

GLOSAS

ACADEMIA NORTEAMERICANA
DE LA LENGUA ESPAÑOLA



Director: **D. Gerardo Piña-Rosales**

Comisión de Traducciones
Presidente: D. Joaquín Segura

Vocales, por orden de antigüedad de nombramiento:

D. Emilio Bernal Labrada, Dr. Antonio Culebras, D.^a Elsa Nadal, D.^a Leticia Molinero, D. Francis D. Gómez,
Dr. Mordecai Rubín, Dra. Beatriz Varela, D. Guido Félix, Dr. José Manuel Gómez y Méndez, Dra. Estelle Irizarry,
Dr. Wenceslao Carlos Lozano, Lic. David J. Deferrari, Dra. Rima de Vallbona, Dra. Silvia Faitelson-Weiser,
Dr. Ignacio Soldevila-Durante, Dr. Ubaldo Di Benedetto, Dr. Carlos Abad, Dr. Valentín García Yebra, Dr. Fernando A.
Navarro, Embajador Tomás Rodríguez-Pantoja, Dr. Raúl Miranda Rico, Lic. Domingo Prieto, Dr. Eugenio Chang-
Rodríguez, Dr. Gustavo A. Silva, Dr. José Luis Sierra-Ponce de León, Dr. Antonio Garrido Moraga, Prof. Joaquín Badajoz,
Dr. Juan Manuel Pascual, Dr. Daniel Fernández Martín, Dr. Gerardo Piña Rosales, Dr. Nicolás Toscano Liria,
Dr. Francisco Marcos Marín, Rolando Hinojosa-Smith

Redacción: J. Segura, 606 Gateway, Valley Cottage, NY 10989. Teléf. y Fax: 845-268-4080
Correo electrónico: j.segura3@verizon.net

Volumen 6, Número 7

Junio de 2008

LA RAE PONE EN MARCHA EL CORPES (Corpus del Español de Siglo XXI)

Desde la aprobación de este proyecto en el Congreso de Academias de la Lengua Española celebrado en Medellín en marzo de 2007, la Real Academia Española, junto con la Asociación de Academias, han venido trabajando en la preparación del *Corpus del español del siglo XXI* (CORPES). El proyecto, que cuenta con el patrocinio del Banco

de Santander, prevé en su primera fase la incorporación de 300 millones de formas ["papeletas" electrónicas] procedentes de textos producidos durante los años 2000 a 2011 y seleccionados según diferentes parámetros de distribución, para conseguir un corpus de referencia representativo del español de los primeros años del siglo XXI.

Antecedentes: El CREA y el CORDE

En el proyecto del nuevo corpus académico (el CORPES) se aprovecha la perspectiva que proporcionan los más de diez años transcurridos desde el inicio del CREA (*Corpus de Referencia del Español Actual*) en 1996 y, pocos meses después, del CORDE (*Corpus Diacrónico del Español*). Hoy es absolutamente evidente que la puesta en marcha de esos dos corpus supuso una modificación radical en el sistema de trabajo de los equipos técnicos de la Real Academia Española, de todas las corporaciones que integran la Asociación de Academias de la Lengua Española y, en general, de cuantos se dedican a la investigación del español. Todas las obras publicadas por la RAE desde ese momento (la vigésima segunda edición del DRAE, el *Diccionario Panhispánico de Dudas*, el *Diccionario del Estudiante* y, sobre todo, el *Diccionario Esencial*), así como las que se encuentran actualmente en fase de redacción (la próxima edición del DRAE, la *Nueva gramática de la lengua española*, la próxima edición de la *Ortografía* y la del *Diccionario Académico de Americanismos*), se han beneficiado de los datos contenidos en el CORDE y, sobre todo, en el CREA ¿ más los extraídos del *Diccionario panhispánico de dudas* y el *Diccionario esencial*? y también se han basado para el contenido de sus entradas precisamente en la información contenida en estos antecesores.

En conjunto, el CREA y el CORDE, al abarcar datos de todas las épocas del español, procedentes de todos los países de habla hispana y de los más diversos tipos, aportaron los materiales pertinentes que necesitaban los lexicógrafos y los gramáticos para llevar a cabo su trabajo. La suma de ambos corpus (cerca de 300 millones de formas desde los orígenes del idioma hasta 1974 en el CORDE y algo más de 150 millones de formas desde 1975 hasta 2004 en el CREA) proporcionó y sigue proporcionando a todos los investigadores o a los simples interesados en nuestra lengua un recurso esencial para documentar con comodidad, rapidez y seguridad la mayor o menor frecuencia con que se utiliza una palabra, su distribución por países, años, tipos de texto, áreas temáticas, etc. En definitiva, estos dos corpus contienen cuanto se necesita para trabajar sobre bases sólidas, tanto en la línea estrictamente científica como en la que fundamenta la toma de decisiones normativas que la Asociación de Academias tiene a su cargo en todo el mundo hispánico.

El nuevo corpus

El diseño del CORPES tiene muy en cuenta la experiencia adquirida con los dos corpus anteriores. El CREA, científicamente adecuado y muy ambicioso cuando se formuló, se ha quedado corto para las necesidades actuales. En su concepción original, el período de cinco años en que se articula, contenía, en los tramos más recientes, 37,5 millones de formas, esto es, 7,5 millones de formas para cada año, distribuidas por tipos de textos, soportes, áreas

temáticas, etc., y repartidas al 50% entre España y América. Estas cifras resultan claramente insuficientes para fundamentar en ellas la enorme cantidad de decisiones que las Academias han de tomar para llevar a cabo las próximas ediciones del DRAE.

El nuevo corpus constará de 25 millones de formas para cada uno de los años comprendidos entre 2000 (para ajustar los períodos al CREA actual) y 2011 (fecha calculada de cierre de la adquisición de datos para la próxima edición del DRAE). En total, pues, son 300 millones de formas, cifra que garantiza la fundamentación empírica exigible a todas las obras académicas. A ella habrá que añadir los 125 millones de la parte del CREA actual comprendida entre 1975 y 1999. El conjunto resultante consistirá por tanto en un total de 425 millones de formas para el período comprendido entre 1975 y 2011, a los que se pueden añadir los casi 300 millones procedentes del CORDE cuando sea necesario analizar períodos anteriores.

De otra parte, la distribución de la procedencia de los textos establecida para el CREA (50% de España y 50% de América), única viable en el momento en que se gestó, resulta hoy difícilmente justificable. Teniendo en cuenta los diversos parámetros que es necesario considerar, la repartición que se considera adecuada para el CORPES es el 30% para España y el 70% para América. El nuevo proyecto permitirá hacer realidad un recurso cultural y científico que será sin duda el más importante del que se haya podido disponer jamás para el estudio del español.

Principales detalles

Es evidente que el diseño, control, adquisición y codificación de textos y la construcción de las herramientas informáticas que permitan la consulta de un corpus como el que aquí se describe desde cualquier lugar del mundo requieren una inversión considerable y el trabajo de un grupo de personas muy amplio. La RAE tiene experiencia en la gestión de proyectos de este tipo, derivada de la ampliación del CORDE de 75 a 300 millones de formas en poco más de dos años.

El proyecto prevé la división del trabajo entre un equipo central, situado en la RAE, y un número relativamente amplio de equipos colaboradores, radicados en diferentes universidades españolas y americanas y dirigidos por un investigador perteneciente a ellas. Los equipos podrán especializarse en la codificación de los textos de una cierta zona lingüística o la de unas determinadas características (prensa, textos orales, etc.).

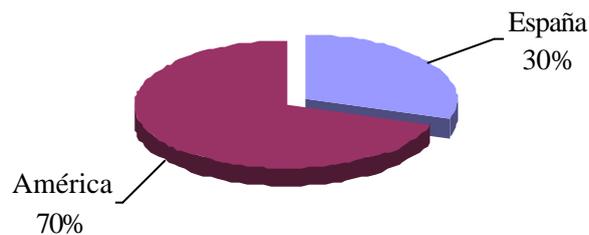
El sistema de trabajo que se ha diseñado en estos meses permite, tras el período de puesta en marcha y preparación del personal, un ritmo rápido y constante de crecimiento de los materiales. Presenta, además, la ventaja adicional de que hace posible mejorar la preparación de un buen número de personas en un grupo amplio de universidades, difundir una tecnología novedosa de acceso a los materiales lingüísticos y homogeneizar los procedimientos de codificación de textos.

El objetivo previsto es la incorporación de 25 millones de formas por cada uno de los años representados en el nuevo corpus (2000 a 2011, ambos inclusive). Los textos procederán de libros, prensa digital, material de Internet y formas orales según los porcentajes que se muestran a continuación:

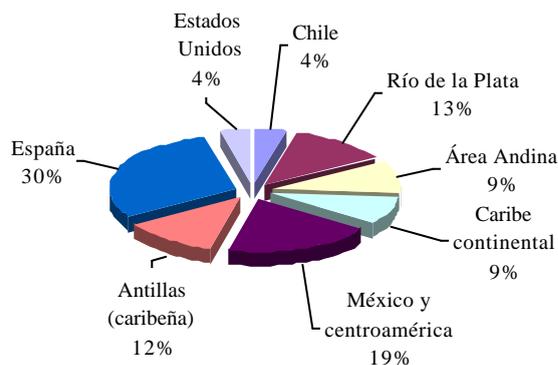
Libros	40%
Periódicos y revistas electrónicas	40%
Material de Internet ¹	10%
Oral	10%
	<hr/>
	100,00%

Los textos se seleccionarán geográficamente de acuerdo con el cruce de una serie de criterios relativos al número de hablantes de cada área lingüística, la publicación de libros y la representación de prensa digital en Internet: el 30% de las formas procederán de textos producidos en España y el 70% restante, de textos americanos.

Distribución prevista para España y América



Distribución prevista por zonas lingüísticas



Temática

Todos los textos, escritos y orales, seguirán la misma clasificación por temas, lo que posibilitará la búsqueda de un registro independientemente de su realización oral o escrita. Los textos se ordenan en dos grandes bloques, ficción y no ficción, y cada uno de estos bloques en hipercampos o grandes zonas temáticas como se indica a continuación:

CLASIFICACIÓN TEMÁTICA		
1. No ficción [75 % del total de los textos]	01. Ciencias y tecnología Astronomía, biología, bioquímica, botánica., ciencias y tecnología, ecología, electrónica, energía, estadísticas, física, geología, industrias diversas, informática, ingenierías, matemáticas, meteorología, química, tecnología, veterinaria, zoología y paleontología.	15%
	02. Ciencias sociales, creencias y pensamiento Antropología, arqueología, astrología y ciencias ocultas, civilización, educación, erotismo, ética, etnología, filosofía, folclore, geografía, historia, lingüística y lenguaje, literatura, mitología, mujer, psicología, religión, sexología, sociología, testimonios varios, urbanismo.	20%
	03. Política, economía, y justicia Ciencia militar, comercio, derecho, desarrollo, economía y hacienda, ejército, empleo, empresa, gobierno, iglesia, industria, justicia, legislación, marketing., negocios, política, protección civil, seguridad social, trabajo, tráfico, unión europea.	20%
	04. Artes, cultura y espectáculos Antigüedades, arquitectura, arte y cultura en general, artesanía, cine y vídeo, danza, diseño, escultura, espectáculos diversos, fotografía, medios de comunicación, música, pintura, publicidad, teatro.	15%
	05. Actualidad, ocio y vida cotidiana Actualidad, aficiones, asuntos domésticos, cocina, costumbres, deportes, gastronomía, jardinería, juegos, moda, tauromaquia, turismo, vivienda.	15%
	06. Salud Biomedicina, farmacología, medicina, medicinas alternativas, nutrición, psiquiatría, salud, sanidad pública.	15%
2. Ficción [25% del total de los textos]	01. Novela	40%
	02. Relatos	40%
	03. Teatro	20%

Codificación

Las normas de codificación se basan en el esquema propuesto por la *Text Encoding Initiative*, en lenguaje XML. Cada palabra incorporada al CORPES estará perfectamente documentada (país, zona geográfica, tema, tipo de texto, etc.) a través de la codificación contenida en la cabecera del texto, en la que se recogen los rasgos correspondientes a cada uno de los parámetros utilizados y también los datos de control interno necesarios. †

RESEÑA DE LIBROS

LUQUE DURÁN, J.d.D. & PAMIES BERTRÁN, A. (eds.): *Interculturalidad y lenguaje. Vol. I.: El significado como corolario cultural*, Granada, Método ISBN 978-84-7933-483-5.
Reseña de Wenceslao Carlos Lozano

Los dos primeros volúmenes de esta serie, dedicada a la interculturalidad desde el punto de vista lingüístico, recogen una selección de trabajos presentados en el *II Congreso Internacional sobre Lenguas y Culturas del Mundo* (Granada, noviembre de 2006).

Una parte de los trabajos de este primer volumen se centra en la relación lengua-cultura como tal, con la sugerente conferencia de Ángel López García *¿Alianza de lenguas, de culturas o de civilizaciones?*, cuyo provocador título anuncia efectivamente un análisis muy lúcido de los fundamentos teóricos de la temática de este tipo de estudios. Le sigue un no menos brillante artículo teórico de Gerd Wotjak, titulado *Algunas reflexiones acerca de lengua y cultura*, que acaba de situar y describir muy acertadamente el problema en la actualidad.

Una gran parte de los trabajos está dedicada a la demostración de la propia dialéctica entre lengua y cultura, plasmada en facetas puntuales muy concretas, como en el ilustrativo caso de las culturas “orientales”: China (como en los artículos de Zhang Zheng Quan; de Juan José Ciruela Alférez; de Lu Jia; y de Li Mei Liu Liu); Japón (Francisco Javier Tablero), o Corea (Antonio J. Doménech del Río). Desde un ángulo más explícitamente contrastivo (bilingüe), estos enfoques son complementados por trabajos como el de Luis Luque Toro sobre los *Parámetros socioculturales en la equivalencia léxica*, (italiano-español) o el de Sündüz Östürk Kasar sobre las discordancias semánticas culturalmente motivadas (turco-francés), o el de Antonio M. González Carrillo (*Condición e hipótesis. Consideraciones lingüísticas y culturales*).

La alta productividad de la motivación cultural en el diseño léxico-semántico de las lenguas tiende a manifestarse de una forma especialmente espectacular en las unidades fraseológicas, en sentido amplio, sobre todo locuciones y paremias (refranes, adagios, etc.); de ahí la importancia concedida a este campo de pruebas. La particular relevancia de la fraseología? en el sentido de modismos? queda bien demostrada en la

conferencia *Idiom semantics from a cognitive perspective*, de Dmitrij Dobrovol'skij, cofundador junto a Elizabeth Piirainen de la corriente que, dentro de los estudios fraseológicos, ha defendido el llamado “componente cultural” de las metáforas, como alternativa al cognitivismo de base psicósomática que imperaba en este campo desde que apareciera la teoría de Lakoff & Johnson sobre la metáfora.

Hay una colección de trabajos de sólida base empírica que apoyan esta idea, como el de Véselka Ángelova Nénkova (*Factores socioculturales para la formación de las unidades fraseológicas*), y muchas lenguas y parejas de lenguas son examinadas desde este punto de vista contrastivo-cultural, en busca de un modelo descriptivo de los mecanismos con los que la fraseología aprovecha, refleja y transmite la herencia cultural, centrándose en campos nocionales concretos.

Unos abordan la temática de la mujer, como los de Isabel Alijo Jiménez y Náder Al Jallad sobre la mujer en la fraseología en árabe; de Paula Santillán Grimm & Naima Ilhami sobre el piropo en árabe dialectal; de M^a Isabel González Rey sobre la imagen de la identidad femenina y masculina en la fraseología francesa; de Antonio J. Doménech del Río sobre la visión femenina del mundo en la fraseología coreana.

Otros tratan el protagonismo de la religión en las metáforas, como el de Carmen Mellado Blanco sobre el trasfondo bíblico de los modismos en español y alemán; de Náder Al Jallad sobre el trasfondo ideológico de frases hechas árabes; el de Marta Kobiela-Kwasniewska sobre la paremiología del calendario festivo en polaco; de Yaroslava Ivchenko sobre la imagen del Demonio en la fraseología y paremiología del ruso, ucraniano, bielorruso y polaco; de Julia Sevilla Muñoz sobre la fiesta de San Juan, contrastando la paremiología española y la francesa; de Limei Liu Liu sobre los valores morales y filosóficos inherentes de la paremiología china y sus (manipuladas) adaptaciones por los primeros acercamientos de los (misioneros) españoles sobre esta temática.

La productividad metafórica de las partes del cuerpo es tratada por Inés Olza Moreno, que aborda la base cognitiva y psicomotriz de las metáforas somáticas españolas, mientras que José Manuel Pazos analiza la base cultural de los somatismos alemanes y M^a Antonella Sardelli compara los somatismos fraseológicos del brazo en italiano y español. A ello se añaden estudios sobre otros campos conceptuales, como las partes de la casa en la fraseología francesa y española (Marina García Yelo), las metáforas del toro en esloveno (Polonca Kocjancic), la imagen de los extranjeros en la fraseología checa (Andrea Skorepova), las constantes figurativas de los insultos chinos (Juan José Ciruela Alférez), los prejuicios de valor en la fraseología inglesa de la biomedicina (Cristina Bayón García), la imagen de la sabiduría en la fraseología china (Zhang Zheng Quan), o el reflejo de la sociedad kurdo-iraquí en el refranero kurdo (Lawin Muhsin Ismail) y la valoración del propio lenguaje en el refranero español (Elvira Manero Richard).

Junto a los estudios empírico-descriptivos, se presentan varios trabajos que desarrollan las bases metodológicas para la elaboración de lo que sería un “diccionario

intercultural”, entendido como herramienta que permita registrar y describir ? desde la realidad de los datos y de manera comparable y translingüística? las conexiones culturales entre conceptos, vehiculadas por la herencia lingüística. Esta parte consta de un importante artículo de Juan de Dios Luque Durán ("La codificación de la información lingüístico-cultural en los diccionarios (inter)culturales") que delimita el campo de estudio, especifica sus metas y necesidades de este novedoso enfoque, así como los obstáculos a los que debe enfrentarse. Basándose en las teorías de Dobrovolskij & Piirainen, Antonio Pamies Bertrán, discute y ejemplifica los límites y subclasificaciones que requiere el estudio del componente cultural en un diccionario interlingüístico a través del estudio exhaustivo de un “culturema” zoomórfico (el lenguaje de la lechuga), en sus múltiples conexiones interconceptuales entre decenas de metáforas complejamente interrelacionadas. Lucía Luque Nadal ("Presupuestos teóricos y metodológicos para la redacción de la entrada ‘sangre’ en un diccionario intercultural") analiza un somatismo de gran productividad a la luz de esta propuesta epistemológica, centrándose sobre todo en el problema de la amplísima polisemia que, incluso en una misma lengua, poseen algunos símbolos culturales, llevando a situaciones límite la tarea del lexicógrafo en este terreno.

Sin duda, todos estos trabajos ? de alto nivel investigador? nos adentran en el corazón de una problemática muy actual y notablemente compleja, como es la de la dialéctica entre lenguas y culturas en un mundo cada vez más globalizado pero a la vez más atento a las peculiaridades y especificidades culturales locales, propias y ajenas, muy concretamente las que se manifiestan en el amplio contexto de la interculturalidad. Por ello, resultan sumamente instructivos no sólo para los especialistas en las materias tratadas aquí, sino para todos aquellos que sientan interés por conocer y entender aspectos generales y parciales de la compleja e imbricada red de relaciones semióticas que unen y separan a los seres humanos.

Nota de la Redacción: No es esta la primera reseña de Wenceslao Carlos Lozano que aparece en *Glosas*. Carlos Lozano es traductor literario, profesor de la Universidad de Granada y miembro numerario de la Academia de Buenas Letras de la misma ciudad, además de miembro correspondiente de la Academia Norteamericana de la Lengua. †

Nuestro Cibersitio

Al cumplirse el 35 aniversario de su incorporación a la Asociación de la Academias de la Lengua Española, la Academia Norteamericana acaba de estrenar cibersitio, todavía en construcción parcialmente, al que se podrá acceder mediante la dirección electrónica www.anle.us. Confiamos en que nuestros lectores difundirán esta noticia del nuevo cibersitio, en donde encontrarán, además de un índice de todos los números de *Glosas* publicados hasta la fecha, gran cantidad de datos y fotos sobre todas las Academias asociadas. A partir de este número, *Glosas* estará accesible, a título gratuito, en la sección correspondiente del nuevo cibersitio. Se seguirá enviando la versión impresa a los suscriptores que no tengan acceso a la Internet, o que prefieran recibir *Glosas* en versión impresa, aunque al renovarse su actual suscripción se les cobrarán solamente los gastos de impresión y envío postal. Les agradeceremos que nos hagan saber la preferencia de cada suscriptor. †