Teorización sobre la construcción de corpus orales en la adquisición del lenguaje: una propuesta metodológica para el procesamiento de lenguas con soporte tecnológico

Asier Romero, Irati de Pablo, Aintzane Etxebarria y Ainara Romero

(a.romero@ehu.eus)

UNIVERSIDAD DEL PAÍS VASCO / EUSKAL HERRIKO UNIBERTSITATEA

Resumen

Hay una preocupación cada vez mayor por las condiciones de obtención y manipulación de los datos en la lingüística de corpus. Este trabajo analiza las facetas metodológicas en la constitución de corpus orales para investigar la adquisición del lenguaje, y valora las consecuencias de las decisiones que atañen a las técnicas de registro de muestras, y al tratamiento y codificación de los datos.

Palabras clave

Corpus
Lengua oral
Metodología
Procesamiento de lenguas

Abstract

There is a growing concern because of the conditions of obtaining and handling the collected corpus data in the linguistics of corpus. This paper discusses the methodological facets applied in the constitution of oral corpus in the investigation of the acquisition of the language, with the aim of assessing the consequences of the decisions pertaining to techniques for registration of samples and the treatment and coding of data.

Key words

Corpus
Oral language
Methodology
Language processing

***AnMal Electrónica* 42 (2017)**
ISSN 1697-4239

INTRODUCCIÓN

Desde mediados del siglo XX, se ha producido un cambio en la orientación teórico-metodológica de la lingüística, que ha generado un importante cambio a la hora de utilizar y manipular datos reales y contextualizados como fundamento de la

investigación. Esta nueva orientación se fundamenta en la convicción, cada vez más generalizada, de que para comprender el funcionamiento de las lenguas es necesario abandonar posicionamientos idealistas y reduccionistas en el acercamiento del investigador/a a las lenguas sujeto del análisis. Esta nueva convicción ha llevado a numerosos investigadores/as a cuestionar la validez de los estudios lingüísticos fundamentados sobre ejemplos surgidos de la autoobservación e introspección del lingüista, y ha posibilitado el nacimiento de nuevas disciplinas, como la sociolingüística variacionista norteamericana (Labov 1972, 1981 y 2001), el contextualismo británico (Firth 1957; Halliday y Hasan, 1985), la lingüística del texto (Hymes 1964b), el interaccionismo simbólico de Goffman (1981), la antropología lingüística (Hymes 1964a) o la etnometodologia de Garfinkel (1967). Todas estas nuevas disciplinas tienen en común el abandono del método hipotético-deductivo y la intuición del investigador/a como evidencias probatorias de los fundamentos teóricos, para basar sus afirmaciones en la observación y el análisis de datos del uso lingüístico real, extraído de contextos reales y representativos de hablantes reales. Por tanto, en base a esta nueva perspectiva metodológica, la lingüística ha pasado a ser una ciencia empírica, en la que el estudio sobre datos reales es uno de sus fundamentos argumentativos, incorporando además la variación social, geográfica y estilística y el modo oral como objetos de pleno derecho del estudio y análisis lingüístico.

El objetivo de esta investigación es describir las diferentes implicaciones metodológicas que se producen en la constitución de corpus orales en la investigación de la adquisición del lenguaje, con el objetivo de valorar las consecuencias de las decisiones que atañen a las técnicas de registro de muestras y al tratamiento y codificación de los datos grabados en soporte tecnológico. A la luz de estos datos, se propone una metodología observacional contextualizada sobre la base tecnológica de varias herramientas para el tratamiento de la información audiovisual.

SOBRE LA LINGÜÍSTICA DE CORPUS: LA NECESIDAD DE DATOS REALES Y SU CODIFICACIÓN

La necesidad de contar con datos reales para el análisis lingüístico ha supuesto un importante impulso en la construcción de corpus, convertidos en la actualidad en la base de la descripción, explicación y teorización lingüística en cualquiera de sus múltiples perspectivas o enfoques.

Ahora bien, la artificiosidad de los métodos utilizados en la recogida de lengua oral ha originado entre los investigadores/as una reflexión en torno a la problemática de si los *corpora* lingüísticos obtenidos con estos métodos sirven para extraer conclusiones sobre el funcionamiento de la interacción oral natural en situaciones no controladas por el investigador/a; o, por el contrario, tienen un alcance más limitado debido a que las conclusiones lingüísticas están lógicamente influidas por el propio método empleado, y cuya validez por tanto, es cuestionable.

Los efectos de la metodología utilizada sobre los datos obtenidos han sido analizados por numerosos autores. Si durante la primera mitad del siglo XX la lingüística estructural americana sentó las bases de la lingüística de corpus como metodología empírica basada en la observación de datos, con la aparición de la chomskyana a finales de los años 50 se impondrá un racionalismo como base para toda investigación en los estudios sobre el lenguaje (Chomsky 1957).

A partir de los años 60, se empezó a gestar una nueva corriente dentro de la lingüista de corpus caracterizada por dos elementos fundamentales: la presencia del ordenador y el carácter representativo de los datos. Los trabajos de Labov (1972) ya mostraban el alto porcentaje de secuencias gramaticales en un corpus aplicando la entrevista sociolingüística bien planificada y representativa, y con la finalidad de obtener muestras más próximas al habla vernácula que las producidas en el seno de la entrevista propiamente dicha.

El resurgir de la lingüística de corpus tuvo lugar a partir de la década de los años 80. En esta nueva perspectiva hay que destacar a varios autores, entre los que sobresale Leech (1992), que rebatió las críticas teóricas y prácticas que se habían formulado contra la primera lingüística de corpus. Los argumentos de Leech se centraron en considerar a la lingüística de corpus como una metodología científica, caracterizada por la gramaticalidad de los enunciados, la utilidad de los datos cuantitativos y el uso del ordenador para procesar los datos con un coste reducido.

AnMal Electrónica 42 (2017)
ISSN 1697-4239

Eisenstein y Bodman (1993) ofrecen datos sobre las repercusiones del método en la longitud y complejidad de las expresiones de gratitud en inglés, correspondiendo a los datos auténticos los que daban lugar a muestras más largas y complejas. Beebe y Cummings (1996) concluyen que existen ciertos aspectos en los que los métodos artificiales no reflejan el habla natural, en su mayoría elementos correspondientes a la prosodia y a la pragmática comunicativa. También Félix-Brasdefer (2007) alude a los efectos de la metodología de obtención de datos, subrayando la importancia del reconocimiento de los interlocutores o el espacio social y su influencia en las características de los actos de habla investigados frente a los métodos no naturales, como por ejemplo la representación de roles.

Con todo, los investigadores/as reconocen la utilidad de los métodos artificiales en la formación de *corpora* y su proximidad a los datos naturales; además, se admiten también las ventajas de utilizar este tipo de datos cuando se trata de describir estudios relacionados con la producción espontánea. En este punto, nos encontramos con la opinión generalmente asumida de considerar a la conversación coloquial como un género, en el que error o la inconsistencia constituían sus descriptores habituales. Ahora bien, recientes estudios parecen contradecir este tópico de desorden y simplicidad organizativa (Villayandre Llamazares 2008). Warren (2006) describe el papel fundamental que juegan las conversaciones en su papel de intercambio comunicativo y en la consecución de los objetivos de intercambio en la dicotomía función transaccional / función interactiva.

Moreno y Urresti (2005), al describir los problemas que plantea la transcripción de conversaciones frente a textos más formales, subrayan que los discursos formales presentan mayores problemas de transcripción frente a los diálogos espontáneos. Así, la dificultad de transcripción de la producción espontánea radica en los rasgos de interacción como palabras por turno, solapamientos, velocidad de elocución, etc.

Ochs (1979) y Siguán (1983), al referirse a la transcripción de producciones infantiles, destacan la tarea fundamental de seleccionar un adecuado sistema de transcripción, calificándola como decisiva e influyente en todo proceso investigador. Además, el tratamiento que reciban las producciones va a permitir reducir o ampliar la distancia entre la realidad lingüística y la descripción ofrecida.

Por su parte, López Morales (1994) señala que la confección de un corpus oral encierra ciertas complicaciones que no aparecen en la elaboración de un corpus de textos escritos. Las dificultades están ligadas a la riqueza informativa que contiene;

por tanto, intentar reflejar todos los elementos que integran la comunicación es, sin lugar a dudas, una tarea sumamente compleja. Partir de un documento escrito supone, en cierto modo, partir de una primera versión de la transcripción. Lógicamente, es necesario pulir los datos registrados, es decir, etiquetar los comportamientos susceptibles de convertirse en objeto de estudio.

En la actualidad el concepto de corpus ha cambiado notablemente con respecto al que manejaban los primeros lingüistas, que lo empleaban como recurso para sus investigaciones. Así, hoy en día se considera que los *corpora* deben cumplir con 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 Llamazares 2008).

Así mismo, hay distintas clasificaciones para establecer tipologías de corpus (Sinclair 1996; Torruella y Llisterri 1999). Atendiendo a los objetivos de este estudio nos centraremos en los *corpora* orales y en las dos tipologías en las que se suelen clasificar: corpus orales y corpus de lengua oral. En primer lugar, los «*corpora* orales» están orientados a la descripción fonético-fonológica de la lengua o al desarrollo de sistemas relacionados con el ámbito de las tecnologías del habla. Los *corpora* de este tipo se graban en condiciones muy controladas por el investigador/a y, especialmente, se centran en segmentos o frases aisladas, textos leídos o grabaciones y transcripciones de diálogos entre personas. En general, se diseñan con detalle para recoger el fenómeno objeto de estudio y tienen un tamaño reducido, al no utilizar un número elevado de hablantes. En segundo lugar, se encuentran los *corpora* desarrollados en el marco de la lingüística de corpus, que reciben la denominación de «corpus de lengua oral». En este caso, las grabaciones se realizan en entornos naturales y se favorecen las muestras espontáneas, no planificadas. El objetivo principal de este tipo de estudios no es tanto el análisis de las características de tipo segmental o suprasegmental, sino contar con una transcripción ortográfica de la lengua hablada, y efectuar diferentes análisis lingüísticos sobre el texto transcrito (Tabla 1).

AnMal Electrónica 42 (2017)
ISSN 1697-4239

	Corpus de lengua oral (Lingüística de corpus)	Corpus orales (Fonética y Tecnologías del habla)
Materiales	Habla espontánea (*unelicited speech*)	Corpus controlado (*elicited speech*)
Nivel de análisis	Discurso, diálogo	Enunciado
Obtención de datos	Entorno natural	Entorno controlado
Transcripción	Transcripción ortográfica	Transcripción fonética y ortográfica alineada con la señal sonora
Orientación	Representación simbólica, categorial	Señal sonora, representación temporal

Tabla 1. Principales características que diferencian los corpus de lengua oral de los corpus orales (fuente: Llisterri 1996)

Pese a existir un notable consenso en la lingüística moderna sobre la importancia de analizar el funcionamiento de la conversación en la descripción y caracterización de la lengua, no ha podido imponerse a los lógicos obstáculos éticos, técnicos y metodológicos que se producen en la recogida de una muestra de lengua espontánea oral. Esta circunstancia ha generado el desarrollo de diferentes métodos, caracterizados por su distinto grado de artificiosidad, que han propiciado un intenso debate sobre la validez y fiabilidad de los datos obtenidos. Se trata de analizar si los instrumentos utilizados para la formación de corpus lingüísticos sirven para obtener exactamente el tipo de datos que se desea obtener o, dicho con otros términos, «si los datos obtenidos a través de métodos artificiales son un reflejo fiable de su distribución real en el universo poblacional» (Recalde y Vázquez Rozas 2009: 56).

METODOLOGÍAS PARA LA RECOGIDA DE DATOS

En general, los métodos de formación y creación de corpus de lengua oral se clasifican en dos grupos: no intrusivos e intrusivos (Tabla 2). Lógicamente, hay una relación entre las deficiencias que presenta la lingüística de corpus y los problemas metodológicos derivados de la obtención de los datos que caracterizan a dicho

corpus. Esta elección metodológica afectará a la validez y fiabilidad de los datos obtenidos, y tendrá un alcance inevitable sobre las conclusiones extraídas de su análisis. Así, la mayor o menor validez otorgada a un estudio suele depender de la solidez del entramado metodológico sobre el que se sustenta y la confianza que suscitan tanto el proceso de selección del material como las técnicas empleadas en su análisis, factor influyente en la valoración de toda actividad científica (Fernández Pérez 1986).

No cabe duda de que las mejores muestras de habla natural son aquellas que se recogen reduciendo lo máximo posible la intromisión del investigador/a en la dinámica de la producción comunicativa. La base de estas técnicas no intrusivas reside en el principio de la observación etnográfica por parte del investigador/a, es decir, dado que el método y el investigador/a pueden interferir en la recogida de datos, es necesario controlar los efectos reduciendo su presencia todo lo posible. Por tanto, con esta técnica la observación es directa y el registro de la producción de los interlocutores queda al margen de cualquier intromisión del investigador/a. La no intromisión del investigador/a tiene importantes ventajas ya que permite a los propios interlocutores llevar a cabo su intervención, desarrollando con libertad la interacción oral, el intercambio de roles y las metas comunicativas. De esta forma, son los propios interlocutores los que estructuran su interacción en función de sus necesidades comunicativas y no el investigador/a en función de los objetivos y propósitos de la investigación.

En la última década, esta técnica de observación directa se ha visto beneficiada por importantes avances tecnológicos en la recogida de datos, y ha permitido, mediante el uso de sofisticados grabadores de audio y vídeo, obtener datos de gran calidad, lo que no sólo garantiza la fiabilidad de los datos obtenidos sino también su conservación, recuperación y tratamiento codificado de la información para su posterior análisis (Golato 2005). Este método ha sido ampliamente utilizado en la lingüística de corpus para recoger datos del habla natural; ahora bien, ha recibido también críticas, sobre todo en dos direcciones: la dificultad de gestionar las implicaciones éticas relativas al tratamiento de los datos personales de los informantes y la consideración por parte del análisis del discurso de la conversación coloquial como un género de segunda división, frente al uso transaccional e informativo de otros géneros más especializados (Brown y Yule 1983; Warren 2006). Además, a este tipo de corpus compuestos por datos naturales se les

ha achacado, también, su carácter asistemático, la escasa representatividad de las muestras, la dificultad de obtener con ellos la información necesaria sobre los interlocutores para un adecuado análisis lingüístico, y la generalización por parte del investigador/a de datos procedentes de su red social para caracterizar el conjunto de las normas sociolingüísticas de la comunidad de habla: «the family, colleages, friends, and acquaintances, not to mention the associated strangers, around a researcher are not necessarily a "speech community"» (Beebe y Cummings 1996: 68).

Estas dificultades para lograr una cantidad suficiente de datos naturales y auténticos y que permita el análisis lingüístico de fenómenos poco frecuentes, ha impulsado a los investigadores/as a desarrollar técnicas *ad hoc* para la recogida de muestras. Estas técnicas se basan en el diseño de un evento comunicativo artificial, a semejanza de aquellos que surgen espontáneamente en contextos naturales. Entre estas técnicas hay que destacar la representaciones de roles, las entrevistas semiestructuradas, etc. (Felix-Brasdefer 2007).

A tenor del objetivo de este estudio, centrado en la recogida de datos en el proceso de adquisición de la lengua oral en los niños/as, se analizará la técnica intrusiva de la representación de roles, uno de los métodos más frecuentemente utilizados para la obtención de datos en este tipo de investigaciones. En este caso, a los participantes (familiar, tutor/a, maestro/a y/o educador/a) se les dan instrucciones para guiar la comunicación con el niño/a, para reaccionar ante una descripción situacional y responder oralmente tal y como lo harían en una interacción espontánea conversacional. Los autores que defienden este tipo de técnica se basan en la semejanza existente entre la producción conversacional resultante con la representación de un rol y el discurso natural. Además, se resalta el carácter común de la interacción, la posibilidad de examinar un amplio rango de rasgos discursivos, la distribución de turnos, la coordinación entre interlocutores y la consecución conjunta de metas interpersonales derivadas del contrato conversacional (Kasper y Blum-Kulka 1993).

Las críticas a este método fueron numerosas y la mayoría centradas en la artificiosidad del contexto ficticio que genera este juego de roles (*role-play*), y que es además una situación de investigación creada por el investigador/a. En la representación de roles, la creación conversacional no existe y por tanto tampoco los diversos componentes contextualizados que condicionan el *output* lingüístico, por lo que su éxito dependerá en gran medida de la capacidad que el hablante tenga de

abstraerse de la realidad (Recalde y Vázquez Rozas 2009). Otro de los puntos en el que se centraron las críticas fue en la calificada como «paradoja del observador», o la dificultad de conseguir que los informantes que participan en el estudio se comporten o actúen como lo harían cuando no son observados. Las soluciones para minimizar este efecto han sido variadas; por ejemplo, mediante la construcción de relaciones solidarias con el informante o la presencia del investigador/a en un segundo plano (Labov 2001). Con todo, cuando se observa se influye en la acción, aunque el investigador/a se retire a un segundo plano, porque inexorablemente los interlocutores seguirán la presencia del investigador/a; por lo que en lugar de evitar este efecto habrá que trabajar con él, integrándolo y siendo consciente de su naturaleza insoslayable (Duranti 2000).

Método intrusivo	Método no intrusivo
Específicamente diseñado para recoger muestras sustitutivas del habla natural	Reduce la intromisión del investigador/a
Entrevista sociolingüística: posibilita la recopilación de muestras amplias, representativas, socialmente estratificadas y de buena calidad	Mayor fidelidad de los datos obtenidos
Permite la observación sincrónica del cambio lingüístico	Implicaciones éticas respecto al tratamiento de los datos
	Ampliamente utilizado en lingüística de corpus para obtener datos de habla natural
Dificultad para conseguir que los informantes se comporten como lo harían cuando no son observados («paradoja del observador»).	Dificultad para obtener datos naturales suficientes que permitan un análisis lingüístico de fenómenos poco frecuentes
Críticas a la entrevista semidirigida o a la representación de rol, por su contexto artificioso, género híbrido, ambiguo y poco apropiado para el estudio de la variación lingüística	Carácter asistemático y con escasa representatividad de las muestras

Tabla 2. Características de los métodos intrusivos y no intrusivos

Las investigaciones sobre el desarrollo del lenguaje han utilizado principalmente diseños transversales (se emplean dos o más grupos de sujetos al mismo tiempo y se comparan), longitudinales (se realiza un seguimiento de uno o más sujetos durante un tiempo definido, ya sean semanas, meses o años) o la combinación de ambos, que se ha denominado «diseño secuencial» (Baltes, Reese y Nesselroade 1977). Independientemente al tipo de diseño, la metodología puede ser observacional —investigando el comportamiento lingüístico espontáneo— o experimental —mediante la manipulación rigurosa de las variables de estudio—. La mejor cualidad que ofrece el método observacional reside en la no intervención del investigador/a, es decir, nada restringe al sujeto observado para su producción lingüística. Ahora bien, en el otro lado de la moneda, el investigador/a se debe conformar con lo que la situación le ofrece y por tanto sólo obtiene una muestra parcial del sistema que pretende investigar.

Por otra parte, el método experimental intenta provocar un determinado comportamiento verbal en el sujeto y de esta forma ofrecernos información más ajustada sobre un aspecto particular de la competencia lingüística. No hay que olvidar que con este método la obtención de datos proviene del control exhaustivo de las circunstancias de análisis, de manera que puedan llegar a verificarse las hipótesis de investigación planteadas. La dificultad de aplicar este método experimental en la investigación sobre adquisición del lenguaje, reside en conseguir que los sujetos de edad temprana colaboren en el estudio, ya que la motivación o la atención resultan claves. Además, junto con la dificultad anterior se suma una limitación importante, la poca naturalidad que plantea el contexto de análisis, es decir, los datos se recogen a menudo en situaciones poco naturales, de manera que la ansiada validez metodológica resulta cuestionada para muchos investigadores/as.

En el contexto de este método experimental, los investigadores/as utilizan distintos procedimientos para intentar obtener datos válidos y fiables sobre la producción, percepción o comprensión que están investigando. Entre estos procedimientos destacan: (1) pruebas estandarizadas del lenguaje; (2) tareas de imitación o técnicas de elicitación para estudiar la producción 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), y (5) actividades de comprensión para estudiar la comprensión lingüística.

Además de estos dos enfoques (observacionista y experimental), que surgen del comportamiento prelingüístico o lingüístico de los informantes, nos encontramos también con el método basado en la simulación. Este enfoque se basa en obtener la información a partir de datos proporcionados por la simulación del comportamiento lingüístico humano mediante ordenadores. En el campo de la adquisición del lenguaje, este enfoque 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).

Asimismo, tal y como expone Anguera (1990), se distinguen cuatro niveles en lo que respecta al nivel de participación del investigador: (1) en la *observación no-participante o externa*, el investigador/a no interacciona con los sujetos observados y conserva una distancia con ellos/as; (2) la *observación participante* es aquella en que el investigador/a toma parte en las actividades diarias de los sujetos observados para recolectar datos teniendo un contacto directo con ellos/as y tratando de causar la menor distorsión posible (Kluckholm 1940); (3) en la *participación-observación*, un miembro del grupo adopta el papel de observador participante de otros/as integrantes/sujetos del propio grupo, y (4) en la *auto-observación*, el propio observador es al mismo tiempo sujeto y objeto. En el ámbito de estudio que tratamos, la investigación del corpus oral de los niños/as, la auto-observación (4) no sería factible; de las tres restantes, opinamos que (2) crearía mayores distorsiones comparándolas con las otras, ya que el investigador/a necesariamente tiene que ser sujeto-paciente, y, respecto a la (3), es muy difícil encontrar dos docentes o familiares en que el primero/a actúe como investigador/a y el segundo como interlocutor/a, y se les forme previamente sobre cómo grabar, etc., debido a que no pueden desentenderse de las demandas y responsabilidades que tienen en el centro escolar y, en el caso de los familiares, no siempre se tiene la disponibilidad como para realizar éstas. Así, creemos que la más factible y adecuada es (3), la participación-observación, habiéndose dado un período previo de habituación con los niños/as, donde el investigador/a no sea un elemento desconocido, sino uno más.

AnMal Electrónica 42 (2017)
ISSN 1697-4239

PROPUESTA METODOLÓGICA: LA «OBSERVACIÓN CONTEXTUALIZADA»

El análisis de las producciones espontáneas de un grupo heterogéneo de niños/as plantea necesariamente la exposición de ciertas consideraciones metodológicas. La mayor o menor validez otorgada a un estudio suele depender de varios condicionantes, entre los que destacan: (1) la solidez del entramado metodológico sobre el que se sustenta, y (2) la confianza que suscitan tanto el proceso de selección del material como las técnicas empleadas en su análisis. Por tanto, atendiendo a estas consideraciones, presentar y explicar la metodología es, en cierto modo, describir y explicar todo el estudio.

En primer término y, en el ámbito del lenguaje infantil, es necesario reflexionar sobre las peculiaridades del objeto de estudio, es decir, hay que determinar cuáles son los caminos de análisis adecuados en cada caso. Consiguientemente, se debe conocer el qué para decidir el cómo. En palabras de Taylor y Shanker,

> While different conceptions of what the child acquires lead ineluctably to contrasting accounts of how it is acquired, the reverse is not always true. […] However, the salient differences in their account of how the child acquires language are not matched by fundamental differences concerning their conceptions of what is acquired (2003: 155).

La reflexión sobre los métodos de análisis de datos trae inexorablemente una visión excluyente entre las investigaciones que emplean técnicas cuantitativas y aquellas que recurren a procedimientos metodológicos cualitativos. En el ámbito de la lingüística esta visión se ha visto reforzada por el diferente devenir histórico que han tenido ambas técnicas metodológicas. La investigación cuantitativa ha estado relacionada con el análisis estadístico, mientras que la metodología cualitativa se ha situado en la órbita de la antropología y etnografía. Lógicamente, el punto de partida de ambas metodologías es diferente pero su aplicación no debe seguir caminos diferentes. Por tanto, la visión integradora y no excluyente es la que se propone actualmente como mayoritaria en la literatura científica. Ésta es conocida como *metodología mixta*, en la que los métodos cualitativos y los cuantitativos se complementan entre sí para la recogida de datos y el análisis de estos (Ugalde y Balbastre Benavent 2013). Un investigador puede utilizar principalmente un único

modo de análisis como herramienta metodológica básica, sin omitir o invalidar los resultados que le muestren otros métodos (Ruiz Olabuénaga 1999). La perspectiva que se debería seguir es la de adaptar la metodología al problema de análisis, para dar la mejor respuesta posible a las necesidades de la investigación (Reichardt y Cook 1982).

En la medida en que la investigación sobre la adquisición del lenguaje sea de naturaleza empírica, estamos refiriéndonos a un ámbito de conocimiento en el cual las hipótesis y predicciones han de ser contrastadas sistemáticamente con los datos procedentes del comportamiento lingüístico de los sujetos. En este contexto, las investigaciones sobre el desarrollo del lenguaje han utilizado, principalmente, diseños transversales o longitudinales. Por lo tanto, en las investigaciones centradas en el lenguaje infantil se puede lograr un importante grado de integración metodológica. Así, en primer lugar se puede analizar un determinado aspecto del proceso adquisitivo a través de diferentes técnicas cuantitativas y añadiendo a continuación las valoraciones cualitativas oportunas. Lógicamente, puede dar la impresión de que la metodología cualitativa posee un valor subsidiario tras la detallada exposición cuantitativa de los datos obtenidos; pero, como ya hemos explicitado, creemos que es posible alcanzar un alto grado de integración, sin omitir otras posibilidades de análisis. La literatura científica ya ha señalado la pertinencia de utilizar enfoques cuantitativos y cualitativos en un mismo trabajo, permitiendo de esta forma salvar determinados obstáculos (Bryman 1992; Bericat 1998).

El empleo de una u otra metodología en el estudio de los *corpora* lingüísticos está caracterizado por una serie de ventajas y limitaciones. Así, por ejemplo, una metodología cuantitativa en el estudio de un corpus fonológico amplio permite determinar el peso estadístico de ciertas variables, las frecuencias relativas que caracterizan un determinado fenómeno, la comparación de resultados de diversos trabajos, la generalización de los resultados obtenidos y su asociación con los representativos de todo un grupo, etc. Como limitación se podría señalar la dificultad para presentar a través de cifras ciertos factores implicados en la interacción comunicativa.

En cuanto a la ventaja de aplicar una metodología cualitativa, recae sobre todo en la posibilidad de precisar valorativamente la influencia que poseen ciertos elementos contextuales, interpretando los hechos y profundizando en el análisis de algunos aspectos que en otros estudios son presentados como tangenciales. Entre las

desventajas en la aplicación de este tipo de metodología se han señalado la imposibilidad de aplicar un enfoque estadístico que permita un control más riguroso de las variables y la imposibilidad de generalizar los resultados. Esta valoración no debe llevarnos a interpretar el estudio de las variables como una actuación enriquecedora tan sólo desde un punto de vista cuantitativo, ya que tanto los nuevos programas y aplicaciones de análisis cualitativo como los estudios multimétodo de investigación cualitativa pueden facilitar la interpretación de relaciones entre variables (Bericat 1998).

La aplicación de una observación contextualizada en el análisis de la adquisición del lenguaje parte 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. La 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 que efectúa el investigador/a, el lugar en el que realiza la investigación o el tipo de actividad que realiza. La observación contextualizada vendría a ser el modelo de observación participante utilizada en el ámbito educativo, y la génesis de esta técnica de recogida de información se localiza en la antropología y la sociología. 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 e ir recogiendo datos sobre su vida cotidiana. Así, compartimos las ideas de que «el conocimiento científico es antes que nada observable» (Ramos, Catena y Trujillo 2004: 24), y de que «la adquisición del lenguaje no debe verse como un campo racional, sino como terreno de hechos» (Tomasello 2003: 328). Se trata, por tanto, de una técnica en que lo prioritario son los datos de adquisición y en la que prima la conveniencia de un análisis próximo a la realidad lingüística.

La planificación de la observación contextualizada tiene un carácter inductivo y versátil, donde el investigador/a se integra en la situación natural de grabación, la cual va acomodando a medida que avanza con el objetivo de dar respuesta a los interrogantes de la investigación. Hay dos elementos fundamentales en el diseño de la observación contextualizada en el entorno de una investigación sobre adquisición del lenguaje con niños/as: (1) el grado de participación del investigador/a, y (2) la

estrategia a desarrollar para introducirse en el contexto natural de observación. El grado de participación está referido al papel que tiene que adoptar el investigador/a en el escenario de investigación. Lógicamente, la investigación con bebés y niños/as implica una acomodación más implícita, ya que es necesario lograr una ambiente de seguridad y confianza en los niños/as. En este sentido, es recomendable un periodo de adaptación del investigador/a con los informantes, con el objeto de que tanto el investigador/a como el instrumental asociado a la investigación puedan integrarse con una equilibrada naturalidad. En esta línea, el investigador/a puede optar por encubrir su status; ahora bien, esta perspectiva no eliminará otra serie de riesgos, como los límites en la posibilidad de observar fuera del contexto de investigación, la pérdida de una perspectiva global, la dificultad para mantener la objetividad o los problemas éticos que se puedan derivar (Ruíz Olabuénaga 1999). La integración secuenciada del investigador/a en el escenario de investigación permitirá una adecuación positiva de los bebés o los niños/as. Además, no hay que olvidar que el objetivo de la investigación es obtener un corpus prelingüístico y/o lingüístico por lo que lograr un clima de confianza y seguridad resultará básico.

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 confidencialidad 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 y docentes), básicos en la tarea fundamental de introducir al investigador/a en el contexto de observación. Esta función esencial de los informadores se refiere a su función de intermediario para que el investigador/a no sea visto como persona no objetable y, por tanto, como suministrador de la información necesaria sobre el niño/a para que el registro audiovisual de la observación sea lo más relevante posible.

Tratamiento de los datos

Las pautas que guían la construcción del corpus sobre adquisición del lenguaje recaen en cuatro elementos fundamentales: los informantes, los criterios de

grabación, la transcripción y su codificación y, por último, el análisis de los resultados.

A) Los/as informantes. Los datos son un requisito indispensable para iniciar cualquier proceso de investigación. Si la compilación de datos se realiza de forma individual el proceso se complica, lógicamente. El uso de corpus colectivos —nos referimos a proyectos que poseen bases de datos sobre adquisición del lenguaje infantil, como Child Language Data Exchange System (CHILDES) y Corpus de Habla Infantil Espontánea del Español (CHIEDE)— reduce considerablemente la amplitud y minuciosidad necesaria respecto al recabado de forma individual. Con todo, este estudio parte con la idea de que es el propio investigador el que inicia una base de datos con producción de lenguaje infantil. Por tanto, hay diferentes parámetros fundamentales a la hora de iniciar la investigación y que están centrados en los informantes del corpus: adscripción geográfica, edad, número y situación social de los informantes. El rigor en la aplicación de estos parámetros será clave, ya que no hay que olvidar que estos elementos pueden formar parte de las variables de la investigación. Por tanto, la amplitud y dispersión de la muestra será crucial para lograr que el corpus seleccionado sea representativo.

El primer elemento a tener en cuenta es la selección del lugar de grabación: localidad y tipo de centro escolar. Los objetivos de la investigación se centrarán de una forma más certera en la selección geográfica, sin olvidar que si entre los objetivos se encuentra la descripción de la lengua materna de los informantes, la selección geográfica vendrá dada por el desarrollo propio de la lengua atendiendo a la localidad de residencia del informante. Se piensa, por ejemplo, en una investigación que trate de analizar la adquisición del vasco como lengua materna. Lógicamente, la selección geográfica en este caso vendrá dada en primer término por el desarrollo sociolingüístico del vasco en cada localidad. En cuanto a la selección del centro escolar, debe estar relacionada con las etiquetas de familiar y habitual, además de tener en cuenta la tipología del centro, público o concertado. Además, distintas contingencias —enfermedades, periodos vacacionales…— pueden motivar que la investigación tenga que continuar en el domicilio del informante.

El segundo parámetro concierne a la edad; como es lógico, el objetivo de la investigación fijará los márgenes cronológicos del estudio. Las edades elegidas para el registro de producciones pueden justificarse atendiendo al análisis de muy diversos

aspectos de la lengua del niño/a. Ahora bien, es importante fijar de forma homogénea el comienzo y el final cronológico de las grabaciones, y escalar o fijar de forma flexible un número de intervenciones en cada etapa cronológica del niño/a. De esta forma, a través unas tablas, por ejemplo, se apreciará con claridad la importancia concedida a un periodo concreto, si adoptamos un enfoque transversal, o a una evolución, si estamos ante un estudio longitudinal. Lógicamente, se intentará buscar un equilibrio y regularidad en el número de grabaciones para cada tramo o periodo, aunque teniendo en cuenta que las oscilaciones en el número de producciones registradas en uno y en otro momento es una situación, que en cierto modo, se presenta como inevitable al estar sometiendo a examen producciones espontáneas (Tabla 3).

Siglas que identifican al informante	Edad de la grabación	Número de grabaciones en que participa	Duración de todas las grabaciones del informante
MBR [masc.]	1;01.20	2	53’ 26”
ELB [masc.]	1;08.17	1	19’ 11”
ALL [masc.]	1;07.2	2	1h. 01’ 09”
ARL [fem.]	1;3.23	2	48’ 19”
LBG [fem.]	1;02.11	1	21’ 11”
LBR [fem.]	1;06.5	2	35’ 17”

Tabla 3. Ejemplo de cómo se pueden detallar los datos de los informantes

En cualquier estudio sobre adquisición del lenguaje, la edad de los informantes influye de forma directa en el comportamiento de los niños/as y, por tanto, en la recogida de los datos. En consecuencia, hay distintos elementos que pueden condicionar y distorsionar los datos: la paradoja del observador, la pérdida de naturalidad en las grabaciones, la imposibilidad de realizar una sesión programada, etc. Al lado de estas desventajas, debemos señalar una ventaja, y es precisamente la corta edad de los informantes seleccionados, ya que ese periodo de adaptación y acomodación al contexto que ha realizado el investigador/a, le permitirá disminuir los recelos entre los niños/as y, por tanto, recoger datos más naturales y fiables. En todo caso, el alto grado de improvisación con el que tiene que actuar el investigador/a resultará clave en esta metodología.

El tercer parámetro se refiere a la elección del número de informantes, que está influida por la necesidad de mantener un equilibrio en uno de los parámetros que se presentan como básicos en casi todos los estudios sobre adquisición del lenguaje. Nos referimos al sexo al que pertenecen los informantes. Por tanto, la búsqueda de un equilibrio entre niños y niñas resulta fundamental. Ahora bien, la exigencia metodológica de que los datos, para poder considerarse representativos, deben ser amplios, nos obliga a ser exigentes en la selección extensa, pero sin olvidar que cada informante representa un número importante de horas de grabación, repartidas en diferentes sesiones. En estas circunstancias es aconsejable limitar el número de informantes y decidir sobre la suficiencia de las grabaciones analizadas. Esta limitación no es sencilla. La literatura científica (Tognini-Bonelli 2001; Tomasello y Stahl 2004; Rowland, Fletcher y Freudenthal 2008, entre otros) ya ha señalado las dificultades existentes para determinar la representatividad de un corpus. Además, las trabas aumentan cuando trabajamos con este tipo de informantes; por ejemplo, la imposibilidad de establecer un seguimiento individual en diferentes momentos, o la baja frecuencia de aparición de un fenómeno estudiado. Esta circunstancia obliga a actuar con precaución al generalizar los resultados obtenidos, contrastando los resultados con los realizados por otros autores en otros estudios. En todo caso, la elección del número de informantes tiene que estar relacionada con el tipo de análisis que se vaya a realizar: transversal, longitudinal o una combinación de ambos.

Por último, se analizará la situación social de los informantes; aunque no todos los estudios sobre adquisición del lenguaje lo hacen, pensamos que su toma en consideración es interesante, ya que estos factores sociales intervienen significativamente en el modo y en la proporción de desarrollo de la lengua en las distintas etapas (Fernández Pérez 2003). En la selección de informantes, la preocupación por las condiciones sociales en las que se encuentra inmerso el niño/a puede atenderse desde: (1) la estructura social en la que se encuentra inmerso el niño/a (estrato social, nivel cultural, perspectiva sociolingüística, etc.), y (2) las capacidades socio-cognitivas del niño/a para integrarse en la estructura social (Tomasello 2003). La selección de los informantes debe tener en cuenta estas condiciones sociales para fijar los rasgos que determinen la heterogeneidad social de los informantes. Por el contrario, no tener en cuenta el elemento social, puede incidir en la representatividad de los datos y distorsionar los resultados alcanzados.

B) Criterios de grabación. Estos afectan inexorablemente a todo el estudio. Al describirlos, estaremos reflejando los controles y las normas de actuación de todos los participantes en el proceso de recogida de los datos. La minuciosidad tenida en cuenta al revisar las condiciones que influyen en el registro de los datos repercutirá en los parámetros de fiabilidad del estudio. Principalmente, hay que tener en cuenta cuatro factores: la elección del material, el tipo de producciones registradas, la elección de las actividades y los interlocutores, y la duración y frecuencia de las grabaciones.

El equipo tecnológico necesario para recoger los datos es uno de los aspectos que ha permitido avanzar de una forma espectacular con relación a la obtención de información. Las actuales técnicas modernas posibilitan la recogida rigurosa de datos sobre el uso gestual, prelingüístico o lingüístico de los niños/as. Las innovaciones, como las videocámaras con micrófonos incorporados para grabar audio de alta calidad, minicámaras que pueden estar localizadas en lugares estratégicos, la posibilidad de incorporar a éstas dispositivos como micrófonos externos de solapa o micrófonos de cañón con pértiga, etc., han simplificado enormemente la recogida de datos en este campo. Lógicamente, el objetivo de la investigación motivará la elección de un material más o menos cualificado, aunque atendiendo a lo que nos aporta el mercado, la oferta de dispositivos es muy variada y con un nexo en común, la calidad técnica de los dispositivos. Ahora bien, la calidad de los dispositivos no disminuye los problemas metodológicos, es decir, la utilización de unos dispositivos de grabación tiene que ir acompañado de un periodo de adaptación en que el niño/a se habitúe y tome conciencia de su presencia[1].

En cuanto al tipo de producciones registradas, si utilizamos tanto el método observacional como el experimental, serán producciones orales en el contexto de un ambiente natural o no. Lógicamente, acercarse a la lengua en su contexto natural acrecienta la cantidad de datos registrados y, sobre todo, su fiabilidad (Serra *et al*.

[1] Al margen de este material tradicional empleado en los estudios experimentales u observacionales clásicos, los nuevos procedimientos para estudiar los mecanismos cognitivos y cerebrales asociados a la adquisición del lenguaje y a su procesamiento, han obligado a dotar a los laboratorios de una tecnología sofisticada y muy costosa. Nos referimos a diferentes técnicas como la TEP (tomografía por emisión de positrones), la RMF (resonancia magnética funcional), la TAC (tomografía tranxacial computerizada) o la MEG (magnetoencefalografía).

AnMal Electrónica 42 (2017)
ISSN 1697-4239

2013). Para el examen de fenómenos de baja frecuencia en el corpus se recurre por lo general a la utilización de técnicas elicitativas, que imposibilitan en este caso la observación del niño/a en un contexto natural, pero logran resultados que cumplan con los objetivos de la investigación. Además, la observación en contextos naturales ofrece datos muy desordenados, lo que dificulta ciertos análisis. En todo caso, algunos investigadores/as ya han señalado, con independencia del método a utilizar, la pertinencia de crear un ambiente de interacción con el niño/a que tenga la mayor naturalidad posible (Bosch 2004; Fernández López 2009; Monfort y Juárez 1989).

Las investigaciones sobre adquisición del lenguaje con niños/as imposibilitan una rígida planificación de las sesiones con un estricto patrón organizativo de las actividades. Así, la actitud tiene que ser flexible, con el objetivo de establecer una interacción dinámica y abierta, en la que el planteamiento de las actividades no parezca de obligado o forzoso cumplimiento, sino como algo opcional. La tipología de actividades puede ser muy variada: narrar cuentos (cuentos con animales desplegables), jugar con objetos/estímulos familiares para los niños/as, verbalizando sus acciones (globos, pompas, set de juego simbólico, juguetes autopropulsados, muñecas, etc.), dar de comer, juegos de baño, etc. (Belinchón Carmona 1985). En todo caso, dentro de esta heterogénea diversidad se plantea un elemento en común: siempre se tiene que intentar mantener un diálogo con el niño/a, tal como se hace normalmente en situaciones cotidianas de interacción. En cuanto al tránsito de una a otra actividad, el ritmo de la propia situación interactiva y el interés del niño/a hacia cada una de las situaciones deben anteponerse a los criterios temporales.

Este proceso interactivo es básico para que el niño/a verbalice con naturalidad, evitando posiciones en las que el interlocutor puede resultar pasivo u observador. Además, hay que señalar que la corta edad de los informantes no es un obstáculo a la hora de planificar actividades basadas en el diálogo. La literatura científica (Bruner 1984; Miller y Lossia 2013; Mundy y Gomes 1998; Tomasello y Farrar 1986; Tomasello 1988) proporciona evidencias directas sobre la capacidad comunicativa multimodal y las rutinas de alternancias de turnos en los bebés a los pocos meses de vida, y retrasándose hasta el año y medio las primeras técnicas conversaciones, coincidiendo con la etapa lingüística de las primeras palabras (Puyuelo Sanclemente 2000; Vihman y Miller 1988). Por tanto, es factible trabajar con informantes de tan corta edad con la interacción como procedimiento para la obtención de datos, aunque la decisión de interactuar con un tipo u otro de actividad no está exenta de dificultades; pero estos

obstáculos se centran principalmente en la codificación del material y en el análisis de las grabaciones (Aguado 2010; Fernández López 2009).

En los estudios sobre el lenguaje infantil, el interlocutor es una de las piezas clave en el proceso de investigación. A este respecto, la bibliografía existente nos muestra que la presencia del adulto, conocido para el niño/a, influye de forma positiva en la calidad del material registrado[2]. La riqueza que surge de la interacción niño/a-adulto es quizá una consecuencia del valor intencional que inevitablemente poseen las intervenciones del adulto. Tanto en las grabaciones como en la vida diaria, actúa como organizador de las interacciones, y, por supuesto, como modelo que marca pautas de aprendizaje. La presencia del adulto se constituye como un elemento habitual en el proceso adquisitivo. No hay que olvidar que para que éste se desarrolle con normalidad debe llegar a establecerse una «interacción íntima con adultos ya hablantes» (Clemente Estevan 1998: 42).

Finalmente, haremos referencia a la duración y frecuencia de las grabaciones. Al iniciar un estudio sobre la adquisición del lenguaje hay que tener en cuenta que registrar todas las producciones del niño/a es imposible, por lo que la selección de grabaciones es inevitable. Además, el carácter longitudinal o transversal del estudio, la edad de los informantes o los objetivos concretos del estudio, influirán en la toma de una decisión final. En todo caso, la búsqueda de un equilibrio en la frecuencia de las grabaciones es necesaria, ya que la observación diaria generaría una acumulación de material de compleja gestión, y por otra parte, renunciar a un seguimiento detallado (semanal o quincenal) imposibilitaría observar elementos lingüísticos que necesitan de un seguimiento más cercano. En cuanto a la duración, lo normal es que tiempo de cada grabación dependa de la mayor o menor actitud de colaboración que muestre el informante en cada momento. Así, la decisión que planteamos de evitar un estricto patrón organizativo repercutirá en la calidad de la sesión, evitando centrarse en las divergencias sobre la duración de la grabación para centrarse en las posibilidades productivas del niño/a en cada momento.

[2] Las referencias que podemos citar al respecto son numerosas. Sin salirnos de la década de los 80, encontramos alusiones a la enriquecedora presencia del adulto en Siguán (1983: 9), Bruner (1984: 217), Man Shum (1986: 288) y Boada (1986: 17-18). Incluso, desde que Newport, Gleitman y Gleitman (1977) acuñaron el término *motherese* para el habla materna, las peculiaridades del lenguaje que el adulto dirige al niño/a han sido el tema central de diversos trabajos (Rondal 1979 y 1980; Snow 1986; Moerk 1989; Díez-Itza 1995; Lieven 1994).

C) Tratamiento de los datos: transcripción y codificación. Las posibilidades técnicas actuales permiten la integración de diferentes formatos de registro, junto con versiones escritas del componente verbal de interacción, facilitando la tarea de los investigadores/as, que podrán tomar en consideración el carácter multimodal de la comunicación oral (elementos lingüísticos o gestuales, por ejemplo). Lógicamente, resultará imposible transcribir toda la amplia gama de variedades y peculiaridades de la lengua oral, incluso utilizando un software específico, la fijación gráfica del habla implica transformar un proceso dinámico en un producto textual estático. Además, implica atribuir secuencialidad a lo simultáneo, e inevitablemente conlleva perder numerosos elementos comunicativos presentes en el habla.

Atendiendo a estas circunstancias, y siendo conscientes de las limitaciones, la transcripción y el etiquetado de codificación se pueden preparar de manera que sea posible utilizarlas como material para ser analizado mediante programas informáticos. De esta forma, es posible la extracción de montajes de los diversos aspectos estudiados, etiquetando los comportamientos susceptibles de convertirse en objeto de estudio. La elección de uno u otro sistema de transcripción está relacionado con los objetivos o el tipo de análisis que se pretende realizar. La elección es crucial, ya que la fijación de unos criterios y los resultados de una transcripción nunca son neutrales, sino que reflejan presuposiciones sobre el análisis, prejuicios e interpretaciones previas al análisis (Ochs 1979).

En la transcripción, codificación y análisis de datos, una de las propuestas más utilizadas en la investigación sobre la adquisición y desarrollo del lenguaje, es la base de datos de lenguaje infantil del ya citado proyecto CHILDES (Child Language Data Exchange System). Junto con la base de datos (Talkbank), CHILDES ofrece un formato de transcripción CHAT, Codes for the Human Analysis of Transcripts (Figura 1) y un conjunto de programas para el análisis de la lengua, denominado CLAN (Child Language Analysis Programs). Las ventajas del sistema CHILDES son básicamente las siguientes: permite compartir datos entre los investigadores/as, facilita la recogida de datos, permite incrementar la precisión y la estandarización de la codificación, y también facilita la automatización de muchos procedimientos de codificación y análisis (MacWhinney 2000).

Junto con las opciones que proporciona la plataforma CHILDES, hay otra serie de herramientas para el análisis de los corpus orales: (1) herramientas para el análisis acústico del habla y para el etiquetado de corpus orales: PRAAT y PHON; (2)

herramientas para la transcripción y anotación de corpus de lengua oral y corpus multimodales: ELAN. Llisterri (2007) ofrece una amplia lista de recursos.

Grabación: HAURNET_13
Participantes: ARL (niño), ARA (adulto)
Edad: ARL, 2 años y dos meses
*ARA: Ea esaidazu zer ikusten duzun hemen? (venga, dime, ¿qué ves aquí?).
%act: ha cogido un muñeco del suelo.
*ARL: Hau atoa da (esto es un gato).
%par: atoa = katua $PHO.
%par: señala el dibujo del gato en el cuento.
*ARA: Ziur zaude? (¿estás seguro?)
*SPB: /bai gatua da/
%par: gatua = katua $PHO.

Figura 1. Ejemplo de transcripción en formato CHAT

Los programas PRAAT y PHON permiten analizar las señales acústico-visuales, síntesis articulatoria, procesamiento estadístico de los datos de la voz, edición y manipulación de señales de audio, etc. El programa PRAAT es una herramienta gratuita para el análisis fonético del habla desarrollada por en la Universidad de Ámsterdam (Boersma y Weenink 2016). Este programa es una herramienta de gran utilidad para los estudios fonéticos del habla, ya que posibilita la observación de las características de los parámetros de emisión de voz y, especialmente, en la evaluación y observación de las características del timbre por medio del análisis del espectograma de producciones acústicas grabadas. La implementación de este programa en las investigaciones sobre adquisición del lenguaje es muy habitual; así, no sólo nos encontramos con trabajos que estudian la fonética o la fonología con este software, sino que el estudio de los patrones prosódicos es también uno de los campos que más se han beneficiado de PRAAT (Figura 2).

El programa PHON (Rose *et al*. 2007) es un software gratuito y de soporte digital que, diseñado para estudiar el componente fonológico de una lengua, posibilita el estudio de la adquisición de L1 en un intento de conocer y comprender los mecanismos que subyacen al proceso de adquisición y los factores que inciden en él. El programa permite seguir la metodología RETHAME, de registro, transcripción y

AnMal Electrónica 42 (2017)
ISSN 1697-4239

análisis de muestras de habla espontánea. Además, es un software que incluye cinco niveles de representación fonético-fonológica que permiten reflejar con exactitud el grado de desarrollo fonético-fonológico del niño/a en cada etapa de desarrollo lingüístico. Permite también la alineación de estos niveles con los archivos de audio correspondientes, dando cobertura tanto a los datos segmentales como a los prosódicos o suprasegmentales. En definitiva, el programa PHON posee múltiples funciones que permiten estudiar la evolución del lenguaje en la adquisición de una L1. (Más información sobre este programa en Ramírez Cruz [s. f.].)

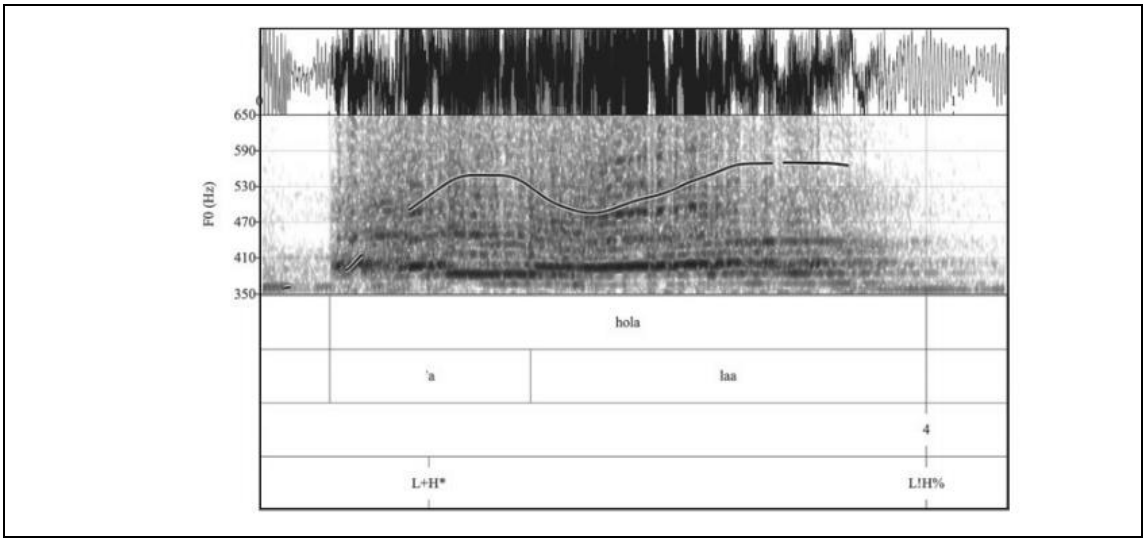

Figura 2. Ejemplo de espectograma del contorno de F0 y etiquetado prosódico para la palabra *hola* (Prieto *et al.* 2011)

Por otra parte, el software de transcripción multimodal ELAN (Lausberg y Sloetjes 2009) es una herramienta profesional con licencia freeware para la creación de anotaciones complejas sobre los recursos de audio y vídeo. Con ELAN se puede agregar un número ilimitado de anotaciones (palabras, frases, comentarios, etc.) de las características audiovisuales observadas. Estas anotaciones se pueden codificar en diferentes capas o tiras, y pueden estar interconectadas jerárquicamente (vocalizaciones, gestos, funciones pragmáticas, etc.). Las ventajas que ofrece este programa en la investigación sobre la adquisición del lenguaje son múltiples: la posibilidad de visualizar la onda de los archivos .wav, un potente sistema de búsqueda, la opción posibilidad de importar y exportar archivos desde los programas CHAT y PRAAT (Figura 3).

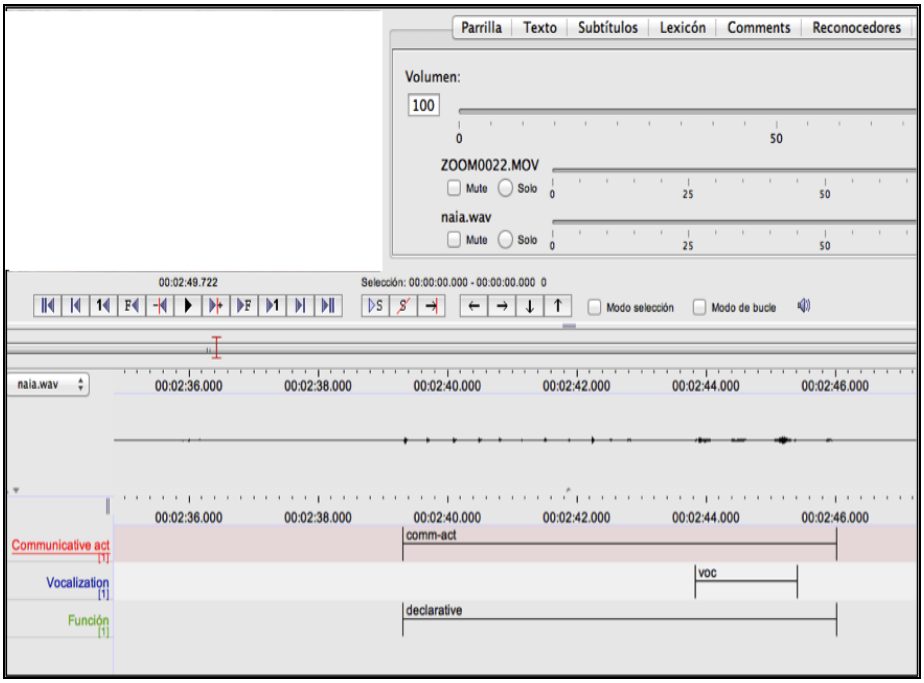

Figura 3. Imagen de ELAN con tiras de anotación

D) Análisis estadístico de los resultados. En la última década, el análisis de resultados en las investigaciones sobre el lenguaje infantil ha experimentado un importante cambio a través del análisis estadístico de los datos. Cuando los resultados de la investigación son verificados por investigadores/as que utilizan un método o diseño distinto, la confianza en los resultados también queda reafirmada. De hecho, cuantos más medios se utilicen para una verificación, más fiables son los resultados. Es más, esta fiabilidad se acrecienta si las nuevas técnicas de obtención de datos corroboran los resultados previos.

El procedimiento estadístico utilizado en la interpretación de datos está relacionado con el marco metodológico de la investigación. Así, unos mismos datos pueden recibir diferentes interpretaciones en función del marco teórico o de la perspectiva teórica que se asuma. Pero también recibirán interpretaciones diferentes según sea el procedimiento estadístico utilizado. Por tanto, el rigor metodológico es la vía para responder a las cuestiones principales que preocupan a quienes estudian la adquisición del lenguaje; entre otras, aquellas que se centran en el curso del desarrollo del prelenguaje, lenguaje y su interrelación con la comunicación intencional y multimodal, o las que se interesan por las causas del cambio evolutivo.

Para estas dos líneas, la literatura científica utiliza el procedimiento estadístico para la interpretación de resultados.

No es el objetivo de este estudio reflejar toda la amplia gama de recursos estadísticos utilizados en el análisis de resultados de las investigaciones centradas en la adquisición de la lengua, aunque sí señalar los más comunes: (1) con el objetivo de calcular la fiabilidad de la codificación o transcripción realizada, se puede utilizar el procedimiento de acuerdo inter-jueces con dos evaluadores externos e independientes y mediante el cálculo del índice Kappa (Cohen 1960), y sobre un conjunto que oscila entre el 15-20%; (2) con el objetivo de determinar si existe una relación significativa entre dos (o más variables), se emplean diferentes pruebas estadísticas, entre las que destacan: Chi-cuadrado de independencia y el Coeficiente de correlación de Pearson. Lógicamente, para la comparación de medias de dos o más muestras, y teniendo en cuenta el nivel de relación o independencia de las mismas, se utilizan también, con frecuencia, el Análisis de la Varianza o ANOVA o la prueba T de Student; y, finalmente, (3) en los estudios de carácter longitudinal, se utiliza específicamente el Modelo Lineal Mixto, LMM (West *et al*. 2007). La aplicación de este modelo es sin duda el más apropiado para analizar datos de carácter longitudinal (correlacionados, incompletos y con intervalos entre observaciones no constantes). Por tanto, los modelos lineales mixtos permiten analizar este tipo de datos, modelando, por un lado, la evolución de la respuesta promedio en función del tiempo y otras covariables mediante efectos fijos (estructura media) y, por otro lado, la variabilidad entre las medidas repetidas dentro y entre sujetos por medio del error y los efectos aleatorios (estructura de covariancia).

CONCLUSIONES

Desde mediados del siglo XX, se ha descrito la necesidad de reflexionar si realmente las metodologías que actualmente se están utilizando para el registro del corpus oral de niños y niñas reflejan la realidad diátopica lingüística o, por el contrario, son un cómputo de datos obtenidos a través de situaciones influidas por el propio método y, por tanto, artificiales. Las repercusiones de la metodología empleada han sido examinadas por múltiples autores: Chomsky (1957), Abercrombie (1965), Beebe y Cummings (1996), entre otros, que han expuesto distintos

razonamientos sobre las metodologías utilizadas, con el nexo común entre las mismas de presentar carencias que hacen que los resultados logrados tengan siempre algún tipo de hándicap metodológico.

Por tanto, todas las nuevas disciplinas, como la sociolingüística variacionista norteamericana (Labov 1972, 1981 y 2001), el contextualismo británico (Firth 1957; Halliday y Hasan 1985), la lingüística del texto (Hymes 1964b), el interaccionismo simbólico de Goffman (1981), la antropología lingüística (Hymes 1964a) o la etnometodologia de Garfinkel (1967), ponen de manifiesto que, para que un resultado sea fiable, es esencial basarse en la observación y el análisis de datos de uso lingüístico real y representativo. Por todo ello, aquí se ha hecho referencia a diferentes metodologías con las que registrar el corpus oral natural de los niños/as, a través de un soporte tecnológico y a las consecuencias que contrae la utilización de una u otra técnica.

Primero, hay que tener presente que el corpus seleccionado debe contar con las siguientes características: (1) el lingüista debe valerse de un formato electrónico con el que computarizar todas las tareas, y (2) es necesario garantizar la veracidad de los datos seleccionando los criterios en función de la finalidad que prevalezca el corpus y que permita la obtención de una muestra significativa que responda a los objetivos expuestos (Villayandre Llamazares 2008).

Por otra parte, como señalan Sinclair (1996) y Torruella y Llisterri (1999), hay diferentes formas de describir tipológicamente los *corpora* y, teniendo en cuenta los objetivos de esta investigación, nos hemos regido por la distinción que hace Listerri (1996) de los *corpora* de la lengua oral y los corpus orales. Además, se ha priorizado las metodologías no-intrusivas en la creación de *corpora* de lengua oral. Asimismo, se han descrito las distintos tipos metodológicos que nos podemos encontrar: observacional, experimental, cuantitativa, cualitativa o tipologías que aúnan diferentes conceptualizaciones.

Con el propósito de describir y priorizar una metodología para la adquisición de la lengua, se ha propuesto una observación contextualizada como metodología más apropiada para el registro del corpus oral natural de los niños/as, en la cual se observan uno o varios sujetos a través de la observación participante y condicionada por el lugar, las actividades, etc. La necesidad de adentrarse en el contexto para observar es corroborada Guasch (1997) cuando expone que, para comprender la cultura y forma de vivir de los grupos humanos, se requiere adentrarse en ellos para

recoger datos sobre su vida diaria. Gracias al avance tecnológico que se ha dado en la última década, la observación directa ha experimentado una notable mejoría en la actualidad, dado que las nuevas herramientas tecnológicas permiten grabar de una forma menos intervencionista y analizar los datos con múltiples técnicas descriptivas.

No obstante, a la hora de observar, se influye en la acción, ya que los interlocutores son conscientes de la presencia del investigador/a, de manera que, en vez de evitar este efecto, habrá que integrarlo (Duranti 2000). Por lo tanto, no debemos olvidar que, en la medida que hacemos una investigación con bebés y/o niños/as, es imprescindible un periodo previo a las grabaciones de acomodación donde el investigador/a consiga ser visto por los niños/as no como un elemento desconocido, sino un elemento más del contexto de grabación.

La propuesta de una metodología concreta para el análisis del corpus de oralidad en contextos iniciales de la lengua, hace también referencia al tratamiento que hay que realizar de los datos, prestando principalmente atención a cuatro elementos: los informantes, los criterios de grabación, la transcripción y su codificación y el análisis de los resultados; y, dentro de cada uno de ellos, a las diversas variables y características que condiciona cada situación. Por tanto, a través de trabajo se ha presentado una secuencia de distintas implicaciones metodológicas para la constitución de corpus orales en la investigación de la adquisición del lenguaje, y la necesidad de obtener estos datos atendiendo a las distintas variables que pueden condicionar el contexto de grabación; es decir, es necesario cuidar los parámetros que afectan a los informantes, a los interlocutores y a las actividades.

BIBLIOGRÁFÍA EMPLEADA

G. AGUADO (2010), *El desarrollo del lenguaje de 0 a 3 años. Bases para un diseño curricular en la Educación Infantil*, Madrid, CEPE.

M. ANGUERA (1990), «Metodología observacional», en *Metodología de la investigación en ciencias del comportamiento*, ed. J. Arnau *et al.*, Murcia, Compobell, pp. 125-238.

P. B. BALTES, H. W. REESE y J. R. NESSELROADE (1977), *Life-span developmental psychology*, Monterrey (CA), Brooks/Cole.

L. M. BEEBE y M. C. CUMMINGS (1996), «Natural speech act data versus written questionnaire data: How data collection method affects speech act performance», en *Speech Acts Across Cultures: Challenges to Communication in a Second Language*, ed. S. M. Gass y J. Neu, Berlin, Mouton de Gruyter, pp. 65-86.

M. BELINCHÓN CARMONA (1985), «Adquisición y evaluación de las funciones pragmáticas del lenguaje: un estudio evolutivo», *Estudios de Psicología*, 19-20, pp. 35-49.

E. BERICAT (1988), *La integración de los métodos cuantitativo y cualitativo en la investigación social. Significado y medida*, Barcelona, Ariel.

H. BOADA (1986), *El desarrollo de la comunicación en el niño*, Barcelona, Anthropos.

P. BOERSMA y D. WEENINK (2016), Praat: doing phonetics by computer, University of Amsterdam.

R. BROWN y G. YULE (1983), *Discourse Analysis*, Cambridge, University.

J. BRUNER (1984), *Acción, pensamiento y lenguaje*, Madrid, Alianza.

A. BRYMAN (1992), «Integrating quantitative and qualitative research: how is it done?», *Qualitative Research*, 6.1, pp. 97-113.

N. CHOMSKY (1957), *Syntactic Structures*, The Hague, Mouton.

R. A. CLEMENTE ESTEVAN (1998), «El papel del adulto en la adquisición del lenguaje. Reflexiones sobre los valores diferenciales entre interlocutores», en Desarrollo del lenguaje y cognición, ed. M. Peralbo Uzquiano *et al.*, Madrid, Pirámide, pp. 41-52.

J. COHEN (1960), «A coefficient of agreement for nominal scales», *Educational and Psychological Measurement*, 20.1, pp. 37-46.

E. DÍEZ-ITZA (1995), «Procesos fonológicos en la adquisición del español como lengua materna», en *Actas del XI Congreso Nacional de Lingüística Aplicada*, ed. J. M. Ruiz *et. al.*, Valladolid, Universidad, pp. 225-264.

A. DURANTI (2000), *Antropología lingüística*, Cambridge, University.

M. EISENSTEIN y J. W. BODMAN (1993), «Expressing gratitude in American English», en *Interlenguage Pragmatics*, ed. G. Kasper y S. Blum-Kulka, New York, Oxford University, pp. 64-78.

C. FELIX-BRASDEFER (2007), «Natural speech vs. elicited data. A comparison of natural and role play requests in Mexican Spanish», *Spanish in Context*, 4.2, pp. 159-185.

AnMal Electrónica 42 (2017)
ISSN 1697-4239

I. FERNÁNDEZ LÓPEZ (2009), *¿Cómo hablan los niños? El desarrollo del componente fonológico en el lenguaje infantil*, Madrid, Arco/Libros.

M. FERNÁNDEZ PÉREZ (1986), *La investigación lingüística desde la Filosofía de la ciencia (A propósito de la lingüística chomskiana)*, Santiago de Compostela, Universidad.

M. FERNÁNDEZ PÉREZ (2003), «Dinamismo construccional en el lenguaje infantil y teoría lingüística», *Estudios de Lingüística de la Universidad de Alicante*, 17, pp. 273-287.

J. R. FIRTH (1957), *Papers in Linguistics*, London, Oxford University.

H. GARFINKEL (1967), *Studies in Ethnomethodology*, Englewood Cliffs (NJ), Prentice-Hall.

E. GOFFMAN (1981), *Forms of Talk*, Philadelphia, University of Pennsylvania.

A. GOLATO (2005), *Compliments and Compliment responses: Grammatical Structure and Secuencial Organization*, Amsterdam-Philadelphia, John Benjamins.

O. GUASCH (1997), *Observación participante*, Madrid, CSIC.

M. A. K. HALLIDAY y R. HASAN (1985), *Language, context and text: a social semiotic perspective*, Victoria: Deakin University.

D. H. HYMES (1964a), *Language in Culture and Society: A Reader in Linguistics and Anthropology*, New York, Harper & Row.

D. H. HYMES (1964b), «Introduction: Toward Ethnographies of Communication», *American Anthropologist*, 66.6, pp. 1-34.

G. KASPER y S. BLUM-KULKA (1993), *Interlenguage Pragmatics*, New York, Oxford University.

F. R. KLUCKHOLM (1940), «The participant-observer technique in small communities», *American Journal of Sociology*, 46.3, pp. 331-343.

W. LABOV (1972), *Sociolinguistic patterns*, Philadelphia, University of Pennsylvania.

W. LABOV (1981), *Field methods of the project on linguistic change and variation*, Sociolinguistic Working Paper, 81, Southwest Educational Development Laboratory, Austin, Texas.

W. LABOV (2001), «The anatomy of style-shifting», en *Style and Sociolinguistic Variation*, ed. P. Eckert y J. R. Rickford, Cambridge, University, pp. 85-108.

H. LAUSBERG y H. SLOETJES (2009), «Coding gestural behavior with the NEUROGES-ELAN system», *Behavior Research Methods*, 41.3, pp. 841-849.

G. LEECH (1992), «Corpora and theories of linguistic performance», en *Directions in Linguistics: Proceedings of Nobel Symposium 82 (Stockholm, 4-8 August 1991)*, ed. J. Svartvik, Berlin, Mouton de Gruyter, pp. 105-122.

E. LIEVEN (1994), «Crosslinguistic and crosscultural aspects of language addressed to children», en *Input and Interaction in Language Acquisition*, ed. C. Gallaway y B. J. Richards, Cambridge, University, pp. 56-73.

J. LLISTERRI (1996), EAGLES. Preliminary recommendations on Spoken Texts. EAG-TCWG-STP/P.

J. LLISTERRI (2007), Speech analysis and transcription tools, Barcelona, Universitat Autònoma.

H. LÓPEZ MORALES (1994), *Métodos de investigación lingüística*, Salamanca, Colegio de España.

B. MACWHINNEY (2000), *The CHILDES Project: Tools for Analyzing Talk*, Hillsdale (NJ), Lawrence Erlbaum Associates.

G. M. MAN SHUM (1986), *Psicolingüística aplicada al estudio de adquisición del lenguaje en niños institucionalizados y niños no institucionalizados*, Madrid, Universidad Complutense.

J. L. MILLER y A. K. LOSSIA (2013), «Prelinguistic infants' communicative system: Role of caregiver social feedback», *First Language*, 33, pp. 524-544.

E. T. MOERK (1989), «Verbal interaction between children and their mothers during the preschool years», *Developmental Psychology*, 11, pp. 788-794.

M. MONFORT y A. JUÁREZ (1989), *Registro fonológico inducido*, Madrid, CEPE.

A. MORENO y J. URRESTI (2005), «El proyecto C-ORAL-ROM y su aplicación a la enseñanza del español», *Oralia*, 8, pp. 81-104.

P. MUNDY y A. GOMES (1988), «Individual differences in joint attention skill development in the second year», *Infant Behavior and Development*, 21, pp. 468-482.

E. NEWPORT, L. A. GLEITMAN y H. GLEITMAN (1977), «Mother I'd rather do it myseft: Some effects and non effects of maternal speech style», en *Talking to Children: Language Input and Language Acquisition*, ed. C. E. Snow y C. Ferguson, Cambridge, University, pp. 109-149.

E. OCHS (1979), «Transcription as theory», en *Developmental Pragmatics*, ed. E. Ochs y B. Schieffelin, New York, Academic Press, pp. 43-72.

AnMal Electrónica 42 (2017)
ISSN 1697-4239

T. K. PLUNKETT (1995), «Connectionist approaches to language acquisition», en *The handbook of child language*, ed. P. Flechter y B. MacWhinney, Oxford, Blackwell, pp. 84-102.

M. PUYUELO SANCLEMENTE (2000), «Aspectos generales de la evaluación del lenguaje», *Evaluación del lenguaje*, Barcelona, Masson, pp. 29-130.

M. M. RAMÍREZ CRUZ (s. f.), «El uso de *PHON* en la adquisición de L1», s.l., 11 pp.

M. M. RAMOS, A. CATENA y H. M. TRUJILLO (2004), *Manual de métodos y técnicas de investigación en ciencias del comportamiento*, Madrid, Biblioteca Nueva.

M. RECALDE y V. VÁZQUEZ ROZAS (2009), «Problemas metodológicos en la formación de corpus orales», en *A survey of corpus-based research*, ed. P. Cantos Gómez y A. Sánchez Pérez, Murcia, Asociación Española de Lingüística del Corpus, pp. 51-64.

S. C. REICHARDT y T. D. COOK (1982), «Más allá de los métodos cualitativos versus los cuantitativos», *Estudios de Psicología*, 11, pp. 40-55.

J. A. RONDAL (1979), *L'interaction adulte-enfant et la construction du language*, Bruxelles, Pierre Mardaga.

J. A. RONDAL (1980), «Father's and mother's speech in early language development», *Journal of Child Language*, 7, pp. 353-369.

Y. ROSE *et al*. (2007), «Phon 1.2: A computational basis for phonological database elaboration and model testing», en *Proceedings of the Workshop on Cognitive Aspects of Computational Language Acquisition, 45th Annual Meeting of the Association for Computational Linguistics*, ed. P. Buttery *et al*., Stroudsburg (PA), ACL, pp. 17-24.

C. F. ROWLAND, S. L. FLETCHER y D. FREUDENTHAL (2008), «How big is big enough? Assessing the reliability of data from naturalistic samples», en *Corpora in Language Acquisition Research. History, methods, perspectives*, ed. H. Behrens, Amsterdam (Philadelphia), John Benjamins, pp. 1-24.

J. I. RUIZ OLABUÉNAGA (1999), *Metodología de la investigación cualitativa*, Bilbao, Universidad de Deusto.

M. SERRA *et al*. (2013), *La adquisición del lenguaje*, Barcelona, Ariel.

M. SIGUÁN (1983), *Metodología para el estudio del lenguaje en la infancia*, Barcelona, Universidad.

J. SINCLAIR (1996), EAGLES. Preliminary recommendations on Spoken Texts. EAG-TCWG-STP/P.

C. E. SNOW (1986), «Conversations with children», en *Language Acquisition*, ed. M. Garman y P. Fletcher, Cambridge, University, pp. 363-375.

T. J. TAYLOR y S. SHANKER (2003), «Rethinking language acquisition: what the child learns», en *Rethinking Linguistics*, ed. H. J. Davis y T. J. Taylor, London, Routledge Curzon, pp. 151-170.

E. TOGNINI-BONELLI (2001), *Corpus Linguistics at Work*, Amsterdam, John Benjamins.

M. TOMASELLO (1988), «The role of joint attentional process in early language development», *Language Sciences*, 10, pp. 69-88.

M. TOMASELLO (2003), *Constructing a Language. A Usage-based Theory of Language Acquisition*, Cambridge, Harvard University.

M. TOMASELLO y J. FARRAR (1986), «Joint Attention and Early Language», *Child Development*, 57.6, pp. 1454-1463.

M. TOMASELLO y D. STAHL (2004), «Sampling children's spontaneus speech: how much is enough?», *Journal of Child Language*, 31, pp. 101-121.

J. TORRUELLA y J. LLISTERRI (1999), «Diseño de corpus textuales y orales», en *Filología e informática. Nuevas tecnologías en los estudios filológicos*, ed. J. M. Blecua, *et al.*, Barcelona, Universidad Autónoma-Milenio, pp. 45-77.

N. UGALDE BINDA y F. BALBASTRE BENAVENT (2013), «Investigación cuantitativa e investigación cualitativa: Buscando las ventajas de las diferentes metodologías de investigación», *Revista de Ciencias Económicas*, 31.2, pp. 179-187.

M. M. VIHMAN y R. MILLER (1988), «Words and bable at the threshold of language acquisition», en *The emergent lexicon: The child's development of a linguistic vocabulary*, ed. M. D. Smith y J. L. Locke, New York, Academic Press, pp. 189-205.

M. VILLAYANDRE LLAMAZARES (2008), «Lingüística con corpus (I)», *Estudios Humanísticos. Filología*, 30, pp. 329-349.

M. WARREN (2006), *Features of Naturalness in Conversation*, Amsterdam, John Benjamins.

B. WEST *et al.* (2007), *Linear Mixed Models: A Practical Guide Using Statistical Software*, New York, Chapman & Hall/CRC.