

LAS VARIEDADES DEL DISCURSO EN LA LINGÜÍSTICA DEL CORPUS:  
*WORKING WITH SPANISH CORPORA* DE GIOVANNI PARODI (COORD.)

Ángeles Romero Cambrón

Universidad de Castilla-La Mancha

Angeles Romero EN uclm es

Giovanni Parodi, ed.

*Working with Spanish corpora*

Londres, Continuum, 2007, 259 páginas

Colección Research in Corpus and Discourse

ISBN: 978-08264-9483-2

[http://www.continuumbooks.com](http://www.continuumbooks.com/Books/detail.aspx?ReturnURL=/Search/default.aspx&CountryID=1&ImprintID=2&BookID=129934)



*Working with Spanish corpora* constituye un volumen de conjunto que recoge nueve trabajos precedidos de una introducción. Abre el libro un prefacio del editor, la presentación de los colaboradores y la lista de siglas empleadas, y lo cierra las habituales referencias bibliográficas y un índice de materias. Todas las colaboraciones reunidas abordan materias de investigación próximas y comparten unos planteamientos metodológicos coincidentes o similares.

© 2009 Ángeles Romero Cambrón

*CÍRCULO de Lingüística Aplicada a la Comunicación (clac)* 37, 64-70

Universidad Complutense de Madrid. ISSN 1576-4737. <http://www.ucm.es/info/circulo>

La relativa novedad del tipo de enfoque adoptado en la obra reseñada acaso aconseje aclarar a qué clase de trabajos se están refiriendo los autores en el título. No se trata, evidentemente, de seleccionar estudios lingüísticos basados en un corpus, sea este de la índole que sea. La investigación en determinadas áreas de la Lingüística, por su propia naturaleza, necesariamente ha de apoyarse en un corpus: ese es el caso de la Dialectología y, particularmente, de la Lingüística Histórica. En sincronía, y pienso de manera especial en Gramática, los datos analizados pueden obtenerse de dos fuentes: el recurso a la competencia del hablante y la recopilación de un corpus. Esta última fue la opción elegida por Salvador Fernández Ramírez para su *Gramática. Corpora* teníamos antes, cuando era preciso disfrutar los textos mediante la lectura, y *corpora* tenemos ahora, en un nuevo formato digital. Ni que decir tiene, claro, que la naturaleza y las posibilidades del corpus digital han revolucionado en nuestros días el trabajo en Lingüística.

El título de nuestra compilación no apunta, obviamente a la utilización en general de un corpus, ni de un corpus digital, en Lingüística. El volumen se inscribe en un tipo de estudio cultivado ya desde hace tiempo y que cada día cobra más auge: la llamada Lingüística del Corpus, coincidente en parte o buena medida con la Lingüística Computacional. Esta nueva rama de nuestra ciencia despierta un interés creciente, como muestra la proliferación de reuniones científicas monográficas: la última de que tengo noticia es el I Congreso Internacional de Lingüística del Corpus (Murcia, del 7 al 9 de mayo de 2009). En esta disciplina se trabaja con *corpora* digitalizados en los que se procesa automáticamente los textos seleccionados y se someten a tratamiento matemático o estadístico los resultados. En ella es por tanto importante no solo dar cuenta de los textos o tipo de lengua que componen el corpus sino atender a cómo se realizan los procesos de transcripción, etiquetado y lematización en su constitución y procesamiento dado que estos son esenciales para los estudios que sobre los *corpora* se quieren realizar.

Los trabajos recopilados por Parodi se circunscriben precisamente al ámbito de la Lingüística del Corpus así entendida, tal como se explica en la Introducción. El editor presenta este tipo de trabajo como novedoso en nuestro idioma, tal vez exagerando un tanto. Lo que sí debe ser poco frecuente es aplicar en este tratamiento automático de los textos el método que en el volumen se llama *multidimensional*. Este método parte del

procesamiento automático del lenguaje, para en un segundo paso realizar un recuento estadístico y finalmente utilizar una serie de haces de rasgos lingüísticos coincidentes en un tipo de texto (65 en total, consignados en la pág. 19), rasgos relacionados con determinadas propiedades comunicativas, al objeto de caracterizar de esta forma diferentes registros. Los trabajos se llenan así de toda clase de esquemas y gráficos. El volumen atiende al español peninsular y al de otros países hispanoamericanos como Argentina, Chile, Estados Unidos y Venezuela. Estos son los estudios agrupados en *Working with Spanish corpora*:

En “Variation across registers in Spanish: Exploring the El Grial PUCV Corpus”, Parodi analiza este corpus chileno utilizando el método multidimensional. El autor considera demostrada la validez de este y subraya la posible aplicación de las observaciones obtenidas a la enseñanza de la lengua. Entre las conclusiones que destaca está la caracterización del discurso oral por su interactividad, contextualidad y las relaciones personales que se entablan, y la universalidad de los rasgos lingüísticos que singularizan a un registro dado. Mi pregunta es si buena parte de estas conclusiones no eran del todo esperables.

Douglas Biber y Nancy Tracy Ventura, en “Dimensions of register variation in Spanish”, utilizan otro corpus distinto, el de Davies ([www.corpusdelespanol.org](http://www.corpusdelespanol.org)). Hablan también de “dimensiones”, si bien las “dimensiones” que distinguen son diferentes a las de Parodi. Su conclusión más destacada es que hay una gran parte de registros compartidos entre todas las lenguas y otra parte característica de cada lengua o cultura en particular.

El objeto de la contribución de Guiomar E. Ciapuscio queda perfectamente concretado en el título: “Epistemic modality and academic orality: Pilot study for COTECA (Corpus Textual del Español Científico de la Argentina)”. La autora, en su análisis de los textos académicos orales, comprueba que en ellos se dan juicios subjetivos. También se potencia en los mismos la expresión de la factualidad, tanto con la variedad epistémica personal como con la no personal (esta última señalada como característica del discurso académico escrito), ambas realizadas mediante los *modality's qualifying schemata* —así lo plantea la autora. De los numerosos marcadores epistémicos se ofrece una minuciosa ejemplificación y clasificación. En suma, concluye Ciapuscio, el nivel gramatical puede ser reinterpretado en función de las dimensiones funcionales y situacionales del género considerado.

Omar Sabaj, en “Multi-register analysis of prepositional schemes in communication verbs in Spanish”, se ocupa del suplemento (o, por mejor decir, de verbos *dicendi* a los

que sigue una preposición) que puedan llevar los verbos que expresen exclusivamente comunicación (“decir”, “preguntar”, “contestar”...) o aquellos otros que transmiten información pero matizada por una actitud del hablante (“asegurar, enfatizar”...). Se someten a observación nueve registros. Tras una larga introducción en la que define este tipo de verbos, concluye Sabaj, de un lado, que el primer tipo de verbos, el que transmite solo información, es el más frecuente y, de otro, que la preposición que lleve el verbo está relacionada con la categoría cognitiva que esta introduce: por ejemplo, la preposición “a” se emplea con “modo”, “entidades” y “acontecimientos”. Es una lástima que el autor no se haya decidido a utilizar en su análisis el concepto de *suplemento* en vez de “verbo seguido de una preposición”, ignorando así las enseñanzas más elementales de la Gramática sobre el régimen verbal. Se añade a esta circunstancia la ausencia casi total de ejemplificación en la exposición: no se proporciona, como hubiera sido de desear, ninguna ilustración de verbos de comunicación que lleven suplemento con “a”, “en”, “de” o cualquier otra preposición, lo que hubiera servido para seguir mejor sus razonamientos y afianzar sus conclusiones.

Mercedes Sedano presenta un interesante estudio sobre la expresión del futuro en “Future tense expressions in several Spanish corpora”. En primer lugar trata el reparto dialectal en español del futuro morfológico, *cantaré*, y del perifrástico, *voy a cantar*: el primero es más frecuente en los textos peninsulares, más en autores instruidos o situaciones formales y en la lengua escrita; sin embargo, tienen un peso muy importante en la elección entre uno y otro futuro las preferencias individuales de cada autor. De todos estos factores, el decisivo es el empleo del registro escrito (futuro morfológico) frente al registro oral (futuro perifrástico). Además hay otros factores que determinan la distribución de los dos futuros: la distancia temporal (el morfológico es más frecuente cuanto mayor es la distancia); la persona gramatical (la forma perifrástica es más frecuente con la primera persona); y los marcadores de modalidad epistémica (la forma perifrástica es más frecuente cuando se expresa certeza). La investigación se sustenta en el aprovechamiento de dos *corpora* venezolanos.

Giovanni Parodi y Aída Gramajo son los autores de “Technical-professional discourse: Specialized and dissemination text types”. En este estudio se centran en el examen de este tipo de textos, en su variedad escrita, partiendo del corpus El Grial. Para ellos el discurso científico es un *continuum* que se descompone en una serie de variedades

definibles por una serie de rasgos concurrentes (*syndromes*). La variedad técnico-profesional aparece cuando el tema abordado es muy específico y se requiere un conocimiento previo de este por parte de emisor y receptor. Los autores presentan una clasificación de los textos de acuerdo con tres criterios: el situacional (entorno en el que es producido el texto, receptor y emisor); el funcional (entendiendo función a la manera de Jakobson); y textual, ya se trate de rasgos derivados de la estructura (argumentativa, descriptiva, expositiva...), o de que sean uno o varios los temas considerados. En el discurso técnico-profesional se aúnan una serie de propiedades: predominio de la función referencial, estructura expositiva, naturaleza monotemática... A partir de las características compartidas se pueden trazar rasgos diferenciadores entre las distintas clases de textos englobados en este discurso: los de naturaleza didáctica divergen de los puramente técnicos (manual, directiva, glosario, descripciones técnicas).

La contribución de Carmen López Ferrero está dedicada a la descripción del lenguaje de los alumnos de secundaria que ingresan en la universidad (lenguaje académico), tal como se recoge en el Corpus 92 (disponible en [www.iula.upf.edu/rec/corpus92/](http://www.iula.upf.edu/rec/corpus92/)): “Academic writing: *Exploring Corpus 92*”. La autora insiste en que su trabajo tiene una utilidad práctica, ya que se propone servir de guía para la enseñanza de estos alumnos. Por esta razón, ese corpus que refleja los usos lingüísticos de los estudiantes se compara con otro dedicado a recoger los de los especialistas. López Ferrero realiza un examen muy pormenorizado de las conjunciones paratácticas, en sus varias funciones (unión de oraciones y conectores que estructuran el texto). En este sentido, López Ferrero subraya que los estudiantes no distinguen niveles (formal e informal) a la hora de utilizar la parataxis, empleando esta en lugar de la hipotaxis en escritos del primer tipo. Tampoco los estudiantes son capaces de usar las conjunciones coordinantes que permiten articular cohesionadamente el texto académico. Dadas las carencias expresivas de los alumnos españoles, no coge por sorpresa que no usen correctamente las conjunciones coordinantes ni tampoco estructuren bien la información ni el texto como unidad expresiva. Si estudios como el presente, basados en el despojo de *corpora*, facilitan el desarrollo de técnicas didácticas eficaces para hacerles superar esas deficiencias, bienvenidos sean.

En “Using Latent Semantic Analysis in a Spanish research article corpus”, René Venegas se propone hacer un análisis del discurso científico, concebido este no como

un tipo de texto bien caracterizado sino como un *continuum* de tipos de textos. El estudio está dedicado en particular a los *abstracts* (resúmenes), *keywords* (palabras clave) y el contenido de los artículos científicos. El método que emplea el autor (*latent semantic analysis*), considerando la proximidad semántica entre palabras, mide estadísticamente la frecuencia en que estas aparecen, sin tener ningún otro tipo de información lingüística; un primer procesamiento del texto mediante este sistema se corrige mediante otro más detallado en el que sí se incluye esa clase de información. Los resultados del trabajo me resultan francamente poco iluminadores: son más fuertes las relaciones de (similitud semántica) entre el *abstract* y los contenidos, lo que implica que los primeros realizan satisfactoriamente la función que les es propia, la macrosemantizadora, esto es, los *abstracts* dan cuenta razonablemente bien del asunto tratado en el texto. Por el contrario, semejante objetivo no lo cumplen las *keywords*.

Cierra el volumen “Lexical bundles in speech and writing”, artículo escrito por Nicole Tracy Ventura, Viviana Cortés y Douglas Biber, y dedicado al estudio de lo que podríamos traducir literalmente como “haz” y mejor como “combinación de palabras” (pues no sirven los términos usuales de locución, frase hecha o “lenguaje repetido”). Se entiende por este término toda secuencia de palabras que se repita muy frecuentemente en un registro. Una combinación de palabras sería, para los autores, tanto “me parece que es”, “el hecho de que” como “a lo largo de”. A mi modo de ver, los autores se desentienden en su planteamiento, injustificadamente, de cuanto en este punto la Gramática enseña para formular nuevos conceptos que se avengan con el instrumento de análisis utilizado, en este caso el ordenador. El estudio establece tres funciones para las combinaciones de palabras: expresiones de punto de vista (“me parece que”), organizadores del discurso (“en cuanto a”) y expresiones referenciales (“eso es lo que”). De todas estas funciones se ofrece amplia ejemplificación.

A modo de conclusión general, sirvan estas reflexiones sobre la obra reseñada. Las primeras no guardan relación directa con ella. Personalmente, hoy por hoy, no es que muestre escepticismo con respecto al trabajo en Lingüística del Corpus; más bien, me muestro expectante, aguardando que estos primeros ensayos juveniles den lugar a obras más reposadas y granadas. Y ello porque soy consciente de que en un adecuado

desarrollo de la Lingüística del Corpus se dirime gran parte del futuro de la Lingüística en general. Concedo a este aspecto, por tanto, una importancia máxima. De ahí que me esfuere más en formular observaciones y críticas que en sumar elogios. Como último comentario general, me llama la atención el propio enfoque y nombre de la disciplina. Hasta ahora, las partes de la Lingüística se habían establecido de acuerdo con las partes del objeto estudiado (Gramática, Fonética...), de la perspectiva (Lingüística General, Histórica...): con Lingüística del Corpus es el “instrumento” empleado en el análisis el criterio distinguidor. Sin duda esto se debe a la relevancia de este instrumento.

Tal vez por esa importancia concedida al “instrumento”, todavía no dominado en toda su potencialidad, las introducciones teóricas y metodológicas de estos diez trabajos comentados sean tan extensas. Acaso porque en más de una ocasión el objeto de estudio se elige en función del “instrumento”, y no en función de su propio interés para el lingüista, las conclusiones de estos estudios me hayan resultado en más de una ocasión predecibles y obvias. Lo que, a mi modo de ver, me parece seguir un camino infructuoso es investigar de espaldas a todos aquellos conocimientos bien asentados en la Lingüística llamémosla “nuclear”. Es la dirección que toman algunos autores: pienso, por referirme solo a dos casos, en el trabajo de Sabaj (donde no se aprovecha el concepto de *suplemento*) y en de Tracy Ventura, Cortés y Biber, escrito sin tener en cuenta nociones como las de *locución*. Como gran virtud, habría que reconocerle al volumen editado por Parodi que es muy homogéneo con respecto a los temas tratados, y al modo de hacerlo: casi todos los trabajos incluidos se ocupan de variedades de discurso. Por eso creo que su lectura será de gran provecho para los interesados en la Lingüística del texto, la Lingüística del discurso e incluso para quienes se ocupen de Retórica. Asoan también en algunos estudios incluidos novedosas aportaciones a la Lingüística aplicada (aprendizaje de una primera o segunda lengua), verdaderamente prometedoras.

Recibido 6 enero 2009

Aceptado 9 enero 2009

Publicado 28 febrero 2009