

PRESEEA

MARCAS Y ETIQUETAS MÍNIMAS OBLIGATORIAS



Índice

	Pág.
Introducción	
A. Textos etiquetados	2
Formato	2
Configuración de los textos	2
<i>Cabecera</i>	2
<i>Etiquetado de texto</i>	5
<i>Marcas y etiquetas comunes de PRESEEA</i>	7
<i>Comentarios complementarios</i>	8
B. Textos sin etiquetas	8
C. Apéndices	9
Texto de muestra etiquetado	9
Texto de muestra sin etiquetas	11
Códigos de ciudades de PRESEEA	13
Autoría	14

Introducción

Los materiales reunidos dentro de PRESEEA y destinados a los fines comunes del proyecto se presentan en dos tipos formatos:

1. Archivos sonoros
2. Archivos de texto en transcripción ortográfica enriquecida

Los archivos sonoros ofrecen la grabación de las entrevistas semidirigidas realizadas en las comunidades de habla objeto de investigación sociolingüística (www.linguas.net/preseea). Los archivos de texto ofrecen la transcripción ortográfica enriquecida de las entrevistas grabadas. Estos archivos de texto se presentan en dos formatos, preparados con fines diferentes:

- Texto etiquetado, destinado a su uso informático con diversos fines.
- Texto sin etiquetar, destinado a su publicación impresa y a la lectura convencional.

Este documento de “Marcas y etiquetas mínimas obligatorias” ofrece la información esencial sobre cómo elaborar y presentar los archivos de texto en transcripción ortográfica enriquecida.

A. TEXTOS ETIQUETADOS

Formato

Los textos etiquetados han de elaborarse y almacenarse en formato electrónico, en archivos de texto con extensión .TXT. Para su creación y manejo se requiere un editor de textos.¹

La denominación de los archivos ha de seguir unas pautas comunes. Se utiliza un sistema que comienza con cuatro caracteres identificativos de la comunidad estudiada (p.e. MALA: Málaga),² seguidos del código sociolingüístico del informante (p.e. H23. Sexo/género: H(ombre), M(ujer); grupo de edad: 1, 2, 3; nivel educativo: 1, 2, 3) y de un número de tres cifras, comprendido entre el 001 y el número máximo de informantes utilizados (p.e. 054 ó 108). Cada equipo podrá adjudicar a los archivos otros códigos y numeraciones para sus fines particulares, pero los materiales presentados para el proyecto común deberán ajustarse al sistema general de identificación.³

El posterior tratamiento informático de los textos etiquetados se hará utilizando el lenguaje de codificación XML, versión adaptada y simplificada de SGML, que facilitará el intercambio de datos y la recuperación selectiva de información.

Configuración de los textos

Los textos etiquetados presentan dos partes bien diferenciadas: la cabecera y el texto propiamente dicho. Para la elaboración de la cabecera, se hará uso de una plantilla que habrá de rellenarse con los datos específicos.

Cabecera

La cabecera está formada por una serie de campos que proporcionan información sobre los siguientes aspectos:

- Datos del propio archivo.
- Datos de la grabación de la entrevista.
- Datos sobre la transcripción y revisión de la entrevista.
- Datos sobre los hablantes participantes en la entrevista.

Esos datos han de ajustarse a un formato común que garantice la homogeneidad de su tratamiento por parte de todos los equipos investigadores y los procesos comunes.⁴ Además, la plantilla tendrá un formato compatible con XML, para asegurar la recuperación de la información. Los datos específicos se

¹ El proyecto PRESEEA no requiere el uso de unos productos comerciales determinados ni de unos sistemas operativos específicos. Como se ha señalado, basta con que los textos tengan el formato TXT. Es preciso asegurarse de que los textos están codificados en ISO-8859-1, también conocido como "Latin.1".

² Véase Apéndice correspondiente.

³ Tras la identificación correspondiente, los archivos de texto recibirán la extensión .TXT. Los archivos de audio recibirán la misma denominación y se distinguirán de los de texto por llevar la extensión .MP3 o .WAV. En caso de que la grabadora utilizada realice su registro en un formato diferente (p.e. analógico, WMA), el archivo debe pasarse a formato digital WAV, recomendado por PRESEEA para el almacenamiento de grabaciones. Para facilitar los envíos o para colgar los archivos en la página electrónica, puede utilizarse el formato MP3.

⁴ Este modelo de cabecera, que recoge los datos previstos en otros tipos de cabecera manejados por los equipos de PRESEEA, se ha preparado buscando su plena compatibilidad con el sistema que se maneja para la elaboración del "Corpus del Español del siglo XXI", de las Academias de la Lengua Española.

complimentan entre las comillas correspondientes en los espacios que aparecen sombreados en esta plantilla de muestra.⁵

```
<Trans audio_filename="MALA_H23_001.mp3" xml:lang="español">
<Datos clave_texto="MALA_H23_001" tipo_texto="entrevista_semidirigida">
<Corpus corpus="PRESEEA" subcorpus="ESESUMA" ciudad="Málaga" pais="España"/>

<Grabacion resp_grab="Matilde Vida" lugar="domicilio informante" duracion="07'16"
fecha_grab="1998-01-01" sistema="WAV"/>
<Transcripcion resp_trans="Matilde Vida" fecha_trans="1999-01-01"
numero_palabras="1586"/>
<Revision num_rev="1" resp_rev="Antonio Ávila" fecha_rev="2005-01-02"/>
<Revision num_rev="2" resp_rev="Juan Villena" fecha_rev="2007-01-03"/></Datos>

<Hablantes>
<Hablante id="hab1" nombre="MALA_H23_001" codigo_hab="I" sexo="hombre"
grupo_edad="2" edad="44" nivel_edu="alto" estudios="derecho" profesion="abogado"
origen="Málaga" papel="informante"/>
<Hablante id="hab2" nombre="Matilde Vida" codigo_hab="E" sexo="mujer" grupo_edad="2"
edad="36" nivel_edu="alto" estudios="filología" profesion="profesora" origen="Málaga"
papel="entrevistador"/>
<Hablante id="hab3" nombre="LPM" codigo_hab="A1" sexo="mujer" grupo_edad="2"
edad="40" nivel_edu="alto" estudios="filología" profesion="profesora" origen="Málaga"
papel="audiencia"/>

<Relaciones rel_ent_inf="desconocidos" rel_inf_aud1="desconocidos"
rel_ent_aud1="conocidos" rel_inf_aud2="no" rel_ent_aud2="no"/>
</Hablantes> </Trans>6
```

Es importante tener en cuenta que, en XML, las mayúsculas y las minúsculas son diferentes, por lo que su uso dentro de la cabecera debe ajustarse al modelo que acaba de exponerse. A continuación se presentan los datos contenidos en la plantilla con las explicaciones necesarias y los requisitos que han de cumplirse.

⁵ Los equipos de PRESEEA podrán incluir más información en esta plantilla, de acuerdo con los intereses derivados del estudio de su comunidad. Ahora bien, la plantilla que ha de acompañar a los materiales comunes de PRESEEA ha de ajustarse estrictamente al formato que aquí se presenta.

⁶ Los datos que figuran en esta plantilla proceden del texto de muestra aportado por el equipo PRESEEA-MÁLAGA-ES. Los datos se especifican en el ejemplo para facilitar su comprensión.

Datos del propio archivo

nombre del archivo de audio (audio_filename): MALA_H23_001.mp3
clave de texto (clave_texto): MALA_H23_001 [Código con formato general de PRESEEA: código de ciudad – 4 caracteres⁷ –, código de informante – sexo/género, grupo de edad, nivel educativo⁸ – número de entrevista, dado por cada equipo]
tipo de texto: entrevista_semidirigida
corpus: PRESEEA
subcorpus: ESESUMA [nombre del corpus de cada equipo, si lo hay]
ciudad: Madrid
país: España

Datos sobre la grabación de la entrevista

responsable de grabación (resp_grab): Matilde Vida
lugar de grabación: domicilio informante
duración: 07'16''
fecha de grabación (fecha_grab): 1998-01-01 [las fechas que aparecen en la cabecera han de tener el formato aaaa-mm-dd]⁹
sistema: WAV [tipo de archivo de grabación original; si la grabación se hizo en cinta, debe anotarse “analógico”]

Datos sobre la transcripción y revisión de la entrevista

responsable de transcripción (resp_trans): Matilde Vida
fecha de transcripción (fecha_trans): 1999-01-01 [las fechas que aparecen en la cabecera han de tener el formato aaaa-mm-dd]
extensión (numero_palabras): 1586 [el número de palabras corresponde al texto transcrito, excluidas la cabecera completa y las etiquetas]
revisión 1 (resp_rev): Antonio Ávila
fecha revisión 1 (fecha_rev): 2005-01-02 [las fechas que aparecen en la cabecera han de tener el formato aaaa-mm-dd]
revisión 2 (resp_rev): Juan Villena
fecha revisión 2 (fecha_rev): 2007-01-03 [las fechas que aparecen en la cabecera han de tener el formato aaaa-mm-dd]

Datos sobre los hablantes participantes en la entrevista

nombre de informante (nombre): MALA_H23_001 [se utilizará el mismo código que figura en el campo “clave de texto”; en este caso, cada equipo podría añadir, tras un nuevo guion bajo, otro código que permitiera su identificación con fines particulares]
código hablante (codigo_hab): I [los códigos de hablante son I (Informante), E (Entrevistador) y A1 (Audiencia)¹⁰]
sexo: hombre [hombre | mujer]
grupo de edad (grupo_edad): 2 [1 | 2 | 3]
edad: 44
nivel educativo (nivel_edu): alto [bajo | medio | alto]
estudios: derecho
profesión: abogado
origen: Málaga
papel: informante

⁷ Véase apéndice de códigos de ciudades.

⁸ Según criterios de la metodología general de PRESEEA.

⁹ Si se desconociera el día o el mes exactos, utilícese 01 en el lugar correspondiente (p.e. 2006-01-01).

¹⁰ Si hubiera más hablantes, recibirían el código A2 o A3, por ejemplo.

nombre de entrevistador (nombre): Matilde Vida
código hablante (codigo_hab): E
sexo: mujer [hombre | mujer]
grupo de edad (grupo_edad): 2 [1|2|3]
edad: 36
nivel educativo (nivel_edu): alto [bajo|medio|alto]
estudios: filología
profesión: profesora
origen: Málaga
papel: entrevistador

nombre de audiencia 1: [puede haber más de un interlocutor como audiencia; los datos que se desconozcan reciben el valor “desc” (desconocido); si no existe audiencia, se elimina de la cabecera el segmento correspondiente al “hab3”]
código hablante (código_hab): A1 [en caso de haber más de un hablante como “audiencia”, pueden utilizarse los códigos A2, A3, ...]
sexo: mujer [hombre | mujer]
grupo de edad (grupo_edad): 2 [1|2|3 | desc]
edad: 40
nivel educativo (nivel_edu): alto [bajo|medio|alto |desc]
estudios: filología [también podría ser “desc”]
profesión: profesora [también podría ser “desc”]
origen: Málaga [también podría ser “desc”]
papel: audiencia

relación EI (rel_ent_inf): desconocidos [relación entre entrevistador e informante: conocidos | desconocidos]
relación IA1 (rel_inf_aud1): desconocidos [relación entre informante y audiencia: conocidos | desconocidos|no] [si no existe A1, debe anotarse la opción “no”]
relación EA1 (rel_ent_aud1): conocidos [relación entre entrevistador y audiencia: conocidos | desconocidos|no] [si no existe A1, debe anotarse la opción “no”]
relación IA2 (rel_inf_aud2): desconocidos [relación entre informante y audiencia: conocidos | desconocidos|no] [si no existe A2, debe anotarse la opción “no”]
relación EA2 (rel_ent_aud2): conocidos [relación entre entrevistador y audiencia: conocidos | desconocidos|no] [si no existe A2, debe anotarse la opción “no”] [para la cabecera, no se tendrá en cuenta la posible presencia de otras personas que actúen como “audiencia”]¹¹

Etiquetado de texto

El texto correspondiente a cada entrevista ha de transcribirse en ortografía convencional, incluida la acentuación. Las palabras que presenten elisiones se completan en su escritura.

No obstante, el texto transcrito también debe ajustarse a ciertos criterios gráficos que no son los habituales en la escritura ordinaria, sino que responden a unas necesidades propias de la representación escrita de la lengua oral. Estos criterios afectan especialmente a la puntuación y a la forma de representar las pausas, pero también a otros aspectos, que quedan aquí convenientemente precisados.

¹¹ Se consideran “conocidos” los familiares, los amigos, los compañeros de trabajo cercanos o los vecinos con los que se tiene comunicación frecuente; se consideran “desconocidos” las personas con las que no se establece comunicación habitualmente, incluidos los vecinos o compañeros con los que no se suele tener trato. Téngase en cuenta que los tipos de relaciones entre interlocutores pueden ser muy diversos. No obstante, las características de las relaciones entre los hablantes han de quedar bien detalladas en las fichas de entrevista que maneje cada equipo. La información que se refleja en la cabecera de los textos tan solo pretende fijar unos parámetros mínimos para la comparación de materiales. Cualquier duda exigirá la consulta de los datos completos de las entrevistas.

El marcado y etiquetado del texto debe hacerse utilizando una serie de signos, marcas, etiquetas y convenciones que serán de uso común y obligatorio por parte de los equipos integrantes de PRESEEA.¹²

Es muy importante que los investigadores conozcan cuáles son los objetivos generales del marcado y etiquetado de textos, puesto que esos objetivos son los que deben orientar las decisiones que se tomarán en los casos de duda o de dificultad imprevista. Ha de tenerse en cuenta que es imposible prever en un documento de estas características todas las circunstancias posibles en el proceso de transcripción y etiquetado de un texto, por lo que resulta imprescindible recurrir a unos objetivos comunes y unos criterios generales. Esos criterios generales, a partir de los cuales se han seleccionado las marcas y etiquetas mínimas y obligatorias de PRESEEA, se pueden resumir en los siguientes puntos:

- Las marcas y etiquetas han de reflejar aquellos aspectos fónicos y ambientales cuyo conocimiento resulte imprescindible para una correcta comprensión del contenido de la entrevista.
- Las marcas y etiquetas han de ser económicas y deben evitar las redundancias y los esfuerzos desproporcionados en el proceso de transcripción.
- Las marcas y etiquetas deben buscar la fidelidad y la objetividad de lo registrado en la entrevista, minimizando o eliminando el componente interpretativo, especialmente en lo relacionado con la intención comunicativa de los hablantes.
- Las marcas y etiquetas no pretenden reflejar toda la información que ofrece un archivo sonoro, cuya audición será imprescindible para proceder al análisis de determinados aspectos lingüísticos y comunicativos.
- Las marcas y etiquetas han de ser compatibles con tratamientos informáticos orientados a diversos fines (p.e. alineamiento texto-voz; integración de los textos en el "Corpus del Español del siglo XXI").
- Las marcas y etiquetas no tienen como finalidad señalar posibles objetos de estudio.

A continuación se presentan los signos, convenciones, marcas y etiquetas que han de aplicarse en la transcripción de los textos que integran el proyecto PRESEEA. Junto a cada uno de ellos, aparece una breve descripción de su función; también se marca, cuando es pertinente, si deben aparecer antes (A) o después (D) de la secuencia etiquetada o si reflejan un hecho producido en el instante (I) en que se anota la etiqueta.¹³

Las etiquetas propiamente dichas pueden clasificarse de la siguiente forma: de ruidos, fónicas, léxicas, de dinámica discursiva, de lengua y de transcripción. Las etiquetas siempre van precedidas y seguidas de un espacio, excepto la etiqueta <alargamiento/>. Tras el cuadro general, se hacen otras aclaraciones y comentarios complementarios.

¹² Aquellos equipos que decidan trabajar con el programa "Word" y guardar sus textos con formato TXT, dispondrán de una macro con las etiquetas preparadas para su rápida anotación.

¹³ El lugar en que aparece la etiqueta, así como las marcas de cierre, son importantes para evitar conflictos entre etiquetas a la hora de tratar los textos con cualquier procedimiento informático. Un fragmento de discurso puede recibir más de un tipo de etiqueta.

Marcas y etiquetas comunes de PRESEEA

ORTOGRAFÍA Y PUNTUACIÓN	
i!	Enunciados exclamativos
¿?	Enunciados interrogativos
/	Pausa mínima
//	Pausa
:	Tras código de hablante (I: E: A1:)
Mayúsculas	Inicial de nombres propios y siglas
Elementos cuasi-léxicos funcionales	Interjecciones; apoyos. Escritura ortográfica (ah, ay, aha, mmm, eeh, pff, bah)
Onomatopeyas	Escritura ortográfica (zas, bum, plas)

ETIQUETADO DE RUIDOS	
<ruido = " "/>	Ruido, con especificación de tipo (p.e. <ruido = "chasquido boca"/>) I
<ruido_fondo> </ruido_fondo>	Ruido continuo de fondo AD
<risas = " "/>	Risas, con especificación de emisor/es (p.e. <risas = "E"/>, <risas = "todos"/>) I
<entre_risas> </entre_risas>	Risas simultáneas con el habla AD
<registro_defectuoso> </registro_defectuoso>	Fragmento de la grabación de mala calidad AD
<interrupción_de_grabación/>	Interrupción de la grabación I

ETIQUETADO FÓNICO	
<énfasis> </énfasis>	Fragmento con pronunciación claramente enfática AD
<alargamiento/>	Alargamiento de sonido D (sin espacios)
<silencio/>	Silencio de un segundo o más I
<palabra_cortada/>	Palabra cortada D
<vacilación/>	Vacilación; titubeo breve I
<sic> </sic>	No es descuido de transcripción AD
<ininteligible/>	Fragmento ininteligible I

ETIQUETADO LÉXICO	
<término> </término>	Lexía claramente usada como uso especializado AD
<extranjero> </extranjero>	Extranjerismo (excepto usos de la L2 del hablante) AD
<siglas = []> </siglas>	Siglas; incluye pronunciación AD

ETIQUETADO DE DINÁMICA DISCURSIVA	
<cita> </cita>	Cita, estilo directo AD
<simultáneo> </simultáneo>	Solapamiento (traslape). También se usa en turnos de apoyo, si fuera necesario AD

ETIQUETADO DE LENGUA	
<lengua = " "> </lengua>	Cambio de lengua (léxico, oracional, ...), especialmente L2 del hablante, con indicación desarrollada de lengua (p.e. <lengua = "gallego"> </lengua> AD

ETIQUETADO DE TRANSCRIPCIÓN	
<transcripción_dudosa> </transcripción_dudosa>	Transcripción dudosa para transcriptor y revisores AD
<tiempo = " "/>	Anotación de minuto y segundo de grabación. (p.e. <tiempo = "02:45"/>) I
<observación_complementaria = " "/>	Observación complementaria I

Comentarios complementarios

- Los dos puntos (:) se utilizan para marcar inicio de turno, tras el código del hablante.¹⁴
- Las mayúsculas solamente se utilizan en la inicial de nombres propios y en las siglas.
- La pausa mínima en ocasiones es casi imperceptible; puede representar una pausa que coincida con una frontera de entonación. El silencio es claramente perceptible (generalmente más de un segundo). La pausa es intermedia entre los anteriores.
- No se utiliza punto a final de turno. La ausencia de punto equivale a abandono o suspensión, sin especificarse si ha sido voluntario o involuntario. Si el final de un turno resulta de una interrupción por parte del interlocutor y se corta una palabra, se marca <palabra_cortada/>, sin embargo no se marca nada si corta una oración.
- Las pausas entre turnos se marcan al final del turno, no al inicio del siguiente.
- Los elementos cuasi-léxicos o paralingüísticos se representan en ortografía ordinaria, de acuerdo con las convenciones más habituales, cuando existan (uhum, ah, eeh, bah). Los ejemplos del cuadro no constituyen una lista cerrada.
- La etiqueta <vacilación/> se usa para marcar pequeños titubeos que no constituyen palabras.
- La etiqueta <sic> no marca usos erróneos o vulgarismos, ni llama la atención sobre formas especiales; solamente se usa cuando la transcripción pudiera interpretarse como error del transcriptor.
- La etiqueta <término> se aplica exclusivamente cuando el hablante haga uso de un tecnicismo, claramente especializado.
- La etiqueta <cita> se usa tanto para las citas textuales como para el estilo directo, usos que en la ortografía convencional se marcarían mediante comillas.
- El solapamiento o traslape de turnos (<simultáneo>) se marca cuando comienza y cuando termina en el discurso del hablante y también se marca cuando comienza y cuando termina en el discurso del interlocutor.
- La etiqueta <tiempo = " " /> se inserta siempre a comienzo de turno, en lugares correspondientes a intervalos de dos minutos aproximadamente o cuando se inicia una fase diferenciada en el desarrollo de la entrevista.
- La etiqueta de <observación_complementaria = " " /> puede incluir cualquier información no prevista en la relación mínima de etiquetas.
- Los equipos que trabajen en comunidades bilingües pueden incluir en la relación de etiquetas otras que tengan que ver con el uso de dos lenguas en la conversación, además del cambio de una lengua a otra, ya recogido en la serie común. En este caso, será imprescindible una descripción detallada del significado de cada etiqueta.

B. TEXTOS SIN ETIQUETAS

La versión de los textos sin etiquetar tiene como finalidad la publicación impresa de los materiales y su lectura convencional. La presentación de un texto sin etiquetar consiste básicamente en disponer una cabecera con los datos esenciales del informante, el número de entrevista, la fecha de la grabación y el texto de la transcripción desprovisto de las etiquetas, tanto las de apertura y cierre, como las aisladas, excepto <risas = " " /> y <silencio/>

¹⁴ El posterior tratamiento informático de los turnos exigirá el uso de etiquetas del tipo <turno = " " />, que sustituirán al sistema de distribución E., I.:. El mantenimiento, en la fase de transcripción, de este último sistema simplemente pretende facilitar la legibilidad del texto por parte de los transcriptores y el paso a un formato sin etiquetas.

C. APÉNDICES

TEXTO DE MUESTRA ETIQUETADO (SIN CABECERA)¹⁵

- E.: y en cuanto a<alargamiento/> / cambiando de tema completamente / la ciudad / Málaga /
- I.: uhum
- E.: tú que eres de Málaga / eeh ¿cómo la describirías / cómo <vacilación/> // eeh / o cómo ha cambiado Málaga en todos los años que llevas tú viviendo aquí?
- I.: ¡buf! muchísimo / Málaga muchísimo yo recuerdo<alargamiento/> // pues<alargamiento/> / <vacilación/> porque la verdad es que me encantaba también / el el salir cuando salía con mi padre con el coche / una historia / que iba porque tenía él tenía una pequeña finca en Cártama // y salíamos mucho de<alargamiento/> / a Cártama o a Marbella que mi abuelo tenía allí un chalecito o<alargamiento/> / o a Benalmádena que tenían mis tíos / estaban allí también veraneando íbamos mucho allí también / entonces<alargamiento/>s / recorrimos mucho <ininteligible/> / a Cártama a la Estación de Cártama<alargamiento/> bueno muchos pueblos la verdad es que era / era habitual el que saliéramo<alargamiento/>s con el coche mucho por lo menos en aquella época de pequeñitos muy pequeñitos // y<alargamiento/> y yo recuerdo eso que las carreteras eran minúsculas / no es que no había carreteras es que había / poquísimas carreteras / <risas = "l"/> eeh la costa era horrible porque la costa había que cruzar el puente de la Azucarera que aquello era / una <risas = "l"/> una travesía como<alargamiento/> / increíble porque era un solo carril en cada sentido / y la verdad es que<alargamiento/> <ruido = "chasquido boca"/> tuvo / bueno incluso en la propia ciudad una ciudad muy pequeñita yo creo que era / Málaga era un pueblo / Málaga no es / yo creo que no tenía por lo menos no no la é <palabra_cortada/> la época que me ha tocado a mí vivir <vacilación/> // no le / yo no le daba la consideración de ciudad / sin embargo<alargamiento/> los últimos años ha avanzado notablemente es decir ha sido // no tanto / como debiera <risas = "l"/> / por supuesto porque creo que debiera debía haber avanzado bastante más / pero sí ha avanzado muchísimo y Málaga se ha convertido en una ciudad / una gran ciudad / una gran ciudad / y ahora se nota / ahora gracias a Dios se nota porque aquí tenemos empezamos a tener teatro empezamos a tener cines empezamos a tener / eeh muchas cosas que tú<alargamiento/> / bueno que cuando<alargamiento/> / <vacilación/> que hemos / que hemos hecho yo creo que todo el mundo lo ha hecho <vacilación/> para para ver una buena obra de teatro te tienes que coger el avión o el metro o el tren / bueno el metro no perdón el tren <risas = "todos"/> / el metro ¡ojalá lo tengamos! / y te tienes que ir a Madrid o te tienes que ir a Barcelona / ¡para ver una obra de teatro! lo que/ lo que tú decías <cita> bueno es que es una barbaridad que yo en mi ciudad no pueda ver </cita> y sin embargo no hoy día / gracias a Dios pues sí tenemos / algunas posibilidades ya hay buenos restaurantes aquí al lado / que eso estaba reservado antes solamente a bueno a la costa a alguno<alargamiento/>s buenos hoteles que tenían espectáculo a / al casino a Torrequebrada o alguna / no había nada lo que era Málaga ciudad no tenía<alargamiento/> // y va cambiando / la verdad es que va cambiando y<alargamiento/> / hombre // y luego tenemos <vacilación/> tenemos el clima especial que tenemos no <risas = "l"/> que quizá lo único que eso / como eso no se lo pueden llevar nunca / por ahora / <risas = "l"/> eso es lo que nos mantiene como ciudad y <vacilación/> bueno / y el tirón que está pegando últimamente de también de a nivel a cultural / que quizá lo que yo echaba mucho en falta en Málaga era<alargamiento/> / el nivel cultural bueno con los museos el Museo Picasso el <siglas = [kák]> CAC </siglas> / el Centro de Arte Contemporáneo / eeh aparte de los museos nuevos aparte de lo<alargamiento/>s / de los edificio<alargamiento/>s / históricos que tenemos aquí / pero eso ha dado mucho<alargamiento/> mucho renombre y quizá<alargamiento/> / o sea que ahora / Málaga es una ciudad donde cualquiera que venga de fuera / eeh y te diga <cita> oye ¿qué hacemos un sábado / qué hacemos? </cita> y lo puedes llevar a cualquier sitio / es decir antes no antes era muy complicado decir <cita> ¿qué hacemos? </cita> pues tenemos nos vamos a la playa <risas = "l"/> o nos vamos a un chiringuito a tomar / a toma<alargamiento/>r espeto<alargamiento/>s / un vino lo que sea y y eso se echaba en falta entonces / en aquella época por eso te digo que yo consideraba a Málaga más que como una ciudad como un pueblo /
- E.: uhum <ruido = "chasquido boca"/>
- I.: <tiempo = "03:29"/> hoy día ya te digo hoy en día sí / hoy día // aunque a pesar de eso te digo mm / creo que no se le ha tratado todo lo bien que se / que debiera / <vacilación/> también hemos vivido la época esta de las autonomías donde bueno // el centralismo de Madrid pasó a ser centralismo de las capitales de las autonomías con lo cual / pues más o menos hemos estado / parecido / bueno
- E.: <risas = "E"/> sí
- I.: pero se ha avanzado<alargamiento/> / notablemente ya <vacilación/> en Málaga toda a nivel de infraestructura a nivel
- E.: y<alargamiento/> en cuanto a los valores culturales <vacilación/> la feria / la Semana Santa <vacilación/> ¿tú participas activamente? ¿te gusta
- I.: bueno / he tenido mis épocas también // <ruido = "chasquido boca"/> la feria no / la feria ya / la Semana Santa eeh quizá / bueno no tanto por el sentimiento religioso sino por el sentimiento social / he participado desde

¹⁵ Véase la cabecera más arriba.

- pequeño / no desde pequeño / desde que tenía dieciocho años / creo que fue la primera vez que saqué un trono La Piedad / eeh luego con la facultad de derecho / sacaba todos los años la virgen de los Estudiantes / pero ya te digo más que como un fenómeno religioso que como un fenómeno social un fenómeno de bueno
- E.: sí cultural de la ciudad ¿no? </simultáneo>
- I.: <tiempo = "04:33"/> cultural de la ciudad pero que no era no tenía ese sentimiento // o por lo menos yo no se lo encontraba / aunque en algunas ocasiones es verdad que // que una vez que te incorporas / a cualquier actividad de la Semana Santa la verdad es que ahí hay momentos hay ráfagas donde sí tienes / surge algún sentimiento o pensamiento ¿no? de estos de
- E.: <risas = "E"/> místicos
- I.: místicos <risas = "I"/> entonces sí es verdad que tiene algo <vacilación/> incluso no / la verdad es que impresiona / yo mismo a mí mismo me ha impresionado siempre la Semana Santa me ha impresionado desde pequeño / eeh y hasta ahora mismo cualquier trono que pasa en el recorrido oficial que está todo el mundo callado tú lo ves / o sea en / es algo que sí que <vacilación/> que mantengo como sentimiento <alargamiento/> // y por supuesto me gustaría que se mantuviera <alargamiento/> / de por vida ¿no? porque creo que es un fenómeno / que quizás ya te digo / más que como religioso como cultural por lo menos social / eeh la feria menos / la feria la entiendo menos / la feria es una cosa que <alargamiento/> / me gusta / me ha gustado mucho // quizá mm la feria del centro pues era más cómoda / o es más cómoda la feria del centro ¿no? pues bajas te tomas unas copas / te vas cuando quieres no hay tanto polvo no hay tanto /
- E.: uhum
- I.: lo de la feria de <vacilación/> del real donde se / la verdad y las casetas sí ya te digo he participado muchísimo hasta hasta con el Colegio de Abogados mon <palabra_cortada/> teníamos antes una caseta / y participábamos en ella <vacilación/> en muchas actividades / pero no no me ha atraído por lo menos nunca me ha atraído y entonces y desde desde hace unos años / lo que hago es que / precisamente <risas = "I"/> en la época de feria / me voy de vacaciones <risas = "todos"/> / que es / que es <vacilación/> las dos semanas o la semana esta larga que <vacilación/> / que está todo cerrado seguro / con lo cual aprovecho incluso para irme de vacaciones porque no no me llega a atraer <vacilación/> no sé / más de la cuenta / hoy día también ten en cuenta que los / para los niños / creo que es mucho más seguro y mucho más / un parque de atracciones que la feria / entonces para los niños es un / les gusta la feria por los carricoches y los los aparatos estos de feria pero no / me merece más la pena irme a Port Aventura o irme a Tívoli o irme adonde sea y lo pasan mucho mejor más seguro más más limpio / que estar allí con // <vacilación/> = "chasquido boca"/>
- E.: <ininteligible/> con los carricoches
- I.: <tiempo = "06:37"/> con los carricoches ¿no? y luego no también el temor que ha habido / la feria de Málaga se ha masificado hay muchísima gente <alargamiento/> / bueno hay gente de todo tipo / y quieras que no todos los años pues ocurre algún tipo de <alargamiento/> // de pelea algún tipo de / de agresión <ininteligible/> no sé / no no me hace mucha gracia / ya te digo que lo mismo que la he disfrutado durante muchos años ahora mismo / procuro quitarme de enmedio <risas = "I"/> / y lo utilizo como excusa para eso para saber que / en esos días / nadie te llama porque todo el mundo está en feria y nadie quie <palabra_cortada/> nadie quiere nadie quiere saber nada / pero vamos
- E.: muy bien / pues ya está
- I.: ¿ya está?
- E.: ya // voy a apagar esto

TEXTO DE MUESTRA SIN ETIQUETAS

ENTREVISTA 52 - H23
HOMBRE DE 44 AÑOS, NIVEL DE ESTUDIOS ALTO, ABOGADO

FECHA DE GRABACIÓN: 01-01-1998

- E.: y en cuanto al / cambiando de tema completamente / la ciudad / Málaga /
- I.: uhum
- E.: tú que eres de Málaga / eeh ¿cómo la describirías / cómo // eeh / o cómo ha cambiado Málaga en todos los años que llevas tú viviendo aquí?
- I.: ¡buf! muchísimo / Málaga muchísimo yo recuerdo // pues / porque la verdad es que me encantaba también / el salir cuando salía con mi padre con el coche / una historia / que iba porque tenía él tenía una pequeña finca en Cártama // y salíamos mucho de / a Cártama o a Marbella que mi abuelo tenía allí un chalecito o / o a Benalmádena que tenían mis tíos / estaban allí también veraneando íbamos mucho allí también / entonces / recorrimos mucho / a Cártama a la Estación de Cártama a bueno muchos pueblos la verdad es que era / era habitual el que saliéramos con el coche mucho por lo menos en aquella época de pequeñitos muy pequeñitos // y yo recuerdo eso que las carreteras eran minúsculas / no es que no había carreteras es que había / poquíssimas carreteras / <risas = "l"/> eeh la costa era horrible porque la costa había que cruzar el puente de la Azucarera que aquello era / una <risas = "l"/> una travesía como / increíble porque era un solo carril en cada sentido / y la verdad es que tuvo / bueno incluso en la propia ciudad una ciudad muy pequeñita yo creo que era / Málaga era un pueblo / Málaga no es / yo creo que no tenía por lo menos no no la é la época que me ha tocado a mí vivir // no le / yo no le daba la consideración de ciudad / sin embargo los últimos años ha avanzado notablemente es decir ha sido // no tanto / como debiera <risas = "l"/> / por supuesto porque creo que debiera debía haber avanzado bastante más / pero sí ha avanzado muchísimo y Málaga se ha convertido en una ciudad / una gran ciudad / una gran ciudad / y ahora se nota / ahora gracias a Dios se nota porque aquí tenemos empezamos a tener teatro empezamos a tener cines empezamos a tener / eeh muchas cosas que tú / bueno que cuando / que hemos / que hemos hecho yo creo que todo el mundo lo ha hecho para para ver una buena obra de teatro te tienes que coger el avión o el metro o el tren / bueno el metro no perdón el tren <risas = "todos"/> / el metro ¡ojalá lo tengamos! / y te tienes que ir a Madrid o te tienes que ir a Barcelona / ¡para ver una obra de teatro! lo que/ lo que tú decías bueno es que es una barbaridad que yo en mi ciudad no pueda ver y sin embargo no hoy día / gracias a Dios pues sí tenemos / algunas posibilidades ya hay buenos restaurantes aquí al lado / que eso estaba reservado antes solamente a bueno a la costa a algunos buenos hoteles que tenían espectáculo a / al casino a Torrequebrada o alguna / no había nada lo que era Málaga ciudad no tenía // y va cambiando / la verdad es que va cambiando y / hombre // y luego tenemos tenemos el clima especial que tenemos no <risas = "l"/> que quizá lo único que eso / como eso no se lo pueden llevar nunca / por ahora / <risas = "l"/> eso es lo que nos mantiene como ciudad y bueno / y el tirón que está pegando últimamente de también de a nivel a cultural / que quizá lo que yo echaba mucho en falta en Málaga era / el nivel cultural bueno con los museos el Museo Picasso el CAC / el Centro de Arte Contemporáneo / eeh aparte de los museos nuevos aparte de los / de los edificios / históricos que tenemos aquí / pero eso ha dado mucho mucho renombre y quizá / o sea que ahora / Málaga es una ciudad donde cualquiera que venga de fuera / eeh y te diga oye ¿qué hacemos un sábado / qué hacemos? y lo puedes llevar a cualquier sitio / es decir antes no antes era muy complicado decir ¿qué hacemos? pues tenemos nos vamos a la playa <risas = "l"/> o nos vamos a un chiringuito a tomar / a tomar espetos / un vino lo que sea y y eso se echaba en falta entonces / en aquella época por eso te digo que yo consideraba a Málaga más que como una ciudad como un pueblo /
- E.: uhum
- I.: hoy día ya te digo hoy en día sí / hoy día // aunque a pesar de eso te digo mm / creo que no se le ha tratado todo lo bien que se / que debiera / también hemos vivido la época esta de las autonomías donde bueno // el centralismo de Madrid pasó a ser centralismo de las capitales de las autonomías con lo cual / pues más o menos hemos estado / parecido / bueno
- E.: <risas = "E"/> sí
- I.: pero se ha avanzado / notablemente ya en Málaga toda a nivel de infraestructura a nivel
- E.: y en cuanto a los valores culturales la feria / la Semana Santa ¿tú participas activamente? ¿te gusta
- I.: bueno / he tenido mis épocas también // la feria no / la feria ya / la Semana Santa eeh quizá / bueno no tanto por el sentimiento religioso sino por el sentimiento social / he participado desde pequeño / no desde pequeño / desde que tenía dieciocho años / creo que fue la primera vez que saqué un trono La Piedad / eeh luego con la facultad de derecho / sacaba todos los años la virgen de los Estudiantes / pero ya te digo más que como un fenómeno religioso que como un fenómeno social un fenómeno de bueno
- E.: sí cultural de la ciudad ¿no?
- I.: cultural de la ciudad pero que no era no tenía ese sentimiento // o por lo menos yo no no se lo encontraba / aunque en algunas ocasiones es verdad que // que una vez que te incorporas / a a cualquier actividad de la

Semana Santa la verdad es que ahí hay momentos hay ráfagas donde sí tienes / surge algún sentimiento o pensamiento ¿no? de estos de

E.: <risas = "E"/> místicos

I.: místicos <risas = "I"/> entonces sí es verdad que tiene algo incluso no / la verdad es que impresiona / yo mismo a mí mismo me ha impresionado siempre la la Semana Santa me ha impresionado desde pequeño / eeh y hasta ahora mismo cualquier trono que pasa en el recorrido oficial que está todo el mundo callado tú lo ves / o sea en / es algo que sí que que mantengo como sentimiento y // y por supuesto me gustaría que se mantuviera / de por vida ¿no? porque creo que es un fenómeno / que quizás ya te digo / más que como religioso como cultural por lo menos social / eeh la feria menos / la feria la entiendo menos / la feria es una cosa que / me gusta / me ha gustado mucho // quizá mm la feria del centro pues era más cómoda / o es más cómoda la feria del centro ¿no? pues bajas te tomas unas copas / te vas cuando quieres no hay tanto polvo no hay tanto /

E.: uhum

I.: lo de la feria de del real donde se / la verdad y las casetas sí ya te digo he participado muchísimo hasta hasta con el Colegio de Abogados mon teníamos antes una caseta / y participábamos en ella y en muchas actividades / pero no no me ha atraído por lo menos nunca me ha atraído y entonces y desde desde hace unos años / lo que hago es que / precisamente <risas = "I"/> en la época de feria / me voy de vacaciones <risas = "todos"/> / que es / que es la las dos semanas o la semana esta larga que / que está todo cerrado seguro / con lo cual aprovecho incluso para irme de vacaciones porque no no me llega a atraer no sé / más de la cuenta / hoy día también ten en cuenta que los / para los niños / creo que es mucho más seguro y mucho más / un parque de atracciones que la feria / entonces para los niños es un / les gusta la feria por los carricoches y los los aparatos estos de feria pero no / me merece más la pena irme a Port Aventura o irme a Tívoli o irme adonde sea y lo pasan mucho mejor más seguro más más limpio / que estar allí con //

E.: con los carricoches

I.: con los carricoches ¿no? y luego no también el temor que ha habido / la feria de Málaga se ha masificado hay muchísima gente y / bueno hay gente de todo tipo / y quieras que no todos los años pues ocurre algún tipo de // de pelea algún tipo de / de agresión no sé / no no me hace mucha gracia / ya te digo que lo mismo que la he disfrutado durante muchos años ahora mismo / procuro quitarme de enmedio <risas = "I"/> / y lo utilizo como excusa para eso para saber que / en esos días / nadie te llama porque todo el mundo está en feria y nadie quiere nadie quiere saber nada / pero vamos

E.: muy bien / pues ya está

I.: ¿ya está?

E.: ya // voy a apagar esto

CÓDIGOS DE CIUDADES DE PRESEEA

PRESEEA – ALCALÁ DE HENARES - ES	ALCA
PRESEEA – BARCELONA - ES	BARC
PRESEEA - BARRANQUILLA - CO	BARR
PRESEEA - BOGOTÁ - CO	BOGO
PRESEEA - CÁDIZ - ES	CÁDI
PRESEEA - CARACAS - VE	CARA
PRESEEA - CIPOLLETTI - AR	CIPO
PRESEEA - CULIACÁN - MX	CULI
PRESEEA - GRANADA - ES	GRAN
PRESEEA - GUATEMALA - GU	GUAT
PRESEEA - LAS PALMAS - ES	LASP
PRESEEA – LÉRIDA/LLEIDA - ES	LERI
PRESEEA - MADRID - ES	MADR
PRESEEA - MÁLAGA - ES	MALA
PRESEEA - MEDELLÍN - CO	MEDE
PRESEEA - MÉRIDA - MX	MERI
PRESEEA - MÉXICO -MX	MEXI
PRESEEA - MIAMI - EU	MIAM
PRESEEA - MONTERREY - MX	MONR
PRESEEA - MONTEVIDEO - UR	MONV
PRESEEA - OVIEDO - ES	OVIE
PRESEEA - PALMA DE MALLORCA - ES	PALM
PRESEEA - PEREIRA - CO	PERE
PRESEEA - QUITO - EC	QUIT
PRESEEA - SAN JUAN - PR	SANJ
PRESEEA - SAN MIGUEL - SA	SANM
PRESEEA - SANTIAGO DE CHILE - CH	SCHI
PRESEEA - SANTIAGO DE COMPOSTELA - ES	SCOM
PRESEEA - SEVILLA - ES	SEVI
PRESEEA - VALENCIA - ES	VALE
PRESEEA - VALLEDUPAR - CO	VALL
PRESEEA - VALPARAÍSO - CH	VALP
PRESEEA - ZARAGOZA - ES	ZARA

AUTORÍA

Este documento se ha elaborado tomando como base el trabajo de Antonio Ávila, Matilde Vida y Mari Cruz Lasarte “Propuesta de transliteración y etiquetado del Macrocorpus PRESEEA” (2006). Una vez difundida esta propuesta entre los equipos que integran el proyecto y experimentadas las ventajas y las dificultades de su aplicación en el proceso de investigación, la asamblea general de PRESEEA decidió elaborar un documento de marcas y etiquetas mínimas, de uso obligado en todos los materiales que integran PRESEEA y destinados a los fines comunes del proyecto. Para la redacción de este documento, se ha contado con informes y propuestas aportados por los equipos de las siguientes ciudades: Alcalá de Henares, Barcelona, Bogotá, Medellín, México D.F., Palma de Mallorca y Santiago de Compostela, además de Málaga. La aplicación de estas normas comunes para el marcado y etiquetado de los textos de PRESEEA no ha de impedir su complementación o ampliación con otras normas que sirvan a los intereses específicos de cada equipo de la red, para lo cual se requiere simplemente que los criterios comunes y los específicos no sean incompatibles metodológicamente.

En la elaboración de este documento de marcas y etiquetas obligatorias han participado las siguientes personas:

Antonio Ávila (redactor de documento base, consultor de XML y revisor)
Laura Camargo (revisora de marcas y etiquetas; consultora de cuestiones discursivas)
Ana M. Cestero (revisora de marcas y etiquetas; consultora de cuestiones conversacionales)
M. Claudia González Rátiva (revisora de marcas y etiquetas)
M. Clara Henríquez (revisora de marcas y etiquetas)
M. Cruz Lasarte (redactora del documento base y revisora)
Pedro Martín Butragueño (revisor de marcas y etiquetas; consultor de cuestiones fónicas)
Francisco Moreno Fernández (redactor del documento y revisor)
Montserrat Recalde (revisora de marcas y etiquetas)
Guillermo Rojo (asesor para cuestiones informáticas, consultor de XML y revisor)
José María Sánchez Sáez (consultor de XML y revisor)
María Sancho Pascual (creadora de macro de etiquetas y revisora)
Antonio Torres (revisor de marcas y etiquetas)
Victoria Vázquez (revisora de marcas y etiquetas)
Matilde Vida (redactora del documento base y revisora)
Juan Villena Ponsoda (consultor de XML y revisor)

A esta relación han de sumarse las contribuciones realizadas de forma oral en los distintos encuentros de investigadores de PRESEEA. La experiencia acumulada por todos los equipos del proyecto, especialmente los que hasta la fecha han reunido y tratado un mayor volumen de materiales (Alcalá, Granada, Guatemala, Málaga, México, Valencia), ha sido esencial en la preparación de este documento de marcas y etiquetas mínimas obligatorias. Para esos equipos, además, la adaptación a los criterios finales resulta especialmente penosa por tener ya concluidos o avanzados sus trabajos de transcripción de acuerdo con criterios diferentes. Su esfuerzo en beneficio del grupo es digno de admiración y agradecimiento.

17 de febrero de 2008

Cómo citar este documento:

PRESEEA (2008): “Marcas y etiquetas mínimas obligatorias”. Vers. 1.2. 17-02-2008.
<<http://www.linguas.net/preseea>>