

La definición del corpus en los estudios del discurso

ADRIANA BOLÍVAR

En los editoriales anteriores hemos hablado de la teoría y del método en la investigación. Es natural ahora tocar el tema del corpus puesto que estos tres elementos van estrechamente unidos. De la manera más sencilla, el corpus de la investigación se define como el conjunto de materiales lingüísticos o no, que conforman el objeto de estudio en una investigación o en una parte de ella. No obstante, las decisiones sobre cómo definir o constituir un corpus son más complejas puesto que detrás de todo corpus hay una disciplina, una posición teórica y un problema de investigación que puede ser de naturaleza muy variada. Son muchos los aspectos que podemos considerar al hablar de corpus y, aunque se puede llenar un libro con el tema, podemos destacar algunos puntos que interesan a los analistas del discurso tanto desde una perspectiva descriptiva como crítica.

El primer punto es que no cualquier colección de materiales es un corpus. Una cosa es coleccionar textos u objetos porque nos gustan o para practicar distintos tipos de análisis (gramatical, semántico, estilístico, retórico, etc.) y otra es recoger textos con alguna pregunta de investigación en mente para explicar un fenómeno discursivo o un problema social. Tampoco es un corpus cualquier colección de ejemplos inventados para argumentar o apoyar una propuesta de análisis pues toda propuesta o modelo de análisis está precedido por la observación o exploración de textos en contextos. Además, es un requisito fundamental que los materiales, lingüísticos o no, deben encontrarse en su estado natural y no haber sido manipulados previamente por el analista. Frecuentemente se procede por etapas. Se recogen materiales siguiendo ciertos supuestos básicos y luego el estudio se concentra en aspectos más precisos del problema cuando se analiza una parte o partes de ese gran corpus que suele denominarse el corpus “mayor”, el “archivo” o el “acervo”. Los materiales escogidos de estos corpus más grandes se denominan también corpus, sub-corpus o muestras, pero en todos los casos subyacen supuestos teóricos que, en gran medida, condicionan la interpretación de los datos obtenidos a partir de textos producidos o puestos a circular por actores sociales en contextos específicos.

El segundo punto es que hay una diferencia entre los corpus y el corpus de la investigación. En el caso de los corpus se trata de los materiales que emplean las distintas disciplinas, y sus formas de recogerlos, pero cuando hablamos de “el corpus” estamos refiriéndonos a los materiales que se usan en una investigación particular, de cualquier disciplina, y que debe justificarse en el contexto de la investigación. En el caso de la lingüística, se usa a menudo la definición ofrecida por Sinclair (1991: 171): “A corpus is a collection of naturally occurring language text, chosen to characterise a state or variety of a language”. De hecho, Charaudeau y Mainguenu (2002) proponen una definición muy similar, inspirados por Sinclair: “un corpus est une collection de données langagières choisies et organisées selon de critères linguistique explicites pour servir d'échantillon du langage” (p. 148). Es importante señalar que estos últimos autores hacen la diferencia entre los corpus lingüísticos, los de las otras ciencias y los del análisis del discurso. Todos sabemos que nuestro interés como estudiosos del discurso se concentra

en los corpus que nos permitan aproximarnos a los usos del lenguaje en la interacción social.

De la misma forma en que la invención de la grabadora significó grandes cambios para el estudio del lenguaje (particularmente la fonología), la tecnología moderna y los medios de comunicación han revolucionado la forma de hacer lingüística y de estudiar el discurso. Ahora es posible almacenar en las computadoras grandes cantidades de textos (orales, escritos, multimodales) y, también, emplear distintos tipos de programas computarizados para su análisis cuantitativo y cualitativo. Dichos corpus responden a presuposiciones sobre cómo funcionan las lenguas en la comunicación humana y, en consecuencia, surgen preguntas de investigación muy concretas. Tal es el caso de la lingüística de corpus (*corpus linguistics*) en la que se compete por construir corpus más y más grandes, de millones de palabras, basados en el supuesto de que mientras más grande sea el corpus más posibilidades existen de conocer y explicar cómo funciona una lengua y el discurso sobre la base de datos obtenidos en la interacción natural en una cultura determinada (asumiendo el término cultura en un sentido amplio que incluye naciones y pequeños grupos). Estos corpus gigantes, especialmente en lengua inglesa, pero también en español y portugués (nuestras lenguas oficiales en ALED) tienen el gran valor de servir como corpus de referencia para contrastar resultados obtenidos en estudios pequeños y evitar las generalizaciones apresuradas. Así como varían los corpus (diacrónicos, sincrónicos, sociolingüísticos, de lenguajes especializados, etc.) también varían los métodos; se puede comenzar con categorías léxico-gramaticales definidas a priori (*corpus-based research*) o pueden surgir del texto mismo (*corpus-driven linguistics*). Una no excluye la otra, y en la práctica a menudo se combinan. De manera similar, podríamos decir que en discurso es posible combinar los métodos que empiezan con categorías predeterminadas y los que se apoyan en las categorías que van surgiendo de los datos.

Visto lo anterior, nos podemos preguntar si los corpus para estudiar el lenguaje como sistema se pueden usar para estudiar el discurso. La respuesta es sí, definitivamente, siempre que contengan textos completos porque el objeto de estudio del discurso son los textos. Lo importante es tener claro que lingüistas y analistas del discurso pueden trabajar con datos distintos porque los corpus de lengua hablada o lengua escrita se leen de manera desarticulada como un bloque de datos que nos proporciona información sobre el lenguaje como un todo, sobre el sistema mismo desde distintas posiciones teóricas; mientras que los corpus discursivos se leen de acuerdo con los enfoques o teorías que adopta el analista sobre la forma en que los actores sociales se relacionan en distintos contextos y en diferentes culturas, sobre cómo construyen discursivamente el conocimiento, a los otros y a sí mismos. La búsqueda de datos variará de acuerdo con la preguntas de la investigación y según el alcance del contexto y, de la misma forma en que los corpus lingüísticos pueden permitirnos el estudio del léxico referencial y evaluativo, los corpus discursivos nos pueden ayudar a relacionar las prácticas sociales con las prácticas discursivas, a explicar cómo surgen los géneros discursivos en contextos micro y macro, cuáles son los temas preferidos, las formas de argumentar, las ideologías, las estrategias discursivas, etc.

En discurso necesitamos ampliar la definición de corpus para incluir otros lenguajes y otros métodos, por lo menos para diferenciar entre los objetivos de la lingüística y del análisis del discurso, y para comprender por qué en un campo multidisciplinar como el nuestro se abordan los datos de manera diferente. El “giro lingüístico” que han tomado las distintas disciplinas humanísticas y científicas al reconocer que la realidad se construye

socialmente, en y con el lenguaje, impone a todos un mayor conocimiento sobre el lenguaje situado en su contexto histórico y social. Por eso es importante que las investigaciones en la lingüística se conozcan y se tomen en cuenta, ya que un mayor conocimiento del sistema de la lengua nos permitirá hacer afirmaciones de significado mucho más confiables. Esto aplica tanto a la lingüística crítica como al análisis crítico del discurso. Una de las áreas que más apoyo necesita de la lingüística de corpus es el análisis crítico del discurso debido a los ataques que reciben los analistas por su supuesta falta de rigor metodológico ya que, a menudo, hacen explícita su posición ante el abuso de poder, la discriminación, las desigualdades, etc. El análisis de los problemas sociales necesita contar con evidencia lingüística porque, cuando las afirmaciones están apoyadas en datos confiables, es más fácil sustentar los argumentos que si las propuestas provienen de opiniones o de comentarios sobre textos. En estos casos la lingüística de corpus es una gran aliada para responder a las críticas (ver por ejemplo Bolívar, 2009; Orpin, 2005).

Un punto importante en relación con el corpus es diferenciar entre el proceso de construcción por el que pasa el investigador (durante el cual se hace muchas preguntas teóricas y busca el camino para responderlas) y otra el acto de reportar la investigación que se plasma en un artículo de investigación, el género académico clave para legitimar la investigación. Los procesos de construcción pueden variar según el tipo de estudio, pero en el reporte para los pares es obligatorio que se presenten y expliquen detalladamente los materiales empleados en la investigación, la justificación de por qué esos y no otros materiales, la forma en que se recogieron los datos, cómo se codificaron o catalogaron, los procedimientos seguidos para cumplir con los objetivos de la investigación. Uno de los mayores problemas en los artículos de investigación rechazados o aceptados con modificaciones mayores se debe a la falta de claridad en el momento de contextualizar el problema, definir el tipo de investigación y en poner por escrito la relación entre los conceptos centrales, las categorías de análisis y los procedimientos. La investigación puede ser impecable pero si no se explica bien pierde fuerza e impacto, y no llega a ser conocida.

Un problema adicional es pensar que los corpus son cerrados, y que al analizar un problema se ha descrito “la realidad”. En lingüística y en análisis del discurso los corpus solamente pueden ser abiertos porque en la dinámica de la comunicación humana y de los cambios sociales estamos en capacidad de captar momentos de un proceso mayor, pero nunca la totalidad. En todo este proceso, es muy relevante el punto de partida y la perspectiva ¿vemos el mundo a través del lenguaje? O ¿vemos el lenguaje a través del mundo? ¿es el lenguaje el objeto de estudio? O ¿es el uso del lenguaje en un contexto histórico, cultural, social y político determinado con propósitos muy particulares? Lo ideal es combinar ambas visiones. Ese es el mayor reto que tenemos como analistas.

En este número se incluyen cinco artículos y dos reseñas. El primer artículo, de Thays Adrián S., se titula *Divide y vencerás: la antonimia como estrategia ideológica de polarización en el discurso de Hugo Chávez Frías*. La autora se pregunta cómo se construye discursivamente la polarización en un momento histórico determinado en Venezuela y analiza un corpus de siete textos producidos por Hugo Rafael Chávez Frías entre 1999 y 2003 de acuerdo con un enfoque hermenéutico (Thompson, 1993). Dicho análisis le permite mostrar la construcción discursiva de tipos de polarización mediante el recurso de la antonimia, y afirmar que dicha estrategia lingüística discursiva fue empleada por el expresidente venezolano para legitimar al propio grupo y deslegitimar al contrario, marcar límites, valorar o denostar y destacar desigualdades.

El segundo artículo, titulado *Una mirada a la discriminación en México desde el discurso de la prensa escrita*, pertenece a Adriana Aguayo, quien se pregunta cómo la prensa escrita mexicana aborda el problema de la discriminación en momentos en que se reportan incidentes de este tipo. Se presentan los resultados de un análisis cuantitativo aplicado a un corpus que contiene 373 textos publicados en tres diarios de circulación nacional en México. El estudio ofrece información relacionada con la cantidad de textos que publicaron los diarios sobre el problema de la discriminación, cuáles géneros discursivos periodísticos mencionaban algún tipo de discriminación, cuáles eran los temas más frecuentes y cuáles eran los grupos más mencionados. Los resultados muestran que los diarios pocas veces señalan responsables o toman postura ante los hechos que narran.

En el tercer trabajo, de Mercedes Duarte, titulado *Redes de metáforas cognitivas en el discurso político: “el socialismo del siglo XXI” de Hugo Chávez*, la autora se pregunta cómo se construyen las redes de metáforas asociadas con la noción de ‘socialismo’ en el discurso del ex presidente Hugo Chávez. Para ello recurre primero al análisis cognitivo propuesto por Lakoff y Johnson ([1980] 2007) y luego al análisis lingüístico según Cameron (1999). El corpus lo conforman cuatro emisiones del programa *Aló Presidente* del año 2005 relacionados con los temas dominantes: educación, salud, cultura y economía. El estudio muestra cómo las redes metafóricas construidas en el discurso del expresidente se asocian y combinan como familias semánticas para cumplir con los fines ideológicos del “socialismo del siglo XXI”.

El cuarto artículo, *Paradigmas en disputa, presupuestos compartidos*, corresponde a Gabriel Dvoskin, quien analiza las representaciones sobre la sexualidad en un debate parlamentario en la Argentina a raíz del *Proyecto de Educación Sexual Integral*. El autor sigue la propuesta de la Lingüística Crítica (Hodge y Kress, 1979, Raiter, 1999a) con el propósito de averiguar de qué forma el Proyecto está sustentado en un discurso que cuestiona el eje de referencias impuesto por el discurso dominante que marca la sexualidad con valores represivos. En el estudio se analizan las intervenciones de dos diputados, uno en favor de la sanción del Proyecto y otra en contra, y se discute en qué medida ambos paradigmas presuponen valoraciones diferentes sobre el signo ideológico “sexualidad”.

El quinto artículo, titulado *La ciudad y el campo como referentes de identidad en adolescentes mapuches de Temuco y Santiago*, escrito en colaboración por Chery Pérez Arce, Sandra Garrido Osses, Gabriel Llanquinao Llanquinao, Héctor Turra Chico, y María Eugenia Merino Dickinson, se concentra en el rol que juegan los espacios ‘ciudad’ y ‘campo’ en la construcción de identidad. El estudio se aborda desde el análisis del discurso y el enfoque de identidades espaciales de Proshansky (1978). Los datos se recogieron mediante entrevistas semiestructuradas y los resultados indicaron que las construcciones discursivas se enmarcan principalmente en las modalidades de identificación por *pertenencia* y *apego* (Dixon y Durrheim (2000)). Se muestra que la noción ciudad evoca un contexto en el cual se relevan aspectos sociales, culturales y lingüísticos, mientras la noción de campo se asocia a modos de apego vinculados a tipos de afectividad en prácticas de socialización mapuche.

Las dos reseñas fueron enviadas por colegas de Argentina. El primer libro (Bolívar, A. y Beke, R. (Comp.) (2011). *Lectura y escritura para la investigación*) tiene que ver con la formación de formadores en el campo del discurso académico y fue reseñado por Laura Ferrari. El segundo (Londoño Zapata, O. I. (2012). *Los estudios del discurso: miradas latinoamericanas 1*) es la segunda reseña que publicamos sobre un libro de este autor. En el libro anterior entrevistó a analistas europeos y en este caso se trata de entrevistas a un

grupo de analistas del discurso en América Latina. La reseña estuvo a cargo de Juan Eduardo Bonnín.

En general, los artículos y las reseñas dan una muestra de la intensa actividad que se lleva a cabo en nuestra Asociación y de la forma en que se abordan los problemas lingüísticos y sociales.

REFERENCIAS

- BOLÍVAR, A. (2009). “Democracia” y “revolución” en Venezuela: un análisis crítico del discurso político desde la lingüística de corpus. *Oralia*, 12, 27-54.
- CHARAUDEAU, P., y MAINGUENAU, D. (2002). *Dictionnaire d'analyse du discours*. Paris: Éditions du Seuil.
- ORPIN, D. (2005). Corpus linguistics and critical discourse analysis. Examining the ideology of sleaze, *International Journal of Corpus Linguistics* 10(1): 37-61.
- SINCLAIR, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

AB