

Dolores Azorín Fernández
M^a Antonia Martínez Linares
Juan Luis Jiménez Ruiz (Eds.)

**ESTUDIOS PARA UN CORPUS
DEL ESPAÑOL
HABLADO EN ALICANTE**

UNIVERSIDAD DE ALICANTE
INSTITUTO DE CULTURA JUAN GIL-ALBERT

ÍNDICE GENERAL

1.- EL CORPUS DEL ESPAÑOL HABLADO EN ALICANTE EN EL CONTEXTO DE LAS INVESTIGACIONES ACTUALES SOBRE EL LENGUAJE ORAL	11
1.1.- La lingüística de la oralidad	13
1.2.- Los <i>corpora</i> lingüísticos	14
1.3.- La problemática específica de los <i>corpora</i> orales	17
1.4.- Los <i>corpora</i> orales del español	19
1.5.- Referencias bibliográficas	20
2.- PROBLEMAS EPISTÉMICOS Y METODOLÓGICOS DE LA INVESTIGACIÓN	23
2.1.- Introducción: lingüística teórica <i>versus</i> lingüística aplicada. Hacia una concepción integral	25
2.2.- Bases glotológicas de la investigación lingüística. La aparición de la sociolingüística	27
2.3.- Demarcación disciplinaria y propuestas de caracterización	31
2.4.- Fundamentos críticos de sociolingüística urbana contemporánea	33
2.5.- Referencias bibliográficas	37
3.- CRITERIOS METODOLÓGICOS	41
3.1.- Introducción	43
3.2.- La variación diastrática en el Corpus del español hablado en Alicante	45
3.2.1.- El sexo	45
3.2.2.- La edad	47
3.2.3.- El nivel sociocultural	49
3.2.4.- La procedencia	53
3.2.5.- El barrio	55
3.2.6.- La situación de contacto de lenguas	57
3.2.7.- Referencias bibliográficas	59
3.3.- La variación lingüística situacional	63
3.3.1.- Los conceptos de estilo y registro	66

3.3.2.- Las variables situacionales: las relaciones entre los participantes	71
3.3.3.- La entrevista sociolingüística y la variación situacional	74
3.3.4.- Referencias bibliográficas	78
4.- ALICANTE Y SUS BARRIOS: ESTUDIO SOCIODEMOGRÁFICO DEL ESPACIO URBANO APLICADO AL DISEÑO DE UN CORPUS ORAL DEL ESPAÑOL	81
4.1.- Introducción	83
4.1.1.- Configuración urbana	83
4.1.2.- Aproximación sociodemográfica	85
4.1.3.- La cuestión del bilingüismo en Alicante	87
4.2.- Los barrios de la zona centro y periferia residencial de la ciudad ..	89
4.2.1.- Casco antiguo y Santa Cruz – Ayuntamiento	90
4.2.2.- Raval Roig-Virgen del Socorro	93
4.2.3.- Centro	95
4.2.2.- Ensanche-Diputación	98
4.2.5.- Mercado	100
4.2.6.- San Antón	102
4.2.7.- Campoamor	105
4.2.8.- Alipark	107
4.3.9.- El Palmeral-Urbanova	108
4.2.10.- Vistahermosa	110
4.2.11.- Playa de San Juan	112
4.2.12.- Albufereta	114
4.2.13.- Cabo de las Huertas	116
4.2.14.- Ciudad Jardín	118
4.3.- Barrios del primer extrarradio de la ciudad	121
4.3.1.- Carolinas Altas	122
4.3.2.- Carolinas Bajas	125
4.3.3.- El Pla del Bon Repós	128
4.3.4.- Altozano-Conde de Lumiares	131
4.3.5.- Ciudad de Asís	134
4.3.6.- La Florida	136
4.3.7.- Benalúa	139
4.3.8.- San Blas-Santo Domingo	141
4.3.9.- Polígono de San Blas	144
4.3.10.- Polígono de Babel	146
4.4.- Los barrios del conjunto urbano industrial	151
4.4.1.- San Gabriel	152
4.4.2.- Tómbola	154
4.4.3.- Rabasa	156

4.4.4.- Divina Pastora	158
4.4.5.- Los Ángeles	160
4.4.6.- San Agustín	163
4.4.7.- Lo Morant-San Nicolás de Bari	165
4.4.8.- Mil Viviendas	167
4.4.9.- Sidi Ifni-Nou Alacant	169
4.4.10.- Cuatrocientas Viviendas	171
4.4.11.- Juan XXIII	173
4.4.12.- Colonia Requena	175
4.4.13.- Garbinet	177
4.4.14.- Virgen del Remedio	179
4.5.- Referencias bibliográficas	181
5.- PLANIFICACIÓN DE LAS ESTRATEGIAS METODOLÓGICAS EN LA RECOGIDA DEL CORPUS DEL ESPAÑOL HABLADO EN ALICANTE	
	183
5.1.- La elección de los informantes	186
5.1.1.- Factores demográficos generales	186
5.1.2.- La muestra	189
5.1.3.- La muestra por barrios	191
5.2.- Método de recolección de los datos	192
5.2.1.- Técnicas específicas	192
5.2.2.- Nuestra propuesta	194
5.3.- Referencias bibliográficas	196
6.- EL USO DE CORPORA DIGITALIZADOS EN LA INVESTIGACIÓN LINGÜÍSTICA: ETIQUETADO Y ESTRUCTURA EN LENGUAJE SGML DEL CORPUS ORAL DE ESPAÑOL HABLADO DE ALICANTE (ALCORE)	
	199
6.1.- Introducción	201
6.2.- La cientificidad de la gramática: epistemología y ciencia de la lengua	202
6.2.1.- Evolución del concepto de cientificidad en la Lingüística actual	202
6.2.2.- El método de análisis lingüístico	202
6.2.2.1.- Crítica de la Lingüística logicista o referencial.	203
6.3.- Los <i>corpora</i> como herramienta metodológica de análisis lingüístico	204
6.3.1.- Concepto de corpus	204
6.3.2.- Características y clasificación del corpus	205
6.3.3.- Utilidades del corpus	205
6.3.4.- Evolución y desarrollo de los corpora informatizados	206

6.3.5.- Programas informáticos para el análisis lingüístico de los corpóra	207
6.4.-Etiquetado del corpus de español hablado de alicante (alcore) en lenguaje sgml	209
6.4.1.- Introducción a las etiquetas SGML	209
6.4.2.- Clasificación de las etiquetas	209
6.4.2.1.- Casos especiales	211
6.4.2.1.1.- Etiquetas de los hablantes	211
6.4.2.1.2.- Fáticos:	212
6.4.2.1.3.- Ruidos	212
6.4.2.1.4.- Uso de las mayúsculas	212
6.4.2.1.5.- Estilo directo	213
6.5.- Referencias bibliográficas	213
APÉNDICE	215
• FICHA DEL INFORMANTE	
• GUIÓN PARA LA ENTREVISTA	
• INFORME DEL ENTREVISTADOR	

1.- EL CORPUS DEL ESPAÑOL HABLADO EN ALICANTE EN EL CONTEXTO DE LAS INVESTIGACIONES ACTUALES SOBRE EL LENGUAJE ORAL

DOLORES AZORÍN FERNÁNDEZ

1.1.- La lingüística de la oralidad

En los últimos años, la dimensión oral de las lenguas ha llegado a convertirse en uno de los objetivos prioritarios de la investigación lingüística. Así parecen demostrarlo las numerosas contribuciones que, desde distintos ámbitos disciplinarios y/o teórico-metodológicos, coinciden en destacar la indistricutable relevancia que la oralidad tiene dentro del complejo funcionalmente polimórfico de la comunicación humana.

Sin salirnos del marco de la investigación en España, panorámicas como la que presenta L. Cortés (1994) o recopilaciones de estudios como las de este mismo autor (Cortés, 1995) o las de Briz, Gómez, Martínez y Grupo Val.Es.Co (1996) bastan para mostrar el gran impulso con que esta nueva corriente ha hecho acto de presencia en el panorama lingüístico actual.

Sin embargo, sería un tanto ingenuo por nuestra parte pensar que el interés por la oralidad es un fenómeno de nuestros días surgido, sin transición previa, como abrupta alternativa al modelo inmanentista representado por las dos grandes corrientes que han marcado el rumbo de la lingüística del siglo XX: el estructuralismo y la gramática generativa. Como es sabido, en ambos casos —y salvando las indudables distancias que los separan— no se toma en cuenta, como afirma Poch Olivé (1996: 193) «el habla (en el sentido saussureano del término) sino que el interés de los diversos autores se centra en la lengua [...] o dicho en otras palabras [...] en aquello que es «invariable» en las lenguas».

Este modelo asocial del lenguaje sería, no obstante, objeto de revisión por parte de otras corrientes que no tardaron en oponer una nueva concepción más próxima a los intercambios lingüísticos reales, a sus participantes y al contexto social en que se desenvolvía este proceso.

La alternativa al modelo asocial constituyó un movimiento heterogéneo, de base interdisciplinar, abierto por tanto a las aportaciones de otras esferas del conocimiento, cuya intersección con la lingüística iba a propiciar la apertura de nuevas vías para el estudio del lenguaje que el inmanentismo se había autovedado.

De este modo, como afirma Cortés Rodríguez (1996: 51), vino a producirse una «especie de consenso interdisciplinar» en el que:

Desde antropólogos, Gumperz y Hymes, a sociólogos, Goffman o Berstein, desde etnometodólogos, Sacks y Shegloff, a filósofos, Austen, Searle o Grice, pasando obviamente por lingüistas como Benveniste o Labov, vieron la necesidad de revisar una lingüística que se encontraba en un callejón sin salida, incapaz de una consideración dialéctica realmente explicativa de los hechos del lenguaje en su contexto social. (Cortés Rodríguez: *Ibid.*).

A pesar de la disparidad de enfoques y de objetivos, aproximaciones como las representadas por la sociolingüística, la etnografía de la comunicación, el análisis del discurso, la pragmática de los actos de habla, el análisis de la interacción conversacional etnometodológico, etc. constituyen los precedentes inmediatos de esa incipiente lingüística de la oralidad que, como hemos señalado anteriormente, ha comenzado a despuntar con fuerza en la década de los noventa, siendo uno de sus pilares más consolidados el análisis de la conversación. Dirección ésta en la que han confluído investigaciones de muy variada orientación, quizás por ser el marco conversacional el esquema más elemental y universalmente válido donde se muestra el lenguaje en acción (Cestero, 1994) a través de ciertos datos susceptibles de análisis sistemático (Stubbs, 1987: 31).

Los avances realizados, en más de treinta años de tentativas, no han conseguido aquilatar en el presente esa lingüística de la oralidad a la que antes nos referíamos, cuya consolidación ha de pasar, no sólo por el perfeccionamiento de las técnicas de análisis del discurso, sino también por el necesario acopio de datos que revelen el comportamiento lingüístico de los hablantes en situaciones reales de comunicación. Es decir, por la recopilación de grandes *corpora* orales de referencia capaces de proporcionar, de manera organizada y accesible, amplias muestras representativas de la variada tipología que presenta el texto oral.

1.2.- Los *corpora* lingüísticos

Desde hace algunos años, el acopio sistemático de materiales sobre el uso de las lenguas en sus más variadas manifestaciones, es tarea que viene ocupando a un buen número de investigadores procedentes tanto del sector público —universidades, instituciones científicas, etc.— como del privado —empresas editoriales, tecnologías de la comunicación, etc.—. Gracias a esta labor, hoy se puede disponer de diversos *corpora* de referencia, especialmente de la lengua escrita, cuyas variadas aplicaciones sobrepasan los límites de las distintas especialidades lingüísticas.

Debido a la multiplicidad de aplicaciones que estos documentos pueden llegar a tener, no es extraño que últimamente su cultivo haya llegado a

convertirse en objetivo prioritario de muchas líneas de investigación subvencionadas por organismos estatales o supraestatales. Un ejemplo en este último ámbito lo tenemos en el proyecto N.E.R.C. —Network for European Reference Corpora—, financiado por la Dirección General XIII de la Comisión de las Comunidades Europeas (Alvar-Villena, 1994: 6).

Fruto de sus variadas utilidades es la diversidad que, en sus realizaciones concretas, adquieren los *corpora*. De manera que podemos encontrar colecciones de documentos completos frente a colecciones de fragmentos; colecciones de textos orales, frente a las de escritos o a las que combinan ambos; recopilaciones de la lengua general, frente a las que se basan en algún tipo de sublengua; compilaciones básicamente monolingües frente a las bilingües o multilingües; etc., etc. El amplio espectro en que se puede concretar la producción de *corpora* hace difícil hallar una definición englobadora de su diversidad tipológica. Así, para F. Marcos Marín (1991: 1):

Un corpus de referencia es, esencialmente, una gran base de datos textual o, lo que es equivalente, un conjunto de bases de datos textuales unidas en un sistema de estructuración de datos, textos, referencias y utensilios informáticos para su tratamiento, bien en línea, en conexión directa, bien telemáticamente.

Definición en la que se priorizan los aspectos de gestión informática de la información contenida en los *corpora*, pero en la que no se abordan otros rasgos relacionados directamente con la naturaleza lingüística de los datos que se almacenan. Otros autores destacan, sin embargo, el carácter de representatividad que, respecto de un conjunto de hechos lingüísticos determinados, ha de poseer todo *corpus*. John M. Sinclair (1991: 171) parece referirse a esa característica cuando afirma: «A corpus is a collection of naturally-occurring language text, chosen to characterize a state or variety of a language». Más pormenorizada es la definición que ofrecen Alvar *et al.* (1994: 9-10), para quienes un *corpus* se define básicamente por ser «un conjunto homogéneo de documentos lingüísticos de cualquier tipo (orales, escritos, literarios, coloquiales, etc.) que se toman como modelo de un estado o nivel de lengua, al cual representan o se pretende que representen». Idéntico planteamiento encontramos en la definición del concepto de *corpus* que presentan Bach, Cabré, Saurí y Vivaldi (1997: 2) como punto de partida para la presentación del *Corpus de Llenguatges Especialitzats* del IULA (Institut Universitari de Lingüística Aplicada):

Un corpus lingüístic de dades textuais és una col.lecció organitzada de textos seleccionats sobre la base d'uns criteris determinats. Aquets criteris s'estableixen prèviament de manera acurada, per tal que els materials que formin part del corpus siguin representatius del conjunt d'actualitzacions, en un domini concret, d'una llengua —en el cas dels corpus monolingües— o de diverses llengües —en el cas dels corpus multilingües—.