

ROCÍO CARAVEDO

# Gramática española: Enseñanza e investigación

## I

### APUNTES METODOLÓGICOS

#### 6. Lingüística del Corpus

Cuestiones teórico-metodológicas aplicadas al español



EDICIONES UNIVERSIDAD DE SALAMANCA

# ÍNDICE

<i>Introducción</i> .....	13
---------------------------	----

## CAPÍTULO I MARCO HISTÓRICO-CONCEPTUAL

1. DEFINICIÓN DE LA LA LINGÜÍSTICA DEL CORPUS.....	17
1.1. <i>Dependencia empírica</i> .....	20
1.2. <i>Producción</i> .....	21
1.2.1. La fuente de información.....	21
1.2.2. Tipos de fuente de información.....	22
1.3. <i>Realismo</i> .....	25
1.3.1. Oralidad frente a escritura.....	27
1.3.1.1. Realismo en los textos literarios.....	32
1.3.2. La escritura de la oralidad.....	33
1.4. <i>Explicitud</i> .....	35
1.5. <i>Síntesis</i> .....	36
2. LA NOCIÓN DE CORPUS Y LA LINGÜÍSTICA MODERNA.....	37
2.1. <i>El corpus en el estructuralismo descriptivista</i> .....	38
2.2. <i>El corpus en la dialectología</i> .....	44
2.3. <i>El corpus en la sociolingüística</i> .....	46
2.4. <i>El corpus en la lingüística histórica</i> .....	49
2.5. <i>El corpus en la «lingüística de corpus»</i> .....	53

## CAPÍTULO II ASPECTOS METODOLÓGICOS

1. DATOS, CORPUS Y VALOR PREDICTIVO.....	61
1.1. <i>La homogeneidad</i> .....	63
1.2. <i>La representatividad</i> .....	66
1.2.1. Representatividad cualitativa.....	67
1.2.2. Representatividad cuantitativa.....	68
2. LO CUANTITATIVO EN EL ANÁLISIS.....	70
2.1. <i>Tipos de cuantificación</i> .....	72
2.1.1. Conexiones entre los tipos de cuantificación.....	76
2.2. <i>Lo cuantitativo como inherente</i> .....	76
2.3. <i>Cuantificación y frecuencia</i> .....	83
2.4. <i>Cuantificación y tipo de fenómeno</i> .....	86

CAPÍTULO III  
FORMACIÓN Y CARACTERÍSTICAS DE UN CORPUS PARTICULAR

1.	TIPOS DE CÓRPORA.....	93
1.1.	<i>Córpورا abiertos</i> .....	94
1.2.	<i>Córpورا cerrados</i> .....	98
2.	UN CORPUS DIALECTOLÓGICO.....	101
3.	UN CORPUS SOCIOLINGÜÍSTICO.....	104
4.	INTEGRACIÓN DEL CORPUS DIALECTOLÓGICO Y DEL SOCIOLINGÜÍSTICO.....	111
4.1.	<i>Marco general: una interpretación socio-espacial</i> .....	112
4.2.	<i>Metodología de la integración: conceptos operativos</i> .....	117

CAPÍTULO IV  
PRINCIPIOS DEL ANÁLISIS

1.	LA VISIÓN DICOTÓMICA DE LO INVARIABLE Y LO VARIABLE.....	127
1.1.	<i>Invariante y variantes</i> .....	130
1.2.	<i>Tipos de variantes: lingüísticas y extralingüísticas</i> .....	131
2.	VARIACIÓN FUNCIONAL Y VARIACIÓN NO FUNCIONAL.....	135
2.1.	<i>Modalidades de la variación</i> .....	137
2.1.1.	Invariación.....	138
2.1.2.	Variación no funcional.....	139
2.1.3.	Variación funcional.....	139
2.1.3.1.	Reducción.....	139
2.1.3.2.	Ampliación.....	144
2.1.3.3.	Conexión diacrónica de los procesos.....	148
2.2.	<i>Función y significado</i> .....	151

CAPÍTULO V  
ANÁLISIS CASUÍSTICO: VARIACIÓN FONOLÓGICA

1.	ESPACIOS DE VARIABILIDAD.....	155
2.	VARIACIÓN FUNCIONAL EN LAS SIBILANTES.....	161
2.1.	<i>La variación no funcional de las sibilantes en Hispanoamérica</i> .....	167
2.2.	<i>¿Distinción o seseo? Un caso conflictivo de variación funcional</i> .....	171
2.2.1.	Microcorpus.....	174
2.2.2.	El papel del léxico en la variación funcional fonológica.....	176
2.2.3.	Análisis cuantitativo.....	178
2.2.4.	Un tipo de diferencia situacional: la lectura.....	183
2.2.5.	Hermenéutica.....	184
3.	VARIACIÓN FUNCIONAL EN LAS PALATALES SONORAS.....	188
3.1.	<i>El caso del español andino</i> .....	190
3.1.1.	Microcorpus.....	191
3.1.2.	Análisis cuantitativo.....	194
3.1.3.	Hermenéutica.....	199
3.2.	<i>El caso del español amazónico</i> .....	205
3.2.1.	Descripción del fenómeno.....	208
3.2.2.	Microcorpus.....	209
3.2.3.	Análisis cuantitativo.....	212
3.2.4.	Hermenéutica.....	219

CAPÍTULO VI  
ANÁLISIS CASUÍSTICO: VARIACIÓN MORFOSINTÁCTICA

1. LA VARIACIÓN FUNCIONAL EN LA MORFOSINTAXIS.....	223
2. EL FACTOR ESPACIAL EN LA VARIACIÓN FUNCIONAL DE LOS PRONOMBRES OB- JETO EN ESPAÑOL.....	226
2.1. <i>¿Distinción o indistinción de objeto en el español andino?</i> .....	231
2.1.1. Microcorpus.....	234
2.1.2. Análisis cuantitativo.....	236
2.1.2.1. Evaluación de las unidades referenciales de con- tabilidad .....	246
2.1.3. Hermenéutica .....	247
2.1.4. Lo cualitativo frente a lo cuantitativo.....	251
2.2. <i>El caso del español amazónico</i> .....	253
2.2.1. Microcorpus.....	253
2.2.2. Análisis cuantitativo.....	255
2.2.3. Hermenéutica.....	264
BIBLIOGRAFÍA.....	271

## INTRODUCCIÓN

El título del presente libro podría entenderse como una alusión a la tendencia desarrollada en los últimos años en el ámbito de la lingüística anglosajona denominada «corpus linguistics», tendencia que se ha concentrado especialmente en la recopilación de material lingüístico, de preferencia escrito (aunque sin descartar el material oral) almacenado mediante procesamiento informático, y difundido en soporte magnético para análisis o estudios de naturaleza diversa. Pero es necesario despejar tal posible malentendido. Por «lingüística *del corpus*» (en vez de «lingüística *de corpus*», nombre que hemos reservado a la traducción de «corpus linguistics») aludimos a un campo más amplio y más antiguo en la lingüística, el cual si bien engloba distintas escuelas, supone un mismo tipo de actitud científica respecto de la relación directa entre teoría y datos.

Tal actitud implica el reconocimiento de la prioridad de la observación en la actividad analítica y sintética del razonamiento científico, y concomitantemente de la necesidad de un proceso de selección y de reconstrucción de la porción de realidad investigada en la forma de una nueva totalidad denominada corpus<sup>1</sup>. Obviamente, la llamada «corpus linguistics» supone esa misma actitud y —en consecuencia— forma parte de lo que se engloba aquí en la «lingüística del corpus». No son, pues, los medios materiales (informáticos o no), siempre cambiantes y provisorios, ni de difusión de la información, criterios determinantes en la identificación de una actividad reflexiva disciplinaria. Por lo tanto, incluiré dentro de la lingüística del corpus las perspectivas lingüísticas que —valiéndose o no de medios informáticos o de procesamientos cuantitativos— han hecho depender sus hipótesis, conjeturas, métodos y estrategias analíticas de la observación de la realidad a través de un corpus predefinido. Quedan agrupados, pues, estudios en sectores tan diferentes, como la lingüística descriptivista, la dialectología, la sociolingüística, la filología, la lingüística histórica, además de la lingüística de corpus en sentido restrictivo, representativos todos de esta modalidad de investigación lingüística.

Por otra parte —como se desprende del subtítulo— el presente libro no constituye en modo alguno un estudio pormenorizado que abarque todo lo que entiendo como lingüística del corpus. Tampoco se trata de un manual informativo ni de un estudio historiográfico que examine de forma exhaustiva las investigaciones realizadas en este campo<sup>2</sup>. Se trata, más bien, de un

1. En el presente estudio utilizo los términos *corpus* y *córpura* como préstamos técnicos integrados a la lengua española. Por consiguiente, pongo tilde en el segundo, y empleo ambos sin marcarlos con letra cursiva.

2. De allí que la bibliografía diste mucho de ser exhaustiva: es más bien resultado de una selección de solo los estudios que he conectado directamente con los principales puntos discutidos.

estudio específico que aspira —en un primer paso— a exponer de modo razonado un conjunto de problemas teórico-metodológicos que se plantean en toda lingüística del corpus, y, en un segundo, a enfrentar algunos de esos problemas en un corpus particular en relación con el análisis de ciertos fenómenos elegidos con un carácter ejemplificador. En ambos momentos se replantean o retoman cuestiones generales, reiteradas en el diálogo disciplinario a través del cual se construye el discurso científico, válidas para cualquier tipo de aproximación semejante a la representada aquí. La necesidad de aplicar las cuestiones generales a un corpus específico nace de la convicción de que una reflexión sobre el corpus en abstracto, sin conexión con problemas surgidos de la porción de realidad contenida en él, resultaría casi un contrasentido, pues la definición y alcances del propio concepto exigen un cierto nivel de particularización, si no se quiere caer en una determinación demasiado laxa y, a fin de cuentas, trivial.

No se trata, por cierto, de limitarse a un empirismo particularizador y detallista: todo lo contrario. Al detenerse a precisar la noción de corpus en un marco histórico-conceptual, y su aplicación en las principales líneas de investigación lingüística, la primera parte del libro (capítulos I a IV) constituye un intento de lograr tal objetivo. Hay en ella capítulos centrados en las cuestiones metodológicas implicadas en el carácter predictivo o de proyección de los análisis que se basan en corpóra, como la homogeneidad, la representatividad, la cuantificación. Este último aspecto, considerado fundamental en el manejo de una realidad empírica como la que presenta un corpus, será cuidadosamente examinado en relación con su situación teórica, y su capacidad descriptiva —cuando no, explicativa— ligada a la naturaleza de los fenómenos tratados. Se cierra esta parte con un capítulo centrado en la presentación esquemática y general de tipos de ordenamiento de la variación lingüística accesible a través del corpus, que vendrán posteriormente aplicados a fenómenos concretos en la última parte. Se reformulan aquí los principios tradicionales de la delimitación restrictiva de la variación para extenderlos de modo que incluyan —sin atentar contra la coherencia entre objetivos y métodos— las fluctuaciones de índole funcional.

La base conceptual propuesta no tiene un carácter especulativo, pues se concreta en la configuración de un corpus particular, cuyo punto de partida es una realidad compleja correspondiente al dominio del español, la cual exige ciertas decisiones teóricas y metodológicas. En primer término, el carácter de la realidad abordada en su concretización espacial y social requiere una integración de dos perspectivas: dialectológica y sociolingüística, integración capaz no solo de subsumir sino de recrear en una nueva síntesis interpretativa métodos tradicionales —que se han mostrado relativamente eficaces para cierto tipo de análisis— correspondientes a diferentes parcelas del pensamiento disciplinario. Por ello, en un capítulo independiente se desarrollan los principios generales que guiarán el análisis y que persiguen, a través de un replanteamiento de la línea funcionalista y sociolingüística, pero aprovechando mucho de ambas, dotar al análisis casuístico de un soporte de alcances generales. La prioridad de la observación no implica un inductivismo, puesto que toda observación debe guiarse de principios y propósitos presta-

blecidos, pero que lejos de ser fijos e inmutables sean permeables a ella. La idea subyacente a muchas críticas del trabajo en esta línea reside en la concepción ingenua de que los datos hablan por sí mismos, y que el investigador va sin teorías previas a extraer de ellos lo fundamental. Tal concepción, verdaderamente ingenua, difícilmente se defiende. Los datos hablan solo para quienes sean capaces de entenderlos y de interpretarlos. Y esto supone un conocimiento metalingüístico técnico muy profundo no solo de los fenómenos, sino de las teorías y metodologías que han tratado de explicarlos. En verdad, áreas como la sociolingüística, deben considerarse casi de índole metateórico y hasta metametodológico, en la medida en que constituyen desarrollos, extensiones o aplicaciones de conceptos tradicionales largamente trabajados en el ámbito disciplinario.

Como no se busca una exposición de planteamientos sin soporte en la realidad, la segunda parte del libro consiste en un análisis casuístico en dos capítulos finales (V y VI), análisis que concretan, replantean y aplican los presupuestos generales, las estrategias metodológicas y los principios analíticos discutidos en la primera parte. Seleccionados del mismo corpus, se trata de fenómenos internamente vinculados, centrales y problemáticos en la historia de la lengua española, extraídos del plano fonológico y del morfosintáctico: el caso de las diferencias sistemáticas entre indiferenciación y diferenciación (aplicado a las sibilantes y a las palatales sonoras) de una parte, y de la diferenciación e indiferenciación de los pronombres objeto, por otra. La elección no es naturalmente ingenua, en la medida en que persigue, de un lado, retomar planteamientos antiguos y someterlos a confrontación empírica, y, de otro, ensayar las propuestas interpretativas de la primera parte. No es de ningún modo accidental que tal elección haya recaído en fenómenos conflictivos que conmueven las definiciones tradicionales y constituyen un reto a los enunciados de una teoría, pues son los que precisamente permiten su reformulación, su desarrollo o su eventual descarte. La circunscripción de estos fenómenos al ámbito espacial cualquiera que este sea —referida en este caso circunstancial al español del Perú aunque situada en un contexto diatópico más amplio— es fundamental para lograr los objetivos de este estudio, pues constituye el cumplimiento de una exigencia mostrativa y comprobatoria de los enunciados de índole general de los primeros capítulos, imprescindible en el enfrentamiento de conceptos particularizantes como el de corpus. Una problemática sobre el corpus sin corpus constituiría la negación misma del concepto.

La circunstancia de poner como centro la variación de tipo funcional lleva de modo natural a otra selección metodológica, que en este caso —dada mi propia experiencia en la investigación— ha recaído, aunque no de modo acrítico, en la lingüística de la variación. No es necesario abundar en las razones individuales, justificadas en la propia historia personal, pero tampoco en las objetivas —sin duda más pertinentes— para aceptar la relevancia de esta línea en la esfera de la lingüística del corpus, concebida no como una técnica recolectora de datos o como mera heurística, sino como un tipo de hermenéutica. Aun mostrando mi acuerdo con los principios fundamentales de este enfoque, a saber, los centrados en la necesidad de des-

cubrir la organización y la pertinencia de los fenómenos variables como parte del funcionamiento de una lengua, aceptaré solo críticamente, de modo razonado y dependiendo de la naturaleza de los hechos observados, la imprescindibilidad de la medición cuantitativa. Así, cuento también con la existencia de hechos variables no cuantificables, pero no por ello desorganizados, que suponen una dimensión perceptiva de lo singular, más que de lo recurrente, y que pueden ser determinantes en la dirección que tomen los procesos evolutivos de una lengua. En el análisis he combinado la percepción y ordenamiento cuantitativos con la percepción cualitativa. Tampoco restringiré el estudio de la variación a las diferencias de estrato social —con lo cual coincido con la línea variacionista— sino que lo expandiré para abarcar todo fenómeno entendido como variable, cuyo comportamiento se ha mostrado (o se muestra) fluctuante e inestable en algunos puntos de la historia de una lengua. Vistos los hechos en el terreno de lo concreto, no cabrá tampoco la restricción terminológica que distingue variables lingüísticas puras (y en esto me distancio de la línea canónica), no conectadas con alguna instancia de la manifestación. Para sostener una posición semejante he reconsiderado algunos planteamientos sociolingüísticos, y los he reformulado y extendido, conforme a las necesidades de coherencia teórica y de comprensión empírica de los fenómenos conflictivos. La elección de tales fenómenos desempeña, pues, un papel crucial para someter a crítica las convenciones metodológicas y teóricas, y encaminar la investigación en nuevas direcciones que permitan corregir, reformular o —si acaso— abandonar antiguas premisas y conceptos, y quizás también falsas contradicciones provenientes de una confrontación insuficiente con la realidad. La casuística solo adquirirá sentido si se interpreta como una mera ejemplificación de posibilidades más ricas de análisis, que trasciendan los fenómenos aquí enfrentados y propongan otro tipo de problemas y de métodos de solución, más allá de los límites del presente estudio.