

**RESEÑA AL III COLOQUIO INTERNACIONAL SOBRE CORPUS DIACRÓNICOS  
EN LENGUAS IBERORROMÁNICAS (CODILI 3). UNIVERSITÄT ZÜRICH, 23-25  
DE JUNIO DE 2014**

ELENA DIEZ DEL CORRAL ARETA  
*Université de Lausanne (Switzerland)*  
elena.diezelcorralareta@unil.ch

Del 23 al 25 de junio de 2014 se celebró en la Universidad de Zúrich el III Congreso Internacional sobre Corpus Diacrónicos en Lenguas Iberorrománicas, también conocido como *CODILI*, acrónimo que se acuñó a partir de este tercer congreso. Las dos ediciones previas tuvieron lugar en Mallorca en el año 2007 (organizada por Andrés Enrique-Arias y su equipo) y en Barcelona en 2011 (por Joan Torruella y sus colaboradores)<sup>1</sup>. La puesta en marcha de esta tercera edición, a la que asistieron muchos investigadores presentes en las dos primeras y otros más que se sumaron, corrió a cargo de la Cátedra de lingüística iberorrománica de la Universidad de Zúrich, formada por el profesor Johannes Kabatek, sus asistentes Carlota de Benito Moreno y Albert Wall, y por la secretaria del departamento, María Luisa Gago.

El programa<sup>2</sup> constó de tres conferencias plenarias pronunciadas por Inés Fernández-Ordoñez, Andrés Enrique-Arias y Guillermo Rojo; de dos mesas redondas dedicadas a la presentación de proyectos de corpus históricos; de treinta comunicaciones basadas en la teorización de corpus, en el análisis de frecuencias, en el estudio fonético, morfológico, sintáctico y discursivo de algunos corpus y en la presentación de otros nuevos, y, por último, de nueve pósteres, presentados en una sesión conjunta.

El lunes 23, tras unas palabras de bienvenida y una introducción al congreso por Kabatek, De Benito y Wall, se inauguró el coloquio con la primera conferencia plenaria a cargo de la profesora y académica Inés Fernández Ordoñez: «El corpus Sonoro Oral del Español Rural (*COSER*) y la investigación diacrónica». El *COSER*<sup>3</sup> es un corpus oral sincrónico, pero con un interés diacrónico evidente, ya que recoge en sus grabaciones una variación dialectal de tal calibre que permite documentar muchas formas de difícil registro, consideradas incluso como perdidas o de antaño. Además, frente a los atlas y

---

<sup>1</sup> Estos dos investigadores, además, a partir de las reflexiones sobre la lingüística histórica basada en corpus y contando con la colaboración de otros grupos de investigación como la Red Internacional *CHARTA* (coordinada por Pedro Sánchez-Prieto), crearon la revista *Scriptum digital* ([www.scriptumdigital.org](http://www.scriptumdigital.org)), dedicada a difundir trabajos sobre aspectos teóricos y metodológicos relacionados con la edición digital y los corpus diacrónicos informatizados en las lenguas iberorrománicas.

<sup>2</sup> Tanto el programa como el libro de resúmenes del coloquio pueden consultarse todavía en su página web, cuya navegación está disponible en cuatro idiomas (español, portugués, alemán e inglés): [www.codili.ch](http://www.codili.ch). La presencia de lenguas iberorromances en el congreso fue, además, muy variada. Hubo presentaciones en gallego, catalán, portugués y español, con sus respectivas variedades.

<sup>3</sup> En su página web, se pueden encontrar varias muestras del corpus: [www.uam.es/coser](http://www.uam.es/coser).

monografías dialectales de finales de los 90, el *COSEER* permite estudiar la variación gramatical del español, como mostró Fernández Ordoñez en su conferencia con algunos ejemplos de *leísmo*, *laísmo* y *loísmo* o de la estructura constituida por artículo + posesivo + sustantivo.

Las primeras grabaciones de este corpus se realizaron en 1990, siguiendo el ejemplo de Diego Catalán Menéndez-Pidal y de Flora Klein-Andreu, con el fin de analizar el uso de los pronombres átonos. A partir de entonces, Inés Fernández Ordoñez ha promovido y desarrollado este gran proyecto de corpus, junto a muchos otros investigadores y estudiantes que han participado en la recopilación, edición y análisis de los datos. A día de hoy, el *COSEER* cuenta con grabaciones de 1069 localidades y con 1360 entrevistas semidirigidas, a las que siguen sumándose otras paulatinamente. Entre los objetivos próximos del proyecto se encuentran el aumento de las localidades (como, por ejemplo, las zonas bilingües no contempladas hasta ahora y las Islas Canarias), la transcripción de todas las grabaciones, la lematización y etiquetación de las transcripciones y el desarrollo de un buscador para poder utilizar fácilmente el corpus.

Tras la conferencia inaugural de Inés Fernández-Ordoñez, comenzó la sesión de comunicaciones con dos secciones paralelas: una dedicada a estudios sintácticos basados en corpus y otra a la presentación de corpus (en la que se habló de un corpus de documentos medievales de Miranda de Ebro, del corpus textual *DITECA* y de un corpus de cartas de particulares del siglo XVI). A las 13h tuvo lugar la comida, tras la cual comenzaron de nuevo dos secciones paralelas de comunicaciones, dedicadas esta vez a estudios morfológicos y a la presentación de nuevos corpus (del Gondomar: *Corpus dixital de textos galegos da idade moderna* y del Fondo Balbuena de la Universidad de Lausana). Seguidamente, después de la pausa de café, hubo un periodo en el que se habló de corpus (del *CORHIJE*: Corpus Histórico Judeoespañol) y otro sobre análisis fonéticos, para reunirnos todos a continuación en la primera mesa redonda del congreso sobre el estado actual de algunos proyectos de corpus históricos. Esta mesa versó concretamente sobre el *Corpus del català antic*, del *CIPM* de Portugal y del *TMILG* de Galicia, presentados por Joan Torruella, Maria Francisca Xavier y Eduardo Moscoso Mato, respectivamente.

El martes 24, la jornada comenzó con la segunda conferencia plenaria del congreso «La variación en la expresión de la posesión en el castellano medieval: nuevos métodos y perspectivas», pronunciada por el profesor Andrés Enrique-Arias. Este investigador, director del conocido proyecto *Biblia Medieval*<sup>4</sup>, habló sobre algunos nuevos métodos en la lingüística de corpus y sobre las nuevas perspectivas que se abren, entre otros, para el estudio del castellano medieval. En primer lugar, se detuvo en comentar las limitaciones de los corpus convencionales frente a los corpus paralelos. Los corpus convencionales son limitados en su función heurística y en la comparabilidad de datos, mientras que los corpus paralelos, al contener textos equivalentes, ofrecen una dimensión variacional con una perspectiva abierta que nos permite observar el desarrollo histórico de diferentes fenómenos lingüísticos.

Al contrastar estos dos tipos de corpus, Andrés Enrique-Arias se detuvo en señalar una de las paradojas en la composición de corpus como es la heterogeneidad de las fuentes que debe conformarlos (incluyendo diversos registros, dialectos, etc.) y la

---

<sup>4</sup> Este corpus está compuesto de cinco millones de palabras y diecisiete mil imágenes digitales, y permite la descarga de todos los textos que lo conforman. Puede accederse directamente a él en la página web: [www.bibliamedieval.es](http://www.bibliamedieval.es).

homogeneidad que deben presentar los diferentes cortes sincrónicos para poder ser comparables. De ahí resulta, como destacó el conferenciante, la necesidad de componer corpus paralelos como el de la *Biblia Medieval*. Para demostrar la utilidad del mismo, este investigador ofreció un análisis de la posesión en español medieval, explicando cómo se pueden obtener todas las variantes y las ocurrencias de los posesivos.

Después de la conferencia matutina se inició la sesión de comunicaciones con dos secciones paralelas: la primera dedicada a frecuencias y a la teorización de corpus, y la segunda a sintaxis y a tradiciones discursivas. Tras la comida, la primera sesión continuó con la teorización de corpus y con la presentación de nuevos corpus (en esta ocasión, el *COREECOM: Corpus Electrónico del Español Colonial Mexicano*), y la segunda con trabajos basados en tradiciones discursivas.

Después de la pausa-café tuvo lugar la segunda mesa redonda del coloquio sobre el estado actual de algunos proyectos de corpus históricos, en concreto: el *CHEM/México*, presentado por Carlos Méndez-Cruz; el *Tycho Brahe/Brasil*, expuesto por Charlotte Galves, y el *PHPB/Brasil*, del que hablaron José da Silva Simões y Alessandra Castilho. Como colofón a esta segunda jornada, se celebró la cena oficial del congreso en el restaurante Uniturm, un lugar con unas vistas excepcionales de la ciudad de Zúrich y con una cocina excelente.

El último día del congreso, el miércoles 25 de junio, se inició con la conferencia plenaria del profesor y académico Guillermo Rojo, titulada «*Citius, maius, melius*. Los corpus de referencia en la actualidad». Esta versó sobre los grandes corpus textuales del español como son el *CREA*, el *CORDE* y, más recientemente, el *CORPES*, en cuya construcción y desarrollo ha estado participando Rojo desde hace años.

El conferenciante señaló la relevancia de los corpus de referencia, ubicándolos entre los corpus pequeños muy especializados que contienen codificación detallada, y los corpus extraídos directamente de la web que no ofrecen diseño alguno ni herramientas de recuperación selectiva de datos, aunque nos permiten obtener muchas veces fenómenos de baja frecuencia. En relación con ello, nos ofreció un recorrido retrospectivo sobre los cambios que se han producido en la investigación lingüística, en la que destaca la lingüística de corpus como una auténtica revolución instrumental que nos ha proporcionado nuevas herramientas de análisis. Asimismo, Guillermo Rojo se detuvo en comentar algunos conceptos importantes como el de la 'representatividad' de los corpus (que, al fin y al cabo, es un concepto estadístico), y en mencionar los nuevos desafíos de los corpus de referencia que deben proporcionar más velocidad, más memoria, una mayor difusión, etc.

En cuanto a los corpus de referencia del español, además de señalar sus ventajas (como, por ejemplo, el hecho de haber sido diseñados con la finalidad de recuperar la información de manera selectiva), Guillermo Rojo habló de los progresos metodológicos del *CORPES*, frente al *CREA* y al *CORDE*, como son el diseño, la distribución de los textos (que es más representativa: un 70% de América y un 30% de España), la inclusión de textos de zonas poco contempladas como Filipinas y Guinea Ecuatorial, la anotación morfosintáctica y su lematización, la proporción de una tipología textual más compleja y extensa, y la aplicación para una consulta flexible que permite una recuperación selectiva de la información, coapariciones por lemas y la posibilidad de seleccionar varios parámetros al mismo tiempo. Además, anunció las innovaciones que están programadas

para el verano del 2014 en el *CORPES*, y para el *CREA* y una parte del *CORDE* durante el año 2015.

Tras la última conferencia se iniciaron las sesiones paralelas de comunicaciones sobre: 1) teorización de corpus y 2) estudios sintácticos, para pasar a una sesión de pósteres, acompañada, además, de algunas deliciosas tapas para picar que amenizaron el intercambio de ideas, mientras se escuchaba, se leía y se observaban los diferentes paneles.

Para terminar el congreso, tuvo lugar un periodo de sesiones paralelas sobre: 1) estudios sintácticos y 2) sobre corpus (en concreto, uno del español peruano para el estudio histórico del léxico del siglo XIX). Después de una breve pausa de café, se cerró el congreso con una discusión final, centrada en la siguiente pregunta: ¿hacia dónde va la lingüística histórica basada en corpus? En ella se recapitulaban los puntos más importantes de las tres jornadas de trabajo y se pusieron de relieve muchas de las tareas que todavía nos quedan por acometer.

Por último, es de recibo apuntar que, además de todos los actos académicos de calidad reseñados, el congreso se caracterizó por el cuidado de cada detalle: desde la recepción calurosa o *warming up* del domingo por la noche en el *biergarten* del *Fischstube*; los abundantes ágapes y cafés (los aperitivos de inauguración y clausura que se ofrecieron fueron tan copiosos que bien se hubieran podido llamar cenas); la mesa de profesores donde comimos todos los días con unas vistas singulares de la ciudad; la cena oficial del congreso ya mencionada, hasta incluso la transmisión de partidos del mundial, una vez acabadas las jornadas de trabajo, para que los más futboleros pudieran disfrutar de ellos (el lunes, por ejemplo, tuvimos la posibilidad de ver los partidos de España/Chile y México/Brasil en el *Romanisches Seminar*, mientras se servía el aperitivo de inauguración).

Finalmente, como información tanto para aquellos que asistieron como para los que les hubiera gustado asistir, se acordó que la próxima cita, es decir, la cuarta edición de este congreso, tuviera lugar en Alcalá de Henares en el año 2016, ya que es necesario dejar al menos dos años de diferencia para poder observar los cambios en el desarrollo de los corpus existentes y la creación de otros nuevos. Después de esta exitosa edición, estamos seguros de que no faltarán participantes entusiastas que sigan compartiendo los avances y el desarrollo de corpus ya conocidos, el nacimiento de otros y los diversos análisis lingüísticos que seguiremos haciendo, gracias, en gran parte, a estos Corpus Diacrónicos en Lenguas Iberorrománicas.