

En torno a *PRESEEA*: Notas de investigación y de sociología de la ciencia

*Francisco Moreno Fernández**

Universidad de Alcalá – Instituto Cervantes en la Universidad de Harvard

La llamada “lingüística de corpus” tiene a sus espaldas una larga historia cuyos límites, en el espacio y en el propio tiempo, no se concretan con facilidad. Puede afirmarse, sin embargo, que la concepción moderna de esta manifestación de la lingüística comenzó su más nítido desarrollo a partir de los años sesenta (Francis y Kucera 1964) y que su consolidación dentro de la especialidad se conoció durante la última década del pasado siglo. Si bien la utilidad de los corpus lingüísticos ha acabado reconociéndose prácticamente desde cualquier especialidad en el estudio de la lengua, en la década de los sesenta la lexicografía anglosajona lideró el avance metodológico y técnico de los corpus (Sinclair 1991, Nunberg 1992). Esta misma área contribuyó también a la creación de corpus de lengua española, que resultaron fundamentales para la elaboración de los diccionarios de la serie VOX (Alvar y Ezquerro y Villena Ponsoda 1994) y de la editorial SGEL (Sánchez *et al.* 1995), en los años ochenta y noventa, o de los diccionarios académicos ya en el siglo XXI (Moreno Fernández 2004, López Morales 2005, Rojo 2016).

* Para correspondencia, dirigirse a: Francisco Moreno Fernández (francisco.moreno@uah.es), Departamento de Filología, Comunicación y Documentación, Facultad de Filosofía y Letras - Edif. Caracciolos Universidad de Alcalá. 28801 - Alcalá de Henares (Madrid) - España.

En relación con esta época, corría el año 1999 cuando el gramático belga Josse De Kock, en una de sus múltiples visitas a Madrid, quiso que nos viéramos para hablar de una nueva entrega de su *Gramática española*, aunque, sobre todo, quería que conversáramos de lingüística. De Kock siempre ha sido un hombre con preguntas en la cartera, cuestiones que se plantea a sí mismo y que propone a colegas y amigos, como juego unas veces y otras, como reto. En esa ocasión, la pregunta era la siguiente: ¿debe decirse “lingüística de corpus” o “lingüística del corpus”? Evidentemente, no se trataba de un simple juego preposicional, sino de una cuestión de fondo que a la postre se ha discutido muy ampliamente: la investigación realizada a partir de corpus (Parodi 2008): ¿debe entenderse como una disciplina lingüística o como un método? De Kock optó por una interpretación diferente: entenderla como instrumento de investigación, por eso tituló un libro editado por él en 1999 recurriendo a otra preposición: *Lingüística con corpus* (De Kock 2001). Comparto la interpretación.

Ya desde los años sesenta la lingüística del español ha contado con una experiencia previa en la elaboración de corpus, una experiencia originada desde la dialectología o, si se quiere, desde la dialectología social: el proyecto para el estudio de la norma culta de la lengua española, conocido como PILEI (Lope Blanch 1986). En realidad, en aquellos años y con este fin, no se hacía referencia explícita a una lingüística de corpus propiamente dicha, pero el caso es que el PILEI entraba tan de lleno en este terreno de la lingüística que, cuando en 1998 se reunieron en un mismo formato todos sus materiales, recibieron el nombre de “Macrocorpus de la norma lingüística culta de las principales ciudades del mundo hispánico” (Samper, Hernández y Troya 1998). Se inauguraba de este modo, en un sentido contemporáneo, una línea de trabajo que ha sumado notables contribuciones para el estudio de la lengua española hablada. Los primeros trabajos del PILEI revelaban una interpretación limitada de la realidad sociolingüística del español, pero enseguida dieron lugar a colecciones enriquecidas desde tal perspectiva. Así hay que entender el corpus del español de Caracas (Venezuela), dirigido por Paola Bentivoglio en los años setenta (Bentivoglio y Sedano 1993), y el corpus del español de Monterrey (México), dirigido por Lidia Rodríguez Alfano en los años ochenta (Rodríguez Alfano 2005).

La experiencia hispánica en el ámbito de la creación de corpus de lengua hablada evolucionó en paralelo a las investigaciones de otras lenguas, especialmente del inglés y del francés canadiense. A comienzos de los años setenta, alrededor de las figuras de Gillian Sankoff, David Sankoff y Henrietta Cedergren, se elaboró el corpus del francés de Montreal, que fue ampliamente explotado para investigaciones de todo tipo, incluidas las dedicadas a la entonación (Thibault y Vincent 1990). Más adelante, ya

en los años ochenta, quedó concluido el corpus bilingüe de Ottawa-Hull, en el que participó activamente Shana Poplack y con el que trabajó Sali Tagliamonte en sus años de formación (Poplack 1989). Fue Tagliamonte precisamente la que me explicó la configuración de este corpus en una estancia de investigación que pude disfrutar en Canadá durante el curso 1990-1991. El corpus Ottawa-Hull pasaba entonces por ser el más grande del mundo (3,5 millones de palabras procedentes de 120 hablantes), hasta el punto de que la mayoría de los análisis se realizaban a partir de submuestras y no del corpus en su totalidad. No se tenían en cuenta, sin embargo, las dimensiones del corpus de Monterrey, que sobrepasaba en 500 entrevistas al canadiense y que probablemente siga siendo el que aporte materiales de un mayor número de hablantes.

Todo lo mencionado anteriormente, además de las aportaciones que la sociolingüística anglosajona había ido haciendo desde 1964, constituyó el caldo de cultivo idóneo para el inicio de un proyecto que ha marcado la sociolingüística hispánica desde su presentación pública en 1996: el proyecto para el estudio sociolingüístico del español de España y de América (PRESEEA). Las características de este proyecto, así como las aportaciones que de él se han derivado y los resultados que ha aportado a la lingüística hispánica en su conjunto han sido explicados en diversos medios y diferentes ocasiones, por lo que no es necesario reiterar aquí lo sabido o lo que, por otro medio, puede conocerse fácilmente (Moreno Fernández 2005, 2006). No obstante, siempre es oportuno insistir en lo que no se ha entendido de un modo adecuado, aclarar las cuestiones menos evidentes o matizar mejor aquello que no se ha explicado suficientemente.

En relación con el origen y el desarrollo de PRESEEA, es importante mencionar que el proyecto, por decisión colectiva, se diseñó para incluir materiales lingüísticos producidos entre los años 1996 y 2010. De hecho, el procesamiento de esos materiales concretos es una de las tareas específicas que aún está en desarrollo, incluida la edición electrónica a través de la página *preseea.linguas.net*. Al margen de la edición propiamente dicha, PRESEEA se impuso como objetivo, a partir del 2010, el análisis coordinado de diversos aspectos de la lengua española, en sus niveles fonético (p.e. /d/ intervocálica, /s/ final de sílaba), gramatical (p.e. *haber*, dequeísmo) y discursivo (p.e. atenuación). Los trabajos reunidos en este volumen del *Boletín de Filología* son muestra de ello. Sin embargo, el inicio formal de esta segunda fase “analítica” no detuvo el interés externo por el proyecto, de modo que desde entonces han continuado llegando peticiones de incorporación a la red, con tal insistencia que hemos decidido seguir creciendo y sirviendo a quien quiera obtener ventaja de las particulares condiciones de PRESEEA.

En cuanto a la aportación del proyecto a la sociolingüística general y del español, merecen comentarse, en mi opinión, desde una perspectiva contemporánea, cuatro aspectos, referidos a cuestiones de metodología lingüística y de sociología de la ciencia (Bloor 2003), dado que PRESEEA es un proyecto colectivo que responde tanto a una dinámica científica como a una dinámica social. Esos aspectos afectan a las muestras, las entrevistas, la distribución geográfica y la arquitectura social del proyecto.

1. *La muestra.* Es bien sabido que PRESEEA nació en pleno apogeo de una sociolingüística urbana y variacionista; una sociolingüística de “regla variable” que ponía especial interés en la correlación de unas variables lingüísticas, habitualmente fonomorfológicas, con unos factores sociales. Tales factores venían fuertemente condicionados por una visión de la sociedad en que los conceptos de “clase” y “estrato” eran poco menos que hegemónicos y que privilegiaba las posiciones sobre las funciones sociales, así como las estructuras sociales sobre la dinámica de los mercados y los intercambios comunicativos (Tagliamonte 2016). La evidencia de estas realidades, cautivas de una época, no restó significación, sin embargo, a otros hechos manifiestos, como los siguientes: a) los individuos que conforman una muestra poseen valores intrínsecos que pueden analizarse desde nociones diferentes a las de clase y estructura; b) cualquier uso lingüístico puede ser interpretado desde cualquiera de los factores cualitativos que concurren en un hablante individual, incluidos los relativos a mercados, funciones, redes u otros tipos de agrupaciones. Estos hechos se fundamentan en la realidad de que cualquier grupo de hablantes puede ser descompuesto en sus individualidades y, por lo tanto, es susceptible de reconfigurarse de acuerdo a criterios sociológicos diferentes (Giddens 1995, Simmel 2015).
2. *La entrevista.* Las muestras de lengua hablada que conforman PRESEEA se reúnen mediante entrevistas semidirigidas. Se trata de la técnica de mayor recorrido en la sociolingüística internacional y la que más resultados ha aportado en relación con distintos objetos de estudio y propuestas epistemológicas. Las ventajas e inconvenientes que supone el empleo de las entrevistas semidirigidas han sido analizados con profusión en distintas ocasiones (Fernández Sanmartín 2008; Pitkowski 2010). Es evidente que una entrevista no produce el mismo tipo de lengua que una conversación y que el carácter semidirigido de una interacción no puede ofrecer el mismo tipo de materiales que una interacción ordenada en forma de diálogo público,

que una alocución política o que una conversación en un contexto profesional. Pero nada coloca a la entrevista semidirigida por debajo del nivel de validez o interés de otras modalidades discursivas. Si la entrevista semidirigida no es capaz de ofrecer muestras equivalentes a las de otras manifestaciones lingüísticas, lo mismo podría decirse de cada una de las demás alternativas. En unos casos y en otros, son múltiples las posibilidades de análisis sociolingüístico, en ocasiones absolutamente divergentes. De hecho, los objetivos de muchos de los trabajos realizados sobre el corpus PRESEEA en los primeros años del siglo XXI no coinciden con los objetivos de los trabajos realizados durante la segunda década. Sea como sea, los materiales de un corpus, obtenidos mediante una modalidad determinada no son responsables del tipo de análisis que los investigadores practiquen sobre ellos.

3. *Distribución geográfica*. Hace un tiempo, el colega mexicano Raúl Ávila me apuntaba, no sin un punto de razón, que el proyecto PRESEEA debería incluir en su denominación, no la referencia a “España y América”, sino a “Europa y América”. Bien pensado, probablemente en su origen, el nombre debería haber aludido solamente a la “lengua española”, prescindiendo de la mención de territorios, nacionales o transnacionales. Ocurre, sin embargo, que el español tiene una implantación internacional tan dispersa y diversa que exige precisar la geografía y los contextos que interesan en cada caso. Así las cosas, la referencia al español de “Europa” se haría extraña siendo España el único país hispanohablante del continente, además de Andorra, y teniendo España significativos enclaves en la geografía africana. Además, el emparejamiento de “América” con otro país singular (pensamos en Filipinas o en los Estados Unidos), por diversas razones, no resulta excepcional. Con todo, una posible sigla PRESOLE, referida al “estudio sociolingüístico de la lengua española” no le habría caído mal al proyecto, aunque tampoco en este caso habrían faltado las críticas, tal vez por hablar de “lengua española” y no de “lengua castellana”.

Por otro lado, la geografía representada en PRESEEA no responde a un diseño espacial de laboratorio. La adscripción a PRESEEA por parte de los equipos que analizan la sociolingüística de cada ciudad es absolutamente voluntaria y de iniciativa propia. Por lo tanto, la red final de comunidades investigadas es fruto de circunstancias universitarias muy diversas, no de una planificación externa, excepto en el rastreo de nuevos equipos que pudieran interesarse por el

estudio sociolingüístico de áreas que no podían descuidarse. Ese rastreo se hizo en Argentina, donde la sociolingüística urbana no goza de tantos practicantes como otras disciplinas lingüísticas; así se hizo en Cuba, cuyas relaciones internacionales durante una época fueron complicadas; y así se sigue haciendo con Bolivia, donde la sociolingüística se interesa más por las condiciones sociopolíticas de las lenguas originarias que por la variación del español o castellano urbano. Así pues, las iniciativas universitarias particulares son las que han llevado a que, dentro de PRESEEA, en unos países exista interés por el estudio sociolingüístico de varios núcleos urbanos; en otros, apenas por uno solo, y en otros, por ninguno

4. *Arquitectura social*. Desde sus inicios, PRESEEA ha contado con una “coordinación general”, complementada desde 2013 con una coordinación técnica, y ambas tienen su base en la Universidad de Alcalá (España). Esta coordinación consiste básicamente en el cumplimiento de cuatro tareas: registrar, editar, formar e informar. Se registra a los equipos que desean adscribirse al proyecto, garantizando el cumplimiento de unas características operativas e institucionales básicas; se forma a los equipos en los aspectos técnicos relativos a cualquier fase del proceso investigador; se editan en formato electrónico los materiales aportados por los equipos para sumarse al corpus común; y se informa a la comunidad PRESEEA de todos aquellos asuntos que son de interés parcial o general.

Ahora bien, la existencia de una coordinación general no supone que los equipos coordinados se organicen en una estructura jerárquica o piramidal. Los equipos de PRESEEA configuran una auténtica red social de investigación, en la que la coordinación se sitúa como un elemento auxiliar de servicio y en la que cualquier equipo puede tener protagonismo para determinadas tareas de interés general. La sociología interna de esta red revela la existencia de equipos –o de coordinadores de equipos– que a menudo asumen responsabilidades generales; también de otros que nunca las han asumido, por las circunstancias que sea, aunque no se ha desvirtuado por ello la conciencia de red social. Las ventajas de trabajar desde una estructura en forma de red se aprecian en muchos ámbitos, como el de la financiación o el de la presencia internacional, dentro y fuera del espacio hispanohablante. PRESEEA, como red, ha recibido financiación de la Fundación Comillas de España y de la Real Academia Española, a la que se ceden los materiales comunes para

su inclusión en el Corpus del Español del siglo XXI (CORPES) (Rojo 2016). Del mismo modo, la red ha servido de base para la solicitud de financiación por países, bien para uno solo de los equipos, bien para el apoyo económico de equipos subcoordinados. Finalmente, esta labor en red está siendo un factor de enorme utilidad para los procesos de internacionalización de las universidades implicadas, tan importante en la actual dinámica mundial de investigación.

Para concluir estas notas, quisiera recordar que el proyecto acaba de cumplir 20 años y que, afortunadamente para todos sus integrantes, sigue activo y productivo. No es nada fácil semejante continuidad en una dinámica universitaria e investigadora, que prima el resultado inmediato y que favorece las acciones individuales sobre las colectivas. Pero PRESEEA encuentra su fuerza en su configuración reticular, porque hoy día quien participa en una red seria y estable tiene un tesoro. A la vez, el proyecto común ofrece unas condiciones óptimas para servir de motor de iniciativas cuyas últimas consecuencias y resultados probablemente serán tan ricos como imprevisibles.

REFERENCIAS BIBLIOGRÁFICAS

- ALVAR EZQUERRA, MANUEL Y JUAN VILLENA PONSODA. 1994. *Estudios para un corpus del español*. Málaga: Universidad de Málaga.
- BENTIVOGGIO, PAOLA Y MERCEDES SEDANO. 1993. *Investigación sociolingüística: sus métodos aplicados a una experiencia venezolana*. Caracas: Universidad Central de Venezuela.
- BLOOR, DAVID. 2003. *Conocimiento e imaginario social*. Barcelona: Gedisa.
- FERNÁNDEZ SANMARTÍN, ALBA, MARCOS GARCÍA SALIDO, MONTSERRAT RECALDE, VICTORIA VÁZQUEZ ROZAS. 2008. Reflexiones metodológicas sobre la técnica de la entrevista semidirigida. *Actas del XV Congreso Internacional de ALFAL*. Montevideo, Uruguay. Consulta en línea: http://gramatica.usc.es/~vvazq/pdf_public/entrevista_semidir_alfal_2008.pdf [22-08/2016]
- FRANCIS, NELSON Y HENRY KUCERA. 1964 *A Standard Corpus of Present-Day Edited American English, for use with Digital Computers*. Providence, Rhode Island: Brown University.
- GIDDENS ANTHONY. 1995. *La constitución de la sociedad. Bases para la teoría de la estructuración*. Buenos Aires: Amorrortu.
- KOCK, JOSSE DE. 2001. *Lingüística con corpus. Catorce aplicaciones sobre el español*. Salamanca: Universidad de Salamanca.
- LOPE BLANCH, JUAN M. 1986. *El estudio del español hablado culto. Historia de un proyecto*. México: UNAM.
- LÓPEZ MORALES, HUMBERTO. 2005. Los corpora lingüísticos hispanoamericanos: situación actual. En L. Rodríguez Alfano (coord.). *Investigación sociolingüística. El habla de Monterrey*, pp. 31-47. México: Trillas.
- MORENO FERNÁNDEZ, FRANCISCO. 2004. Corpora of Spoken Spanish Language - The Representativeness Issue. En Y. Kawaguchi, S. Zaima, T. Takagaki, K. Shibano y M.

- Usami (eds.). *The First International Conference on Linguistic Informatics. State of the Art and the Future*, pp. 49-76. Tokio: Tokyo University of Foreign Studies.
- _____. 2005. Corpus para el estudio del español en su variación geográfica y social. El corpus "PRESEEA". *Oralia* 8: 123-139.
- _____. 2006. Información básica sobre el Proyecto para el Estudio Sociolingüístico del Español de España y de América – PRESEEA (1996-2010). *Revista Española de Lingüística* XXXVI: 385-392.
- NUNBERG, GEOFFREY. 1992. Usage in the American Heritage Dictionary: The Place of Criticism. *The American Heritage Dictionary of the English Language*. En A. Soukhanov (ed.). 3rd ed. New York: American Heritage and Houghton, pp. xxvi-xxx.
- PARODI, GIOVANNI. 2008. Lingüística de corpus: una introducción al ámbito. *Revista de Lingüística Teórica y Aplicada* 46: 93-119
- PTKOWSKI, ELENA. 2010. *La entrevista sociolingüística y su valor como método de investigación*. Montréal: Université de Montréal.
- POPLACK, SHANA. 1989. The Care and Handling of a Mega-corpus. En Ralph W. Fasold y Deborah Schiffrin (eds.). *Language Change and Variation*, pp. 411-451. Amsterdam & Philadelphia: John Benjamins.
- RODRÍGUEZ ALFANO, LIDIA (coord.). 2005. *Investigación sociolingüística. El habla de Monterrey*. México: Trillas.
- ROJO, GUILLERMO. 2016. *Citius, maius, melius: del CREA al CORPES XXI*. En J. Kabatek (ed.). *Lingüística de corpus y lingüística histórica iberorrománica*. Berlín: de Gruyter. En prensa.
- SAMPER PADILLA, JOSÉ ANTONIO, CLARA HERNÁNDEZ CABRERA Y MAGNOLIA TROYA. 1998. *Macrocorpus de la norma lingüística culta de las principales ciudades del mundo hispánico*. Las Palmas de Gran Canaria: Universidad de Las Palmas.
- SÁNCHEZ, AQUILINO, RAMÓN SARMIENTO, PASCUAL SANTOS, JOSÉ SIMÓN. 1995. *Cumbre. Corpus lingüístico del español contemporáneo. Fundamentos y aplicaciones*. Madrid: SGEL.
- SIMMEL, GEORG. 2015. *Sociología: Estudios Sobre las Formas de Socialización*. México: Fondo de Cultura Económica.
- SINCLAIR, JOHN. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- TAGLIAMONTE, SALI. 2016. *Making Waves. The Story of Variationist Sociolinguistics*. Oxford: Wiley Blackwell.
- THIBAUT, PIERRETTE Y DIANE VINCENT. 1990. *Un corpus de Français parlé. Montréal 84: Historique, Méthodes et Perspectives de Recherche*. Québec: Université Laval.