Desarrollo de un módulo de asignación de parámetros prosódicos para la versión en español del sistema de conversión texto-habla ACTOR®

Juan M. Garrido Almiñana, Isabel Ortín Blanco
Departament de Filologia Espanyola, Universitat Autònoma de Barcelona
juanma@liceu.uab.es, isabel@liceu.uab.es

Silvia Quazza, Pier Luigi Salza, Franca Mancini
Centro Studi e Laboratori Telecomunicazioni s.p.a. (CSELT)
silvia.quazza@cselt.it, pierluigi.salza@cselt.it, franca.mancini@cselt.it

Resumen

El objetivo de esta comunicación es describir los módulos de análisis prosódico y de asignación de curvas melódicas desarrollados para la versión en español del sistema de conversión texto-habla ACTOR®. En primer lugar se hace una presentación general de la estructura del conversor, y se describe de forma general el proceso de adaptación al español. A continuación se describe la aproximación empleada en el desarrollo del módulo de segmentación prosódica, que incluye la segmentación de los enunciados en grupos acentuales, en grupos tónicos y en grupos entonativos. Finalmente, se definen los principios teóricos en los que se basa el funcionamiento del módulo de asignación de curvas de F0.

Introducción

El objetivo de esta comunicación es describir los módulos de análisis prosódico y de asignación de curvas melódicas desarrollados para la versión en español del sistema de conversión texto-habla ACTOR®. De este sistema existen en la actualidad dos versiones, una en italiano y otra en español.

Descripción general del conversor

La estructura general de ACTOR® se organiza en dos grandes bloques, tal como se observa en la figura 1:

1) Un módulo de análisis lingüístico-prosódico, encargado de transformar el texto ortográfico en su correspondiente transcripción fonética, enriquecida con información lingüística.

2) Un módulo de síntesis, que convierte la cadena de símbolos obtenida a la salida del módulo anterior en la onda sonora correspondiente.

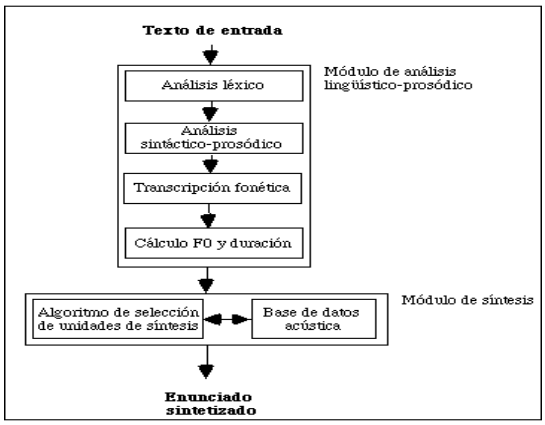

Figura 1. Esquema general del funcionamiento del sistema ACTOR®

El módulo de análisis lingüístico-prosódico realiza básicamente tres funciones. En primer lugar se procede al análisis léxico de las palabras del texto, con el objetivo de determinar su categoría gramatical y las sílabas con acento léxico. En segundo lugar, se asignan etiquetas de límite prosódico a partir de un análisis sintáctico-prosódico previo. Finalmente, se lleva a cabo la transcripción fonética y se calculan los parámetros suprasegmentales de F0 y duración para cada sonido.

El módulo de síntesis utiliza una técnica basada en la concatenación de unidades extraídas de un corpus amplio (*corpus-based*

concatenative synthesis). A diferencia de los enfoques tradicionales, basados en el uso de unidades fijas de tamaño reducido (difonemas, trifonemas, etc), esta técnica de síntesis intenta utilizar las unidades de la base de datos acústica almacenada más largas y más adecuadas al contexto prosódico, con el objeto de reducir tanto el número de concatenaciones como la magnitud de la modificación prosódica necesaria, dos de los factores principales que inciden en la degradación del habla sintética generada. Esta técnica se basa en la premisa de obtener calidad gracias a la cantidad: cuanto mayor es la base de datos acústica, mayores posibilidades hay de obtener la unidad de síntesis más adecuada. Una descripción más detallada de este procedimiento puede encontrarse en Balestri *et al.* (1999).

Adaptación de ACTOR® al español

A partir de la estructura del sistema ACTOR® italiano, se planeó su adaptación al español, desarrollando los módulos lingüísticos y fonéticos específicos para esta lengua. Esta adaptación ha implicado las siguientes tareas:

a) Definición del inventario de sonidos del español para ACTOR®

b) Adaptación del módulo de preprocesado (expansión de siglas, símbolos, abreviaturas y números)

c) Definición e implementación de reglas de transcripción fonética, silabificación y acentuación

d) Definición de reglas de segmentación prosódica

e) Definición y grabación de la base de datos de unidades de síntesis

f) Definición e implementación de reglas de asignación de curvas de F0, pausas y duración.

g) Evaluación de los diferentes módulos y de la calidad del habla generada.

La versión actual del sistema se puede considerar completa por lo que se refiere a los módulos de expansión de los números y abreviaciones, acentuación lexica, transcripción fonética y realización de las unidades acústicas de síntesis. Los módulos prosódicos, desarrollados en las tareas d) y f), están todavía en fase de implementación. La metodología utilizada en su desarrollo y los resultados alcanzados hasta ahora constituyen el objeto del presente trabajo.

Desarrollo del módulo de asignación de curvas de f0

La nueva técnica de síntesis adoptada en ACTO® reduce la necesidad de modificar de forma artificial la prosodia de la señal concatenada, porque las unidades de síntesis están constituidas por largos fragmentos de señal extraída del contexto prosódico apropiado: donde es posible, pues, las duraciones y la F0 de la señal original se mantienen inalteradas. Esto sólo es válido, sin embargo, para un tipo de prosodia enunciativa “neutra”, con la que está grabada la base de datos acústica.

Para modalidades oracionales diferentes, por ejemplo en las frases interrogativas, es necesario que la señal original, con entonación enunciativa, se modifique de manera que su curva melódica corresponda a la entonación deseada. Para alcanzar este objetivo, es necesario que las distintas modalidades entonativas estén representadas por modelos adecuados, que permitan calcular la curva de F0 deseada y superponerla a la señal original.

Una de las tareas necesarias para el desarrollo de la versión española de ACTOR® es, pues, la realización de un módulo para calcular las curvas de F0, indispensable para la realización de las interrogativas y útil, en segundo lugar, también para un mejor control de las enunciativas.

El modelo se ha desarrollado a partir del análisis de un corpus de oraciones interrogativas diseñado especialmente con este objetivo, que fue grabado por el mismo locutor empleado para la grabación de la base de datos acústica. Por el momento el modelo sólo permite la generación de curvas de F0 para oraciones interrogativas, aunque está previsto ampliarlo con el análisis de oraciones enunciativas y exclamativas.

Para la definición del modelo se ha utilizado un marco teórico de referencia basado en la descripción de las curvas melódicas del español presentada en Garrido (1996), semejante al aplicado en la versión italiana de ACTOR®. Este marco teórico es el que se describe a continuación.

La representación estilizada de las curvas melódicas

La primera premisa que asume el modelo es que las curvas melódicas se pueden representar

de forma ‘estilizada’ como una serie de puntos que representan los cambios de dirección relevantes en la curva melódica o puntos de inflexión (Garrido, 1991). De esta forma se pueden obtener representaciones que contengan las variaciones relevantes de la curva melódica, eliminando las irrelevantes (variaciones micromelódicas, por ejemplo).

El análisis acústico que se llevó a cabo sobre las curvas melódicas del corpus consistió en obtener representaciones estilizadas de las mismas, determinando los puntos de inflexión relevantes, tal como se ejemplifica en la figura 2. Consecuentemente, el modelo desarrollado a partir de este análisis permite generar curvas melódicas en forma de series de puntos unidos mediante líneas.

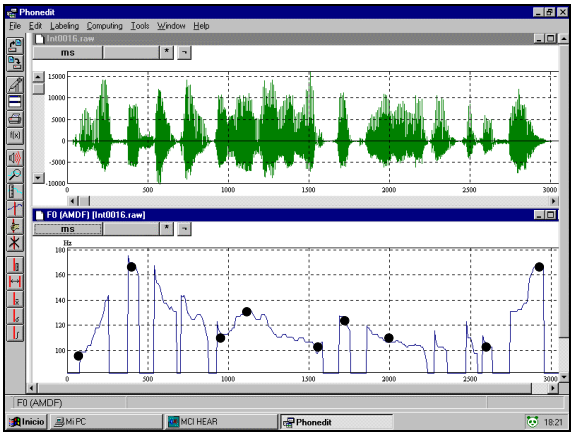

Figura 2: Curva melódica del enunciado ‘¿Conoces el contenido del artículo de la Constitución?’ en el que aparecen los puntos de inflexión considerados durante el análisis.

Niveles tonales

El modelo asume igualmente que en las curvas melódicas es posible diferenciar cuatro niveles tonales diferentes:

a) un nivel P, correspondiente a los puntos de inflexión que coinciden con máximos -o picos- en la curva melódica;

b) un nivel V, correspondiente a los mínimos, o valles;

c) un nivel M, o medio, que aparece al inicio de aquellas curvas melódicas que se inician con una sílaba tónica;

d) un nivel P+, por encima del nivel P, que alcanzan ciertos picos a lo largo de la curva melódica.

El desarrollo del modelo ha implicado asignar los puntos de inflexión detectados en las curvas melódicas del locutor a uno de estos cuatro niveles tonales, tal como se ejemplifica en la figura 3.

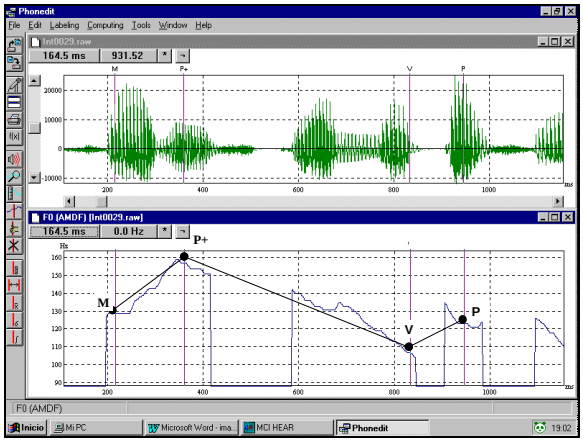

Figura 3: Curva melódica del segmento ‘Sabes cuando pasan…’ perteneciente al enunciado ' ¿Sabes cuándo pasan a recoger los muebles viejos este mes?' en la que aparecen representados los puntos M, P+, V y P.

La estructura jerárquica de las curvas melódicas

La tercera premisa en que se basa este modelo es que en las curvas melódicas es posible reconocer formas recurrentes, que se repiten independientemente del enunciado y del locutor. Son los llamados patrones melódicos. Las curvas melódicas son el resultado de la suma de una serie de estos patrones.

La elaboración de un modelo de las curvas melódicas de un locutor implica, por tanto, la definición de un inventario de patrones melódicos que este locutor utiliza en la construcción de sus curvas melódicas.

En este modelo se asume además que las curvas melódicas son el resultado de la superposición de patrones de ámbito diferente para obtener la curva melódica final: un nivel global, que da cuenta de la forma general de la curva melódica, y un nivel local, que define las variaciones locales de la misma (picos y valles asociados con las sílabas acentuadas, tonos de límite), tal como se ilustra en la figura 4.

Esta idea, aplicada al desarrollo del modelo, ha implicado la definición de dos tipos de patrones melódicos:

1) Patrones locales: referidos al nivel local (acentos, tonos de límite)

2) Patrones globales: referidos al nivel global.

Cada uno de estos dos tipos de patrones tienen ámbitos de aplicación diferentes: el grupo acentual en el caso de los patrones locales, y el grupo entonativo y la oración en el caso de los

patrones globales.

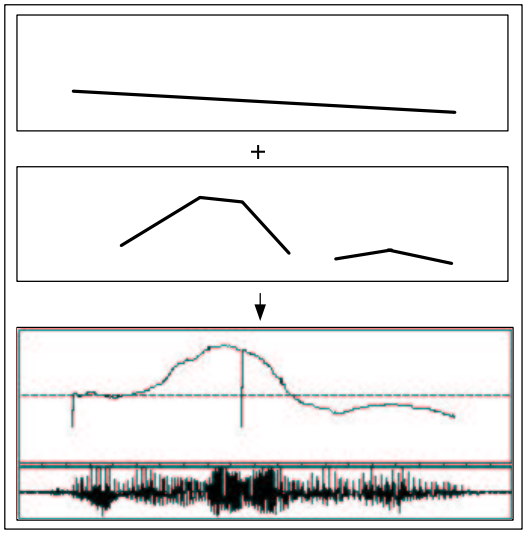

Figura 4: Descomposición en patrones melódicos superpuestos de la curva melódica del enunciado ‘Ramón llegó en avión’, pronunciada por un locutor masculino.

Patrones locales

Los patrones locales se definen como series de puntos de inflexión, etiquetados como P+, P, M o V, y asociados a puntos determinados del enunciado, que se dan en el ámbito de un grupo acentual (GA). Un GA se define como la secuencia formada por una sílaba tónica (con acento primario) y todas las sílabas átonas (sin acento primario) que la siguen hasta la siguiente sílaba tónica.

De acuerdo con el modelo presentado aquí, las curvas melódicas se construyen combinando patrones locales de cuatro tipos:

a) iniciales: aparecen al principio de un grupo entonativo

b) interiores: aparecen en interior de grupo entonativo

c) finales: aparecen al final de un grupo entonativo

d) iniciales-finales: abarcan por sí mismos un solo grupo entonativo

En las figuras 5 y 6 se ilustran los diferentes tipos de patrones locales.

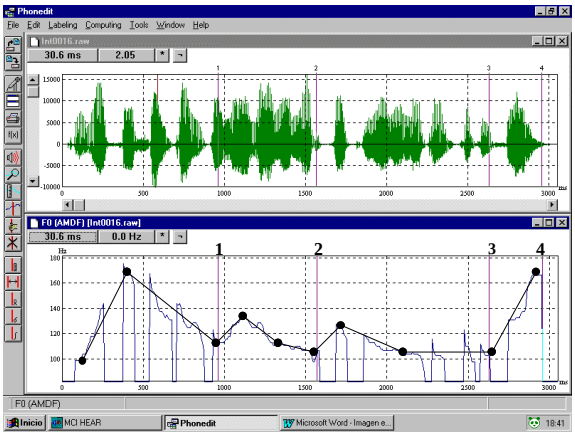

Figura 5: Curva melódica del enunciado ‘¿Conoces el contenido del artículo de la Constitución?’ en el que aparecen estilizados el patrón inicial (1), dos intermedios (2,3) y el patrón final (4).

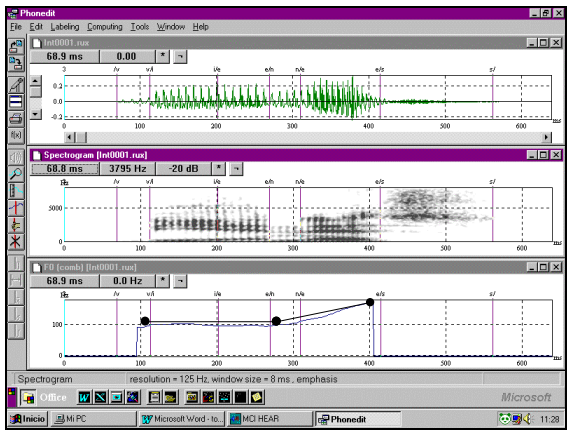

Figura 6: Curva melódica del enunciado ¿Vienes?, que se realiza con un solo patrón inicial-final.

El proceso de desarrollo del modelo ha implicado, por lo que se refiere a la definición de los patrones locales, el establecimiento de los patrones locales empleados por el locutor de referencia, y la selección de los más frecuentes para su inclusión en el mismo.

Patrones globales

En la aproximación adoptada aquí los patrones globales son los que se aplican en el ámbito del grupo entonativo (GE) y la oración[1].

En el modelo se asume que la forma global de las curvas melódicas a lo largo de un grupo

[1]La oración se define aquí, siguiendo un criterio ortográfico, como una porción de texto cuyo límite final se marca con uno de los siguientes signos de puntuación: <!> <?>, <.>, <:>, <…>.

entonativo puede representarse mediante una serie de 'líneas de referencia', que indican la evolución de los picos (P) y los valles (V) a lo largo de la curva, como ejemplifica la figura 7.

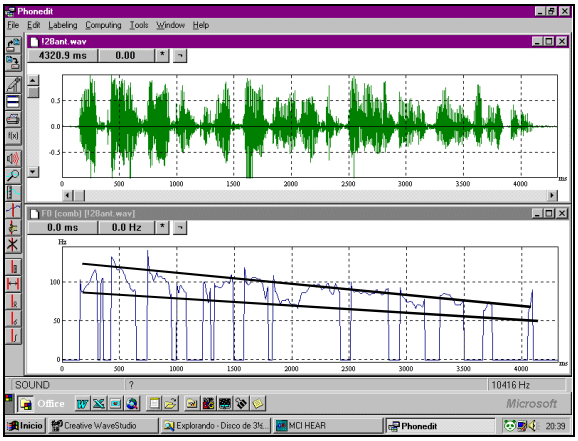

Figura 7: Curva melódica del enunciado 'El acto de la firma transcurrió tal y como habían pactado ambas delegaciones la víspera', pronunciada por un locutor masculino. Las líneas rectas representan las 'líneas de referencia' correspondientes a los puntos P (superior) y V (inferior) de la curva melódica.

El desarrollo de los patrones globales ha implicado, de acuerdo con esto, el cálculo de las líneas de referencia de los diferentes grupos entonativos del corpus de referencia, así como la definición de una serie de líneas-patrón, formuladas por medio de ecuaciones del tipo:

$$y = ax + b$$

En la que 'y' equivale al valor de F0 en un instante determinado de tiempo, 'x' equivale al tiempo, y 'a' y 'b' son constantes que se calcularon en cada caso.

Cuando un enunciado está compuesto por más de un grupo entonativo, cada uno tiene sus propias líneas de referencia, a una altura tonal diferente. Entre las líneas de referencia de dos grupos entonativos consecutivos se produce el fenómeno del 'reajuste de F0' (*F0 reset*), que eleva el nivel tonal de las líneas de referencia del nuevo grupo por encima del nivel de las líneas de referencia del grupo anterior (Garrido, 1996). El reajuste de F0 puede presentar diversos grados o 'niveles', que parecen relacionarse con factores sintácticos y pragmáticos (Garrido, 1996), aunque se trata de un fenómeno todavía poco estudiado. La figura 8 ilustra este fenómeno.

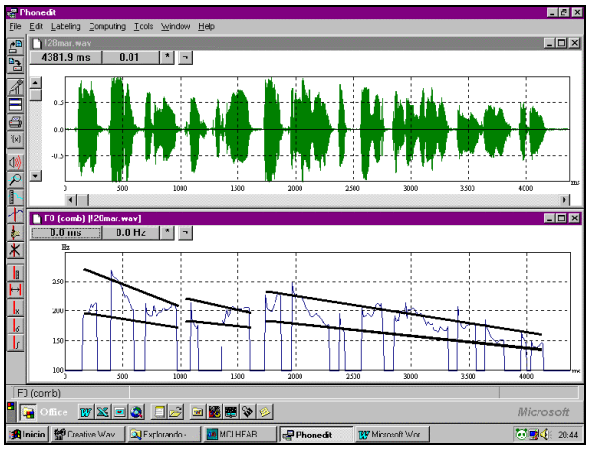

Figura 8: Curva melódica correspondiente al enunciado 'El acto de la firma transcurrió tal y como habían pactado ambas delegaciones la víspera' pronunciado por un locutor femenino. Sobre cada grupo entonativo se han dibujado
las líneas de referencia correspondientes.

El fenómeno de reajuste de F0 se recoge en este modelo en términos de 'reajuste medio' (nivel medio de reajuste entre los diferentes grupos entonativos), sin tener en cuenta, por el momento, diferentes niveles de reajuste. Este reajuste medio se obtuvo a partir del análisis de los niveles de reajuste que realizó el locutor entre los diferentes grupos entonativos.

Desarrollo del módulo de segmentación prosódica

Para la asignación de curvas de F0, pausas y duración es necesario que previamente el texto de entrada se haya segmentado en las unidades prosódicas pertinentes, dado que, como es bien sabido, estas unidades son el ámbito natural de asignación de los fenómenos suprasegmentales, especialmente la entonación. El reconocimiento de estas unidades es llevado a cabo en ACTOR® por el módulo de segmentación prosódica.

Como ocurre con el módulo de asignación de curvas de F0, el módulo de segmentación prosódica está aún en fase de desarrollo, por lo que la versión actual sólo tiene implementado parcialmente el modelo presentado a continuación.

El enfoque empleado para el desarrollo de este módulo, semejante al aplicado en la versión italiana de ACTOR® (Quazza y Gili Fivela, 1996), se basa parcialmente en el modelo de fonología prosódica de Nespor y Vogel (1982, 1983, 1986). De acuerdo con este modelo los enunciados se estructuran en una jerarquía de unidades prosódicas que son el ámbito natural de diversos fenómenos fonológicos, entre ellos

los suprasegmentales como el acento y la entonación. En este caso, se ha adoptado de la propuesta de Nespor y Vogel su principio de estructuración jerárquica de las unidades prosódicas pero no el inventario estricto de unidades propuestas. Las unidades prosódicas que se han considerado para el desarrollo de este modelo han sido las siguientes:

1. El grupo acentual (GA).
2. El grupo tónico (GT)
3. El grupo entonativo (GE)

El grupo acentual

Tal como se ha explicado anteriormente, el grupo acentual se considera el ámbito natural de asignación de los llamados patrones melódicos locales, y se define como la secuencia formada por una sílaba tónica (con acento primario) y todas las sílabas átonas (sin acento primario) que la siguen hasta la siguiente sílaba tónica. Así, una oración como la que aparece a continuación se segmentaría en grupos acentuales de la forma siguiente:

¿El dis[quete] [viejo que me] [has] [dado] [es para] [este ordena][dor]?

Sin embargo, en algunos casos, la segmentación en grupos acentuales presenta algunas particularidades. Así, cuando el primer grupo acentual va precedido de una o varias sílabas átonas, éstas se agrupan con el primer grupo acentual en un único patrón melódico, tal como se observa en el ejemplo de la figura 9.

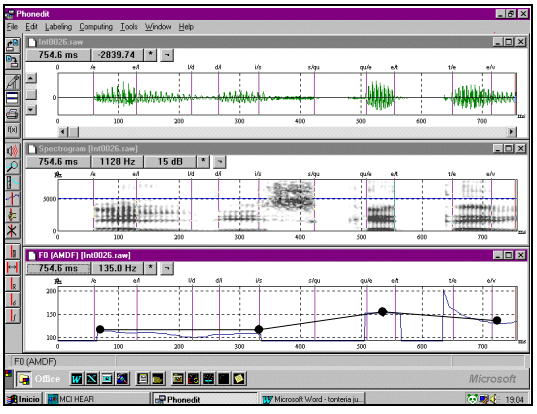

Figura 9: Curva melódica perteneciente al segmento 'El disquete' extraído del enunciado '¿El disquete viejo que me has dado es para este ordenador?', pronunciado por el locutor JR.

De acuerdo con esto, el primer grupo acentual se restructuraría de la siguiente forma:

¿**[El disquete]** [viejo que me] [has] [dado] [es para] [este ordena][dor]?

Por otro lado, en ocasiones, un patrón melódico puede extenderse a dos o más grupos acentuales. Así ocurre, por ejemplo, en ciertos casos de choque acentual (aparición de dos sílabas tónicas consecutivas), tal como se observa en la figura 10.

Figura 10: Curva melódica perteneciente al segmento 'has dado' extraído del enunciado '¿El disquete viejo que me has dado es para este ordenador?', pronunciado por el locutor JR.

Para dar cuenta de este hecho el segmentador prosódico agrupará en un solo grupo acentual los dos grupos que contienen las sílabas tónicas contiguas:

¿[El disquete] [viejo que me] **[has dado]** [es para] [este ordena][dor]?

El grupo entonativo

El grupo entonativo se define como la porción de enunciado que lleva asociado un patrón entonativo determinado. Generalmente suele coincidir con el llamado grupo fónico, es decir, el fragmento de enunciado comprendido entre dos pausas. Por ello, la segmentación de los enunciados en grupos entonativos resulta doblemente necesaria en un conversor texto-habla, tanto para la asignación de la curva melódica como para la inserción de pausas.

Como es sabido, la segmentación de un enunciado en grupos entonativos es el resultado de la interacción de diversos factores:

a) Factores fisiológicos:

La necesidad de tomar aire por parte del locutor es, evidentemente, uno de los factores que limitan la duración de los grupos entonativos. Así, se ha observado que un lector difícilmente puede pronunciar grupos fónicos de más de 30 sílabas (Garrido, 1996).

b) Factores sintácticos:

La segmentación en grupos entonativos está también condicionada por su estructura sintáctica. Se ha señalado, por ejemplo, la existencia de una serie de límites sintácticos en los que resulta incorrecto insertar una pausa (Canellada y Kuhlmann, 1987). Estos límites (artículo/sustantivo, perífrasis verbales, conjunción 'que'/oración que le sigue, etc.) implican una unión sintáctica fuerte entre los elementos. En cambio, hay una serie de límites en los que sí es posible insertar un límite de GE: final de oración, límite sujeto/predicado, final de un elemento antepuesto… (Garrido, 1996). En este segundo caso, se trata de límites sintácticos con un grado de unión débil, que suelen aparecer en los nudos superiores de la estructura sintáctica de los enunciados.

Partiendo de esta idea, los diferentes límites sintácticos se pueden clasificar, de cara a la segmentación prosódica, en tres grupos, en función de sus posibilidades de presentar un límite de GE:

a) **Límites sintácticos que exigen un límite de GE.** Todo hablante los debe realizar para que el enunciado se considere correcto. La gran mayoría de límites marcados con signos ortográficos de puntuación pertenecen a este grupo.

b) **Límites sintácticos que aceptan opcionalmente un límite de GE.** Los hablantes pueden hacerlos coincidir o no con límites de GE, sin que el enunciado deje de ser correcto. El límite sujeto/predicado es un ejemplo de ello.

c) **Límites sintácticos que no pueden coincidir nunca con un límite de GE.** Si se hacen coincidir con un límite de GE, el enunciado resulta incorrecto o, como mínimo, extraño, de manera que se dificulta gravemente la comprensión del mensaje: el límite artículo/sustantivo sería un ejemplo.

c) Factores semánticos:

En tercer lugar, parecen existir condicionantes de tipo semántico que determinan la segmentación de los enunciados en grupos entonativos, ya que parece que éstos deben poseer una cierta unidad semántica. Selkirk (1984) formalizó esta restricción en la llamada '*sense unit condition*'.

En la conversión texto-habla, dado que se parte de un texto escrito, se suele distinguir entre pausas marcadas ortográficamente y pausas no marcadas ortográficamente. Las primeras, que son las que están representadas por signos de puntuación, no presentan ningún problema para su detección. Las segundas, en cambio, dado que no hay ninguna marca explícita en el texto y existe cierta libertad por parte de los locutores para su inserción, son las que requieren un análisis contextual más complejo.

Intentar aplicar todos los factores que influyen en la inserción de pausas en un sistema de conversión texto-habla implica disponer de sistemas de análisis sintáctico y semántico actualmente no existentes. Por ello, en este caso se ha preferido una aproximación que, aunque no recoge todos los factores descritos, resulta implementable en un sistema de conversión texto-habla con relativa facilidad. Esta aproximación no tiene en cuenta los factores semánticos y recoge los factores sintácticos de una forma simplificada, por lo que la segmentación de los enunciados obtenida no siempre tiene por qué coincidir con la considerada más correcta. Se ha intentado, de todas formas, evitar la inserción de límites ' prohibidos' , y que, en todo caso, los errores se produzcan en la selección de los límites opcionales más adecuados.

En la aproximación adoptada aquí, se asume que:

a) los límites de GE obligatorios coinciden con los signos de puntuación de un texto;

b) los límites de GE ' prohibidos' corresponden a aquéllos que están en interior de un grupo tónico (GT), que se define en el apartado siguiente;

c) todos los límites sintácticos que coinciden con un límite de GT y no están marcados con signo de puntuación son candidatos a la inserción de un límite de GE opcional.

El grupo tónico

El grupo tónico (GT), aunque no es el ámbito en sí de ningún fenómeno entonativo, se considera necesario, como ya se ha dicho, para la segmentación de los grupos entonativos y, por lo tanto, para la asignación de pausas. Se define como la agrupación de una palabra tónica con todas las átonas que la preceden. Así, el enunciado anterior se segmentaría en grupos tónicos de la siguiente forma:

¿[El disquete] [viejo] [que me has] [dado] [es] [para este] [ordenador]?

La forma de incluir la información sintáctica

en el módulo de segmentación prosódica ha sido la utilización de grupos tónicos marcados con etiquetas de tipo categorial, que informan sobre la clase de palabras que se incluye en cada grupo, de una manera parecida a como se lleva a cabo en Amades *et al.* (1995). Así, por ejemplo, la etiqueta 'GP' (grupo preposicional) se aplicaría a aquellos grupos tónicos introducidos por una preposición distinta de 'de', y la etiqueta 'GPart' a aquellos grupos que contienen un participio, como se muestra en el siguiente ejemplo:

¿[El disquete]$_{GN}$ [viejo]$_{GA}$ [que me has]$_{GComp}$ [dado]$_{GPart}$ [es]$_{GV}$ [para este]$_{GP}$ [ordenador]$_{GN}$?

El tipo de etiquetas empleado está muy orientado a la tarea específica de la inserción de pausas, y no coincide con las etiquetas tradicionales en alguno de los casos.

Esta aproximación implica la existencia de un módulo de análisis categorial que está siendo desarrallado actualmente, que analice, y en la medida de lo posible desambigüe, la categoría gramatical de las palabras ortográficas.

Evaluación

Para verificar la adecuación del modelo de asignación de curvas de F0 implementado en la versión actual del sistema se ha llevado a cabo una prueba perceptiva, cuyo objetivo era evaluar hasta qué punto las curvas generadas artificialmente para las frases interrogativas se asemejan a las del locutor de referencia.

La prueba incluía 24 pares de estímulos. En cada par se presentaban dos versiones de la misma oración, la primera pronunciada por el locutor de referencia y la segunda sintetizada con el sistema ACTOR®. 30 sujetos, 10 expertos (con conocimientos especializados de prosodia) y 20 no expertos, participaron en la prueba. Tras escuchar cada par de estímulos los sujetos debían puntuar, en una escala del 0 al 4, el grado de semejanza de la curva de F0 sintetizada respecto a la curva natural.

Los resultados, que se resumen en la tabla 1, indicaron que, en líneas generales, los sujetos consideraron que las curvas sintetizadas eran relativamente semejantes a las naturales (2,58 de media, en una escala de 0 a 4). Sin embargo, se observan diferencias significativas en función del tipo de interrogativa (valores más altos en las interrogativas absolutas que en las parciales) y del tipo de locutor (valores más altos entre los no expertos que entre los expertos).

	Tipo interrogativa		
Locutor	Absoluta	Parcial	**Total**
Experto	2,6	1,85	**2,22**
No experto	3,23	2,3	**2,76**
Total	**3,02**	**2,15**	**2,58**

Tabla 1. Valores medios de las puntuaciones asignadas por los sujetos a cada una de los pares de la prueba de evaluación, presentados en función del tipo de interrogativa y del tipo de locutor

Referencias

AMADES, L. - FORNS, G. - ORTIN, I. (1995) *Pausas no marcadas ortográficamente en la lectura de textos en castellano*, Manuscrito no publicado, Bellaterra, Departament de Filologia Espanyola, Universitat Autònoma de Barcelona.

BALESTRI, M. - PACCHIOTTI, A. - QUAZZA, S. - SALZA, P. L. - SANDRI, S. (1999).- "Choose the best to modifify the least: a new generation concatenative synthesis system", *Proceedings Eurospeech 99, Budapest, August 1999*, vol. 5, pp. 2291-2294.

CANELLADA, M. J. - KUHLMANN, J. (1987).- *Pronunciación del español*, Madrid, Castalia.

GARRIDO, J. M. (1991).- *Modelización de patrones melódicos para la síntesis y el reconocimiento de habla*, Bellaterra, Departament de Filologia Espanyola, UAB.

GARRIDO, J. M. (1996).- *Modelling Spanish intonation for text-to-speech applications*, Tesis doctoral, Bellaterra, Departament de Filologia Espanyola, UAB.

NESPOR, M. - VOGEL, I. (1982).- "Prosodic Domains and External Sandhi Rules", en v. d. HULST, H. - SMITH, N. (eds.).- *The Structure of Phonological Representations*, Part I, Dordrecht: Foris; pp. 225-255.

NESPOR, M. - VOGEL, I. (1983).- "Prosodic Structure Above the Word", en CUTLER, A. - LADD, D.R. (eds.).- *Prosody: Models and Measurements*, Berlin: Springer-Verlag; pp. 123-140.

NESPOR, M. - VOGEL, I. (1986).- *Prosodic Phonology*, Dordrecht, Foris, Studies in Generative Grammar, 28.

QUAZZA, S. – GILI FIVELA, B (1996).- **"**A Prosodic Parser for an Italian Text-to-Speech System**"**, *Actas del XII Congreso de la SEPLN (Sociedad Española para el Procesamiento del Lenguaje Natural), Sevilla, España, 11-13/09/1996.*