Corpus histórico del español de México

Constituir un *corpus* diacrónico para el español de México del siglo XVI al XIX, que pueda ser consultado mediante herramientas computacionales en una interfaz de Internet, es el objetivo del *Corpus histórico del español de México* (CHEM) proyecto desarrollado por el Grupo de Ingeniería Lingüística del Instituto de Ingeniería.

Ésta es una compilación de documentos históricos escritos en la Nueva España y el México independiente a los que se podrá acceder electrónicamente. Con ella, se abren las puertas a diversos tipos de investigación, desde aquellos relacionados con la inteligencia artificial y el procesamiento del lenguaje natural, hasta los de corte lingüístico que típicamente involucran alguno de los niveles de estudio del lenguaje, pasando por los relacionados con la extracción y recuperación de información y la minería de textos.

Los corpus históricos electrónicos para la lengua española de hoy en día no exhiben un interés particular por los sistemas dialectales de América. En los hechos, la representatividad de los corpus para estos sistemas y el cuidado para compilarlos descansa sobre todo en sus hablantes. En México, el CHEM constituye el espacio multidisciplinario para albergar documentos rescatados por especialistas del lenguaje y permitir su análisis mediante herramientas desarrolladas en el II UNAM.

El beneficio más importante de este trabajo es haber desarrollado una herramienta que se puede utilizar para investigaciones de diversos tipos, tanto en lingüística, como en desarrollos para tecnologías del lenguaje, de hecho se considera un repositorio documental, un patrimonio cultural.

A lo largo de la investigación se diseñó el mapa del corpus por cada siglo, tomando en cuenta géneros literarios (prosa, poesía, ensayo, etc) y temáticos (literatura, gobierno, religión, ciencia, etc), tipos de textos (libros, artículos, recetas, periódicos), autores prominentes, colecciones y archivos disponibles; también se planearon, diseñaron y desarrollaron herramientas para hacer exploraciones sincrónicas y diacrónicas del corpus. Se puso especial cuidado en localizar y digitalizar documentos clave para cada siglo del mapa del corpus, tanto de textos ya capturados electrónicamente, como de textos no digitalizados; se construyó también

CHEM Corpus Histórico del Español

Consulta de concordancias en el CHEM

Palabra: servjcio

Siglo: (•) XVI ( ) XVII ( ) XVIII ( ) XIX Ventana: 110 caracteres

Resultados de la búsqueda

#	Siglo		Pal.	
1	XVI	lacion de lo que alcançava a conosçer y cunplia al	**servjcio**	de vuestra majestad --aunqu
2	XVI	andasse proveer como viesse que convenia a su real	**servjcio**	y nos enbiasse a mandar a l
3	XVI	lo que en todo aviamos de hazer en las cosas de su	**servjcio**	y hazienda, que no tenian nj
4	XVI	erra enbié, y avrá proveydo como mejor cunpla a su	**servjcio.**	En ésta haré saber a vuestra
5	XVI	oveer de remedio en esta tierra como cu[m]ple a su	**servjcio.**	\ El qual dicho Ordas entró c
6	XVI	en lo de adelante no estaria tan obediente en	**servjcio**	de vuestra majestad, como c

una base bibliográfica para monitorear el contenido del corpus, mediante informes que permitan visualizar la distribución de los documentos según área temática y tipo textual, y se codificó información lingüística y extralingüística directamente en los documentos mediante etiquetado XML y de otros tipos (por ejemplo, POS). Este tipo de marcaje permite hacer búsquedas sofisticadas sobre la estructura de los textos y el lenguaje utilizado en ellos.

Este proyecto se realiza con financiamiento del Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica, de la DGAPA, UNAM. Durante el primer año, se desarrolló la primera versión del generador de concordancias de palabras gráficas que puede ser consultada, después de registrarse, en http://www.iling.unam.mx/chem.

Participaron en este proyecto Carlos Méndez Cruz, Laura Chavarría Cruz, Carlos Fonseca Rojas y Teresita Adriana Reyes Careaga, entre otros, bajo la dirección del doctor Alfonso Medina Urrea