No obstante, solo con el rasgo de la inflexión final se identifica a estos contornos como interrogativos.

En la Fig. 1 se encuentra un contorno sin primer pico y una inflexión final de 85,8%.

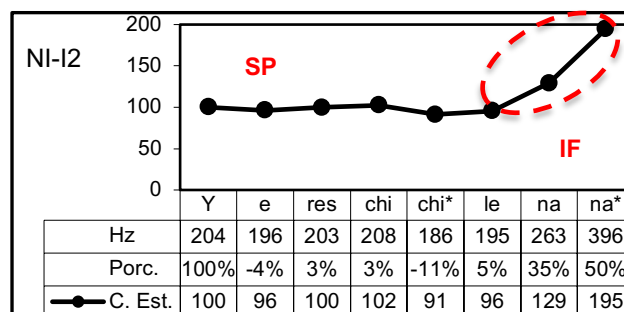


Figura 1: Análisis del gráfico NI-I2 ¿Y eres chilena?

3.1.2. Patrón melódico III. IF ascendente (40% ~70%)

El patrón melódico III se define por un primer pico desplazado a una vocal átona posterior, un cuerpo en declinación suave y constante y una inflexión final ascendente de entre 40% y 70%. Los 3 enunciados obtenidos en el corpus, un 3,7%, siguen este patrón.

Como se muestra en la Figura 2, el primer pico cae en la vocal *-je*, de *personaje*. Se trata de un primer pico desplazado a una vocal átona posterior y tiene una inflexión final acumulada de 46,3%.

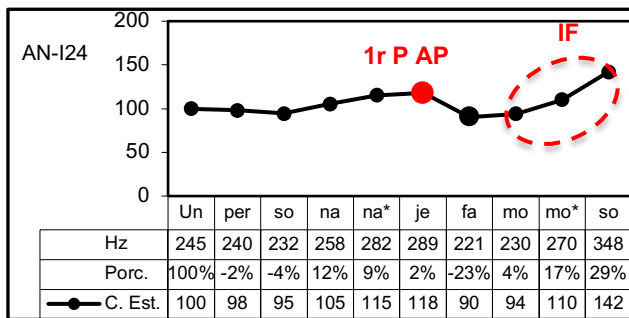


Figura 2: Análisis del gráfico AN-I24 ¿Un personaje famoso?

3.1.3. Patrón melódico IV. IF circunfleja ascendente-descendente

El patrón melódico IV se caracteriza por presentar una inflexión final circunfleja ascendente-descendente, en la que el ascenso de la inflexión final debe alcanzar un 30% seguido de un descenso (Font-Rotchés y Mateo, 2017). En nuestro corpus, encontramos 3 enunciados, uno similar a este patrón (véase Fig. 3) y otros en los que el ascenso en la inflexión final es inferior al 30%.

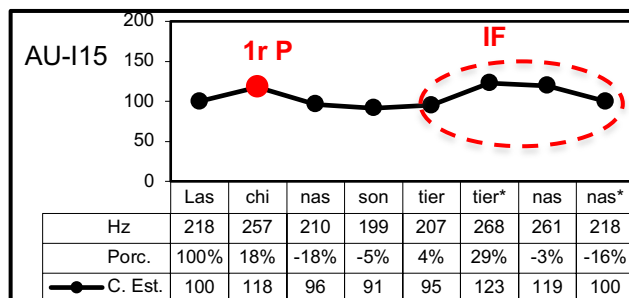


Figura 3: Análisis del gráfico AU-I15 ¿Las chicas son tiernas?

3.1.4. Patrón melódico XIII. Cuerpo e inflexión final ascendentes

Normalmente, el patrón melódico XIII no tiene primer pico, y el cuerpo y la inflexión final son ascendentes, cuyo ascenso total tiene que ser igual o superior a un 60%. Encontramos 7 enunciados en el corpus, un 8,5%, que corresponden a dicho patrón.

A modo de ejemplo, el contorno AN-I23 (véase en la Fig. 4.) sigue a este patrón, porque el cuerpo y la inflexión final son ascendentes y el ascenso suma en total un 76%.

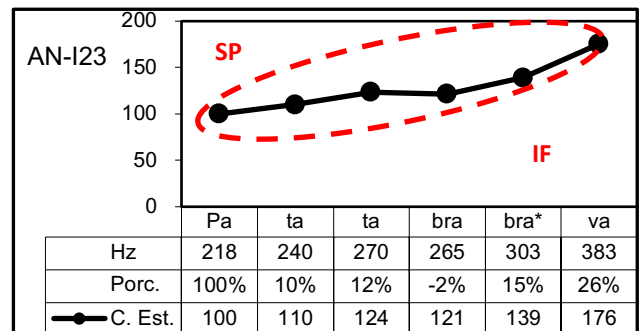


Figura 4: Análisis del gráfico AN-I23 ¿Patata brava?

3.2. Patrones no interrogativos

En nuestro corpus, hay 62 enunciados que no siguen una entonación interrogativa, un 75,6%. La mayoría, 53 (64,6%), presentan rasgos de la entonación suspendida; 5 (6,1%), de la entonación neutra; y 4 (4,9%) de la entonación enfática.

3.2.1. La entonación neutra: Patrón melódico I. IF ascendente (10%~15%) y descendente (10%~30%)

El patrón melódico I del español es el propio de la entonación neutra y se encuentra en las declarativas. Se caracteriza por un primer pico en la primera vocal tónica del contorno, que es el punto más alto; un cuerpo con una declinación constante hasta el núcleo o última sílaba tónica y una inflexión final con un descenso (entre un 10~30%) o un ascenso de entre un 10~15%. Entre los 5 enunciados (un 6,1%) de la interlengua que corresponden al patrón I, solo en 2 el primer pico cae en la primera vocal tónica, 2 no tienen primer pico y 1 cuenta con el primer pico desplazado. Constatamos, pues, que pese al intento de producir preguntas en los 5 enunciados, lo que los informantes producen es una declarativa. Veamos un ejemplo en la Fig. 5: no tiene primer pico y su inflexión final es de un 13%.

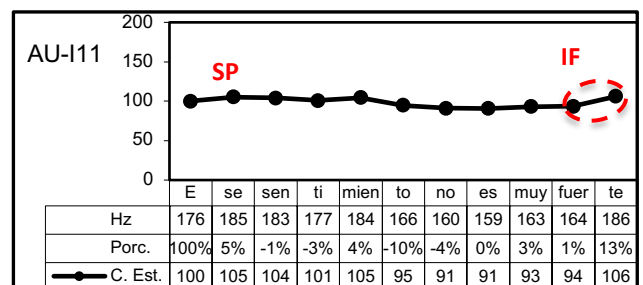


Figura 5: Análisis del gráfico AU-I11 ¿Ese sentimiento no es muy fuerte?

3.2.2. La entonación suspendida: Patrones melódicos V. Sin IF y VI. IF ascendente (15%~70%)

El patrón melódico V se caracteriza por ser suspendido y su rasgo principal es que carece de inflexión final; cuenta con un primer pico situado en la primera vocal tónica del contorno y un cuerpo en declinación suave y constante. Hemos clasificado los contornos con inflexión final de entre +10% y -10% en este apartado, puesto que presentan una terminación plana, que se percibe inacabada.

Encontramos 5 enunciados (un 6,1%) que siguen este patrón: 2 no tienen primer pico, 2 tienen el primer pico desplazado y 1 en la primera vocal tónica. En la Figura 6, aportamos un contorno que no tiene primer pico y que su inflexión final presenta un ascenso leve (+8.6%), característico de la entonación suspendida del español.

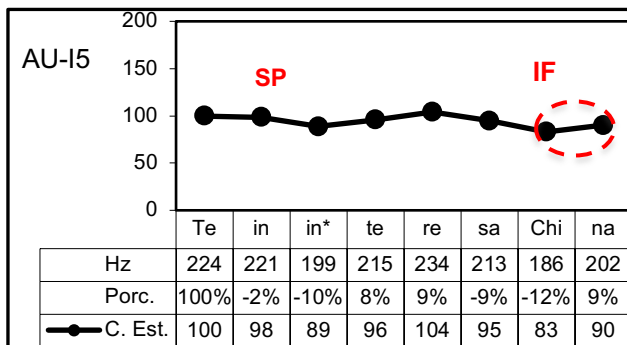


Figura 6: Análisis del gráfico AU-15 ¿Te interesa China?

El patrón melódico VI, cuya entonación responde a un enunciado suspensivo, se presenta con dos variaciones: el VIa y el VIb. VIa presenta un ascenso hasta el primer pico o primera vocal tónica de un 40% como máximo; una declinación suave y constante en el cuerpo; y una inflexión final ascendente entre 15 y 70% (Véase el contorno de la Fig. 7). En cambio, el VIb, presenta un primer pico desplazado a una vocal átona posterior y una inflexión final menos ascendente, entre 15% y 40%.

Uno de los rasgos llamativos es que, a diferencia del español peninsular, 24 enunciados de los 48 clasificados en este grupo, un 50%, no tienen primer pico. A pesar de este rasgo, las inflexiones finales permiten identificarlos como suspendidos.

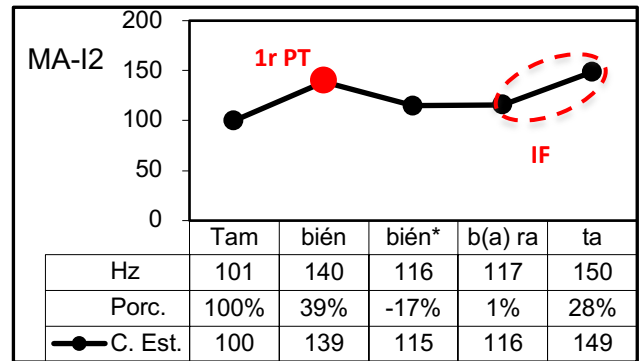


Figura 7: Análisis del gráfico MA-12 ¿También barata?

3.2.3. La entonación enfática: Patrones melódicos VII. IF de núcleo elevado y X. IF ascendente-descendente

El patrón melódico VII se caracteriza por un primer pico desplazado a la vocal átona posterior; un cuerpo en declinación que se resitúa al alza para posibilitar que el núcleo alcance el nivel del primer pico; y una inflexión final descendente hasta el final del contorno. En nuestro corpus, existen 3 casos, un 3,7%, que se asemejan a dicho patrón, es decir, no algunos rasgos son distintos, como la ausencia de primer pico o el ascenso leve en la inflexión final que no retorna al punto de inicio. Véase el contorno de la Figura 8, sin primer pico, el cuerpo bastante plano con un énfasis de palabra en *-nas*, un ascenso hasta el núcleo *ca-* muy leve (13,4%) y un descenso posterior de 3,6%. Observamos que los rasgos de estos contornos presentan una leve tendencia a expresar énfasis, y están lejos de presentar una melodía de pregunta.

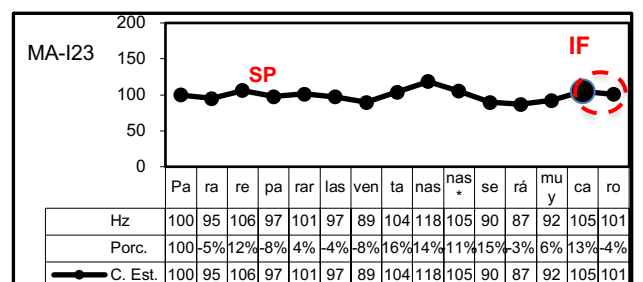


Figura 8: Análisis del gráfico MA-123 ¿Para reparar las ventanas será muy caro?

El patrón melódico X, con sus dos variantes, Xa con inflexión final ascendente-descendente y Xb con inflexión final descendente-ascendente, tiene como rasgo que es enfático. Ambas variantes tienen en común el primer pico desplazado a la siguiente vocal átona.

En el corpus, hemos encontrado 1 enunciado similar al patrón melódico Xb (véase en la Fig. 9.). El contorno no tiene primer pico ni anacrusis pero

presenta una inflexión final descendente-ascendente. Lo más probable es que se interpretara como un enunciado enfático en vez de una pregunta.

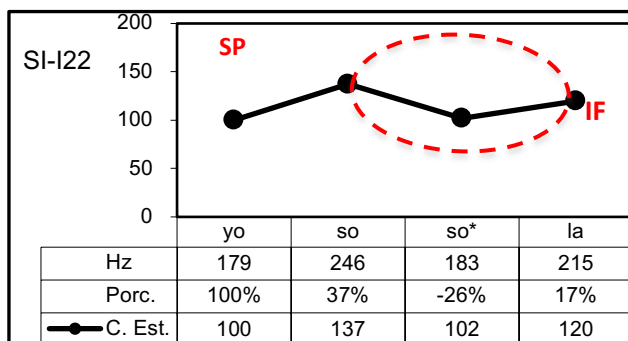


Figura 9: Análisis del gráfico SI-I22 ¿Yo sola?

4. CONCLUSIONES

Después del análisis y comparación de los contornos emitidos por chinos con los patrones del español peninsular, hemos obtenido las siguientes conclusiones:

1. La mayoría no presentan una melodía propia de las preguntas. De los 82 enunciados del corpus, solo 20 (24,4%) se corresponden a un patrón interrogativo: 7 contornos al patrón melódico II (IF ascendente +70%); 3 contornos al P. III (IF ascendente +40% ~ +70%); 3 al patrón IV (IF ascendente-descendente) y 7 contornos al patrón melódico XIII (cuerpo e IF ascendentes).
2. Los 62 enunciados restantes (75,6%) presentan mayoritariamente un patrón suspendido, un 64,6%, o en menor grado, un patrón neutro o enfático. Por consiguiente, en la mayoría de situaciones comunicativas, se podrían dar malentendidos por el hecho de que las preguntas no serían interpretadas como tales, sino como enunciados suspensos o enfáticos.
3. Alrededor de un 50% de los enunciados no presentan primer pico.
4. Respecto a la inflexión final, más de la mitad de los contornos del corpus presentan inflexiones inferiores a un 40%. Esto explicaría por qué no se interpretarían como preguntas.

Este trabajo nos indica unas tendencias de la entonación del español hablado por chinos que queremos ahondar en un futuro ampliando el número de informantes, las zonas de donde proceden y la tipología de enunciados. De la misma manera, consideramos fundamental continuar investigando en la interlengua español hablado por chinos, puesto que en la medida que se obtengan más conocimientos respecto a este tema, los docentes que enseñan español a chinos podrán

ayudar más a sus alumnos, con el fin de que ellos emitan enunciados con una entonación que responda de forma efectiva a sus intenciones comunicativas.

5. BIBLIOGRAFÍA

- Boersma, P., y Weenink, D. (1992-2019). Praat: doing phonetics by computer [Computer program]. Version 6.1.02. <http://www.praat.org/>.
- Cantero, F. J. (2002). *Teoría y análisis de la entonación*. Barcelona: Edicions de la Universitat de Barcelona.
- Cantero, F. J., y Font-Rotchés, D. (2007). Entonación del español peninsular en habla espontánea: patrones melódicos y márgenes de dispersión. *Moenia*, 13, 69-92.
- Cantero, F. J., y Font-Rotchés, D. (2009). Protocolo para el análisis melódico del habla. *Estudios de Fonética Experimental*, 18, 17-32.
- Cantero, F. J. y Font-Rotchés, D. (2020). Melodic Analysis of Speech (MAS). *Phonetics of Intonation*. En Abasolo, J.; de Pablo, I.; Ensunza, A.: Contributions on education (pp. 20-47). Universidad del País Vasco.
- Cao, Y., y Rius-Escudé (2019): Caracterización acústica de las vocales del español hablado por chinos. *Phonica*, 15, 3-22. Doi: 10.1344/phonica.2019.15.3-22
- Chen, X. (2017). 中国西班牙语专业学生语音层面语用失误及原因分析 (Los errores pragmáticos de los estudiantes chinos de español en el habla y sus causas). *Journal of Guangdong University of Foreign Studies*, 28, 54-62.
- Cortés, M. (2006). Análisis acústico de la transferencia de rasgos del sistema tonal chino al habla en español como lengua extranjera. *Estudios de Fonética Experimental*, XV, 43-65.
- Font-Rotchés, D. (2007). *L'entonació del català*. Barcelona: Publicacions Abadia de Montserrat (Biblioteca Milà i Fontanals, 53).
- Font-Rotchés, D., y Mateo, M. (2013). Entonación de las interrogativas absolutas del español peninsular del sur en habla espontánea. *Onomázein*, 28, 256-275. <https://doi.org/10.7764/onomazein.28.17>.
- Font-Rotchés, D., y Mateo, M. (2017). Melodías para confirmar, preguntar, sugerir o pedir en español. *Phonica*, 13, 49-67. <https://doi.org/10.1344/phonica.2017.13.49-67>
- Herrero, C., Planelles, M. y Alhmoud, Z. (2020). Perception of L2 Spanish polite requests and impolite commands by Chinese migrant workers living in Spain. *Lengua y migración* 12:2: 65-85.
- Liu, Y. H. (2005). *La entonación del español hablado por taiwaneses*. Tesis doctoral. Biblioteca Phonica, 2. http://www.publicacions.ub.edu/revistes/phonica-biblioteca/esp_taiw/esp_taiw.pdf. Consultado el 29.01.2020.

- Lu, J. S. (1991). 汉语和西班牙语语音对比 (Comparación de la fonética en chino y español). *Journal of Shanghai International Studies University*, 58-73.
- Planas Morales, S. (2012): Consideraciones y recursos para la enseñanza de oraciones interrogativas españolas en hablantes chinos. En: Cabedo, A.; Infante, P.: *Lingüística XL. El lingüista del siglo XXI*. Madrid: SEL ediciones, 63-70.
- Planas Morales, S. (2015): Percepción de los enunciados interrogativos de duda y de enfado sin apoyo visual en alumnos chinos de ELE. *Revista española de lingüística*, 45, 153-174.
- Shen, Y. B., y Li, M. (2014). A Study of Negative Transfer of Chongqing Dialect on Spanish Phonetics Learning and Countermeasures. *Foreign Language Teaching and Research (bimonthly)*, 30, 175-179.
- Yue, Y. Y. (2016). 汉语和英语在西班牙语语音教学阶段的迁移 (La interferencia lingüística de chino e inglés en la etapa de enseñanza de la fonética del español). *Contemporary Education Research and Teaching Practice*, 171-172.

