

## **PROTOCOLO PARA EL ANÁLISIS MELÓDICO DEL HABLA**

FRANCISCO JOSÉ CANTERO SERENA  
*Universitat de Barcelona*  
cantero@ub.edu

DOLORS FONT ROTCHÉS  
*Universitat de Barcelona*  
dolorsfont@ub.edu

### **RESUMEN**

En este artículo presentamos un protocolo de análisis melódico del habla con medios instrumentales que permite hacer descripciones completas y objetivas de la entonación de una lengua, tanto desde el punto de vista fonético (melódico) como fonológico. Está basado en el método de base acústico-perceptiva de Cantero (2002) y Font (2007a), que ya se ha empleado en diversas investigaciones sobre la prosodia de diferentes lenguas.

Palabras clave: *entonación, análisis melódico, fonética, fonología, habla espontánea.*

### **ABSTRACT**

In this work, we present a protocol in melodic analysis of speech with instrumental means that allows us to carry out complete and objective descriptions of a language intonation, from a phonetic (melodic) point of view as well as from a phonological point of view. It is based on the acousticperceptive method by Cantero (2002) and Font (2007a), which has already been used in investigations about prosodic in different languages.

Keywords: *intonation, melodic analysis, phonetics, phonology, spontaneous speech.*

## 1. INTRODUCCIÓN

El método de análisis acústico-perceptivo que nos proponemos describir<sup>1</sup>, expuesto en Cantero (1995, 1999, 2002), se ha experimentado con éxito en diversos trabajos sobre la entonación del español (Cantero, 2002; Cantero & Font, 2007; Cantero et al., 2005), del catalán (Font, 2005, 2006, 2007a; Font et al., 2002), del francés (Espuny, 1997), así como en estudios sobre la adquisición del español (Cortés, 2000, 2004, 2005; Liu & Cantero, 2002; Liu, 2003) y en estudios multisistémicos (Torregrosa, 1999, 2006).

Frente a otros métodos de análisis de la entonación, nuestro método:

1. ofrece un criterio de segmentación de las melodías del habla exclusivamente fónico, independiente de cualquier otro nivel de análisis, con lo que podemos afrontar incluso el análisis de habla espontánea genuina.
2. presenta un sistema de procesamiento de los datos acústicos que nos permite obtener los valores relativos que constituyen las melodías, para compararlas, clasificarlas, reproducirlas con toda fidelidad, experimentar con ellas mediante la síntesis de voz, someterlas al análisis perceptivo y hacer generalizaciones lingüísticas.

Esto nos permite poder contar con múltiples informantes y un gran número de contornos, que podemos emplear, por ejemplo, para establecer con precisión los patrones entonativos de una lengua, aptos para ser aplicados en diversos ámbitos, como la enseñanza de lenguas, la reeducación de la voz y los trastornos del habla, así como la síntesis de voz, el reconocimiento del habla y los sistemas de diálogo, entre otros.

El protocolo de análisis melódico del habla que presentamos ha sido elaborado a partir de la propuesta de Cantero (2002), e implementado en la investigación de Font (2007a) sobre la entonación de la lengua catalana.

---

<sup>1</sup> Algunos aspectos del marco teórico de este artículo ya fueron publicados en Font Rotchés (2007b).

---

## **2. IDENTIFICACIÓN DE LAS UNIDADES MELÓDICAS EN UN CORPUS DE HABLA ESPONTÁNEA**

A menudo, los estudios de la entonación se plantean a partir de un corpus de «habla de laboratorio» (frases preparadas y leídas o, como mucho, inducidas por el investigador), para tener identificados y separados los enunciados cuya entonación quiere describirse. Esta dependencia de otros niveles de análisis (sintáctico o pragmático) tiene que ver con unos modelos teóricos que no han definido con suficiente claridad las unidades fónicas del habla, sino que más bien se limitan a describir las propiedades fónicas de las unidades gramaticales (por ejemplo, de las oraciones sintácticamente bien formadas).

Todo ello lleva a un tipo de análisis que no puede describir la realidad lingüística tal cual es, sino únicamente una realidad lingüística creada en el laboratorio, que, obviamente, no sólo no es muy «real», sino que además reproduce los propios planteamientos teóricos de partida del investigador.

En nuestros trabajos hemos desarrollado un modelo teórico que nos permite identificar con toda precisión las unidades fónicas del habla, independientemente de cualquier otro nivel de análisis. Ello nos permite describir la entonación de un corpus de habla no preparado ni inducido, totalmente ajeno a la intervención de los investigadores, con informantes anónimos y con habla espontánea genuina. Un corpus donde no podemos controlar los enunciados emitidos ni su forma, en ningún sentido, y donde nuestra intervención se limita a «seleccionar» los enunciados, sin ninguna necesidad de influir en ellos.

Nuestro modelo teórico (expuesto en Cantero, 2002, y en Font, 2007a) se basa en el concepto de jerarquía fónica, según el cual el habla está formada por unidades fónicas trabadas, bien jerarquizadas: la sílaba, el grupo rítmico (o palabra fónica) y el grupo fónico.

La unidad de análisis de la entonación es el grupo fónico (con una melodía determinada); la unidad de análisis del ritmo es el grupo rítmico (cuya melodía también es relevante dentro del contorno); y la unidad de análisis de la melodía es el segmento tonal (es decir, el valor tonal, relativo, del núcleo silábico: la vocal). Cada vocal, entonces, constituye un segmento tonal, excepto las vocales tónicas, que pueden constituir una inflexión tonal, es decir, dos (o más) segmentos tonales: como ocurre en el acento de frase (o acento sintagmático), que es el núcleo del grupo fónico y, por tanto, también el núcleo de la melodía: la inflexión final del contorno.

Así, para segmentar el habla espontánea en unidades melódicas se parte de un criterio formal: la presencia de una inflexión tonal que delimita el grupo fónico.

Ante un corpus de habla espontánea, y en un primer momento, conviene seleccionar los enunciados que coinciden con un turno de palabra del diálogo (normalmente breves y fáciles de identificar) hasta que el investigador no haya adquirido la suficiente confianza en la identificación de los grupos fónicos y en la delimitación de las melodías del habla.

Una vez identificadas las unidades melódicas (los grupos fónicos), hay que tratarlas como enunciados autónomos (aunque siempre manteniendo bien identificado el contexto tonal donde se ha producido cada uno, para poder relativizarlo más adelante, si conviene). Por ejemplo, a partir de una grabación original analógica, pueden digitalizarse y guardarse como archivos independientes (a partir de una grabación digital, bastará con seleccionar cada enunciado y guardarlo, también, como un archivo independiente). Esto nos facilitará mucho el análisis, ya que podremos tratar cada contorno independientemente de los otros, como una unidad de habla autónoma.

Cada contorno contendrá, en una ocasiones, sintagmas más o menos definidos y, en otras, oraciones más o menos completas gramaticalmente (ya que trabajamos con lengua oral espontánea): pero en ningún caso dependemos de estas unidades gramaticales para el análisis, porque las unidades que analizamos son unidades fónicas (contengan o no unidades gramaticales, enteras o no). Nosotros creemos que son más bien las unidades gramaticales las que ubican y se adaptan a su contenedor melódico, que es el auténtico estructurador del discurso oral: lo que llamamos entonación prelingüística.

### **3. FASE ACÚSTICA: DETERMINACIÓN DE LOS VALORES FRECUENCIALES RELEVANTES Y ESTANDARIZACIÓN**

El primer paso en el análisis melódico de un enunciado consiste en discriminar los valores frecuenciales relevantes (los segmentos tonales) de los valores irrelevantes (la  $F_0$  de las consonantes sonoras y de las glides). Es decir, una vez separados los enunciados determinamos el valor de  $F_0$  de cada vocal, de cada segmento tonal.

Utilizando un instrumento de análisis acústico fiable (nosotros utilizamos el software de análisis y síntesis Praat), identificamos las vocales (por ejemplo,

guiándonos con el sonograma) y determinamos su valor central (o bien la moda, o bien la media de los valores de  $F_0$  de la vocal, cuando es lo suficientemente estable).

Cuando se trata de una vocal tónica que contiene una inflexión tonal, hemos de determinar el valor de los dos segmentos tonales que constituyen la inflexión (o de los tres segmentos, si la inflexión es circunfleja). Estos valores se calculan a de los valores extremos de la inflexión. Véase dos ejemplos: uno del *castellano* *¡Y le daba cierto pudor!* en el que la vocal tónica final tiene dos valores (tabla 1), y otro del *atalán* en el que la vocal tónica final tiene tres valores *¡Bé crisi és això!* 'Crisis es esto' (tabla 2):

Enunciado	pu	<b>dor</b>	<b>dor</b>
Herzios	246	<b>315</b>	<b>270</b>

Tabla 1. *Vocal tónica final con dos valores.*

Enunciado	tre	<b>ba</b>	lla
Herzios	311	<b>204</b>	334

Tabla 2. *Vocal tónica final con tres valores.*

Otras veces, la inflexión tonal final comienza en una vocal tónica y acaba en la vocal átona, como ocurre en las inflexiones finales que coinciden con una palabra llana. Véase el ejemplo del *castellano* *Hombre, yo espero que sea pronto* (tabla 3) y el del *atalán* *Vostè treballa? '¿Usted trabaja?'* (tabla 4):

Enunciado	<b>pron</b>	to
Herzios	<b>86</b>	143

Tabla 3. *Vocal tónica final en palabra llana.*

---

Enunciado	ai	xò	xò	xò
Herzios	329	<b>301</b>	<b>394</b>	<b>276</b>

Tabla 4. *Vocal tónica final en palabra llana.*

En el caso de que la inflexión final acabe en una vocal seguida de una consonante nasal o lateral, esta consonante sonante podría constituir, por sí misma, el último segmento tonal de la inflexión. Véase el ejemplo del castellano *Es muy difícil* (tabla 5) y el del catalán *És cantant?* ‘¿Es cantante?’ (tabla 6):

Enunciado	Es	muy	di	fí	cil
Herzios	177	230	176	279	<b>160</b>

Tabla 5. *Vocal final seguida de lateral.*

Enunciado	És	can	tant	Nt
Herzios	231	251	218	<b>426</b>

Tabla 6. *Vocal final seguida de nasal.*

Los valores absolutos obtenidos en esta primera fase del análisis (la extracción y la determinación de la  $F_0$  de los segmentos tonales, medida en Hz), no constituyen aún, sin embargo, la melodía del contorno, sino que sólo son datos en bruto que han de procesarse adecuadamente.

El segundo paso en el análisis melódico es, pues, la estandarización de los datos frecuenciales. La melodía del contorno no es la mera sucesión de valores frecuenciales, sino su relación. Es decir, la melodía no es una sucesión de valores absolutos, sino una sucesión de valores relativos: una sucesión de intervalos.

Los datos en Hz, por tanto, han de relativizarse para poder describir la melodía del

contorno. La melodía que constituye la sucesión de valores 100Hz–200Hz no es igual que la melodía que constituyen los valores 200Hz–300Hz, aunque su diferencia es la misma: en los dos casos, 100Hz. En términos absolutos, la diferencia es la misma (el mismo número de Hz), pero no en términos relativos: el intervalo tonal no es el mismo. En el primer caso, la diferencia es de un 100%; pero en el segundo caso la diferencia es de un 50%. Así pues, entre 100Hz y 200Hz hay el mismo intervalo que entre 200Hz y 400Hz (un 100%), y entre 200Hz y 300Hz hay el mismo que entre 300Hz y 450Hz (un 50%).

Algunos autores expresan este intervalo usando como unidad del semitono, que tiene como ventaja ser una unidad directamente logarítmica, empleada ya para medir los intervalos musicales. En la expresión de las melodías del habla, con todo, utilizar semitonos implica trabajar con una fórmula de estandarización algo compleja, y con intervalos expresados, normalmente, con decimales (cosa que no ocurre en música, donde los intervalos siempre son los mismos y donde la unidad mínima es, por definición, el semitono).

La ventaja de utilizar los porcentajes es que son mucho más intuitivos, ya que permiten expresar linealmente un fenómeno logarítmico: calculamos (con una sencilla regla de tres) el porcentaje de variación de cada valor absoluto respecto del valor anterior, expresando el ascenso como porcentaje positivo y el descenso como porcentaje negativo (véase la tabla 7):

Enunciado	És	muy	di	<b>fi</b>	cil
Herzios	177	230	176	<b>279</b>	160
Porcentaje	100%	29,9%	-23,5%	<b>58,5%</b>	-42,6

Tabla 7. Intervalo tonal entre una vocal y la siguiente expresada en porcentajes.

Con este cálculo obtenemos los valores relativos de este contorno, que constituyen el algoritmo: 100% +29,9% -23,5% +58,5%-42,6%. Este algoritmo es la expresión de la melodía.

Para hacer la representación gráfica de esta melodía convertimos los porcentajes obtenidos a valores estándares: comenzando, por ejemplo, por el valor 100 (un

valor arbitrario), le aplicamos el porcentaje de ascenso que ha sufrido el segundo segmento, 29,9%, y obtenemos el segundo valor, 130; a este, se le aplica el porcentaje de descenso del tercero, -23,5%, y obtenemos el tercer valor, 99; y así, sucesivamente, hasta el último segmento. De este proceso resultan los siguientes valores (véase la tabla 8):

Enunciado	Es	muy	di	fi	cil
Herzios	177	230	176	<b>279</b>	160
Porcentaje	100%	29,9%	-23,5%	<b>58,5%</b>	-42,6
Curva estándar	100	130	99	<b>158</b>	90

Tabla 8. *Valores absolutos y estándares o relativos de Es muy difícil.*

Con los valores estandarizados de cada enunciado, hacemos la representación gráfica de cada contorno, con la finalidad de poder compararlos formalmente. Véase, a continuación, la representación gráfica del contorno analizado (figura 1):

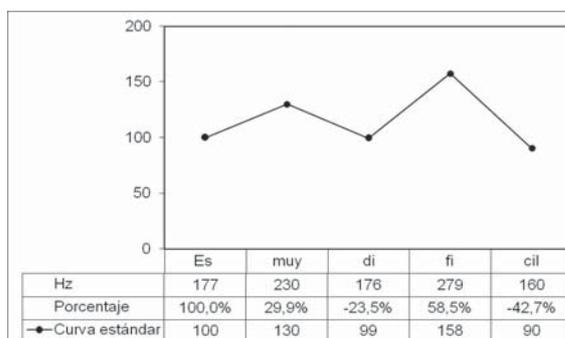


Figura 1. *Curva estandarizada de Es muy difícil.*

Como veremos en la fase perceptiva, los valores estandarizados de un contorno entonativo se validan al comprobar en el sintetizador de voz que constituyen una melodía idéntica a la original, pero con una tesitura de voz diferente.

Con la aplicación de esta primera fase del método, o fase acústica, se obtiene la estandarización de los contornos, que ahora ya son comparables y clasificables, independientemente de la edad, el sexo o cualquier otra característica del informante, ya que se han extraído todas las variaciones micromelódicas y se han normalizado los valores (a partir de un valor arbitrario 100). Tampoco hay que preocuparse por el número de informantes del corpus, ya que con la aplicación del método todos los contornos obtenidos son directamente comparables.

Este proceso no normaliza la duración de los diferentes contornos, sino que la abstrae y la sustituye con el número de segmentos tonales y su valor relativo, único dato relevante en un análisis melódico. Las diferencias entre dos contornos de diferente duración deberemos entenderlas desde una perspectiva melódica, independientemente de si la melodía contiene una palabra o muchas, si contiene un enunciado más o menos largo (o si contiene un enunciado más o menos correcto gramaticalmente), etc.

De hecho, una de las virtudes de nuestro método de análisis es que es exclusivamente melódico.

Veamos, por ejemplo, la representación de dos contornos del catalán bien diferentes, pero con una misma inflexión final (IF) descendente-ascendente, que los caracteriza a ambos como pertenecientes a una misma clase o patrón melódico (figuras 2 y 3). Independientemente de su longitud o del significado de cada enunciado, la melodía es, desde un punto de vista estrictamente melódico, la misma:

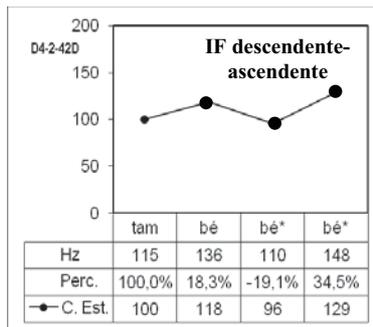


Figura 2. Curva estandarizada del enunciado Tam-bé ‘también’.

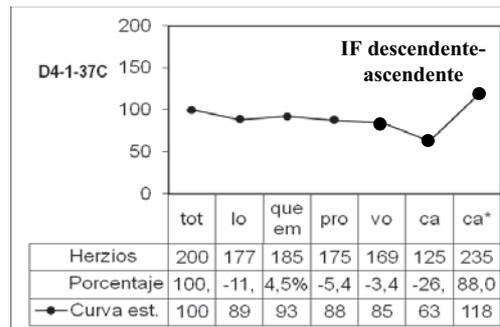


Figura 3. Curva estandarizada del enunciado Tot lo que em provoca! ‘¡todo lo que me provoca!’.

#### 4. FASE PERCEPTIVA: VALIDACIÓN DEL ANÁLISIS E INTERPRETACIÓN DE LOS DATOS

Después de la fase acústica, podemos comprobar la validez de nuestros resultados mediante una serie de pruebas perceptivas en las que sometemos a juicio de los oyentes una copia exacta (por síntesis de voz) de la frase analizada.

Utilizamos el programa Praat para obtener esta copia sintetizada (con el método PSOLA), de la cual borramos todos los datos originales y los sustituimos por nuestros datos estandarizados. Así comprobamos que, efectivamente, el análisis melódico ha sido correcto y refleja la melodía original, sin variaciones micromelódicas y con los valores normalizados.

El siguiente paso es, evidentemente, interpretar adecuadamente la melodía obtenida: extraer los datos melódicos relevantes que permiten una interpretación del contorno, por ejemplo una interpretación fonológica.

En nuestro modelo teórico, distinguimos entre los rasgos melódicos de los contornos (que constituyen el nivel fonético de la entonación) y los rasgos fonológicos (que permiten establecer los tonemas o unidades fonológicas de la entonación).

Los rasgos fonológicos que contemplamos son: /± Interrogativo/ /± Enfático/ y /± Suspendido/, cuya combinación nos ha permitido caracterizar los tonemas de la lengua (Cantero, 2002). Estos rasgos fonológicos han sido suficientes para clasificar todos los contornos de un corpus de 580 enunciados del catalán (Font, 2007a).

Los rasgos melódicos, por su parte, son las características de los elementos funcionales del contorno: el anacrusis, el primer pico, el cuerpo (o declinación) y la inflexión final. Entendemos por anacrusis las sílabas átonas previas a la primera vocal tónica del contorno, llamada primer pico; y por cuerpo las sílabas que van desde el primer pico a la última vocal tónica del contorno a partir de la cual comienza la inflexión final.

La descripción de estos elementos (especialmente, de la inflexión final) nos permite definir la melodía del contorno. También podemos establecer los patrones melódicos típicos de nuestros corpus (como «contornos tipo» de los tonemas) y sus márgenes de dispersión.

Así, por ejemplo, el patrón melódico con una inflexión final ascendente (+80%), que hemos descrito en castellano y en catalán y que se caracteriza por ser

/+interrogativo –enfático –suspensivo/ (véase la figura 4) es representativo de todos los contornos de las dos lenguas que empiezan en un ascenso hasta al primer pico (opcional) de un 40% como máximo, continúan con una declinación suave y constante hasta la última sílaba tónica y acaban en una inflexión final con un ascenso de un 80% o superior.

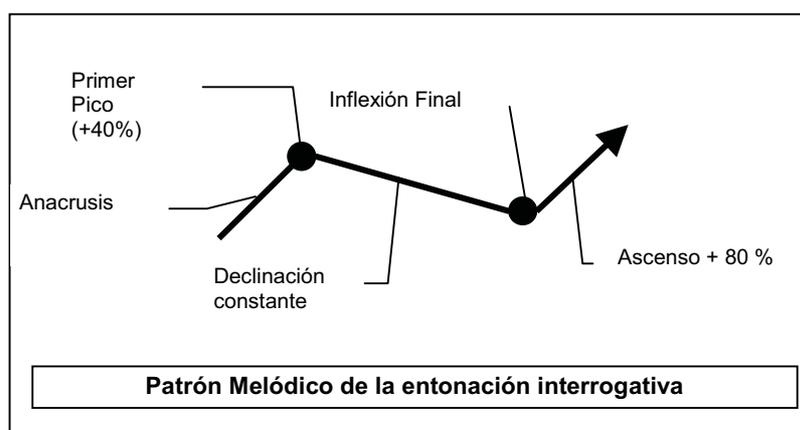


Figura 4. Representación del patrón ascendente (+80%) y sus rasgos melódicos.

Mediante la síntesis de voz, podemos modificar cada uno de estos rasgos con mucha precisión (gracias a la objetividad del método de análisis), para establecer los límites a partir de los cuales los cambios melódicos implican cambios de sentido para los oyentes.

Según nuestro método, el patrón melódico no es una mera representación de una línea con ascensos y descensos, sino que es una abstracción de la realidad hablada, representativa de múltiples melodías con los rasgos melódicos comprendidos entre sus márgenes de dispersión (definidos con claridad y cuantificados objetivamente).

Estos márgenes de dispersión del patrón nos permiten entender los múltiples contornos en la lengua que tienen una inflexión similar: son las variaciones de que nos servimos los hablantes para transmitir intenciones, emociones y otros contenidos expresivos o bien para evidenciar diferencias socioculturales o dialectales (aunque su valor fonológico continúa siendo, en todo caso, el mismo).

Como resultado de la aplicación del método, hemos obtenido diversos patrones para el castellano y para el catalán y cabe decir que algunos de ellos existen en ambas lenguas. Son coincidentes el patrón descendente-ascendente, patrón enfático que hemos ejemplificado con enunciados del catalán en los gráficos 2 y 3, y el también patrón enfático ascendente-descendente, el patrón ascendente (de 10% a 80%), que es un patrón suspenso, y el patrón descendente (que presenta en la inflexión final un ascenso hasta un 10 % o un descenso), que es el patrón neutro por antonomasia. En castellano hemos definido hasta 12 patrones (véase Cantero & Font, 2007) y en catalán hasta ocho (véase Font 2007a).

Ahora bien, los contornos pueden presentar otros rasgos melódicos (ascensos o descensos pronunciados en el cuerpo, declinación plana, ausencia de inflexión final, etc.) que constituyen alteraciones de los rasgos melódicos típicos del patrón y suponen un cambio de signo fonológico, de tal manera que un contorno /-interrogativo/ puede convertirse en uno /+interrogativo/, o un contorno /-enfático/ en uno /+enfático/, etc. Estos rasgos son variantes melódicas que deben definirse mediante la comparación formal y validarse perceptivamente.

Por ejemplo, un contorno del patrón descendente caracterizado por ser (-interrogativo -enfático -suspendido/, si presenta un declinación muy pronunciada o un ascenso superior al 30% en cualquier palabra del cuerpo, cambia de signo: se convierte en /+enfático/; si, en cambio, no tiene inflexión final, eso también implica un cambio de signo: en este caso, se convierte en /+suspendido/.

Por consiguiente, el método no sólo nos permite definir los patrones melódicos de una lengua con unos márgenes de dispersión establecidos con unos límites precisos, sino que también nos permite determinar con la misma precisión las variantes que puede presentar cada patrón.

## 5. REFERENCIAS BIBLIOGRÁFICAS

CANTERO SERENA, F. J. (1995): *Estructura de los modelos entonativos: interpretación fonológica del acento y la entonación en castellano*, tesis doctoral publicada en microforma (1997), Universitat de Barcelona.

CANTERO SERENA, F. J. (1999): «Análisis melódico del habla: principios teóricos y procedimientos», *Actas del I Congreso de Fonética Experimental*, Tarragona, pp. 127-133.

- 
- CANTERO SERENA, F. J. (2002): *Teoría y análisis de la entonación*, Barcelona, Edicions de la Universitat de Barcelona.
- CANTERO SERENA, F. J.; R. ALFONSO; M. BARTOLÍ; A. CORRALES y M. VIDAL (2005): «Rasgos melódicos de énfasis del español en habla espontánea», *Phonica*, vol. 1, en línea.  
<http://www.ub.es/lfa>
- CANTERO SERENA, F. J. y D. FONT ROTCHÉS (2007): «Entonación del español peninsular en habla espontánea: patrones melódicos y márgenes de dispersión», *Moenia*, 13.
- CORTÉS MORENO, M. (2000): *Adquisición de la entonación española por parte de hablantes nativos de chino*, tesis doctoral publicada en microforma, Universitat de Barcelona.
- CORTÉS MORENO, M. (2004): «Análisis acústico de la producción de la entonación española por parte de sinohablantes», *Estudios de Fonética Experimental*, XIII, pp. 79-110.
- CORTÉS MORENO, M. (2005): «Análisis experimental del aprendizaje de la acentuación y la entonación españolas por parte de hablantes nativos de chino», *Phonica*, vol. 1, en línea.  
<http://www.ub.es/lfa>
- ESPUNY, J. (1997): «Aspects prosodiques du discours hétérogène diaphonique», *Estudios de Fonética Experimental*, VIII, pp. 271-295.
- FONT ROTCHÉS, D. (2005): *L'entonació del català. Patrons, tonemes i marges de dispersió*, tesis doctoral, Universitat de Barcelona.  
<http://www.tesisenxarxa.net/TDX-0802106-114003/>.
- FONT ROTCHÉS, D. (2006): *Corpus oral de parla espontània. Gràfics i arxius de veu*, *Phonica*, 4, en línea  
<http://www.ub.edu/lfa/biblioteca.htm>.
- FONT ROTCHÉS, D. (2007a): *L'entonació del català*, Barcelona, Publicacions de l'Abadia de Montserrat, Biblioteca Milà i Fontanals, 53.

FONT ROTCHÉS, D. (2007b): «El mètode Anàlisi melòdica de la Parla aplicat al català», en J. Carrera y Claudia Pons (eds.): *Aplicacions de la fonètica*, Barcelona, Promociones y Publicaciones Universitarias.

FONT ROTCHÉS, D.; A. CANALS; G. ESTER; A. HERMOSO y F. J. CANTERO (2002): «Patrones melódicos de la entonación interrogativa del catalán en habla espontánea», *Actas del II Congreso de Fonética Experimental*, Sevilla, pp. 192-197.

LIU, Y.H. (2005): *La entonación del español hablado por taiwaneses*, *Biblioteca Phonica*, 2, pp. en línea.  
<http://www.ub.es/lfa>

LIU, Y.H. y F. J. CANTERO (2002): «La entonación prelingüística del español hablado por taiwaneses: establecimiento de un corpus», *Actas del II Congreso de Fonética Experimental*, Sevilla, pp. 238-242.

TORREGROSA, J. (1999): «Correlación de patrones entonativos y kinésicos: análisis de un debate televisado», *Actas del I Congreso de Fonética Experimental*, Tarragona, pp. 317-323.

TORREGROSA, J. (2006): «Análisis multisistémico de la comunicación humana», *Phonica*, vol. 2, pp. en línea.  
<http://www.ub.es/lfa>