

La transcripción de los elementos prosódicos en corpus de habla coloquial espontánea¹

ANTONIO HIDALGO NAVARRO
Universitat de València
antonio.hidalgo@uv.es

CARLOS CASTELLÓ VERCHER
Universitat de València
carlos.castello@uv.es

Resumen: La lingüística de corpus tiene como objetivo la generación de bases de datos procedentes de la lengua real dispuestos para su análisis. En este artículo sintetizamos los conceptos básicos sobre transcripción y elaboración de corpus textuales y orales, centrándonos en estos últimos, así como en los sistemas de transcripción de los elementos prosódicos en la conversación coloquial. Tras esta revisión, proponemos un sistema de transcripción prosódica a partir del modelo del Grupo Val.Es.Co. con una serie de criterios e instrucciones objetivas, posibles gracias al empleo de los programas informáticos ELAN y Praat. El sistema propuesto se centra en la identificación y descripción de límites y contornos prosódicos tras el análisis de los fenómenos *pausa* o *alargamiento* y *reajuste tonal*, *inflexión final*, *tonema* o *F0 inicial* y *final*, enmarcados en las cualidades del sonido *duración* y *F0*, respectivamente.

Palabras clave: prosodia, corpus orales, transcripción, conversación coloquial.

Transcription of prosodic elements in spontaneous colloquial speech corpora

Abstract: The aim of corpus linguistics is the elaboration and analysis of real language texts. In this article we synthesise the basic concepts

¹ Esta investigación ha sido posible gracias a la ayuda recibida por el Ministerio de Ciencia, Innovación y Universidades para el proyecto ECOS-C/N, Estudio de los condicionantes sociales del español actual en el centro y norte de España: nuevas identidades, nuevos retos, nuevas soluciones, (ref. PID2023-148371NB-C42).

of transcription and elaboration of textual and oral corpora, focusing on the latter, as well as on the systems of transcription of prosodic elements in colloquial conversation. Following this review, we propose a prosodic transcription system based on the Val.Es.Co. system with a series of objective criteria and instructions, made possible using the ELAN and Praat software. This system focuses on the identification and description of prosodic boundaries and contours after the analysis of the phenomena *pause* or *lengthening* and *tonal readjustment*, *final inflection*, *toneme* or *initial* and *final F0*, framed in the sound qualities *duration* and *F0*, respectively.

Keywords: prosody, oral corpus, transcription, colloquial conversation.

1. Lingüística de corpus y transcripción de la lengua oral

La moderna lingüística de corpus se ha interesado entre otras cuestiones por la elaboración de métodos para la recopilación y organización de muestras de habla (orales o escritas) en soporte digital. Su propósito es obtener datos objetivos, empíricos, de modo que cada corpus lingüístico pueda llegar a representar a escala el funcionamiento real del lenguaje natural.

Tales corpus han permitido disponer de bases de datos amplias para proceder al estudio de las lenguas convencionales y de sus características intrínsecas (prosodia, léxico, morfología, sintaxis, aspectos históricos, etc.) y valorar otros factores como, por ejemplo, la variación lingüística; por lo demás, la disposición de corpus digitalizados permite someter los datos a diversas formas de procesamiento cualitativo y/o cuantitativo.

1.1. *Corpus orales y corpus textuales*

Existen dos tipos de corpus lingüísticos: los *textuales* y los *orales*. Los primeros contienen muestras de la lengua escrita, los segundos, muestras de la lengua oral (transcripción ortográfica o grabaciones acompañadas de transcripción). Por lo demás, en líneas generales, la necesidad de obtener modelos estadísticos de la lengua para su aplicación en el desarrollo de sistemas de reconocimiento (aplicables a trabajos como el dictado automático) ha llevado a un uso cada vez más frecuente de los corpus textuales y de las transcripciones del registro oral espontáneo. Además, el creciente interés por los aspectos prosódicos y suprasegmentales de la conversación y del discurso oral ha obligado a disponer del texto transcrito y de la grabación.

Por razones de espacio no tratamos aquí sobre los problemas generales relativos al diseño, confección o aprovechamiento de un corpus, ni tampoco describimos sus tipos o sus criterios de construcción (véanse al respecto Llisterri 1997; Torruella y Llisterri 1999; Alvar Ezquerra y Villena Ponsoda 1994; Ávila Muñoz 1996; Pino y Sánchez 1999). En lo que sigue nos ceñimos más precisamente a tratar sobre los corpus orales (transcripciones ortográficas de lengua hablada²) y, dentro de estos, un subgrupo más específico: el de corpus orales de conversación espontánea³ (Briz y Grupo Val.Es. Co. 1995 y 2002). En este marco conversacional trataremos de manera específica sobre algunos problemas habituales que afectan a la interpretación de los datos transcritos, particularmente en lo que respecta a los fenómenos suprasegmentales.

1.2. *Sistemas de transcripción de corpus orales*

La transcripción de conversaciones se ha convertido en una práctica común en diferentes disciplinas: antropología, lingüística, psicología, derecho... No extraña, pues, que a objetos de estudio distintos correspondan sistemas de transcripción también diferentes.

Al respecto, Payrató (1995) hacía referencia a distintos sistemas de transcripción y codificación: los procedentes de la etnometodología, interesados en reflejar la interacción verbal (Atkinson y Heritage, Eds. 1984; Button y Lee, Eds. 1987), los desarrollados por la etnografía de la comunicación o la sociolingüística interaccional (Ochs 1979; Tannen 1987; DuBois 1991; DuBois et al. 1993; Gumperz y Berenz 1993) o la propuesta de CHILDES (MacWhinney 1991) utilizada para el estudio

² De acuerdo con Torruella y Llisterri (1999:53-59) podemos catalogar un corpus de muy diversas formas, tanto mediante criterios generales:

- según el porcentaje y la distribución de los diferentes tipos de textos,
- según la especificidad de los textos,
- según la cantidad de texto recogida en cada documento,
- según la codificación y la anotación,
- según la documentación que acompaña a los textos,

como mediante criterios específicos, de acuerdo con los objetivos específicos de cada corpus:

- corpus para la descripción fonética de la lengua,
- corpus para el desarrollo de sistemas en el ámbito de las tecnologías del habla,
- transcripciones ortográficas de lengua hablada.

³ Para el Grupo Val.Es.Co. (Valencia, Español Coloquial) la conversación es un tipo de discurso que se caracteriza por los siguientes rasgos: es oral, es decir, se articula a través del canal fónico; dialogal, lo que implica, frente al monólogo, sucesión de intercambios; inmediato, puesto que, a diferencia de un informativo o un mensaje pregrabado, se desarrolla en la coordenada espacio-temporal aquí-ahora-ante ti; retroalimentado y cooperativo, puesto que se obra juntamente con otro y su intervención; dinámico, como demuestra la alternancia de turnos, que además es no predeterminada, a diferencia de otros discursos dialogales tales como el debate, la entrevista, etc. La conversación es coloquial cuando presenta además los rasgos no planificado, lo que implica un escaso control de la producción de habla, que favorece la presencia de reinicios, vacilaciones y vueltas atrás; no transaccional, es decir, orientada a un fin interpersonal, de comunión fática, frente a la conversación transaccional, constituida como medio para obtener un fin específico. Consecuentemente, el tono de dicha conversación es informal. En suma, en la llamada conversación coloquial se reconocen, por un lado, rasgos conversacionales, relativos al tipo de discurso y, por el otro, rasgos coloquiales, propios del registro de uso.

del lenguaje infantil. Otros sistemas extendidos han sido los presentados por Blanche-Benveniste y Colette (1987) para la transcripción de un corpus sobre el francés hablado, o por Stenström (1994), Cestero (1994) y Tusón (1995) orientados al análisis de la conversación, e incluso la propuesta del propio Payrató (1995) los resume así:

- Neutralidad o fidelidad: la transcripción no debe ser interpretativa.
- Globalidad o complejidad: deben incluirse todos los fenómenos que aparecen en el discurso oral.
- Omnifuncionalidad: la transcripción debe permitir diversos usos y aplicaciones.
- Claridad: en cuanto al aprendizaje del sistema y a la legibilidad de la representación.
- Universalidad: compatibilidad entre sistemas informáticos.

En definitiva, un sistema de transcripción debe ser interpretativo de los datos, selectivo en cuanto a los fenómenos que se transcriben, pertinente para el objeto de investigación, coherente con la base teórica adoptada por el investigador, fiel en cuanto a la representación de los datos y flexible para que sea posible su utilización en estudios diversos. Por lo demás, la simbología utilizada debería ser clara, económica, sencilla, exenta de ambigüedad y compatible con sistemas estandarizados internacionalmente. En cualquiera de los casos, un sistema de transcripción será adecuado siempre que se ajuste al objeto de estudio y a la finalidad para la que se emplee, y cumpla los principios de exhaustividad y pertinencia de los signos.

En cuanto a la consideración de métodos de transcripción de uso extendido, cabe destacar el sistema EAGLES (Expert Advisory Group on Language Engineering Standards), donde se revisan diferentes elementos propios de la lengua oral, como puede observarse en la Tabla 1:

Nivel de análisis	Elementos transcritos, marcados o codificados
Nivel silábico	Alargamiento, timbre, acento, reconstrucción de segmentos elididos.
Nivel léxico	Fronteras de palabras, palabras truncadas, formas no estándar, formas onomato-péyicas, formas deletreadas, acrónimos, abreviaturas, cambios entonativos en la palabra, acento léxico, pausas percibidas entre palabras o en el interior de una palabra.
Nivel sintáctico	Fronteras entre enunciados, modalidad, interrupciones en el enunciado con o sin presencia de pausas.

Nivel de análisis	Elementos transcritos, marcados o codificados
Nivel suprasegmental	<p>Unidades entonativas: Fronteras entre unidades entonativas o entre unidades menores, unidades tonales incompletas o truncadas, reajustes (resets) tonales, junturas, índices de cohesión, contornos tonales terminales.</p> <p>Tono: Cambios melódicos en el enunciado o en parte del enunciado, nivel tonal, rango tonal, registro, movimiento tonal en la palabra o en el enunciado.</p> <p>Acento: Acento de palabra, acento de frase, acento tonal, niveles de acento, prominencia, énfasis, acento contrastivo, tensión, propiedades rítmicas,</p> <p>Intensidad: Intensidad absoluta o relativa de partes del enunciado</p> <p>Velocidad de elocución: Cambios en la velocidad de elocución, velocidad de elocución relativa o absoluta.</p> <p>Pausas: Pausas silenciosas, pausas vocalizadas, duración absoluta o relativa de las pausas.</p>
Nivel paralingüístico	Vocalizaciones semi-léxicas, vocalizaciones no léxicas, timbre de la voz, otros elementos vocalizados (canto, gritos, etc.).
Nivel discursivo	Turnos de palabra, tipo de transición entre turnos, superposición de turnos.
Nivel contextual	Fenómenos no comunicativos no léxicos y no vocales, información kinésica.

Tabla 1. Elementos transcritos, codificados o marcados en el estudio de la lengua oral (EAGLES 1996).

Por lo que atañe al ámbito hispánico, al margen del corpus de conversaciones coloquiales de Val.Es.Co., cuya versión actualizada puede consultarse en https://www.valesco.es/#/pages/cod_hj3y7hwvuuajtlk-q0ik/cod_fa393ih5l4jx9zssv7, cabe destacar el Corpus de Variedades Vernáculas Malagueñas (Alvar y Villena 1994), constituido con un objetivo predominantemente sociolingüístico, el Corpus de Referencia del Español Actual (CREA) o el Corpus del Español del Siglo XXI (CORPES XXI).

El Corpus de Variedades Vernáculas Malagueñas se basa en la ortografía convencional, pero incorpora convenciones y rasgos específicos que facilitan la reconstrucción por el lector y la inclusión de características fónicas, discursivas y estilísticas (Ávila 1996:103). Su codificación sigue los estándares de la TEI (*Text Encoding and Interchange*), utilizando para ello SGML (*Standard Generalized Markup Language*).

Las etiquetas utilizadas en este corpus codifican información sobre hablantes y turnos de palabra (simultaneidad o interrupción), rasgos prosódicos (tono, intensidad, entonación, tempo, diversos tipos de pausa en función de su duración relativa y énfasis), acciones no verbales y fenómenos no vocales, actuación lingüística e incidencias que tienen lugar en la grabación, e incluso se han introducido etiquetas

para indicar características fonéticas propias de las variedades de Málaga (Ávila 1996:106).

En cuanto al Corpus de Referencia del Español Actual (CREA)⁴, desarrollado por el Instituto de Lexicografía de Real Academia Española, introduce una serie de convenciones para el tratamiento de los problemas relacionados con formas reducidas de palabras, abreviaturas y acrónimos, palabras deletreadas, secuencias numéricas, interjecciones, fenómenos comunicativos no vocales, fenómenos no comunicativos no vocales, errores de producción, repeticiones, rectificaciones e interrupciones en el discurso, titubeos, etc. Se trata fundamentalmente de un corpus textual que incluye, en todo caso, transcripciones de lengua oral. Emplea los estándares de la TEI y, una vez transcrito y codificado, el texto se almacena en formato SGML. Se utilizan asimismo signos ortográficos habituales de acuerdo con la normativa de puntuación en español, excepto en el caso del punto y coma, que no se emplea en la transcripción, y de las comillas, cursivas y mayúsculas que se usan como medio tipográfico de realce.

En esta línea, el Corpus del Español del Siglo XXI (CORPES XXI)⁵ divide las transcripciones de lengua oral según se ofrezcan o no alineadas con el audio. En el caso de las alineadas, el corpus codifica la transcripción con XML y tiene presente elementos de la lengua oral tales como la pausa y su duración, las palabras cortadas, y elementos paralingüísticos como la risa, el habla solapada y la vacilación.

Por su parte, el corpus de conversaciones coloquiales del Grupo Val. Es.Co. (Briz y Val. Es.Co. 1995 y 2002) se ajusta a los principios que debe cumplir todo sistema de transcripción ya señalados previamente, particularmente a los de exhaustividad (cada signo representa un único fenómeno) y de pertinencia (cada fenómeno aparece codificado mediante una única convención). Este corpus intenta reproducir fielmente la conversación, por lo que incorpora aspectos como la alternancia de turnos, la sucesión inmediata de emisiones, los solapamientos, reinicios y autointerrupciones, las escisiones conversacionales, las pausas y silencios, la entonación (inflexiones finales que influyen en el curso de la conversación con cambios respecto a la prosodia convencional), los fenómenos de énfasis, problemas relacionados con emisiones dudosas o indescifrables, aspectos de fonosintaxis, alargamientos fonéticos, etc. Se combina el método ortográfico con el propuesto por el Análisis de la Conversación, y resulta suficientemente estrecho para conseguir que el lector pueda reproducir con bastante aproximación la conversación

⁴ Para más detalles sobre los criterios de diseño y confección de este corpus véase Pino y Sánchez (1999).

⁵ Para más información sobre su codificación véase el enlace <https://www.rae.es/corpes/assets/rae/files/corpes/codOral.pdf>.

original, y suficientemente ancho para permitir una lectura fluida de la misma⁶.

2. La transcripción de datos prosódicos

La incorporación de información sobre elementos suprasegmentales en un corpus oral plantea diversos problemas, que se concretan en las variaciones continuas de F0, de intensidad o de duración. Por tanto, es necesario un proceso de abstracción para determinar cuáles de estas variaciones son lingüísticamente significativas y cómo se relacionan con categorías discretas. Finalmente, tales categorías deberían ser representadas en un sistema de notación.

En conjunto, pues, un sistema ideal de transcripción prosódica debería permitirnos la representación en varios niveles, ser compatible con el intercambio electrónico de datos y cubrir las necesidades del mayor número de lenguas posible. Ahora bien, en ausencia de un sistema que reúna tales características, parece adecuado elaborar mecanismos de compatibilidad entre los ya existentes a fin de facilitar la reutilización de los datos.

Al respecto, en el marco del proyecto ESPRIT SAM se han desarrollado sistemas de transcripción prosódica compatibles con las necesidades de anotación de bases de datos en soporte electrónico (Llisterri 1997), resultando de ello propuestas como PROSPA, SAMSINT o SAMPROSA (Gibbon 1989; Wells *et al.* 1992). De todos ellos, un conjunto de símbolos que recibió bastante aceptación para la transcripción prosódica fue el de SAMPROSA (*SAM Prosodic Alphabet*), propuesto por Gibbon (1989) y desarrollado por Wells *et al.* (1992) hasta llegar al formato que se presenta en la Tabla 2, tomada de Wells (1995):

SAM-PROSA	ASCII	Definition	SAM-PROSA	ASCII	Definition
Local tone					
H	72	High pitch	M	77	Mid pitch
L	76	Low pitch	+	43	High pitch
T	84	Top pitch (extreme H)	++	43,44	Much higher pitch
B	66	Bottom pitch (extreme H)	+-	43,45	Peak (upward- downward)

⁶ Las conversaciones transcritas por Val.Es.Co. constituyen de hecho una forma de transcripción semiestrecha. En función del objetivo perseguido, la transcripción puede estrecharse (integrando, p. e., aspectos prosódicos ausentes antes) o ensancharse (prescindiendo, p. e., de reinicios y solapamientos, si no son pertinentes para el estudio). Por ejemplo, Hidalgo (1997), aplicando el sistema de base de Val.es.Co. (véase en Anexo las convenciones de transcripción), desarrolla una versión estrecha de este sistema, que incluye, además de los rasgos generales otros como la frecuencia fundamental (F0) de cada uno de los grupos entonativos segmentados.

SAM-PROSA	ASCII	Definition	SAM-PROSA	ASCII	Definition
-	45	Lower pitch	^^	94,94	Wide upstep
--	45,45	Much lower pitch	!	33	Downstep
-+	45,43	Trough (downward-upward)	!!	33,33	Wide downstep
^	94	Upstep	= or > or S	61,62,83	Level or same tone
Global tone: from local and nuclear tone repertoire					
Terminal tone: from local and nuclear tone repertoire					
Nuclear tone			Pause		
-	45	Level tone (before tone group boundary)	...	46,46,46	Silence
' or / or R	39,47,82	Rising tone	Boundary		
` or \ or F	96,92,70	Falling tone	\$	36	Syllable boundary
'' (etc.)	96,39 (etc.)	Fall-rise	#	35	Word boundary
'' (etc.)	39,96 (etc.)	Rise-fall		124	Tone group boundary / non-directional)
Length			[91	Tone group boundary (left)
:	58	Segmental length mark]	93	Tone group boundary (right)
Stress			Metasymbols		
"	34	Primary stress	-	45	Separator (the underscore, _✓ ASCII 95, may replace this owing to ambiguity with level tone)
%	37	Secondary stress	*	42	Conjunctive

Tabla 2. Transcripción de datos prosódicos en SAMPROSA (Wells, 1995).

En todo caso, el sistema de notación prosódica más seguido en la actualidad es, sin duda, ToBI (*Tone and Break Index*). Este método incluye dos tipos de símbolos, los que representan la estructura tonal subyacente (*tones*) y los que marcan los límites entre unidades prosódicas (*break indices*)⁷. ToBI recoge el inventario de unidades entonativas propuestas

⁷ Sin embargo, se han dirigido ciertas críticas hacia este sistema ToBI entre las cuales figura, por una parte, su estrecha dependencia del modelo fonológico desarrollado por Pierrehumbert (1980)

por Pierrehumbert (1980) e introduce algunas especificaciones respecto del comportamiento tonal que se traducen en diferentes componentes (Hualde 2003: 157-173):

1. **Tono:** H si es alto (*high*), L si es bajo (*low*) y M si es medio (*mid*).
2. **Acento tonal:** tono o secuencia de tonos fonológicamente asociado con una sílaba acentuada, que viene señalado con un asterisco (*).
3. **Tono de frontera:** tono asociado con el límite de una frase y que va marcado con el símbolo %.
4. **Escalonamiento:** los picos pueden estar distribuidos de manera descendente (*downstep*) de manera predecible o por algún contraste pragmático (!) o de manera ascendente (*upstep*), si marcan un pico significativamente superior al previo (j).
5. **Frasas prosódicas:** pueden ser entonativas o intermedias. La frase entonativa está constituida por frases intermedias con una separación menor entre sí y constituidas por elementos tonales llamados acentos de frase (L-/H-). Estos acentos combinados con los tonos de frontera dan lugar a cuatro posibilidades: descenso (L-L%), descenso con ascenso final (L-H%), suspensión (H-L%), ascenso (H-H%) (Beckman *et al.* 2002).

El sistema ToBI ha tratado de superar la disparidad de soluciones a la hora de transcribir la entonación mediante un inventario cerrado de acentos tonales. Para el español (Sp_ToBI) trabajos como los de Sosa (2003) o Estebas-Vilaplana y Prieto (2008) tratan de precisar la propuesta de Beckman *et al.* (2002), e incluso se ha llegado a proponer una herramienta de transcripción automática en Praat (Elvira-García *et al.* 2015). Asimismo, una de las principales virtudes de ToBI es permitir la representación por separado de las unidades prosódicas en que se organizan los enunciados (*jerarquía entonativa*) y los fenómenos entonativos (en términos de tonos).

No obstante lo anterior, uno de los problemas que plantea el sistema ToBI es que se trata de un procedimiento de anotación fonológica y no fonética que, además, resulta compleja en su proceso de aprendizaje y no recoge de modo preciso las diferencias de rango tonal (fenómenos

y, por otra, su mejor adaptación al inglés que a otras lenguas, lo que explicaría su amplia utilización en Estados Unidos. No obstante, se han realizado trabajos en italiano, alemán, húngaro y español, entre otras lenguas, usando ToBI como sistema de representación. Un inconveniente más importante es que la anotación mediante ToBI requiere un cierto conocimiento previo de los patrones entonativos de la lengua. Aun así, es un sistema que ofrece indudables ventajas, tales como la presentación de una estructura jerárquica, que permite seleccionar de entre subconjuntos o conjuntos mayores de símbolos, la posibilidad de representar problemas que aparecen en la transcripción o el hecho de que existan experimentos que muestran un elevado grado de acuerdo entre transcripores diferentes (Pitrelli *et al.*, 1994).

de altura, fenómenos de reajuste tonal, etc.) ni otros aspectos como los alargamientos, etc.

Otro sistema de transcripción prosódica vinculado a la lengua oral espontánea corresponde a Chafe (1993), con un modelo de unidades basado en su segmentación prosódica, donde se utiliza una serie de símbolos para representar fenómenos como los alargamientos (=), los acentos primario (^) y secundario (^) y los límites entre unidades de acento (|), además de la coma y el punto que diferencian tipos de contornos melódicos de continuación o cierre.

Por su parte, Mertens (2004) propone un modelo de transcripción prosódica denominado prosograma. Este sistema de transcripción intenta cumplir varios criterios:

1. Es objetivo, robusto y fácil de interpretar, además de la representar la entonación perceptiva.
2. Fija la evolución de largos fragmentos de habla.
3. Esta fijación es cuantificada para ser estimada por intervalos melódicos.
4. Se preserva la organización temporal con el fin de observar pausas e interrupciones.
5. La transcripción es semi-automática.
6. Es neutral, es decir, independiente de cualquier modelo con el que pueda ser usada por investigadores de distintos marcos teóricos.
7. Hay una alineación entre sonido y texto.
8. Permite la manipulación y la evaluación de la transcripción.

El sistema de Mertens permite, además, cuatro variantes según se quiera (1) representar la curva entonativa perceptiva, (2) añadir la curva de intensidad y de F0, (3) añadir la cuantificación temporal y de F0 o (4) incluir toda la información prosódica (curva entonativa perceptiva, curva de intensidad y de F0 y cuantificación temporal y de F0).

Finalmente, Cantero (2002), Cantero y Font (2007 y 2009), Cantero y Mateo (2011) y Mateo (2010) han desarrollado el modelo de Análisis Melódico de Habla, que conceptualiza la entonación como complejo prelingüístico, lingüístico y paralingüístico. El modelo de transcripción prosódico de Cantero y Font (2007) distingue rasgos melódicos y fonológicos. Los rasgos melódicos se analizan a partir de la estilización de la curva entonativa entendida como un conjunto de elementos estructurales. Como modelo de representación, el protocolo del AMH plantea dos pasos. El primer paso consiste en la extracción del valor de

F0 de cada vocal. El segundo paso consiste en estandarizar los datos de F0, de manera que la primera vocal de un segmento dado toma un valor igual a 100 y se recalculan mediante reglas de tres las subidas y bajadas tonales de manera porcentual; los valores positivos implican una subida tonal y los negativos, bajadas (Cantero y Font 2009). Este modelo ha sido complementado recientemente con el denominado Análisis Prosódico del Habla (APH) (Cantero 2019), que integra simultáneamente el tono, la intensidad y la duración.

3. Propuesta de implementación de datos prosódicos a partir del sistema de transcripción de Val.Es.Co.

Al margen de lo dicho hasta aquí, la transcripción de un corpus de lengua oral plantea muy diversos problemas, tales como la existencia de variaciones fonéticas no recogidas en diccionarios normativos, el uso de signos de puntuación o la representación de siglas, las palabras deletreadas, las secuencias numéricas, etc. A ello se añaden otros problemas derivados de la necesidad de delimitar enunciados y unidades tonales, la variación entre elementos suprasegmentales, el establecimiento oportuno de pausas (silencios), las pausas oralizadas o llenas, la representación de elementos paralingüísticos o cuasi-léxicos (risas, toses, etc.), los cambios de turno, las intervenciones simultáneas de dos o más hablantes, las palabras o construcciones truncadas, las repeticiones o errores de producción, etc. Sin duda, la codificación de todos estos elementos obliga a enriquecer la transcripción ortográfica, lo que resulta imprescindible para proceder adecuadamente al uso y análisis de lo transcrito, esto es, a su interpretación.

Quedan, pues, sin resolver algunos problemas vinculados a la falta de objetividad en cuanto a la información y al análisis, y a la dificultad de la representación y de lectura de datos prosódicos. Presentamos por ello a continuación una propuesta de transcripción prosódica que atiende a los principios siguientes:

- a. Criterios claros y objetivos basados en datos acústicos.
- b. Objetivación de los resultados mediante el empleo de herramientas informáticas adecuadas.
- c. Facilidad en la lectura.
- d. Posibilidad de elaboración del corpus de manera individual y autónoma.

Con el fin de satisfacer estos principios, nuestra propuesta de transcripción prosódica (3.4.) va precedida de tres apartados que afectan

a la segmentación en grupos entonativos (3.1.), a la identificación y representación de la curva entonativa (3.2.) y a las herramientas de transcripción y análisis (3.3.).

3.1. Criterios para la segmentación en grupos entonativos

El concepto de grupo entonativo y su identificación ha sido ampliamente por Quilis, Cantarero y Esgueva (1993), Cabedo (2011a y 2011b) e Hidalgo (2018 y 2019).

En primer lugar, Quilis, Cantarero y Esgueva (1993) proponen que los fenómenos pausa y/o inflexión de F0 son criterios claros para la identificación de límites prosódicos. En segundo lugar, Cabedo (2011a) delimita la rentabilidad de los fenómenos pausa, reajuste tonal, inflexión y alargamiento en su fórmula estadística (MESTEL) para la identificación de límite prosódico, tal como se indica en la Figura 1:

Fórmula	Resultado
$1 + 2.71^{(4.97 - (0.01 \times 60) - (1.88 \times 1.4) - (0.004 \times 70) - (17.7 \times 0.2))}$	= 0.89 (89%)

Figura 1. Ejemplo de aplicación de la fórmula MESTEL⁸

En tercer lugar, Hidalgo (2019: 98) propone varios criterios operativos para la identificación de límites prosódicos⁹:

- Pausa mayor o igual de 300 ms¹⁰.
- Reajuste tonal positivo o negativo mayor o igual de 3 st.

⁸ MESTEL contiene una serie de constantes asociadas a parámetros que miden la importancia de cada variable para el reconocimiento de fronteras demarcativas:

- constante matemática e (2.71);
- constante establecida por la prueba estadística (4.97);
- constante de duración (1.88);
- constante de reajuste (0.004);
- constante de inflexión (0.01);
- constante de pausa (17.7).

Todos estos valores se relacionan en última instancia con los de cada punto concreto del discurso sobre el que queramos determinar su carácter de frontera/no frontera. Con ello se obtiene un valor probabilístico de importancia para cada factor: puede calcularse así la probabilidad de que una frontera melódica se constituya como frontera efectiva de grupo entonativo. Por ejemplo, para un determinado segmento de habla, si existe una pausa posterior de 0,2 segundos, una duración del segmento entonativo de 1.4 segundos, una inflexión melódica del 60 % y un reajuste tonal del 70 %, la probabilidad de que dicho segmento sea grupo entonativo sería del 89 %. Este es el caso del ejemplo formulado en la Figura 1.

⁹ El Corpus Val.Es.Co. 3.0 emplea de hecho los criterios de la pausa (≥ 300 ms) y el reajuste (3 st) para la identificación de límites prosódicos (Pons 2022).

¹⁰ Se mantienen así los criterios aplicados en el marco del sistema de transcripción del Corpus Val.Es.Co. 3.0: pausa (≥ 300 ms) y reajuste (3 st) para la identificación de límites prosódicos (Pons 2022).

- Inflexión tonal mayor o igual de 3 st.
- Duración de la sílaba final del GE (grupo entonativo) cuando sea al menos el doble respecto de la media del GE.

Así pues, los fenómenos que indican la presencia de un límite y que delimitan un GE se pueden agrupar en dos bloques según su naturaleza: duración (duración de la pausa y alargamiento) y frecuencia fundamental (reajuste tonal e inflexión).

Por lo que respecta a la representación del límite prosódico, el sistema de transcripción Val.Es.Co. (Briz y Val.Es.Co. 2002) ofrece una serie de convenciones que permiten señalar la presencia de estos fenómenos demarcativos, a excepción del fenómeno de reajuste tonal, que no cuenta con una forma de transcripción, pero sí ha sido incluido en nuestra propuesta como se verá en 3.4.:

a. En primer lugar, la pausa se puede señalar con / (300-500 ms), // (500-1000 ms) y /// (>1000 ms). En este último caso, se señala entre paréntesis la duración de la pausa en segundos¹¹.

b. En segundo lugar, el alargamiento de la vocal se señala con la duplicación de la vocal alargada (aa, ee, ii, oo, uu).

c. En tercer lugar, la inflexión tonal se marca con flechas ↑ (entonación ascendente), ↓ (entonación descendente), → (entonación mantenida o suspendida), y el acento circunflejo ^ (entonación circunfleja).

3.2. *Criterios para la identificación y la representación de la curva entonativa*

Como hemos visto previamente en § 2, para la representación de la curva entonativa es especialmente productivo el sistema de transcripción Tone and Break Index (ToBI), que permite representar, con un número reducido de símbolos, la diversidad entonativa. No obstante, su lectura resulta poco transparente para usuarios no expertos. Por su parte, el modelo de Mertens o el Análisis Melódico del Habla permiten visualizar de manera más clara la curva entonativa. Sin embargo, tanto estos modelos como el sistema ToBI, necesitan de una segunda línea de transcripción para su clara visualización, lo que va en detrimento de la claridad en la transcripción y de la facilidad de lectura.

Por su parte, el sistema de Val.Es.Co. permite transcribir las inflexiones finales mediante las flechas ↑, ↓ y → y el acento circunflejo ^, pero no ofrece información sobre el inicio o el cuerpo del GE. Por ello,

¹¹ Las razones que justifican esta decisión metodológica se derivan de la experiencia de transcripción de años llevada a cabo en el marco del grupo Val.Es.Co, tal como se refleja en la introducción de Briz y Val.Es.Co. (2002).

Hidalgo (1997, 2002) propone transcribir la F0 inicial y final de cada GE en hercios. Esta información permite indicar la dirección tonal del GE y ofrece información acerca del reajuste tonal entre GE sucesivos.

La cuantificación de la F0 en hercios, sin embargo, supone un problema a nivel técnico, puesto que los hombres presentan una media de 120 hercios frente a los 200 que suelen presentar las mujeres (González et al. 2002), por lo que habría que identificar el sexo de los hablantes en la transcripción de la conversación. Así pues, el empleo de cifras numéricas en hercios dificulta la labor de transcripción frente a otras propuestas para la identificación del límite prosódico que emplean semitonos, por dos motivos: por una parte, la extracción de valores en hercios requiere de una estandarización posterior para evitar las diferencias de sexo que ya viene dada con la unidad semitono y, por otra parte, la extracción de valores en semitonos facilita la identificación de límites prosódicos y el empleo de la misma unidad resulta más operativa para la obtención de todos los datos prosódicos, evitando así la variedad de unidades de medida para un mismo fenómeno.

3.3. Herramientas de transcripción y análisis

Los criterios previos necesitan de herramientas informáticas que objetiven la transcripción. En este sentido, el programa ELAN¹² resulta operativo para la transcripción prosódica, debido a que permite, por una parte, sincronizar el audio y el texto y, por otra, interactuar con el programa Praat (Boersma y Weenink 2022), que facilita la información prosódica relativa a la duración, la F0 y la intensidad. ELAN es una herramienta de transcripción offline que emplea el lenguaje XML y que permite añadir capas o niveles de anotación con los que añadir texto o comentarios de una manera jerarquizada, alineada con el audio y, además, fácilmente exportable a otros formatos. Por su parte, Praat puede automatizar la extracción de datos acústicos en análisis más precisos mediante el uso de archivos Textgrid y el empleo de scripts como `Analyze_tier_modify`¹³.

Con todo ello, la propuesta de transcripción prosódica que presentamos aquí obedece al siguiente protocolo:

1. Grabación de audio según el objetivo de la investigación.
2. Importación del archivo de audio en formato .wav al programa ELAN.

¹² El siguiente enlace ofrece información detallada sobre el funcionamiento de este programa: <https://archive.mpi.nl/tla/elan>

¹³ Disponible en: http://uk.groups.yahoo.com/group/praat-users/files/Daniel_Hirst/analyze_tier.praat

3. Creación de líneas de transcripción: una por hablante y otras que se adapten al objetivo de la transcripción.

4. Transcripción del texto por cada grupo fónico o espiratorio.

5. Observación y cuantificación de la duración de los silencios dentro de cada grupo fónico. De esta manera, los silencios iguales o superiores a 300 ms constituyen una pausa, hecho que señala un límite prosódico, por lo que la caja de transcripción deberá ser dividida, como se observa en la Figura 2:

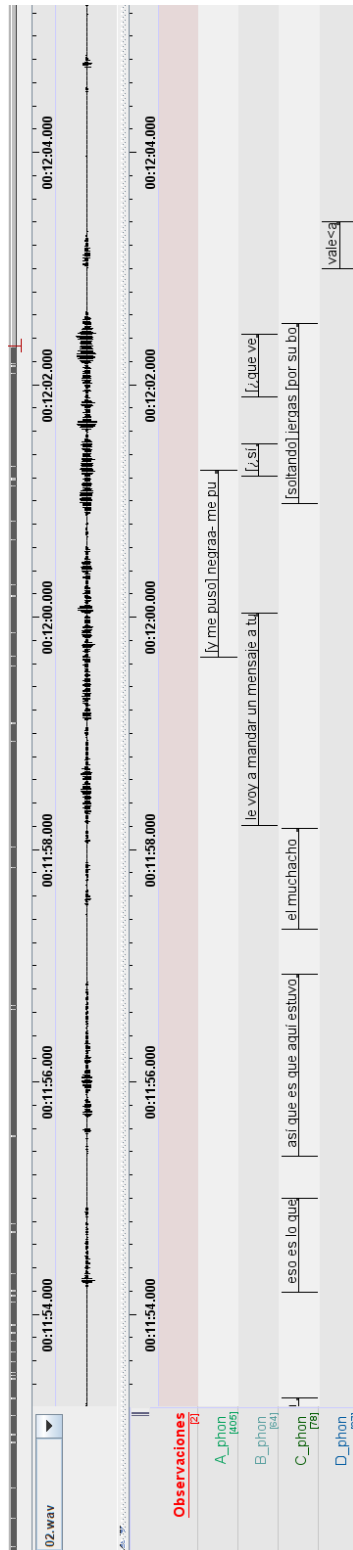


Figura 2. Fragmento de transcripción en ELAN.

Llegados a este punto, dispondremos de una transcripción en GE que solo tienen presente la pausa como límite prosódico, por lo que la identificación de los demás fenómenos deberá llevarse a cabo de manera o bien *manual*, o bien *automatizada*.

En primer lugar, para la identificación manual proponemos los siguientes pasos:

Abrir cada caja de ELAN en Praat, observar los posibles saltos tonales existentes entre palabras y dividir las cajas en los casos en los que se den saltos iguales o superiores a 3 st, hecho que señala un reajuste tonal. Este salto se mide tomando el valor en semitonos del final de una palabra y restando el valor inicial en semitonos de la palabra siguiente. Véase al respecto la Figura 3:

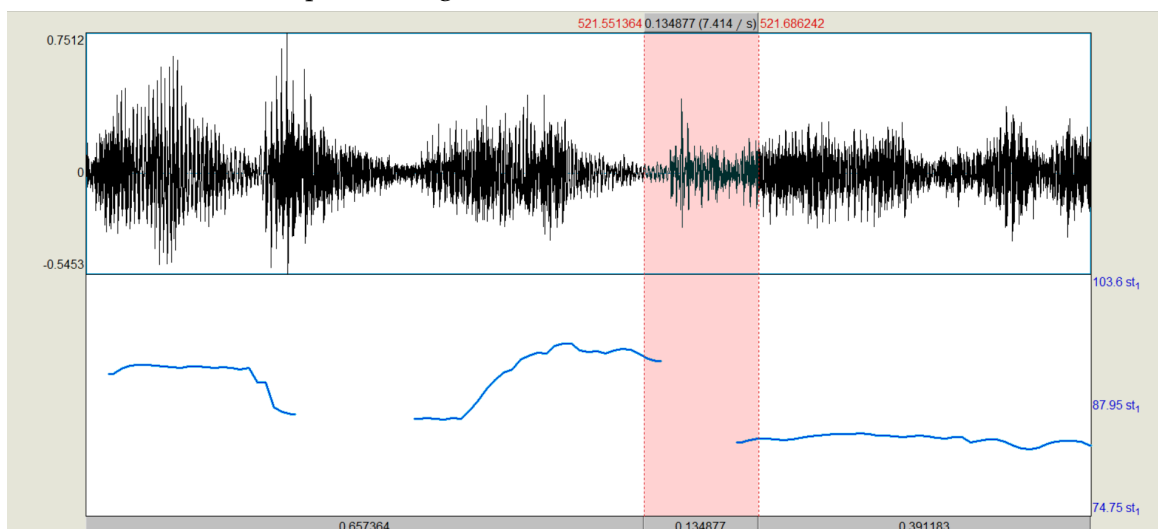


Figura 3. Representación sombreada del reajuste tonal en Praat.

Observar la presencia de inflexión tonal final de GE, esto es, cuando la diferencia entre la sílaba tónica y el final de la palabra del GE es igual o superior a 3 semitonos. En este caso, podemos señalar la presencia de un tonema ascendente o descendente según la dirección de la curva. Véase al respecto la Figura 4:

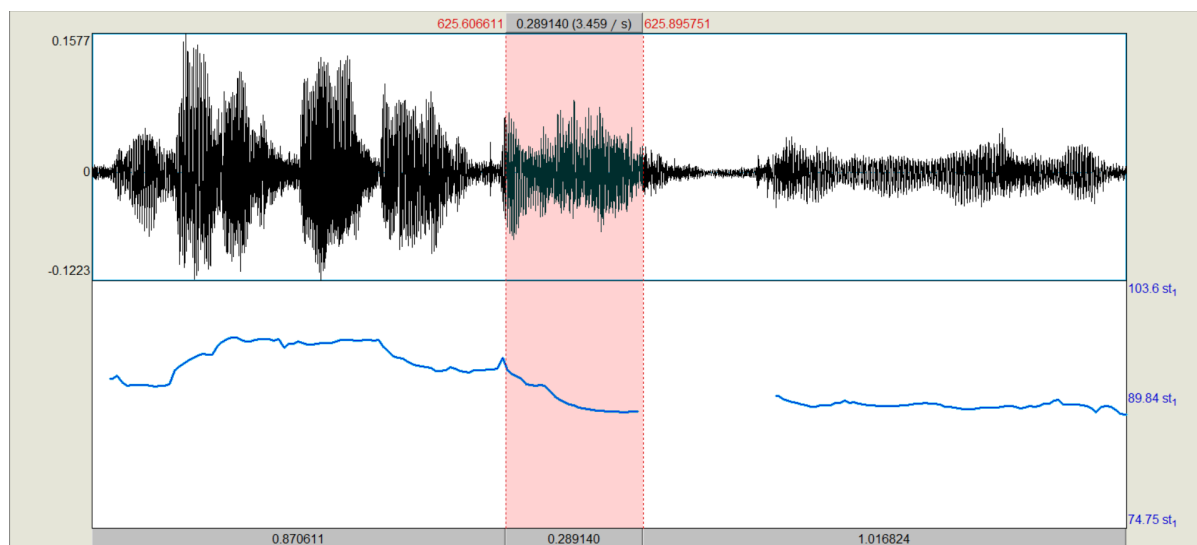


Figura 4. Representación sombreada de la inflexión tonal en Praat.

La identificación del alargamiento puede hacerse de modo automatizado mediante el uso de Textgrid y el script mencionado anteriormente (`Analyze_tier_modify`), que permiten obtener la información de manera más objetiva y completa. Así pues, el uso de Textgrid exige la selección de «películas»¹⁴ en Praat con las que se crea un archivo de audio de manera automática que permite la edición y la anotación. Para crear el Textgrid en Praat debemos clicar en la opción ANNOTATE > To TEXTGRID y crear las líneas PALABRA y SÍLABA.

Después de crear los TextGrid de todas las cajas de ELAN, hay que guardar los archivos de audio (*sound*) en una carpeta y los de texto (*TextGrid*) en otra. Posteriormente, se instala el script `Analyze_tier_modify` y se abren las carpetas con todos los fragmentos anotados. El programa nos ofrecerá un archivo .txt que puede ser importado a una hoja Excel con los valores totales de duración, medias máximas y mínimas de F0 e intensidad de cada fragmento, tal como puede observarse en la Figura 5:

¹⁴ ELAN permite la extracción de fragmentos de audio ya editables gracias a la opción SELECCIÓN DE «PELÍCULAS» CON PRAAT.

	A	B	C	D	E	F	G	H	I	J
1		file	label	duration	f0_min	f0_mean	f0_max	int_min	int_mean	int_max
2	# file : 02_1									
3	# min_pitch: 135; median_pitch: 203 max_pitch 336;									
4	1 02_1	di		115.741	88.579	90.129	92.291	53.102	60.016	62.750
5	2 02_1	solapado		1.427.201	83.976	92.626	97.242	53.978	70.023	75.201
6	3 02_1	al		174.516	89.643	90.767	92.929	46.640	58.499	62.755
7	4 02_1	sue		162.229	92.394	92.952	94.094	46.772	57.698	63.262
8	5 02_1	gro		186.563	93.037	93.603	94.753	45.095	60.686	64.402
9	6 02_1	por		75.031	92.815	93.183	93.974	44.547	54.972	58.948
10	7 02_1	que		139.922	91.194	91.822	93.799	48.077	56.154	59.080
11	8 02_1	van		141.950	91.064	91.717	92.071	49.834	58.993	61.944
12	9 02_1	a		40.557	91.299	92.683	94.294	51.930	54.512	55.801
13	10 02_1	rre		119.644	91.582	91.839	93.049	56.972	59.840	61.496
14	11 02_1	glar		451.657	91.481	92.056	92.869	49.380	59.862	63.340
15										
16	# min_pitch: 134; median_pitch: 190 max_pitch 300;									
17	12 02_2	el		124.219	88.572	90.616	96.558	51.055	54.128	56.503
18	13 02_2	sue		117.317	91.480	92.075	94.400	48.458	61.986	63.839
19	14 02_2	gro		97.012	91.706	92.197	92.652	60.827	64.315	66.346
20	15 02_2	lo		74.451	91.957	92.116	92.197	56.906	59.513	61.688
21	16 02_2	tie		49.634	91.758	92.860	94.530	50.620	57.189	59.437
22	17 02_2	ne		133.109	90.237	91.474	92.764	51.424	54.256	57.184
23	18 02_2	ca		49.634	90.598	92.124	93.702	53.660	56.123	58.710
24	19 02_2	si		115.061	90.278	92.116	93.673	48.243	54.311	57.973
25	20 02_2	a		56.402	89.661	89.954	90.278	50.698	57.104	59.873
26	21 02_2	rre		157.926	89.504	90.639	95.395	47.672	57.013	60.331

Figura 5. Datos importados en Excel.

Sobre los datos obtenidos se pueden realizar operaciones matemáticas para la obtención de valores necesarios para la identificación de fenómenos prosódicos. A continuación, explicamos brevemente dichas operaciones y los valores que permiten obtener:

a. Recolocación: el script utilizado ofrece valores máximos y mínimos de F0, sin embargo, son los valores inicial y final los que aportan información relevante para la identificación de la inflexión y el reajuste tonal. Por ello, la dirección de la curva se cruza con los valores máximos y mínimos, de manera que en una curva ascendente el valor mínimo es inicial y el máximo, final y en una curva descendente el valor máximo es inicial y el mínimo, final.

b. Resta: una vez identificados los valores iniciales y finales de cada segmento se restan los segundos a los primeros y se obtiene el valor del salto tonal con lo que se identificarán las posibles inflexiones o reajustes tonales.

c. Porcentaje: para obtener el alargamiento de una sílaba cabe obtener la media de duración de las sílabas del GE y calcular mediante una regla de tres si alguna sílaba final de palabra tiene un valor igual o superior a 200 % de la media (véase al respecto lo indicado en §3.1. en este sentido).

Obtendremos de esta manera los datos relativos a la identificación de límites prosódicos (pausa, alargamiento, reajuste tonal e inflexión) y del contorno entonativo (tonema y valores inicial y final).

En cualquier caso, a pesar de las ventajas de objetividad que supone el análisis acústico, sea manual o automatizado, a la hora de transcribir el discurso oral, existen ciertas limitaciones para la segmentación y anotación prosódica:

1. En los casos de pérdida de línea de frecuencia fundamental, solapamiento, habla dudosa, elementos paralingüísticos (risas, toses, bostezos) o en el habla entre risas, solo es posible identificar de manera manual la pausa y no se puede identificar ningún otro fenómeno.

2. Hay ocasiones en los que la curva final del GE no toma una dirección clara por lo que es difícil determinar el tipo de tonema; en este caso, la opción Stylize Pitch de Praat puede ayudar a determinar la dirección tonal precisa.

3. El tonema circunflejo no puede identificarse de manera automatizada, por lo que se requiere una revisión de los finales de GE para identificar la posible presencia de estos tonemas.

4. Los monosílabos tónicos y las palabras agudas suelen presentar una curva tonal propia pero corta, por lo que cabría observar la dirección general del GE y sobre todo de la sílaba previa. En el caso de los monosílabos átonos, estos forman parte de la coda del GE.

3.4. *Propuesta de transcripción*

Vistas las consideraciones previas, nuestra propuesta de transcripción incluye información sobre límites y sobre el contorno prosódicos. Los fenómenos tenidos en cuenta son la pausa, el alargamiento, la inflexión, el tonema, la F0 inicial y final de grupo entonativo y el reajuste tonal. Algunos fenómenos sirven exclusivamente para la identificación de límites (pausa) o para la descripción del contorno (F0 inicial y final), pero otros sirven para ambos, como es el caso de la inflexión y el tonema, que sirven para la identificación del final de un grupo entonativo (límite) y para la representación del curva melódica (contorno), o el alargamiento final de grupo entonativo que da información sobre el final de esta unidad entonativa y sobre su ritmo interno. Finalmente, como se ha señalado, el sistema Val.Es.Co. actualmente no ofrece ningún símbolo para la representación del reajuste tonal, por lo que los valores iniciales y finales de F0 (en st) transcritos en el superíndice y subíndice, respectivamente, señalarán el salto tonal existente entre grupos entonativos. La Tabla 3 sintetiza la información prosódica que refleja nuestra propuesta y su modo de representación:

fenómenos prosódicos	Forma de representación
Presencia de silencio no pausa	(< 300 ms)
Pausa corta, inferior al medio segundo	/ (300-500 ms)
Pausa entre medio segundo y un segundo	// (500-1000 ms)
Pausa de un segundo o más	/// (> 1000 ms)
Alargamiento	aa, ee, ii, oo, uu
Inflexión ascendente / Tonema ascendente	↑
Inflexión descendente / Tonema descendente	↓
Tonema suspendido	→
Inflexión circunfleja / Tonema ascendente	^
F0 inicial de grupo entonativo	^{90.12} digo
F0 final de grupo entonativo	arreglaar _{92.05}

Tabla 3. Sistema de transcripción de datos prosódicos

Véase en el siguiente fragmento de una conversación coloquial cuál es el resultado final de la transcripción según nuestra propuesta:

A: ^{90.58}lo que no me hace gracia_{79.35}↓(045) ^{92.22}es que con mi
hija- mi hija me pone muy_{87.81}↓ ^{75.19}negra_{87.60}↑ ^{75.51}¿eh? _{80.50}↑ (012)
^{92.50}te lo juro_{88.36}↑//(679) ^{90.13}yo de todas se lo dije digo vamos a
ver_{89.13}→/(496) ^{93.42}¿qué me dijiste? digo yo_{88.69}→ ^{93.09}que no caí en
ese momento bien que tiene que dar las nóminas ella_{95.99}→//(595)
^{90.12}digo [¿por qué tienes que dar las nóminas?]

B: [¿ella tiene que dar las nóminas?]

A: §^{90.76}al suegro
porque van a arreglaar_{92.05})^o→^{90.61}el suegro lo tiene casi
arregla(d)o paraa_{87.06})^o→

B: ^{85.52}la hipoteca_{95.96} (213) [para que la
hipo]teca sea para ella solo_{79.80}↓

A: [digo](1.098)
^{95.75}no_{89.31}↓(179) ^{91.89}para ampliar se ve_{75.56}↓(298) ^{86.37}algo_{93.67}↑//
(629) ^{89.80}digo [mira Jimena]

C: [yo creo que para] ampliarlo_{81.65}↑

A: ^{94.01}digo vamos
a ver ^{90.95}(no es que tenemos que_{89.50})^o↓ ^{90.95}digo_{88.17}↓(032) ^{93.71}vamos
a ver_{91.55}↓//(922) ^{93.11}digo haz el favor_{89.15}↓(057) ^{94.88}de explicar un
poco tu padre sabe muy bien de bancos_{87.89}↓ (042) ^{89.04}digo a ver
si vaa a encargarse el Manuel de TODO_{97.06}↑//(1.221) ^{91.50}que yo
quiero confiaar^{96.97}↑/(313) ^{89.64}pero a ver si va arreglar_{95.74}↑(152)
^{94.80}cara a que su hiijo salga beneficia(d)o_{97.98}↓/(347) ^{91.28}y tú no eres
capaz de decirle a tu padre na(da)↓//(928) ^{99.13}no_{99.13}→ ^{90.81}porque

él ya se arreglaará lo que tenga que arreglar digo ¡hombre! ^{91.76}↑/
 (315) ^{89.33}*haz el favor ¿eh?* ^{98.53}↑/(355) ^{98.53}*así que tú coges y le pre-*
guntas ^{87.47}→(234) ^{93.30}*es que está en un plan* ^{95.10}↑///(1.316) ^{95.83}*es*
que está en un ^{88.43}→(044) ^{92.39}*plaan quee-* ^{86.22}→///(598) ^{91.55}*que*
no ^{86.31}↓(079)

(Conversación 2016.PT.(20).S6 del Corpus Val.Es.Co. 3.0)

4. Conclusiones

Para concluir, el sistema propuesto en este trabajo parece mejorar sistemas anteriores de transcripción donde también se trataban de incorporar los aspectos prosódicos del habla. Ello se debe a varias razones objetivas, entre las que destacamos las siguientes:

– No resulta tan confuso o prolijo como otros sistemas de transcripción prosódica (SAMPROSA, ToBI...).

– Permite una lectura directa de las convenciones empleadas para representar los datos suprasegmentales.

– Cumple los requisitos exigibles a todo sistema de transcripción, particularmente los de exhaustividad y pertinencia.

– Incorpora parámetros no considerados por otros sistemas de transcripción, como es el caso del reajuste tonal.

– Sustituye la cuantificación en hercios de la F0 por su valoración en semitonos (st), lo que hace posible el estudio simultáneo de datos procedentes de individuos de diferente sexo.

– Emplea aplicaciones informáticas útiles que permiten la exportación de los datos y favorecen la realización de un análisis semiautomático de los mismos; así, para la medición del reajuste tonal y la determinación del tipo de tonema se puede llevar a cabo una medición manual, mientras que para la identificación de los alargamientos cabe aplicar un método automatizado.

En todo caso, dada la naturaleza específica del tipo de discurso que se está transcribiendo, el registro coloquial, no siempre el análisis acústico permite obtener datos inequívocos válidos para una transcripción «fiel» de la prosodia discursiva. Los solapamientos de habla, los ruidos, las risas, ciertas alteraciones «extrañas» de la melodía debidas a causas idiosincrásicas o contextuales, etc., son todos ellos factores que no podemos obviar, por cuanto el sistema aquí propuesto, pese a sus ventajas, deberá ser implementado en fases posteriores con objeto de mejorar en lo posible sus prestaciones. Es necesario, pues, probar todavía

su rentabilidad en sucesivos trabajos de transcripción prosódica que permitan determinar los implementos metodológicos necesarios para su progresivo perfeccionamiento.

BIBLIOGRAFÍA

- Alvar Ezquerro, Manuel, y Juan Andrés Villena Ponsoda (coords.) (1994), *Estudios para un corpus del español*, Málaga: Universidad de Málaga.
- Atkinson, J. Maxwell, y John Heritage (eds.) (1984), *Structures of social action: studies in conversation analysis*, Cambridge/París, Cambridge University Press/Éditions de la Maison des Sciences de l'Homme.
- Ávila Muñoz, Antonio Manuel, y Juan Andrés Villena Ponsoda (2010), *Variación social del léxico disponible en la ciudad de Málaga*, Málaga, Sarriá.
- Ávila Muñoz, Antonio Manuel (1996), «Problemas prácticos en la realización de corpus orales: la transliteración del corpus oral del proyecto de investigación de las variedades vernáculas malagueñas (VUM)», en Juan de Dios Luque Durán y Antonio Pamiés Bertrán (eds.), *Actas del Primer Simposio de Historiografía Lingüística*, Granada, Método Ediciones: 103-112.
- Beckman, Mary, Manuel Díaz-Campos, Julia Tevis McGory y Terrell Morgan (2002), «Intonation across Spanish, in the Tones and Break Indices framework», *Probus*, 14: 9-36.
- Blanche-Benveniste, Claire, y Jeanjean Colette (1987), *Le français parlé: transcription et édition*, París, Didier Érudition.
- Boersma, Paul, y David Weenink (2022), Praat: doing phonetics by computer [Programa informático]. Version 6.2.14. Disponible en <http://www.praat.org/>.
- Briz, Antonio (1998), *El español coloquial en la conversación: esbozo de pragmatogramática*, Barcelona, Ariel.
- Briz, Antonio, y Grupo Val.Es.Co. (1995), *La conversación coloquial: materiales para su estudio*, Anejo XVI de *Cuadernos de Filología*, Valencia, Universitat de València.
- Briz, Antonio, y Grupo Val.Es.Co. (2002), *Corpus de conversaciones coloquiales*, Anejo I de la revista *Oralia*, Madrid, Arco Libros.

- Button, Graham, y John Lee (eds.) (1987), *Talk and social organization*, Clevedon, Multilingual Matters.
- Cabedo, Adrián (2011a), «Hacia un modelo predictivo para la segmentación prosódica del discurso oral coloquial: MESTEL (Modelo Estadístico para la Selección de Términos Entonativos Ligados)», *Oralia*, 14: 85-104.
- Cabedo, Adrián (2011b), «El reajuste tonal en la delimitación de grupos entonativos», en Antonio Hidalgo Navarro, Yolanda Congosto Martín y Mercedes Quilis Merín (eds.), *El estudio de la prosodia en España en el siglo XXI, perspectivas y ámbitos*, Valencia, Universitat de València: 209-222.
- Cantero, Francisco José (2002), *Teoría y análisis de la entonación*, Barcelona, Edicions Universitat de Barcelona.
- Cantero, Francisco José (2019), «Análisis prosódico del habla: más allá de la melodía», en María Rosa Álvarez, Silva Álex Muñoz y Alvarado Leonel Ruiz Miyares (eds.) *Comunicación social: lingüística, medios masivos, arte, etnología, folclor y otras ciencias afines*, Volumen II, Santiago de Cuba, Ediciones Centro de Lingüística Aplicada: 485-498.
- Cantero, Francisco José, y Dolors Font (2007), «Entonación del español peninsular en habla espontánea: patrones melódicos y márgenes de dispersión», *Moenia*, 13: 69-92.
- Cantero, Francisco José, y Dolors Font (2009), «Protocolo para el análisis melódico del habla», *Estudios de Fonética Experimental*, 18: 19-32.
- Cantero, Francisco José, y Miguel Mateo (2011), «Análisis melódico del habla: complejidad y entonación en el discurso», *Oralia*, 14: 105-127.
- Cestero, Ana María (1994), *Análisis de la conversación: alternancia de turnos en la lengua española*, tesis doctoral, Universidad de Alcalá de Henares.
- Chafe, Wallace (1993), «Prosodic and functional units of language», en Jane A. Edwards y Martin D. Lampert (eds.), *Transcription and coding in discourse research*, Nueva Jersey, Lawrence Erlbaum Associates.
- DuBois, John W. (1991), «Transcription design principles for spoken discourse research», *Pragmatics*, 1:71-106.

- DuBois, John W., Stephan Schuetze-Coburn, Susanna Cumming, y Danae Paolino (1993), «Outline of discourse transcription», en Jane A. Edwards y Martin D. Lampert (eds.), *Talking data: transcription and coding in discourse research*, Hillsdale, Lawrence Erlbaum Associates: 45-90.
- EAGLES (1996), *Preliminary recommendations on spoken texts*, EAGLES Document EAG-TCWG-STP/P.
- ELAN (Version 6.8) [Computer software]. (2024), *Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive*. Recuperado de <https://archive.mpi.nl/tla/elan>.
- Elvira-García, Wendy, Paolo Roseano, Ana Fernández Planas y Eugenio Martínez Celdrán (2015), «Una herramienta para la transcripción prosódica automática con etiquetas Sp_ToBI en Praat», en Antonio Hidalgo y Adrián Cabedo (eds.), *Perspectivas actuales en el análisis fónico del habla: tradición y avances en la fonética experimental*, Valencia, Universitat de València: 455-464.
- Estebas Vilaplana, Eva, y Pilar Prieto (2008), «La notación prosódica del español: una revisión del Sp-ToBI», *Estudios de Fonética Experimental*, 17: 264-283.
- Garrido, Juan María (1991a), *Modelización de patrones melódicos del español para la síntesis y el reconocimiento de habla*, Bellaterra, Universitat Autònoma de Barcelona.
- Garrido, Juan María (1991b), «Modelización de patrones melódicos del español para sistemas de conversión texto-habla», *Procesamiento del Lenguaje Natural*, 11: 209-219.
- Gibbon, Dafydd (1989), *Survey of Prosodic Labelling for EC Languages. SAM-UBI-1/90, 12 february 1989; Report e.6, en ESPRIT 2589 (SAM) Interim Report, Year 1. Ref. SAM-UCL G002*, University College London.
- González, Julio, Teresa Cervera y José Luis Miralles (2002), «Análisis acústico de la voz: fiabilidad de un conjunto de parámetros multidimensionales», *Acta Otorrinolaringológica Española*, 53 (4): 256-268.
- Gumperz, John J., y Norine Berenz (1993), «Transcribing conversational exchanges», en Jane A. Edwards y Martin D. Lampert (eds.), *Talking data: transcription and coding in discourse research*, Hillsdale, Lawrence Erlbaum Associates: 91-122.

- Hidalgo, Antonio (1997), *La entonación coloquial: función demarcativa y unidades de habla*. Valencia, Anejo XXI de *Cuadernos de Filología*, Valencia, Universitat de València.
- Hidalgo, Antonio (2002), *Comentario fónico de textos coloquiales*, Madrid, Arco Libros.
- Hidalgo, Antonio (2018), «Unidades discursivas mínimas en la conversación: una aproximación de base prosódico-contextual», en Ester Brenes Peña, Marina González-Sanz y Francisco Grande Alija (eds.), *Enunciado y discurso: estructura y relaciones*, Sevilla, Universidad de Sevilla: 229-250
- Hidalgo, Antonio (2019), *Sistema y uso de la entonación en español hablado: aproximación interactivo-funcional*, Santiago de Chile, Universidad Alberto Hurtado.
- Hualde, José Ignacio (2003), «El modelo métrico y autosegmental», en Pilar Prieto (coord.), *Teorías de la entonación*, Barcelona, Ariel: 155-184.
- Llisterri, Joaquim (1997), *Transcripción, etiquetado y codificación de corpus orales*, Seminario de Industrias de la Lengua, Curso Etiquetado y extracción de información de grandes corpus textuales, Soria, Fundación Duques de Soria.
- MacWhinney, Brian (1991), *The CHILDES project: tools for analysis talk*, Hillsdale, Lawrence Erlbaum.
- Mateo, Miguel (2010), «Protocolo para la extracción de datos tonales y curva estándar en análisis melódico del habla (AMH)», *Phonica*, 6: 49-90.
- Mertens, Piet (2004), «The prosogram: semi-automatic transcription of prosody based on a tonal perception model», en Bernard Bel y Isabelle Marlien (eds.), *Proceedings of Speech Prosody*, Nata: 23-26.
- Navarro Tomás, Tomás (1944), *Manual de entonación española*, Nueva York, Hispanic Institute.
- Navarro Tomás, Tomás (1982), *Manual de pronunciación española*, Madrid, CSIC.
- Ochs, Elinor (1979), «Transcription as a theory», en Elinor Ochs y Bambi B. Schieffelin (eds.), *Developmental pragmatics*, Nueva York, Academic Press: 43-72.
- Payrató, Lluís (1995), «Transcripción del discurso coloquial», en Luis Cortés Rodríguez (ed.), *El español coloquial: actas del I*

- Simposio sobre Análisis del Discurso Oral*, Almería, Universidad de Almería: 43-70.
- Pierrehumbert, Janet B. (1980), *The phonology and phonetics of English intonation*, tesis doctoral, MIT.
- Pino, Marta, y Mercedes Sánchez (1999), «El subcorpus oral del banco de datos CREA-CORDE (Real Academia Española), procedimientos de transcripciones y codificación», *Oralia*, 2:83-138.
- Pitrelli, John, Mary E. Beckman, y Julia Hirschberg (1994), «Evaluation of prosodic transcription labelling reliability in the ToBI framework», en *Proceedings of the Third International Conference on Spoken Language Processing*, Yokohama, ICSLP, vol. 2: 123-126.
- Pons Bordería, Salvador (2022), *Creación y análisis de corpus orales*, Berna, Peter Lang.
- Quilis, Antonio (1975), «Las unidades de la entonación», en *Revista Española de Lingüística*, 5 (2): 261-280.
- Quilis, Antonio, Margarita Cantarero, y Manuel Esgueva (1993), «El grupo fónico y el grupo de entonación en español hablado», *Revista de Filología Española*, 73: 55-64.
- Sosa, Juan Manuel (2003), «La notación tonal del español en el modelo Sp-ToBI», en Pilar Prieto (ed.) *Teorías de la entonación*, Barcelona, Ariel: 185-208
- Stenström, Anna-Brita (1994), *An introduction to spoken interaction*, Londres/Nueva York, Longman.
- t'Hart, Johan, René Collier, y Antoine Cohen (1990), *A perceptual study of intonation: an experimental-phonetic approach to intonation*, Cambridge, Cambridge University Press.
- Tannen, Deborah (1987), *Conversational style: analyzing talk about friends*, Norwood, Ablex.
- Torruella, Joan, y Joaquim Llisterri, (1999), «Diseño de corpus textuales y orales», en José Manuel Blecua et al. (eds.), *Filología e informática: nuevas tecnologías en los estudios filológicos*, Barcelona, Editorial Milenio/Universitat Autònoma de Barcelona.
- Tusón, Amparo (1995), *Anàlisi de la conversa*, Barcelona, Empúries.
- Wells, John (1995), SAMPROSA (SAM Prosodic Transcription), Disponible en: <http://www.icp.grenet.fr/SpeechDat/home.html>.

Wells, John, William Barry, Martine Grice, Adrian Fourcin, y Dafydd Gibbon (1992), *Standard Computer-Compatible Transcription. SAM Stage Report Sen. 3 SAM UCL-037, 28 February 1992, en SAM (1992) ESPRIT PROJECT 2589 (SAM) Multilingual Speech Input/Output Assessment, Methodology and Standardisation. Final Report. Year Three: I.III.91-28.II.92*, Londres, University College London.