



**José María Lahoz-Bengoechea
and Rubén Pérez Ramón (Eds.)**

SUBSIDIA:

Tools and resources for speech sciences

José María Lahoz-Bengoechea
and Rubén Pérez Ramón (Eds.)

Subsidia: Tools and Resources for Speech Sciences

Subsidia: Herramientas y Recursos para las Ciencias del Habla

Libro publicado por la Universidad de Málaga (UMA)

Subsidia: Tools and Resources for Speech Sciences is a scientific publication arising from the organisation of a congress with the same name, carried out in the city of Málaga (Spain) in June, 2017. Its main goal is to give voice to tools and resources developed with the aim of facilitating research in the field of speech sciences. This framework embraces subjects such as phonetics, experimental phonetics, phonology, discourse analysis or dialectology, among others. This book, outcome of the collaboration of expert researchers on their respective areas, aims to be an aid to the scientific community in the sense that it compiles and depicts a series of materials that, we hope, may result beneficial to keep moving forward in the research.

The papers collected in this volume are a selection of those submitted to the above-mentioned congress, and have undergone peer review.

Subsidia: Herramientas y Recursos para las Ciencias del Habla es una publicación científica resultante de la organización del congreso del mismo nombre desarrollado en la ciudad de Málaga (España) en junio del año 2017. Su objetivo principal es dar a conocer herramientas y recursos desarrollados con el objetivo de facilitar la investigación en el campo de las ciencias del habla. Dentro de este marco se engloban disciplinas tan variadas como la fonética experimental, la fonología, el análisis del discurso o la dialectología, entre otras. Este libro, resultado de la colaboración de investigadores expertos en sus respectivas áreas, pretende ser una ayuda a la comunidad científica en tanto en cuanto recopila y describe una serie de materiales que esperamos resulte provechoso para continuar avanzando en la investigación.

Los artículos recogidos en este volumen son una selección de los que se presentaron en dicho congreso y han pasado una evaluación por pares.

EDITORIAL BOARD / CONSEJO DE REDACCIÓN

Chair / Presidenta del congreso: Juana Gil Fernández

Co-chair / Copresidenta del congreso: Inés Carrasco Cantos

Editors / Editores: José María Lahoz-Bengoechea & Rubén Pérez Ramón

SCIENTIFIC COMMITTEE / COMITÉ CIENTÍFICO

Chair of the scientific committee / Presidente del comité científico: Joaquim Llisterri Boix

TECHNICAL EDITION / EDICIÓN TÉCNICA

© Universidad de Málaga, 2019

© Authors on their chapters / Los autores de sus respectivos capítulos

Cover design / Diseño de la cubierta: Rubén Pérez Ramón & José María Lahoz-Bengoechea.

This is an open-access publication distributed under the terms of the Creative Commons Attribution-Non Commercial (by-nc) 3.0.

Esta es una publicación de acceso abierto distribuida bajo los términos de la licencia Creative Commons Reconocimiento - No comercial (by-nc) 3.0.

The opinion and facts stated in each article are the exclusive responsibility of the authors. The Universidad de Málaga is not responsible in any case of the credibility and authenticity of the works.

The manuscripts published in this book are the property of the Universidad de Málaga, and quoting this source is a requirement for any partial or full reproduction.

Las opiniones y hechos consignados en cada artículo son de exclusiva responsabilidad de sus autores. La Universidad de Málaga no se hace responsable, en ningún caso, de la credibilidad y autenticidad de los trabajos.

Los originales publicados en este libro son propiedad de la Universidad de Málaga, siendo necesario citar la procedencia en cualquier reproducción parcial o total.

Subsidia: Tools and Resources for Speech Sciences

CREDITS

ARTICLES

| | |
|---|-------|
| Bringing together tools and resources for speech sciences JOSÉ MARÍA LAHOZ-BENGOCHEA & RUBÉN PÉREZ RAMÓN | p. 1 |
| Aalto Aparat: A freely available tool for glottal inverse filtering and voice source parametrization PAAVO ALKU, HILLA POHJALAINEN, & MANU AIRAKSINEN | p. 5 |
| The phonetic approach of voice qualities: challenges in corresponding perceptual to acoustic descriptions ZULEICA CAMARGO, SANDRA MADUREIRA NATHALIA DOS REIS, & ALBERT RILLIARD | p. 11 |
| The analysis of facial and speech expressivity: tools and methods SANDRA MADUREIRA & MARIO AUGUSTO DE SOUZA FONTES | p. 19 |
| TransText, un transcriptor fonético automático de libre distribución para español y catalán JUAN MARÍA GARRIDO, MARTA CODINA, & KIMBER FODGE..... | p. 27 |
| dVoice: doing phonetics by smartphones FRANCESCO CUTUGNO, ENRICO LEONE, ANTONIO ORIGLIA, & RENATA SAVY .. | p. 33 |
| MWN-E: a graph database to merge morpho-syntactic and phonological data for Italian ANTONIO ORIGLIA, GIULIO PACI, & FRANCESCO CUTUGNO | p. 37 |
| Methodological issues in the assessment of cross-language phonetic similarity JULI CEBRIAN..... | p. 47 |
| Exploiting a multimedia academic corpus for learning Spanish as a Foreign Language: <i>Video4ELE-UNED</i> VICTORIA MARRERO & VÍCTOR FRESNO | p. 55 |
| Plataforma interactiva para el autoaprendizaje de la pronunciación inglesa: la enseñanza de la entonación EVA ESTEBAS VILAPLANA..... | p. 59 |

Subsidia: Tools and Resources for Speech Sciences

| | |
|---|--------|
| <i>Dumloquor hora fugit: aprendizaje autónomo y autorregulado de la pronunciación del catalán a través de las <i>Guies de pronunciació del català</i></i> JOSEFINA CARRERA-SABATÉ, JESÚS BACH MARQUÉS, & MAR MIR CAMPILLO..... | p. 65 |
| Explicit and implicit training methods for the learning of stress contrasts in Spanish SANDRA SCHWAB & VOLKER DELLWO | p. 75 |
| Els sons del català, una herramienta digital para aprender fonética y fonología catalanas en la red CLÀUDIA PONS-MOLL & JOSEFINA CARRERA-SABATÉ..... | p. 81 |
| Bayesian strategies for likelihood ratio computation in forensic voice comparison with automatic systems DANIEL RAMOS, JUAN MAROÑAS-MOLANO, & ALICIA LOZANO-DIEZ | p. 89 |
| EMULANDO: Corpus de habla con acento no nativo auténtico y disimulado JOSÉ MARÍA LAHOZ-BENGOECHEA, JUANA GIL FERNÁNDEZ, & CLARA LUNA GARCÍA GARCÍA DE LEÓN | p. 97 |
| Detecting neuromotor disease in speech articulation PEDRO GÓMEZ, DANIEL PALACIOS, ANDRÉS GÓMEZ, CRISTINA CARMONA, ANA R. LONDRAL, VICTORIA RODELLAR, VÍCTOR NIETO, MIGUEL A. FERRER, & AGUSTÍN ÁLVAREZ..... | p. 103 |
| Perceptual experiments in Praat: beyond the standards RUBÉN PÉREZ RAMÓN..... | p. 109 |
| VILE-P: un corpus para el estudio prosódico de la variación inter e intralocutor JOAQUIM LLISTERRI, MARÍA J. MACHUCA, & ANTONIO RÍOS..... | p. 117 |
| Génesis y aspectos fundamentales de ProDis ANA MARIA FERNÁNDEZ PLANAS, PAOLO ROSEANO, WENDY ELVIRA-GARCÍA, & SIMONE BALOCCO..... | p. 125 |
| FonetiToBI, una herramienta para la anotación prosódica automática de corpus WENDY ELVIRA-GARCÍA & JUAN MARÍA GARRIDO..... | p. 133 |

Génesis y aspectos fundamentales de ProDis

Ana Maria Fernández Planas¹, Paolo Roseano¹, Wendy Elvira-García¹ y Simone Balocco¹

¹ Universitat de Barcelona
e-mail: anamariafernandezp@ub.edu

Citation / Cómo citar esta publicación: Fernández Planas, A. M., Roseano, P., Elvira-García, W., & Balocco, S. (2019). Génesis y aspectos fundamentales de ProDis. In J. M. Lahoz-Bengoechea & R. Pérez Ramón (Eds.), *Subsidia. Tools and resources for speech sciences* (pp. 125–132). Málaga: Universidad de Málaga.

RESUMEN: El objetivo de este trabajo consiste en presentar ProDis, una herramienta informática para el análisis dialectométrico de la entonación creada por el equipo del *Laboratori de Fonètica* de la *Universitat de Barcelona*. El trabajo presenta un breve estado de la cuestión sobre los sistemas de dialectometría disponibles en la actualidad y se hace especial hincapié en los sistemas capaces de tratar datos prosódicos numéricos y en la necesidad del Atlas Multimedia de la Entonación del Espacio Románico (AMPER) de crear una herramienta propia que sirva para gestionar los datos recogidos mediante el corpus fijo del proyecto en cuestión. Asimismo, se detallan los sistemas anteriores que han perimido el nacimiento de la herramienta. Posteriormente, se realiza una presentación de ProDis en la que se abordan el método que usa para calcular las distancias prosódicas, las salidas del programa y la información que se puede obtener de dicho análisis.

Palabras clave: dialectometría; entonación; lenguas romances.

ABSTRACT: The aim of this paper is to present ProDis, a software for the dialectometric analysis of intonation. In the first part of the article we present a state of the art of currently available dialectometrical tools. Since the prosodic data that are collected within the *Atlas Multimedia de la Entonación del Espacio Románico* (AMPER) are numeric, we pay special attention to the tools that can dialectometrize numeric prosodic data. After resuming the features of the existing tools that, to a certain extent, can carry out the dialectometric analyses of numeric data, we present more in detail the characteristics of ProDis, the prosodic dialectometrical tool created at the Phonetics Laboratory of the University of Barcelona. The aspects of ProDis that are described are the method used to calculate prosodic distances, the outputs of the program, and the information that can be obtained thanks to the prosodic dialectometry.

Keywords: dialectometry; intonation; Romance languages.

1. INTRODUCCIÓN

Los atlas lingüísticos clásicos constituyen una fuente maravillosa de datos fonéticos, morfológicos, sintácticos y, sobre todo, léxicos (recuérdese por ejemplo el ALiR —Atlas Linguistique Roman—, el ALPI —Atlas Lingüístico de la Península Ibérica—, el ALE —Atlas Linguistique de l'Europe—, o el ALAC —Atlas Lingüístico de América Central—, por ejemplo, entre otros muchos). El estudio de sus datos fue el punto de partida para el establecimiento de isoglosas y fronteras dialectales que clasificaban los puntos de encuesta en grupos a partir de rasgos considerados muy relevantes de forma cualitativa por los investigadores.

Desde los años 70 y 80 del siglo XX se ha producido un paso natural en el desarrollo dialectal de la mano de la llamada dialectometría, que pretende establecer agrupaciones entre la masa de datos empíricos obtenidos en grandes bases de datos a partir de criterios

cuantitativos y de procedimientos estadísticos objetivos. El término dialectometría se debe a Séguy (1973), uno de los padres de dicha disciplina junto con Guiter, aunque fue Goebel quien le dio un impulso definitivo. Goebel (1981) define la dialectometría como una alianza metodológica entre la geolingüística y la taxonomía numérica como disciplina matemática. Exactamente, el autor lo expone de forma sintética de la siguiente manera: dialectometría = geografía lingüística + taxonomía numérica (Goebel, 1981, p. 349).

Ciertamente, lo que los estudios dialectométricos pretenden es utilizar una enorme cantidad de datos que se han generado a través de los estudios dialectológicos y los atlas lingüísticos para establecer agrupaciones entre la masa de datos empíricos disponibles y obtener una distribución en el espacio virtual de los datos (Fernández Planas, Roseano, Martínez Celadrán y Romera Barrios, 2011, p. 145), o también en forma de agrupaciones reflejadas en dendrogramas. Sus resultados permiten una rápida asociación entre los

elementos considerados a partir de su cercanía o su lejanía —es decir, de sus semejanzas o de sus diferencias— y posibilitan condensar una gran cantidad de información cuantitativa en un espacio relativamente reducido.

La dialectometría no pretende eliminar el estudio dialectológico tradicional, sino que busca completarlo y erigirse como una herramienta esencialmente útil cuando se manejan cantidades enormes de datos a partir de grandes bases. Sin embargo, ofrece ciertas ventajas respecto a la dialectología tradicional: (1) permite gestionar sin gran esfuerzo por parte del investigador grandes cantidades de datos, de donde se puede inferir que permite llegar a conclusiones estadísticamente fiables, en el sentido de que van más allá de las meras intuiciones de los investigadores; (2) no hay apriorismos en el tratamiento de los datos, ya que el estudio se centra en valoraciones cuantitativas y no cualitativas. Este hecho supone un cambio radical respecto a la dialectología tradicional, más bien cualitativa, y constituye el fundamento de la reticencia de algunos autores; y (3) la forma de presentación de los datos (análisis de clúster en dendrogramas y escalamiento multidimensional, básicamente) es totalmente visual y favorece una comprensión bastante rápida de los hechos.

Por lo que respecta a las lenguas romances, el método se ha aplicado principalmente a las áreas lingüísticas del ladino (Goebel, 1993; Bauer, 2005), el italiano (Bauer, 2003), el francés (Séguy, 1973; Verlinde, 1988; Goebel, 1987; Goebel, 2003), el gallego (Álvarez Blanco, Dubert, y Sousa, 2006; Sousa, 2006; Saramago, 2002), el bable (D'Andrés Díaz, Álvarez-Balbuena García, y Suárez Fernández, 2003) o el catalán (Clua, 2004; Polanco, 1992). Fuera de la Rumania se utiliza también en estudios dialectológicos de lenguas como el holandés (Heeringa y Nerbonne, 2001), el inglés (Goebel y Schiltz, 1997), o el euskara (Aurrekoetxea, 1992). Normalmente se ha trabajado con datos fonético-segmentales, morfológicos o léxicos, como hemos dicho. El estudio de los aspectos prosódicos de diferentes variedades se ha trabajado muchísimo menos.

En el seno de macroproyecto AMPER, Atlas Multimedia de Prosodia del Espacio Románico (Contini, 1992; Contini *et al.*, 2002; Romano y Contini, 2001; Contini, Lai y Romano, 2002; Romano, 2003; Fernández Planas, 2005), tras tener muy avanzada una enorme base de datos prosódicos acústicos, se impone trabajar en la comparación y clasificación de las variedades románicas. De hecho, que en el marco de AMPER se llegara a utilizar la dialectometría era también un desarrollo natural y un paso esperable. Así pues, el llamado “corpus fijo” en el proyecto constituye un terreno óptimo para el estudio dialectométrico. AMPER trabaja con habla cercana a habla de laboratorio en el llamado corpus fijo que cuenta con frases enunciativas e interrogativas absolutas que presentan la misma estructura SVO y número de sílabas en todas las lenguas (más o menos), con todas las combinaciones acentuales posibles en todas las

posiciones de la frase salvo en el verbo, con dos hablantes por punto de encuesta (como mínimo) que repiten cada uno de ellos tres veces cada frase.

2. LA DIALECTOMETRÍA Y SUS DATOS

En realidad, cuando hablamos de dialectometría no estamos refiriéndonos a una única técnica, sino a un paraguas metodológico que incluye técnicas distintas que trabajan con el mismo objetivo (la aplicación de técnicas estadísticas a grandes bases de datos para averiguar cómo se agrupan a partir de las distancias que mantienen los elementos entre sí a partir de sus características) pero con diferentes algoritmos y con diferentes tipos de datos.

Desde sus inicios, igual que la dialectología clásica, la dialectometría ha usado datos fonético-fonológicos segmentales, morfológicos, léxicos o sintácticos que se almacenan en bases de datos alfabéticos. Desde el punto de vista estadístico, eso implica que se necesitan algoritmos capaces de establecer distancias cuantitativas a partir de variables nominales, principalmente Levenshtein (Kessler, 1995). El uso de datos prosódicos provenientes de análisis acústicos, más recientemente, ha planteado una cuestión metodológica crucial: conviene operar con variables numéricas, lo cual implica el hecho de necesitar otro tipo de métricas para trabajar. Si los datos prosódicos se transcriben con símbolos alfabéticos, tanto en un nivel más superficial y cercano a las melodías acústicas, como en un nivel mucho más profundo o fonológico, de acuerdo con los postulados para el sistema, conseguimos una cadena alfabética que vuelve a necesitar algoritmos que trabajen con datos alfabéticos o nominales.

Existen herramientas disponibles para trabajar en dialectometría tanto con datos numéricos como con datos alfabéticos. Gabmap (Nerbonne, Colen, Gooskens, Kleiweg, y Leinonen, 2011) puede trabajar con ambas, pero con muchas limitaciones porque el programa no permite vectores y reduce cada variable a un único valor. VisualDialectometry (Goebel y Haimmerl, 2004) o DiaTech (Aurrekoetxea, Fernández-Aguirre, Rubio, Ruiz, y Sánchez, 2013) operan con datos nominales pero también ofrecen restricciones severas en el tratamiento de datos prosódicos porque, por ejemplo, no aceptan caracteres que se utilizan en el etiquetaje con los sistemas ToBI (“%” o “*”, por ejemplo). Además, los algoritmos no son suficientemente robustos como para hacer frente a diferencias en variabilidad en la longitud de las etiquetas prosódicas.

3. ¿POR QUÉ CREAR UNA HERRAMIENTA NUEVA PARA EL TRATAMIENTO PROSÓDICO DE LAS DISTANCIAS ENTRE LOS DATOS?

Necesitamos una herramienta que trabaje con los datos prosódicos obtenidos en el marco AMPER y que refleje las especificidades de dicho tipo de datos.

La prosodia, por una parte, se manifiesta, fundamentalmente, en tres parámetros: f_0 , duración e intensidad, de forma numérica; respectivamente, en Hz o semitonos, en segundos (o milésimas de segundo) y en

decibelios. Por otra parte, la prosodia vehicula información sobre la modalidad oracional, el acento léxico, la estructura sintáctica o la manifestación del foco. Finalmente, la prosodia expresa diferencias diafásicas, diastráticas o diatópicas (en terminología coseriana). En este último terreno, AMPER (Martínez-Celdrán y Fernández Planas, 2003–2016), junto con el IARI (Prieto, Borràs-Comes y Roseano, 2010–2014), ha ido conformando una enorme base de datos prosódicos acústicos que es susceptible de ser sometida a estudios dialectométricos a partir de algoritmos que trabajen con datos numéricos.

En el seno de este proyecto hubo una propuesta de herramienta dialectométrica, Stat-Distances (Rilliard y Lai, 2008), que no se acabó de desarrollar del todo y, a pesar de proporcionar resultados interesantes, no ofrecía datos imprescindibles como la matriz de distancias o la explicación de los algoritmos que usaba. Nuestro primer contacto con la metodología dialectométrica fue de la mano de este programa y nos sirvió para empezar a constatar que nuestros resultados no siempre eran totalmente coincidentes con los obtenidos por la dialectología tradicional, lo cual es plausible ya que la dialectología clásica nunca había tratado este tipo de datos. En seguida Stat-Distances dejó de estar disponible para los investigadores que trabajamos en el proyecto. Entre los diferentes grupos implicados se han sucedido otras propuestas. A saber: un script en R (Martínez Calvo y Fernández Rei, 2015) y el programa que se presenta en este trabajo, que es la versión mejorada de unas rutinas previas que llamamos Calcu-Dista (Roseano, Elvira-García, y Fernández Planas, en revisión; Elvira-García, 2014). La especificidad de los datos prosódicos y la falta de una herramienta útil para operar con ellos nos llevó a proponer nuestra herramienta, a la que llamamos ProDis, procedente de la expresión *Prosodic Distances* (Elvira-García, Balocco, Fernández Planas, Roseano, y Martínez Celdrán, 2015; Fernández Planas, 2016a, 2016b), que es la versión mejorada de Calcu-Dista (Roseano, Fernández Planas, Elvira-García, Cerdà Massó, y Martínez Celdrán, 2015) y también se inspira en Stat-Distances. ProDis funciona, como es habitual en los trabajos en el seno de AMPER, en el entorno MatLab.

Así pues, nuestra herramienta constituye nuestra contribución para construir un programa potente capaz de trabajar en dialectometría a partir de los datos prosódicos numéricos que obtenemos en nuestros análisis en el seno de AMPER. En concreto, responde, en esta primera fase, a las exigencias técnicas que habían surgido dentro del marco del proyecto. En concreto, se necesitaba que nuestra herramienta fuera “amigable” y flexible, que fuera capaz de trabajar con datos numéricos, de solucionar los problemas debidos a las diferencias en número de sílabas entre frases en distintas lenguas en el marco AMPER (por ejemplo, *O pássaro gosta de Renato*, 10 sílabas, vs. *La guitarra se toca con paciencia*, 11 sílabas), de considerar de cada frase 3 repeticiones, de ponderar f_0 por duración, por intensidad o por ambos parámetros a la vez y,

finalmente, de trabajar con nuevas lenguas adicionalmente.

4. LA GÉNESIS DE PRODIS: DE CALCUL-DISTA A PRODIS

ProDis, como antes Calcu-Dista y antes todavía Stat-Distances, utiliza la fórmula (1), que se inspira en la que propuso Hermes (1998) y que resulta ser una fórmula sencilla para calcular distancias entre datos acústicos numéricos.

$$(1) \quad RMS = \sqrt{\frac{\sum_{i=1}^N (f_0x_i - f_0y_i)^2}{N}}$$

El antepasado más cercano de ProDis, conocido como Calcu-Dista, era, más que una herramienta, una rutina para el cálculo de distancias prosódicas a partir de los datos numéricos de las melodías en semitonos. Esa rutina se fundamentaba en tres programas bien conocidos: Praat v. 5.4.01 (Boersma y Weenink, 2014), Excel (Microsoft Office 2007) y SPSS Statistics 20 (IBM). Los programas en cuestión tomaban con punto de partida los datos acústicos previamente procesados por tres programas creados en el seno del Laboratori de Fonètica de la UB y circunscritos al ámbito AMPER: AMPER-Reno, AMPER-Extra y AMPER-Eti (Roseano, 2012).

En primer lugar, un script de Praat creado *ad hoc* extraía, a partir de los archivos txt de cada repetición de las frases proporcionados por AMPER-2006 (López Bobo, Muñiz Cachón, y Díaz Gómez, 2007), los valores de f_0 en semitonos y los colocaba en una matriz de datos comparando cada repetición de una frase en un mismo hablante y entre hablantes distintos considerando tres valores por vocal.

En segundo lugar, un análisis en Excel sobre la salida de Praat aplicaba la fórmula de las distancias escogida. Se escogió como índice de la distancia entonativa entre dos frases, que podemos llamar x e y, la media cuadrática de la diferencia entre los valores de f_0 de la frase x y de la frase y en cada uno de los puntos de medición. Para los dos conjuntos x e y de valores de f_0 $\{f_0x_1, f_0x_2, \dots, f_0x_N\}$ y $\{f_0y_1, f_0y_2, \dots, f_0y_N\}$, donde N es el número de puntos de medición de f_0 en cada una de las dos frases, mientras que f_0x_i y f_0y_i son los valores de f_0 en semitonos en cada uno esos puntos.

Esta fórmula proporcionaba la distancia entre dos frases con la misma estructura (por ejemplo, entre dos declarativas SVO con sujeto llano, verbo llano y objeto esdrújulo) de dos puntos de encuesta. Para determinar la distancia general entre todas las frases de dos puntos de encuesta, puesto que la distribución de las distancias no es normal, de acuerdo con De Castro Moutinho, Coimbra, Rilliard, y Romano (2011, p. 44) se escogió la mediana de las RMS calculadas por cada pareja de frases x e y. A partir de las medianas de las distancias entre cada par de puntos de encuesta se pudo construir la matriz de distancias correspondiente.

En tercer lugar, la matriz de distancias constituía, a su vez, la base para la fase final del proceso de análisis,

que se efectuaba con SPSS y consistía en lo siguiente. En primer lugar, en un análisis de conglomerados clúster, técnica multivariante cuya finalidad es clasificar los puntos de encuesta en grupos a partir de la semejanza entre sus características entonativas donde como método de comparación se utiliza la media de las distancias entre los grupos tratando de lograr la máxima homogeneidad en cada grupo y la mayor diferencia entre los grupos. El resultado se expresa en forma de dendrograma que permite ver, en forma de árbol invertido, cómo se agrupan los datos hasta el nivel que se considera oportuno. Y en segundo lugar, un análisis de tipo escalamiento multidimensional (EMD o MDS) que representa bidimensional o tridimensionalmente de forma gráfica las distancias entre los sujetos o puntos de encuesta de la manera más objetiva posible en un espacio virtual. Este método estadístico pretende construir un espacio métrico con el menor número de dimensiones posibles, de tal manera que permite representar las proximidades o preferencias entre los objetos con el mayor grado de fidelidad. Desde un conjunto de objetos se establecen sus propiedades numéricas a partir de las cuales se elaboran las tablas de proximidad (o de similitud) y, finalmente, se trasladan estas proximidades a un espacio, un mapa de objetos (Matas Crespo, 2006). En realidad, ambos tipos de gráficos —dendrogramas y espacios MDS— proporcionan la misma información, y así se puede comprobar en el apartado de resultados. La ventaja de ambas formas de representación es la de permitir captar la distribución y la agrupación de los datos sin necesidad de tener que recurrir a una matriz de distancias numérica de proporciones enormes.

Como medida utilizamos el intervalo de distancia euclidiana (2).

$$(2) \quad d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

El funcionamiento de Calcu-Dista, como era esperable, ha sido validado estadísticamente mediante la comparación con resultados obtenidos con métodos comparables (Fernández Planas *et al.*, 2015).

5. CARACTERÍSTICAS FUNDAMENTALES DE PRODIS

La idea de partida era: (1) cómo poder establecer una matriz de distancias por informantes y por puntos de encuesta (reuniendo en un bloque los distintos informantes del mismo punto de encuesta) a partir de una matriz inmensa de datos obtenidos en AMPER-2006 en txt de datos en semitonos; y (2) cómo ver reflejadas las matrices de distancias (o de proximidades) de forma gráfica en forma de dendrogramas y de distribución en espacios virtuales.

ProDis realiza la media y la mediana de correlación por informantes y por punto de encuesta. A continuación, a partir de estos datos realiza un análisis de clúster que permite clasificar los informantes o los puntos de encuesta en diferentes grupos, tanto en forma

de dendrograma como de distribución en un espacio virtual, según su semejanza o su proximidad.

Concretamente, a partir de los datos en semitonos de la f_0 de las frases, computa las correlaciones entre ellos, analiza la media y la mediana de las correlaciones para cada hablante, construye la matriz de correlaciones entre hablantes y entre puntos de encuesta, lleva a cabo el análisis de clúster entre hablantes y entre puntos de encuesta y prepara las diferentes salidas gráficas (los dendrogramas y EMD), pero también los gráficos que permiten establecer la validación estadística de los datos como los mapas de correlación, de desviación estándar, los gráficos de silueta o los gráficos de Shepard.

En la Figura 1 aparece una imagen de la interfaz de la herramienta.

Para el cálculo de las correlaciones se utilizan las de Pearson con cuatro métricas diferentes para cada análisis: (1) sin ponderar; (2) ponderadas por la intensidad (a la manera de Hermes, 1998), ponderadas por la duración, y ponderadas por la intensidad y la duración. Existen diferentes métodos para mediar las distancias en dialectometría y la correlación de Pearson (la que hemos estimado mejor) es una de ellas. Se presentó una revisión crítica de ellos en Elvira-García, Balocco, Roseano, y Fernández Planas (2016) y también en Hermes (1998). Serían, entre otros, el algoritmo de Levenstein (Kessler, 1995), la distancia euclidiana (Nerbonne *et al.*, 2011; Roseano *et al.*, 2015); Mahalanobis (Wouters y Macon, 1998), correlación de Spearman (Hermes, 1998), correlación de Pearson (Heeringa y Gooskens, 2003), tau de Kendall (Hermes, 1998).

Para las correlaciones, el programa compara cada repetición de una frase de un hablante de un punto de encuesta con las otras dos repeticiones del mismo informante de la misma frase y con las tres repeticiones de la frase con la misma estructura sintáctica de otro informante. Ello permite obtener una matriz de correlaciones como la que aparece en la Figura 2. Los valores, lógicamente, van de -1 a 1.

Los mismos datos los podemos ver de forma completa y más fácilmente aprehensible en un mapa de correlaciones como el que aparece en la Figura 3. Cabe destacar que en el gráfico en cuestión no aparecen valores numéricos, sino gradaciones de colores que

Figura 1: Interfaz de ProDis.

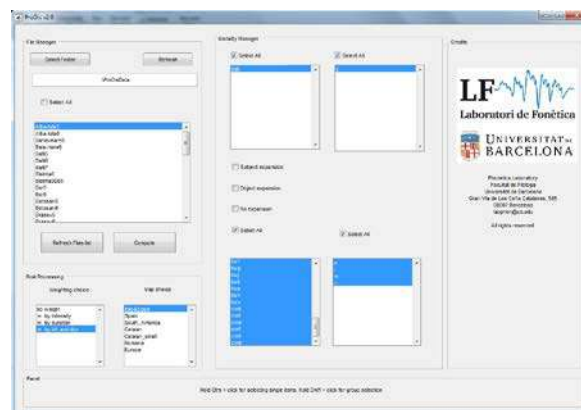


Figura 2: Ejemplo de matriz de correlaciones.

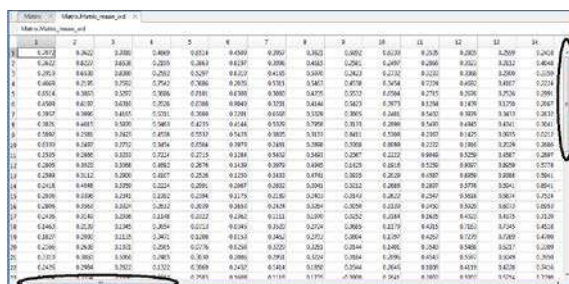


Figura 3: Ejemplo de mapa de correlaciones.

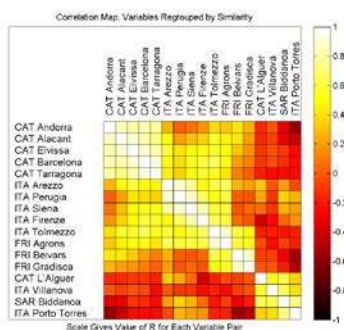
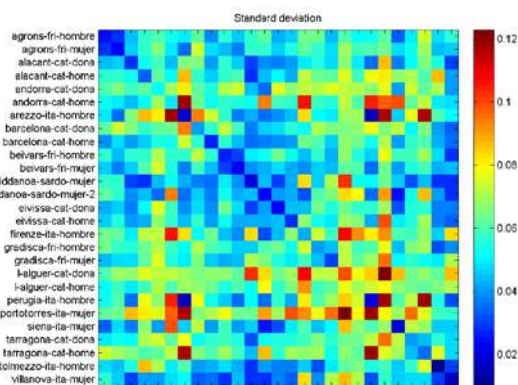
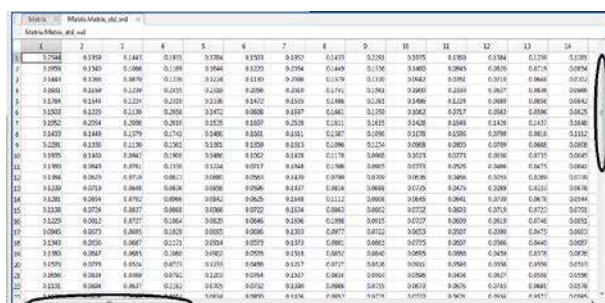


Figura 4: Matriz y mapa de desviación estándar.



femenina de L'Alguer en catalán, en el mapa de ejemplo) permiten adivinar que el informante (o el punto) en cuestión es menos homogéneo en los patrones melódicos que utiliza.

ProDis proporciona también gráficos de silueta que, de alguna manera, indican los grupos que matemáticamente sería óptimo establecer en los dendrogramas que agrupan progresivamente los elementos hasta unirlos todos, aunque es la decisión del investigador la que establece el límite de agrupaciones. Para las agrupaciones se toma como referencia el elemento más lejano. Podemos ver un ejemplo de dendrograma en la Figura 5.

El dendrograma funciona con una técnica multivariante que crea clústering aglomerativo jerárquico usando un método de agrupación completo, grupos basados en el elemento más lejano.

El gráfico de Shepard, que se refiere a los gráficos en EMD, nos demuestra cuando tiende a una línea, que los gráficos son fiables y válidos.

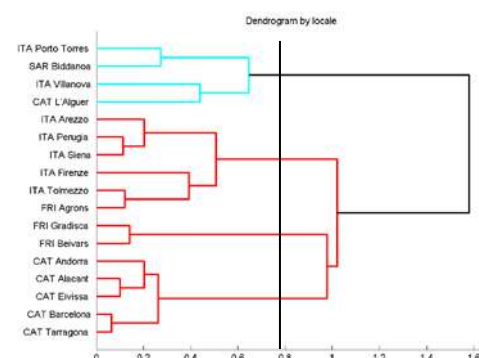
En la Figura 6 vemos dos formas de visualizar los gráficos EMC en ProDis: a) en dos dimensiones; b) en tres dimensiones. Aunque el gráfico en tres dimensiones es más recomendable porque el valor de Stress tiende a ser más bajo que en el gráfico en dos dimensiones, en ocasiones puede ser más difícil de interpretar.

Finalmente, mediante ProDis podemos obtener mapas geográficos donde se representen en colores coincidentes con los dendrogramas y con los EMD las localizaciones de los informantes o de los puntos de encuesta. Véase un ejemplo en la Figura 7.

6. CONCLUSIÓN. MEJORAS DE PRODIS RESPECTO A STAT-DISTANCES Y LÍNEAS DE FUTURO

ProDis satisface con creces la idea de partida que señalábamos en el inicio del aparatado anterior, ya que por un lado permite establecer una matriz de distancias por informantes y por puntos de encuesta a partir de los datos numéricos obtenidos en AMPER-2006 y, por otra parte, transforma las matrices de distancias en gráficos de más fácil interpretación. Además, mejora y supera en prestaciones las que ofrecía Stat-Distances porque: (1) considera repertorios de datos no coincidentes en el

Figura 5: Ejemplo de dendrograma con una línea que establece el límite de agrupaciones que se sería óptimo tener en cuenta.



indican grados de correlación distintos. De esa manera, la matriz resulta de comprensión más inmediata.

La herramienta ofrece también la matriz y el mapa de desviación estándar por informante y por punto de encuesta (Figura 4), lo cual es muy interesante para comprobar la coherencia intrasujetos y entre sujetos, por un lado, y también intrapuntos de encuesta y entre puntos de encuesta, por otro lado.

Se aprecia fácilmente cómo los puntos más tendientes a rojo (por ejemplo los valores de la voz

- D'Andrés Díaz, R., Álvarez-Balbuena García, F., y Suárez Fernández, X. M. (2007). Proxecto ETLEN para o estudo dialectográfico e dialectométrico da zona Eo-Navia, Asturias: fundamentos teóricos. Actas VII Congreso Internacional de Estudos Galegos: mulleres en Galicia: Galicia e os outros pobos da península (pp. 749–759). A Coruña: Edicións do Castro.
- De Castro Moutinho, L., Coimbra, R. L., Rilliard, A., y Romano, A. (2011). Mesure de la variation prosodique diatopique en portugais européen. *Estudios de Fonética Experimental*, 20, 33–55.
- Elvira-García, W. (2014). Calcu-Dista scripts package. Praat script. Disponible en <http://stel.ub.edu/labfon/en/praat-scripts>
- Elvira-García, W., Balocco, S., Fernández Planas, A. M., Roseano, P., 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. Presentado en el VII Workshop sobre prosodia del catalán, Universitat de Barcelona, 22/06/2015.
- Elvira-García, W., Balocco, S., Roseano, P., y Fernández Planas, A. M. (2016). Comparació de mesures de distància prosòdica entre varietats dialectals. Presentado en el VIII Workshop sobre prosodia del catalán, Universitat Pompeu Fabra, 04/07/2016.
- Elvira-García, W., Roseano, P., Fernández Planas A. M. y Martínez Celdrán E. (2016). A tool for automatic transcription of intonation: Eti-ToBI a ToBI transcriber for Spanish and Catalan. *Language Resources and Evaluation*, 50(4), 767–792.
- Fernández Planas, A. M. (2005). Datos generales del proyecto AMPER en España. *Estudios de Fonética Experimental*, 14, 13–27.
- Fernández Planas, A. M. (2016a). Aspectos de ProDis, una nueva herramienta para el análisis dialectométrico prosódico. Presentado en el Workshop «Approaches to Sociolinguistic Aspects of Romanian and Spanish Intonation», Alexandru Ioan Cuza University of Iasi (Rumanía), 21/10/2016.
- Fernández Planas, A. M. (2016b). Características generales de ProDis (herramienta para analizar distancias prosódicas). Presentado en el Servei de Tractament de la Parla i el So (STPS) de la Universitat Autònoma de Barcelona, 18/11/2016.
- Fernández Planas, A. M., Dorta, J., Roseano, P., Díaz, X., Elvira-García, W., Martín Gómez, J. A., 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(2), 13–45.
- Fernández Planas, A. M., 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, 141–178.
- Goebel, H. (1981). Eléments d'analyse dialectométrique (avec application à l'AIS). *Revue de Linguistique Romane*, 45, 349–420.
- Goebel, H. (1987). Encore un coup d'oeil dialectométrique sur las Tableaux phonétiques de patois suisses romands (TPPSR). *Vox Romanica*, 46, 91–125.
- Goebel, H. (1993). Dialectometry: A short overview of the principles and practice of quantitative classification of linguistic atlas data. En R. Köhler y B. B. Rieger (Eds.), *Contributions to quantitative linguistics* (pp. 277–315). Dordrecht: Springer.
- Goebel, H. (2003). Regards dialectométriques sur les données de l'Atlas linguistique de la France (ALF): relations quantitatives et structures de profondeur. *Estudis Romànics*, 25, 60–117.
- Goebel, H. y Haimmerl, E. (2004). Visual Dialectometry. <http://www.dialectometry.com/dmdocs/index.html> [28/11/2016].
- Goebel, H. y Schiltz, G. (1997). Dialectometrical compilation of CLAE 1 and CLAE 2. Isoglosses and dialect integration. En W. Viereck, H. Ramisch, H. Händler, y C. Marx (Eds.), *The computer developed linguistic Atlas of England, Vol. 2* (pp. 13–21). Tübingen: Niemeyer.
- Heeringa, W. y Gooskens, C. (2003). Norwegian dialects examined perceptually and acoustically. *Computers and the Humanities*, 37(3), 293–315.
- Heeringa, W. y Nerbonne, J. (2001). Dialect areas and dialect continua. *Language Variation and Change*, 13, 375–400.
- Hermes, D. J. (1998). Measuring the perceptual similarity of pitch contours. *Journal of Speech Language and Hearing Research*, 41(1), 73–82.
- Kessler, B. (1995). Computational dialectology in Irish Gaelic. En *Proceedings of the Seventh Conference of the European Chapter of the Association for Computational Linguistics (EACL)* (pp. 60–66).
- López Bobo, M. J., Muñiz 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 Ediciones.
- Martínez Calvo, A., y Fernández Rei, E. (2015). Unha ferramenta informática para a análise dialectométrica da prosodia. *Estudios de Fonética Experimental*, 24, 289–303.
- Martínez Celdrán, E. y Fernández Planas, A. M. (Coords.) (2003–2016). *Atlas Multimèdia de la Prosòdia de l'Espai Romànic*. http://stel.ub.edu/labfon/ampcr/cast/index_ampcrat.html
- Matas Crespo, J. (2006). La técnica del Escalamiento Multidimensional en el vocalismo: un análisis comparativo (Tesis Doctoral). Universitat de Barcelona.

- Nerbonne, J., Colen, R., Gooskens, C., Kleiweg, P., y Leinonen, T. (2011). Gabmap: A web application for dialectology. *Dialectologia*. Special issue II, 65–89.
- Polanco, L. (1992). Lengua y dialecto: una aplicación dialectométrica a la lengua catalana. *Miscelánea*, 3, 5–28.
- Prieto, P., Borràs-Comes, J., y Roseano, P. (Coords.) (2010–2014). Interactive Atlas of Romance Intonation. <http://prosodia.upf.edu/iari/>.
- 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, Rumanía: Editura Universității Al. I. Cuza.
- Romano, A. (2003). Un projet d'Atlas multimédia prosodique de l'espace roman (AMPER). En F. Sánchez Miret (Ed.), *Atti del XXIII CILFR, Vol. 1* (pp. 279–294). Tübingen: Niemeyer.
- Romano, A. y Contini, M. (2001). Un progetto di Atlante geoprosodico multimediale delle varietà linguistiche romanze. En E. Magno Caldognetto y P. Cosi (Eds.), *Multimodalità e Multimedialità nella Comunicazione. Atti delle XI Giornate di Studio del "Gruppo di Fonetica Sperimentale" dell'Associazione Italiana di Acustica* (pp. 121–126). Padova: Unipress.
- Roseano, P. (2012). La prosòdia del friulà en el marc de l'Atles Multimèdia de Prosòdia de l'Espai Romànic (Tesis Doctoral). Universitat de Barcelona.
- Roseano, P., Elvira-García, W., y Fernández Planas, A. M. (en revisión). Calcu-Dista: A Tool for Dialectometric Analysis of Intonational Variation. En I. Feldhausen, M. M. Vanrell y U. Reich (Eds.), *Empirical Methods in Romance Prosody Research*. Language Science Press.
- Roseano, P., Fernández Planas, A. M., Elvira-García, W., Cerdà Massó, R., y Martínez Celdrán, E. (2015). Contacto lingüístico y transferencia prosódica: El caso del alguerés. *Dialectologia et Geolinguistica*, 23(1), 95–123.
- Saramago, J. (2002). Diferenciação lexical interpontual nos territórios galego e português (Estudo dialectométrico aplicado a materiais galegos do ALGa). En R. Álvarez, F. Dubert García, y X. Sousa Fernández (Eds.), *Dialectoloxía e léxico* (pp. 41–68). Santiago de Compostela: Instituto da Lingua Galega – Consello da Cultura Galega.
- Séguy, J. (1973). La dialectométrie dans l'Atlas linguistique de la Gascogne. *Revue de Linguistique Romane*, 37, 1–24.
- Sousa, X. (2006). Análise dialectométrica das variedades xeolingüística galegas. En M. C. Rolão Bernardo y H. Mateus Montenegro (Eds.), *Actas do I Encontro de Estudos Dialectológicos* (pp. 345–362). Ponta Delgada, Portugal: Instituto Cultural de Ponta Delgada.
- Verlinde, S. (1988). La dialectométrie et la détection des zones dialectales: L'architecture dialectale de l'Est de la Belgique romane. *Revue de Linguistique Romane*, 51, 151–172.
- Wouters, J., y Macon, M. W. (1998). A perceptual evaluation of distance measures for concatenative speech synthesis. En *Proceedings of the Fifth International Conference on Spoken Language Processing*.

FonetiToBI, una herramienta para la anotación prosódica automática de corpus

Wendy Elvira-García¹, Juan María Garrido¹

¹ Universidad de Barcelona

e-mail: wendyelvira@ub.edu, juanmaria.garrido@ub.edu

Citation / Cómo citar esta publicación: Elvira-García, W. & Garrido, J. M.. (2019). FonetiToBI, una herramienta para la anotación prosódica automática de corpus. In J. M. Lahoz-Bengoechea & R. Pérez Ramón (Eds.), *Subsidia. Tools and resources for speech sciences* (pp. 133–142). Málaga: Universidad de Málaga.

RESUMEN: El objetivo de esta comunicación es presentar FonetiToBI, una herramienta basada en Praat para la anotación prosódica automática de enunciados en español y catalán en el marco del modelo métrico autosegmental. Se describen brevemente sus funcionalidades y su estructura, y se explican de forma general las reglas, basadas en conocimiento lingüístico-fonético, que se incluyen en los módulos que la componen. Se presenta también la evaluación realizada con enunciados del español y del catalán, los resultados de la cual muestran que se trata de una herramienta con un grado de fiabilidad suficiente para su uso en tareas de anotación automática de corpus.

Palabras clave: Anotación prosódica, corpus, modelo autosegmental, Sp_ToBI, Cat_ToBI

ABSTRACT: The paper aims at describing FonetiToBI, a Praat-based tool for the automatic prosodic annotation of Spanish and Catalan utterances within the Autosegmental Metrical model. It describes briefly the structure of the script and its functionalities. Moreover, an overview of the knowledge-based linguistic-phonetic rules making up its modules is also presented. Finally, the evaluation of the tool carried out with utterances of Spanish and Catalan is detailed. The results of the evaluation show that the tool is reliable enough to be used in tasks of automatic annotation of corpora.

Keywords: Prosodic annotation, corpora, Autosegmental-Metrical Model, Sp_ToBI, Cat_ToBI

1. INTRODUCCIÓN

El objetivo de este trabajo es presentar FonetiToBI, una herramienta diseñada para la anotación prosódica automática de habla para el español y el catalán en el marco del Modelo Métrico Autosegmental (AM) (Pierrehumbert, 1980).

En los últimos años, se ha producido un auge en el desarrollo de herramientas para la anotación automática de corpus de habla, tanto a nivel segmental como suprasegmental. En el caso de la anotación segmental, herramientas como MAUS (Schiel, 1999), EasyAlign (Goldman, 2011) o SPPAS (Bigi, 2012) ofrecen resultados aceptables, aunque en diferente grado según la lengua y la herramienta. En el caso de la anotación suprasegmental, existen también desde hace años herramientas de segmentación automática de las unidades prosódicas (SPPAS, Bigi, 2012; SegProso, Garrido, 2013) y de la entonación (MoMel, Hirst y Espesser, 1993; MelAn, Garrido, 2010). En el marco del Modelo AM, existen también algunas herramientas, como AuToBI (Rosenberg, 2010), que permite la

anotación automática de la entonación de enunciados del inglés, o Eti_ToBI (Elvira-García, Roseano y Fernández Planas, 2015; Elvira-García, Roseano, Fernández Planas, y Martínez Celdrán, 2016), diseñado para la anotación de enunciados en español y catalán.

Por otro lado, también está habiendo un genuino interés en definir sistemas de transcripción prosódica más objetivos y basados en los fenómenos fonéticos (Roseano y Fernández Planas, 2015; Hualde y Prieto, 2016).

La herramienta presentada aquí permite la anotación automática completa de los enunciados en el marco de las convenciones ToBI (*Tones and Break Indices*). Es el resultado de la integración de dos herramientas existentes previamente, aunque modificadas para el desarrollo de esta aplicación, SegProso y EtiToBI, la primera orientada a la anotación automática de unidades entonativas, y la segunda a la anotación de los eventos tonales. Las dos tienen en común que han sido desarrolladas, a diferencia de otras herramientas, implementando el conocimiento fonético y lingüístico necesario en forma de reglas.

Figura 1: Estructura y esquema de funcionamiento de FonetiToBI.

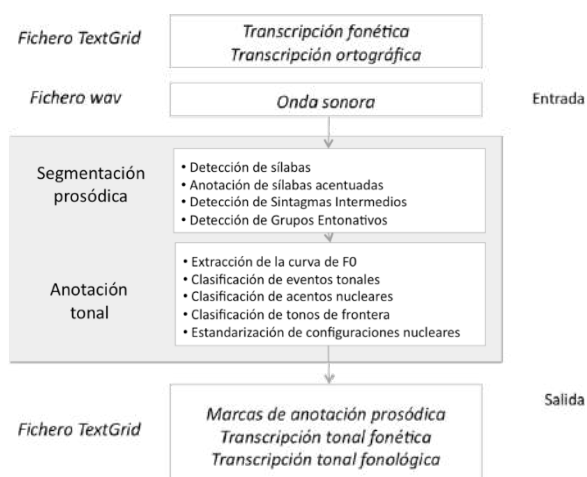
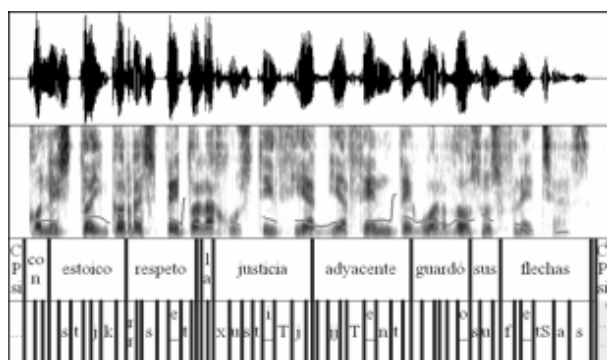


Figura 2: Ejemplo de TextGrid de entrada de FonetiToBI.



2. DESCRIPCIÓN DE LA HERRAMIENTA

FonetiToBI es una herramienta basada en Praat (Boersma, 2001) que permite obtener automáticamente una anotación ToBI completa (tonos y límites prosódicos) a partir de una transcripción ortográfica y fonética de los enunciados de entrada. La Figura 1 presenta su estructura y flujo de funcionamiento.

FonetiToBI está compuesto de dos módulos principales: el primero es una versión modificada de SegProso, y se encarga de identificar los límites prosódicos (*Break Indices*, BI); el segundo, EtiToBI, se encarga de la anotación de los tonos.

2.1. Entrada

FonetiToBI necesita como entrada dos ficheros: un fichero wav con la onda sonora, y un fichero TextGrid de Praat con la transcripción ortográfica y fonética del enunciado, alineada temporalmente con la señal de habla. El TextGrid debe contener, pues, al menos dos capas (*tiers*), una que contenga la transcripción ortográfica, palabra a palabra, y otra con la transcripción fonética, alófono a alófono, ambas alineadas temporalmente con la señal de habla. La Figura 2 presenta un ejemplo de TextGrid de entrada.

FonetiToBI admite como entrada transcripciones fonéticas realizadas utilizando tanto el alfabeto IPA (IPA, en línea) como SAMPA (Wells, 1995). Esto permite, por ejemplo, emplear como entrada ficheros

TextGrid generados automáticamente con alguna herramienta de segmentación automática, como SPPAS, que generan la transcripción fonética con símbolos SAMPA.

Otros parámetros de entrada que deben especificarse al inicio son la versión de ToBI que se empleará para la anotación tonal (Sp_ToBI, para la anotación de enunciados en español, Prieto y Hualde, 2015; o Cat_ToBI, para la anotación del catalán, Prieto y Cabré, 2013), o si se llevará a cabo una revisión manual de la anotación al final del proceso.

2.2. Segmentación prosódica

El módulo de segmentación prosódica se encarga de generar una capa de segmentación prosódica (*Break Index Tier*), que contiene las marcas correspondientes a los diferentes tipos de límites prosódicos contemplados en sistema ToBI para el español: grupo clítico (0), palabra (1), sintagma intermedio (3) y grupo o frase entonativa (4). La identificación de estos límites en el enunciado de entrada se lleva a cabo, a partir de la transcripción fonética y ortográfica del enunciado, en tres fases: la primera, de identificación de las sílabas tónicas; la segunda, de identificación de los grupos entonativos; y finalmente, la identificación de los límites de sintagma intermedio. Después, en la última fase de procesado, la salida de estos tres pasos se integra y modifica para obtener una representación de la anotación prosódica en un formato compatible con las convenciones ToBI. Aunque no la única, esta ha sido la principal modificación llevada a cabo en SegProso para su integración en FonetiToBI.

2.2.1. Detección de las sílabas tónicas

La detección de las sílabas tónicas es un paso previo necesario para el etiquetado ToBI que se lleva a cabo en la segunda fase, pero se requiere también para la identificación de los límites de palabra fonológica (los etiquetados con '1' según las convenciones ToBI). Para su identificación, es necesario llevar a cabo previamente una agrupación de los alófonos en sílabas, que también es necesaria para la anotación de los tonos.

La detección de los límites silábicos se lleva a cabo a partir del análisis de la segmentación en palabras y alófonos proporcionada en el TextGrid de entrada, mediante la aplicación de una serie de reglas implementadas en un script de Praat cuyo flujo de funcionamiento puede resumirse de la siguiente manera:

- Se localizan en primer lugar los límites de palabra en la capa que contiene la segmentación ortográfica. Se asume que estos límites actúan como una barrera para la agrupación de los alófonos en sílabas, por lo que no se contemplan los fenómenos de resilabificación entre palabras.
- Se buscan los alófonos que representan núcleos silábicos en la capa de transcripción fonética. El procedimiento encargado de esta tarea comprueba si los alófonos se encuentran en la lista implementada de 'alófonos nucleares' (que contenía inicialmente solo las vocales del español y el catalán, pero que ha

sido ampliada posteriormente con vocales de otras lenguas, como el portugués de Brasil o el francés; Silva y Garrido, 2016).

- Se buscan los límites de la sílaba correspondiente a cada núcleo silábico detectado. Las reglas de silabificación que se aplican analizan los alófonos que aparecen en la transcripción entre el núcleo de la sílaba actual y el siguiente, y establecen el límite de sílaba al final del alófono identificado como final de la coda. Como ya se ha explicado, el límite de palabra se considera una barrera para la silabificación, por lo que cuando se alcanza un límite de palabra se asigna automáticamente un límite de sílaba en esa posición.

Finalmente, una vez establecidos los límites de la sílaba, si el alófono identificado como núcleo silábico lleva una marca de acento, la sílaba se anota como tónica. Es muy importante, por tanto, que la transcripción fonética de entrada incluya marcas de acento para el correcto funcionamiento de la herramienta.

2.2.2. Detección de sintagmas intermedios

Los sintagmas intermedios (*intermediate phrases*, ip) son, en el modelo AM, aquellas unidades entonativas cuyo límite no se marca con una pausa, sino únicamente con un tono de frontera y un alargamiento de la sílaba final de unidad. Su detección automática en la onda sonora de los enunciados debe orientarse, por tanto, a la identificación de estos dos indicios acústicos (movimientos tonales indicativos de la presencia de un tono de límite y alargamientos silábicos). El límite final de estas unidades se marca en la capa de segmentación prosódica con un '3'.

Las reglas implementadas en FonetiToBI para la detección de los límites de ip están orientadas únicamente a la detección de posibles tonos de frontera al final de las palabras tónicas. Requiere, por tanto, la segmentación en palabras proporcionada en el TextGrid de entrada y la identificación de las palabras tónicas llevada a cabo en la fase anterior. Dichas reglas intentan detectar dos tipos de indicios acústicos relacionados con la presencia de tonos de frontera: por un lado, la presencia en la curva de f_0 de determinados movimientos ascendentes indicativos de la existencia de un tono de frontera intermedia; y por otro, la presencia de reajustes de f_0 , indicadores del inicio de una nueva unidad entonativa.

- Las reglas de identificación de tonos de frontera intermedia tratan de localizar movimientos ascendentes de f_0 con una pendiente lo suficientemente pronunciada como para ser interpretadas perceptivamente como tonos de frontera. Básicamente, estas reglas comparan el valor de f_0 en el centro de cada sílaba tónica con el valor de f_0 en la última sílaba de la palabra, o, si la sílaba tónica es la última, al final de la misma. Si la diferencia entre los dos valores supera un umbral (actualmente establecido en el 5 % del valor de f_0 en el centro de la sílaba tónica), se considera que existe

Figura 3: Onda Sonora, espectrograma, curva de f_0 y anotación correspondiente al enunciado del catalán 'Alt, fort, i amb expressió salvatge', pronunciado por un hablante femenino. El intervalo seleccionado fue marcado como ip por las reglas de identificación de tonos de frontera.

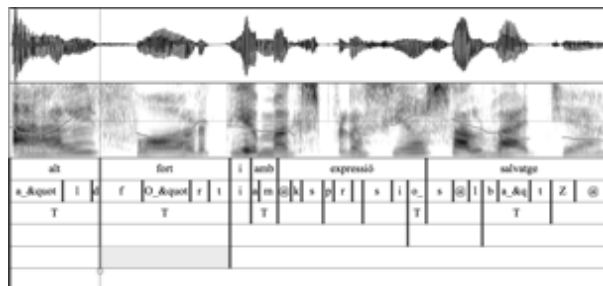
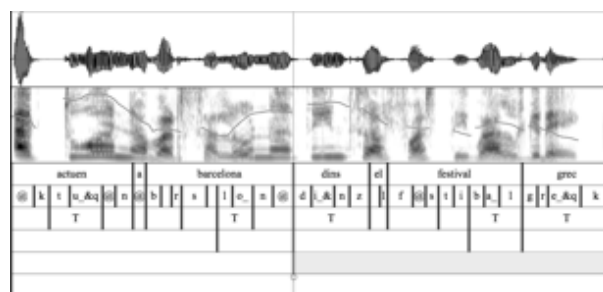


Figura 4: Onda Sonora, espectrograma, curva de f_0 y anotación correspondiente al enunciado del catalán 'Actuen a Barcelona dins el festival Grec', pronunciado por un hablante masculino. El intervalo seleccionado fue marcado como ip por las reglas de reajuste de f_0 .



un límite de frase intermedia al final de esa palabra. La Figura 3 presenta un ejemplo de límite detectado con este tipo de reglas.

- Las reglas de reajuste de f_0 tratan de identificar diferencias significativas de f_0 entre sílabas tónicas consecutivas: se miden los valores de f_0 en el centro de ambas sílabas, y si el valor de la segunda resulta ser significativamente mayor que el de la primera (la diferencia entre ambas debe ser al menos un 5 % del valor de f_0 en la primera tónica), se establece un límite de ip entre las dos palabras. La Figura 4 presenta un ejemplo de límite correspondiente a este segundo tipo.

2.2.3. Detección de grupos entonativos

Los grupos entonativos (*Intonational Phrases*, IP) son unidades prosódicas marcadas por un tono de límite y una pausa. Se marcan mediante un '4' en la capa de segmentación prosódica.

Las reglas encargadas en FonetiToBI de la detección de límites de IP asumen una equivalencia plena entre IP y grupo fónico, un supuesto que, como es bien sabido, es cierto en la mayoría de los casos, pero no en todos. Dicho de otra manera, las reglas asumen que existe un límite de IP siempre que se produce una pausa silenciosa (y por tanto se cierra un grupo fónico), y no comprueban si antes de la pausa existe efectivamente un tono de frontera.

El procedimiento de detección de estos límites en FonetiToBI es, pues, sencillo: simplemente se localizan en la capa de la segmentación fonética de entrada los

segmentos etiquetados como silencio, y se asigna una marca de límite al final de la palabra que lo precede. Es muy importante, por tanto, que la anotación de los silencios sea lo más correcta posible en el TextGrid de entrada.

2.2.4. Conversión a formato ToBI

En su versión anterior, SegProso generaba salida diferentes capas (*tiers*), cada una con la anotación correspondiente a una unidad prosódica diferente. En las convenciones ToBI, sin embargo, esta información se muestra en un solo tier (*Break Index tier*). La última fase del proceso de segmentación prosódica consiste entonces en convertir la salida de SegProso, en diferentes *tiers*, a un formato compatible con las convenciones ToBI, en un solo *tier*. De la salida original de SegProso, se mantiene también el *tier* correspondiente a la segmentación silábica, con indicación de las sílabas tónicas (etiqueta ‘T’)

2.3. Anotación tonal

El módulo de anotación tonal se encarga de generar una capa de tonos (*Tones*), que contiene las marcas correspondientes a los movimientos tonales de la curva de f_0 tal y como se explicitan en las convenciones para acentos tonales y tonos de frontera de los sistemas de transcripción Cat_ToBI (Prieto y Cabré, 2013) y Sp_ToBI (Hualde y Prieto, 2016).

La transcripción prosódica tiene dos niveles. El primero, que llamaremos transcripción fonética estrecha, es una transliteración de los movimientos tonales de f_0 en términos AM, es decir en L y H. Esta transcripción sigue las convenciones de transcripción prosódica fonética estrecha expuestas en Martínez Celdrán y Fernández Planas (2003) y Roseano y Fernández Planas (2013). En este nivel, el procedimiento de anotación empleado puede generar en algún caso etiquetas no totalmente acordes con las convenciones estándar ToBI, como son los tonos tritones. El segundo nivel es una transcripción fonética ancha en la que los movimientos de la transcripción fonética estrecha se simplifican para dejar solo aquellos susceptibles de ser tonos fonológicos, es decir, los recogidos en las versiones actuales de Sp_ToBI y Cat_ToBI.

La clasificación de los movimientos tonales se lleva a cabo a partir de:

- la segmentación silábica;
- las marcas de segmentación prosódica (*Break Indices*);
- las marcas de acento léxico.

Esta información es proporcionada por el módulo de segmentación descrito en el apartado 2.2.

2.3.1. Extracción de la curva de f_0

Además de los datos provenientes de la salida del módulo de segmentación, para poder transcribir la entonación el script necesita los datos de f_0 . Estos datos se extraen del sonido proporcionado en el input usando

el método de autocorrelación propuesto por Boersma (1993).

Los valores mínimo y máximo de f_0 utilizados como parámetros por el método de extracción de f_0 (*pitch floor* y *pitch ceiling*) se calculan para cada IP también automáticamente. Para su obtención se aplica el método en dos pasos expuesto en Hirst (2011), que se basa a su vez en las investigaciones presentadas en De Looze (2010). Este método extrae un primer objeto ‘Pitch’ con un rango amplio, para después buscar el f_0 mínimo y máximo en ese objeto y crear un objeto nuevo con el valor mínimo (que se establece multiplicando por 0.75 el primer cuartil del rango anterior) y el máximo (obtenido multiplicando por 1.5 el tercer cuartil del rango).

El objeto ‘Pitch’ resultante es el que se usa para detectar los movimientos tonales y clasificarlos, es decir para realizar la transcripción tonal. Por lo tanto, si dicha curva contiene errores de detección de f_0 , la transcripción resultante será errónea. Por este motivo, el transcriptor funciona mejor con grabaciones de calidad alta, un alto porcentaje de segmentos sonoros y pocos sonidos o elementos que puedan causar errores o problemas de detección en el contorno de f_0 , como serían fricativas sibilantes o barras de explosión.

2.3.2. Clasificación de eventos tonales

La teoría AM prevé que los movimientos tonales estén asociados a algún elemento prominente en el enunciado (sílabas tónicas o fronteras prosódicas). Por lo tanto, el módulo usa la información contenida en el TextGrid para determinar la posibilidad de prominencia, y, a partir de ahí, analiza los movimientos de f_0 para encontrar eventos tonales.

El script considera que ha habido un movimiento tonal entre dos momentos del enunciado (por ejemplo, entre dos sílabas contiguas o entre el inicio de la sílaba tónica y el final) cuando la diferencia entre esos dos momentos sobrepasa los 1.5 semitonos. El uso de este umbral está avalado por trabajos que muestran que las diferencias menores de 1.5 semitonos no se usan con significado lingüístico (Pamies, Fernández Planas, Martínez Celdrán, Ortega-Escandell, & Amorós Céspedes, 2002; Rietvelt y Gussenhoven, 1985). Si existe un movimiento que sobrepasa los 1.5 semitonos alrededor de una sílaba prominente, el script detecta que ha habido un evento tonal y procede a clasificarlo.

Los eventos tonales en ToBI se clasifican en:

- acentos prenucleares, es decir, que no sea el último acento de grupo entonativo;
- acentos nucleares, el último de grupo entonativo;
- tonos de frontera.

Dado que los eventos tonales posibles en cada uno de estos casos son diferentes, las fórmulas que se tienen que aplicar para clasificarlos también lo son.

2.3.3. Acentos prenucleares

En el caso de los acentos prenucleares el script busca tres tipos de movimientos (diferencias) distintos:

- diferencias entre el centro de la sílaba pretónica, tónica y postónicas;
- diferencias entre el inicio y final de la tónica;
- diferencias entre el pico y el valle más próximos.

Como se ha explicado, las diferencias entre estos puntos se calculan en semitonos. Para ello, se sigue la fórmula propuesta por Noteboom (1997).

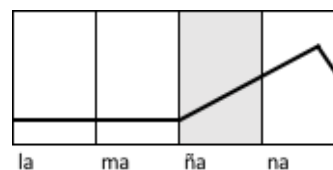
Si no hay ningún movimiento significativo (que sobrepase el umbral de 1.5 semitonos) entre ninguno de estos puntos, el script asigna un acento prenuclear monotonal. Este puede ser L* o H*. Si el acento es el primero de la frase, el nivel (L o H) se determinará calculando el rango del hablante en ese IP. Si la frecuencia de la sílaba que se está analizando es superior al 66 % del rango, el acento tonal (*pitch accent*) se etiqueta como H*; si no alcanza ese nivel, como L*. En el caso de que no sea el primer acento tonal de la IP, el script busca el acento anterior para determinar el tono dependiendo de esta etiqueta. Por ejemplo, si no hay movimiento tonal tras un acento L+H* o H*, el script etiquetará el tono como H*, pero si el tono anterior era L* o H+L* se etiquetará como L*.

En el caso de que haya un movimiento tonal el script aplica una serie de reglas (condiciones *if-then*) que determinan la transcripción del movimiento. En la actualidad el script contiene 64 reglas y 26 subreglas que permiten transcribir en español y en catalán, incluyendo los dialectos que han sido descritos para para el catalán en Prieto y Cabré (2013) y para el español en Prieto y Roseano (2010).

El funcionamiento del script se entiende mejor mediante un ejemplo. Imaginemos un movimiento tonal como el que aparece en la Figura 5, en la que cada recuadro representa una sílaba y el recuadro sombreado representa la sílaba tónica.

El script encontraría un movimiento significativo entre la pretónica y la tónica y adjudicaría el tono L*+H. Como la diferencia en la tónica también es significativa, sustituiría ese tono por L+H*. Después buscaría un movimiento entre tónica y postónica, y encontraría un descenso; por lo tanto, como los dos movimientos son significativos, adjudicaría un nuevo tono, con lo que quedaría un tono tritonal L+H*+L. En el caso de que aparezca un movimiento tritonal (como este), el script comprueba cuál de los dos movimientos que lo conforman es de menor rango y lo coloca entre paréntesis. Por lo tanto la transcripción fonética estrecha del tono sería L+(H*+L). Pero esta etiqueta no está recogida en el sistema Sp_ToBI¹, por lo que en la transcripción fonética ancha el movimiento de menor rango (el descenso después de la tónica) se simplificaría y pasaría a ser L+H*. Sin embargo, queda un último grupo de reglas que aplicar, las que calculan la diferencia entre pico y valle y la posición del pico. Al aplicar estas reglas, como el pico está en la postónica, la etiqueta volvería a sustituirse y se transcribiría como

Figura 5: Esquema de un movimiento tonal modelo en el pretonema. Cada recuadro representa una sílaba, y la tónica aparece sombreada.



L+<H*. En resumen, el script detectaría mediante una serie de reglas que el movimiento es un ascenso en la sílaba tónica con el pico en la postónica y lo transcribiría en consecuencia: en su transcripción fonética estrecha L+H*+L y en la transcripción fonética ancha L+<H*.

En el caso de los acentos descendentes, las reglas más complejas tienen que ver con el descuento de la declinación. Muchos de los movimientos descendentes que se pueden observar en una curva de f_0 no están causados por la existencia de una diana tonal baja en la curva, sino por la declinación. El módulo incluye un conjunto de reglas para evitar estos “falsos” tonos descendentes en la transcripción fonética ancha. En el prenúcleo, el problema radica en que algunos movimientos descendentes son en realidad sílabas tónicas desacentuadas. Para resolver este problema, cuando el módulo encuentra un acento descendente, exige que se cumpla además una de las dos condiciones siguientes: (1) que haya una diana tonal alta en la pretónica, es decir, que la pretónica no sea también descendente; (2) que haya un *plateau* –una ‘meseta’ alta– en las sílabas anteriores. Este supuesto se puede comprobar calculando si desde la sílaba anterior a la sílaba en la que se encuentra la tónica analizada no ha habido un descenso de más de 1.5 semitonos.

Además de estos movimientos, el módulo de transcripción fonética también etiqueta los casos en los que el rango aumenta considerablemente. Así, se han considerado como movimiento tonal superalto (¡H) los casos en que la diferencia entre dos puntos supera los 6 semitonos. Sin embargo, las cuestiones de rango son fonológicas en español y catalán en contadas ocasiones, por lo que en la capa fonética ancha el símbolo que indica nivel superalto desaparece la mayoría de veces. Se mantiene en el caso de L+¡H*, ya que está demostrado que contrasta fonológicamente con L+H* (Borràs-Comes, Vanrell, y Prieto, 2014).

Por otro lado, la etiqueta ¡ y, más concretamente, ¡H* también se utiliza en la transcripción para marcar que un movimiento tonal es un ascenso desde un punto ya alto, es decir, para transcribir ascensos desde una meseta alta a un punto aún más alto, como sería el caso de las interrogativas informativas en canarias (Vizcaíno Ortega, Cabrera Abreu, Dorta, & Hernández Díaz, 2007).

Los símbolos ¡ y ! no se usan en este transcriptor para marcar casos de escalonamiento tonal (*upstep*, *downstep*), por lo que el transcriptor se aparta en este aspecto de las convenciones aplicadas habitualmente en ToBI para el uso de estos símbolos.

¹La etiqueta L+H*+L aparece en el español de Argentina.

Los acentos monotonaes son los mismos, tanto en la transcripción estrecha como en la ancha.

2.3.4. Clasificación de acentos nucleares

Las reglas que se han explicado hasta ahora se podían aplicar sin tener en cuenta el tipo de acento de la palabra. Sin embargo, para los acentos nucleares esto no es posible, puesto que los puntos entre los que el script busca diferencias significativas tienen que ser distintos en el caso de las agudas y en el de las llanas y esdrújulas.

En el caso de las palabras llanas y esdrújulas el script busca diferencias entre 3 puntos de la pretónica (inicio, centro y final), 5 de la tónica (inicio, centro, final, mínimo y máximo) y 5 de la postónica (inicio, centro, final, mínimo y máximo). También toma el máximo valor en la tónica y el mínimo en las postónicas.

Las reglas que se aplican funcionan como en el caso de los acentos nucleares, con la única diferencia de que el inventario tonal es algo diferente. Por ejemplo, en posición tonemática no se pueden encontrar los tonos $L+<H^*$ o L^*+H , por lo que cuando estos aparecen en la transcripción fonética estrecha se reescriben en la transcripción ancha teniendo en cuenta el tono de frontera posterior.

Sin embargo, en el caso de las agudas, al no tener postónicas, el script solo puede tomar valores de la pretónica y de la tónica. Se considera que los primeros valores de la tónica se corresponderían con los movimientos tonales del acento nuclear tónico y los últimos, con los del tono de frontera, por lo que su clasificación se verá en la sección siguiente. De esta manera se resuelven los casos de compresión tonal.

2.3.5. Clasificación de tonos de frontera

Después de transcribir el acento nuclear, el script transcribe el tono de frontera. Para ello, las sílabas postónicas se dividen en 6 intervalos de tiempo iguales de los que se extrae la f_0 . También se identifican los valores mínimo y máximo de f_0 en las postónicas.

El script recupera para esta transcripción algunos de los datos de la última sílaba tónica (datos de la f_0 máxima y un dato nominal que indica si la sílaba tónica acabó en un nivel alto o bajo). El script calcula las diferencias en semitonos entre estos puntos y procede a realizar la transcripción. Por ejemplo, ante una diferencia mayor de 1.5 semitonos entre el primer punto de las postónicas y el último, el script considerará que ha habido un ascenso y transcribirá $H\%$. Pero si no ha habido ningún movimiento que pase el umbral pero la tónica acabó en un tono alto (por ejemplo, H^* o $L+H^*$), el tono de frontera resultante también será $H\%$.

Para poder transcribir los tonos medios el script usa el rango de la IP. El rango del hablante se divide en tres niveles (L , $!H$ y H) y si el final del movimiento tonal acaba dentro del nivel medio se transcribe como tono medio.

Como se avanzaba en la sección anterior, este análisis es diferente en el caso de las palabras agudas. Cuando el script detecta que la última sílaba acentuada es también la última sílaba de IP, la transcripción del

acento nuclear y el tono de frontera se realiza de manera conjunta.

La sílaba se divide en 12 partes iguales y se toma un valor de la sílaba pretónica y 8 de la tónica. De los 8 valores que se toman de la tónica, los 3 primeros corresponderían al acento nuclear y los restantes al tono de frontera. A partir de estos datos, las fórmulas asignan directamente una configuración nuclear completa.

Como se ha dicho, esta estrategia resuelve los casos de compresión y de diferencias en el alineamiento tonal pero no resuelve los casos de truncamiento, ya que en estos los tonos subyacentes no se pueden recuperar a partir de los datos acústicos. Estos casos requieren un conocimiento fonológico de la lengua que solo puede tener un investigador humano o un transcriptor prosódico que incluya información semántica del contexto para realizar desambiguaciones.

2.3.6. Estandarización de configuraciones nucleares

El último paso del módulo de anotación tonal es la estandarización de las configuraciones nucleares. El script usa 23 reglas para realizar esta tarea.

Los tonos nucleares y de frontera incluyen los tonos medio (*mid*) y superalto en la capa fonética estrecha siempre que se cumplan las condiciones acústicas necesarias, como por ejemplo que el rango supere los 6 semitonos.

Sin embargo, estos tonos de frontera solo son susceptibles de ser fonológicos cuando aparecen al lado de algunas configuraciones nucleares. Por ejemplo, en posición nuclear, $L+!H^*$ y $L+H^*$ solo pueden ser susceptibles de ser fonológicos si van seguidos de un tono $L\%$. Igualmente, el tono de frontera $!HH\%$ solo es posible cuando va precedido de $L+H^*$.

2.4. Salida

El script tiene dos salidas. La primera, por defecto, es el TextGrid con la transcripción ToBI y la segmentación prosódica. La segunda, opcional, es una figura con el resultado.

La anotación tonal resultante del proceso anterior se guarda en el TextGrid de entrada (Figura 6) en el que aparecen las transcripciones fonéticas estrecha y ancha en capas separadas, además de la segmentación prosódica y la segmentación en sílabas generadas por el primer módulo.

Figura 6: Onda sonora, espectrograma, curva de f_0 y anotación generado por FonetToBI correspondiente al enunciado del español 'fue inyectado en el abdomen y en una pierna', pronunciado por un hablante femenino.

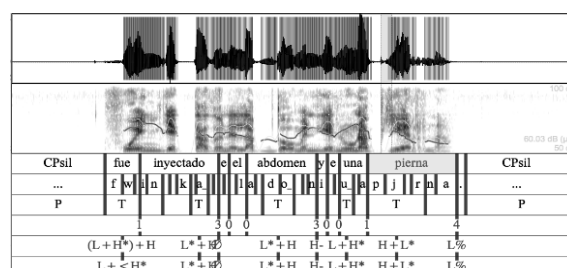


Tabla 2: Coincidencia en porcentaje, valor kappa y valoración delacoincidencia entre la transcripción de FonetitoBI y la del transcriptor humano del corpus del español.

| Evento tonal | n | % | Kappa | Evaluación |
|--------------|------|---------|-------|------------|
| APN | 1660 | 94.94 % | 0.907 | muy buena |
| AN | 1186 | 88.11 % | 0.831 | muy buena |
| TF | 1186 | 81.28 % | 0.756 | buena |

Tabla 3: Coincidencia en porcentaje, valor kappa y valoración delacoincidencia entre la transcripción de FonetitoBI y la de los transcriptores humanos del corpus del catalán.

| Evento tonal | T | % | Kappa | Evaluación |
|--------------|---|---------|-------|------------|
| AN | 1 | 85.71 % | 0.772 | buena |
| | 2 | 85.71 % | 0.770 | buena |
| | 3 | 82.65 % | 0.722 | buena |
| | 4 | 78.79 % | 0.657 | buena |
| TF | 1 | 92.86 % | 0.884 | muy buena |
| | 2 | 92.86 % | 0.885 | muy buena |
| | 3 | 93.88 % | 0.900 | muy buena |
| | 4 | 90.82 % | 0.851 | buena |

emisión, ya que el corpus estaba formado por frases aisladas. En final de emisión, las vocales pueden aparecer ensordecidas y, en general, hay una pérdida sustancial de cualidad de voz con la aparición de fenómenos como la voz rota (*creaky voice*) que pueden hacer que el algoritmo de detección de f_0 de Praat no funcione correctamente y aparezcan fallos.

3.3.2. Resultados para el catalán

Los resultados para el catalán (Tabla 3) varían dependiendo del transcriptor, pero aun así todos los valores obtenidos se sitúan en niveles buenos o muy buenos. Para los acentos nucleares los valores en porcentaje oscilan entre el 78 % y el 85 %, mientras que los kappa tienen valores de entre 0.65 y 0.77. Los tonos de frontera, sin embargo, muestran en catalán mayor porcentaje de acierto (entre el 90 % y el 93 %) y sus valores kappa se mueven entre el 0.85 y el 0.88.

3.3.3. Análisis cualitativo de los errores

En cuanto al tipo de errores que comete el script, se ha podido observar que para algunas configuraciones nucleares hay más discrepancias entre el script y los etiquetadores humanos que para otras.

En la matriz de confusiones de la Tabla 4, obtenida a partir de los resultados de la evaluación para el catalán, se puede comprobar que, en la mayoría de casos, las configuraciones nucleares han sido transcritas igual por el transcriptor automático y los humanos (diagonal de la tabla)

En la Tabla 4 se observa también que las configuraciones nucleares que suponen más problemas corresponden a, por un lado, confusiones entre L+H* L% y H* L% y, por otro, confusiones entre L+H* L% y L+_jH* L%. La primera confusión tiene que ver con la implementación fonética de los tonos fonológicos. Cuando en el acento L+H* hay un escalonamiento ascendente del tono L debido a un tono alto anterior,

dado que el tono fonético resultante es alto, el script no puede detectar la diana tonal baja por lo que transcribe solo los tonos altos que sí se han realizado. La segunda confusión tiene que ver con el rango de los eventos tonales. El requisito para etiquetar un tono como extra-alto se marcó en el script como una diferencia de más de 6 semitonos con la sílaba anterior. Este número se obtuvo a partir de la comparación de algunos estudios de percepción de la literatura. Sin embargo, viendo la cantidad de discrepancias entre el script y los etiquetadores parece claro que hay que revisar ese dato.

4. CONCLUSIONES Y POSIBLES MEJORAS

A la vista de los resultados obtenidos en las pruebas de evaluación, puede afirmarse que FonetitoBI es capaz de anotar prosódicamente enunciados simples en español y catalán con un nivel de precisión aceptable, similar al de los etiquetadores humanos, por lo que puede aplicarse con garantías al etiquetado automático de grandes corpus, y con un considerable ahorro en el tiempo de etiquetado. Sin embargo, hay que tener en cuenta que los test se llevaron a cabo con enunciados grabados en laboratorio y en los que la mayoría de sonidos eran sonoros, por lo que la fiabilidad del script podría disminuir si se usan corpus con calidad inferior o con abundantes sonidos sordos.

Otro aspecto que es importante resaltar es el hecho de que ofrezca, con un mismo sistema de transcripción, dos niveles diferentes de anotación tonal: uno, el de transcripción fonética estrecha, más basado en la forma acústica de la curva de f_0 , con un inventario de marcas tonales más rico (es decir, más fonético); y otro, el denominado aquí como transcripción fonética ancha, con un inventario de símbolos más reducido y más cercano a las convenciones establecidas en Cat_ToBI y Sp_ToBI, que pretenden recoger solo los eventos tonales contrastivos (más fonológico, por tanto). Esta posibilidad de transcribir fonética y fonológicamente con un mismo inventario de símbolos los eventos tonales de una lengua, existente ya desde hace tiempo para la transcripción de los elementos segmentales con el Alfabeto Fonético Internacional, aumenta de forma importante las aplicaciones potenciales de esta herramienta y del sistema mismo de anotación. Por ello, una de las tareas previstas para un futuro próximo es mejorar el inventario de etiquetas usado en el nivel fonético para permitir una transcripción lo más fiel posible de los eventos tonales observados en los contornos.

El hecho de que FonetitoBI admita también como entrada transcripciones fonéticas realizadas en SAMPA abre la posibilidad, como se ha mencionado al principio, de utilizar herramientas automáticas, como SPPAS o el propio Praat, para obtener el fichero TextGrid que requiere como entrada. Se podría llegar a obtener así, de forma completamente automática, una transcripción fonética completa (segmental y suprasegmental) a partir de un fichero wav con la onda sonora de un enunciado y de un fichero de texto con su transcripción ortográfica. Otra de las líneas en las que se está trabajando

Tabla 4: Matriz de confusiones de las diferencias entre el etiquetaje de FonetiToBI y los transcripores humanos en la evaluación del corpus en catalán.

| | H* L% | H+L* L% | L* H% | L+H* L% | L+ _i H* L% | L* HL% | L+H* L!H% | L+H* LH% | L* L% |
|-----------------------|-------|---------|-------|---------|-----------------------|--------|-----------|----------|-------|
| H* L% | 52 | 6 | | 4 | 1 | | | 1 | 4 |
| H+L* L% | 6 | 12 | | | | | | | 1 |
| L* H% | | | 179 | 4 | | 2 | | | |
| L+H* L% | 11 | | 4 | 7 | 20 | | | | |
| L+ _i H* L% | 1 | | | 20 | 11 | | | | |
| L* HL% | | | 2 | | | 18 | | | 3 |
| L+H* L!H% | | | | | | | 0 | 3 | |
| L+H* LH% | 1 | | | | | | 3 | 24 | |
| L* L% | 4 | 1 | | | | 3 | | | 8 |

actualmente es en el desarrollo de una herramienta que permita realizar esta tarea.

Finalmente, hay que destacar el hecho de que FonetiToBI es una herramienta de libre distribución, disponible bajo licencia GNU de manera gratuita en:

- <http://stel.ub.edu/labfon/en/praat-scripts>;
- <https://sites.google.com/site/juanmariagarrido/research/resources/tools/fonetitobi>

5. REFERENCIAS

- Bigi, B. (2012). SPPAS: a tool for the phonetic segmentations of speech *The Eight International Conference on Language Resources and Evaluation, Istanbul (Turkey)* (pp. 1748–1755). Recuperado de http://www.lrec-conf.org/proceedings/lrec2012/pdf/1116_Paper.pdf.
- Blum-Kulka, S. (1982). 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.
- Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *IFA Proceedings* 17, 97–110.
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International*, 5, 341–345.
- Borràs-Comes, J., Vanrell, M. del M., & Prieto, P. (2014). The role of pitch range in establishing intonational contrasts. *Journal of the International Phonetic Association*, 44, 1–20.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46.
- De Looze, C. (2010). *Analyse et interprétation de l'empan temporel des variations prosodiques en français et en anglais* (Tesis Doctoral). Aix-en-Provence: Université Aix-en-Provence.
- Elvira-García, W., Roseano, P., & Fernández Planas, A. M. (2015). Una herramienta para la transcripción prosódica automática con etiquetas Sp_ToBI en Praat. En A. Cabedo Nebot (Ed.) *Perspectivas actuales en el análisis fónico del habla: tradición y avances en la fonética experimental* (pp. 455–464). València: Universitat de València.
- Elvira-García, W., Roseano, P., & Fernández Planas, A. M. y Martínez Celadrán E. (2016). A tool for automatic transcription of intonation: Eti-ToBI a ToBI transcriber for Spanish and Catalan. *Language Resources and Evaluation*, 50(4), 767–792.
- Garrido, J. M. (2010). A tool for automatic F0 stylization, annotation and modelling of large corpora. *SpeechProsody 2010, Chicago, May 2010*. Recuperado de <http://speechprosody2010.illinois.edu/papers/100041.pdf>.
- Garrido, J. M. (2013). SegProso: A Praat-based tool for the automatic detection and annotation of prosodic boundaries. *Proceedings of TRASP 2013* (pp. 74–77). Recuperado de <http://www.lpl-aix.fr/~trasp/Proceedings/19864-trasp2013.pdf>.
- Goldman, J. P. (2011). EasyAlign: an automatic phonetic alignment tool under Praat. *Proceedings of InterSpeech, September 2011, Firenze, Italy*.
- Hirst, D. (2011). The analysis by synthesis of speech melody: from data to models. *Journal of Speech Sciences*, 1(1), 55–83.
- Hirst, D., & Espesser, R. (1993). Automatic modelling of fundamental frequency using a quadratic spline function. *Travaux de l'Institut de Phonétique d'Aix*, 15, 71–85.
- Hualde, J. I., & Prieto, P. (2016). Towards an International Prosodic Alphabet (IPrA). *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 7(1).
- International Phonetic Association (en línea). *The International Phonetic Alphabet and the IPA Chart*. Recuperado de <https://www.internationalphoneticassociation.org/content/ipa-chart>.
- Martínez Celadrán, E., & Fernández Planas, A.M. (2003). Taxonomía de las estructuras entonativas de las modalidades declarativa e interrogativa del español estándar peninsular estándar según el modelo AM en habla de laboratorio. En E. Herrera & P. Martín (Eds.), *La tonía: dimensiones fonéticas y fonológicas* (pp. 267–295). México DF: El Colegio

- de México.
- Pamies, A., Fernández Planas, A. M., Martínez Celdrán, E., Ortega-Escandell, A., & Amorós Céspedes, M. C. (2002). Umbrales tonales en español peninsular. *Actas Del II Congreso de Fonética Experimental* (pp. 272–278).
- Prieto, P. & Cabré, T. (Eds.) (2013). *L'entonació dels dialectes catalans*. Rubí: Publicacions de l'Abadia de Montserrat.
- Prieto, P., & Roseano, P. (Eds.) (2010). *Transcription of intonation of the Spanish language*. München: Lincom Europa.
- Rietveld, A., & Gussenhoven, C. (1985). On the relation between pitch excursion size and prominence. *Journal of Phonetics*, 13, 299-308.
- Roseano, P., & Fernández Planas, A. M. (2013). Transcripció fonètica i fonològica de l'entonació: una proposta d'etiquetatge automàtic. *Estudios de Fonética Experimental*, 22, 275–332.
- Vizcaíno Ortega, F., Cabrera Abreu, M., Dorta, J., & Hernández Díaz, B. (2007). La entonación de enunciados declarativos e interrogativos absolutos de Lanzarote. En J. Dorta (Ed.), *La prosodia en el ámbito lingüístico románico* (pp. 347–369). Santa Cruz de Tenerife: La Página Ediciones.