

ADAPTACIÓN DEL FORMATO CHAT PARA LA TRANSCRIPCIÓN Y CODIFICACIÓN DE LOS DATOS DE LENGUAJE PATOLÓGICO DEL CORPUS PERLA¹

MAITE FERNÁNDEZ-URQUIZA
Universidad de Oviedo

BEATRIZ GALLARDO PAÚLS
Universitat de València

RESUMEN

En este trabajo se ofrece una adaptación del formato de transcripción y codificación CHAT (Codes for the Human Analysis of Transcripts) que ha sido específicamente desarrollada para posibilitar el trasvase del corpus PerLA de lenguaje patológico a la base de datos internacional TALKBANK. Para ello, se requiere la adopción de un formato estandarizado de transcripción y codificación compatible con el software CLAN (Computerized Language ANalysis) de análisis automatizado de lenguaje. Este formato admite, sin embargo, un grado considerable de variación en la elección de las convenciones empleadas, en función de la naturaleza de los datos, la perspectiva del estudio para el que el corpus haya sido elaborado, y los fenómenos que se considere relevante codificar. La adaptación que ofrecemos persigue resultar de utilidad para la codificación de muestras de lenguaje que reflejan un rango considerablemente amplio de déficits lingüísticos, con el fin de realizar estudios de naturaleza pragmática.

PALABRAS CLAVE: *lingüística de corpus, lingüística clínica, pragmática, corpus PerLA, CHAT, Talkbank.*

ABSTRACT

The aim of this paper is to describe a version of the CHAT (Codes for the Human Analysis of Transcripts) transcription format developed with the purpose

¹ Esta publicación se realiza en el marco de los proyectos de investigación “Capacidades metalingüísticas y función ejecutiva en afasias y trastornos relacionados. Estudio interdisciplinar y evaluación clínica” (Ref. FFI2011-25947) y “Ampliación y adaptación del corpus PERLA de datos clínicos en el marco del proyecto internacional CHILDES/TALKBANK: perfiles pragmáticos y propuestas de intervención” (Ref. FFI2012-39325-C03-01), ambos del Ministerio de Economía y Competitividad.

of enabling open-access to the PerLA corpus of language pathologies through the Talkbank International Database. In doing so, adapting PerLA to a transcription and coding format compatible with CLAN (Computerized Language ANalysis) software for automatic language analysis was required. However, CHAT conventions are not compulsory, but they allow for a wide range of variations depending on the nature of data, the scope of the study for which the data have been collected, and the kind of phenomena that are considered relevant for coding. The adaptation offered here has been envisaged to allow the transcription and coding of data showing a wide variety of language and communication disorders, in order to carry out further studies from a pragmatic perspective.

KEYWORDS: *corpus linguistics, clinical linguistics, pragmatics, PerLA corpus, CHAT, Talkbank.*

0. INTRODUCCIÓN

En este trabajo proponemos una adaptación del formato CHAT (MacWhinney, 2014a) específicamente desarrollada para llevar a cabo la transcripción y codificación del corpus PerLA de lenguaje patológico (Gallardo Paúls, 2002) con el objeto de ponerlo a disposición de la comunidad de investigadores en la base de datos internacional TALKBANK (<<http://talkbank.org/>>), lo que requiere la adopción de un formato estandarizado compatible con el software CLAN (Computerized Language ANalysis) de análisis automatizado de lenguaje (MacWhinney, 2014b).

A lo largo de la última década, y en el marco de la elaboración, transcripción y codificación del corpus PerLA de datos clínicos, hemos desarrollado los ajustes necesarios para cada uno de los niveles de representación de los datos. El primer sistema adoptado fueron las convenciones del modelo de los analistas de la conversación norteamericanos, propuestas por Jefferson (1984), adaptadas para reflejar con precisión los abundantes fenómenos de *code-switching* presentes en un corpus bilingüe español-catalán.

El proceso de adaptación del formato CHAT (Codes for the Human Analysis of Transcripts) para PerLA comenzó en enero de 2013. Dado que gran parte de los símbolos disponibles en CHAT son opcionales, en el sentido de que el sistema ofrece diversas posibilidades para codificar un mismo fenómeno, y dado que el corpus PerLA recoge muestras de lenguaje que reflejan un rango considerablemente amplio de déficits lingüísticos, ha sido necesario realizar una selección de las convenciones de CHAT encaminada a conseguir la consistencia en la transcripción y codificación de datos ciertamente heterogéneos. Al mismo tiempo, se han perseguido la claridad y la sencillez tanto en la transcripción como en la codificación de categorías, de modo que la legibilidad de los datos se vea facilitada.

La funcionalidad de este documento, en el que explicitamos la motivación para el uso de cada una de las convenciones y categorías, reside en constituir una guía para los investigadores que, en un futuro próximo, decidan trabajar con muestras de PerLA², permitiendo la correcta interpretación de los datos ofrecidos. Por otra parte, nuestra propuesta de transcripción y codificación queda a disposición de los investigadores que deseen ofrecer sus propios datos clínicos en Talkbank, con el único requerimiento de que hagan mención a la misma tal y como establecen los procedimientos de cita habituales.

1. SELECCIÓN DE CONVENCIONES PARA LA DESCRIPCIÓN DE CADA TRANSCRIPCIÓN

1.1. *Denominación unificada de archivos*

El trabajo coordinado entre equipos de diferentes universidades³ para la adaptación y ampliación de PerLA requiere, en primer lugar, la adopción de un sistema unificado de denominación de archivos. CHAT genera archivos diseñados para ser analizados con el software CLAN (MacWhinney, 2014b), cuya extensión es .cha.

El formato de denominación de cada una de las muestras de PerLA proporciona información sobre los siguientes elementos:

1. La pertenencia del archivo al corpus PerLA se indica mediante el código PERLACH.
2. A continuación se incluye el número de la grabación original en PerLA, el cual depende del tipo de déficit. Así, las afasias se numeran en el rango del 100, el TDAH (Trastorno por Déficit de Atención y/o Hiperactividad) en el 200, los LHD (Lesionados de Hemisferio Derecho) en el 300, el Síndrome de Asperger en el 400, las DTA (Demencias Tipo Alzheimer) en el 500, los Trastornos de Espectro Autista (TEA) en el 600, los déficits ocasionados por síndromes de base genética como Down o Williams con el 700, las disfemias con el 800, y otras situaciones con el 900.
3. Lo siguiente en aparecer son las tres iniciales del informante, en letras mayúsculas.
4. El tipo de déficit se especifica a continuación, del modo siguiente:

² Estimamos que el corpus estará disponible en TALKBANK a mediados de 2015.

³ En la actualidad son siete equipos los que trabajan de manera coordinada en las universidades de Almería, Bristol, Burgos, Castilla-La Mancha, Florida-Universitaria, Oviedo y Valencia.

- Aflu: Afasia fluente
- ANoflu: Afasia no fluente
- Lhd: Lesionados de hemisferio derecho
- Dta: Demencias tipo Alzheimer
- SAsp: Síndrome de Asperger
- Tdah: Trastorno por Déficit de Atención y/o Hiperactividad

Así, por ejemplo, el nombre de archivo PERLACH302.IMM.Lhd.cha nos permite saber que estamos manejando la transcripción en formato CHAT de los datos correspondientes a la segunda grabación de lesionados de hemisferio derecho del corpus PerLA, cuyo informante es IMM.

1.2. Las líneas de cabecera (headers)

Son líneas que codifican los datos principales en relación con la grabación, como los nombres y roles de los participantes en la conversación, el lugar en que se realizó, las actividades llevadas a cabo durante la misma, su duración, etc. Las líneas de cabecera se identifican porque van precedidas del símbolo @⁴. A continuación se ofrece un ejemplo del aspecto que debe tener todo inicio de transcripción en PERLACH:

```

/Users/maite/Dropbox/PROYECTO VALENCIA
@Begin
@Languages: spa
@Participants: IMM Target Adult, HIJ Isabel Adult, MAR Adult, BGP
                BeatrizGallardo Investigator
@Options: CA
@ID: spa|PerLA|IMM|70;05.10|female|Lhd||Target_Adult||
@ID: spa|PerLA|HIJ||female||Adult||
@ID: spa|PerLA|MAR||male||Adult||
@ID: spa|PerLA|BGP||female||Investigator||
@Transcriber: Beatriz Gallardo-Paúls, Maite Fernández-Urquiza
@Time Duration: 00:09:52
@Date: 16-FEB-2009
@Location: Valencia, Comunidad Valenciana, Spain
@Situation: Hospital room
@Activities: Colloquial conversation, Boston cookie theft test
@Comment: Right Brain Damage

```

En la anterior imagen se puede observar la ventana del editor de textos de CHAT/CLAN. La línea de cabecera @Begin indica al programa que se trata de un documento que contiene datos para el análisis. El final de la transcripción se señala mediante una línea de cierre @End. Los idiomas

⁴ Prescindimos del uso de comillas y cursivas en la denominación de las convenciones de transcripción en este trabajo, ya que ambas pueden constituir parte de los códigos, por lo que resulta necesario evitar equívocos.

utilizados en la grabación se codifican en la línea @Languages: mediante los códigos disponibles en el documento ISO-639.cut que se encuentra por defecto en la carpeta *lib* de CLAN, dentro de la subcarpeta *fixes*. Así, por ejemplo, *spa* (español), *eng* (inglés), *por* (portugués), *glg* (gallego), etc. La línea @Participants codifica las iniciales de los participantes y el rol desempeñado por los mismos. La lista de roles puede consultarse en el *depfile.cut* disponible en la carpeta *lib* de CLAN. La línea @Options: CA indica al sistema que van a utilizarse algunos símbolos específicos del Análisis Conversacional. Las líneas @ID codifican datos relativos a cada participante: lengua(s) empleada(s), nombre del corpus de origen, iniciales, edad, sexo, y rol desempeñado durante la conversación. En la línea @Transcriber señalamos en primer lugar el nombre de quien realizó la transcripción original según las convenciones PerLA y, a continuación, el nombre de la persona responsable de la transcripción y codificación en CHAT. Esto, además de servir como reconocimiento de su trabajo, facilitará las consultas que otros investigadores puedan tener sobre los datos en el futuro, cuando se encuentren disponibles en TALKBANK. Las líneas @Time Duration y @Date indican respectivamente la duración total de la grabación y la fecha en que fue realizada, mientras que la línea @Location proporciona los datos sobre la ciudad, la región y el país. La línea @Situation describe las características del lugar en que se efectúa la grabación, mientras que la línea @Activities nos permite señalar las tareas y actividades realizadas durante la misma. Por último⁵, la línea @Comment puede incluir cualquier información que se considere relevante. Aquí incluimos el tipo de déficit y, en caso de que este sea adquirido, el tiempo transcurrido desde el accidente cerebrovascular, si es que se dispone de ese dato, como en el ejemplo siguiente:

```

/Users/maite/Dropbox/PROYECTO VA
@Begin
@Languages:    spa, cat
@Participants: DMP Target_Adult, HIJ Luis Adult, BGP
               BeatrizGallardo Investigator
@Options:      CA
@ID:           spa, cat|PerLA|DMP|77;0.|female|Lhd||Target_Adult|||
@ID:           spa, cat|PerLA|HIJ||male|||Adult|||
@ID:           spa, cat|PerLA|BGP||female|||Investigator|||
@Transcriber:  Beatriz Gallardo Paúls, Maite Fernández-Urquiza
@Time Duration: 00:08:49
@Date:         13-FEB-2009
@Location:     Valencia, Comunidad Valenciana, Spain
@Situation:    Informant's home
@Activities:   Colloquial conversation, Boston cookie theft test,
               storytelling
@Comment:     Date of stroke JAN-2009. Right Brain Damage.

```

⁵ CHAT ofrece la posibilidad de incorporar más líneas de cabecera, que han de ser seleccionadas de acuerdo a los objetivos de la investigación y características del corpus. Pueden consultarse en MacWhinney (2014a: 24-35).

2. SELECCIÓN Y ADAPTACIÓN DE CONVENCIONES EN LÍNEAS PRINCIPALES

2.1. Las líneas principales (*main tiers*)

Son las líneas que contienen las intervenciones de los participantes en la conversación. Cada una de ellas va introducida por * seguido de las iniciales del informante, dos puntos, y tabulador. El final de cada una de ellas ha de ir delimitado por un signo de final de enunciado. Estos signos son la exclamación, la interrogación y el punto final, siempre separados por espacio. Ante la dificultad de unificar criterios en torno a lo que puede considerarse un enunciado o no desde el punto de vista pragmático, en los archivos PERLACH hemos decidido segmentar la intervención de un mismo hablante en no más de dos renglones por línea principal, de modo que se pueda etiquetar posteriormente cada fenómeno con claridad en líneas dependientes⁶.

Así, por ejemplo, en el siguiente fragmento se observa cómo se ha segmentado un turno de habla de la informante *FVC:

```
*MVR: eso no es un fallo de memoria eso es que tenemos veintimil cosas en
la cabeza .
*fvc: claro sí es que tengo uno que es (...) [/ -] que está en la mi [/ -]
ay:@i (.) es (.) soldado convencional &*MVR: uhum .
%xepr: $a:vl vp vp vl vp $i:rp rp rp
*fvc: y va allí y [/] y viene como está en marines (.) <viene y va>
[/] viene y va (...) y muchas veces me dice +"/.
%xepr: $a:vp vp vp vp vp $i:rp rp rp rp rp
*fvc: +" mamá † voy a comer .
%xepr: $a:vp $i:rp
*fvc: y yo a lo mejor no he hecho comida pa@d él .
%xepr: $a:vp $i:rp
%xipr: $t:ast ret rpt
*MVR: ya no cuenta con él claro (.) .
```

A continuación comentamos en detalle las convenciones de transcripción adoptadas para las líneas principales.

⁶ Un sistema de etiquetado pragmático compatible con CLAN y basado en el *Perfil PerLA de Evaluación Pragmática* (Gallardo Paúls, 2006) está siendo desarrollado paralelamente. Se trata de un modelo de corte funcional y perceptivo que organiza las categorías pragmáticas en tres niveles, a saber: enunciativo, textual, e interactivo (Gallardo Paúls, 2007: 45). Se ha completado ya el desarrollo de los códigos correspondientes a las etiquetas de los niveles enunciativo e interactivo (Gallardo Paúls y Fernández-Urquiza, 2014, en prensa), mientras que los del nivel textual se encuentran en desarrollo.

2.2. *La introducción de símbolos CA (Conversational Analysis)*

CHAT proporciona la posibilidad de utilizar algunas de las convenciones de transcripción propias del análisis conversacional etnometodológico (MacWhinney, 2014a: 86). En PerLA hemos seleccionado el símbolo ↑ para las anticadencias, ↓ para las cadencias, y → para las entonaciones suspendidas. El empleo de estos códigos se encuentra condicionado a la aparición de patrones entonativos en la grabación que violenten las expectativas prosódicas habituales. En otras palabras, los delimitadores de final de enunciado codifican por sí mismos un patrón entonativo determinado, a saber: ascendente para la interrogación, ligeramente ascendente o sostenido para la exclamación, y descendente para el punto (MacWhinney, 2014a: 60). Nuestra decisión obedece, por tanto, al deseo de eliminar las marcas de entonación redundantes con los patrones melódicos esperables. Esto puede ocurrir tanto a final de enunciado como a lo largo del mismo, como cuando un patrón melódico queda suspendido durante una pausa oralizada, por ejemplo.

2.3. *Marcadores en el nivel de la palabra*

Para sistematizar la codificación de las diversas situaciones en que el audio resulta al transcriptor un discurso carente de sentido, hemos incorporado el símbolo @n para los neologismos.

Para la codificación de los elementos cuasi-léxicos (Poyatos, 1995), hacemos uso del símbolo @i para las interjecciones, y @o para las onomatopeyas (MacWhinney, 2014a: 37-41). Ambos se colocan al final de la palabra en cuestión, del modo siguiente:

*BGP: setenta (.) muy bien lleva(d)os@d „ eh@i ?

[PERLACH302.IMM.Lhd: 23]

Cuando una onomatopeya se compone de varios elementos, puede optarse por vincularlos mediante el símbolo +, así: guau+guau@o.

Por otra parte, los significantes que son manifestación de una determinada variedad y/o registro lingüísticos, se marcan mediante el código @d, como ocurre frecuentemente en nuestro corpus con la elisión de [d] intervocálica en los participios pasados, fenómeno muy habitual en el habla coloquial. A la par, los casos de elisión pueden ser completados por el transcriptor mediante el uso de paréntesis, como se observa en el ejemplo propuesto arriba.

2.4. *Codificación de grabaciones con cambios de código*

Cuando el hablante realiza cambios habituales de una lengua a otra, intercalando palabras o frases de su segunda lengua, estas pueden marcarse con @s, del modo siguiente⁷:

*CHI: this is my juguetes@s .

En este ejemplo, el símbolo @s está indicando que el hablante ha introducido un enunciado en la segunda lengua de las que han de ir listadas en la línea de cabecera @Languages. Es decir, el símbolo @s alcanza aquí al enunciado completo, y su presencia indica un cambio al inglés con respecto al enunciado precedente, a pesar de que la última palabra esté en español. Una forma más clara de notar estos cambios consiste en utilizar el precódigo [- eng] para indicar a qué lengua cambia el hablante, y el signo @s para indicar un cambio subsumido en el anterior, así:

*CHI: [- eng] this is my juguetes@s .

En este caso, el signo @s alcanza solo a la palabra a la que se anexa, e indica que hay un cambio al español subsumido en un enunciado que, a su vez, supone un cambio al inglés con respecto a la línea precedente. Para mayor claridad aún, puede especificarse a qué lengua se realiza el cambio añadiendo dos puntos y el código ISO de la lengua en cuestión a continuación de @s:

*CHI: [- eng] this is my juguetes@s:spa .

*CHI: this is my juguetes@s:eng .

Hemos seleccionado esta convención por su utilidad para indicar los cambios de una lengua a otra que se producen en algunos casos de afasia, en los que el hablante puede alternar hasta entre cuatro lenguas diferentes.

2.5. *La codificación de parafasias*

El capítulo 15 del manual CHAT (MacWhinney, 2014a: 100), recoge un sistema de codificación de errores que permite dar cuenta tanto de parafasias fonológicas como semánticas, proporcionando también la opción de distinguir subclases dentro de cada categoría. Cualquier tipo de error en el nivel de la palabra se codifica en CHAT mediante la presencia de [*] a continuación del ítem léxico en cuestión. Este uso del asterisco entre corche-

⁷ Reproducimos un ejemplo de MacWhinney (2014a: 26), por su claridad explicativa.

tes puede especificarse añadiendo una letra identificadora del tipo de error que ha cometido el hablante, ya sea fonológico [* p] o semántico [* s].

De este modo, las parafasias fonémicas quedan recogidas por [* p]. Ahora bien, si debido al tipo de déficit o a los objetivos de la investigación, interesase especificar subtipos de parafasia fonémica, existen las siguientes posibilidades:

- Cuando la sustitución se realiza por una palabra conocida, se utiliza [* p:w].
- El marcador [* p:n] indicaría que la parafasia da lugar a una pseudo-palabra (*non-word*).
- El marcador [* p:m] señala la existencia de metátesis.

En todos los casos, el target (en caso de que se pueda inferir) puede indicarse en una línea dependiente %err:, así:

```
*FVC: lo he pasado muy mal (..) los pocos días que estaba.
*fvc: y gracias a mi hija (.) la pequeña (.) que <me ha du> [/] me ha du
      [/-] dudao [* p:w] mucho (.).
%err: dudao = ayudado
```

[PERLACH303.FVC.Lhd: 192-195]

Las parafasias semánticas quedan recogidas por [* s]. Al igual que en el caso anterior, es posible realizar una codificación más detallada:

- El código [* s:r] indica que la parafasia producida es una palabra real, conceptualmente relacionada con el target (*related*).
- El símbolo [* s:ur] indica que la parafasia producida es una palabra real, pero sin relación con el target (*unrelated*).
- El uso del marcador [* s:uk] indica que la parafasia es una palabra real, pero sin sentido en el contexto, aunque el target se desconoce (*unknown*).
- Por último, el código [* s:per] señala las perseveraciones.

En todos estos casos, el target puede indicarse, si se conoce, en una línea %err:, como ya se ha ejemplificado.

2.6. La codificación del déficit morfológico

Los errores morfológicos en el nivel de la palabra se codifican en CHAT mediante el código [* m]. Dado que los submarcadores propuestos por CHAT son específicos para la morfología de la lengua inglesa, en PERLACH prescindimos de ellos y señalamos el target en línea %err:, al igual que para los componentes fonológico y semántico. Por ejemplo:

- *FVC: era mi yerno (...) [=! cough].
 *FVC: y ya voy a beber agua (.) el vaso me se sr [/ -] resbaló y se
 caí [* m] (.). [+ gram]
 %err: caí = cayó

[PERLACH303.FVC.Lhd: 165-168]

El postcódigo [+ gram] indica la presencia a nivel de enunciado de algún error de tipo gramatical. El análisis detallado de estos fenómenos, si interesa, se explicita posteriormente en la fase de etiquetado en una línea dependiente %mor:, %syn:, o %gra:.

2.7. *Otras convenciones frecuentes en nuestro corpus*

2.7.1. Habla ininteligible

El habla ininteligible se transcribe mediante el símbolo xxx. El manual de CHAT (MacWhinney, 2014a: 41) señala que este símbolo corresponde a situaciones en que no se oye o no se entiende lo que dice el hablante, y apunta la posibilidad de incluir varias cadenas xxx si se cree que la emisión puede segmentarse en varias palabras, como en:

- *HIJ: tú <xxx xxx> [?].

[PERLACH302.IMM.Lhd: 154]

2.7.2. Casos de transcripción dudosa

En casos de transcripción dudosa, se añade [?] al lado del fragmento objeto de duda. Si el fragmento consta de más de una palabra, estas han de ir entre corchetes triangulares, indicando así el alcance del símbolo [?]. Por ejemplo:

- *IMM: <casi [/] casi> [?] nunca [=! shakes her head no].

[PERLACH302.IMM.Lhd: 424]

2.7.3. Fragmento no transcrito

En algunas ocasiones, el transcriptor puede decidir omitir un fragmento de la conversación grabada, bien por considerarlo de escaso interés de cara a la detección de fenómenos de estudio, o bien porque en él se abordan temas cuya publicación podría vulnerar la intimidad de los informantes. En este caso, el símbolo www ocupa el lugar de los enunciados realmente emitidos por cada informante, así:

- *IMM: www .

2.7.4. Turnos sin habla

Los turnos sin habla (o “acciones sin habla”, en MacWhinney, 2014a: 42), se codifican mediante el símbolo 0 (numeral cero). De este modo se indica la existencia de una toma efectiva de turno, a pesar de que su valor interactivo no se apoya en la emisión verbal.

Para que estos turnos sean tenidos en cuenta en los análisis con CLAN, se utiliza el postcódigo “incluir turno” [+trn], después del marcador de final de enunciado, separado por un espacio. Esto indica a los programas de análisis que contabilicen los actos no verbales realizados por el informante como turno conversacional. En el siguiente ejemplo se observa también que los actos no verbales constitutivos de turno se codifican en línea principal entre corchetes, introducidos por el símbolo =! :

*HIJ: mira con qué peloh@d te van a sacar .

*IMM: 0 [=! tries to fix her hair up] . [+trn]

[PERLACH302.IMM.Lhd: 31-32]

2.7.5. Fenómenos intra-turno

CHAT ofrece una serie de símbolos para codificar las repeticiones, reinicios, reformulaciones, y falsos comienzos, que resultan claramente redundantes. Los símbolos [/], [//], [///] y [/-] descritos en MacWhinney (2014a: 73-75) presentan funcionalidades superpuestas. Así, por ejemplo, se propone [//] para *Retracing*, es decir, casos en que el hablante inicia su emisión, se detiene, y repite lo esencial aunque cambia la sintaxis, pero mantiene la misma idea o intención discursiva, ya sea incluyendo autorrectificaciones o sin ellas. El símbolo [///] se asigna a las *Reformulations*, para las que se señala también la doble posibilidad de incluir o no autorrectificaciones explícitas. El símbolo [/] indica la repetición de una palabra o cadena de palabras (el alcance del código se delimita siempre acotando el material repetido entre corchetes triangulares). Por último, el símbolo [/-] se atribuye a *False Start Without Retracing*, es decir, al abandono de cierta emisión para comenzar un nuevo hilo discursivo.

En PERLACH hemos decidido simplificar estas convenciones y utilizar el código [/-] para los reinicios de cualquier tipo y [/] para las repeticiones, como en los ejemplos siguientes:

*FVC: +^ y lo tuve [/-] (.) me na [/-] el primero me nació muerto (.) .

[PERLACH303.FVC.Lhd: 37]

*FVC: y va allí y [/] y viene como está en marines (.) <viene y va>
[/] viene y va (..) y muchas veces me dice +”/.

[PERLACH303.FVC.Lhd: 137-138]

Para los prefacios (vocativos iniciales, temas en estructuras informativas del tipo tema/rema) y las preguntas confirmatorias de final de turno, decidimos utilizar los que el manual designa como *marcadores para elementos satélite* (MacWhinney, 2014a: 59).

Los prefacios se señalan mediante el código ‡, del modo siguiente:

*FVC: +” (.) mama ‡ hoy no vengo a comer .
[PERLACH303.FVC.Lhd: 125]

Las preguntas confirmatorias se delimitan con el símbolo „ :

*MVR: pues nada usted ya ha tenido lo suyo „ eh@i ?
[PERLACH303.FVC.Lhd: 549]

2.7.6. Omisiones, sustituciones y pausas

Las omisiones se codifican mediante el símbolo 0word, aunque conviene señalar que cualquier juicio sobre la ausencia de cierto elemento no deja de ser una suposición por parte de quien transcribe, muchas veces cuestionable (MacWhinney, 2014a: 43). La omisión de un elemento puede señalarse en línea principal indicando la palabra que se considera que falta, o bien su categoría gramatical (por ejemplo: 0de o 0prep). La segunda opción resulta más interesante de cara a la fase de etiquetado de este tipo de fenómenos, que está en vías de desarrollo y posibilitará un recuento automatizado de las categorías etiquetadas, de manera que podamos calcular cuáles son las que se ven más afectadas según el tipo de déficit.

Sin embargo, salvo casos en los que la ausencia de ciertas partes de la oración es muy clara, puede ser muy arriesgado tratar de codificar en detalle qué es lo que está pasando a nivel morfosintáctico. Por ello, en MacWhinney (2014a: 104) se propone, para los casos de enunciados con errores en el orden de palabras u otros errores de tipo morfosintáctico (incluyendo el discurso agramático y paragramático) utilizar el postcódigo [+ gram], como se puede ver en el ejemplo siguiente, que presenta una inversión del orden de los pronombres átonos:

*FVC: y ya voy a beber agua (.) el vaso me se sr [-] resbaló y se
caí [* m] (.) . [+ gram]
[PERLACH303.FVC.Lhd: 165-166]

En PERLACH hemos decidido utilizar este postcódigo hasta que hayamos desarrollado un sistema de etiquetado del déficit morfosintáctico acorde a las categorías funcionales del *Perfil PerLA de Evaluación Pragmática* (Gallardo Paúls, 2006).

Como ya hemos comentado en detalle en 2.5. y 2.6., las sustituciones se codifican marcando la palabra errónea con [*] en línea principal, y añadiendo a continuación el símbolo del tipo de error (fonético, semántico, morfológico) cuyo target puede especificarse en línea dependiente %err, como ya se ha explicado.

Las pausas vacías se transcriben utilizando de uno a tres puntos entre paréntesis, en función de su duración estimada. Pueden indicarse también los segundos de duración de la pausa cuando son excepcionalmente largas, también entre parentésis. Por ejemplo:

*FVC: sí (...) ah@i (.) cuarta la de blanco .
[PERLACH303.FVC.Lhd: 55]

Respecto a los signos de puntuación, conviene recordar que las transcripciones no deben contener puntos, comas, puntos y comas, ni dos puntos, tal y como se usan en la ortografía convencional. El uso del punto se reserva como marcador de final de enunciado. Las comas pueden aparecer combinadas con otros símbolos que iremos viendo, para codificar fenómenos específicos. Los dos puntos se utilizan para indicar los alargamientos vocálicos. Las letras mayúsculas sólo se utilizan para la primera letra de los nombres propios.

2.7.7. Algunos elementos suprasegmentales

Como acabamos de señalar, los alargamientos vocálicos se codifican mediante el uso de los dos puntos ortográficos a continuación de la vocal, como en:

*FVC: <sí> [<] hoy este mediodía he subido a: [/] al del siete (.) un kilo de moros (.) y un kilo de: chuletas (.) un kilo de: xxx.
[PERLACH303.FVC.Lhd: 466-467]

De este modo se evita que la aplicación MOR contabilice las palabras con un número de vocales diferente como ítems distintos, o que un análisis con FREQ arroje un número de *types* mayor del real⁸.

La pronunciación silabeada se indica con el símbolo ^ entre cada una de las sílabas:

*MCP: Torre^fiel [=! effortful speech] .
[PERLACH118.MCP1.ANoflu: 117]

⁸ MOR y FREQ son funcionalidades de análisis incorporadas en CLAN (MacWhinney, 2014b), que facilitan el análisis cuantitativo de los datos.

La pronunciación enfática se codifica acotando el fragmento entre corchetes triangulares <> seguidos del símbolo [!]. Si el énfasis recae solo en una palabra los corchetes no son necesarios:

*ANG: <pero lo he superao:> [!] .

[PERLACH103.ANG.Aflu: 59]

2.7.8. Información paralingüística y no verbal

La multiplicidad de posibilidades que CHAT proporciona para codificar información paralingüística y no verbal, tanto en líneas dependientes específicas como en líneas principales (precedida en cada caso de diferentes símbolos) ha motivado que nos decantemos por la que consideramos más sencilla. Así pues, decidimos codificar toda la información paralingüística en línea principal, introduciéndola mediante el símbolo [=]. La información paralingüística incluye las miradas [=! looks at BGP], risas [=! laughs], llantos [=! cries], suspiros [=! sighs], como en:

*ANG: a mí era el único elemento que no me robaban la [/] la moto
[=! laughs] (.) en casa .

[PERLACH103.ANG.Aflu: 277-278]

También incluimos en esta categoría de información cuasi-lingüística los gestos emblemáticos, como asentimientos [=! nods] y negaciones con manos o cabeza [=! shakes his/her head no], así como los gestos deícticos (ilustradores) que permiten identificar los referentes del acto verbal y completar su significado, como por ejemplo:

*ANG: pero: (..) el laboratorio lo tengo aquí [=! points at his head] .

[PERLACH103.ANG.Aflu: 101]

Por tanto, los fragmentos que en PerLA transcribíamos como voz susurrada con el símbolo °()°, pasan a acotarse entre paréntesis triangulares <> seguidos de [=! whispers]:

*ANG: hablo mal (.) <muy mal> [=! whispers] [>] .

[PERLACH103.ANG.Aflu: 64]

Lo mismo ocurre con la pronunciación esforzada, transcrita en PerLA con *()*, que pasamos a transcribir como [=! effortful speech]:

*IMM: (0.6) un <aro> [=! effortful speech] <un a> [/] un aro
[=! points to the picture once and again]?

[PERLACH302.IMM.Lhd: 244-245]

Decidimos codificar la información proxémica en línea dependiente %gpx: . Entendemos que incluye los gestos que acompañan al acto verbal, pero sin un significado *cuasi-lingüístico* en el sentido de Cosnier, por ejemplo:

*IMM: casi siempre: (0.3) la paella cómo (0.2) .

%gpx: moves her hand

[PERLACH302.IMM.Lhd: 403-404]

Por el contrario, los gestos con significado emblemático (como los asentimientos y negaciones con manos o cabeza), o los gestos ilustradores que ayudan a identificar el referente del acto verbal y, por tanto, a completar su significado, se codifican en línea principal mediante el símbolo [=!], como acabamos de señalar.

Decidimos señalar los continuadores del sistema secundario de toma de turno (Gallardo Paúls, 2007) con el símbolo & seguido de las iniciales del informante precedidas de asterisco, y seguidas de dos puntos (&*DMP:) y a continuación la forma que actúa como continuador:

*IMM: casi todos los domingos <me la hace> [/] me la hace ella

[=! points to HIJ] &*BGP: uhhum .

[PERLACH302.IMM.Lhd: 470-471]

Es importante recordar que estas intervenciones secundarias se sitúan en la línea principal del turno precedente, con el que se encuentran en relación, y no en una línea independiente, a diferencia de lo que ocurre en AC.

2.7.9. Fenómenos inter-turno

En este apartado codificamos dos tipos de interrupciones. La auto-interrupción, o abandono de la propia intervención, se indica mediante el símbolo +//., mientras que la interrupción ajena lo hace mediante +/. .

En ambos casos, cuando se trata de sintaxis interrogativa se añade el signo ortográfico convencional de interrogación ?. Por ejemplo:

*MVR: y gracias a ese esfuerzo se ha superado y ha conseguido +//.

[PERLACH.303.FVC.Lhd: 334]

*MVR: ah muy bien (.) yo [-] pues yo pensaba no sé por qué (.) <pensaba que vivía usted sola y yo digo oh@i pues> [>] +/.

*FVC: <no [/] no (..) y mi marido> [<].

*MVR: +, se arregla muy bien para estar sola .

[PERLACH.303.FVC.Lhd: 385-390⁹]

⁹ Ocultamos en el ejemplo las líneas dependientes que no interesan a los propósitos explicativos de este artículo. De ahí que la numeración de las líneas no coincida con el número de líneas reales del ejemplo.

Decidimos no utilizar el símbolo +., asignado en MacWhinney (2014a: 65) a la *transcription break*, pues resulta redundante con la codificación de los solapamientos, que se indican con los símbolos [>] y [<] de apertura y cierre, como puede observarse en el ejemplo anterior. También preferimos no utilizar las *indicaciones visuales opcionales*, es decir, las tabulaciones para situar las cadenas de habla solapadas de manera alineada en la transcripción (MacWhinney, 2014a: 72), ya que parecen variar según el ancho de pantalla. Igualmente prescindimos del uso de *solapamientos solapados* numerados (MacWhinney, 2014a: 72), en virtud de la legibilidad de los datos.

Utilizamos el marcador de autoterminación tras interrupción o *self-completion* +, (MacWhinney, 2014a: 67) en casos en que el interlocutor interviene con un turno competitivo, o en los que interrumpe para introducir una aclaración, y en los que el hablante en posesión del turno continúa hablando después del solapamiento, como también se observa en el ejemplo anterior. Esto es lo que indicábamos en PerLA con el signo = al final y al inicio de los turnos en cuestión. Cuando la terminación del turno es producida por otro interlocutor, es decir, en el caso de *turnos colaborativos* (Gallardo Paúls, 2007), el marcador utilizado es ++, así:

*FVC: digo y: +/.

*MVR: ++ no se lo creía nadie .

[PERLACH303.FVC.Lhd: 573-576]

Los encadenamientos de dos hablantes sin pausa apreciable, que señalábamos en PerLA con el signo §, se codifican en CHAT con el símbolo +^, como en:

*MVR: ++ no se lo creía nadie .

*FVC: +^ claro (.) y a mi marido decía papa (.) .

[PERLACH303.FVC.Lhd: 576-577]

2.7.10. Fenómenos discursivos

Para el estilo directo y el discurso referido de cualquier extensión, utilizamos los símbolos +"/. y +". Este tipo de usos es frecuente en tareas de narración, donde el estilo directo se utiliza para reproducir el discurso de los personajes. Sin embargo, es muy habitual encontrarlo también en conversación espontánea, especialmente en secuencias de historia (Gallardo Paúls y Hernández Sacristán, 2013: 153) desatadas por preguntas abiertas:

*FVC: después cuando me pasó esto (.) no me dijo mi [/] mi
[/-] mis hijos +"/.

*FVC: +” no tienes que cuidar ninguno más (.) .

[PERLACH303.FVC.Lhd: 16-19]

2.7.11. Postcódigos para fenómenos a nivel de enunciado y/o turno

Adoptamos el postcódigo disponible en CHAT para la codificación de perseveraciones que afectan al enunciado completo, y que son especialmente relevantes en el corpus de afasia. Las perseveraciones en este nivel son definidas por Brookshire (1997) como la repetición de un enunciado cuando ha dejado de ser contextualmente apropiado.

En relación con los circunloquios, que se producen por lo general como estrategia compensatoria de fallos de acceso léxico, es frecuente que afecten al turno completo, como se observa en el ejemplo siguiente, donde se aprecia la posición que el postcódigo ([+ cir]) adoptan en la transcripción:

*ANG: +, en el [-] en [-] el [-] en el fenómeno de: (.) e:h e:h (.)
 mecanizar las cosas (..) y esforzarme (.) en el sentido↑ . [+ cir]
 [PERLACH103.ANG.Aflu: 138-139]

3. CONCLUSIONES

En este trabajo hemos ofrecido una adaptación del formato de transcripción y codificación CHAT específicamente diseñada para representar, con la mayor sencillez y claridad posibles, el amplio espectro de fenómenos lingüísticos y comunicativos recogidos por el corpus PerLA de lenguaje patológico.

Por un lado, hemos propuesto una sistematización selectiva en los casos en que existían diversas posibilidades de codificación de un mismo fenómeno y, por otro lado, hemos establecido opciones concretas derivadas de las peculiaridades de los datos clínicos.

El trasvase del corpus a este formato mediante el software CLAN permitirá hacer accesible en abierto nuestra base de datos para la comunidad internacional de investigadores por medio de la plataforma TALKBANK.

De este modo, el documento que presentamos constituye una guía para los investigadores que decidan trabajar con muestras PerLACH, permitiendo la correcta interpretación de los datos ofrecidos, a la par que una propuesta de transcripción y codificación que facilitará la labor de otros investigadores en el ámbito de la lingüística clínica que decidan utilizar CHAT/CLAN.

BIBLIOGRAFÍA

- BROOKSHIRE, R. (1997): *An Introduction to Neurogenic Communication Disorders*, St. Louis, Mosby.
- GALLARDO PAÚLS, B. (2002): “Fronteras disciplinarias: pragmática y patología del lenguaje”, en C. Hernández y E. Serra (coords.), *Estudios de lingüística clínica*, Valencia, Nau Llibres, págs. 129-174.
- GALLARDO PAÚLS, B. (2006): “Más allá de las palabras y su estructura: las categorías del componente pragmático”, en E. Garayzábal (ed.), *Lingüística clínica y logopedia*, Madrid, Antonio Machado Libros, págs. 81-196.
- GALLARDO PAÚLS, B. (2007): *Pragmática para logopedas*, Cádiz, UCA.
- GALLARDO PAÚLS, B. y FERNÁNDEZ-ÚRQUIZA, M. (2014): “Etiquetado pragmático de datos clínicos”, en prensa.
- GALLARDO PAÚLS, B. y HERNÁNDEZ SACRISTÁN, C. (2013): *Lingüística Clínica. Un enfoque funcional sobre las alteraciones del lenguaje*, Madrid, Arco/Libros.
- JEFFERSON, G. (1984): “Transcript Notation”, en J. M. Atkinson y J. Heritage (eds.), *Structures of Social Action: Studies in Conversation Analysis*, Cambridge, Cambridge University Press, págs. 9-17.
- MACWHINNEY, B. (2014a): *The CHILDES Project. Tools for Analyzing Talk-Electronic Edition. Part 1: The CHAT Transcription Format*, Carnegie Mellon University. <<http://childes.talkbank.org/manuals/CHAT.pdf>>
- MACWHINNEY, B. (2014b): *The CHILDES Project. Tools for Analyzing Talk-Electronic Edition. Part 2: The CLAN Programs*, Carnegie Mellon University. <<http://childes.talkbank.org/manuals/CLAN.pdf>>
- POYATOS, F. (1995): *La comunicación no verbal II. Paralenguaje, kinésica e interacción*, Madrid, Istmo.

ANEXOS

TABLA-RESUMEN DE CÓDIGOS EN LÍNEAS PRINCIPALES

CÓDIGO	SIGNIFICADO	EJEMPLO/COMENTARIOS
↑	Entonación ascendente	Sólo en casos en que se violenten los patrones melódicos esperables en el enunciado
↓	Entonación descendente	
→	Entonación suspendida	
@n	Neologismo	*CHI: breaked@n
@i	Interjección	*CHI: ay@i
@s	Cambio a segunda lengua	*HIJ: s [/-] suro es corcho@s:spa
@o	Onomatopeya	*ANP: guau@o .
@d	Forma específica de variedad o registro	*ANP: acabao@d .
+	Vincula los elementos de una onomatopeya	*ANP: guau+guau+guau@o .
()	Permite completar las palabras que el informante emite incompletas	*ANP: cole(gio) .
www	Fragmento de conversación no transcrito	*ANP: www .
[* p]	Parafasia fonémica	
[* p:w]	Parafasia fonémica: palabra real	*ANG: dudao [::ayudao] [* p:w]
[* p:n]	Parafasia fonémica: pseudopalabra	*ANG: lecha [:leche] [* p:n]
[* p:m]	Parafasia fonémica: metátesis	*MPT: murciégalo [:murciélago] [* p:m]
[* s]	Parafasia semántica	

CÓDIGO	SIGNIFICADO	EJEMPLO/COMENTARIOS
[* s:r]	Parafasia semántica: palabra real conceptualmente relacionada con target	*ANG: agua [::leche] [* s:r]
[* s:ur]	Parafasia semántica: palabra real sin relación con target (incluye proformas)	*ANG: trabajos [::galletas] [* s:ur]
[* s:uk]	Parafasia semántica: palabra real, target desconocido	*ANG: caballos [* p:uk]
[* s:per]	Perseveraciones (en el nivel de la palabra)	*PAR: The boy kicked the ball through the ball [::window] [* s:per] .
[* m]	Error morfológico a nivel de palabra	*FVC: y ya voy a beber agua (.) el vaso me se sr [/-] resbaló y se caí [* m] (.) . [+ gram] %err: caí = cayó
Postcódigo [+ gram]	Error gramatical a nivel de enunciado	En el ejemplo anterior, alteración del orden normativo de los pronombres átonos
Postcódigo [+ per]	Perseveración a nivel de enunciado	Repetition of an utterance when it is no longer appropriate (Brookshire, 1997)
Postcódigo [+ cir]	Circunloquio	*ANG: +, <en el> [/] <en el> [/] en el fenómeno de: (.) e:h [/] e:h (.) mecanizar las cosas (..) y esforzarme (.) en el sentido ↑ . [+ cir]
xxx	Habla ininteligible	*FVC: xxx que (.) tengo yo unos pocos .
[?]	Transcripción dudosa	*HIJ: ey@i home [?] però això seria per a vore com anaves (.) ,, no ?
0 . [+trn]	Turnos sin habla	*IMM: 0 [=] tries to fix her hair up] . [+ trn]
0word	Omisiones de palabras o categorías morfosintácticas	*BGP: a mitad 0de camino ? *BGP: a mitad 0prep camino ?

CÓDIGO	SIGNIFICADO	EJEMPLO/COMENTARIOS
(.) (..) (...) (4.3)	Pausas de menor a mayor duración, o de duración medida en segundos	*IMM: (3.0) uy@i (.) ahora vivo (..) en Masamagrell .
vowel:	Alargamiento vocálico	*IMM: casi siempre: .
[/-]	Reinicio	*FVC: +^ y lo tuve [/-] (.) me na [/-] el primero me nació muerto (.) .
[/]	Repetición	*FVC: y va allí y [/] y viene como está en marines (.) <viene y va> [/] viene y va (..) y muchas veces me dice +”/.
‡	Prefacio	*FVC: +” (.) mama ‡ hoy no vengo a comer .
„	Pregunta confirmatoria	*MVR: se acuerda usted de todo „ eh@i ?
syllable^syllable	Pronunciación silabeada	*FVC: produc^to .
[!]	Pronunciación enfática	*FVC: +” <a:h a:h> [!] .
[=! text]	Información paralingüística	*FVC: +” qué haces que no vas a abrir la puerta [=! laughs] .
&*INF	Sistema secundario de toma de turno (continuadores)	*IMM: sí [=! whispers while nodding] &*BGP: hmm ?
+//.	Autointerrupción	*SAR: smells good enough for +//. *SAR: what is that?
+/.	Interrupción ajena	*MVR: ah muy bien (.) yo [/-] pues yo pensaba no sé por qué (.) <pensaba que vivía usted sola y yo digo oh@i pues> [>] +/. *FVC: <no [/] no (..) y mi marido> [<].

CÓDIGO	SIGNIFICADO	EJEMPLO/COMENTARIOS
++	Terminación ajena del turno tras interrupción (turnos colaborativos).	*FVC: digo y: +/. *MVR: ++ no se lo creía nadie .
+,	Autoterminación del turno tras interrupción	*MVR: +, se arregla muy bien para estar sola .
[>]	Solapamiento sigue	*MAR: hay otro am<biente [=! laughs]> [>] .
[<]	Solapamiento precede	*IMM: <porque> [<] hay otro ambiente y todo .
+”/.	El turno siguiente es una cita	*FVC: después cuando me pasó esto (.) no me dijo mi [/] mi [/-] mis hijos +”/.
+”	Lo que sigue es una cita (se usa en combinación con el anterior).	*FVC: +” no tienes que cuidar ninguno más (.) .
+^	Encadenamiento de los turnos de diferentes hablantes	*MVR: la del medio . *FVC: +^ del medio .

OTROS CÓDIGOS

CÓDIGO	SIGNIFICADO	EJEMPLO/COMENTARIOS
%gpx:	Información proxémica en línea dependiente	*IMM: casi siempre: (0.3) la paella cómo (0.2) . %gpx: moves her hand