

I Jornadas Filológicas Internacionales-UR

**Corpus, bases de datos y diccionarios para
la investigación de la Lengua y la Literatura**

LIBRO DE RESÚMENES

30 septiembre - 2 octubre 2015

www.unirioja.es/fiur/

jornadas.fiur@gmail.com

Universidad de La Rioja - Cilengua

© Universidad de La Rioja, 2015

Título

Libro de resúmenes de las I Jornadas Filológicas Internacionales-UR: *Corpus, bases de datos y diccionarios para la investigación de la Lengua y la Literatura.*

Editoras

Sara Gómez Seibane

Rebeca Lázaro Niso

Zaida Vila Carneiro

Logroño, septiembre de 2015

ISBN 978-84-608-1721-5

PROBLEMAS METODOLÓGICOS EN LOS CORPUS DE LENGUAJE INFANTIL

Nuria Polo

Universidad Nacional de Educación a Distancia (UNED)

Se conoce como *lenguaje o habla infantil* al lenguaje de los menores de 5 años, es decir, al tipo de lenguaje que aparece durante el desarrollo de la lengua materna, con características fonológicas, morfológicas y sintácticas especiales, similares en todas las lenguas. En este trabajo concretamente se repasan los distintos tipos de corpus creados para el estudio de este tipo de habla, desde los primeros y más tradicionales como los diarios (Stern y Stern, 1907), hasta los actuales basados en técnicas digitalizadas (Rose *et al.*, 2006). En este sentido, la parte central del trabajo versa sobre la reflexión acerca de los problemas metodológicos que están presentes en la creación y uso de estos corpus.

Por tanto, el objetivo principal de este trabajo es presentar este tipo de corpus y reflexionar críticamente sobre las características que los corpus para el estudio lingüístico del lenguaje infantil deben tener, tales como la exhaustividad y la sistematicidad, generales a cualquier otro tipo de corpus, y específicas de estos como la accesibilidad o la viabilidad de los datos.

El estudio se centra específicamente sobre los problemas concretos de las intervenciones infantiles. Fundamentalmente dichos problemas se refieren, primero, a cómo realizar la recogida de datos, pues este grupo de análisis lingüístico responde a unas características especiales (MacWhinney y Snow, 1985); segundo, a la representatividad de los datos (Tomasello y Stahl, 2004); y tercero, a los problemas derivados del análisis de los datos, relacionados, por ejemplo, con las transcripciones y la fiabilidad de dichas transcripciones (Escudero *et al.*, 2012).

En resumen, se trata de reflexionar sobre cómo conseguir muestras de habla de niños muy pequeños, pues en ocasiones estos apenas tienen unos meses de vida, que sean válidas para un análisis científico y riguroso sobre el desarrollo del lenguaje. Finalmente, se repasan también algunos de los aspectos éticos que hay que tener en cuenta cuando se trabaja con una población vulnerable como lo es esta.

PALABRAS CLAVE: corpus habla infantil, metodología, fiabilidad, accesibilidad, representatividad.

BIBLIOGRAFÍA

- Escudero, D., Aguilar, L., Vanrell, M. M., y Prieto, P. (2012): "Analysis of intertranscriber consistency in the Cat_ToBI prosodic labelling system", *Speech Communication*, 54, 566-582.
- MacWhinney, B. y Snow, C. (1985): "The child language data exchange system", *Journal of Child Language* 12, 271-296.

- Rose, Y., B. MacWhinney, R. Byrne, G. Hedlund, K. Maddocks, P. O'Brien y T. Wareham (2006): "Introducing Phon: A Software Solution for the Study of Phonological Acquisition", Bamman, D., T. Magnitskaia y C. Zaller (eds.), *Proceedings of the 30th Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press, 489-500.
- Stern, C. y Stern, W. (1907): *Die Kindersprache: Eine psychologische und sprachtheoretische Untersuchung*. Leipzig Barth.
- Tomasello, M. y Stahl, D. (2004): "Sampling children's spontaneous speech: How much is enough?", *Journal of Child Language* 31, 101-121.