

LOS CORPUS DE HABLA INFANTIL METODOLOGÍA Y ANÁLISIS

Marta Garrote Salazar



LOS CORPUS DE HABLA INFANTIL
Metodología y análisis

©2010 Marta Garrote Salazar

©2010 Ediciones de la Universidad Autónoma de Madrid.
Ciudad Universitaria de Cantoblanco. 28049 Madrid.
www.uam.es/publicaciones // servicio.publicaciones@uam.es

Reservados todos los derechos. Está prohibido, bajo las sanciones penales y el resarcimiento civil previsto en las leyes, reproducir, registrar o transmitir esta publicación, íntegra o parcialmente (salvo en este último caso, para su cita expresa en un texto diferente, mencionando su procedencia), por cualquier sistema de recuperación y por cualquier medio, sea mecánico, electrónico, magnético, electroóptico, por fotocopia o cualquier otro, sin la autorización previa por escrito de Ediciones de la Universidad Autónoma de Madrid.

Maquetación y diseño: M. A. Tejedor López

ISBN: 978-84-8344-187-9

Depósito Legal:

Realiza:

Impreso en España.

LOS CORPUS DE HABLA INFANTIL
Metodología y análisis

Marta Garrote Salazar



Servicio de Publicaciones de la Universidad Autónoma de Madrid

ÍNDICE GENERAL

Introducción	7
1. La Lingüística de Corpus	11
1.1. Antecedentes.....	11
1.2. La Lingüística de Corpus en la actualidad	13
1.2.1. Principales corpus actuales.....	16
1.3. Conceptos básicos	18
1.3.1. Definición de corpus.....	18
1.3.2. Criterios y dificultades de recopilación.....	21
1.3.3. Tipología de corpus	22
1.3.4. Ventajas.....	25
1.4. Perspectivas de futuro	26
2. La ontogénesis del lenguaje	31
2.1. Reseña histórica sobre el estudio del lenguaje infantil	31
2.2. Principales teorías	34
2.3. Evolución del lenguaje infantil	38
3. Los corpus y la ontogénesis del lenguaje.....	43
3.1. Estado de la cuestión.....	43
3.2. Los corpus infantiles en la actualidad	45
3.3. CHILDES.....	47
3.3.1. Presentación.....	47
3.3.2. Contenidos de la página.....	49
3.3.3. El corpus español.....	53
3.4. Análisis del método CHILDES	56
4. Metodología	65
4.1. Aspectos generales	65

4.2. Grabaciones y digitalización	73
4.3. Transcripción	74
4.4. Conversión a XML	79
4.5. Anotación morfosintáctica del corpus	81
4.5.1. Etiquetario	85
4.6. Transcripción fonológica	87
5. Análisis	89
5.1. Listado de frecuencias de las unidades	89
5.1.1. Datos extraídos del etiquetador morfosintáctico	90
5.1.2. Datos extraídos del transcriptor fonológico	100
5.2. Comparación de CHIEDE con C-ORAL-ROM	108
5.3. Estudio pragmático sobre los marcadores discursivos en el lenguaje infantil	110
5.4. Extracción de unidades distintivas del lenguaje infantil	114
5.5. Estudio fonológico del habla infantil	120
5.5.1. Extrapolación de los resultados	122
5.5.2. Frecuencia de uso y dificultades de adquisición	125
6. Conclusiones	129
7. Bibliografía	131
Apéndice A	141
Ejemplo de transcripción ortográfica de CHIEDE	141
Apéndice B	147
Frecuencias de formas en CHIEDE	147