TIC98-0423-C06

SISTEMA DE DIÁLOGO PARA HABLA ESPONTÁNEA EN UN DOMINIO SEMÁNTICO RESTRINGIDO.

REPORT DE INVESTIGACIÓN

Número de identificación	BS12BV30
Título	Manual para el etiquetado de disfluencias
Módulo	M1
Tareas	1.2, 1.4
Fecha	09 de mayo de 2000
Versión	3.0
Número de páginas	16
Autores	Luis Javier Rodríguez Fuentes (EHU) Inés Torres Barañano (EHU) Amparo Varona Fernández (EHU)
Responsable módulo	M1 UPC-T
Estado	Definitivo
Distribución	Interna
Palabras clave	Disfluencias, Anotación, XML
Persona de contacto	Luis Javier Rodríguez Fuentes tfno. 94 6012716 e-mail: luisja@we.lc.ehu.es

EVOLUCIÓN DEL DOCUMENTO

Versión	Fecha	Estado	Notas
1.0	20/03/2000	En revisión	Procedimiento de anotación de disfluencias para Basurde. Propuesta inicial, a falta de validarla (contrastarla, corregirla y/o aumentarla) frente a un conjunto pequeño de diálogos.
2.0	20/04/2000	Revisado	Procedimiento de anotación de disfluencias para Basurde. Propuesta revisada.
3.0	09/05/2000	Definitiva	Procedimiento de anotación de disfluencias para Basurde. Versión definitiva.

CONTENIDO

Manual para el etiquetado de disfluencias

1.- Introducción.

Como primer objetivo nos planteamos anotar todos aquellos fenómenos acústicos, léxicos o sintácticos que no aparecen en habla leída, y que pueden dificultar la tarea de reconocimiento del habla espontánea. El resultado de la anotación será, para cada diálogo, un fichero de texto con la transcripción ortográfica y marcas que caracterizarán los fenómenos de habla espontánea observados. Como es lógico, sólo se anotarán las intervenciones -o turnos- del usuario. Hemos supuesto que la transcripción ortográfica de partida es correcta, pero lo cierto es que adolece de ciertas deficiencias, en la puntuación y en la caracterización de determinados fenómenos léxicos y acústicos: palabras mal pronunciadas, ruidos, etc. Así pues, como segundo objetivo nos planteamos corregir o detallar -según el caso- la transcripción ortográfica.

Es obvio que la tarea de anotar requiere escuchar las señales, quizá varias veces, para que el resultado sea preciso. El criterio fundamental será, pues, generar una transcripción ortográfica y de disfluencias que refleje la acústica, aunque deberá ser posible filtrar dicha transcripción para obtener una versión "limpia", es decir, gramaticalmente correcta. Así, por ejemplo, si encontramos una palabra no terminada, anotaremos la versión ortográfica de lo que se ha oído, pero también -como atributo- la palabra completa. Los fenómenos han de anotarse exactamente donde suceden, ya sea entre dos palabras, o en medio de una palabra. Algunos fenómenos, como ruidos ambientales, pueden afectar a una o más palabras, por lo que se anotarán abarcando el segmento de texto correspondiente. En definitiva, la transcripción será un correlato ortográfico de los eventos acústicos, tan preciso como sea posible.

En lo que sigue se describen los distintos fenómenos que, después del estudio preliminar, se han considerado relevantes, más los signos de puntuación, que reciben un tratamiento bastante caótico en la transcripción actual. En cada caso se propone una marca y un inventario de atributos. Se explica en qué consiste el fenómeno, y qué variedad define cada uno de los valores de los atributos, todo ello ilustrado con ejemplos.

El formato de anotación que se plantea en primera instancia está inspirado claramente en el XML, con idea de realizar una migración en el futuro. Las marcas pueden definir sucesos puntuales, en cuyo caso se anotan como sigue:

<MARCA ATRIBUTO=VALOR/>

Cuando las marcas afectan a un intervalo de señal, un fonema, una palabra, una frase o incluso un turno, se anotarían como sigue:

<MARCA ATRIBUTO=VALOR>TEXTO MARCADO</MARCA>

De momento no se dispone de una herramienta gráfica de anotación, así que deberán ponerse las marcas directamente sobre el fichero de texto, lo cual hace que el proceso sea lento y tedioso. Para agilizarlo se utilizará un formato *simplificado*, que después se puede traducir directamente al formato *largo*. Adoptamos este

procedimiento como solución de compromiso, ya que tan sólo se trata de 227 diálogos. Cada uno de estos diálogos deberá ser procesado por dos anotadores distintos, con objeto de verificar la robustez del procedimiento. Un tercer anotador deberá resolver los casos de conflicto.

2.- Fenómenos.

2.1.- Alargamiento de sonidos.

Este fenómeno cumple la misma función que lo que en la literatura se conoce como *filled pauses*, es decir, el locutor alarga una vocal para darse tiempo mientras planea lo que va a decir a continuación. Los datos indican que ambas estrategias suceden indistintamente y no dependen del locutor, es decir, insertar tiempos entre palabras o entre frases, mediante las citadas *filled pauses*, o alargar una vocal final o intermedia de una palabra, son dos opciones igualmente probables y no dependen, quizá, más que del contexto de la frase

Anotamos los alargamientos con objeto de comprobar la eficacia de las distintas técnicas que pueden aplicarse para detectarlos, o la robustez de unos ciertos modelos duracionales. Por otro lado, encontramos una fuerte correlación entre disfluencias sintácticas y alargamiento de sonidos. Y es que tanto las *filled pauses* como los alargamientos pueden señalar una reformulación del discurso, es decir, una interrupción de la frase tal como estaba siendo formulada, a la que seguirá el comienzo de una nueva frase, o al menos una corrección dentro de la misma frase. De ahí que sea importante detectar estos fenómenos.

Normalmente es una vocal lo que se alarga y el marcado consistirá en parentizar dicha vocal entre las marcas *<a>* y **, aunque el segmento podría ser más largo, es decir, una vocal más una consonante, como veremos en un ejemplo. Otras veces, aunque con menos frecuencia, es una consonante (habitualmente l, n, s) lo que se alarga. En cualquier caso, el marcado tendría el siguiente aspecto:

<a>TEXTO

En el ejemplo que sigue, la “e” y la "l" que van en negrita son “e” y "l" alargadas, posiblemente porque el locutor estaba recordando la fecha. La anotación de este fenómeno consiste simplemente en parentizar el segmento que se alarga con las marcas *<a>* y **:

Ejemplo 2.1.1 (turno U3 del diálogo A520):

```
quisiera saber los horarios del quince de agosto
quisiera saber los horarios d<a>el</a> quince de agosto
```

Podríamos haber definido un atributo que caracterizara la duración del alargamiento, pero dar un valor a dicho atributo dependería demasiado del anotador, así que se ha optado por indicar simplemente la existencia de un alargamiento. Falta comprobar hasta qué punto es subjetiva también la percepción de un alargamiento. En el siguiente ejemplo anotamos varios alargamientos casi consecutivos.

Ejemplo 2.1.2 (turno U4 del diálogo A550):

```
bueno, pues de Teruel a Zaragoza para el día ocho
bueno, pues d<a>e</a> Teruel a Zaragoz<a>a</a> para el dí<a>a</a> ocho
```

2.2.- Ruidos.

Distinguimos, desde el punto de vista de la duración, dos tipos de ruidos: puntuales y prolongados. Los primeros afectan a una parte pequeña de la señal, como mucho a una palabra, o incluso a ninguna si suceden en intervalos de silencio. Los últimos afectan a un tramo grande de la señal y pueden afectar a varias palabras o incluso a un turno completo. En la práctica, un usuario podría llamar desde la calle o desde cualquier otro entorno con ruido de fondo fuerte, o utilizar una línea telefónica defectuosa, en cuyo caso el ruido prolongado afectaría a todos los turnos. También puede suceder que mientras se desarrolla la consulta se produzca un ruido prolongado o una secuencia de ruidos puntuales que afecten a un solo turno, por ejemplo un bocinazo, uno o varios timbrazos, una televisión o una radio, etc. En las señales examinadas, tomadas del Mago de Oz, no se observa ningún caso de ruidos prolongados, debido a que los usuarios llamaban desde entornos relativamente silenciosos. Se observan, eso sí, secuencias de ruidos puntuales que pueden abarcar varias palabras y que se anotarían como si se tratase de ruidos prolongados (véase ejemplo 2.2.3).

Desde el punto de vista del origen del ruido, distinguimos también dos tipos: los producidos por el locutor (aspiraciones, chasquidos de labios, etc.), que suelen ser puntuales, y los procedentes de fuentes externas. Como se ha dicho, salvo pequeños tics del auricular, y en algún caso voces o débiles ruidos de fondo, no se han observado fuentes externas de ruido significativas en las grabaciones del Mago de Oz. Por ello, una gran parte de los ruidos que aparecen en dichas señales corresponden a ruidos producidos por el usuario, normalmente al tomar aire para comenzar una nueva frase, al respirar después o durante un turno, o al soplar sobre el auricular. Los sonidos del habla que no forman palabras -nos referimos a las *filled pauses*- no entran en la categoría de ruidos, ya que su distribución no es aleatoria, sino que responde a un patrón semántico que hemos descrito en el apartado anterior.

Para describir esta fenomenología, se propone una sola marca de ruido: *n*, con dos atributos: *source* y *type.* Así pues, la misma marca sirve para ruidos puntuales y ruidos prolongados, producidos por el usuario o producidos por el mundo exterior. A los ruidos producidos por el usuario les corresponde una marca como la que sigue:

<n source=speaker type=TIPO/>	ruidos puntuales
<n source=speaker type=TIPO>TEXTO</n>	ruidos prolongados

TIPO puede tomar los valores *air* (para cualquier soplido o aspiración), *lips* (para los tics producidos por los labios del locutor) y *cough* (para toses y carraspeos). Este conjunto de valores podría ser ampliado si se observaran otros casos.

En cuanto a los ruidos externos, aunque podrían definirse varios tipos distintos, de momento se catalogarán todos bajo una sola denominación genérica:

<n source=world type=generic/>	ruidos puntuales
<n source=world type=generic>TEXTO</n>	ruidos prolongados

Ejemplo 2.2.1 (turno U2 del diálogo A620):

```
[spk] sí [int] .
<n source=speaker type=air/> sí <n source=world type=generic/> .
```

Ejemplo 2.2.2 (turno U6 del diálogo A620):

```
[spk] precio [int] .
<n source=speaker type=lips/> precio <n source=world type=generic/> .
```

Ejemplo 2.2.3 (turno U17 del diálogo C240):

```
no eso es todo muchas gracias .
no . eso es todo . muchas <n source=world type=generic> gracias </n> .
```

2.3.- Pausas de silencio.

Las pausas de silencio tienen una gran importancia en el habla espontánea, ya que señalan puntos de ruptura entre dos frases o entre dos unidades semánticas. De hecho, pueden ser utilizadas, por ejemplo, para segmentar un turno en dos o más unidades. A veces tales pausas coinciden con lo que entenderíamos como un “punto y seguido” o un “punto y aparte” (dependiendo de la duración de la pausa), y las unidades que definen podrían entenderse como frases. Una segmentación como ésta facilitaría enormemente la tarea del reconocedor. Sin embargo, con mucha frecuencia se insertan pausas de silencio en mitad de una frase gramatical, bien separando la frase en unidades con una cierta coherencia semántica interna, bien rompiendo su estructura gramatical, abandonando la frase inicial y comenzando una frase nueva.

En cualquier caso, se ha estimado importante marcar todas las pausas cuya duración exceda de lo razonable en discurso continuo. No se marcarán como pausas los silencios que preceden y siguen a cada turno del usuario. En esos tramos iniciales y finales sólo se anotarán los ruidos que sucedan muy cerca de la señal.

Actualmente la transcripción ortográfica no marca las pausas de silencio, y por tanto será necesario marcar todas ellas a partir de la señal. El objetivo sería, por un lado, comprobar la eficacia del reconocedor en la detección de tales pausas, y por otro, facilitar el análisis y posterior modelado de disfluencias sintácticas, ya que -al igual que las *filled pauses* o los alargamientos- las pausas de silencio pueden señalar el comienzo de una reformulación.

Las pausas de silencio se marcarán con la secuencia <p/>.

Ejemplo 2.3.1 (turno U3 del diálogo A730):

```
podría decirme el tren que sale más próximo a las <p/> quince horas
```

Ejemplo 2.3.2 (turno U4 del diálogo A750):

```
no, <p/> me gustaría viajar <p/> el siete de agosto del dos mil .
```

2.4.- Pausas habladas.

Nos referimos a lo que hasta ahora hemos venido llamando *filled pauses*, es decir, pausas “rellenas” de señal, normalmente una vocal o una nasalización, que el locutor utiliza para darse tiempo de planear lo que va a decir a continuación. Rellenar la pausa responde al propósito de mostrar al interlocutor su intención de seguir hablando. Si el locutor simplemente interrumpiera su discurso con una pausa de silencio, el interlocutor podría verlo como una oportunidad de intervenir. Así pues, pausas habladas y pausas de

silencio se producen por una misma necesidad de planear el discurso, y es quizá la percepción de una mayor o menor urgencia lo que determina el uso de una u otra estrategia.

Una pausa hablada puede realizarse fonéticamente de distintas formas y puede tener distintas duraciones. No trataremos de caracterizar la duración, ya que cada anotador puede percibirla de manera distinta. Sin embargo, sí definiremos un atributo que permita identificar fonéticamente cada pausa hablada:

<f type=TIPO/>

TIPO puede tomar los valores *a*, *e* y *m*, correspondientes a las dos realizaciones vocales más frecuentes y a la realización en forma de nasalización. Cuando la pausa hablada no encaja con ninguno de los casos anteriores, o cuando se tengan dudas sobre la identidad del sonido, se etiquetará como *trash* (basura).

Ejemplo 2.4.1 (turno U1 del diálogo A530):

```
no, no, <f type=m> sí , el sábado treinta de octubre , <f type=m> hay un tren
```

Ejemplo 2.4.2 (turno U0 del diálogo A640):

```
quisiera saber <f type=e> el precio de un viaje
```

Ejemplo 2.4.3 (turno U9 del diálogo A740):

```
<f type=a> no , ya está todo
```

Ejemplo 2.4.4 (turno U1 del diálogo B220):

```
<f type=trash> llegar a Tarragona sobre las nueve
```

2.5.- Disfluencias léxicas.

En esta categoría se incluyen tres tipos de fenómenos: palabras no terminadas, palabras *mal* pronunciadas y sonidos guturales de afirmación o de negación. Se han reunido bajo la denominación de disfluencias léxicas debido a que en todos los casos se produce una realización acústica incompleta, marginal o poco ortodoxa de una palabra. Anotamos estos fenómenos para comprobar la eficacia del reconocedor, por un lado; y por otro, para tenerlos en cuenta en el análisis de disfluencias sintácticas, ya que pueden producir o pueden señalar una reformulación. También pueden ayudar a realizar un estudio sobre determinadas variedades de pronunciación habituales en habla espontánea, y en último extremo, a modelar dichas variedades.

Para representar la fenomenología descrita se utilizará una sola marca *<l>*, y serán necesarios dos atributos: *type* y *word*. El atributo *type* podrá tomar los valores *unfinished, mispronounced* y *gutural.* Este último valor se ha incluido para dar cobertura a este tipo de fenómenos pero no aparecerá en los diálogos del Mago de Oz, debido a la baja interactividad de los mismos. El atributo *word* se utilizará para indicar la versión ortográfica correcta de la palabra. Las marcas *<l>* y *</l>* acotarán el correlato ortográfico de la parte acústica, salvo cuando se trata de afirmaciones o negaciones guturales, en los que no puede darse un correlato ortográfico preciso. En estos casos, el valor del atributo *word* será *sí* o *no*, según se trate de afirmaciones o negaciones.

Ejemplo 2.5.1 (turno U6 del diálogo B020):

```
en <l type=unfinished word=intercity> interci </l> bueno ¿ cuál es el más rápido ?
```

Ejemplo 2.5.2 (turno U6 del diálogo B020):

```
<l type=mispronounced word=bueno> ueno </l> ¿ cuál es el más rápido ?
```

2.6.- Disfluencias sintácticas.

Esta es la categoría más compleja y también la más difícil de anotar. Se incluyen en ella repeticiones, reformulaciones con borrado, inserción o sustitución de elementos, y frases abandonadas. Repeticiones y reformulaciones se funden en una sola categoría porque pueden ser descritas, como veremos, mediante un único formalismo. Desde el punto de vista de su amplitud, las reformulaciones pueden afectar a toda la frase desde su inicio, o pueden afectar sólo a una parte de la misma. En el primer caso suele hablarse de inicios fallidos o reinicios, y en el segundo caso se habla propiamente de reformulaciones. En cuanto a las frases abandonadas, o más bien segmentos de frases abandonados, podrían tratarse como reformulaciones con sustitución, pero tras unas primeras pruebas de etiquetado, se ha optado por definir una categoría específica con objeto de simplificar la estructura sintáctica de muchas frases.

2.6.1.- Repeticiones y reformulaciones.

Las repeticiones responden -como las pausas habladas- a la necesidad de planear el discurso, es decir, el locutor repite un segmento mientras piensa lo que va a decir a continuación. En lugar de pensar en silencio, habla para evitar la apropiación del turno por parte del interlocutor. Desde el punto de vista de éste, una repetición debe interpretarse como una señal de espera. En cuanto a las reformulaciones, constituyen el único mecanismo de que dispone el locutor para ir corrigiendo la construcción de frases. A diferencia del lenguaje escrito, en el lenguaje hablado es imposible borrar lo que ya se ha dicho. Sólo hablaremos de reformulaciones en aquellos casos en los que se recompone la frase repitiendo parte del segmento inicial y borrando, insertando o sustituyendo elementos. Cuando el segmento inicial es abandonado y la estructura gramatical se recompone sustituyéndolo por otro completamente distinto, hablaremos de segmentos de frase abandonados.

Anotamos este tipo de fenómenos para analizar su correlación con determinadas pistas acústicas: entonación, presencia de pausas habladas o alargamiento de vocales, etc. que nos van a permitir plantear después un modelo específico para detectar este tipo de disfluencias. Una vez detectadas, podremos integrar las reformulaciones en el análisis sintáctico, lo cual debería facilitar la eficacia del módulo de comprensión, ya que, por ejemplo, podríamos eliminar los inicios fallidos, conservando únicamente los segmentos introducidos como corrección.

Para detectar muchas de estas disfluencias es preciso escuchar la señal, ya que sólo a partir de la transcripción ortográfica sería imposible decir si se ha producido una reformulación.

Ejemplo 2.6.1.1 (turno U5 del diálogo B020):

```
el precio el precio del viaje
el precio (PS) el precio del viaje
```

En el ejemplo 2.6.1.1 parece que se ha producido una repetición de la secuencia *el precio*, pero lo cierto es que existe una pausa de silencio (PS) entre las dos repeticiones

de la secuencia, con lo cual obtenemos dos frases independientes. En todo caso podríamos ver la segunda frase como una extensión o matización de la primera. Este fenómeno se conoce como *right dislocation* y no está claro si debe considerarse o no como reformulación. Nuestro criterio aquí es no considerarlo como tal.

Ejemplo 2.6.1.2 (turno U12 del diálogo C310):

```
llegar a Madrid antes como muy tarde a las siete
llegar a Madrid (PS) antes (PH) como muy tarde a las siete
```

En el ejemplo 2.6.1.2 nos encontramos con un caso similar. La transcripción ortográfica parece indicar que se trata de *llegar a Madrid antes*, y que este objetivo se concreta en llegar *como muy tarde a las siete*. Aparentemente no hay reformulación, sólo falta una coma. Sin embrago, una pausa de silencio separa las palabras *Madrid* y *antes*, y una pausa hablada (PH) separa la palabra *antes* de la secuencia *como muy tarde*. Si escuchamos la señal, nos damos cuenta de que en realidad se trata de llegar a Madrid, inicialmente *antes de las siete* y luego (el locutor se lo piensa) *como muy tarde a las siete*, es decir, se produce una sustitución de la forma *antes de* (que no llega a formularse del todo) por la forma *como muy tarde a*. Por tanto, anotaríamos este caso como una reformulación.

Las disfluencias sintácticas pueden representarse mediante una estructura muy simple que consta de tres elementos: *reparandum*, *señal* y *corrección*. La parte que denominamos *reparandum* corresponde al segmento identificado como erróneo, que puede ser el comienzo de una frase, una parte de un sintagma o simplemente una palabra. Lo que llamamos *corrección* es el segmento que sustituye al segmento erróneo. En cuanto a la *señal*, no siempre está presente. En ocasiones, la *corrección* sigue al *reparandum* sin solución de continuidad, pero es más habitual que el locutor utilice una pausa hablada o quizá un marcador de discurso específico, como “es decir”, “quiero decir”, “perdón”, “esto”, “sí”, “no”, etc. para señalar el punto en que ha decidido realizar una corrección. De hecho, la *señal* puede estar formada por una secuencia de estos elementos, por ejemplo una pausa de silencio seguida de una pausa hablada y de un marcador de discurso.

En cuanto a cómo identificar cada una de los elementos, se aplicarán simultáneamente dos criterios:

a) tomar los segmentos *reparandum* y *corrección* mínimos alrededor de la *señal*, de manera que la frase que resulte de sustituir *reparandum* por *corrección*, borrando la *señal* y dejando inalterado el resto, sea gramaticalmente correcta,

b) los segmentos *reparandum* y *corrección* han de mantener una cierta coherencia gramatical, es decir, al menos en parte deben encajar dentro de la misma categoría gramatical.

Esto lo vemos mejor con algunos ejemplos, donde hemos marcado los segmentos que identificamos como *reparandum* (RM), *corrección* (RR) y *señal*, cuando ésta aparece:

Ejemplo 2.6.1.3 (turno U3 del diálogo A520):

```
trenes talgo  de  que van de  Barcelona
            | RM |     RR   |
```

Ejemplo 2.6.1.4 (turno U5 del diálogo A630):

```
pero  estas ho   estas horas  de llegada
     |   RM    |     RR     |
```

Ejemplo 2.6.1.5 (turno U0 del diálogo A720):

```
horarios  a Ci   desde Barcelona a Ciudad_Real
        |  RM  |            RR              |
```

Ejemplo 2.6.1.6 (turno U0 del diálogo A750):

```
resido  en Granada  perdón   en Málaga
       |    RM    |  Señal |    RR   |
```

Ejemplo 2.6.1.7 (turno U0 del diálogo A730):

```
 quería    lla   llamaba    para pedir información
|  RM   |       RR       |
        |  RM  |   RR    |
```

Ejemplo 2.6.1.8 (turno U5 del diálogo A630):

```
¿qué son, del sábado o del domingo?  del   su    si    supongo  que del  sábado
                                    | RM |                RR           |
                                         | RM | Señal |  RR   |
```

Cada uno de los ejemplos anteriores viene a ilustrar alguno de los casos que podemos encontrar al anotar disfluencias sintácticas, aunque no todos. Los ejemplos 2.6.1.3, 2.6.1.4 y 2.6.1.5 tienen en común que el locutor no utiliza una *señal* explícita para separar el *reparandum* de la *corrección*. El ejemplo 2.6.1.4 muestra un caso de repetición y los ejemplos 2.6.1.3 y 2.6.1.5 sendos casos de reformulación con inserción. El ejemplo 2.6.6 muestra un caso de sustitución con *señal* explícita, en concreto el marcador de discurso “perdón”. Finalmente los ejemplos 2.6.1.7 y 2.6.1.8 muestran dos casos de anidamiento de reformulaciones. En el ejemplo 2.6.1.7 la palabra “quería” es sustituida por la palabra “llamaba”; luego vemos que la palabra “llamaba” aparece repetida, si bien incompleta en el primer caso. En el ejemplo 2.6.1.8 el locutor realiza una pregunta y se responde a sí mismo, en principio comienza la respuesta con la palabra “del” (parece que para decir “del sábado”), pero inmediatamente, sin solución de continuidad, reformula la respuesta, sustituyendo la palabra “del” por “supongo que del” (reformulación con inserción). Lo que pasa es que dentro de la *corrección* se repite la palabra “supongo”, con una *señal* (el marcador de discurso “sí”) que marca la separación entre la primera (incompleta) y segunda de las apariciones de la palabra “supongo”.

Para anotar convenientemente este tipo de fenómenos, se definen cuatro marcas distintas, una marca primaria *<r>* que acota la reformulación completa, y tres marcas secundarias, *<m>*, *<s>* y *<c>* que distinguen cada uno de los elementos descritos anteriormente (*reparandum*, *señal* y *corrección*). La marca primaria *<r>* dispondrá de un atributo *type* con el que podremos especificar el tipo de disfluencia sintáctica. Los valores posibles del atributo *type* son: *repetition, deletion, insertion* y *substitution.* Veamos cómo quedarían los ejemplos anteriores:

Ejemplo 2.6.1.3 (turno U3 del diálogo A520):

```
trenes talgo
<r type=insertion>
    <m>de</m>
    <c>que van de</c>
</r>
Barcelona
```

Ejemplo 2.6.1.4 (turno U5 del diálogo A630):

```
pero
<r type=repetition>
     <m>estas ho</m>
     <c>estas horas</c>
</r>
de llegada
```

Ejemplo 2.6.1.5 (turno U0 del diálogo A720):

```
horarios
<r type=insertion>
     <m>a Ci</m>
     <c>desde Barcelona a Ciudad_Real</c>
</r>
```

Ejemplo 2.6.1.6 (turno U0 del diálogo A750):

```
resido
<r type=substitution>
     <m>en Granada</m>
     <s>perdón</s>
     <c>en Málaga</c>
</r>
```

Ejemplo 2.6.1.7 (turno U0 del diálogo A730):

```
<r type=substitution>
     <m>quería</m>
     <c><r type=repetition>
            <m>lla</m>
            <c>llamaba</c>
        </r>
     </c>
</r>
```

Ejemplo 2.6.1.8 (turno U5 del diálogo A630):

```
<r type=insertion>
     <m>del</m>
     <c><r type=repetition>
            <m>su</m>
            <s>si</s>
            <c>supongo</c>
        </r>
        que del
     </c>
</r>
sábado
```

2.6.2.- Segmentos de frase abandonados.

Se ha definido una etiqueta específica para aquellos casos en que una parte de una frase es abandonada y su estructura gramatical se recompone después de un modo totalmente distinto. Esto sucede con frecuencia al principio de las frases, en lo que hemos denominado *inicios fallidos*. Son casos que podrían interpretarse como reformulaciones con sustitución de elementos. Sin embargo, de hacerlo así resultarían etiquetados muy complicados, en los que *reparandum* y *corrección* no cumplirían la misma función gramatical, sino que simplemente serían frases distintas. Etiquetando sólo el segmento de frase abandonado se logra simplificar el proceso de anotación, puesto que ya no es necesario identificar la parte de la frase que actua como *corrección*.

Así pues, los segmentos de frase abandonados se etiquetarán como inserciones erróneas, es decir, no se establecerán relaciones con otras partes de la frase o con otras frases. Se etiquetarán entre las marcas ** y ** de forma que la frase que resulte al eliminar el segmento abandonado sea gramaticalmente correcta:

ANTES SEGMENTO ABANDONADO DESPUÉS

ANTES DESPUÉS frase gramaticalmente correcta

A continuación se muestran dos ejemplos que se han marcado primero como reformulaciones con sustitución y después como segmentos de frase abandonados. En ambos casos las sustituciones propuestas no parecen muy lógicas y se puede observar cómo se simplifica el etiquetado:

Ejemplo 2.6.2.1 (turno U0 del diálogo C040):

```
me gustaría
<r type=susbstitution>
    <m>que</m>
    <c>viajar en coche cama</c>
</r>

me gustaría <b>que</b> viajar en coche cama
```

Ejemplo 2.6.2.2 (turno U9 del diálogo C040):

```
sí . a ver .
<r type=substitution>
    <m>para</m>
    <c>yo quiero</c>
</r>
estar el jueves siete de octubre

sí . a ver . <b>para</b> yo quiero estar el jueves siete de octubre
```

2.7.- Marcadores de discurso.

Ciertas palabras de semántica imprecisa cumplen, sin embargo, una función importante dentro de los diálogos, marcando la apropiación o la cesión del turno de intervención, confirmando la comprensión de un enunciado, marcando el inicio de una corrección, pidiendo una respuesta, etc. Se denominan marcadores de discurso. Aunque en muchos casos no pueden catalogarse propiamente como disfluencias, se trata de fenómenos específicos del habla espontánea, y como tales hemos creído importante anotarlos. De este modo podremos reconocerlos y tenerlos en cuenta en el modelo de lenguaje.

Para describir estos fenómenos se define una marca *<d>* y un atributo *type*, con el que podremos especificar la función que cumple cada marcador de discurso. Se proponen los siguientes valores:

VALOR	**DESCRIPCIÓN**	**EJEMPLOS**
open	apertura	“hola”, “buenos días”, “buenas”
close	cierre	“adiós”, “gracias”, “buenos días”, “hasta luego”
accept	aceptación o acuerdo	“sí”, “bueno”, “vale”, “ya”, “de acuerdo”
reject	rechazo o desacuerdo	“no”, “para nada”, “nunca”
explain	explicación o corrección	“es decir”, “o sea”, “es que”, “perdón”, “esto”
request	solicitud de intervención	“por favor”, “¿no?”, “¿sí?”, “¿eh?”, “¿vale?”
fill	relleno	“pues”, “entonces”, “mire”, “verá”
exclaim	exclamación	“¡ah!”, “¡oh!”, “¡vaya!”, “¡ay!”

Las palabras que en alguna circunstancia actúan como marcadores de discurso no siempre lo hacen, es decir, podemos encontrarlas en el texto cumpliendo otras funciones. Por otro lado, ciertas palabras pueden cumplir distintas funciones como marcadores de discurso. Así, la palabra “bueno”, que puede actuar como adjetivo, cuando lo hace como marcador de discurso puede significar aceptación o puede introducir una corrección; la secuencia “buenos días” no la encontramos más que como marcador de discurso, pero puede servir como fórmula de apertura o como fórmula de cierre, etc.

Las palabras que actúan como marcadores de discurso son fáciles de identificar: son siempre las mismas y suceden en situaciones del diálogo muy concretas. La decisión dependerá de muchos factores, como la entonación, el contexto de la frase, la situación del diálogo, etc. Pero en la mayor parte de los casos la regla más fiable consiste en eliminar de la frase la palabra sospechosa: si la frase mantiene el mismo significado, podemos concluir que la palabra borrada era semánticamente innecesaria y cumplía, por tanto, una cierta función como marcador de discurso.

Ejemplo 2.7.1 (turno U0 del diálogo A720):

```
<d type=open>hola</d> , <d type=open>buenos días</d> , <d type=open>mire</d>
```

Ejemplo 2.7.2 (turno U4 del diálogo A620):

```
estación de salida , <d type=request>por favor</d>
```

Ejemplo 2.7.3 (turno U3 del diálogo A630):

```
<d type=accept>bueno</d> , y ¿ a qué hora llegan estos trenes a Vigo ?
```

Ejemplo 2.7.4 (turno U8 del diálogo A720):

```
no , ya está bien , <d type=close>gracias</d> , <d type=close>buenos días</d> .
```

Ejemplo 2.7.5 (turno U0 del diálogo A750):

```
<d type=open>hola</d> , resido en Granada , <d type=explain>perdón</d> , en Málaga
```

Ejemplo 2.7.6 (turno U0 del diálogo A530):

```
me gustaría saber <d type=fill>pues</d> a qué hora
```

Ejemplo 2.7.7 (turno U2 del diálogo A540):

```
que esté cerca entre ellas , <d type=explain>o sea</d> , que esté cerca la ciudad
```

3.- Signos de puntuación.

Como ya se ha dicho, la transcripción ortográfica de la que partimos no se ajusta exactamente a la señal, es decir, no es correlato exacto de lo que se ha dicho. Algunas veces faltan detalles, por ejemplo cuando se utiliza el asterisco (*) para indicar cualquier anomalía de pronunciación, repetición de palabras, etc. Otras veces están ausentes de la transcripción palabras que se han dicho, o se incluyen palabras que no aparecen en la señal. Por tanto, será necesario repasar la transcripción ortográfica actual para corregirla o aumentarla, según el caso. En concreto, los asteriscos serán eliminados y sustituidos por el correlato ortográfico de lo que el anotador oiga.

En lo que respecta a la puntuación, distinguimos dos criterios, uno para acentos, comas y puntos, y otro para signos de admiración e interrogación. No se marcarán punto_y_comas (;). Los acentos, obviamente, irán colocados donde corresponda en el sentido ortográfico. Comas y puntos se colocarán no como correlato de la señal, es decir,

no como reflejo de pausas: las pausas de silencio se marcan explícitamente. Tampoco puntuaremos como si las intervenciones de los usuarios fueran texto escrito normal. Aplicaremos dos criterios:

a) minimizar el uso de comas, y

b) dividir cada turno cuanto sea posible en frases independientes con sentido.

Así pues, utilizaremos el punto para dividir un texto en dos segmentos siempre que resulten frases con sentido. En concreto, marcadores de afirmación y negación ("sí", "bien", "no", etc.) y frases similares constituidas por otros marcadores de discurso ("buenos días", "muchas gracias", etc.) irán separados siempre mediante puntos, salvo que vayan seguidos de una frase subordinada del tipo “sí, pero...” o “no, porque...”, en cuyo caso irán separados por comas. Precisamente, en cuanto a las comas, se utilizarán principalmente con dos objetivos:

a) parentizar o separar segmentos que el usuario inserta en la frase para aclarar o matizar contenidos (véase ejemplo 3.1), y
b) marcar el comienzo de una frase subordinada o coordinada (véase ejemplo 3.2).

Recuérdese, por otra parte, que se ha convenido separar las comas y puntos de las palabras adyacentes mediante un espacio en blanco.

En cuanto a los signos de admiración e interrogación, acotarán fragmentos de texto cuya entonación los identifique claramente como exclamaciones o como preguntas, respectivamente. Es decir, serán correlato ortográfico de la entonación. Así, aunque la construcción sintáctica no indique la existencia de una pregunta, si la entonación es de pregunta, anotaremos ese texto entre signos de interrogación. Inversamente, aunque la construcción de la frase sugiera el uso de interrogantes, si no existe entonación no los colocaremos. No se colocarán ni coma ni punto delante de una interrogación de apertura ni tampoco detrás de una interrogación de cierre. Es decir, los símbolos de interrogación actuarán como puntos o comas, según convenga, separando frases o bloques semánticos.

A continuación se presentan varios ejemplos en los que se muestra por un lado la transcripción actual, y por otro la transcripción corregida según los criterios mencionados. Para clarificar el proceso, no se han añadido marcas de disfluencias.

Ejemplo 3.1 (turno U0 del diálogo A530):

```
sí . mire es que prontamente *de hecho el sábado me voy *a Port_Aventura
sí . mire . es que prontamente , de de hecho el sábado , me voy a Port_Aventura
```

Ejemplo 3.2 (turno U5 del diálogo A720):

```
sí . no recuerdo que tipo de tren era me ha dicho que no intercity pero *no *me
acuerdo , me lo podría decir por favor .

sí . no recuerdo qué tipo de tren era . me ha dicho que no había intercity , pero
no me no me acuerdo ¿ me lo podría decir , por favor ?
```

Ejemplo 3.3 (turno U1 del diálogo A530):

```
no no [fil] si el sábado treinta de octubre , [fil] hay un tren
no . no . [fil] sí . el sábado treinta de octubre [fil] hay un tren
```

Ejemplo 3.4 (turno U5 del diálogo A630):

```
pero estas o esas horas de llegada
pero estas ho estas horas de llegada
```

Ejemplo 3.5 (turno U6 del diálogo A640):

```
¿ quisiera saber el horario ?
quisiera saber el horario
```

Ejemplo 3.6 (turno U1 del diálogo A720):

```
sí . me lo podría mirar *en intercity .
si me lo podría mirar en intercity .
```

Ejemplo 3.7 (turno U0 del diálogo A730):

```
[spk] quería *llamaba para pedir información
[spk] quería lla llamaba para pedir información
```

4.- Formato simplificado y procedimiento de anotación.

Después de presentar la gran variedad de fenómenos que han de ser anotados, así como las etiquetas, atributos y valores definidos al efecto, no parece conveniente anotar directamente en el formato tipo XML definido en primera instancia, dado que no disponemos de una herramienta gráfica adecuada. En lugar de eso, se utilizará un formato simplificado que acorte y facilite el proceso de anotación. También se escribirá un programa-filtro que transforme el fichero en formato simplificado a formato tipo XML. Finalmente, también parece conveniente detallar los pasos que el anotador deberá seguir para colocar las marcas sobre la transcripción ortográfica.

4.1.- Formato de anotación simplificado.

En adelante nos referiremos al formato tipo XML que hemos definido para anotar fenómenos específicos del habla espontánea como *Disfluent Speech Markup Language (DSML)*. En cuanto al formato simplificado, basándonos en la estructura del DSML, trataremos de hacerlo más ergonómico. En primer lugar, los segmentos a marcar se acotarán mediante paréntesis de apertura y cierre. Tras el paréntesis de apertura colocaremos la marca DSML correspondiente, y tras ella, si existen atributos, un carácter que identificará de forma inequívoca los valores de los atributos:

(MARCA IDENTIFICADOR SEGMENTO)

La única excepción a esta regla se dará con las disfluencias léxicas. En este caso, tras la marca de disfluencia léxica (*l*) y una letra (*u, m, g*) que identifica el valor del atributo *type*, irá −separado por un espacio en blanco− el valor del atributo *word*, esto es, la versión ortográfica correcta de la palabra, y tras ésta −de nuevo tras un espacio en blanco− la versión ortográfica de lo que se ha oído:

(lu VERSIÓN CORRECTA SEGMENTO)

(lm VERSIÓN CORRECTA SEGMENTO)

(lg sí)

(lg no)

En la Tabla I se muestran las equivalencias para todas las marcas que llevan atributos:

Tabla I. Equivalencias entre DSML y el formato de anotación simplificado.

Marca	Atributo: Valor	Atributo: Valor	Letra
n	source: speaker	type: lips	l
		type: air	a
		type: cough	t
	source: world	type: generic	w
f		type: a	a
		type: e	e
		type: m	m
		type: trash	b
l		type: mispronounced	m
		type: unfinished	u
		type: gutural	g
r		type: repetition	r
		type: deletion	d
		type: insertion	i
		type: substitution	s
d		type: open	o
		type: close	c
		type: accept	a
		type: reject	r
		type: explain	e
		type: request	q
		type: fill	f
		type: exclaim	x

Ejemplo 4.1.1 (turno U6 del diálogo B020):

```
en (lu intercity interci) (lm bueno ueno) ¿ cuál es el más rápido ?
```

Ejemplo 4.1.2 (turno U4 del diálogo A550):

```
bueno . (df pues) d(a e) Teruel a Zaragoz(a a) para el dí(a a) ocho
```

Ejemplo 4.1.3 (turno U6 del diálogo A620):

```
(nl) precio (nw) .
```

Ejemplo 4.1.4 (turno U17 del diálogo C240):

```
no . eso es todo . muchas (nw gracias) .
```

Ejemplo 4.1.5 (turno U4 del diálogo A750):

```
no . (p) me gustaría viajar (p) el siete de agosto del dos mil .
```

Ejemplo 4.1.6 (turno U1 del diálogo A530):

```
no . no . (fm) sí . el sábado treinta de octubre (fm) hay un tren
```

Ejemplo 4.1.7 (turno U0 del diálogo A750):

```
resido (rs(m en Granada)(s perdón)(c en Málaga))
```

Ejemplo 4.1.8 (turno U5 del diálogo A630):

```
(ri(m del)(c(rr(m su)(s si)(c supongo)) que del)) sábado
```

Ejemplo 4.1.9 (turno U0 del diálogo A720):

```
(do hola) . (do buenos días) . (do mire) .
```

Ejemplo 4.1.10 (turno U4 del diálogo A620):

```
estación de salida , (dq por favor) .
```

Ejemplo 4.1.11 (turno U3 del diálogo A630):

```
(da bueno) ¿ y a qué hora llegan estos trenes a Vigo ?
```

Ejemplo 4.1.12 (turno U0 del diálogo A530):

```
me gustaría saber (df pues) a qué hora
```

Ejemplo 4.1.13 (turno U0 del diálogo C040):

```
me gustaría (b que) viajar en coche cama
```

4.2.- Procedimiento de anotación.

En el proceso de anotación de disfluencias se distinguirán 5 niveles: ortográfico, acústico, léxico, pragmático y sintáctico, que serán anotados en ese orden preciso.

El ***nivel ortográfico*** recoge las consideraciones hechas en el apartado 3, así que el primer paso consistirá en corregir la transcripción ortográfica de partida, con el criterio de que exista máxima correspondencia entre la señal y su transcripción. En particular, los asteriscos -que indican una repetición o una anomalía en la pronunciación- serán eliminados y sustituidos por una representación más fiel de la señal. Recuérdese además la importancia de ubicar adecuadamente puntos, comas, signos de admiración y signos de interrogación.

En el ***nivel acústico*** se incluyen los fenómenos descritos en los apartados 2.2, 2.3, 2.4 y 2.1, en este orden. En primer lugar se localizarán todos los sonidos extralinguísticos -lo que denominamos ruidos-, etiquetándolos de acuerdo a su procedencia y características. Nótese que ya existen en las transcripciones actuales marcas de ruidos que será necesario contrastar con la señal. Es posible que algunas de esas marcas estén mal colocadas, y por supuesto, será necesario añadir algunas otras. A continuación se localizarán las pausas de silencio, después las pausas habladas y finalmente los alargamientos de vocales. Las pausas habladas también han sido marcadas en la transcripción original, pero no de forma muy precisa. Al igual que los ruidos, su presencia y ubicación habrán de ser contrastadas con la señal.

A ***nivel léxico*** se anotarán los fenómenos descritos en el apartado 2.5, es decir, palabras no terminadas y palabras mal pronunciadas, así como, si las hubiera, afirmaciones y negaciones guturales.

Hemos denominado ***nivel pragmático*** al conjunto formado por los marcadores de discurso, ya que no son fenómenos léxicos, ni sintácticos, ni semánticos. Su existencia se debe a la necesidad de expresar determinados contenidos metalingüísticos: gestión de turnos, protocolos de apertura y de cierre, etc. No debería ser difícil identificar las palabras o expresiones que funcionan como marcadores de discurso, ya que suelen ser siempre las mismas y suceden en situaciones del diálogo muy concretas. Como primera aproximación se considerarán únicamente las subcategorías descritas en el apartado 2.7, aunque podrían aparecer marcadores con otras funciones.

Finalmente, por tratarse del nivel de anotación más complejo, se anotarán las ***disfluencias sintácticas***. Repeticiones y reformulaciones han sido descritas en el apartado 2.6.1 bajo un mismo formalismo. Una vez entendida la relación entre los tres elementos de una reformulación: *reparandum*, *señal* y *corrección*, la mayor dificultad puede encontrarse, por un lado, en identificar los segmentos que corresponden a cada elemento, y por otro, en definir con exactitud los niveles de anidamiento. Son dos los criterios que han de seguirse para identificar los segmentos: minimizar el tamaño de los mismos alrededor de la *señal*, de modo que al eliminar *reparandum* y *señal* quede una frase gramaticalmente correcta, y la equivalencia gramatical entre *reparandum* y *corrección*. En cuanto a la definición de los niveles de anidamiento, en la práctica es raro encontrar más de dos niveles, y suele tratarse siempre de una reformulación en la que se anidan una o más repeticiones. En cuanto a los segmentos de frase abandonados, objeto del apartado 2.6.2, es importante recordar que se pueden identificar como reformulaciones con sustitución en las que *reparandum* y *corrección* no cumplen la misma función gramatical, es decir, representan cambios abruptos en la estructura gramatical. En estos casos el etiquetado se reduce a identificar el segmento abandonado.

El anotador utilizará únicamente dos fuentes de información: la señal y la transcripción ortográfica original, y generará un unico producto: la transcripción ortográfica corregida y aumentada con marcas de disfluencias. Para insertar las marcas utilizará un editor de texto simple. El fichero de texto resultante tendrá la extensión *.dis* para distinguirlo del fichero con la transcripción original, cuya extensión es *.txt*.