Una herramienta para la transcripción prosódica automática con etiquetas Sp_ToBI en Praat

ELVIRA-GARCÍA, WENDY; ROSEANO, PAOLO; FERNÁNDEZ PLANAS, ANA MA.; MARTÍNEZ CELDRÁN, EUGENIO

1 Introducción

Los estudios lingüísticos de las últimas décadas han resaltado la importancia de la prosodia y, en especial, la entonación, dentro del conjunto de los sistemas lingüísticos. Este hecho, junto con las posibilidades técnicas que ahora se abren para su estudio, ha posibilitado el aumento de los estudios sobre entonación. El carácter complejo de la entonación, donde hay que tener en cuenta simultáneamente cuestiones como la amplitud y la alineación de los movimientos tonales, conlleva que la manera idónea de estudiarla sea a través de transcripciones de la curva entonativa, para lo que se han venido creando durante el siglo XX diferentes modelos de entonación (para una reseña véase Prieto, 2002). Sin embargo, la transcripción prosódica exige una cantidad de recursos humanos y de pericia del investigador (Syrdal, Hirschberg, McGory & Beckman, 2001) que muchas veces no es rentable para el objetivo que persigue la investigación.

Esto ha hecho que desde la misma invención de los modelos de transcripción se haya intentado crear paralelamente instrumentos informáticos capaces de acelerar la labor del investigador (Lea, 1980; Rietveld, 1984). Por una parte, se han creado interfaces de usuario que ayudan a acelerar la transcripción manual (Syrdal *et al*., 2001). Y por otra, transcriptores prosódicos automáticos que

realizan un análisis de la curva de F0 y la etiquetan de acuerdo con las convenciones del sistema transcripción para el que fueron creados (Hirst y Espesser, 1993; Mertens, 2004; Rosenberg, 2010).

La transcripción prosódica automática, como el resto de los componentes de las tecnologías del habla, se puede abordar desde dos perspectivas. Por un lado, es posible llevarla a cabo usando modelos estadísticos predictivos y técnicas de *machine learning*. Por otro, se puede realizar implementando el conocimiento lingüístico disponible en la actualidad sobre una lengua.

Si bien es cierto que, hasta el momento, los sistemas de reconocimiento y síntesis de habla basados en estadística han demostrado ser los más productivos y son, de hecho, los que se han implementado los sistemas que se comercializan hoy en día, el análisis basado únicamente en el conocimiento lingüístico es posible y, aunque hasta ahora había sido menos fiable, permite arrojar luz sobre los fenómenos lingüísticos que todavía desconocemos.

Este trabajo tiene dos objetivos. Por una parte, nos proponemos presentar Eti-ToBI, una herramienta lingüística para la transcripción automática de la entonación del español que usa las convenciones de notación Sp_ToBI.

Por otro lado, este trabajo pretende mostrar que un transcriptor basado en el conocimiento lingüístico es posible y, por tanto, aplicable al reconocimiento de voz. De esta manera, se quiere estimular el desarrollo de los sistemas de reconocimiento de voz basados en conocimiento de los sistemas lingüísticos, para revalorizar la labor del lingüista en el campo de las tecnologías del habla.

1.1 Los transcriptores prosódicos automáticos basados en sistemas ToBI

De los modelos de transcripción prosódica existentes en la actualidad, las convenciones Tones and Break Indices (ToBI) constituyen, seguramente, el modelo que goza de más popularidad. Los sistemas ToBI son un conjunto de normas de transcripción prosódica específicas para cada lengua que están basadas en el modelo métrico-autosegmental o AM (Pierrehumbert, 1980). Esta teoría entiende la entonación como una sucesión de tonos en la que estos son autosegmentos que se anclan a las posiciones prominentes de la frase entonativa (sílabas acentuadas y límites de frase). A partir de estos fundamentos teóricos, se han creado una serie de convenciones para la transcripción prosódica fonológica de corpus, en un primer momento, para el inglés, MAE_ToBI (Beckman y Elam, 1997), y, más tarde, para otras lenguas entre las que se encuentra el español, Sp_ToBI (Beckman, Díaz-Campos, McGory y Morgan, 2002; Estebas-Vilaplana y Prieto, 2008; Face y Prieto, 2007; Prieto y Roseano, 2010).

Los sistemas de anotación ToBI, en principio, constan de cuatro niveles (*tiers*) de transcripción. En el primer nivel, se detalla la sucesión de tonos de la frase; en el segundo, la transcripción segmental; en el tercero, la separación prosódica de los componentes de la frase (*Break Indices*); y en el cuarto se anotan posibles comentarios. No obstante, en algunos sistemas ToBI (Jun, Lee, Kim & Lee, 2010), se ha decidido incluir un nivel más para posibilitar una transcripción fonética de la curva entonativa. Para el español, la propuesta de cariz fonético convive con la fonológica desde 2002 (Dorta, 2013; Fernández Planas, Martínez Celdrán, Salcioli Guidi, Toledo & Castellví Vives, 2002; Martínez Celdrán & Fernández Planas, 2003; Roseano & Fernández Planas, 2013).

En Sp_ToBI, se etiquetan dos tipos de eventos tonales, los acentos tonales (*pitch accents*) y los tonos de frontera (*boundary tones*). Los acentos tonales están

relacionados con la sílaba tónica y en la notación se marca con un asterisco (*) el tono que coincide con la misma. Los tonos de frontera pueden estar relacionados con el límite de una frase intermedia (-) o de una frase entonativa (%). El sistema ToBI (como el modelo AM) está, en principio, basado en niveles, es decir, anota los tonos dependiendo de si estos son altos (H) o bajos (L). Para el español, además de los dos niveles básicos propuestos en las primeras versiones del sistema, se contempla un nivel medio (!H) en los tonos de frontera y uno extraalto (¡H) en los acentos tonales. Así, la curva entonativa se configura como una sucesión de tonos de diferentes niveles. En Sp_ToBI, se prevén dos tipos de acentos tonales: monotonales (T*) y bitonales (T*+T o T+T*) y para los tonos de frontera, tres tipos: monotonales (T%), bitonales (TT%) y tritonales (TTT%). El último acento tonal de la frase entonativa junto con el tono de frontera configuran el núcleo o configuración nuclear, mientras que todos los acentos anteriores a ese se clasifican como prenucleares.

Este sistema, debido a su generalización de uso y a su especificidad para cada lengua, es probablemente el sistema que más aplicaciones de transcripción automática tiene. La primera de ellas que conviene considerar es anterior al propio nacimiento del sistema. Pierrehumbert (Pierrehumbert, 1983) creó un sistema de reconocimiento para el modelo AM que buscaba las inflexiones de la curva entonativa a través de derivadas para después aplicar unas condiciones que atribuían a cada movimiento su acento tonal correspondiente.

A partir de ese momento los sistemas de transcripción basados en AM dejaron de basarse solamente en el conocimiento lingüístico de la lengua que describen y pasaron a basarse en modelos de predicción estadística. Estos modelos predictivos, basados en técnicas de regresión y *machine learning* exigen un etiquetaje manual previo de un gran número de datos que usan como entrenamiento para predecir la transcripción de nuevos datos. Pero, por grande que sea la cantidad de datos que manejan, esta siempre es finita, lo que hace que sea virtualmente imposible conseguir un modelo de predicción capaz de prever todos los casos con estas técnicas. Los modelos de predicción utilizados dependían en gran parte del avance del reconocimiento de habla para unidades segmentales. Y esto hizo que las técnicas de reconocimiento de la prosodia utilizadas coincidieran con aquellas que estaban en boga en el mundo del reconocimiento de habla de las unidades segmentales en cada momento.

Así, encontramos modelos basados en árboles de decisión (Black & Hunt, 1996; Lee, Kim & Lee, 2002), regresión lineal (Rietveld, 1984), modelos de Markov (Ross & Ostendorf, 1996; Wightman & Ostendorf, 1994), principio de máxima entropía (Sridhar, 2008) y la combinación de varios de éstos (Wagner, 2008). Pero probablemente el transcriptor prosódico automático MAE_ToBI que más éxito ha tenido es el creado por Rosenberg (2010), programado en Java, no se limita a la transcripción entonativa, sino que funciona de manera completamente autónoma, localiza fronteras, los posibles acentos tonales y los etiqueta.

Además de para el inglés, ha habido intentos de crear transcriptores prosódicos para otras lenguas como el koreano (Lee *et al*., 2002), el japonés (Campbell, 1996; Noguchi & Kiriyama, 1999), el italiano (Savino, Refice & Daleno, 2002) o el sueco (Frid, 1999), entre otros.

1.2 Los transcriptores prosódicos automáticos del español

En la actualidad, también existen transcriptores prosódicos automáticos para el español. De ellos, algunos están basados en otros modelos prosódicos como

el IPO (Garrido Almiñana, 2008) y, en un caso, el transcriptor efectúa un análisis con el sistema Sp_ToBI (Escudero-Mancebo, González-Ferreras, Vivaracho-Pascual & Cardeñoso-Payo, 2014). Este transcriptor, como la mayoría de los mencionados hasta ahora, se apoya en métodos estadísticos, concretamente en la lógica difusa, para realizar una propuesta de etiquetaje que proporciona con un nivel de certeza y en última instancia es el investigador quien decide si esa propuesta de tonos es la más adecuada para cada caso. El transcriptor está siendo utilizado actualmente para acelerar la transcripción del corpus del español Glissando, únicamente con frases declarativas.

2 Metodología

Para cumplir los objetivos que se han expuesto en el apartado 1, se ha creado un script de Praat (Boersma y Weenink, 2014) que etiqueta los tonos en tres niveles: uno superficial o fonético, uno profundo o fonológico y uno estandarizado. Puede etiquetar, en el nivel superficial, 13 tipos de acentos prenucleares, 15 nucleares y 10 tipos de tonos de frontera, lo que supone un total de 150 configuraciones nucleares. En el nivel profundo, distingue entre 9 acentos prenucleares, 8 nucleares y 10 tonos de frontera: un total de 80 configuraciones nucleares que son posteriormente estandarizadas. El script puede reconocer los patrones descritos en Prieto y Roseano (2010); reconoce, por tanto, los patrones de variedades de español habladas en Castilla, Cantabria, islas Canarias, República Dominicana, Puerto Rico, zona andina de Ecuador y Venezuela, Chile, Argentina y México.

El script trabaja a partir de las sílabas tónicas de la frase, por lo tanto, su uso requiere que los audios tengan un TextGrid asociado con una separación de sílabas en un *tier* de intervalos y, en ese *tier*, una marca (a escoger por el investigador) en los intervalos que correspondan a una sílaba tónica.

2.1 Formalización del sistema

El primer paso para el análisis automático es el tratamiento de la señal. En este caso para procesar la señal y facilitar la búsqueda de F0 se han tomado dos medidas cautelares. La primera consiste en pasar un filtro que permite mantener solo las frecuencias inferiores a 1000 Hz; de esta manera se palian las distorsiones que los sonidos de altas frecuencias como las fricativas puedan causar en la curva de F0. La segunda medida consiste en calcular el rango de la frase entonativa automáticamente para cada producción. Esta operación se ha efectuado con la técnica en dos pasos descrita por Hirst (2011) en la que, en el primer paso, se extrae la curva de F0 de la frase con un rango muy amplio para obtener el mínimo y el máximo de la frase. En un segundo paso se extrae un nuevo objeto de F0 donde el mínimo del rango se establece multiplicando por 0,75 el primer cuartil del rango anterior y el máximo multiplicando por 1,5 el tercer cuartil (De Looze, 2010).

Seguidamente, se procede al análisis de la curva entonativa obtenida. Este análisis es posible gracias al uso de una transcripción en tres niveles que permite efectuar una primera transcripción fonética objetiva. El paso del nivel 1 al 3 de análisis nos lleva a una aproximación de la representación fonológica progresiva. Pero hay momentos en los que solo un etiquetador humano puede realizar una transcripción realmente fonológica. Es el caso, por ejemplo, de los procesos de truncamiento (con palabras agudas) en los que el tono de frontera no se llega a realizar en un plano superficial.

Como se ha resaltado para otras lenguas (Lee *et al*., 2002), la transcripción prosódica automática en sistemas ToBI es solo posible si se parte de una transcripción prosódica fonética, que constituye una de las características básicas de las aproximaciones del Laboratori de Fonètica de la UB (Fernández Planas *et al*., 2002; Martínez Celdrán y Fernández Planas, 2003; Roseano y Fernández Planas, 2013). En este primer nivel de etiquetaje, se transcriben todos los movimientos que son susceptibles de ser percibidos de manera lingüística, esto es los movimientos que superan 1,5 st (Pamies, Fernández Planas, Martínez Celdrán, Ortega-Escandell y Amorós Céspedes, 2002). Los movimientos se pueden buscar en puntos fijos determinados *a priori* por el script aplicando los conocimientos sobre la prosodia del español (ej. entre inicio y final de tónica) o en puntos determinados por la misma curva, es decir, diferencias entre el pico y el valle de un movimiento. A partir de estos movimientos, se establece un etiquetaje con etiquetas tritonales que dan cuenta de los mismos[1]. El resultado es una transcripción que podemos equiparar a la transcripción fonética estrecha en fonética segmental. Esta transcripción es independiente de la lengua, en el sentido en que cada movimiento se etiqueta usando símbolos que se corresponden con la información acústica de la curva.

El segundo nivel, el nivel profundo, es la equivalencia suprasegmental a una transcripción fonética ancha. En ella, los datos acústicos se interpretan de acuerdo a la fonología entonativa del español. Es en este nivel donde se tienen en cuenta las reglas de implementación fonológica específicas para el español. Estas reglas permiten simplificar las etiquetas tritonales que pudiera haber en el primer nivel y convertirlas en etiquetas bitonales relevantes para el español. En este nivel, además, se tienen en cuenta los movimientos descendentes de la curva y se comprueba si son fonológicos o causados por el fenómeno de declinación. Esto afecta a dos tipos de movimientos: en el prenúcleo, acentos tonales desacentuados, que fonéticamente se producen como H+L*; y en el núcleo, acentos descendentes fonéticos (H+L* L%) que en realidad no tienen una diana alta fonológica y son por tanto simplemente bajos (L* L%). Estas conversiones son posibles gracias a que en español un acento tonal H+L* consiste en un descenso dentro de la sílaba tónica donde el primer punto de la tónica está todavía en un nivel alto (Estebas-Vilaplana y Prieto, 2010; Prieto, 2014).

En el tercer nivel, el estandarizado, la transcripción del nivel profundo se normaliza para hacerla coincidir con el sistema actual de Sp_ToBI. Esta estandarización es necesaria únicamente para las configuraciones nucleares por: a) razones de convención del sistema (ya que el sistema Sp_ToBI no admite tonos en la cola en la configuración nuclear, así, algo como L*+H L% pasaría a ser L* HL%) o b) contrastes fonológicos que se dan solo para ciertas configuraciones, por ejemplo, !H% solo es fonológico cuando va detrás de L+H*.

3 Resultados

Para evaluar la actuación del script se ha comparado el etiquetaje proporcionado por este (en concreto, el etiquetaje del tercer tier o normalizado) con el realizado por un etiquetador humano. El análisis estadístico del nivel de acuerdo entre etiquetadores se ha llevado a cabo con la técnica de la *kappa* de Cohen, que sirve para medir el nivel de acuerdo entre dos etiquetadores descartando los acuerdos que hayan tenido lugar debido al azar (Cohen 1960;

[1] La última versión del sistema de etiquetaje superficial se puede consultar en Roseano y Fernández Planas (2013).

Cohen 1968), y se ha calculado con una herramienta en línea (GraphPad 2014). Los valores del estadístico oscilan entre -1 (desacuerdo total) hasta 1 (acuerdo perfecto) donde 0 correspondería al nivel del azar. Normalmente, un kappa a partir de .60 se considera bueno.

Para llevar a cabo la comparación se etiquetó un corpus basado en la emisión de 1186 frases producidas por 4 hablantes de 4 variedades diferentes de español peninsular: español de Cantabria, español de Madrid, español de Sevilla y español de Barcelona. Las hablantes, todas mujeres, tenían unas edades comprendidas entre los 21 y 28 años y eran informantes representativas de su punto de encuesta. Los datos se grabaron a través de una grabadora Marantz PMD620 conectada a un micrófono direccional Shure SM58 y las grabaciones se realizaron en casa de los informantes.

La transcripción prosódica manual fue realizada por un único experto transcriptor humano. La decisión de realizar la comparación con un sólo individuo se debe al gran tamaño del corpus. Sin embargo, esto no supone un problema metodológico debido a que el transcriptor ha participado anteriormente en experimentos de nivel de acuerdo entre etiquetadores ToBI y sus resultados han sido coherentes con los del resto de transcriptores (Roseano, Fernández Planas, Elvira García, Cerdà Massó y Martínez Celdrán, aceptado).

La comparación de las transcripciones se realizó dividiendo los eventos tonales en: 1) acentos prenucleares, 2) acentos nucleares, 3) tonos de frontera, ya que el script usa diferentes fórmulas para cada uno de ellos.

Entre todos ellos hay un nivel de acuerdo >80 % y valores de kappa por encima de .7. Si además tenemos en cuenta los valores de kappa ponderados, el nivel de acuerdo sube hasta .8 lo que se considera un muy buen nivel de acuerdo (tabla 1).

Evento tonal	n	% Acuerdo	Kappa	Kappa ponderado
Acentos prenucleares	1660	94.94 %	0.907	0.859
Acentos nucleares	1186	88.11 %	0.831	0.832
Tonos de frontera	1186	81.28 %	0.756	0.822

Tabla 1. Valores de acuerdo y kappa entre el etiquetador automático Eti-ToBI y el etiquetador humano

4 Discusión

Tal y como se ha visto, los resultados de fiabilidad del *script* son muy buenos, pero no alcanzan una fiabilidad del 100 %. Esto tiene dos causas. La primera de ellas es la limitación técnica que supone trabajar con algoritmos de predicción de F0 a partir de datos acústicos, y no trabajar a partir de los datos reales. Esto hace que, en final de emisión, donde hay ensordecimiento, el algoritmo de detección de pitch de Praat pueda interpretar erróneamente periodicidad donde no la hay y en esos casos el script coloca un tono en consecuencia.

La segunda es debida a un problema intrínseco del mismo sistema de transcripción, el Sp_ToBI. Es numerosa la bibliografía (tabla 2) que resalta que, aunque los sistemas ToBI son fiables, el grado de acuerdo entre diferentes etiquetadores nunca es perfecto. Por lo tanto, en un nivel fonológico, no así en el fonético, son abundantes los casos de ambigüedad donde existe más de una etiqueta posible. En estos casos, la etiqueta proporcionada por Eti-ToBI puede no coincidir con la etiqueta que darían los investigadores.

	Acentos tonales		Tonos de frontera		
Sistema	% acierto	Kappa	% acierto	Kappa	Fuente
Cat_ToBI	0.61	0.46	0.86	0.69	Escudero *et al*. 2012
AM_ToBI	0.72	0.67	0.86	0.84	Syrdal y McGory 2000
E_ToBI	0.86	0.51	0.89	0.79	Yoon *et al*. 2004
G_ToBI	0.71	-	0.86	-	Grice *et al*. 1996
K_ToBI	0.52	-	0.82	-	Jun *et al*. 2000

Tabla 2. Acuerdo entre etiquetadores humanos para diferentes sistemas ToBI.

A este último problema también se enfrentan los transcriptores basados en estadística, que en estos casos suelen calcular cuál es la etiqueta más probable y colocarla o dar al investigador los datos para que él elija. En el caso de Eti_ToBI nuestro objetivo no es proporcionar al investigador una etiqueta probable, sino proveer una etiqueta acústicamente posible y fonológicamente descriptiva. En cualquier caso, el único modo de salvar este obstáculo es conseguir una descripción de la entonación del español que dé cuenta de manera detallada de la interacción fonético-fonológica y esto sólo es posible a través de la investigación.

5 Conclusiones

En este trabajo se ha presentado un transcriptor prosódico automático con etiquetas Sp_ToBI basado únicamente en los datos acústicos y el conocimiento fonológico de la entonación, un instrumento que no usa técnicas de predicción estadística. El sistema es válido para un amplio abanico de variedades del español, de las cuales reconoce una gran cantidad de contornos.

El trabajo prueba que, a partir de las formalizaciones del conocimiento lingüístico adecuadas, es posible llegar a un nivel de fiabilidad equiparable al de los transcriptores humanos y los transcriptores automáticos basados en predicción estadística. En este caso, la formalización que ha permitido llegar a ese nivel de fiabilidad es un análisis prosódico fonético que se ha probado como una transcripción objetiva, clara y universal basada únicamente en datos acústicos. Esta transcripción, por no estar condicionada por la lengua, sería aplicable a otros sistemas ToBI.

6 Bibliografía

Beckman, M., Díaz-Campos, M., McGory, J. T. & Morgan, T. A. (2002). Intonation across Spanish, in the Tones and Break Indices framework. *Probus*, *14*, 9-36. doi:10.1515/prbs.2002.008

Beckman, M., & Elam, G. A. (1997). *Guidelines for ToBI Labelling*. The Ohio State University Research Foundation.

Black, A. W., & Hunt, A. J. (1996). Generating F 0 contours from ToBI labels using linear regression. In *ICSLP 96. Fourth International Conference on Spoken Language Proceedings* (pp. 1385-1388). Philadelphia: IEEE. doi:10.1109/ICSLP.1996.607872

Boersma, P., & Weenink, D. (2014). *Praat: doing phonetics by computer*. En http://www.praat.org/

Campbell, N. (1996). Autolabelling Japanese ToBI. En *ICSLP 96. Fourth International Congress on Conference on Language Processing Proceedings*, vol. 4, 2399-2402. Philadelphia: IEEE. En http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=607292

De Looze, C. (2010). Analyse et interprétation de l'empan temporel des variations prosodiques en

français et en anglais. Aix-en-Provence. En http://halshs.archives-ouvertes.fr/tel-00470641/

Dorta, J. (ed.). (2013). *Estudio comparativo preliminar de la entonación de Canarias, Cuba y Venezuela*. Madrid-Sta. Cruz de Tenerife: La Página ediciones.

Escudero, D., Aguilar, L., Vanrell, M. del M., & Prieto, P. (2012). Analysis of inter-transcriber consistency in the Cat_ToBI prosodic labeling system. *Speech Communication*, *54* (4), 566-582. En http://www.sciencedirect.com/science/article/pii/S0167639311001749

Escudero-Mancebo, D., González-Ferreras, C., Vivaracho-Pascual, C., & Cardeñoso-Payo, V. (2014). A fuzzy classifier to deal with similarity between labels on automatic prosodic labeling. En *Computer Speech y Language*, vol. 28, 326-341. doi:10.1016/j.csl.2013.08.001

Estebas-Vilaplana, E., & Prieto, P. (2008). La notación prosódica del español: una revisión del Sp-ToBI. *Estudios de Fonética Experimental*, *17*, 264-283. En http://www.raco.cat/index.php/EFE/article/view/140072/0

Estebas-Vilaplana, E., & Prieto, P. (2010). Castilian Spanish intonation. *Transcription of Intonation of the Spanish Language, Lincom Europa, München*, 17-48.

Face, T., & Prieto, P. (2007). Rising accents in Castilian Spanish: a revision of Sp-ToBI. *Journal of Portuguese Linguistics*, *6* (1), 117.

Fernández Planas, A. M., Martínez Celdrán, E., Salcioli Guidi, V., Toledo, G., & Castellví Vives, J. (2002). Taxonomía autosegmental en la entonación del español peninsular. En *Actas del II Congreso de Fonética Experimental*,180-186. Sevilla.

Frid, J. (1999). An environment for testing prosodic and phonetic transcriptions. En *Proceedings of ICPhS 99*, 2319-2322). San Francisco. En http://lup.lub.lu.se/record/529087/file/1624474.pdf

Garrido Almiñana, J. M. (2008). Modelling Spanish Intonation for Text-to-Speech Applications. Universitat Autònoma de Barcelona. En http://www.tdx.cat/handle/10803/4885

Grice, M., Reyelt, M., Benzmüller, R., Mayer, J., & Batliner, A. (1996). Consistency in transcription and labelling of German intonation with GToBI. In *Proceedings of the ICSLP*, 1716-1719). Philadelphia.

Hirst, D. (2011). The analysis by synthesis of speech melody: from data to models. *Journal of Speech Sciences*, *1*(1), 55-83. En http://www.journalofspeechsciences.org/index.php/journalofspeechsciences/article/viewArticle/21

Hirst, D., & Espesser, R. (1993). Automatic Modelling of Fundamental Frequency Using a Quadratic Spline Function. *Travaux de l'Institut de Phonétique d'Aix-En-Provence*, 75-85.

Jun, S. A., Lee, S. H., Kim, K., & Lee, Y. J. (2000). Labeler agreement in transcribing Korean intonation with K-ToBI. En *Interspeech*, 211-214.

Jun, S. A., Lee, S. H., Kim, K., & Lee, Y. J. (2010). Labeler agreement in transcribing korean intonation with K-ToBI. En *Interspeech '10*, 211-214.

Lea, W. (1980). Prosodic aids to speech recognition. In W. Lea (Ed.), *Trends in Speech Recognition* (pp. 166–205). Englewood: Prentice-Hall.

Lee, J., Kim, B. y Lee, G. (2002). Automatic corpus-based tone and break-index prediction using K-ToBI representation. *ACM Transactions on Asian Language Information Processing (TALIP)*, *1*(3), 207-224. doi:10.1145/772755.772757

Martínez Celdrán, 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 y P. Martín (Eds.), *La tonía: dimensiones fonéticas y fonológicas*, 267-294. México D. F.: El Colegio de México.

Mertens, P. (2004). The Prosogram: Semi-Automatic Transcription of Prosody based on a Tonal Perception Model. En B. Bel y I. Marlien (Eds.), *Proceedings of Speech Prosody* 2004, 23-26. Nara (Japan). En http://bach.arts.kuleuven.be/pmertens/papers/sp2004.pdf

Noguchi, H. y Kiriyama, K. (1999). Automatic labeling of Japanese prosody using J-ToBI style description. En *EUROSPEECH'99. Sixth European Conference on Speech Communication and Technology*, 2259-2262. En http://20.210-193-52.unknown.qala.com.sg/archive/ar

chive_papers/eurospeech_1999/e99_2259.pdf

Pamies, A., Fernández Planas, A. M., Martínez Celdrán, E., Ortega-Escandell, A., & Amorós Cespedes, M. C. (2002). Umbrales tonales en español peninsular. In *Actas del II Congreso de Fonética Experimental*, 272-278.

Pierrehumbert, J. (1980). *The Phonology and Phonetics of English Intonation*. MIT, Cambridge, Massachusetts.

Pierrehumbert, J. (1983). Automatic recognition of intonation patterns. En *Proceedings of the 21st annual meeting on Association for Computational Linguistics*, 85-90. Accesible en línea en http://dl.acm.org/citation.cfm?id=981328

Prieto, P. (2002). *Entonació: models, teoria, mètodes*. Barcelona: Ariel.

Prieto, P. (2014). The intonational phonology of Catalan. En S.-A. Jun (Ed.), *Prosodic typology*, vol. 2, 43-80. Oxford: Oxford University Press. En http://www.elebilab.com/documentos/archivos/publicaciones/3_GGT-08-04.pdf

Prieto, P., & Roseano, P. (eds.). (2010). *Transcription of Intonation of the Spanish Language*. München: Lincom Europa.

Rietveld, A. C. M. (1984). Syllaben, klemtonen en de automatische detectie van beklemtoonde syllaben in het Nederlands. Université de Nijmegen.

Roseano, P. y Fernández Planas, A. M. (2013). Transcripció fonètica i fonològica de l'entonació: una proposta d'etiquetatge automàtic. *Estudios de Fonética Experimental*, *XXII*, 275-332. En http://www.raco.cat/index.php/EFE/article/view/275413

Roseano, P., Fernández Planas, A. M., Elvira García, W., Cerdà Massó, R., & Martínez Celdrán, E. (aceptado). La entonación de las preguntas parciales en catalán. *Revista Española de Linguistica Aplicada*.

Rosenberg, A. (2010). AutoBI - a tool for automatic ToBI annotation. En *INTERSPEECH 2010, 11th Annual Conference of the International Speech Communication Association*, 146-149. Mihama, Japan. En http://eniac.cs.qc.cuny.edu/andrew/papers/autobi-is10.pdf

Ross, K., & Ostendorf, M. (1996). Prediction of abstract prosodic labels for speech synthesis. *Computer Speech and Language*, *10*(3), 155-185. En http://www.sciencedirect.com/science/article/pii/S0885230896900108

Savino, M., Refice, M., & Daleno, D. (2002). Methods and Tools for Prosodic Analysis of a Spoken Italian Corpus. En *Proceedings of the I International Conference on Language Resources and Evaluation*, 307-312. En http://lrec-conf.org/proceedings/lrec2002/pdf/101.pdf

Sridhar, V. (2008). Exploiting acoustic and syntactic features for automatic prosody labeling in a maximum entropy framework. En *IEEE Transactions on Audio, Speech, and Language Processing*, *16*(4), 797-811. En http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4453862

Syrdal, A. K., & McGory, J. T. (2000). Inter-transcriber reliability of ToBI prosodic labeling. En *INTERSPEECH*, 235-238.

Syrdal, A. K., Hirschberg, J., McGory, J., & Beckman, M. (2001). Automatic ToBI prediction and alignment to speed manual labeling of prosody. *Speech Communication*, *33*(1), 135-151. En http://www.sciencedirect.com/science/article/pii/S016763930000073X

Wagner, A. (2008). Automatic labeling of prosody. En *Proceedings of the 2nd ISCA Workshop on Experimental Linguistics, ExLing 2008*, 25-27. Athens, Greece. En http://isca-speech.org/archive_open/archive_papers/exling2008/exl8_221.pdf

Wightman, C., & Ostendorf, M. (1994). Automatic labeling of prosodic patterns. En IEEE Transactions on *Speech and Audio Processing*, vol. 2, 469-481. En http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=326607

Yoon, T., Chavarría, S., Cole, J., & Hasegawa-Johnson, M. (2004). Intertranscriber Reliability of Prosodic Labeling on Telephone Conversation Using ToBI. *Proceedings of ICSA International Conference on Spoken Language Processing, Interspeech 2004*, Jeju, Korea, 2729-2732.

Agradecimientos.

Este trabajo ha sido financiado mediante los fondos concedidos por el Ministerio de Economía y Competitividad para el proyecto AmperCat con referencia FFI2012-35998 y la beca predoctoral APIF-2012 de la Universidad de Barcelona.