

El sistema CHILDES: codificación y búsquedas en el corpus *Koiné*

Isabel Fernández López
(Universidade de Santiago de Compostela)

Guión.

1. Introducción
2. La recopilación de datos desde una perspectiva histórica
3. Los principales avances en el manejo de datos
 - 3.1. Posibilidades de registro
 - 3.2. Posibilidades de acceso
 - 3.3. Posibilidades de tratamiento
 - 3.3.1. Codificación de datos
 - 3.3.2. Rastreo de datos
4. Conclusiones

La recopilación de datos en el siglo XIX

A esta edad, ó algo más tarde, según su grado de desarrollo, los niños empiezan á pronunciar monosílabos aislado de carácter labial; el *pa-pa*, *ma-ma* y pocas veces *me* ó *pe*, ocupan un período más ó menos largo, en que se cansa uno de observar sin notar progresos. El niño parece un alumno torpe que no acierta á juntar las sílabas; parece como el período de las lenguas monosilábicas; el *pa*, el *ma*, sin embargo, y el *ta*, y el *te*, que vienen despues de estar todos los dientes formados, suenan claros, distintos, con individualidad, como ciertas notas y sílabas de los pájaros.

[...]

Al empleo de los monosílabos labiales *pa*, *ma*, *ba* —el último de los cuales da origen al jueguecillo infantil que consiste en dar la madre ó nodriza guantaditas con la palma de la mano abierta sobre la boca del niño, haciéndole decir *ba baa baaa baaa baa baaa*— y de las dentales *tá* ó *té*, sigue la pronunciación de otras consonantes, cuya orden de aparicion no he tenido la suficiente virtud ni paciencia para observar. En este período, ó sea de los diez y ocho meses en adelante, empiezan, no sólo á pronunciar algunas consonantes más que las indicadas (*b*, *p*, *m*, *t*), sino á repetir los monosílabos juntándolos, como por ejemplo: *papa*—*mama*—*tata*. Mi hijo Pepe, á los 20 meses de edad, decia las siguientes palabras, que copia de una nota de mi cartera:

1 papa.	9 fo.
2 mama.	10 osa.
3 tete.	11 oncha.
4 tata.	12 pa.
5 eche.	13 ma.
6 omo.	14 apa.
7 oche.	15 uchacha.
8 fa.	16 aba.

Sobre estas diez y seis palabras que formaban, salvo alguna ligera omision, el *vocabulario* de mi hijo Pepe á los 20 meses de edad, no haré más que esas observaciones, á saber: que con *fo* queria decir *flor*; con *tata*, *bota*; con *uchacha*, *muchachos*; *aba*, *agua*; *pa*, *pan*; *ocha*, *oncha*, *Concha* (nombre de la criada), y con *tete*, *tío Pepe*. La *c*, como se ve, y otras consonantes, no las pronunciaba todavía; la *f* fue, de las citadas, la última que pronunció, y por cierto que tuvo por maestro de ella á un lindo gato negro, el cual, si por milagro, le arañó, pocas veces, sin género de milagro alguno le enseñó la pronunciacion de la *f*, haciendo *ffff*, sonido que el niño pretendia imitar.

Mi hijo Joaquin, que tiene hoy próximamente la edad de su hermano Pepe el año 81, disponía, hace poco más de un mes, ó sea á los diez y nueve meses de edad, del siguiente vocabulario, ya notablemente enriquecido, que transcribo con su significado usual.

1 Papá.		16 cocos.	mocos.
2 Mamá.		17 cocon.	Encarnacion.
3 Papa.		18 cacon.	Encarnacion.
4 Nene.		19 cocó.	Encarnacion.
5 Titín.	Joaquín.	20. cacon.	Encarnacion.
6 Tetín.	Joaquín.	21 ashon.	Encarnacion.
7 Caquín.	Joaquín.	22 mi.	yo.
8 baar.	bajar.	23 Tete.	Pepe y carrete.
9 caé.	caer.	24 no.	
10 Pepe.	Pepe.	25 aiya.	silla.
11 ubí.	subir.	26 erá.	cerrar.
12 riná.	orinar.	27 coca.	toca.
13 ahua.	agua.	28 Quica.	Francisca.
14 aba.	agua.	29 ntá.	sentar.
15 abí.	abrir.	30 Tata.	Zapatos.

Sobre este vocabulario, tambien interrumpido, y el anterior, voy á hacer ligeras observaciones. Ambos niños pronunciaban á la edad de once meses la *p*, la *m*, la *t*; la *f* no la pronunció el primero hasta los veinte meses, y el segundó hasta hará quince días, esto es, á los veinte y dos meses. Aquél pronunció de *ch* en la palabra *eche*, *leche*, y *ocha* (*Concha*), esforzándose por repetir el nombre de la criada de entonces, Concha; éste, en cambio, ha pronunciado antes la *c* que aquél no pronunciaba, esforzándose por repetir el nombre de la criada, á quien llama *Cocon*, *Cancon* y *Ashon* (*Encarnacion*). En el segundo vocabulario falta la *f* y figuran, en cambio, la *c*, que faltaba en aquél, la *r* suave, la *n*, que sólo existia en aquél en la palabra *oncha*, la *sh*, análoga á la *sh* inglesa en las palabras *she*, *short*, etc., y la *y*, en vez de la *ll*, en la palabra *silla* que los andaluces dicen *siya*. En cuanto á los elementos vocales, es mucho más rico el segundo vocabulario que el primero, pues en este pueden reducirse á

a—á—ó		ə o
a a		o e
a a		o a
e e		u aa

mientras que en el segundo

a—i—o	aá	e	ioo
	aa	ai	oo
	ee	ui	ea
	ii	ia	oa

Sobre los anteriores datos, que han de parecer de seguro tan nimios y prolijos á los no aficionados á estas materias como pobres y escasos á los filólogos, descansa la que hubiera debido llamar mejor que hipótesis, explicacion del nombre de *Titin*.

Creo, en efecto, poder afirmar, en vista de los vocabularios mencionados, que en todas las voces infantiles, tanto *monosilábicas* como *bisilábicas*, no entra jamás que una sola consonante como componente, bien aislada, como en *pa*, bien repetida, como en *papa*: los niños en la época á que aludo dicen *papa*, *mama*, *tata*; pero nunca *pata*, *tapa*, *mata*. Para decir Joaquin, por tanto, para imitar esta palabra, mi hijo menor necesitaba emplear una consonante cualquiera repetida.

Antonio Machado y Álvarez, "Titín y las primeras oraciones", *La Tribuna*, agosto de 1883. Se ha manejado la reedición: "Titín. Estudio sobre el lenguaje de los niños", *Boletín de la Institución Libre de Enseñanza*, 1887, XI, pp. 319-320 y 326-329. [pp. cit. 328-329].

Codificación de datos en el sistema CHILDES

▪ Texto 1

una oveja qué nos da la oveja un tiene un tiene un anido aquí qué tiene un anido oye dónde va tu anillo del otro día dónde está el que te dio tu mamá dónde está a ver enseña oh no trajiste anillo hoy qué es eso una pulsera y también te la dio tu mamá ésa no es mía ah ésa es tuya y de qué color es rosa rosa y qué más

▪ Texto 2

ADULTA: Una oveja.

ADULTA: ¿Qué nos da la oveja?

NIÑA: Un ...

NIÑA: Tiene un, un anido aquí.

ADULTA: ¿Qué tiene?

NIÑA: Un anido.

ADULTA: Oye, ¿dónde va tu anillo del otro día?

ADULTA: ¿Dónde está?

ADULTA: El que te dio tu mamá.

ADULTA: ¿Dónde está?

ADULTA: A ver.

ADULTA: Enseña.

ADULTA: Oh!

NIÑA: ¿?

ADULTA: No trajiste anillo hoy.

ADULTA: ¿Qué es eso?

NIÑA: Una pulsera.

ADULTA: ¿Y también te la dio tu mamá?

ADULTA: Ésa.

NIÑA: No.

NIÑA: Es mía.

ADULTA: Ah, es tuya.

ADULTA: ¿Y de qué color es?

NIÑA: Rosa.

ADULTA: Rosa, ¿y qué más?

▪ Estructura de las transcripciones

- (i) Las líneas de encabezado
- (ii) Las líneas principales
- (iii) Las líneas dependientes

(i) Las líneas de encabezado y cierre

- Se inician con el signo @
- En ellas se especifican los datos generales de la transcripción

(ii) Las líneas principales

- Se inician con el signo *
- En ellas se recogen las producciones emitidas por los participantes

(iii) Las líneas dependientes

- Se inician con el signo %
- En ellas se recogen diversas anotaciones e información complementaria

▪ Codificación en las líneas principales

- Códigos constantes:
- asterisco
 - tres letras que identifican al hablante
 - dos puntos
 - espacio de un tabulador
 - producciones emitidas
 - signo de puntuación final

pausa

pausa prolongada

xxx *material ininteligible tratado como palabra o como secuencia*

www *fragmento de conversación que se ha decidido no transcribir*

() *permite completar las palabras que el informante emite incompletas*

[>] *sigue un solapamiento*

[<] *precede un solapamiento*

[/] *repetición sin corrección*

[//] *repetición con corrección*

NOTA: aparecen precedidos siempre de los signos <> para fijar la secuencia a la que remiten

[/?] *repetición indefinida (superior a tres veces)*

[?] *aproximación o conjetura*

[*] *marca de uso peculiar codificado como error*

NOTA: aparecen precedidos de los signos <> si la secuencia a la que remiten es una frase

@o *onomatopeya*

@d *forma dialectal*

@s *segunda lengua*

+ *entre las partes de una onomatopeya*

+// *el emisor se interrumpe*

+/ *al emisor lo interrumpen*

++ *una persona completa la frase que ha iniciado otra*

+ *una persona completa la frase que ha iniciado él mismo en otro turno*

+’/ *el siguiente turno es una cita*

+” *se inicia una cita*

+”. *el turno anterior era una cita*

*ANP: hola # hola .

*ANP: azul # es azul .

*ANP: xxx .

*ANP: www .

*ANP: cole(gio) .

*ANP: es <azul> [>] .

*BRE: <no> [<] # es rojo .

*ANP: <un> [/] un niño .

*BRE: <un> [//] una niña .

*ANP: bueno [/?] # bueno .

*ANP: azul [?] .

BRE: dojo [] # rojo .

%err: dojo = rojo \$PHO ;

*ANP: guau@o .

*ANP: ghuapa@d .

*ANP: os conejos@s .

*ANP: guau+guau+guau@o .

*ANP: ele(fante) +// .

*ANP: ele(fante) +/ .

*ANP: ++ (ele)fante .

*ANP: + (ele)fante .

*ANP: la reina dijo +’/ .

*ANP: +” te perdono .

*ANP: dijo la reina +” .

▪ Codificación en las líneas dependientes

%err: *usos peculiares*

%act: *acción del hablante o del interlocutor*

%add: *receptor de una emisión*

%com: *comentarios de cualquier tipo*

%spa: *imitaciones*

%err: dojo = rojo \$PHO ;

%act: levanta la mano

%add: ANP

%com: caen las fichas

%spa: \$IMIT

▪ Muestra

@Begin
 @Languages: es, gl
 @Participants: ANA Ana Child, PIL Pilar Investigator
 @ID: es|koiné|ANA|3;1.27|female|Vite||Target_Child||
 @ID: es, gl|koiné|PIL||female|Vite||Investigator||
 @Coder: ISA
 @Date: 20-NOV-1998
 @Location: Santiago de Compostela
 @Situation: escuela infantil
 @Activities: comentario de cartas y láminas
 @Tape Location: 2.5
 *PIL: qué te parece ?
 *PIL: y dónde está Breixo ?
 *PIL: cuéntame .
 *PIL: por qué no +/ .
 *ANA: está en casa .
 *PIL: fue a casa ?
 %act: ANP asiente moviendo la cabeza
 *PIL: pero sabes por qué # por qué non está ?
 *PIL: xxx .
 %com: ANP mueve su silla; el ruido impide que escuchemos con claridad
 las palabras de PIL
 *PIL: está enfermo ?
 %act: ANP niega moviendo la cabeza; PIL se sienta a su lado
 *PIL: pois nada # xogamos as duas solas hoxe # vale ?
 *PIL: eh ?
 *ANA: tato [?] .
 %act: ANP ha cogido la baraja de cartas que estaban sobre la mesa
 *PIL: que estuveches facendo agora na clase ?
 %act: PIL le quita las cartas a ANP
 *ANA: xxx con el barro .
 *PIL: estuveches xogando co barro # si ?
 %act: ANP asiente moviendo la cabeza
 *PIL: e como tes así toda a cara ?
 %act: señala su cara
 *PIL: que che pasou ?
 *ANA: nada .
 *PIL: ensuciácheste toda ?
 *ANA: no !
 *PIL: no ?
 ANA: no # juguei [] <con> [/] con el agua de mi casa .
 %err: juguei = jugué \$MOR ;
 %com: el error se explica por la influencia de la morfología gallega
 *PIL: jugaste con el agua de tu casa ?
 *PIL: quen te peinou hoxe ?
 *ANA: mamá .
 *PIL: si ?
 *PIL: que guapa che puxo .
 *PIL: e puxóchete estas pinzas ela ?
 %act: le toca las pinzas que tiene en el pelo
 *PIL: que +/ .
 *ANA: mamá .
 *PIL: que chulis # non ?
 *PIL: gústanche ?
 *PIL: que tes aquí no pantalón ?
 %act: señala una de las piernas de los pantalones de ANP
 *PIL: a ver # que dibuxos tes # a ver ?
 *PIL: deíxame ver .
 *PIL: que dibuxos tes ?
 *PIL: que son ?
 *ANA: <son> [/] son # ositos .
 *PIL: ah # son ositos .
 *PIL: e de que color é este pantalón ?
 *ANA: rojo .
 *PIL: rojo .
 *PIL: que chuli !
 *PIL: a ver # vamos a xogar as cartas un pouquiño # vale ?
 *PIL: tu tes que adivinar unha e eu outra # vale ?
 %act: comienza a seleccionar las cartas
 *PIL: a ver # vamos a coller unhas pouquiñas .
 %act: ANP intenta coger una carta
 *PIL: eh # non se vale velas .
 *PIL: non se vale # non se vale # Ana .
 %act: ANP se agacha para recoger una carta que ha caído
 *PIL: a ver # empezamos así con estas .
 %act: ANP se ha vuelto a sentar
 *PIL: a ver # vamos a mezclalas ben .
 *PIL: a ver # primeiro tu e despois eu # vale ?
 *PIL: vale ?
 %act: ANP asiente con la cabeza
 *PIL: ah # que é isto ?
 %act: le enseña una carta
 *ANA: un elefante .
 *PIL: un ?
 *ANA: un elefante .
 *PIL: un elefante .
 *PIL: chachi .
 *PIL: a ver # agora eu .
 *PIL: uy # que e esto ?
 %act: mira atentamente la carta
 *PIL: non o sei moi ben qué é .
 *ANA: es una vaca .
 *PIL: non # unha vaca non é que ten os cornos moi grandes .
 *PIL: esto é un buei .
 *ANA: no .
 *PIL: si .
 *ANA: porque tiene xxx .
 %act: señala una parte del dibujo
 *PIL: claro .
 *PIL: un toro .
 *ANA: no es toro .
 *PIL: no ?
 *PIL: por que ?
 *ANA: <porque> [/] porque tiene ## los ojos abiertos .
 *PIL: e por eso non é un toro ?
 *ANA: no .
 *PIL: ah # bueno # vale .
 *ANA: es una cabra .
 *PIL: una cabra no .
 @End

Rastreo de datos en el sistema CHILDES

▪ Buscador

La ruta para acceder al buscador es *Window > Commands*. El acceso es inmediato si se pulsan las teclas *Ctrl+D* y salta de modo automático al abrir CLAN.

▪ Orden de búsqueda

Dar una orden supone construir una “frase” con distintas “palabras”. Algunos de los constituyentes básicos de esa supuesta construcción sintáctica son los siguientes:

- (i) Tipo de orden
- (ii) Transcripciones
- (iii) Informantes
- (iv) Secuencias ou palabras

Entre unos constituyentes y otros se deja un espacio en blanco (como entre las “palabras” de la “frase”).

Ejemplos:

```
combo vite2_12.cha +t*BRE +s“casa”
freq vit*.cha +t*ANA
```

▪ Tipo de orden

- Aparece en primer lugar
- No va precedido de ningún código
- Hay cinco tipos básicos:

freq	<i>busca datos de las frecuencias de uso de las palabras</i>
combo	<i>busca datos referidos a palabras o a combinaciones de palabras</i>
mlu	<i>busca datos referidos al dominio morfológico</i>
kwal	<i>busca datos de las frecuencias de uso de las palabras en el contexto</i>
gem	<i>busca pasajes de la transcripción marcados previamente</i>

▪ Transcripciones

- Non van precedidas de ningún código
- Se denominan utilizando el nombre del archivo
- El asterisco permite sustituir cualquier secuencia

```
Vite2_12.cha
Vite*.cha / *.cha
```

▪ Informantes

- Aparecen precedidos de +t
- Se denominan utilizando el mismo código que en la transcripción: mediante un asterisco y las tres letras identificativas
- Se pueden incluir varios informantes en la misma orden
- El asterisco permite sustituir cualquier secuencia
- Se pueden excluir hablantes si se cambia el signo + por el –

```
+t*BRE
+t*BRE +t*ANA
+t*AN* / +t*
-t*PIL -t*ANC
```

▪ Secuencia o palabra

- Aparecen precedidas de +s
- La secuencia va entrecomillada
- Si se busca una frase las palabras deben separarse con el signo ^
- Si se buscan varias palabras puede utilizarse el símbolo +
- El asterisco permite sustituir cualquier secuencia
- Se pueden excluir palabras si se cambia el signo + por el –

```
+s“casa”
+s“casa^de”
+s“mamá+papá”
+s“cas*”
-s “casa”
```

▪ Otros códigos que pueden formar parte de una orden

- Los informantes y las secuencias no son las únicas especificaciones que pueden introducirse en una orden.
- Precedidas del signo + y de una letra, es posible orientar la búsqueda de datos de modomuy diverso.
- Las combinaciones de estos constituyentes con los tipos de órdenes está limitado.
- Para conocer las compatibilidades existentes, basta con teclear un tipo de orden en *Commands* y pulsar *intro*. Inmediatamente se mostrará en *CLAN Output* toda la información sobre la combinación de códigos.
- Como muestra de otros posibles constituyentes, podemos citar los siguientes:

```
+o ordena los resultados según su frecuencia de uso
+zNw busca los usos que se especifiquen en secuencias con una longitud de palabras superiores a N
-wN +wN busca los usos que se especifiquen con un contexto formado por el número de líneas que marca N
```

▪ Muestra

```
> freq vite1.cha +tANA
freq vite1.cha +tANA
Wed Nov 16 19:52:21 2005
freq (10-Feb-2005) is conducting analyses on:
  ONLY speaker main tiers matching: *ANA;
*****
```

From file <Vite1.cha>

```
1 a
1 abiertos
1 agua
1 ahora
1 al
1 así
1 barro
4 bollo
1 cabra
1 caen
2 casa
1 come
3 con
1 de
1 e
4 el
2 elefante
2 en
3 es
1 está
1 gigante
1 hojas
1 juguei
1 los
2 mamá
1 mi
1 mono
1 mí
1 nada
5 no
1 ojos
1 ositos
1 otoño
1 parque
5 porque
2 pátano
1 rojo
1 se
1 sempiente
2 son
2 sí
1 tato
2 tiene
1 toro
9 un
3 una
1 vaca
1 ésta
```

48 Total number of different word types used

84 Total number of words (tokens)

0.571 Type/Token ratio

>

```
> combo vite*.cha +tBRE +s"yo"
(yo)
```

```
combo vite*.cha +tBRE +syo
```

```
Wed Nov 16 19:59:14 2005
```

```
combo (10-Feb-2005) is conducting analyses on:
```

```
  ONLY speaker main tiers matching: *BRE;
```

```
*****
```

From file <Vite1.cha>

Strings matched 0 times

From file <Vite2.cha>

Strings matched 0 times

From file <Vite3.cha>

*** File "Vite3.cha": line 179.

*BRE: y (1)yo también .

*** File "Vite3.cha": line 230.

*BRE: <<xxx> [<] y (1)yo vi> [/] (2)yo vi una ballena # y saltaba mucho .

*** File "Vite3.cha": line 254.

*BRE: y (1)yo # xxx .

*** File "Vite3.cha": line 317.

*BRE: pois <eu teño> [//] (1)yo tengo <un> [/] un cuento <de> [/?] de Blancanieves .

*** File "Vite3.cha": line 320.

*BRE: (1)yo !

*** File "Vite3.cha": line 473.

*BRE: sí # (1)yo .

*** File "Vite3.cha": line 478.

*BRE: (1)yo sé <el> [/] el del gato con botas .

*** File "Vite3.cha": line 480.

*BRE: también # que lo tengo (1)yo en mi casa .

*** File "Vite3.cha": line 484.

*BRE: la gatita [?] presumida # la tengo (1)yo .

Strings matched 10 times