

CORPUS MAVIR

CONVENCIONES USADAS PARA LAS TRANSCRIPCIONES DE HABLA

Las convenciones empleadas, que a continuación explicamos, son una versión revisada de las marcas utilizadas en el corpus C-ORAL-ROM. Consúltense para más información *C-ORAL-ROM. Integrated Reference Corpora for Spoken Romance Languages*, E. Cresti y M. Moneglia (eds.), John Benjamins (Amsterdam: 2005).

El corpus recoge transcripciones de habla oral, sin tener en cuenta las convenciones propias de la lengua escrita: no se emplean puntos, comas, comillas, punto y coma, guión, etc. Las unidades prosódicas que conforman el habla oral (tengan o no sentido completo) son los elementos delimitados en la transcripción según las marcas correspondientes: barra simple (/), triple barra (///), signo de cierre de interrogación (?), etc. Tampoco hemos indicado el tipo de entonación (ascendente, descendente, suspendida...).

Todas las palabras se transcriben en minúscula, con la excepción de los nombres propios (ej. *Universidad Europea de Madrid*). Los nombres de empresas o instituciones mantienen la ortografía del nombre tal y como aparece en la documentación de sus respectivos portales web (ej. *iSOCO, DAEDALUS*).

Las cifras o fechas se transcriben ortográficamente (ej. *dos millones quinientos noventa mil; el veintinueve de septiembre del dos mil seis*).

La cabecera de cada documento incluye la siguiente información metalingüística acerca de la grabación, incluida entre corchetes ([]) en el documento con formato .TXT:

@Title: identifica el contenido del texto.

@File: nombre del archivo (difiere sólo la extensión: .WAV, XML, TXT...)

@Participants: incluye la información acerca de los hablantes que intervienen:

- Identificación del turno del hablante (tres letras en mayúsculas)
- Nombre del hablante
- Entre paréntesis se especifican sus características (si se desconoce alguna, se indica como 'x'):
 - sexo: man / woman
 - edad:
 - A: 18-25 años
 - B: 25-40 años
 - C: 40- 50 años
 - D: más de 60 años

- nivel de estudios:
 - 1: educación primaria
 - 2: bachillerato
 - 3: graduados, licenciados y estudiantes universitarios
- profesión: ocupación, cargo o trabajo que desempeña.
- papel en la grabación: participante, moderador, conferenciante...
- lugar de procedencia geográfica

@Date: fecha de la grabación

@Place: lugar de grabación

@Situation: se expresa la siguiente información:

- Circunstancias de la grabación: conferencia, mesa redonda...
- Lugar de grabación: salón de actos, aula de universidad...
- Condiciones de grabación: micrófono escondido o no oculto; investigador observador o participante en la grabación...

@Topic: tema(s) principal(es) del discurso

@Source: nombre del propietario de los derechos de reproducción.

@Class: conjunto de campos que identifican el tipo de grabación; en el corpus MAVIR tan sólo aparecen los siguientes:

- formal in natural context: registro cuidado y público grabado en condiciones lo más cercanas al habla espontánea.
 - conference: conferencia o discurso expositivo realizado en público.
 - professional explanation: explicación técnica detallada.

A su vez, en estos subcampos se indica el estilo dialógico: monologue (un hablante), dialogue (dos hablantes) o conversation (más de dos hablantes).

@Length: duración temporal de la grabación, en horas, minutos y segundos.

@Words: número de palabras de la transcripción.

@Acoustic_quality: calidad acústica de la grabación: A (grabación digital), B (grabación analógica con micrófono de calidad, sin ruido de fondo...) o C (grabación analógica de mala calidad, con ruido de fondo...). En el corpus MAVIR, realizado con una grabadora digital DAT (Digital Audio Tape), todas las grabaciones poseen una calidad acústica A.

@Transcriber: transcriptor de la grabación

@Revisor: revisor(es) de la transcripción

@Comments: indica algún comentario referente a la transcripción.

Marca	Descripción	Significado	Ejemplo
/	Interrupción prosódica no terminal y no autónoma	Separa unidades tonales que no tienen autonomía sintáctica ni semántica (una palabra o sintagma). La entonación suele ser ascendente, porque la oración está inacabada.	*COR: el ocho por ciento / el siete coma nueve por ciento de los internautas del mundo / son hispanohablantes (mavir03)
//	Interrupción prosódica no terminal y autónoma.	Separan dentro del discurso aquellas unidades que tienen autonomía oracional y semántica. Indica una pausa más marcada, después de la cual se prosigue la unidad informativa. Suele separar fragmentos de estilo directo dentro de un enunciado mayor, preguntas aclaratorias que no rompen el sentido del enunciado en que se enmarcan, incisos y paréntesis, la coda final de un enunciado tras un fragmento que parece haber concluido y se prosigue tras una pausa ...	Ej. *ROD: bueno / muchas gracias a todos // por la presentación (mavir11) *ROD: al usar implicación textual // como he dicho que tenemos texto y -> [/] e hipótesis // vemos que / nos falta la hipótesis (mavir11) *JGO: vais a decir // no hombre hhh {&act: blow} // qué tontería // se hace / traducción automática y ya está (mavir13)
///	Interrupción prosódica terminal; entonación descendente.	Separa unidades informativas completas desde un punto de vista semántico, sintáctico y pragmático, con su propio significado ilocutivo. La entonación suele ser descendente e indica fin de enunciado. La marca sólo se usa en los documentos .TXT.	*COR: y / ayer recordaba una cosa en cuanto a lo del tema del mercado potencial // que ahora hablando Isabel Aguilera / hhh {&act: click} / me ha ayudado a cambiar /// (mavir03)
¡!	Interrupción prosódica terminal; entonación exclamativa.	Indican la modalidad de enunciación exclamativa. Separamos los signos exclamativos del signo contiguo, salvo las interjecciones, que no están separadas de los signos exclamativos.	*COR: ¡ es curioso ! (mavir03) *COR: ¡je! / ¡ si es casi imposible ! (mavir03)

Marca	Descripción	Significado	Ejemplo
¿?	Interrupción prosódica terminal; entonación interrogativa	Indican la modalidad de enunciación interrogativa. Escribimos los signos interrogativos separados del carácter contiguo; pero algunos marcadores discursivos que se acompañan de interrogación aparecen con los signos interrogativos sin espacios separados de la palabra.	*COR: Domingo / pregunta / Domingo / ¿ tenéis herramientas reales para saber / lo que quiere el usuario en Internet ? (mavir03) *COR: porque esos productos se van a comprar también en nuestro país / ¿eh? (mavir03)
...	I Interrupción prosódica terminal; entonación suspendida.	Marcan enunciados que no están acabados sintáctica ni semánticamente porque su significado resulta evidente para los hablantes, o aparece implícito en el contexto.	*COR: explícale tú a ese presidente de compañía lo importante que es tener una web corporativa útil y rastreable ... (mavir03) *COR: esto lo he discutido / de verdad / me [/] me preguntaban al principio // ¿ podemos grabar la... ? (mavir03)
=	Autointerrupción	El hablante interrumpe el enunciado por decisión propia: de repente se le ocurre otra idea, cambia de tema rápidamente sin concluir el anterior...	*ENR: y estamos en una posición fuerte // como empresa // de cara a posibles = yo &cre [/] no [/] no nos hemos preocupado pues sencillamente de no hacer accesible la información para que no sea transmisible {%alt: transmitible} fácilmente (mavir02)
+	Interrupción	El hablante es interrumpido por otra persona. El enunciado puede finalizar ahí, o puede seguir mientras habla quien ha interrumpido, o después de que éste ha concluido. En este caso, empleamos el signo ¬ al principio del turno del hablante interrumpido.	*LRO: o sea como en el caso de xxx + {%com: loud sound emitted by a microphone} *VOX: {%com: sound test} hola / buenas tardes / ¿ me escucháis ? (mavir02)

Marca	Descripción	Significado	Ejemplo
¬	Continuación de turno tras interrupción o solapamiento	Cuando el hablante es interrumpido por otra persona, y el enunciado continúa después de que éste ha concluido, empleamos este signo al principio del turno del hablante interrumpido.	*IRA: y se gastan una porrada de millones / *ENR: no /// no {%com: everybody laughs} /// *IRA: {%com: everybody laughs} ¬ se gastaron una porrada [/] poyada [/] porrada de millones con la gente de AutomaticsTrans / para no conseguir nada /// (mavir02)
[/]	Reformulación o reinicio léxico o sintagmático	Debido a un falso inicio, el hablante repite un término o un sintagma, o lo reconstruye si se ha equivocado.	*COR: y es // en [/] en una [/] en la historia de la música pop / de un escritor / que se llama Nick Cohn / a -> finales de los / sesenta principios de los setenta / planteaba una tesis muy interesante (mavir02)
[///]	Reformulación o reinicio sintáctico.	Marca una interrupción en el plano sintáctico, no en el semántico. El hablante interrumpe el enunciado para comenzar otra estructura (sin repetir la anterior) que expresa de mejor forma lo mismo que pretendía comunicar antes.	*COR: el japonés tiene hoy en día [///] los japoneses tienen una penetración / de Internet / del sesenta y siete por ciento (mavir03)
->	Alargamiento vocálico o consonántico a final de palabra	El sonido se alarga (porque se piensa mientras se habla, por vacilaciones...)	*COR: &eh / quiero agradecer a la UNED que -> / nos invite a participar en [/] en esta jornada MAVIR (mavir03)
< >	Solapamiento	Cuando dos hablantes hablan a la vez, se indican entre estos signos las palabras emitidas al mismo tiempo. El enunciado del segundo hablante que se solapa aparece precedido del signo [<].	*LRO: sí / además el cliente tiene que entender // no solo que tiene un problema // sino que / tu tecnología es capaz *JSL: < es capaz de resolvérselo > *LRO: [<] <de resolvérselo> (mavir02)

Marca	Descripción	Significado	Ejemplo
xxx	Palabras ininteligibles	Señalan pasajes que no han podido transcribirse claramente.	*IRA: xxx es [///] eso es te dura un año no te dura más ¿eh? (mavir03)
&	Ante palabra incompleta o incorrecta.	Precede a la palabra no terminada de pronunciar (debido a una autocorrección, una autointerrupción...), o a una palabra agramatical cuya forma normativa no se puede reconstruir, o no parece apropiado hacerlo.	*COR: y pocas veces &u [/] utilizo una alternativa a Google cuando &ha [/] hago una búsqueda (mavir03) *COR: y estuvimos durante quince días / haciéndole cargas de profundidad // &comparando [/] comparando / diferentes idiomas // el inglés / el francés / el alemán ... (mavir03)
hhh {%act: }	Signos paralingüísticos	Indica un gesto o comportamiento no lingüístico del hablante: tos, risa, asentimiento, chasquido, movimiento de una parte del cuerpo, una señal a otra persona ... Se indica el tipo de signo inmediatamente después de la marca hhh , en un comentario entre llaves, precedido de %act .	*COR: y un segundo ámbito / hhh {%act: click} / los propios buscadores (mavir03) *COR: dos meses y medio -> o / tres hhh {%act: laugh} / en la compra de un producto (mavir03)
&eh &ah &mm &er &uh &um ...	Apoyos vocálicos	El hablante los emplea como forma de mantener el turno de habla mientras se piensa, se encuentra una palabra o una manera correcta de expresarse... Se ha intentado emplear el correspondiente a cada lengua: en español, &eh, &eh o &mm; en inglés, &er, &um, &uh, etc.	*ANT: &eh para registrarlo / lo que te piden es que les des el código fuente // con lo cual / &mm te dicen que lo va a custodiar la administración // (mavir02) *ENR: no yo [///] nosotros nuestros clientes son en general &ah [/] son todos empresas privadas (mavir02)

Marca	Descripción	Significado	Ejemplo
#	Pausa no prosódica	Señala una pausa en el discurso, sin intención expresiva: para consultar notas, mirar el ordenador, dar tiempo al auditorio...	*COR: pero no con herramientas sofisticadas / con herramientas sencillas # {%com: consults the laptop} (mavir03)
{%alt: }	Errores de producción y fenómenos no normativos	Indica cualquier error de pronunciación o palabra agramatical, debido al relajamiento propio del habla oral o al desconocimiento por parte del hablante. Se transcribe correctamente la palabra y se translitera el sonido entre corchetes, indicado por la marca %alt .	*COR: y / yo dije hhh {%act: blow} / esto {%alt: to} [/] esto {%alt: to} tiene alcance (mavir03) *COR: se han dado {%alt: dao} cuenta (mavir03)
{%com: }	Comentarios	Comenta un suceso o una circunstancia del contexto comunicativo: cuando otras personas aparte del hablante ríen, si suena un sonido o ruido imprevisto, si se corta la grabación, si alguien lee un texto...	*COR: es que no me podía resistir / hhh {%act: laugh} {%com: everybody laughs} (mavir03)