

DE MELODÍAS Y VARIEDADES DEL ESPAÑOL

Miguel Mateo Ruiz

Dep. Filología Hispánica, Universitat de Barcelona

miguelmateoruiz@ub.edu

Resumen:

En este artículo presentamos los trabajos de tipo metodológico y de elaboración de corpus de habla espontánea que en el marco del modelo de Análisis Melódico del Habla (AMH) estamos desarrollando. Se centran básicamente en dos ámbitos: por una parte, la confección de *scripts* de PRAAT que permitan la extracción semiautomática de datos tonales (F0); por otra, la constitución de un amplio corpus de las variedades meridionales del español. Se presentan los principales datos del corpus elaborado y las funcionalidades de los programas de PRAAT. También, un breve resumen de los estudios realizados y los que están en curso o planificados, tanto a partir de la explotación y análisis del corpus como de desarrollo de nuevas funcionalidades metodológicas en PRAAT.

Palabras clave: Entonación, Habla espontánea, Corpus, Metodología AMH

Resum:

Aquest treball dona a conèixer els treballs de tipus metodològic i d'elaboració d'un corpus de parla espontània que estem desenvolupant dins del marc teòric del model Anàlisi Melòdic de la Parla (AMP). Dos són els àmbits d'aquests treballs: d'una banda, la construcció de *scripts* de PRAAT que facin possible l'extracció semiautomàtica de dades tonals (F0); d'altra, la creació d'un corpus molt ampli amb dades de totes les varietats meridionals de l'espanyol. Presentem les principals dades del corpus elaborat i les funcionalitats dels programes elaborats amb PRAAT. També, un breu resum dels estudis ja realitzats i els previstos. D'altra banda, les investigacions previstes per nous estudis del corpus obtingut i les noves millores dels programes de PRAAT, esmenes de tipus metodològic, doncs.

Paraules clau: Entonació, Parla espontània, Corpus, Metodologia AMP

Abstract:

This paper will present methodological work and preparation of a corpus of spontaneous speech that are developing within the framework of the model Melodic Analysis of Speech (MAS). Two aspects of this work are: firstly, the construction of PRAAT scripts that enable the semi-automatic extraction of data pitch (F0); on the other, creating a very large corpus with data from all southern varieties of Spanish. We present the main corpus data and functionalities developed programs made with PRAAT. Also, a brief summary of the studies already conducted and planned. Furthermore, research studies planned new corpus obtained and improvements programs PRAAT, methodological amendments.

Keywords: *Intonation, Spontaneous speech, Corpus, Methodology MAS*

1. INTRODUCCIÓN

El análisis melódico de las variedades del español es un proyecto iniciado en el año 2010, en el marco del proyecto *Análisis Melódico del Habla*. Para realizar este análisis de los rasgos de la entonación prelingüística, uno de los tres niveles de análisis de la entonación que contempla el modelo desarrollado a partir de la teoría de Cantero (2002), ampliado en Font-Rotchés (2007) y cuya última articulación teórica se puede encontrar, de forma resumida, en Cantero y Mateo (2011), de las diversas variedades peninsulares y de Canarias, se han dado pasos en dos direcciones: por una parte, la elaboración de amplios corpus de habla espontánea de las variedades del norte (Ballesteros, 2011) y meridionales (Mateo, 2013); por otra, la elaboración de unos programas –*scripts*, en terminología PRAAT– que permitieran realizar la extracción de los datos tonales (F0) de forma semiautomática (Mateo, 2010b).

Aquí describiremos de forma sucinta el proceso de elaboración y la composición del corpus de las variedades meridionales y las características de los programas de PRAAT.

2. ELABORACIÓN DEL CORPUS

El corpus elaborado sigue los criterios que ya se indicaban en Ballesteros, Mateo y Cantero (2011), aquí los presentaremos de forma sucinta y haremos una breve descripción de la composición de las variedades meridionales.¹

2.1. Criterios de elaboración

Una de las principales dificultades que se deben superar en la elaboración de un corpus de habla oral se refiere a la calidad de las grabaciones, se debe asegurar una calidad que permita un análisis fiable de las emisiones de los informantes. Este hecho explica que, de forma casi general en los estudios de la entonación se recurra a corpus creados mediante la denominada habla de laboratorio, con datos obtenidos de manera aislada, leídos –frases con poco significado comunicativo– o inducidos por el investigador. Creemos que esta forma de proceder permite trabajar con numerosos enunciados nítidos pero obvia a riqueza melódica de la conversación espontánea, natural.

Como ya señalamos en el artículo citado en la introducción de este apartado, «entendemos que la realidad es la que norma el acceso metodológico y no al revés» (Ballesteros, Mateo y Cantero, 2011).

Así pues, para que el corpus de trabajo ofrezca unas mínimas garantías deberá presentar las siguientes características:

1- Enunciados orales no planificados (habla espontánea): obtenidos en el marco natural de la entonación, el diálogo; no de manera preparada, artificiosa. Hemos desestimado la locución profesional.

2- Cantidad suficiente y equilibrada de contornos: corpus extenso de cada una de las cinco variedades: Andalucía, Canarias, Castilla La Mancha, Extremadura y Murcia. Con cantidad de contornos *suficiente* de cada una de ellas y del tipo de enunciado (neutro, preguntas, énfasis, suspensos).

¹ Las variedades del norte, en Ballesteros (2011): Castilla y León, Navarra, País Vasco, Asturias y Madrid.

3- Número amplio, representativo y equilibrado de informantes: tanto en cuanto a género como a edad, nivel de estudios, etc.

4- Buena calidad acústica: hemos recurrido a grabaciones de televisión de programas como concursos, tertulias, debates, etc. en las que los participantes no tienen sus intervenciones planificadas (descartamos, normalmente, la primera y la última), como sí puede suceder en un *reality show*. Además, los perfiles de los participantes son muy variados y ellos mismos o los conductores de los programas nos ofrecen datos para identificarlos como informantes (edad, estudios, población, etc.).

2.2. Las variedades meridionales

El corpus de las variedades meridionales está constituido por 3238 grupos fónicos emitidos por un total de 475 informantes, según el desglose que se muestra en la tabla de la figura 1.

Corpus	Grupos fónicos	Informantes
Andalucía Occidental (OCC)	428	84
Andalucía Oriental (ORI)	474	55
Canarias	532	75
Castilla La Mancha	478	65
Extremadura	838	105
Murcia	478	91
Total	3238	475

Figura1. Tabla de grupos fónicos e informantes.

Por lo que respecta al género, presenta también unos datos equilibrados, según la distribución que se muestra en el gráfico de la figura 2.

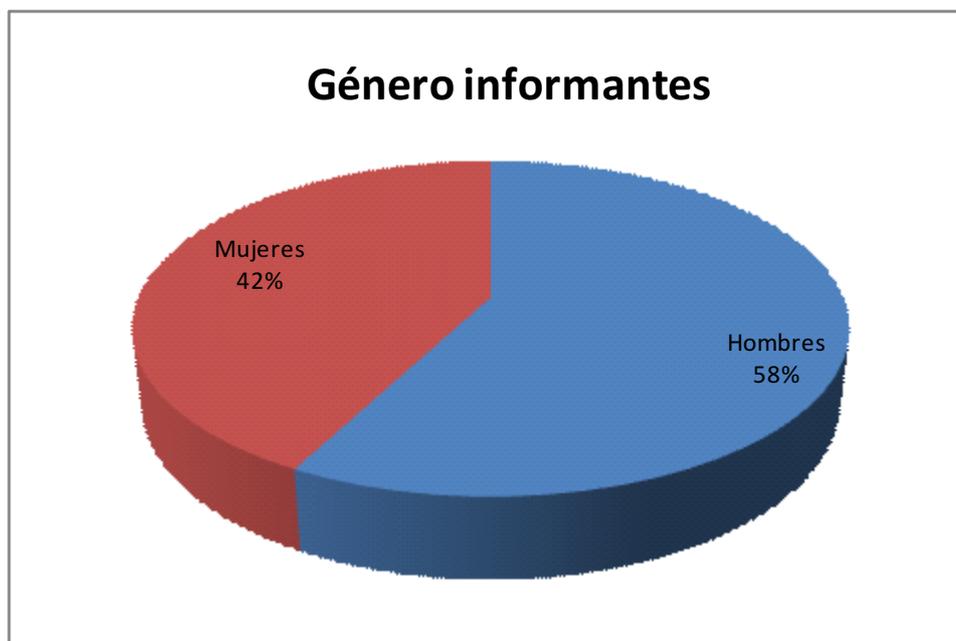


Figura2. Distribución informantes por sexo.

Para finalizar esta breve caracterización numérica del corpus de las variedades meridionales, mostramos la distribución de los informantes, en función de su población natal, en la figura 3, los datos de la zona meridional peninsular; los de Canarias, en la figura 4.

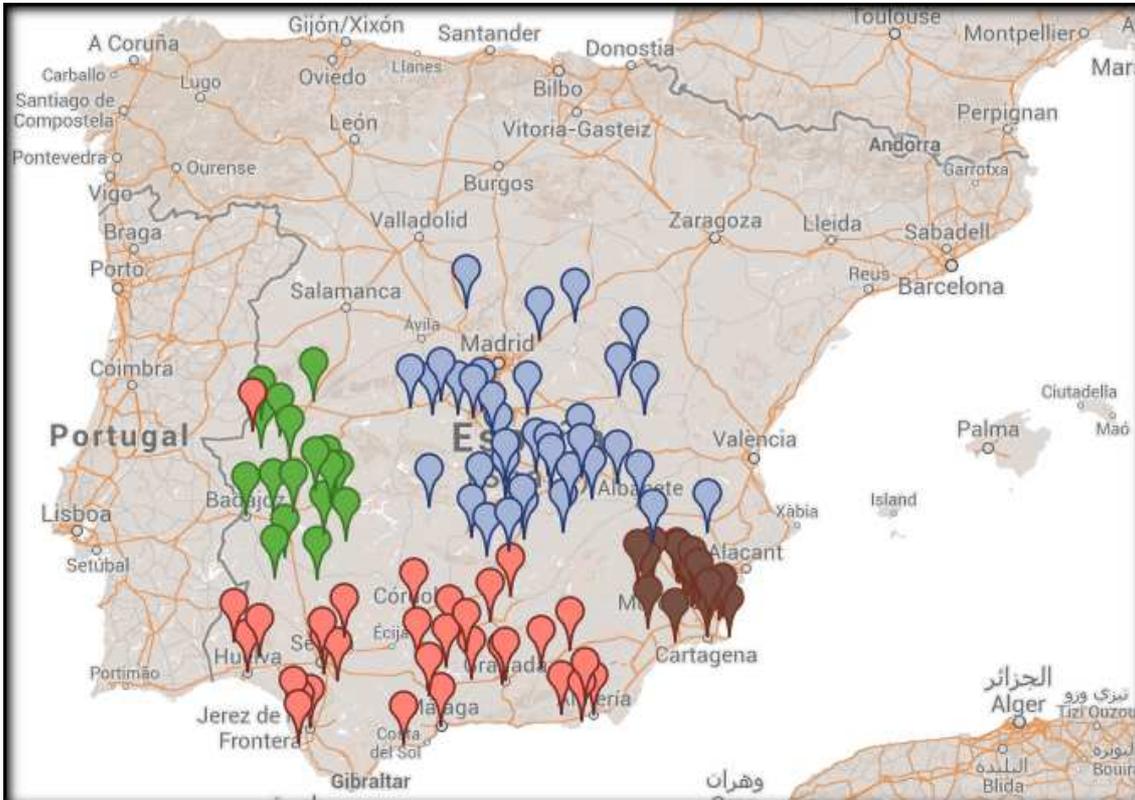


Figura 3. Mapa con las poblaciones de origen de los informantes, español meridional peninsular².



Figura 4. Mapa con las poblaciones de origen de los informantes, español de Canarias.

²©2013 Google, basado en BCN IGN España Imágenes ©2013 TerraMetrics.

3. LOS SCRIPTS

Los scripts desarrollados, descritos ampliamente en Mateo (2010b), han permitido elaborar un nuevo protocolo, Mateo (2010a) que aquí presentamos de forma breve:

- a) *Segmentación y etiquetaje de los segmentos tonales*: Es el punto clave de todo el proceso, el etiquetaje ha de ser preciso, el investigador ha de realizar este proceso con suma atención y la máxima precisión, recurriendo a la ampliación del espectrograma (botón *SEL* de la pantalla con sonido más el *textgrid* en que realizamos la segmentación y etiquetaje), modificación de los límites del rango de hercios, (en la pantalla *pitch settings*), atención a la onda sonora además del sonograma, de modo que pueda establecer una correcta segmentación del enunciado; en esta primera versión nuestra herramienta únicamente realiza la extracción de los datos de los segmentos tonales establecidos por el investigador.
- b) *Extracción de datos y confección de la curva estándar*: este paso lo realiza el primero de los scripts que describiremos; su detalle y la forma de instalarlo, se han incluido en el anexo I.
- c) *Revisión de alertas y corrección de datos*: a partir de la información del *script* anterior, se realizan las correcciones oportunas y se vuelve a generar la curva estándar de los enunciados modificados.
- d) *Generación de gráficos con la cura estándar y comparación de modelos*: este último paso consiste en la generación de los gráficos con la curva estándar, tal y como se hace en el protocolo actual, de recogida manual de datos, mediante documentos de Excel que incorporan el cálculo y confección del gráfico.

De forma gráfica, la implantación de los programas nos ha permitido pasar de la toma manual de datos, a la ejecución de una opción de menú de PRAAT, como podemos ver en la figura 5.

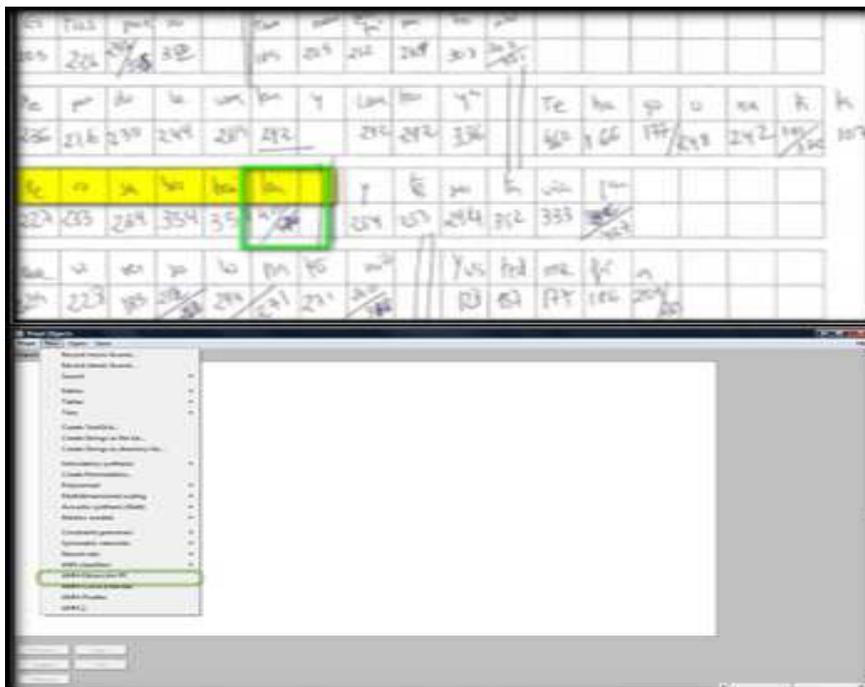


Figura 5. Ejemplo de toma manual de datos vs opción menú PRAAT.

4. CONCLUSIONES

En este apartado presentamos un resumen de los trabajos que el corpus de las variedades del español elaborado en el Laboratorio de Fonética Aplicada ha permitido desarrollar, en el caso del estudio de la entonación, además, estos trabajos se han podido desarrollar de una forma más ágil, con el nuevo protocolo de análisis (Mateo, 2010a) definido a partir de la implementación de los *scripts* de PRAAT construidos.

4.1. Trabajos realizados

A partir de los corpus elaborados, de 10 variedades dialectales, y la utilización de los programas de PRAAT, podemos señalar las siguientes contribuciones:

- 1) Descripción de la entonación prelingüística de las variedades del norte peninsular (Ballesteros, 2011) y de las meridionales (Mateo, 2013).
- 2) El estudio de la entonación lingüística, especialmente las interrogativas (Font-Rotchés y Mateo, 2011 y Rotchés y Mateo, en prensa).
- 3) El estudio de la interlengua, el español hablado por ingleses (Muñoz, 2013).
- 4) El análisis de la entonación paralingüística (Devís, 2011³ o Martínez Hernández, 2013).
- 5) Las implicaciones didácticas, en primer lugar, en el nivel de la entonación prelingüística (Cantero y Mateo, en prensa; Fonseca y Mateo, en prensa).
- 6) Consonantismo del español en habla espontánea: análisis y descripción de las consonantes aproximantes (Sola, en curso).

4.2. Trabajos futuros

Nuestro interés concreto con la elaboración y análisis del corpus de las variedades meridionales ha sido el poder completar una descripción de la entonación prelingüística del español que permita responder a alguna de las múltiples preguntas que C. de la Mota (2009:162-166) se plantea sobre los retos de la investigación fonética y fonológica de la lengua oral, especialmente las relativas a *lo fónico en la enseñanza de lenguas y la enseñanza de la fonética y la fonología*.

En consonancia con los diferentes tipos de objetivos que permiten la constitución de un corpus de las variedades del español y las posibilidades de automatización que ofrece PRAAT, señalamos a continuación las líneas de trabajo futuras que han de permitir seguir avanzando en el análisis y descripción de los diferentes niveles de la entonación.

4.2.1. PRAAT

Las tareas que nos planteamos desarrollar en un futuro están lógicamente relacionadas con las posibilidades que nos ofrece PRAAT y, por lo tanto, con la idea de tratar de simplificar el trabajo de los investigadores que podríamos denominar rutinario. Así, se pueden desarrollar nuevas mejoras en tres ámbitos principales:

³ Entre muchas otras, consultar: <http://www.ub.edu/dll/professorat/devis-herraiz-empar>.

- a) incorporación de nuevas alertas, por ejemplo una que permita identificar aquellos casos que hemos señalado en los que el algoritmo de PRAATse *equivoca*, considera el doble o la mitad de los ciclos.⁴
- b) Estudio y tratamiento de otros parámetros físicos, como por ejemplo la intensidad.

4.2.1. Corpus

Los trabajos que nos hemos planteado, algunos ya en curso, como la elaboración de los corpus señalada en el punto 1 o los análisis contrastivos (punto 5), son los siguientes:

- 1) Elaboración del corpus de Aragón y La Rioja.
- 2) Estudio de la entonación lingüística de las variedades meridionales. Este análisis debe incluir la validación perceptiva de los rasgos hallados y descritos en Mateo (2013), ha de permitir identificar los valores (% de variación tonal) que permiten caracterizarlos de forma relevante.
- 3) Comparación de los resultados con la descripción de la entonación del español del norte efectuada por Ballesteros (2011).
- 4) Estudio de la interacción de los rasgos melódicos en los diferentes niveles de la entonación, especialmente prelingüística y paralingüística.
- 5) Análisis contrastivos, de la entonación del español hablado por españoles con el español de otras comunidades de hablantes que no lo tienen como L1, por ejemplo con el español hablado por brasileños o ingleses.
- 6) Estudio de otros fenómenos acústicos: el amplio corpus constituido ha permitido empezar un análisis sistemático, en habla espontánea, del consonantismo del español, por ejemplo, las vibrantes y las laterales.
- 7) Estudio de características de tipo cinésico en la comunicación, además del corpus oral, disponemos de las grabaciones en vídeo, con la identificación exacta del momento de cada una de las intervenciones.

6. REFERENCIAS BIBLIOGRÁFICAS

- BALLESTEROS, M^a P. (2011): *La entonación del español del norte*. Tesis doctoral. Dep. de Filología Hispánica, Universitat de Barcelona [Documento disponible en: <http://hdl.handle.net/10803/109963>].
- BALLESTEROS, M^a P.; MATEO, M.; CANTERO, F. J. (2011): «Corpus oral para el análisis melódico de las variedades del español». *Actas del XXXIX Simposio de la SEL*. Santiago de Compostela, Universidade de Santiago de Compostela.
- CANTERO, F. J. (2002): *Teoría y análisis de la entonación*. Barcelona: Ed. de la Universitat de Barcelona.
- CANTERO, F. J.; DEVÍS, E. (2011): «Análisis melódico de la interlengua». En HIDALGO, A.; CONGOSTO, Y.; QUILIS, M. (eds.), *El estudio de la prosodia en España en el s. XXI: perspectivas y ámbitos*, València, Universitat de València, 285-299.

⁴ Sugerencia, entre otras, recibida de investigadores que están utilizando los *scripts* desarrollados.

- CANTERO, F. J.; MATEO, M. (2011): «Análisis Melódico del Habla: complejidad y entonación en el discurso». *Oralia*, 14, 105-127.
- CANTERO, F. J.; MATEO, M. (en prensa): «La entonación prelingüística del español: implicaciones didácticas». *XIII Congreso de la SEDLL*, Cádiz: Servicio de Publicaciones de la Universidad de Cádiz.
- DEVÍS, E. (2011): «La entonación de (des)cortesía en el español coloquial». *Phonica*, 7, 36-79 [Documento disponible en: <http://www.publicacions.ub.edu/revistes/documentos/733.pdf>].
- FONSECA, A.; MATEO, M. (en prensa): «A entonaçãõ pré-lingüística do español e do español falado por brasileiros: análise contrastiva». *Actas del VIII Congresso Internacional da ABRALIN*, Natal.
- FONT-ROTCHÉS, D. (2007): *L'entonació del català*. Barcelona: Publicacions de l'Abadia de Montserrat.
- FONT-ROTCHÉS, D.; MATEO, M. (2011): «Absolute interrogatives in Spanish: a new melodic pattern». *Actas do VII Congresso Internacional da ABRALIN*. Curitiba, 1111-1125.
- FONT-ROTCHÉS, D.; MATEO, M. (en prensa): «Entonación de las interrogativas absolutas del español meridional en habla espontánea». *Onomázein*.
- MARTÍNEZ HERNÁNDEZ, D. (2013): *Aproximación a la caracterización prosódica de la ironía en el discurso oral*. Trabajo Final de Máster, “Estudios Fónicos”. UIMP-CSIC.
- MATEO, M. (2010a): «Protocolo para la extracción de datos tonales y curva estándar en análisis melódico del habla (AMH)». *Phonica*, 6, 49-90 [Documento disponible en: <http://www.publicacions.ub.edu/revistes/phonica6/documentos/702.pdf>].
- MATEO, M. (2010b): «Scripts en Praat para la extracción de datos tonales y curva estándar». *PHONICA*, 6, 91-111 [Documento disponible en: <http://www.publicacions.ub.edu/revistes/phonica6/documentos/703.pdf>].
- MATEO, M. (2013): *La entonación del español meridional*. Tesis doctoral. Dept. de Didàctica de la Llengua i la Literatura. Universitat de Barcelona.
- DE LA MOTA, C. (2009): «El análisis fonético y fonológico de la lengua oral. Tendencias y retos actuales». En DE LA MOTA, C.; PUIGVERT, G. (eds.), *La investigación en Humanidades*, Madrid: Biblioteca Nueva.
- MUÑOZ, A. (2013): *Aproximación al perfil melódico de la interlengua de anglófonos que hablan español como segunda lengua*. Trabajo Final de Máster, “Estudios Fónicos”. UIMP-CSIC.
- SOLA, A. (en curso): *Caracterización acústica de las aproximantes /b/, /d/, /g/ en el habla espontánea del español peninsular*. Tesis doctoral. Dept. de Didàctica de la Llengua i la Literatura. Universitat de Barcelona.