

Llisterri, J. (1999). Transcripción, etiquetado y codificación de corpus orales. In J. Gómez Guinovart, A. M. Lorenzo Suárez, J. Pérez Guerra & A. Álvarez Lugrís (Eds.), *Revista Española de Lingüística Aplicada*. Volumen monográfico “Panorama de la investigación en lingüística informática” (pp. 53–82). Logroño: AESLA, Asociación Española de Lingüística Aplicada
http://liceu.uab.cat/~joaquim/publicacions/RESLA_99.pdf

TRANSCRIPCIÓN, ETIQUETADO Y CODIFICACIÓN DE CORPUS ORALES

JOAQUIM LLISTERRI

Universidad Autónoma de Barcelona

RESUMEN. *En este trabajo se abordan tres niveles de representación propios de los corpus orales - transcripción ortográfica, representación fonética y etiquetado - así como cuestiones específicamente centradas en la transcripción y codificación de este tipo de corpus. En cada caso se presentan algunas de las convenciones desarrolladas como un primer paso hacia la estandarización, insistiéndose en la noción de compatibilidad entre sistemas para promover el intercambio y la reutilización de recursos.*

PALABRAS CLAVE. *Corpus orales, transcripción, etiquetado, codificación, estándares.*

ABSTRACT. *This work discusses three levels of representation in spoken corpora -orthographic transcription, phonetic representation and labelling- as well as some approaches to the transcription and encoding of spoken texts. Some proposals which have been put forward as a first attempt to standardization are presented, emphasizing the need of compatibility between different systems in order to promote exchange and reusability of resources.*

KEYWORDS. *Spoken corpora, transcription, labelling, encoding, standards.*

1. INTRODUCCIÓN

La constitución de recursos lingüísticos requiere una reflexión sobre diversos aspectos que inciden de forma directa en el resultado final obtenido y en su posterior utilización. Uno de ellos, especialmente importante en el caso de los corpus orales, es el que se refiere a las cuestiones relacionadas con la propia representación de los datos, tratadas generalmente bajo las denominaciones de “transcripción”, “etiquetado” y “codificación”. En el presente trabajo pretendemos poner de relieve que las decisiones relativas a estos tres aspectos deben basarse en un cuidadoso análisis lingüístico acompañado de unos planteamientos más globales sobre las futuras posibilidades de explotación del corpus. Para ello, se parte de una concepción del

corpus como conjunto de niveles de representación de los datos, en cada uno de los cuales se incorporan descripciones de los mismos que posteriormente deben ser recogidas mediante algún procedimiento formal.

2. NIVELES DE REPRESENTACIÓN, ETIQUETADO Y CODIFICACIÓN

Tanto si es oral como escrito, un corpus puede concebirse como un conjunto estructurado de materiales lingüísticos en el que se distinguen diversos niveles de representación correspondientes a diferentes grados de elaboración de los datos que lo constituyen. El establecimiento de estos niveles es una de las tareas que se llevan a cabo en la fase de definición del corpus y responde, lógicamente, a los objetivos del mismo.

Para cada nivel de representación suele establecerse un conjunto de “etiquetas” que se asocian a un determinado fragmento del corpus - un segmento sonoro, una unidad prosódica, una palabra, etc. - y definen sus propiedades. El etiquetado conlleva un enriquecimiento del corpus mediante información adicional introducida por el investigador en función de sus objetivos y, lo que es más importante, de su interpretación lingüística de los materiales recogidos. El establecimiento del inventario de etiquetas - estrechamente relacionado con los fenómenos que desean anotarse en un corpus - y las características que se señalan con cada una de ellas no es en absoluto una tarea trivial, especialmente en el marco de proyectos multilingües que deben cubrir simultáneamente las necesidades de varias lenguas.

Finalmente, es preciso definir, en el proceso de creación de recursos lingüísticos, un sistema o esquema de codificación tanto de los fenómenos representados en los diferentes niveles del corpus como de sus correspondientes etiquetas. La codificación constituye pues, tal como se concibe en la actualidad, una herramienta que dota al corpus de una estructura. Un esquema de codificación adecuado permite, entre otras cosas, mantener una distinción entre los datos y su forma de representación, que es útil cuando se requiere el intercambio de corpus o su reutilización en otros proyectos. Por ello es esencial, a la hora de diseñar tal esquema, considerar sistemas que cumplan este requisito, entre los cuales destaca el SGML (*Standard Generalized Markup Language*), un lenguaje jerárquico de codificación plenamente estandarizado (ISO 8879; Burnard 1995; URL: <http://www.sil.org/sgml/sgml.html>).

En el caso de un corpus oral, el primer nivel de representación necesario es el de la transcripción. Esto responde a la exigencia de materializar o fijar la información lingüística y comunicativa presente en una señal sonora esencialmente transitoria. Tal como la caracteriza Payrató (1995: 45), la transcripción es “un procedimiento de traslado o transposición a una forma gráfica (escrita) de una producción (lingüística, discursiva) originariamente oral”. Atendiendo a esta definición, podemos referirnos

tanto a la transcripción ortográfica como a la transcripción fonética, que se lleva a cabo mediante un conjunto de símbolos especialmente diseñados para tal fin. En los apartados que siguen centraremos nuestra atención en estos dos niveles de representación, el ortográfico y el fonético, considerando también algunos aspectos relativos al etiquetado en el nivel fonético.

3. REPRESENTACIÓN ORTOGRÁFICA DE CORPUS ORALES

Uno de los primeros problemas a los que se enfrenta un investigador en el momento de constituir un corpus oral es, evidentemente, el de la representación ortográfica. Esta representación constituye una transcripción de los enunciados producidos por los hablantes utilizando la ortografía convencional o, dicho en otras palabras, una transliteración. Puesto que se trata de un nivel de representación común a todo tipo de corpus orales sea cual sea su aplicación, se han realizado una notable cantidad de propuestas en lo que concierne a la forma escrita de los corpus orales. En el informe del Grupo de Trabajo sobre Textos Orales de EAGLES - cuyos principales elementos se retoman a continuación - se revisan algunas de ellas, incluyendo las que proceden del proyecto NERC en la tradición de la lingüística de corpus y las que se derivan de proyectos en el ámbito de las tecnologías del habla como *SpeechDat* o las desarrolladas en el propio ámbito de EAGLES.

3.1. *Propuestas para la representación ortográfica de corpus orales en el ámbito de la lingüística de corpus*

En el marco del proyecto NERC - *Network of European Reference Corpora* (NERC 1994), se adoptaron, para la representación ortográfica de corpus orales, las convenciones desarrolladas por French (1991, 1992), esencialmente concebidas para ser utilizadas en la constitución de la parte oral del corpus COBUILD (Payne 1995; URL:<http://www.cobuild.collins.co.uk>). En ellas se detalla el uso de los signos de puntuación - por ejemplo, no se utilizan comas entre frases y se emplean puntos para delimitar frases -, de las mayúsculas - únicamente al inicio de frase - o de los apóstrofes y las contracciones.

Por su parte, en el informe del Grupo de Trabajo sobre Textos Orales de EAGLES - *Expert Advisory Group on Language Engineering Standards* (EAGLES 1996; URL:<http://www.ilc.pi.cnr.it/EAGLES/home.html>) se proponen unas recomendaciones básicas para la transcripción ortográfica de corpus orales, recogiendo tanto las sugerencias de NERC como algunas de las ideas desarrolladas en el ámbito de las tecnologías del habla que se discuten en el próximo apartado. Las recomendaciones pueden resumirse en tres puntos:

- Se recomienda utilizar, siempre que sea posible, las formas ortográficas que aparecen en los diccionarios normativos o estándares para cada lengua, convención que se aplica también a contracciones, formas reducidas, apóstrofes, formas dialectales, interjecciones y formas semi-léxicas.
- Si puede darse más de una forma ortográfica de la misma palabra o si se introducen representaciones ortográficas que no aparecen en los diccionarios, es recomendable mantener una base de datos con las formas utilizadas en la transcripción.
- Los números, las abreviaturas, los acrónimos y las palabras deletreadas deben representarse ortográficamente tal como son pronunciadas por el hablante, utilizando la forma ortográfica completa.

Como puede observarse, éstas son recomendaciones generales que constituyen principios básicos. Un aspecto que requiere una discusión más detallada es el empleo de los signos de puntuación en la transcripción ortográfica de corpus, tema sobre el que existen diversas posturas, desde las que propugnan su completa eliminación - lo que disminuye notablemente la legibilidad del texto - hasta las que abogan por su uso según las normas habituales de cada lengua. Tal diversidad de opiniones es, obviamente, debida a las dificultades que plantea la introducción de puntuación en transcripciones de habla espontánea, ya que implica necesariamente una segmentación del enunciado que, en algunos casos, puede estar sujeta a la interpretación del transcriptor.

3.2. *Propuestas para la representación ortográfica de corpus orales en el ámbito de las tecnologías del habla*

En el campo de las tecnologías del habla existen diversas propuestas para la representación ortográfica, dos de las cuales se presentan a continuación. Pueden verse otros ejemplos en las Guías de Transcripción del *Center for Spoken Language Understanding* (CSLU, *Oregon Graduate Institute*) (Lander 1997) y en las convenciones desarrolladas para corpus como el *HCRC Map Task* (Anderson *et al.* 1991; URL: <http://www.elsnet.org/resources/mapTaskCorpus.html>) o en el proyecto *Verbmobil* (Hess *et al.* 1995; URL: <http://www.dfki.uni-sb.de/verbmobil/>; <http://www.ims.uni-stuttgart.de/projekte/verbmobil/index-en.html>).

Los sucesivos proyectos LRE-63314 *SpeechDat - Infrastructure for Spoken Language Resources* (URL: <http://www.icp.grenet.fr/SpeechDat/home.html>) y LE2-4001 *SpeechDat II - Speech Databases for the Creation of Voice Driven Teleservices* (URL: <http://www.phonetik.uni-muenchen.de/SpeechDat.html>) tienen entre sus objetivos desarrollar estándares para la creación de recursos lingüísticos especialmente adecuados al entrenamiento y evaluación de sistemas de reconocimiento automático del habla. Para llevar a cabo la transcripción ortográfica se han definido una serie de convenciones - resumidas en dos recomendaciones: “Keep it simple” y “Document

everything adequately” - entre las cuales se contempla el uso de mayúsculas y minúsculas, la transcripción de secuencias de números, de letras, y el empleo de los signos de puntuación; se establecen también procedimientos para la transcripción de errores de pronunciación, de palabras ininteligibles, fragmentos de palabras, elisiones y acontecimientos acústicos no verbales como las denominadas pausas llenas, ruidos producidos por el hablantes, ruidos estacionarios en la grabación o ruidos intermitentes, respondiendo estas últimas categorías a las necesidades de transcripción de una base de datos recogida a través del teléfono (Senia y van Velden 1997).

Los documentos del Grupo de Trabajo sobre Lengua Oral de EAGLES (Winski *et al.* 1995; Gibbon, Moore y Winski, eds. 1997; URL: <http://coral.lili.uni-bielefeld.de/~gibbon/EAGLES>) ofrecen las siguientes recomendaciones para transcripción ortográfica de corpus orales:

- Las formas reducidas de las palabras se transcriben tal como aparecen en los diccionarios normativos.
- Si es necesario, y si se documentan adecuadamente, pueden utilizarse formas que no aparezcan en los diccionarios.
- Se recomienda el uso de formas reducidas si aparecen frecuentemente y si implican elisiones de sílabas.
- Las formas dialectales se marcan en la transcripción.
- Los números se transliteran como palabras.
- En la transcripción ortográfica se utilizan las formas completas de las abreviaturas.
- Las abreviaturas que los hablantes pronuncian como palabras se transcriben de este mismo modo.
- En las transcripciones se indica la aparición de palabras deletreadas.
- Las interjecciones se transcriben con la representación ortográfica que aparece en los diccionarios.

Como puede deducirse fácilmente, la filosofía general de esta propuesta consiste en utilizar siempre las formas que aparecen en los diccionarios normativos y documentar adecuadamente las formas no estándares que se encuentren durante el proceso de transcripción.

4. REPRESENTACIÓN FONÉTICA DE CORPUS ORALES

A la hora de plantearse la elección de un sistema de transcripción fonética para un corpus oral, son útiles los planteamientos que Johansson (1995:93) realiza en el marco de la TEI (*Text Encoding Initiative*): “The degree of phonetic detail given in speech transcription varies from none to a very precise phonetic or phonemic

transcription [...] Where there is a great deal of phonetic or phonemic detail, it will be more convenient to design a specialized writing system”.

Evidentemente, hay casos en los que el objeto de la investigación no requiere la utilización de un sistema de transcripción fonética, pero cuando es necesario disponer de un “sistema especializado de escritura” el AFI (*Alfabeto Fonético Internacional*) aparece como el primer candidato para cubrir las necesidades de la investigación, siendo igualmente el sistema recomendado tanto por la TEI como por el Grupo de Trabajo sobre Textos Orales de EAGLES. Sin embargo, cuando se requiere el intercambio electrónico de textos, SAMPA (*SAM Phonetic Alphabet*) - considerado como una versión informática de parte del AFI (Gibbon, Moore y Winski, eds. 1997) - se presenta como una opción válida, especialmente en lo que se refiere a las lenguas europeas cuando se requiere una transcripción de carácter fonológico. Existe una extensión de SAMPA (X-SAMPA) que ofrece los elementos del AFI inexistentes en SAMPA y que, por lo tanto, puede considerarse como un sistema adecuado para la transcripción de corpus orales (EAGLES 1996). Por otra parte, cabe considerar también Wordlbet como una alternativa cuando la naturaleza del corpus aconseje disponer de un sistema más detallado. Estos procedimientos, comúnmente utilizados en la transcripción de corpus orales desarrollados en el ámbito de la fonética y las tecnologías del habla, se presentan someramente a continuación. Dada la diversidad de métodos existentes, parece clara la conveniencia de desarrollar mecanismos que permitan la compatibilidad, asegurando la fácil traducción de una representación a otra, de modo que diferentes investigadores puedan hacer uso de corpus ya transcritos fonéticamente, incorporándolos a su propio entorno de trabajo.

4.1. *La representación fonética de los elementos segmentales*

El Alfabeto Fonético Internacional (AFI / IPA, *International Phonetic Alphabet*)² es el sistema más ampliamente utilizado para la transcripción fonética. Es bien sabido que se basa en un principio fonológico, según el cual se transcriben aquellos segmentos con valor diferencial, utilizándose los diacríticos para introducir aspectos de tipo fonético (IPA 1995). Sin embargo, a pesar del uso extendido del AFI, existen también diversas tradiciones nacionales resultantes, en general, de las necesidades derivadas de la transcripción estrecha de una determinada lengua o de la orientación de la investigación hacia un determinada área lingüística (Pullum y Ladusaw 1986).

La necesidad de intercambiar electrónicamente textos que contienen transcripciones fonéticas llevó, a finales de los años 80, a plantearse la codificación de los símbolos del AFI, de modo que para cada uno de los símbolos del alfabeto se ha definido un número (denominado “número AFI”) que puede relacionarse también con los códigos ASCII (Esling 1988, 1990; IPA 1989; Esling y Gaylord 1993). Las equivalencias entre símbolos, números AFI y códigos ASCII forman parte de las

convenciones conocidas como CRIL (*Computer Representation of Individual Languages*) surgidas del congreso de la Asociación Fonética Internacional celebrado en Kiel en 1991.

SAMPA (*SAM Phonetic Alphabet*) constituye seguramente el esfuerzo más significativo y extendido por desarrollar un sistema de transcripción fonética informatizado. SAMPA tiene su origen en el proyecto ESPRIT 1541 SAM - *Speech Assessment Methods* (Fourcin *et al.* 1989; Fourcin y Dolmazon 1991) llevado a cabo entre 1987 y 1989. Al final del proyecto existían versiones de SAMPA en danés, neerlandés, inglés, francés, alemán e italiano, incorporándose en 1992 el noruego y el sueco, y el español, el griego y el portugués en 1993. Durante 1996 se extendió a una serie de lenguas de los países de la Europa del Este en el marco del proyecto BABEL (URL: <http://midwich.reading.ac.uk/research/speechlab/babel/>) y, en principio, no hay ningún obstáculo para su adaptación a otros idiomas. En cuanto a su utilización práctica, puede señalarse que es el sistema desarrollado para la transcripción del corpus multilingüe EUROM (Chan *et al.* 1995; URL: <http://www.phon.ucl.ac.uk/resource/eurom.html>) y también el adoptado en el proyecto ONOMASTICA, centrado en la transcripción de nombres propios en diversas lenguas europeas (Schmidt 1993; Trancoso 1995; URL: <http://guagua.echo.lu/langeng/en/lre1/onomas.html>).

SAMPA (Wells 1987, 1989; URL: <http://www.phon.ucl.ac.uk/home/sampa/home.htm>) se basa en un conjunto de equivalencias entre símbolos del AFI y códigos ASCII, restringiéndose al uso de caracteres de 7 bits (códigos 32 a 127). El principio que rige las transcripciones en SAMPA, al igual que en el caso del AFI, es esencialmente fonológico, empleándose únicamente símbolos distintos en el caso de segmentos con valor diferencial. Esta filosofía viene motivada por tres razones: simplicidad de la transcripción sin incorporar ambigüedades, ya que determinadas alofónicas son predictibles a partir del contexto; facilidad de utilización por parte personas con poca formación en fonética; y falta de códigos ASCII para la transcripción alofónica de todas las lenguas. Sin embargo, en algunos casos, se introducen símbolos para la representación de alófonos cuando los investigadores que han desarrollado las adaptaciones para cada lengua lo han considerado necesario.

Una de las limitaciones de SAMPA es que no establece una codificación para todos los símbolos del AFI. Por este motivo, Wells (1994) propone un sistema denominado X-SAMPA (URL: <http://www.phon.ucl.ac.uk/home/sampa/x-sampa.htm>) en el que se prevén equivalencias en códigos ASCII para la totalidad de símbolos del AFI en su revisión de 1993, incluyendo diacríticos y marcas tonales. Una de las aplicaciones propuestas para este sistema es el envío de textos en transcripción fonética a través del correo electrónico.

Por otra parte, Worldbet constituye una propuesta más ambiciosa que SAMPA y que su extensión X-SAMPA, pues parte de la base de que “any spectrally and temporally distinct speech sound (not including pitch) which is phonemic in some

language should have a separate base symbol” (Hieronymus 1994); los símbolos de base consisten normalmente en la concatenación de un símbolo del AFI con una serie de diacríticos, de modo que puede representarse una gran variedad de alófonos hallados en las lenguas del mundo, incluyendo símbolos inexistentes en el AFI. Uno de los objetivos de Wordlbet - que en su versión actual usa 299 símbolos compuestos de dos caracteres ASCII - es facilitar la transcripción de bases de datos multilingües que incluyan lenguas indias, asiáticas y africanas, además de las europeas. Este sistema se ha utilizado, por ejemplo, en la transcripción de los corpus distribuidos por el CSLU (*Center for Spoken Language Understanding, Oregon Graduate Institute*) (Lander 1997).

Existen otros sistemas de transcripción fonética concebidos para el intercambio electrónico de textos, tanto desarrollados desde la perspectiva de las tecnologías del habla - por ejemplo, el CPA (*Computer Phonetic Alphabet*) surgido en el marco del proyecto ESPRIT *Linguistic Analysis of European Languages* (Kluger-Kruse 1987) - como diseñados para otros fines. En el ámbito del estudio del lenguaje infantil, el proyecto CHILDES - *Child Language Data Exchange System* (MacWhinney 1991; URL: <http://poppy.psy.cmu.edu/childes/>) utiliza el sistema conocido como PHONASCII (Allen 1988), consistente en un conjunto de símbolos fonológicos - UNIBET - y un alfabeto fonético que permite la transcripción estrecha.

4.2. La representación fonética de los elementos suprasegmentales

La incorporación de información sobre los elementos suprasegmentales - información prosódica - en un corpus oral plantea diversos problemas, derivados precisamente de la propia naturaleza de estos elementos, que se manifiestan en variaciones continuas en los dominios de la frecuencia, la intensidad y el tiempo. Por tanto, es preciso llevar a cabo una abstracción para determinar, en primer lugar, cuáles de estas variaciones son lingüísticamente significativas y, en segundo lugar, cómo se relacionan con categorías discretas. Finalmente, estas categorías deben representarse en un sistema de notación. No es pues extraño que se hayan desarrollado diversas convenciones para la simbolización de los elementos suprasegmentales (Grønnum-Thorsen 1987; Llisterri 1994), algunas de las cuales se presentan en los apartados siguientes.

En el momento de plantearse la selección de un sistema de transcripción prosódica adecuado a un corpus oral el investigador se enfrenta nuevamente a la diversidad de sistemas, al igual que sucede en el caso de la transcripción segmental. Aunque, como veremos más adelante, ToBI parece que se está convirtiendo rápidamente en un estándar a pesar de su orientación hacia la transcripción del inglés y al modelo fonológico subyacente, SAMPROSA ofrece la ventaja de haberse desarrollado teniendo en cuenta las necesidades de las tecnologías del habla; además, SAMPROSA se enmarca en el contexto de estándares extendidos en Europa como los

del proyecto SAM, recomendados, por ejemplo, por NERC (Teubert 1993; Sinclair 1994). En conjunto, el sistema ideal de transcripción prosódica debería permitir una representación en varios niveles, ser compatible con el intercambio electrónico de datos y cubrir las necesidades del mayor número de lenguas posible, pudiendo aplicarse automáticamente en lugar de basarse en las impresiones intuitivas del transcriptor, con vistas a la transcripción de grandes corpus. En ausencia de un sistema que reúna tales características, parece adecuado establecer mecanismos de compatibilidad entre los existentes a fin de facilitar la reutilización de los datos.

El AFI ofrece, además de las posibilidades de transcripción segmental, un conjunto de símbolos para la representación de los elementos suprasegmentales. Con motivo del congreso para la revisión del AFI celebrado en Kiel en 1989 se creó un grupo de trabajo sobre categorías suprasegmentales (Bruce 1988, 1989), concluyéndose la necesidad de representar las categorías suprasegmentales de un modo más adecuado, especialmente en lo que se refiere la entonación. El conjunto actual de símbolos del AFI para los elementos suprasegmentales permite transcribir el tono, así como la duración y el acento.

Las necesidades derivadas de una descripción detallada de la entonación tanto en el campo de la fonética como en el de las tecnologías del habla han propiciado la creación de diversos sistemas de transcripción prosódica, algunos de los cuales se presentan a continuación.

En el marco del proyecto ESPRIT SAM anteriormente mencionado se llevó también a cabo el desarrollo de sistemas de transcripción prosódica compatibles con las necesidades de anotación de bases de datos en soporte electrónico, resultando de ello propuestas como PROSPA, SAMSINT o SAMPROSA (Gibbon 1989; Wells *et al.* 1992). PROSPA fue originalmente concebido por Selting y Gibbon (Selting 1987, 1988) para el análisis de la conversación y del discurso y, por tal motivo, se basa en criterios auditivos, recogiendo únicamente los elementos necesarios para este fin. Por su parte, SAMSINT (*SAM System for Intonation Transcription*) se plantea como objetivo la transcripción de contornos melódicos en el interior de unidades entonativas, basándose en INTSINT - presentado más adelante - con algunas modificaciones adicionales.

A pesar de la coexistencia de los sistemas anteriormente mencionados en el proyecto SAM, probablemente el conjunto de símbolos más extendido actualmente para la transcripción prosódica sea SAMPROSA (*SAM Prosodic Alphabet*), propuesto inicialmente por Gibbon (1989) y desarrollado por Wells *et al.* (1992) hasta llegar a su forma actual (Wells 1995). SAMPROSA ofrece, al igual que SAMPA, un conjunto de caracteres correspondientes a códigos ASCII, mediante los cuales es posible transcribir tonos globales, locales, terminales y nucleares, la duración, el acento, la pausas y las fronteras entre unidades prosódicas. Aunque presenta rasgos característicos de la tradición británica de análisis de la entonación como, por

ejemplo, la codificación de tonos nucleares, es un sistema que se plantea como adecuado para la transcripción de diversas lenguas.

En el ámbito americano, ToBI (*Tone and Break Index*) se desarrolló para cubrir al mismo tiempo las necesidades de la descripción fonética y de la anotación de corpus adecuados al desarrollo de sistemas de tecnologías del habla (Silverman *et al.* 1992; Beckman y Hirschberg 1994; Beckman y Ayers 1994; URL: <http://ling.ohio-state.edu/Phonetics/ToBI/Tobi.html>). Basado en el inglés americano, ToBI permite principalmente la transcripción de patrones entonativos partiendo de un análisis de la curva melódica representada como evolución temporal de la frecuencia fundamental mediante herramientas de análisis y etiquetado como el programa *Waves*TM. Sin embargo, se han definido también convenciones para la anotación de textos representados ortográficamente. Una transcripción utilizando ToBI consta de cuatro niveles:

- Representación ortográfica del enunciado.
- Nivel de índices de ruptura (*break index*), en el que se introducen los indicadores de límite que marcan las fronteras entre las unidades melódicas estableciendo cuatro grados de cohesión entre las mismas
- Nivel tonal, en el que se marcan los denominados “eventos tonales” (*pitch events*) divididos en tonos de frase al principio y al final de las unidades melódicas y en acentos tonales (*pitch accent*) en las sílabas en las que se produce un movimiento de la frecuencia fundamental
- Nivel misceláneo, que facilita, por ejemplo, introducir los comentarios del transcriptor.

Para cada uno de los niveles se define un conjunto de símbolos, alineados temporalmente con los fenómenos entonativos transcritos si se dispone de una herramienta adecuada como *Waves*TM.

Las críticas que se realizan más frecuentemente a ToBI son, por una parte, su dependencia del modelo fonológico desarrollado por Pierrehumbert (1980) y, por otra, su mejor adaptación al inglés que a otras lenguas, lo que explicaría su amplia utilización en Estados Unidos. Sin embargo, 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 como pueden ser su estructura jerárquica, que permite seleccionar entre subconjuntos o conjuntos mayores de símbolos, la posibilidad de representar problemas que aparecen en la transcripción y el hecho de que existan experimentos demostrando un alto grado de acuerdo entre transcriptores diferentes (Pitrelli *et al.* 1994).

En el *Institute for Perception Research* (IPO, Eindhoven) se ha desarrollado una teoría sobre la entonación basada en el modelado de contornos entonativos, consistentes en representaciones estilizadas perceptivamente equivalentes al contorno

original (Hart *et al.* 1990). Esta teoría ha llevado también a la formalización de un sistema de representación de la curva melódica, definida como un conjunto de movimientos tonales que se agrupan en secuencias de configuraciones tonales; la agrupación de estas configuraciones da lugar, a su vez, a contornos melódicos, definiéndose los patrones entonativos en función de agrupaciones similares de contornos melódicos. Un sistema jerarquizado de esta naturaleza permite una representación de la curva melódica adecuada para la transcripción prosódica, aunque que inicialmente no haya sido diseñado para tales fines.

El objetivo de INTSINT (*International Transcription System for Intonation*), en cambio, es proporcionar un sistema para la comparación interlingüística de sistemas prosódicos. Desarrollado por Hirst (1991; Hirst y Di Cristo, en prensa; Hirst, Di Cristo y Espesser, en prensa; URL: <http://www.lpl.univ-aix.fr/~hirst/intsint.html>) en el marco de un proyecto de análisis automático de la entonación, INTSINT se basa en la estilización de la curva melódica realizada a partir de una interpolación entre puntos en los que se da un cambio significativo en el valor de la frecuencia fundamental (*target points*). Estos puntos, que podríamos denominar “de inflexión” - definidos como la combinación de un valor temporal y un valor frecuencial - son susceptibles de codificarse automáticamente mediante los símbolos de INTSINT, previa introducción de marcas de frontera entre las unidades entonativas. Con ello se obtiene una representación fonológica del contorno melódico, simbolizando los diferentes niveles tonales de la curva original. El sistema se basa en la representación de secuencias tonales, consideradas tanto desde una perspectiva global en función del rango de variación tonal del hablante como desde un punto de vista local en relación con la altura de puntos de inflexión anteriores. En el primer caso se definen tres niveles: T (*Top*), M (*Mid*) y B (*Bottom*), mientras que en el segundo se marcan cinco posibilidades: H (*Higher*), L (*Lower*), S (*Same*), U (*Upstep*) o D (*Downstep*). El proceso de modelización necesario para transcribir mediante INTSINT ha sido aplicado a varias lenguas (Hirst *et al.* 1993) y se emplea también para la anotación prosódica de parte del corpus EUROM en francés, inglés, alemán, sueco y español (Hirst *et al.* 1994; Llisterri, ed. 1996) llevada a cabo en el marco del proyecto LRE 62-050 MULTEXT - *Multilingual Text Tools and Corpora* (URL: <http://www.lpl.univ-aix.fr/projects/multext/index.html>). Es posible también aplicar INTSINT a niveles más altos como el párrafo, tal como se muestra en Nicolas y Hirst (1995).

5. ETIQUETADO FONÉTICO DE CORPUS ORALES

Al igual que en el caso de los corpus textuales, una vez recogidos los materiales de base de un corpus oral, debe llevarse a cabo un tratamiento de los mismos que permita su utilización posterior. El primer paso suele ser la transcripción ortográfica, que en determinado tipo de corpus se acompaña de una transcripción fonética o

fonológica. A continuación, a cada segmento de la onda sonora se le asocia una etiqueta (*labelling*) que lo define en términos fonéticos o fonológicos y se lleva a cabo la alineación (*alignment*) entre la señal sonora y las etiquetas, obteniendo una representación que puede compararse a la partitura musical de una obra para voz y orquesta. El proceso de etiquetado segmental puede llevarse a cabo a varios niveles, como veremos en este apartado, y completarse con una anotación de las características suprasegmentales, codificadas según los diversos sistemas que hemos expuesto anteriormente. Si se cumplen las etapas mencionadas, se llega a disponer de un corpus que contiene la señal sonora sincronizada con la transcripción ortográfica y con la transcripción fonética o fonológica, de modo que, una vez definida una estructura de base de datos, es posible consultar el corpus partiendo de etiquetas fonéticas, de marcas prosódicas o de la transcripción ortográfica, al tiempo que se accede a la grabación correspondiente.

Los corpus de lengua oral que consisten únicamente en transcripciones ortográficas - ya que no suele ser factible realizar una transcripción fonética completa de un número elevado de horas de grabación - conllevan un procesamiento menos complejo, aunque en algunos casos contienen marcas prosódicas útiles para el análisis del discurso o de la conversación, como veremos en el apartado 6.

5.1. Niveles de etiquetado fonético de corpus orales

Una cuestión que se plantea inevitablemente a la hora de introducir información adicional sobre los datos recogidos es el nivel de etiquetado del corpus. Se han realizado varias propuestas sobre los distintos niveles en los que se puede realizar el etiquetado fonético de un corpus, algunas de ellas ligadas al proyecto SAM como la de Barry y Fourcin (1992), en la que se plantean cinco niveles posibles de etiquetado fonético de un corpus:

- Nivel físico, en el que se etiquetan las propiedades acústicas presentes en la señal sonora.
- Nivel acústico-fonético, en el que se marcan las propiedades de los segmentos en términos de categorías fonéticas como oclusión, fricción, sonoridad, nasalización, etc.
- Nivel de transcripción estrecha, en el que se caracterizan los segmentos utilizando los símbolos del AFI o de sus equivalentes computacionales.
- Nivel de transcripción fonológica (o fonémica), señalando únicamente aquellos elementos funcionalmente distintivos que se encuentran en la palabra pronunciada aisladamente en un estilo cuidado (*citation form*).
- Nivel de transcripción ancha, en el que se utilizan los símbolos correspondientes a los elementos fonológicos en una lengua dada para transcribir realizaciones alofónicas condicionadas por el contexto³.

A estos cinco niveles debe añadirse aún la anotación prosódica, considerada de forma independiente por los autores de la propuesta.

Una propuesta similar de etiquetado, también en cinco niveles - características acústicas de naturaleza subfonémica, características segmentales, realizaciones fonéticas, forma canónica de las palabras y transcripción ortográfica - se presenta en Tillmann y Pompino-Marschall (1993) y ha sido utilizada con éxito en el proyecto alemán *PhonDat*. Igualmente pueden verse las propuestas de Autesserre *et al.* (1989), enmarcadas en el contexto del proyecto SAM, y los trabajos de Marchal *et al.* (1995) o Roach *et al.* (1990).

La transcripción y el etiquetado de corpus orales se ha abordado en el marco del proyecto EAGLES desde una doble perspectiva: la de la fonética y las tecnologías del habla en el Grupo de Trabajo sobre Lengua Oral y desde la de la lingüística de corpus en el marco de las actividades del Grupo de Trabajo sobre Corpus Textuales. El trabajo llevado a cabo en el grupo dedicado a la lengua hablada en EAGLES recoge esencialmente las propuestas desarrolladas en el marco de SAM y de *PhonDat*. Los niveles de etiquetado propuesto son los siguientes (Gibbon, Moore y Winski, eds. 1997):

- Nivel ortográfico, en el que se utiliza la representación ortográfica convencional.
- Nivel de formas canónicas (*citation form*), representando la forma de pronunciación aislada de la palabra en estilo cuidado.
- Nivel de transcripción ancha o fonotípica: en este nivel se utilizan los símbolos que representan elementos fonológicos en una lengua dada, pero, al igual que en la propuesta de Barry y Fourcin, se transcriben aquellos fenómenos predictibles propios de la fonética sintáctica. SAMPA sería, pues, un alfabeto fonético adecuado para este nivel de transcripción.
- Nivel de transcripción estrecha, basado en una representación alofónica de las realizaciones fonéticas de los hablantes, para el cual puede utilizarse un alfabeto fonético como X-SAMPA.
- Nivel acústico-fonético, en el que se distinguen características acústicas de los segmentos observables en una representación de la onda sonora
- Nivel físico, representado en términos de parámetros acústicos o articulatorios
- Fenómenos no lingüísticos: la inclusión de este nivel permite representar sonidos de naturaleza comunicativa o no introducidos por el hablante, así como toda aquella información paralingüística que se considere de interés; también permite señalar los ruidos externos en la grabación.

Vale la pena detenerse brevemente en el nivel de las formas canónicas y en el concepto de “transcripción fonotípica” que, como puede verse, aparecen de forma recurrente en diversas propuestas. La inclusión de forma canónicas en el etiquetado de un corpus responde a un doble motivo, especialmente ligado a necesidades tecnológicas: en primer lugar, estas formas permiten derivar automáticamente,

mediante un conjunto de reglas, una transcripción ancha o fonotípica en la que se incluyen los elementos alofónicos cuya aparición esta condicionada por el contexto fonético; en segundo lugar, a partir de las formas canónicas puede crearse de forma automática un diccionario fonético partiendo de un corpus transcrito (Wesenick y Schiel 1995; Senia y van Velden 1997), lo que tiene utilidad para el desarrollo de sistemas de síntesis y reconocimiento del habla y también para la descripción lingüística. La transcripción denominada “fonotípica” en los trabajos de SAM constituye un nivel intermedio entre la realización fonética y la transcripción puramente fonológica. Por otra parte, el nivel de transcripción de formas canónicas está directamente relacionado con la representación ortográfica, facilitando así la transcripción fonética automática de corpus en este nivel mediante el desarrollo de los correspondientes sistemas de reglas.

En el informe del Grupo de Trabajo de Corpus Textuales de EAGLES (EAGLES 1996) se intenta realizar una síntesis entre las necesidades de diversos campos de estudio. Se sugieren, por ello, tres niveles de representación:

- S1: representación ortográfica del texto.
- S2: representación fonológica de las palabras en su forma canónica.
- S3: transcripción fonética, consistente en una representación simbólica discreta de la realización del enunciado tal como es percibido por el transcriptor.

Como puede observarse, estos tres niveles encuentran su equivalente en los niveles propuesto por Barry y Fourcin (1992) y en los definidos por el Grupo de Trabajo de Lengua Oral. Para la transcripción del nivel S2 puede utilizarse SAMPA, mientras que la del nivel S3 requiere el uso de un alfabeto fonético como X-SAMPA.

5.2. *La transcripción y el etiquetado automático de corpus orales*

La tarea de transcribir y etiquetar fonéticamente un corpus oral representa un considerable esfuerzo, tanto en términos de tiempo como de los recursos económicos necesarios para llevar a cabo esta labor. El etiquetado manual requiere, por otro lado, el desarrollo de convenciones detalladas para asegurar la consistencia entre diversos transcriptores (Keating *et al.* 1994; Krot y Taylor 1995; Lander 1997) además del establecimiento de un proceso de validación, comprobando la fiabilidad de los resultados obtenidos (Eisen 1993; Cole *et al.* 1994; Strangert y Heldener 1995). Por estos motivos, uno de los intereses principales de los investigadores en el campo de las tecnologías del habla ha sido llevar a cabo el etiquetado de forma semi-automática o, idealmente, automática. Un primer nivel de transcripción - tanto de formas canónicas como fonotípica - puede obtenerse automáticamente, como acabamos de mencionar, a partir de la representación ortográfica si se dispone de un conjunto de reglas que establezcan las correspondencias entre la representación ortográfica y la fonética.

Sin embargo, un etiquetado fonético en los niveles de transcripción más detallada requiere la posibilidad de acceder a la onda sonora y un procedimiento de análisis de la señal. La primera operación que debe realizar un sistema de etiquetado automático es la segmentación del *continuum* sonoro en unidades menores - segmentos o características acústicas, en función del nivel de transcripción deseado - para pasar después a la asignación de etiquetas - símbolos de transcripción - que describan sus características. Este es el proceso conocido propiamente como etiquetado (*labelling*) que, cuando se realiza de forma automática, debe superar los obstáculos derivados de la naturaleza no discreta de la señal sonora - resultado de mecanismos como la coarticulación - y de la variabilidad de realizaciones fonéticas entre hablantes o incluso en un mismo hablante.

Para que el corpus sea útil, el etiquetado fonético debe estar temporalmente sincronizado con la señal sonora y con la representación ortográfica, operación que se realiza en el proceso conocido como alineación temporal (*time alignment*), llevado a cabo también de forma automática. Esta alineación puede hacerse de un modo global, señalando en la onda sonora las fronteras entre palabras ortográficas - para facilitar el acceso a la grabación a partir de este nivel de representación (Lander 1997) - o los puntos en los que se encuentran sílabas acentuadas tal como sucede, por ejemplo, en el proyecto MULTEXT.

En el campo de las tecnologías del habla se han llevado a cabo numerosísimos trabajos encaminados al diseño de sistemas de segmentación, etiquetado y alineación temporal automáticas de corpus orales, utilizando las técnicas propias del reconocimiento del habla como los Modelos de Markov o las redes neuronales. Cabe destacar que tales técnicas no sólo son útiles en el desarrollo de corpus en este ámbito, sino que constituyen una herramienta importante en otros tipos de corpus; sin embargo, es preciso reconocer que el habla espontánea - especialmente si se ha recogido mediante grabaciones realizadas en un entorno natural - plantea aún problemas en lo que se refiere a su transcripción y etiquetado automáticos pese a los notabilísimos avances que tienen lugar en este campo.

6. TRANSCRIPCIÓN Y CODIFICACIÓN DE CORPUS PARA EL ANÁLISIS DE LA LENGUA ORAL

En este último apartado incidiremos en las cuestiones relacionadas con la transcripción y codificación de corpus orales que pueden presentar un mayor interés para ámbitos como el análisis de discurso o de la conversación y para el estudio de la lengua oral en todos los niveles de la descripción lingüística. Como indicábamos al principio, la transcripción y la codificación de los elementos transcritos son, en principio, dos operaciones conceptualmente diferentes. Sin embargo, en la práctica habitual en las disciplinas a las que se refiere este apartado, suelen ser dos operaciones

que se realizan simultáneamente, ya que al mismo tiempo que se realiza la transliteración se introducen las marcas necesarias para la codificación de los elementos que se ha decidido introducir en el corpus. Por ello, en muchos casos la discusión sobre la transcripción y la codificación se realiza conjuntamente en la bibliografía sobre el tema.

6.1. *Los requisitos de una transcripción*

El investigador que decide constituir un corpus de lengua oral se encuentra, en estos momentos, con una gran variedad de criterios en lo que se refiere a la transcripción y a la selección de los elementos que aparecerán codificados en el corpus. Hemos hecho ya referencia en el apartado 3 a las diversas convenciones para la representación ortográfica y, ciñéndonos al ámbito del análisis del discurso y la conversación, constatamos igualmente la existencia una gran diversidad de propuestas. En una revisión de las mismas, Payrató (1995) hace, por ejemplo, referencia a los sistemas de transcripción y codificación procedentes de la etnometodología, interesados en reflejar la interacción verbal, a los métodos desarrollados en el marco de la etnografía de la comunicación y de la sociolingüística interaccional⁴ o a propuestas utilizadas en el estudio del lenguaje infantil. Ante tal situación, parece conveniente plantearse cuáles son los requisitos que debe cumplir una transcripción de la lengua oral, tema que ha sido abordado por diversos autores (DuBois, 1991; Elich, 1993; Edwards, 1993; O'Connell y Kowal, 1994). Tras revisar aportaciones anteriores, Payrató (1995) resume los requisitos que, teóricamente, deberían exigirse a toda transcripción:

- Neutralidad o fidelidad, de modo que la transcripción no sea interpretativa.
- Globalidad o complejidad, recogiendo la totalidad de los fenómenos que aparecen en el discurso oral.
- Omnifuncionalidad, permitiendo diversos usos y aplicaciones.
- Claridad, tanto desde el punto de vista del aprendizaje del sistema como de la legibilidad de la representación.
- Universalidad y compatibilidad entre sistemas informáticos.

Sin embargo, puesto que en la práctica se transcribe en función de los objetivos para los que se ha definido el corpus, Payrató (1995) considera que un sistema de transcripción debe ser, ante todo, 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 diversos estudios. Por otra parte, en opinión de este autor, la simbología utilizada debe ser clara, económica, sencilla, exenta de ambigüedad y compatible con sistemas internacionales

estandarizados. Este último aspecto nos parece especialmente importante ya que, como señala el mismo Payrató (1995: 58):

[...] en ningún caso puede olvidarse que los datos de una investigación no deberían ser tratados nunca en forma tan parcial o idiosincrásica que se impidiera o dificultara el uso por parte de otros investigadores. La transcripción del discurso oral es una actividad suficientemente costosa y conflictiva para tomar precauciones en este sentido, y ninguna comunidad científica puede permitirse el lujo de particularismos (en aspectos tan elementales) que obstruyan el debate o conviertan en inaprovechables una fuente de información.

6.2. *Propuestas de estandarización para la transcripción y codificación de corpus orales*

Ante la diversidad de propuestas para la transcripción y codificación de la lengua oral que acabamos de constatar, el surgimiento de la lingüística de corpus y, muy especialmente, las necesidades derivadas del intercambio electrónico de textos propiciaron la aparición de proyectos e iniciativas centrados en la creación de estándares, tanto en lo que se refiere a la transcripción como a la codificación. En los próximos apartados revisamos sucintamente algunos de ellos.

Uno de los capítulos de las Guías desarrolladas por la TEI está dedicado a la transcripción de corpus orales (Johansson 1995a, b). En él se describe la estructura básica de la representación de un texto oral - cabecera, texto y divisiones - y se definen los procedimientos, basados en SGML, para la representación de los elementos estructurales básicos: información contextual, información temporal, enunciados, pausas, elementos vocalizados semi-léxicos o no léxicos, elementos kinésicos, otros tipos de eventos comunicativos y texto presentado en forma escrita al hablante. También se ofrecen en este capítulo indicaciones para la segmentación y el alineamiento, así como recomendaciones para la transcripción de superposiciones entre hablantes, de formas de palabras, prosodia, elementos paralingüísticos - *tempo*, intensidad, rango tonal, tensión, ritmo y calidad de voz - y alteraciones en el decurso fónico. Para la representación de la información fonética se recomienda el uso del AFI. En la siguiente tabla se resumen los principales elementos propuestos por la TEI para la codificación de corpus orales considerados específicos de este tipo de texto (Sperberg-McQueen y Burnard, eds. 1994):

Elemento codificado	Marca de codificación en SGML	Definición
Divisiones (<i>division</i>)	<div>	Unidades intermedias entre el texto y el enunciado que permiten delimitar partes diferenciadas en un texto.
Enunciado (<i>utterance</i>)	<u>	Segmento de habla comprendido entre dos pausas o delimitado por un cambio en el turno de palabra; puede incluir además información sobre la superposición (<overlap>) de turnos cuando interviene simultáneamente más de un hablante.
Pausa (<i>pause</i>)	<pause>	Interrupción de la fonación percibida entre dos enunciados o en el interior de los mismos; puede describirse en términos relativos o indicando su duración.
Vocal (<i>vocal</i>)	<vocal>	Elemento vocalizado semi-léxico o no léxico (p.ej. pausas llenas o toses).
Kinésico (<i>kinesic</i>)	<kinesic>	Cualquier fenómeno comunicativo no vocal (p. ej. gestos).
Acontecimiento (<i>event</i>)	<event>	Cualquier fenómeno identificado en la grabación no necesariamente vocalizado ni con valor comunicativo (p. ej. ruidos de fondo).
Texto escrito (<i>writing</i>)	<writing>	Texto escrito que se presenta al hablante durante su intervención.
Cambio (<i>shift</i>)	<shift>	Momento en el que se produce un cambio en alguno de los rasgos paralingüísticos - calidad de voz, intensidad, rango tonal, ritmo y velocidad de elocución -; cada uno de los rasgos puede describirse mediante una lista de características.

Tabla 1: Elementos codificados en las transcripciones de lengua oral según los estándares de la TEI (Adaptada de Sperberg-McQueen y Burnard, eds.1994)

Por otra parte, la TEI ofrece mecanismos para indicar el comienzo y el final de un fenómeno, su duración y su sincronización temporal con otros elementos; presenta también la posibilidad de marcar unidades inferiores al enunciado - segmentos <seg> -

que permiten delimitar, por ejemplo, unidades prosódicas, e incluye también entre sus propuestas las marcas necesarias para señalar fenómenos propios de la lengua hablada como palabras truncadas, repeticiones, falsos principios - marcados como (*deletion*) y especificando el fenómeno -. Las partes de la grabación que el transcriptor no llega a interpretar correctamente - <unclear> - o no puede escuchar - <gap> - o las correcciones de errores de producción realizadas por el transcriptor - <sic> para delimitar el error y <corr> para indicar la corrección - son también elementos para los cuales la TEI establece marcas de codificación.

Aunque las Guías de la TEI pueden considerarse un sistema adecuado para la transcripción y codificación de gran variedad de materiales, los autores propios autores reconocen que:

[...] the present proposals are not intended to support unmodified every variety of research undertaken upon spoken material now or in the future; some discourse analysts, phonologists, and doubtless other may wish to extend the scheme presented here to express more precisely the set of distinctions they wish to draw in their transcriptions. Speech regarded as a purely acoustic phenomenon may well require different methods from those outlined here, as may speech regarded solely as a process of social interaction (Sperberg-McQueen y Burnard, eds. 1994: cap. 11)

El proyecto NERC - *Network of European Reference Corpora* (NERC 1994) se planteó como uno de sus objetivos la viabilidad de desarrollar una serie de recomendaciones para la constitución de corpus y, por tanto, incluye entre sus resultados indicaciones sobre el tratamiento de los corpus orales, tanto en lo que se refiere a su codificación como a la transcripción de los elementos segmentales y suprasegmentales (Sinclair 1994). El sistema adoptado para la transcripción y codificación de corpus se basa, como hemos indicado anteriormente, en el desarrollado por French (1991, 1992) para la transcripción del corpus oral desarrollado en el proyecto COBUILD. La transcripción se plantea como una operación que puede desarrollarse en varios niveles:

- Nivel I: consiste en la representación ortográfica, introduciendo los mínimos signos de puntuación necesarios, sin información sobre la interacción entre hablantes. En este nivel se contemplan convenciones desarrolladas para la representación ortográfica y para precisar el uso de la puntuación.
- Nivel II: consiste en una representación ortográfica enriquecida con información básica sobre la identidad de los hablantes, el turno de palabra y los elementos no verbales.
- Nivel III: contiene toda la información presente en el nivel II además de información sobre la interacción entre hablantes - marcando el principio y el final de los solapamientos - y sobre la entonación - señalando las fronteras

entre unidades tonales y las sílabas acentuadas-. En opinión de French (1992) la transcripción a este nivel requiere una formación específica en fonética.

Nivel IV: constituye el nivel más detallado de representación, incluyendo la información propia del nivel III junto con marcas de tonos, sílabas nucleares y con una transcripción fonológica alineada con una representación acústica - espectrograma y curva melódica - del enunciado.

La recomendación del proyecto NERC es que la transcripción de un corpus para el estudio de la lengua oral se lleve a cabo según las especificaciones del Nivel II, adecuado, según French, para todas aquellas investigaciones que no requieran información prosódica.

En el marco de NERC se llevó también a cabo una evaluación de las Guías de la TEI para la transcripción de corpus orales de la que se concluye que: "The TEI proposals are broadly compatible with current practice in the user community, as represented by J.P. French conventions. Furthermore, in the majority of cases it will be a straightforward matter to link the machine-friendly TEI codes to the more user-friendly encoding systems such as J.P. French conventions by means of a simple conversion programme" (Payne 1992: 60). La idea de una conversión automática entre las transcripciones realizadas según las propuestas de NERC y de la TEI es, naturalmente, interesante, y refuerza la idea de que existe una compatibilidad general entre ambos sistemas (Payne 1992; Johansson 1995a; EAGLES 1996); prueba de ello es que en el informe final de NERC se recomienda el uso de las normas de la TEI.

Uno de los objetivos del Grupo de Trabajo sobre Textos Orales de EAGLES fue proponer un conjunto mínimo de fenómenos que deberían codificarse en cualquier texto oral. La propuesta realizada se describe en EAGLES (1996) y toma elementos del proyecto NERC, de la TEI y del Grupo de Trabajo sobre Lengua Oral de EAGLES. Considerando los elementos comunes a los sistemas que sirven de base, se sugiere que la codificación de un texto oral debe contener, al menos, información sobre los siguientes aspectos:

(1) Elementos vocales y no vocales

- Elementos vocales semi-léxicos

Pertencen a esta categoría los elementos utilizados principalmente en el habla espontánea durante los momentos en que el hablante planifica la continuación de su discurso como las llamadas "pausas llenas" o "sonoras" y los elementos que marcan dudas en el momento de la producción del habla. Corresponden a los elementos que en la TEI se marcan con la etiqueta <vocal>.

- Elementos vocales no léxicos
Se incluyen aquí todo tipo de sonidos producidos por el hablante - toses, risas, bostezos, estornudos, etc. - que no constituyen formas léxicas de la lengua. Corresponden igualmente a los codificados con la marca <vocal> en la TEI.
 - Acontecimientos no vocales y no comunicativos
Comprenden los ruidos producidos por otros hablantes o procedentes del entorno de grabación. Estos acontecimientos se codifican en la TEI mediante la marca <event>.
- (2) Elementos necesarios en la transcripción de interacciones verbales
- Identidad del hablante
Este es un elemento imprescindible en la transcripción de interacciones verbales. La TEI dispone de los mecanismos para documentar información sobre los hablantes en la cabecera de los textos, así como de un sistema para marcar la identidad de cada participante en el interior del texto, aunque pueden utilizarse también otros procedimientos.
 - Turnos de palabra
La finalidad del marcado de los turnos de palabra es indicar un cambio de hablante. Además de las convenciones de la TEI pueden emplearse otros procedimientos, mientras se mantenga esta información, esencial en el estudio de la interacción verbal.
 - Superposición de turnos de palabra
Existen igualmente diferentes procedimientos para marcar la superposición de emisiones de dos hablantes, además de las propuestas de la TEI, que los trata como parte de las estrategias para codificar acontecimientos simultáneos.
- (3) Elementos relacionados con la actuación del hablante
- Omisiones en un texto leído
Si se transcribe un texto leído y se dispone del original, es recomendable anotar las palabras o segmentos omitidos por el locutor.
 - Autocorrecciones
Es propio del habla espontánea, aunque también de la lectura, que los hablantes realicen correcciones de su propia producción lingüística (*self-repairs*). Parece conveniente marcarlas en la transcripción de un corpus oral.
 - Fragmentos de palabras
Los fragmentos de palabras corresponden a uno o más segmentos pertenecientes a una palabra que no ha sido totalmente pronunciada por el hablante en un primer intento y que suelen repetirse cuando éste consigue producir la palabra completa. Es importante indicar este fenómeno en la transcripción.

- Fragmentos ininteligibles

Se trata de partes de la grabación que no son comprendidas por el transcriptor y en la TEI se codifican con la etiqueta <unclear>, aunque si se desea se pueden utilizar marcas más detalladas distinguiendo entre transcripciones completamente ininteligibles y entre una posible interpretación ofrecida por el transcriptor.

Un aspecto conflictivo sigue siendo la identificación de los enunciados que conforman un texto oral (Schuetze-Coburn 1991). Mientras que en un texto escrito, los signos de puntuación constituyen un criterio de delimitación, en el habla disponemos de indicios como las pausas, la entonación o el cambio de turno de palabra que nos permiten, al menos convencionalmente, definir un enunciado. Sin embargo, hasta el momento, no se han desarrollado aún criterios objetivos y consistenes para resolver este problema.

7. CONCLUSIONES

La revisión –sin pretensiones de exhaustividad– de los principales sistemas de transcripción, etiquetado y codificación de corpus orales expuesta en este trabajo pone de manifiesto, en primer lugar, la multiplicidad de propuestas desarrolladas frente a los intentos de estandarización, algunos de los cuales aún no se han consolidado plenamente pese a su creciente aceptación. En segundo lugar, destaca también la heterogeneidad de tradiciones que confluyen en el estudio de la lengua hablada, desde la fonética y las tecnologías del habla hasta el análisis del discurso y de la conversación, diversidad que lleva a la creación de convenciones específicas para responder a las necesidades de diversas comunidades científicas. Ante esta situación, la distinción propuesta por Sinclair (1995:107) entre “conformidad” y “compatibilidad” adquiere todo su sentido, ya que si en ciertos casos la conformidad a un estándar determinado es problemática por las limitaciones inherentes a todo sistema común, la compatibilidad –entendida como la facilidad de conversión entre sistemas– debería ser, en cambio, un requisito imprescindible en el desarrollo de los recursos lingüísticos. De esta manera puede tal alcanzarse el equilibrio entre las necesidades de un proyecto y las indudables ventajas científicas y económicas que resultan de la reutilización de los datos disponibles.

NOTAS

1. Además de las *Guidelines for Electronic Text Encoding and Interchange* (Sperberg-McQueen y Burnard, eds. 1994) puede verse también sobre la TEI la recopilación de Ide y Véronis, eds. 1995. Para

- una introducción general son útiles Burnard (1995a, b). Se encuentra más información en URL: <http://www-tei.uic.edu/orgs/tei/> y en URL: <http://etext.virginia.edu/TEI.html>.
2. La versión más reciente del AFI data de 1993 (IPA 1993) y puede encontrarse en URL: <http://www.arts.gla.ac.uk/IPA/ipachart.html>. El *Summer Institute of Linguistics* (URL: <http://www.sil.org/computing/fonts/>) distribuye tipos de letra (*fonts*) que permiten trabajar con los sistemas operativos más corrientes utilizando los símbolos del AFI. También pueden obtenerse en *University College London* (URL: <http://www.phon.ucl.ac.uk/>), a partir de la página de *Linguist* (URL: <http://linguistlist.org/>), del Yamada Language Center (URL: <http://babel.uoregon.edu/yamada/fonts/phonetic.html>) o de la página de la propia Asociación Fonética Internacional (URL: <http://www.arts.gla.ac.uk/IPA/ipafonts.html>).
 3. La distinción entre los dos últimos niveles puede ejemplificarse con el caso de las asimilaciones: mientras que en la transcripción ancha encontraríamos [um poko] (un poco), en el caso de una transcripción fonológica aparecería [un poko] si se opta por no simbolizar archifonemas en la transcripción. En cambio [mismo] se representaría de la misma forma en ambos niveles, encontrándose la transcripción [mizmo] en el nivel de transcripción estrecha.
 4. Estos sistemas se comparan en O'Connell y Kowal (1994). Pueden encontrarse consideraciones generales sobre la transcripción y codificación de la lengua oral en el análisis del discurso y la conversación en Ochs (1979), Du Bois (1991), Edwards (1992, 1993, 1995), Du Bois *et al.* (1993), Gumperz y Berenz (1993), Cook (1995), Chafe (1995) y en Sinclair (1995).

BIBLIOGRAFÍA

- Allen, G. D. 1988. "The PHONAScii System". *Journal of the International Phonetic Association* 18,1: 9-25.
- Anderson, A. H, M. Badger, E. G Bard, E. Boyle, G. Doherty, S. Garrod, S. Isard, J. Kowtko, J. Mcallister, J. Miller, C. Sotillo, H. S. Thompson y R. Weinert. 1991. "The HCRC Map Task corpus". *Language and Speech* 34,4: 351-366.
- Autesserre, D., G. Pérennou y M. Rossi. 1989. "Methodology for the transcription and labeling of a speech corpus". *Journal of the International Phonetic Association* 19,1: 2-15.
- Barry, W. J. y A. J. Fourcin. 1992. "Levels of Labelling". *Computer Speech and Language* 6: 1-14.
- Beckman, M. E. y G. M. Ayers. 1994. *Guidelines for ToBI Labelling*. Version 2.0, February 1994. [Documento de Internet disponible en <http://ling.ohio-state.edu/Phonetics/ToBI/ToBI.0.html>].
- Beckman, M. E. y J. Hirschberg. 1994. *The ToBI Annotation Conventions*. Apéndice A de M. E. Beckman y G. M. Ayers 1994. *Guidelines for ToBI Labelling*. Version 2.0, February 1994. [Documento de Internet disponible en <http://ling.ohio state.edu/Phonetics/ToBI/ToBI.0.html>].
- Bruce, G. 1988. "2.3. Suprasegmental categories and 2.4. The symbolization of temporal events". *Journal of the International Phonetic Association* 18, 2: 75-76.

- Bruce, G. 1989. "Report from the IPA working group on suprasegmental categories". *Lund University Department of Linguistics, General Linguistics, Phonetics, Working Papers* 35: 25-40.
- Burnard, L. 1995. "What is SGML and how does it help?". *Computers and the Humanities* 29,1: 41-50. *The Text Encoding Initiative. Background and Context*. Eds. N. Ide y J. Véronis. Dordrecht: Kluwer Academic Publishers.
- Burnard, L. 1995a. *Text Encoding for Information Exchange. An Introduction to the Text Encoding Initiative*. TEI Document TEI J31. [Documento de Internet disponible en <http://www.uic.edu/orgs/tei/info/teij31/>].
- Burnard, L. 1995b. "The Text Encoding Initiative: an overview". *Spoken English on Computer: Transcription, Markup and Applications*. Eds. G. Leech, G. Myers y J. Thomas. Harlow: Longman. 69-81.
- Chafe, W. 1995 "Adequacy, user-friendliness, and practicality in transcribing". *English on Computer: Transcription, Markup and Applications*. Eds. G. Leech, G. Myers y J. Thomas. Harlow: Longman. 54-61.
- Chan, D., A. Fourcin, D. Gibbon, B. Granström, M. Huckvale, G. Kokkinakis, K. Kvale, L. Lamel, B. Lindberg, A. Moreno, J. Mouropoulos, F. Senia, I. Trancoso, C. Veld y J. Zeiliger. 1995. "EUROM - A Spoken Language Resource for the EU". *Eurospeech'95. Proceedings of the 4th European Conference on Speech Communication and Speech Technology*. Madrid, Spain, 18-21 September, 1995. Vol 1. 867-870.
- Cole, R. A., B. T. Oshika, M. Noel, T. Lander y M. Fauty. 1994. "Labeler Agreement in Phonetic Labeling of Continuous Speech". *Proceedings of the 1994 International Conference on Spoken Language Processing*, Yokohama, Japan, 18-22 September 1994. [Documento de Internet disponible en <http://www.cse.ogi.edu/CSLU/corpora/papers.html>].
- Cook, G. 1995. "Theoretical issues: transcribing the untranscribable". *Spoken English on Computer: Transcription, Markup and Applications*. Eds. G. Leech, G. Myers y J. Thomas. Harlow: Longman. 35-53.
- Du Bois, J. 1991. "Transcription design principles for spoken discourse research". *Pragmatics* 1: 71-106.
- Du Bois, J. W., S. Schuetze-Coburn, S. Cumming y D. Paolino. 1993 "Outline of discourse transcription". *Talking Data: Transcription and Coding in Discourse Research*. Eds. J. A. Edwards y M. D. Lampert. Hillsdale, N.J.: Lawrence Erlbaum Associates. 45-90.
- EAGLES. 1996. *Preliminary Recommendations on Spoken Texts*. EAGLES Document EAG-TCWG-STP/P, May 1996. [Documento de Internet disponible en <http://www.ilc.pi.cnr.it/EAGLES96/spokentx/spokentx.html>].
- Edwards, J. A. 1992. "Design principles in the transcription of spoken discourse". *Directions in Corpus Linguistics*. Proceedings of Nobel Symposium 82. Stockholm, 4-8 August, 1991. Ed. J. Svartvik. Berlin: Mouton de Gruyter. 129-147.

- Edwards, J. A. 1993. "Principles and Contrasting Systems of Discourse Transcription". *Talking Data: Transcription and Coding in Discourse Research*. Eds. J. A. Edwards y M. D. Lampert. Hillsdale Hillsdale, N. J.: Lawrence Erlbaum Associates. 3-32.
- Edwards, J. A. 1995. "Principles and alternative systems in the transcription, coding and mark-up of spoken discourse". *Spoken English on Computer. Transcription, Markup and Applications*. Eds. G. Leech, G. Myers y J. Thomas. Harlow: Longman. 19-34.
- Ehlich, K. 1993. "HIAT: A Transcription System for Discourse Data". *Talking Data: Transcription and Coding in Discourse Research*. Eds. J. A. Edwards y M. D. Lampert Hillsdale, N.J.: Lawrence Erlbaum Associates. 123-148.
- Eisen, B. 1993. "Reliability of speech segmentation and labelling at different levels of transcription". *Eurospeech'93. 3rd European Conference on Speech Communication and Technology*. Berlin, Germany, 21-23 September 1993. Vol. 1. 673-676.
- Esling, J. H. y H. Gaylord. 1993. "Computer Codes for Phonetic Symbols". *Journal of the International Phonetic Association* 23,2: 77-82.
- Esling, J. H. 1988. "Computer coding of IPA symbols and detailed phonetic representations of computer databases". *Journal of the International Phonetic Association* 18,2: 99-106.
- Esling, J. H. 1990. "Computer Coding of the IPA: Supplementary Report". *Journal of the International Phonetic Association* 20,1: 22-26.
- Fourcin, A. y J. M. Dolmazon (on behalf of the SAM Project). 1991. "Speech knowledge, standards and assessment". *Actes du XIIème Congrès International des Sciences Phonétiques*. 19-24 août 1991, Aix-en-Provence, France. Aix-en-Provence: Université de Provence, Service des Publications. Vol 5. 430-433.
- Fourcin, A., G. Harland, W. Barry y V. Hazan, eds. 1989. *Speech Input and Output Assessment. Multilingual Methods and Standards*. Chichester: Ellis Horwood Ltd.
- French, J. P. 1991. *Updated notes for soundprint transcribers*. Working paper, University of Birmingham, October 1991, NERC-WP4-47.
- French, J. P. 1992. *Transcription proposals: multilevel system*. Working paper, University of Birmingham, October 1992. NERC-WP4-50.
- Gibbon, D. 1989. *Survey of Prosodic Labelling for EC Languages*. SAM-UBI- 1/90, 12 February 1989; Report e.6, in ESPRIT 2589 (SAM) *Interim Report, Year 1*. Ref. SAM-UCL G002. University College London, February 1990.
- Gibbon, D., R. Moore y R. Winski, eds. 1997. *Handbook on Standards and Resources for Spoken Language Systems*. Berlin: Mouton De Gruyter. [Documento de Internet disponible en <http://coral.lili.uni-bielefeld.de/EAGLES/eagbook/eagbook.html>].

- Gronnum Thorsen, N. 1987. "Suprasegmental transcription". *ARIPUC - Annual Report of the Institute of Phonetics University of Copenhagen* 21: 1-27.
- Gumperz, J. J. y J. Berenzd. 1993. "Transcribing Conversational Exchanges". *Talking Data: Transcription and Coding in Discourse Research*. Eds. J.A. Edwards y M. D. Lampert. Hillsdale, N.J.: Lawrence Erlbaum Associates. 91-122.
- Hart, J. T., R. Collier y A. Cohen. 1990. *A Perceptual Study of Intonation. An Experimental - Phonetic Approach to Intonation*. Cambridge: Cambridge University Press.
- Hess, W., K. Kohler y H. G. Tillmann. 1995. "The PhonDat Verbmobil Speech Corpus". *Eurospeech'95. Proceedings of the 4th European Conference on Speech Communication and Speech Technology*. Madrid, Spain, 18-21 September, 1995. Vol 1. 863-866.
- Hieronymus, J. L. 1994. *ASCII phonetic symbols for the world's languages: Worldbet*. AT&T Bell Laboratories, Technical Memo. [Documento de Internet disponible en <http://www.cse.ogi.edu/CSLU/corpora/papers.html>].
- Hirst, D., A. Di Cristo y R. Espesser, en prensa. "Levels of representation and levels of analysis for intonation". *Prosody: Theory and Experiments*. Ed. M. Horne. Dordrecht: Kluwer.
- Hirst, D. J. y A. Di Cristo, en prensa. "A survey of intonation systems". *Intonation Systems. A Survey of 20 Languages*. Eds. D. Hirst y A. Di Cristo. Cambridge: Cambridge University Press.
- Hirst, D. J. 1991. "Intonation models: towards a third generation". *Actes du XIIIème Congrès International des Sciences Phonétiques*, 19-24 août 1991, Aix-en-Provence, France. Aix-en-Provence, Université de Provence, Service des Publications, Vol 1. 305-310.
- Hirst, D. J., A. Di Cristo, M. Le Besnerais, Najim, P. Nicolas y P. Roméas. 1993. "Multilingual modelling of intonation patterns". *Proceedings of an ESCA Workshop on Prosody*. September 27-29, 1993, Lund, Sweden. Eds. D. House y P. Touati. *Lund University Department of Linguistics and Phonetics, Working Papers* 41. 204-207.
- Hirst, D. J., N. Ide y J. Véronis. 1994. "Coding fundamental frequency patterns for multi-lingual synthesis with INTSINT in the MULTEXT project". *Conference Proceedings of the Second ESCA/IEEE Workshop on Speech Synthesis*. September 12-15, 1994. Mohonk Mountain House, New Paltz, New York, USA. 77-80.
- Ide, N. y J. Véronis, eds. 1995. "The Text Encoding Initiative: Background and Contexts". *Computers and the Humanities* 29, 1-3. Publicado en forma de libro en: Dordrecht: Kluwer Academic Publishers.
- IPA. 1989. "The IPA 1989 Kiel Convention Workgroup 9 report: Computer Coding of IPA symbols and Computer Representation of Individual Languages". *Journal of the International Phonetic Association* 19,2: 81-92.

- IPA. 1993. "IPA Chart, revised to 1993". *Journal of the International Phonetic Association* 23, 1. [Documento de Internet disponible en <http://www.arts.gla.ac.uk/IPA/ipachart.html>].
- IPA. 1995. Preview of the IPA Handbook. *Journal of the International Phonetic Association* 25, 1.
- ISO 8879. 1986. *Information Processing. Text and Office Systems. Standard Generalized Markup Language (SGML)*. Geneva: International Organization for Standardization.
- Johansson, S. 1995a. "The approach of the Text Encoding Initiative to the encoding of spoken discourse". *Spoken English on Computer: Transcription, Markup and Applications*. Eds. G. Leech, G. Myers y J. Thomas. Harlow: Longman. 82-98.
- Johansson, S. 1995b. "The Encoding of Spoken Texts". *Computers and the Humanities* 29,1: 149-158. *The Text Encoding Initiative. Background and Context*. Eds. N. Ide y J. Véronis. Dordrecht: Kluwer Academic Publishers. 149-158.
- Keating, P., P. MacEachern, A. Shryock y S. Domínguez. 1994. "A manual for phonetic transcription: Segmentation and labelling of words in spontaneous speech". *UCLA Working Papers in Phonetics* 88: 91-120.
- Kluger-Kruse, M. 1987. *Computer Phonetic Alphabet*. ESPRIT Linguistic Analysis of the European Languages. Report BU-CPA0267, July, 1987.
- Krot, C. y B. Taylor. 1995. *Criteria for Acoustic-Phonetic Segmentation and Word Labelling in the Australian National Database of Spoken Language*. [Documento de Internet disponible en <http://www.shlrc.mq.edu.au/criteria.html>].
- Lander, T. 1997. *The CSLU Labeling Guide*. Center for Spoken Language Understanding, Oregon Graduate Institute. [Documento de Internet disponible en <http://www.cse.ogi.edu/CSLU/corpora/papers.html>].
- Llisterri, J. 1994. *Prosody Encoding Survey*. WP 1 Specifications and Standards. T1.5. Markup Specifications. Deliverable 1.5.3. Final version, 15 September 1994. LRE Project 62-050 MULTEXT.
- Llisterri, J., ed. 1996. *Prosody Tools Efficiency and Failures*. WP 4 Corpus. T4.6 Speech Markup and Validation. Deliverable 4.5.2. Final version.. 15 October 1996. LRE Project 62-050 MULTEXT.
- MacWhinney, B. 1991. *The CHILDES Project: Tools for Analyzing Talk*. Hillsdale, N.J.: Lawrence Erlbaum.
- Marchal, A., N. Nguyen y W. Hardcastle. 1995. "Multitiered phonetic approach to speech labelling". *Levels in Speech Communication. Relations and Interactions*. A Tribute to Max Wajskop / Hommage à Max Wajskop. Eds. C. Sorin, J. Mariani, H. Meloni y J. Schoengen. Amsterdam: Elsevier Science B.V. 149-158.
- NERC. 1994. *NERC- 1. Network of European Reference Corpora. Final Report*. Pisa: Instituto di Linguistica Computazionale - CNR. January 1994; versión en curso

- de publicación: *Towards a Network of European Reference Corpora*. Eds. N. Calzolari, M. Baker y P.G. Kruyt. Pisa: Giardini.
- Nicolas, P. y D. Hirst. 1995. "Symbolic coding of Higher-Level Characteristics of Fundamental Frequency Curves". *Eurospeech'95. Proceedings of the 4th European Conference on Speech Communication and Technology*. Madrid, Spain, 18-21 September, 1995. Vol 2. 989-992.
- O'Connell, D. C. y S. Kowal. 1994. "Some Current Transcription Systems for Spoken Discourse: A critical Analysis". *Pragmatics* 4: 81-107.
- Ochs, E. 1979. "Transcription as Theory". *Developmental Pragmatics*. Eds. E. Ochs, E. y B.B. Schieffelin. New York: Academic Press. 43-72.
- Payne, J. 1992. *Report on the compatibility of JP French's spoken corpus transcription conventions with the TEI guidelines for transcription of spoken texts*. Working paper, COBUILD Birmingham and IDS Mannheim. December 1992, NERC-WP8/WP4-122.
- Payne, J. 1995. "The COBUILD spoken corpus: transcription conventions". *Spoken English on Computer: Transcription, Markup and Applications*. Eds. G. Leech, G. Myers y J. Thomas. Harlow: Longman. 203-207.
- Payrató, Ll. 1995. "Transcripción del discurso coloquial". *El español coloquial. Actas del I Simposio sobre Análisis del Discurso Oral*. Almería, 23-25 de noviembre de 1994. Ed. L. Cortés Rodríguez. Almería: Universidad de Almería, Servicio de Publicaciones. 43-70.
- Pierrehumbert, J. B. 1897 (1980). *The Phonology and Phonetics of English Intonation*. PhD Thesis, Department of Linguistics, MIT. Bloomington: Indiana University Linguistics Club.
- Pitrelli, J., M. Beckman y J. Hirschberg. 1994. "Evaluation of prosodic transcription labelling reliability in the ToBI framework". *Proceedings of the Third International Conference on Spoken Language Processing, Yokohama, ICSLP*, Vol. 2. 123-126.
- Pullum, G. K. y W. A. Ladusaw. 1996 (1986) *Phonetic Symbol Guide*. Chicago: The University of Chicago Press. 2nd. edition.
- Roach, P., H. Roach, A. Dew y P. Rowlands. 1990. "Phonetic analysis and the automatic segmentation and labeling of speech sounds": *Journal of the International Phonetic Association* 20,1: 15-21.
- Schmidt, M. S., C. Scott y M. A. Jack. 1993. "Phonetic transcription standards for European names (ONOMASTICA)". *Eurospeech'93. 3rd European Conference on Speech Communication and Technology*. Berlin, Germany, 21-23 September 1993. Vol. 1. 279-282.
- Schuetze-Coburn, S. 1991. "Units of intonation in discourse: a comparison of acoustic and auditory analysis". *Language and Speech* 34,3: 207-234.
- Selting, M. 1987. "Descriptive categories for the auditive analysis of intonation in conversation". *Journal of Pragmatics* 11: 777-791.

- Selting, M. 1988. "The role of intonation in the organisation of repair and problem handling sequences in conversation". *Journal of Pragmatics* 12: 293-322.
- Senia, F. y J. G. Van Velden. 1997. *Specifications of orthographic transcription and lexicon conventions*. LRE-4001 SpeechDat Technical Report SDI.3.2, Final version, 10 January 1997. [Documento de Internet disponible en <http://www.phonetik.uni-muenchen.de/SpeechDat.html>].
- Silverman, K., M. Beckman, J. Pitrelli, M. Ostendorf, C. Wightman, P. Price, J. Pierrehumbert y J. Hirschberg. 1992. "TOBI: A standard for labelling English prosody". *Proceedings of the Second International Conference on Spoken Language Processing, ICSLP-92*. Banff, October 1992. 867-870.
- Sinclair, J. 1994. "Spoken Language" [3B] "Phonetic/Phonemic and Prosodic Annotation" [5.2]. NERC 1994. *NERC-1 Network of European Reference Corpora, Final Report*. Pisa. Versión en curso de publicación: *Towards a Network of European Reference Corpora*. Eds. N. Calzolari, M. Baker y P.G. Kruyt. Pisa: Giardini.
- Sinclair, J. 1995. "From theory to practice". *Spoken English on Computer: Transcription, Markup and Applications*. Eds. G. Leech, G. Myers y J. Thomas. Harlow: Longman. 99-112.
- Sperberg-McQueen, C. M. y L. Burnard, eds. 1994. *Guidelines for Electronic Text Encoding and Interchange. TEI P3*. Chicago and Oxford: Association for Computational Linguistics / Association for Computers and the Humanities / Association for Literary and Linguistic Computing. [Cap. 11: Transcriptions of Speech] [Documento de Internet disponible en <http://etext.virginia.edu/TEI.html>].
- Strangert, E. y M. Heldner. 1995. "Labelling of boundaries and prominences by phonetically experienced and non-experienced transcribers". *Phonum 3, Reports from the Department of Phonetics, Umeå University*: 85-109. [Documento de Internet disponible en <http://www.ling.umu.se/fon/phonum3.html>].
- Teubert, W. 1993. *Phonetic / Phonemic and Prosodic Annotation*. Final Report, IDS Mannheim. February 1993. NERC-WP8-171.
- Tillmann, H. G. y B. Pompino-Marschall. 1993. "Theoretical Principles Concerning Segmentation, Labelling Strategies and Levels of Categorical Annotation for Spoken Language Database Systems". *Eurospeech'93. 3rd European Conference on Speech Communication and Technology*. Berlin, Germany, 21-23 September 1993. Vol. 3. 1691-1694.
- Trancoso, I. 1995. "The ONOMASTICA Interlanguage Pronunciation Lexicon". *Eurospeech'95. Proceedings of the 4th European Conference on Speech Communication and Technology*. Madrid, Spain, 18-21 September, 1995. Vol. 1. 829-832.
- Wells, J. C. 1987. "Computer Coded Phonetic Transcription". *Journal of the International Phonetic Association* 17,2: 94-114.

- Wells, J. C. 1989. "Computer-coded phonemic notation of individual languages of the European Community". *Journal of the International Phonetic Association* 19, 1: 31-54.
- Wells, J. C. 1994. "Computer-coding the IPA: a proposed extension of SAMPA". *Speech, Hearing and Language, Work in Progress, 1994* (University College London, Department of Phonetics and Linguistics) 8: 271-289.
- Wells, J. C. 1995. *SAMPROSA (SAM Prosodic Transcription* [Documento de Internet disponible en <http://www.phon.ucl.ac.uk/home/sampa/samprosa.html>].
- Wells, J. C., W. Barry, M. Grice, A. Fourcin y D. Gibbon. 1992. *Standard Computer Compatible Transcription*. SAM Stage Report Sen.3 SAM UCL-037, 28 February 1992. SAM (1992) Esprit Project 2589 (SAM) Multilingual Speech Input/Output Assessment, Methodology and Standardisation. *Final Report. Year Three: I. III. 91-28.II. 1992*. London: University College London.
- Wesenick, M.-B. y F. Schiel. 1995. *Feasibility of Automatic Annotation and Building Pronunciation Lexica from Corpus Material*. LRE-63314 SpeechDat, Report D3.1.2.3., Final version, 10 October 1995. [Documento de Internet disponible en <http://www.icp.grenet.fr/SpeechDat/home.html>].
- Winski, R., R. Moore y D. Gibbon. 1995. "EAGLES Spoken Language Working Group: Overview and Results". *Eurospeech'95. Proceedings of the 4th European Conference on Speech Communication and Speech Technology*. Madrid, Spain, 18-21 September, 1995. Vol 1. 841-844. [Documento de Internet disponible en: <http://coral.lili.uni-bielefeld.de/~gibbon/EAGLES/rwpaper/rwpaper.html>].