

[DOSSIER DE PRENSA]

**Primeras jornadas de trabajo del
*Corpus del español del siglo XXI (CORPES XXI)***

Entre los días 14 y 18 de julio de 2008 se celebrarán las Primeras Jornadas de trabajo del *Corpus del español del siglo XXI (CORPES XXI)*, proyecto acometido por la Real Academia Española y la Asociación de Academias de la Lengua Española en 2007, con el patrocinio del Santander, la colaboración de la Fundación Comillas y la participación de las Universidades de Alcalá, Autónoma de Barcelona y León. Las veintiuna Academias de América y Filipinas trabajan coordinadas con la Española y se han constituido equipos de trabajo en la Argentina y la Puertorriqueña.

La primera parte de las Jornadas se celebrará en la Real Academia Española, Madrid, del 14 al 16 de julio. A ellas acudirán los coordinadores de los equipos de la Academia Argentina de Letras, la Academia Puertorriqueña de la Lengua Española, la Fundación Comillas y la Universidad de León. Se conciben como sesiones de formación y trabajo, que tienen como objetivo familiarizarse con la metodología y protocolos de trabajo que llevarán a cabo con sus correspondientes equipos. En la Real Academia Española trabajarán con los miembros del equipo de Madrid.

La segunda parte de las Jornadas tendrá lugar en Comillas (Cantabria, España) y tendrá como objetivo informar de las pautas de comunicación y colaboración entre los equipos. Para ello contaremos con la asistencia y participación de los coordinadores de los equipos que han colaborado en la fase piloto del proyecto, el de la Universidad Autónoma de Barcelona y la Universidad de Alcalá de Henares, así como del Director Académico del Instituto Cervantes, Francisco Moreno, y el académico de la RAE, Guillermo Rojo, coordinador general del proyecto. Las jornadas de trabajo se desarrollarán la tarde del jueves y la mañana del viernes. Terminarán con una sesión oficial de clausura en la segunda parte de la mañana del viernes a la que asistirán el Director de la RAE y Presidente de la Asociación de Academias, don Víctor García de la Concha, y el Director general de la Fundación Comillas, don Ignacio Rodríguez del Bosque, acompañados por los patronos de la Fundación Comillas.

Sobre el Corpus del español del siglo XXI

Desde la aprobación del proyecto en el Congreso de Academias de la Lengua Española celebrado en Medellín, en marzo de 2007, la Real Academia Española ha venido trabajando en la preparación del *Corpus del español del siglo XXI* (CORPES).

El 14 de junio de 2007, don Víctor García de la Concha, director de la Real Academia Española (RAE), y don Emilio Botín, presidente del Santander, firmaron un acuerdo que recoge el apoyo del banco en la preparación del Corpus del español del siglo XXI, un proyecto de la RAE en colaboración con las veintiuna instituciones que forman la Asociación de Academias de la Lengua Española. El trabajo servirá como fuente para los proyectos académicos y proporcionará materiales básicos para la investigación lexicográfica y gramatical del español.

El proyecto prevé, en su primera fase, la incorporación de 300 millones de formas procedentes de textos producidos durante los años 2000 a 2011, seleccionados según los diferentes parámetros de distribución en los que se estructura su diseño, para conseguir un corpus de referencia que constituya una muestra representativa del español de los primeros años del siglo XXI y que sirva para ofrecer el fundamento empírico exigible en la toma de decisiones relativas a las distintas tareas lexicográficas que llevan a cabo la Real Academia Española y la Asociación de Academias de la Lengua Española.

Diseño del Corpus del español del siglo XXI. Criterio de selección de textos

Dimensiones

El objetivo previsto es la incorporación de unos 25 millones de formas por año, adquiridas simultáneamente a la construcción del corpus.

Medio

Los textos proceden de libros (40%), prensa digital (40%), material de Internet (10%) y formas orales (10%)

Geográfico

El 30% de las formas proceden de textos producidos en España y el 70% restante, de textos americanos. Después, por zona lingüística, los textos se seleccionan de acuerdo con el cruce de una serie de criterios relativos al número de hablantes de cada área, la publicación de libros y la representación de prensa digital en Internet.

Temático

Los textos se ordenan en dos grandes bloques, ficción y no ficción, y cada uno de estos bloques en una serie de hipercampos o grandes zonas temáticas en las que se clasifican los textos.

No ficción: Ciencias y tecnología, Ciencias sociales, creencias y pensamiento, Política, economía y justicia, Artes, cultura y espectáculos, Actualidad, ocio y vida cotidiana, Salud

Ficción: Novela, relatos, teatro y guiones cinematográficos.

Sistema de trabajo

El sistema utilizado consiste en la división del trabajo entre un equipo central, situado en la RAE, y un número relativamente amplio de equipos colaboradores, vinculados a diversas universidades de España y Academias de la Lengua americanas. El equipo central se ocupa del diseño, selección de materiales, establecimiento del sistema de codificación, preparación técnica del personal de los equipos colaboradores y control general del trabajo. Los equipos colaboradores, dirigidos por un investigador vinculado a la institución colaboradora, podrán especializarse en la codificación de textos de una cierta zona lingüística o de unas determinadas características (prensa, textos orales, etc.). Su vinculación al proyecto se establece mediante convenios con la RAE. El sistema de trabajo permite, tras el período de puesta en marcha y preparación del personal, un ritmo rápido y constante de crecimiento de los materiales. Presenta, además, la ventaja adicional de que hace posible mejorar la preparación de un buen número de personas en diversos lugares del mundo hispánico, difundir una tecnología novedosa de acceso a los materiales lingüísticos y homogeneizar los procedimientos de codificación de textos.