

Diseño de corpus orales de lenguaje infantil: análisis comparado de CHILDES y el Corpus de Habla Infantil Espontánea del Español desarrollado en el Laboratorio de Lingüística Informática de la Universidad Autónoma de Madrid

Miguel Pérez Milans
Antonio Moreno Sandoval
(Universidad Autónoma de Madrid)

Resumen

El desarrollo de corpus con muestras de habla espontánea ha experimentado un importante auge en la última década. Dentro de éstos, los que se han especializado en el lenguaje infantil son menos numerosos. El Childes Language Data Exchange System (CHILDES) es uno de estos casos, y actualmente se presenta como una referencia internacional. Sin embargo, los avances tecnológicos acontecidos en los últimos años (en especial, la grabación digital y los programas de transcripción), unido a la proliferación de corpus orales y escritos que están aportando nuevas metodologías desde el ámbito de la Lingüística, introducen un panorama innovador.

La experiencia del laboratorio de Lingüística Informática de la Universidad Autónoma de Madrid (LLI-UAM) muestra un claro ejemplo de la evolución de los corpus orales. Así, el desarrollo del Corpus Oral de Referencia de la Lengua Española Contemporánea (CORLEC), a principios de los noventa, y del Corpus Oral de las Lenguas Romances (C-ORAL-ROM), finalizado en el presente año con la financiación de la Unión Europea, permiten la aplicación de las últimas metodologías y herramientas tecnológicas y trasladarlas al diseño de corpus con muestras de lenguaje infantil.

De esta manera, en esta comunicación se arrancará con un breve repaso histórico al estudio de la adquisición del lenguaje y la recogida de datos, el cual nos ayudará a establecer las bases para entender el estado actual de la cuestión en lo que se refiere a los corpus infantiles. Posteriormente, nos detendremos en el análisis del corpus CHILDES así como en el que actualmente se desarrolla en el LLI-UAM: el Corpus de Habla Infantil Espontánea del Español (CHIEDE), dando pie a una comparación final entre ambos que tomará como referencia los medios y la metodología que el último puede introducir en este campo de aplicación.