



Revista de  
**LOGOPEDIA, FONIATRÍA y AUDIOLOGÍA**

[www.elsevier.es/logopedia](http://www.elsevier.es/logopedia)



ORIGINAL

## Comparativa entre procedimientos para el análisis del primer desarrollo lingüístico en niños y niñas con L1 vasco y español



Asier Romero Andonegi\* e Irati de Pablo Delgado

*Didáctica de la Lengua y la Literatura, Facultad de Educación, Universidad del País Vasco (UPV/EHU), Facultad de Educación de Bilbao. Campus de Bizkaia, Universidad del País Vasco (UPV/EHU)*

Recibido el 30 de junio de 2021; aceptado el 15 de enero de 2022

Disponible en Internet el 30 de junio de 2022

### PALABRAS CLAVE

Desarrollo lingüístico temprano;  
Inventario de Desarrollo de Habilidades Comunicativas de MacArthur-Bates;  
Léxico;  
Evaluación del lenguaje

### Resumen

**Antecedentes y objetivo:** La necesidad de realizar estudios sobre el nivel de desarrollo léxico temprano en niños bilingües vasco-español es una realidad plasmada en la literatura científica. En este contexto, tanto los inventarios de desarrollo comunicativo MacArthur-Bates como los métodos basados en la observación temprana han supuesto un importante elemento para el análisis del nivel de desarrollo lingüístico de los niños. Por tanto, el objetivo de este trabajo ha sido el de llevar a cabo una comparativa entre dos instrumentos diferentes sobre una misma muestra: por una parte, el Inventario de Desarrollo de Habilidades Comunicativas de MacArthur-Bates, en sus adaptaciones en español y en vasco, y por otra parte, la observación longitudinal por medio de las grabaciones de los niños en contextos diferentes de comunicación espontánea. **Materiales y método:** Los participantes han sido 8 niños vascos con distinta L1 (vasco o español) analizados desde los 17 a los 30 meses y procedentes de un pequeño pueblo del País Vasco, Bermeo. Se realizaron tres grabaciones mediante la observación contextualizada no-participativa en aquellos meses en los que los niños/as cumplían 17, 25 y 30 meses. Además, las anotaciones y transcripciones se realizaron a través de softwares profesionales: ELAN y CHAT. Asimismo, en el momento de la tercera grabación las familias cumplimentaron el CDI-2 en su adaptación al castellano y al euskera.

**Resultados:** Los resultados muestran la utilidad de los reportes parentales para lograr información sobre el desarrollo de las habilidades comunicativas tempranas. En todo caso, se cree necesaria la necesidad de realizar exploraciones de mayor entidad con el objeto de no caer en el argumento de la «sobrevaloración» de las emisiones infantiles. Asimismo, es interesante subrayar que los niños con L1 euskera y los niños con L1 español presentan resultados diferentes,

\* Autor para correspondencia.

Correo electrónico: [a.romero@ehu.es](mailto:a.romero@ehu.es) (A. Romero Andonegi).

de modo similar a lo observado en niños bilingües de otras lenguas, por lo que se puede afirmar que niños con la misma edad de adquisición temprana de dos lenguas (español y euskera) pueden desarrollar diferentes patrones en su bilingüismo dependiendo del grado de exposición a la lengua.

© 2022 El Autor(s). Publicado por Elsevier España, S.L.U. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## KEYWORDS

Early language development; MacArthur-Bates Communication Skills Development Inventory; Lexicon; Language assessment

## Comparison between procedures for the analysis on the first linguistic development in children with MT Basque and Spanish

### Abstract

**Background and objective:** The need to carry out studies on the level of early lexical development in Basque-Spanish bilingual children is a reality reflected in the scientific literature. In this context, both the MacArthur-Bates communicative development inventories and the methods based on early observation have been an important element for the analysis of the level of linguistic development of children. Therefore, the objective of this work has been to carry out a comparison between two different instruments on the same sample: on the one hand, the MacArthur-Bates Communication Skills Development Inventory, in its adaptations in Spanish and Basque, and, on the other hand, longitudinal observation through recordings of children in different contexts of spontaneous communication.

**Materials and method:** The participants were 8 Basque children with different L1 (Basque or Spanish) analyzed from 17 to 30 months and from a small town in the Basque Country, Bermeo. Three recordings were made through non-participatory contextualized observation in those months in which the children were 17, 25 and 30 months old. In addition, the annotations and transcripts were made through professional software: ELAN and CHAT. Likewise, at the time of the third recording, the families completed the CDI-2 in its adaptation to Spanish and Basque.

**Results:** The results show the usefulness of parental reports to obtain information on the development of early communication skills. In any case, the need for larger explorations is considered necessary in order not to fall into the argument of the «overvaluation» of children's broadcasts. Likewise, it is interesting to note that children with L1 Basque and children with L1 Spanish present different results, in a similar way to what was observed in bilingual children of other languages, so it can be stated that children with the same age of early acquisition of two languages (Spanish and Basque) they can develop different patterns in their bilingualism depending on the degree of exposure to the language.

© 2022 The Author(s). Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Introducción

La irrupción de nuevas disciplinas científicas en la lingüística ha originado un importante cambio en la perspectiva metodológica de la investigación. En esta nueva conceptualización empírica, los estudios lingüísticos y psicolingüísticos tienen que basarse sobre datos empíricos como uno de sus fundamentos argumentativos, incorporando, además, la variación social, geográfica y estilística, así como el modo oral, como objetos de pleno derecho del estudio y análisis lingüístico (Labov, 2001). La sociolingüística variacionista norteamericana (Labov, 1972), el contextualismo británico (Halliday y Hasan, 1985), la etnometodología de Garfinkel (1967) o el interaccionismo simbólico de Goffman (1981) son algunos ejemplos de esta transformación.

Esta necesidad de contar con datos reales para el análisis lingüístico ha supuesto un importante impulso en la construcción de corpus lingüísticos que han servido para extraer conclusiones sobre el funcionamiento de la interacción oral en situaciones controladas o no por el investigador. Ahora

bien, hoy en día se considera que los corpus deben cumplir unas características bien definidas: formato electrónico que permita al lingüista automatizar todas las tareas, autenticidad de los datos, criterios de selección condicionados por la finalidad concreta que priorice el corpus, representatividad que responda a parámetros estadísticos que garanticen que el corpus seleccionado representa y garantiza los objetivos del estudio y el tamaño muestral siempre acorde a la finalidad que exija el corpus (Villayandre, 2008).

En la medida en que la investigación sobre el desarrollo y la adquisición del lenguaje sea empírica, estamos tratando con un ámbito de conocimiento en el cual las hipótesis y las predicciones han de ser contrastadas sistemáticamente con los datos procedentes del comportamiento lingüístico de los sujetos, además de ser replicados por varios estudios. Las investigaciones sobre el desarrollo del lenguaje han utilizado principalmente diseños transversales, en los que se observa un corpus en un determinado momento en el tiempo, o longitudinales, los cuales evalúan el corpus a lo largo del tiempo. Ahora bien, para paliar los puntos débiles

de estos diseños, algunos autores han propuesto combinar ambos en un único estudio (Baltés et al., 1977), a lo que se le ha denominado diseño secuencial. Cualquiera que sea el diseño —transversal, longitudinal o la combinación de ambos—, puede realizarse a partir del estudio del comportamiento lingüístico espontáneo o bien a partir del control o de la manipulación rigurosa de variables, como en el caso del método experimental.

La mejor cualidad que nos ofrece el método observacional reside en la no intervención del investigador, es decir, nada restringe al sujeto observado para su producción lingüística, pudiéndose así conocer la situación real en el contexto natural del día a día del sujeto investigado. Ahora bien, el investigador se debe conformar con lo que la situación le ofrece y, por tanto, solo obtiene una muestra parcial del sistema que pretende investigar. Por otra parte, la utilización de inventarios como instrumento permite convertir a las familias (madres, padres, tutores...) en informantes, lo que facilita enormemente su aplicación, «y en su eficaz manera de preguntarles por el desarrollo lingüístico de sus hijos, lo que se demuestra a través del alto grado de fiabilidad y validez alcanzado» (Mariscal et al., 2006). La ventaja de estos cuestionarios respecto a la observación reside en poder conseguir gran cantidad de información en poco tiempo, ya que, en el caso de la observación naturalista, se necesita un tiempo extenso en el lugar para obtener observaciones representativas.

En este contexto, hay múltiples procedimientos para intentar obtener datos válidos y fiables sobre la producción, la percepción o la comprensión de lo que se está investigando, entre los que se encuentran:

1. Cuestionarios/inventarios estandarizados del lenguaje, entre los que destacan los inventarios MacArthur-Bates.
2. Tareas de imitación o técnicas de elicitación para estudiar la producción del lenguaje; en este sentido, los procedimientos informáticos y unificados de transcripción, y los paquetes de programas para el análisis de corpus lingüísticos, y en particular el sistema CHILDES, han representado un salto cualitativo en el modo de generar, analizar y compartir datos en el campo de la adquisición del lenguaje.
3. Tareas enfocadas al estudio de la percepción del habla.
4. Tareas para investigar la producción multimodal de gestos y vocalizaciones (en su vertiente segmental o suprasegmental).
5. Actividades para estudiar la comprensión lingüística.
6. El método basado en la simulación artificial del comportamiento lingüístico humano mediante ordenadores. En el campo de la adquisición del lenguaje, este método, basado en la simulación artificial del funcionamiento cerebral, se ha utilizado principalmente en la adquisición de la morfología verbal o gramatical, o en la adquisición de elementos prosódicos (Plunkett, 1995).
7. El método de observación contextualizado, mediante el cual las interacciones son registradas con grabaciones audiovisuales. La planificación de este método observacional no-participante contextualizado tiene un carácter inductivo y versátil, donde el investigador se integra en la situación natural de interacción.

Asimismo, en el entorno de una investigación sobre adquisición del lenguaje con niños<sup>1</sup> hay dos elementos fundamentales en el diseño de este tipo de observación: a) el grado de participación del investigador y b) la estrategia a desarrollar para introducirse en el contexto natural de observación. Dentro de los cuatro niveles de Anguera (1990), en la observación contextualizada se hace uso de la observación no-participante o externa, que es cuando el investigador no interacciona con los sujetos observados y conserva una distancia con ellos. Respecto a la estrategia a llevar a cabo, la investigación con bebés y niños implica una acomodación implícita donde es necesario lograr un ambiente de seguridad y de confianza en los niños. En este sentido, es recomendable un periodo de adaptación del investigador con los informantes con el objeto de que tanto el investigador como el instrumental asociado a la investigación puedan integrarse con una equilibrada naturalidad. La integración secuenciada del investigador en el escenario de investigación permitirá una adecuación positiva con los bebés o los niños. Además, no hay que olvidar que el objetivo de la investigación es obtener un corpus prelingüístico y/o lingüístico real y natural, por lo que resultará básico lograr un clima de confianza y de seguridad.

Sin duda, atendiendo a los distintos niveles de intervención e implicación en la investigación contextualizada no-participante, se hace referencia a la modalidad participación-observación en la que previamente se haya producido un período de habituación con los niños/as donde el investigador no sea un elemento desconocido, sino uno más. Esta modalidad se puede llevar a cabo con la aplicación de una metodología cualitativa, en la que se podrán examinar los posibles vínculos existentes entre unas y otras variables para posteriormente describir las variables funcionales en el contexto de una metodología cuantitativa (Romero et al., 2017). Por tanto, este tipo de observación consiste, como su nombre indica, en observar uno o varios sujetos pertenecientes al corpus seleccionado de investigación, y se ha añadido el calificativo de «contextualizada» porque inexorablemente esa observación está condicionada por el tipo de participación previa a la recolección de datos que efectúa el investigador. Guasch (1997) ya señala que la única manera de comprender una cultura y estilo de vida de los grupos humanos es mediante la inmersión en los mismos para ir recogiendo datos sobre su vida cotidiana.

El segundo elemento se refiere a la estrategia a desarrollar para introducirse en el contexto natural de observación. En este sentido hay que señalar dos elementos. En primer término, los aspectos relacionados con la confiabilidad de los datos, respetando los principios éticos establecidos, cumpliendo la normativa vigente y adjuntando el previo y preceptivo informe favorable del Comité de Ética en la Investigación que proporciona cada universidad. En segundo término, se encuentran los interlocutores (familiares, tutores/as y docentes) básicos en la tarea fundamental de introducir al investigador en el contexto de observación. Esta labor esencial de los informadores se refiere a su función de intermediario para que el investigador no sea visto como persona objetable y, asimismo, suministre la informa-

<sup>1</sup> Se mostrará el término en masculino al referirse a colectivos mixtos sin una intencionalidad discriminatoria.

ción necesaria sobre el niño para que el registro audiovisual de la observación sea lo más relevante posible.

Aparte del procedimiento observacional, nos encontramos con pruebas procedentes de la psicolingüística evolutiva y que se basan en la precisión de los padres y madres como informantes del desarrollo de sus hijos/as. Una de las pruebas más utilizadas son los Inventarios MacArthur de Desarrollo Comunicativo o escalas CDI (Fenson et al., 1993) que, basados en informes parentales, constituyen un instrumento de evaluación temprana del desarrollo comunicativo (Mariscal et al., 2010). Frente a métodos basados en la observación, los inventarios de desarrollo comunicativo MacArthur-Bates emplean ítems muy concretos basados en la información de los padres y madres mediante el uso del reconocimiento de comportamientos (Dale et al., 1989).

Estas escalas CDI presentan ciertas ventajas frente a otros instrumentos: el escaso coste del procedimiento, la evaluación del desarrollo comunicativo en muestras muy amplias y la existencia de más de una centena de adaptaciones a diferentes lenguas que facilita las comparaciones interlingüísticas (Pérez-Pereira y García, 2003). Además, distintos estudios realizados en el campo de la psicolingüística evolutiva avalan la precisión de los padres como informantes del desarrollo de sus hijos, siempre que a estos se les pregunte adecuadamente (Bates et al., 1988). Otro tipo de investigaciones centradas en los Inventarios MacArthur de Desarrollo Comunicativo también han avalado la validez y la fiabilidad de este tipo de informes parentales como instrumentos de evaluación temprana del desarrollo comunicativo y lingüístico (Fenson et al., 1993). Ahora bien, estudios posteriores han identificado efectos del nivel socioeconómico y de instrucción de los progenitores sobre la valoración que hacen del lenguaje de sus hijos, ya que esta puede estar condicionada por la falta de conocimiento de los progenitores de aquello que van a evaluar (García et al., 2008). Ahora bien, estos mismos autores señalan que las críticas a este modelo metodológico de obtención de la información se deben aplicar más al modo en que se obtiene la información que a su valor intrínseco (Fenson et al., 2000; García et al., 2008).

Por tanto, el objetivo de este trabajo es realizar una comparativa entre dos instrumentos de análisis de la producción léxica mediante la evaluación del nivel de desarrollo léxico del niño haciendo uso del cuestionario cumplimentado por los padres y madres y de las grabaciones de los niños en contextos diferentes de comunicación espontánea y con la comparativa del español y el euskera. Asimismo, cuando los niños tienen la edad de 30 meses se comparan los resultados de ambas técnicas metodológicas de registro de datos.

## Método

### Participantes

Los participantes de la presente investigación son 8 niños vascos (4 niñas y 4 niños) desde los 17 a los 30 meses (media: 17 meses; desviación típica: 3,634). Los niños proceden de hogares biparentales monolingües (vasco o español) de un pequeño pueblo del País Vasco, Bermeo, situado a 30 km al nordeste de Bilbao. Sus padres y madres, en su mayor parte con formación universitaria, fueron contactados a través de

la red de escuelas infantiles y accedieron a participar de forma voluntaria en la investigación. Tras explicarles el procedimiento, todos dieron su autorización por escrito a la grabación de cada una de las sesiones. En cuanto a la lengua materna, hay que señalar que de los 8 niños, 4 tienen como L1 euskera y los otros 4 español, y todos estaban escolarizados en un programa de inmersión total y precoz en euskera durante el período en el que se llevó a cabo la investigación (tabla 1).

### Materiales y procedimiento

Se realizaron tres grabaciones, correspondientes a aquellos meses en los que los niños cumplían 17, 25 y 30 meses, con un cómputo total de 9 horas de grabación<sup>2</sup>. Asimismo, hay que señalar que este rango de edad fue seleccionado, precisamente, por tratarse del periodo clave en el que se producen los primeros actos comunicativos verbales intencionados mediante el gesto, y su transición desde el balbuceo variado o conversacional al periodo de las primeras palabras. Los gestos comunicativos son de aparición más temprana y las primeras palabras son ubicadas por varios autores hacia los 12 meses, con lo cual las edades analizadas se ubican, en realidad, después de estos hitos; entre otros, Fernández-Feña (2012).

Las sesiones se registraron en vídeo en las casas de los sujetos mediante una metodología de observación contextualizada no-participativa, estando presente en todos los casos la madre o el padre. Cada grabación tenía una duración de aproximadamente 20-30 minutos, si bien la duración real se ajustó al estado emocional y comunicativo de cada niño/a. Hay que señalar que en ningún caso se ha fijado con anterioridad a las sesiones un estricto patrón organizativo de las actividades a realizar. Por tanto, mediante esta actitud flexible se buscaba establecer una interacción dinámica, dialogante, abierta y, en cierto modo, libre con los sujetos del corpus (Bosh, 2004; Fernández López, 2009). Las grabaciones se realizaron utilizando la cámara ZOOM modelo Q4HD con micrófono externo Rode SmartLav con condensador, situado en la ropa del niño/a.

Todos los actos comunicativos verbales producidos por los niños fueron identificados y localizados en las grabaciones, utilizando el software de transcripción multimodal ELAN (Lausberg y Sloetjes, 2009) y dentro de la base de datos *Child Language Data Exchange System* (CHILDES), y se utilizó el sistema CHAT (MacWhinney, 2000). Además, los análisis de las transcripciones codificadas se realizaron utilizando el comando «MLU» del programa CLAN y los criterios descritos por Siguan et al. (1990). Este comando mide para cada uno de los informantes el número de enunciados, el número de turnos y el número de palabras, y sobre ellos calcula la media de palabras por turno, la media de producciones por turno y la longitud media de emisión (LME) sobre todas las emisiones que produce el niño en ese contexto. Además, se ha empleado también el comando «FREQ» de CLAN con el objetivo de realizar sobre las transcripciones recuentos de

<sup>2</sup> En los casos en que aparecen más sesiones, ello se debe a que, por diferentes motivos (cansancio, cambio de escenario, etc.), se tuvieron que realizar más sesiones de grabación.



**Tabla 1** Características de las niñas y niños participantes

Siglas que identifican al informante	L1 del informante	Edad de la primera grabación	Edad de la última grabación	Número de grabaciones en las que participa	Duración de todas las grabaciones del informante
LIA [fem.]	Vasco	1;5	2;6	3	1 h 37' 16''
GBE [fem.]	Vasco	1;4	2;6	3	1 h 21' 22''
GLL [masc.]	Vasco	1;5	2;5	4	1 h 42' 21''
IRM [masc.]	Vasco	1;5	2;7	4	1 h 54' 43''
GTN [fem.]	Español	1;5	2;6	3	1 h 15' 23''
CBV [fem.]	Español	1;4	2;5	5	2 h 21' 28''
JIE [masc.]	Español	1;4	2;6	3	1 h 19' 27''
XPB [masc.]	Español	1;5	2;6	4	1 h 43' 52''

las frecuencias de todas las palabras ofreciendo un listado de las mismas, al tiempo que calcula el índice de diversidad léxica que viene dado por el cociente entre los vocablos diferentes que aparecen en la muestra y el total de palabras de la misma. Por otra parte, el habla ininteligible (por solapamiento con el habla adulta, ruidos, etc.) no se incluyó en las transcripciones. Los criterios para considerar una emisión vocal como palabra fueron los siguientes: la palabra debía parecerse fonológicamente a una palabra objetivo y la palabra debía utilizarse en un contexto plausible. En los casos en que no se pudo describir el estatus de la palabra, se utilizaron criterios de confirmación (identificación por parte de los padres y madres si la palabra estaba marcada en el formulario MacArthur).

Con el objetivo de comprobar la fiabilidad de la codificación realizada, se utilizó el procedimiento de acuerdo inter-jueces con dos evaluadores independientes y entrenados previamente, y mediante el cálculo del índice kappa en un subconjunto del 20% de los casos (Cohen, 1960). El acuerdo obtenido fue del 89% para todos los vocablos ( $k=0,872$ ). Atendiendo a Landis y Koch (1977), que desarrollaron las pautas para la interpretación de este índice, se encontraría en la franja 81-100, correspondiente a una coincidencia casi completa.

Además de la grabación en vídeo a los niños, se utilizaron las versiones estandarizadas en español y en euskera del Inventario de Desarrollo de Habilidades Comunicativas de MacArthur-Bates, más conocido como CDI (Fenson et al., 1993; la versión en español, López-Ornat et al., 2005; la versión en euskera, García et al., 2008). En concreto para este estudio, se seleccionó el inventario CDI-2, que mide la producción léxica, morfológica y sintáctica de niños entre 17 y 30 meses y, dentro de este, el apartado «Lista de vocabulario». Previa explicación del contenido de este instrumento, se les entregó este inventario a los padres y madres en el momento en el que se llevó a cabo la tercera grabación (al cumplir los niños 30 meses). A los padres y madres que empleaban el español en el hogar se les proporcionó el CDI-2 en su adaptación al español, y a aquellos padres y madres que usaban el euskera como lengua familiar se les proporcionó el CDI-2 en su versión al euskera. Transcurridos 15 días, se recogieron los sobres con los inventarios cumplimentados y se procedió a la corrección y la puntuación de estos. Siguiendo a López-Ornat et al. (2005) y García et al. (2008), las respuestas de los

cuestionarios se convirtieron en puntuaciones, de manera que cada ocurrencia se puntuaba como 1. Se contaban todas las puntuaciones, siendo la puntuación mínima 0 y la máxima 588 (CDI-2 versión en español) y siendo la puntuación mínima 0 y la máxima 654 (CDI-2 versión en euskera). Lógicamente, al encontrarnos con dos puntuaciones máximas atendiendo a las diferentes versiones, se ha realizado una correlación con el objetivo de unificar una única puntuación máxima. Además, de la correlación realizada sobre la puntuación directa de vocabulario, se ha tomado también en consideración la longitud media de emisión (LME) sobre la emisión más larga que produce el niño.

## Resultados

En primer lugar, se analizó la longitud media de emisión (LME) mediante el comando «MLU» de CLAN. En la figura 1 se muestra la LME para cada una de las sesiones (representadas en el eje X) de cada niño. Para la media del componente sintáctico se tuvo en cuenta la LME en palabras, sin contabilizar imitaciones, rutinas, repeticiones y respuestas sí/no (Ezeizabarrena y García Fernández, 2018). Es interesante resaltar que los niños de la muestra presentan una importante variación respecto al tiempo en el que alcanzan un nivel de LME de 1,5, entendiendo este número como el inicio de la etapa lingüística de las dos palabras. Ahora bien, lo importante para la investigación es que la mayor parte de los niños, atendiendo a este nivel, producen los enunciados con una media de dos palabras. La figura 1 muestra que para los 17 meses prácticamente ninguno de los niños/as alcanza el nivel de 1,0. Este resultado es muy llamativo, y puede deberse a diferentes factores relacionados con la metodología de grabación, a compartimientos lingüísticos sesgados por las grabaciones o a procesos de carácter individual que afectan a la adquisición temprana del componente lingüístico en niños bilingües. Además, mientras que GBE y GLL alcanzan un nivel de LME de 1,5 para los 25 meses, LIA, GTN y CBV no lo logran hasta los 30 meses. El nivel de 2,0 de LME tan solo lo alcanzan para los 30 meses GBE e IRM; el resto de niños se desenvuelven entre el nivel 1,4 y el 1,9 (tabla 2).

En segundo lugar, se analizó el desarrollo léxico de cada niño/a por medio del comando «FREQ» en CLAN, es decir, el número de palabras grabadas únicas por sesión. Se emplea el comando «FREQ» para el recuento de las frecuencias de

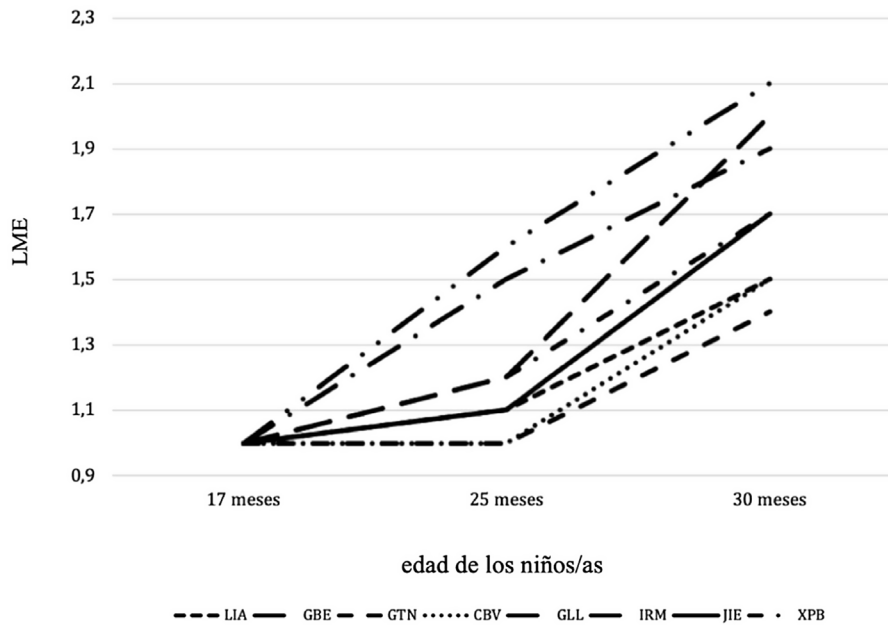


Figura 1 LME para cada una de las sesiones y en cada niño.

Tabla 2 Descriptivos para LME por tramo de edad

Informante	17 meses		25 meses		30 meses	
	Media	DT	Media	DT	Media	DT
LIA	0,8	0,9	1,2	0,9	1,5	1,4
GBE	0,7	0,8	1,5	1,3	2,0	1,6
GLL	0,6	0,7	1,5	1,2	1,4	1,1
IRM	0,7	0,8	1,4	0,9	2,0	1,5
GTN	0,8	0,6	1,2	1,1	1,5	1,1
CBV	0,6	0,5	1,1	0,7	1,5	1,2
JIE	0,5	0,7	0,9	0,6	1,5	1,2
XPB	0,9	0,8	1,5	1,2	1,9	1,6

DT: desviación típica.

todas las palabras, frente al índice TTR, utilizado tradicionalmente en la literatura científica para el cálculo de la diversidad léxica type/token. La figura 2 muestra el número de tipos de palabras distintas encontradas en cada una de las sesiones para cada niño; es decir, se alude a tipos de palabras, no a clases de palabras, en el sentido de tipologías gramaticales. Si atendemos a la propuesta metodológica de Depaolis et al. (2008) en situar y señalar el primer mes en el que el niño usa espontáneamente 25 o más palabras identificables en su interacción con el adulto y en un espacio cronológico de media hora, ninguno de los niños del corpus lo logra en la sesión de los 17 meses. Los datos de la figura 2 muestran que, de manera similar a los datos de LME, los niños GBE y GLL alcanzan un tamaño de vocabulario muy por encima del resto para los 25 meses. Además, atendiendo a los datos extraídos de las videograbaciones, a los 25 meses todavía hay niños que están en parámetros de 30-40 palabras, como GTN, CBV y JIE. A los 30 meses se repiten los cómputos anteriores; así, mientras que los niños GBE, GLL e IRM están por encima de las 200 palabras, otros,

como LIA, GTN, CBV y XPB, están en datos cercanos al centenar de vocablos. Por otra parte, es importante señalar que, aunque los recuentos léxicos fluctúan entre las sesiones, se asume que si un niño produce 25 palabras distintas en una sesión, este ha alcanzado el punto de 25 palabras señalado.

La figura 2 muestra que los datos del tamaño del léxico de los niños no difieren de los datos de LME descritos en la figura 1. Así, mientras que GBE y GLL alcanzan el nivel 1,6 y 1,5 con 188 y 132 palabras, respectivamente, a los 25 meses, otros, como GTN y CBV, están todavía en el nivel 1 a los 25 meses con 38 y 26 palabras, respectivamente. Este paralelismo se vuelve a constatar a los 30 meses con GBE y IRM, cuyos cómputos se mueven más allá de las 200 palabras y, por tanto, sus índices en torno al 2. Estos datos, basados en la observación contextualizada, muestran que los niños del corpus llegan a la etapa de las 2 palabras antes de los 30 meses.

Los datos registrados para estos niños a través de la observación contextualizada no se corresponden con los registrados en otras investigaciones, salvando lógicamente

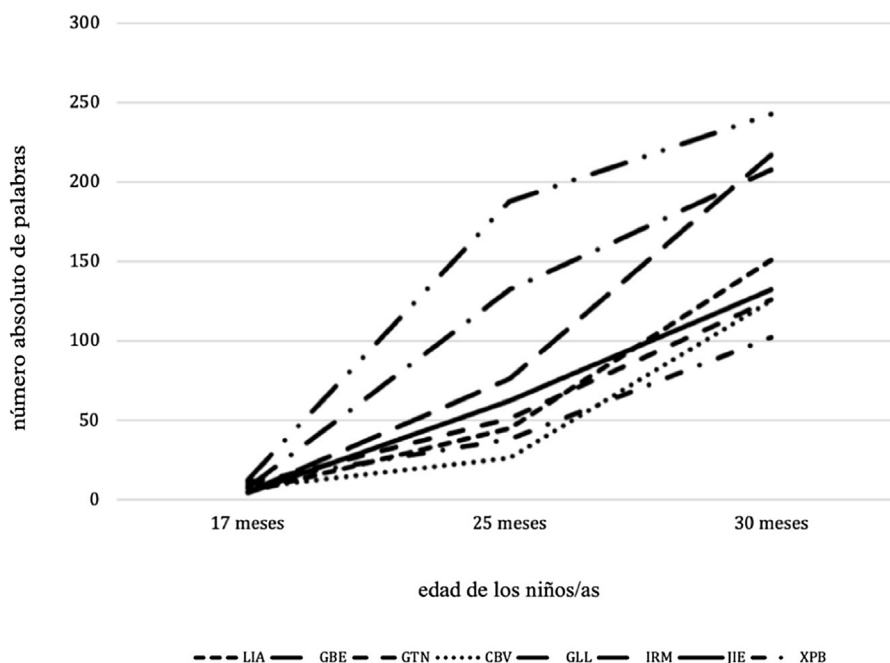


Figura 2 Palabras distintas para cada una de las sesiones y en cada niño.

las diferencias provocadas por otras variables o factores. Así, por ejemplo, en el corpus de referencia Serra-Solé (Serra y Solé, 1995), basado también en transcripciones para el español, se sitúa en torno a las 550 palabras a los 30 meses. Si bien los datos demuestran que, ciertamente, la curva de adquisición de palabras nuevas en muchos niños del corpus tiene un incremento más pronunciado alrededor de los 25 meses o cuando el corpus tiene entre 50 y 75 palabras o más (Dromi, 1987), otros autores han negado este hecho (Fenson et al., 1993) reorientando la cuestión y situándola en el marco de las diferencias individuales en la adquisición. Así, algunos niños y niñas presentan un rápido acrecentamiento de palabras evidente, pero otros muestran un incremento progresivo en el tiempo, sin que se observe un cambio brusco.

Ahora bien, analizados los datos en función del sexo y de la lengua materna (fig. 3), si se perciben variaciones en el número de palabras producidas a favor de los niños y niñas que presentan L1 euskera, es decir, estos niños con L1 euskera producen un mayor número de palabras que los niños del corpus con L1 español. La distribución del número absoluto de palabras diferentes para cada niño resulta evidente, pero no significativa desde un punto de vista estadístico.

Como ya se ha señalado, estos datos fueron obtenidos a través de las transcripciones basadas en la metodología de la observación contextualizada. Ahora bien, atendiendo al objetivo de la investigación, ¿qué sucede si estos resultados se comparan con los datos obtenidos de los Inventarios del Desarrollo Comunicativo MacArthur-Bates? Es decir, a continuación se describen los datos extraídos de las respuestas obtenidas a través del instrumento CDI.

La producción de palabras medida en el CDI-2 en su adaptación al español y al euskera se presenta en las tablas 3 y 4. Estos datos muestran que la interpretación que realizan los padres y madres a través del CDI-2 en sus versiones

idiomáticas no se corresponden con los resultados obtenidos a través de la observación contextualizada. En todos los casos, y atendiendo a las dos adaptaciones del cuestionario, las medidas de producción léxica son superiores para el CDI-2. Además, hay que señalar que la correlación entre las dos medidas de producción léxica, CDI-2 (versión en castellano y en euskera) y observación contextualizada, resultó significativa ( $r=0,82$ ;  $p=0,024$ ). Por último, la correlación entre las medidas de LME, CDI-2 (versión en castellano y en euskera) y observación contextualizada para la medida basada en el número de palabras también resultó significativa ( $r=0,61$ ;  $p=0,008$ ). En todo caso, hay que señalar nuevamente que la longitud media de emisión del CDI-2 se calculó sobre la emisión más larga que producía cada niño, mientras que la LME correspondiente a la transcripción de la muestra de lenguaje espontáneo se realizó sobre todas las emisiones que producía el niño en ese contexto dado.

Por otra parte, hay también una serie de resultados que destacan por garantizar la pertinencia de este tipo de instrumentos. Así, en la mayor parte de los casos analizados, los datos basados en la observación contextualizada se corresponden proporcionalmente con los datos descritos en los cuestionarios CDI-2. Es decir, aunque las diferencias son considerables, las mayores puntuaciones del CDI las obtienen aquellos niños que presentan mejores registros numéricos en las transcripciones realizadas por medio de la observación contextualizada (fig. 4). Esta misma circunstancia se describe en sentido contrario para GTN, que es la niña que presenta la menor producción léxica (con 102 palabras en la observación contextualizada y 119 palabras a través del cuestionario).

Hay también otro elemento a destacar, que es la producción de palabras de los niños con L1 español. Así, los padres y madres que cumplimentaron el CDI-2 en su adaptación al español les otorgaron un número inferior de palabras que los

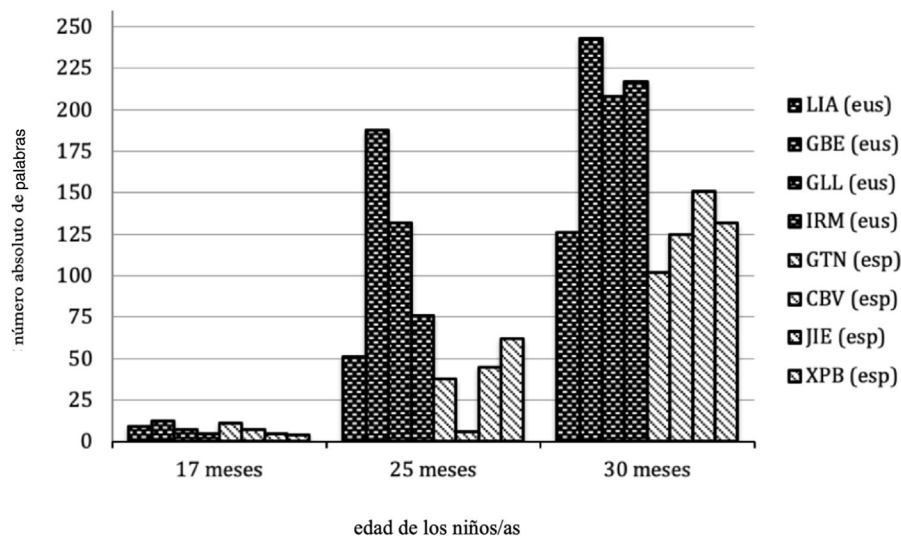


Figura 3 Comparación por la L1 de cada niño.

Tabla 3 Producción de palabras en CDI-2 (L1 euskera)

Informante	Media	DT	Puntuación mínima	Puntuación máxima
LIA [fem.]	295,74	122,34	38	654
GBE [fem.]	303,37	138,53	40	654
GLL [masc.]	304,81	141,05	37	654
IRM [masc.]	323,40	119,56	34	654

DT: desviación típica.

Tabla 4 Producción de palabras en CDI-2 (L1 español)

Informante	Media	DT	Puntuación mínima	Puntuación máxima
GTN [fem.]	141,19	109,52	28	588
CBV [fem.]	171,67	117,48	21	588
JIE [masc.]	186,53	139,08	32	588
XPB [masc.]	194,23	159,31	37	588

DT: desviación típica.

padres y madres que cumplimentaron el CDI-2 en su adaptación al euskera. Esta circunstancia ya se había señalado en los resultados de las transcripciones mediante la observación contextualizada, pero ahora los cuestionarios también lo confirman. A este respecto, hay que señalar que el CDI euskera tiene más ítems que el del español peninsular, por lo que esta podría ser una de las posibles razones de la aparente ventaja del euskera frente al español (Ezeizabarrena, 1996). Con todo, se realizaron dos análisis: por una parte, se dividió la puntuación directa de cada niño por el número posible de ítems, y por otra parte, se realizó una correlación entre medidas de ambas versiones, en ambos casos con el objetivo de atender a esta diferencia en la puntuación existente entre la versión en español y la correspondiente para el euskera. Hay que señalar que la comparativa atendiendo tanto al análisis individual como a la resultante de dicha correlación no resultó significativa ( $r=0,67$ ;  $p=0,38$ ).

## Discusión y conclusiones

Los resultados obtenidos por medio de estas dos diferentes metodologías confirman la validez de ambas para averiguar la media de producción léxica. Los datos proporcionados por una u otra metodología presentan interesantes similitudes en cuanto al cómputo de producción léxica; es decir, los niños que destacan por encima o por debajo en cuanto a su producción comunicativa en las transcripciones de la observación contextualizada tienen una similar actuación a través de las puntuaciones extraídas de los cuestionarios. En todo caso, los datos hacen referencia a índices de una determinada habilidad y, por tanto, no se pueden interpretar como registros exhaustivos de dicha habilidad.

Los Inventarios MacArthur-Bates de Desarrollo Comunicativo (MCDI) han sido reconocidos en la literatura científica como un procedimiento eficaz y válido para evaluar



## Comparación entre ambas metodologías en cada niño a los 30 meses

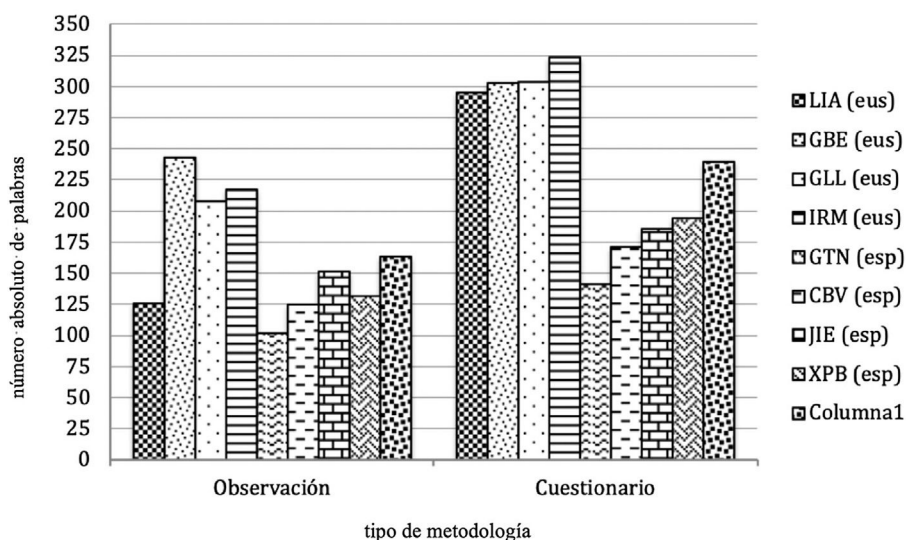


Figura 4 Comparación entre ambas metodologías en cada niño a los 30 meses.

tempranamente el desarrollo lingüístico en general y, en particular, el primer desarrollo gramatical. Los resultados permiten afirmar que las madres y los padres son informantes fiables del desarrollo gramatical de sus hijos cuando estos son todavía menores de 3 años y el formato del instrumento de medida tiene instrucciones claras y de fácil ejecución. En cuanto a la longitud media de emisión (LME), calculada con el programa CLAN y basada en los inventarios y en las muestras de lenguaje espontáneo, sí creemos que hay que realizar exploraciones de mayor entidad con el objeto de no caer en el argumento de la «sobreevaluación» de las emisiones infantiles, ya que los propios datos pueden ser consecuencia del análisis realizado.

Hay que tener en cuenta que en nuestro estudio no se observan variaciones importantes en la evolución de la LME, pero sí un cierto retraso en la evolución lógica atendiendo a lo que acontece en otros estudios similares, ya que, en nuestro caso, para la etapa de 30 meses se describen parámetros de 1,5 para algunos niños. Así, consideramos que este punto requiere mayor profundización y una muestra más amplia, ya que en los datos aportados en otras investigaciones (Pérez-Pereira y García, 2003) sí se muestran puntos de inflexión entre la extensión del enunciado y la complejidad sintáctica.

Por otra parte, hay una serie de elementos que destacan, dado que reflejan grandes diferencias con las encontradas en otras investigaciones. Así, los datos referentes a producción léxica en castellano para la edad entre 16 y 30 meses, analizada por medio de los cuestionarios y observación longitudinal directa, no coinciden con los encontrados en la versión para el castellano (López-Ornat et al., 2005), donde a los 30 meses se superan las 400 palabras. Lógicamente, el estudio de López-Ornat et al. (2005) tiene en cuenta para su muestra niños de toda España, por lo que los resultados son más amplios. Sin embargo, los datos para el euskera recogidos por medio de los cuestionarios sí coinciden con los encontrados en la versión en euskera (Barreña y García, 2009), aunque en su percentil más bajo. Por otro lado, los

datos registrados por medio de la observación contextualizada no coinciden con los cómputos registrados en otros estudios similares. Estas diferencias pueden atribuirse a sesgos de las muestras, a diferencias culturales a la hora de contestar el cuestionario, a alteraciones en el comportamiento lingüístico del niño/a durante las grabaciones, o a procesos de carácter individual que afectan a la adquisición temprana del componente lingüístico en niños bilingües, y que para el caso del bilingüismo vasco-español ya han sido señalados por diferentes investigadores: diferentes grados de exposición en la lengua meta, características tipológicas muy diferentes entre una lengua y otra, etc. (Ezeizabarrena, 1996).

Por lo tanto, y a la luz de estos datos, es necesario realizar análisis adicionales, a partir de datos longitudinales y sobre una muestra mayor, para poder proponer un modelo de función evolutiva de las emisiones verbales en aquellos niños que presentan una L1 español y se encuentran inmersos en una dinámica socioeducativa en la que la lengua vehicular temprana es el euskera, dado que podríamos estar ante bilingüismo consecutivo tardío (Kuhl et al., 2016). De esta forma, podríamos interpretar que estamos ante niños en un proceso de adquisición bilingüe simultánea (2L1) pero con exposición baja al contacto lingüístico con el euskera y, por tanto, atendiendo a investigaciones sobre la adquisición léxica, se podría argumentar que el vocabulario de estos niños es más reducido que el de los monolingües (Pearson et al., 1993). Es precisamente este grado de exposición el que lleva a algunos investigadores a restar protagonismo a la diferenciación categórica entre monolingües y bilingües, y a subrayar el valor individual del grado de exposición a la lengua (Barreña et al., 2011).

A tenor de estos resultados, es interesante subrayar que los niños con L1 euskera y los niños con L1 español (o bilingües simultáneos) presentan resultados diferentes, de modo similar a lo observado en los niños bilingües inglés-español. Los niños con L1 euskera del estudio muestran, tanto por medio de los cuestionarios como por medio de la observación directa contextualizada, un proceso de adquisición

más rápido que los niños con L1 español (o bilingües simultáneos). Por tanto, podríamos concluir afirmando que niños con la misma edad de adquisición temprana de dos lenguas (español y euskera) pueden desarrollar diferentes patrones en su bilingüismo dependiendo del grado de exposición a la lengua. Esta circunstancia se puede dar incluso en zonas sociolingüísticas en las que el euskera es la lengua predominante, como ocurre en este estudio.

## Financiación

Esta investigación se ha realizado dentro de los proyectos IT1208/19 (Gobierno Vasco) y GIU 16/22 (Universidad del País Vasco) y gracias al apoyo económico de la UPV/EHU para la contratación de personal investigador en formación predoctoral (PIF).

## Conflicto de intereses

No tenemos conflicto de intereses

## Bibliografía

- Anguera, M. T. (1990). Metodología observacional. En A. Grass, M. T. Anguera, y J. Gómez (Eds.), *Metodología de la investigación en ciencias del comportamiento* (pp. 125–238). Murcia: Compobell.
- Baltes, P. B., Reese, H. W. y Nesselroade, J. R. (1977). *Life-span Developmental Psychology*. Monterrey, CA: Brooks/Cole.
- Barreña, A. y García, I. (2009). Lehenengo hitzak eta morfologia-markak haur euskaldunengan: elebakarrengan eta elebidunengan. *Euskera*, 54(2), 603–638.
- Barreña, A., Ezeizabarrena, M. J. y García, I. (2011). La influencia del grado de exposición a la lengua en la adquisición del euskera (8-30 meses). *Infancia y Aprendizaje*, 34(4), 393–408. <https://doi.org/10.1174/021037011797898430>
- Bates, E., Bretherton, I. y Snyder, L. (1988). *From First Word to Grammar*. Cambridge: Cambridge University Press.
- Bosh, L. (2004). *Evaluación fonológica del habla infantil*. Barcelona: Masson.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46.
- Dale, P. S., Bates, E., Reznick, J. S. y Morisset, C. (1989). The validity of parent report instrument of child language at 20 months. *Journal of Child Language*, 16(2), 239–250. <https://doi.org/10.1017/S0305000900010394>
- Depaolis, R., Vihman, M. M. y Kunnari, S. (2008). Prosody in production at the onset of Word use: A cross-linguistic study. *Journal of Phonetics*, 36, 406–422. <https://doi.org/10.1016/j.wocn.2008.01.003>
- Dromi, E. (1987). *Early Lexical Development*. Cambridge: Cambridge University Press. CHILDES (Child Language Data Exchange System)., <https://childes.talkbank.org>.
- Ezeizabarrena, M. J. (1996). *Adquisición de la morfología verbal en euskera y castellano por niñas/os bilingües*. Bilbao: Servicio Editorial de la Universidad del País Vasco.
- Ezeizabarrena, M. J. y García Fernández, I. (2018). Length of utterance, in morphemes or in words? MLU3-w, a reliable measure of language development in early Basque. *Frontiers in Psychology*, 8, 1–17. <https://doi.org/10.3389/fpsyg.2017.02265>
- Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J., Pethick, S. y Reilly, J. (1993). *The MacArthur Communicative Development Inventories: User's Guide and Technical Manual*. San Diego: Singular Publishing Group.
- Fenson, L., Bates, E., Dale, P., Goodman, J., Reznick, S. y Thal, D. (2000). Measuring Variability in Early Child Language: Don't Shoot the Messenger. *Child Development*, 71(2), 323–328. <https://doi.org/10.1111/1467-8624.00147>
- Fernández Flecha, M.ª. A. (2012). Evolución de la conducta comunicativa infantil: estudio de caso de un niño castellano-hablante. *Tesis doctoral. Universidad Complutense de Madrid*.
- García, I., Arratibel, N., Barreña, A. y Ezeizabarrena, M.ª. J. (2008). Adaptación de los inventarios MacArthur-Bates al euskara: desarrollo comunicativo entre los 8 y 30 meses. *Infancia y Aprendizaje*, 31(4), 411–424. <https://doi.org/10.1174/021037008786140940>
- Garfinkel, H. (1967). *Studies in Ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.
- Goffman, E. (1981). *Forms of Talk*. Philadelphia: University of Pennsylvania Press.
- Guasch, O. (1997). *Observación participante*. Madrid: CSIC.
- Halliday, M. A. K. y Hasan, R. (1985). *Language, Context and Text: A social Semiotic Perspective*. Deakin University Press.
- Kuhl, P. K., Stevenson, J., Corrigan, N. M., van den Bosh, J. J. F., Deniz Can, D. y Richards, T. (2016). Neuroimaging of the bilingual brain: Structural brain correlates of listening and speaking in a second language. *Brain and Language*, 162, 1–9. <https://doi.org/10.1016/j.bandl.2016.07.004>
- Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, W. (2001). The anatomy of style-shifting. En P. Eckert y J. R. Rickford (Eds.), *Style and Sociolinguistic Variation* (pp. 85–108). Cambridge: Cambridge University Press.
- Landis, J. R. y Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- Lausberg, H. y Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods*, 41(3), 841–849. <https://doi.org/10.3758/BRM.41.3.841>
- López-Ornat, S., Gallego, C., Gallo, P., Karousou, A., Mariscal, S. y Martínez, M. (2005). *Inventario de desarrollo comunicativo MacArthur*. Madrid: TEA ediciones.
- MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk* (3rd edition.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Mariscal, S., Nieva, S., Gallego, C. y López-Ornat, S. (2006). La fiabilidad de los padres como informantes del conocimiento gramatical de sus hijos: los inventarios MacArthur en español. *Lingüística clínica y neuropsicología cognitiva. Actas del I Congreso Nacional de Lingüística Clínica*, 170–181.
- Mariscal, S., Nieva, S. y López-Ornat, S. (2010). Observar y medir el desarrollo gramatical temprano en español. *Psicothema*, 22(1), 51–56.
- Pearson, B. Z., Fernández, S. C. y Kimbrough, O. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning*, 43(1), 93–120.
- Pérez-Pereira, M. y García, X. R. (2003). El diagnóstico del desarrollo comunicativo en la primera infancia: adaptación de las escalas MacArthur al gallego. *Psicothema*, 15(3), 352–361.
- Plunkett, K. (1995). Connectionist approaches to language acquisition. En P. Flechter y B. MacWhinney (Eds.), *The Handbook of Child Language* (pp. 36–72). Oxford: Blackwell.
- Romero, A., de Pablo, I., Etxebarria, A. y Romero, A. (2017). Teorización sobre la construcción de corpus orales en la adquisición del lenguaje: una propuesta metodológica para el procesamiento de lenguas con soporte tecnológico. *Analecta Malacitana*, 42, 123–155.
- Serra, M. y Solé, R. (1995). *CHILDES database. Romance language corpora*. Disponible en: <https://childes.talkbank.org>.
- Siguan, M., Colomina, R. y Vila, I. (1990). *Metodología para el estudio del lenguaje infantil*. Barcelona: Abril Editorial.
- Villayandre, M. (2008). Lingüística de corpus (I). *Estudios Humanísticos Filología*, 30, 329–349.