

Informe de actividades recientes de VARILEX

Antonio Ruiz Tinoco

Durante el año 2005 se han celebrado varios congresos en los que han participado miembros del Equipo VARILEX. A continuación informamos brevemente de la intervención en METHODS XII, celebrado en la Universidad de Moncton, Canadá, entre los días 1 y 5 de agosto, y en el Congreso de ALFAL celebrado en la Universidad Autónoma de Nuevo León, Monterrey, México, entre los días 17 al 21 de octubre.

En METHODS XII se presentaron dos ponencias. En la sesión *Base de datos*, “The VARILEX Project – Spanish Lexical Variation”, de Hiroto Ueda y Antonio Ruiz Tinoco. En el taller organizado por David Heap y Enrique Pato, *New perspectives in Ibero-Romance dialectology*, Antonio Ruiz Tinoco presentó el proyecto COVARES, *Spanish Lexical Variation in Context: COVARES, Joining Lexical Surveys and Variational Corpora*, que se complementa con el proyecto VARILEX.

En la primera ponencia se hizo una presentación general pero breve del proyecto VARILEX, las características de las encuestas, la base de datos utilizable en Internet,

<http://lingua.cc.sophia.ac.jp/varilex/php-atlas/lista3.php>

con especial atención a la metodología de investigación. También se presentó la idea del diccionario panhispánico de variación léxica y algunos resultados en línea que se pueden bajar directamente de la página web del proyecto. Unas muestras de algunos de estos diccionarios en formato PDF se pueden bajar de

<http://lingua.cc.sophia.ac.jp/varilex/php-atlas/index.php>

En la segunda ponencia, se presentó el proyecto COVARES (COrpus de VARIación del ESpaÑol). En este nuevo plan de larga duración, ya se ha empezado a reunir textos en español de distintos tipos de géneros de todas las zonas hispanohablantes a través de Internet. Aparte del uso tradicional de un corpus de tipo general, uno de los objetivos de este proyecto de macrocorpus del español es que se pueda usar en Internet conjuntamente con la base de datos de VARILEX pudiendo hacer búsquedas y obtener el léxico en el contexto necesario para así poder observar mejor su

uso real. Para ello, se hace un uso intensivo de bases de datos relacionales en medio LAMP (Linux, Apache, MySQL y PHP) de manera análoga a como se está haciendo con VARILEX.

Entre las características principales de COVARES podemos mencionar que es posible la búsqueda compleja de ejemplos de uso de vocabulario con indicación de país, género, época e incluso autor. También las búsquedas pueden ser por palabras o frases completas, así como por trozos de palabras. En una etapa más adelante también se podrá hacer búsquedas por lemas. En la Fig.1 se puede ver el interfaz actual de búsquedas¹. Las búsquedas básicas se pueden hacer por cinco campos diferentes, además de la expresión clave, en el centro, también se puede condicionar el contexto y el país.

Corpus de variación del español -- COVARES --

BUSCAR	
zona	<input type="text" value="Chile"/>
contexto izquierdo	<input type="text" value="contains"/> <input type="text" value=""/>
izquierda	<input type="text" value="contains"/> <input type="text" value=""/>
expr. clave	<input type="text" value="is equal to"/> <input type="text" value="luego"/>
derecha	<input type="text" value="starts with"/> <input type="text" value="que"/>
contexto derecho	<input type="text" value="contains"/> <input type="text" value=""/>
<input type="button" value="Anterior"/> <input type="button" value="BUSCAR"/> <input type="button" value="Limpiar"/>	

Búsqueda realizada con COVARES => VARILEX PHP-Atlas VARILEX
 Programa Open Research Center de la Universidad Sophia de Tokio y el Programa B
 Copyright © Antonio Ruiz Tinoco

Fig. 1

En la Fig. 2 se muestra el resultado de una búsqueda, en la que se pueden apreciar campos básicos como la zona y los distintos contextos.

El campo *obra* a la izquierda nos muestra el número de la obra de la que ha sido extraído el ejemplo. Si hacemos clic podemos obtener algunos detalles de la obra en cuestión, tal como se ve en la Fig. 3.

Aunque por el momento los datos de la obra están incluidos en la base de datos, más adelante se podrán utilizar como condiciones de búsqueda. Es decir, se podrán

¹ La URL experimental de COVARES es: <http://lingua.cc.sophia.ac.jp/covares/> y para entrar es necesario introducir login y password, invitado/invitado. Al estar en fase de pruebas todavía, es posible que el contenido cambie de vez en cuando.

hacer búsquedas con condiciones tales como “en cuentos publicados entre los años 1930 y 1990 por autor argentino o chileno”, o “novela de García Márquez”, etc. Esperamos no tener problemas de derechos de autor ya que cumplimos los criterios normalmente “aceptados” como son el uso estrictamente académico y en ningún caso podemos obtener la obra completa. Es decir, el resultado de las búsquedas es tan corto que se puede considerar una cita de tal obra.

Corpus de variación del español -- COVARES --

Elegir Zona: All

[Mostrar todos](#) | [BUSCAR](#) | [Exportar\[XLS\]](#) | [Logout](#)
[Primera hoja](#) | [Anterior](#) | [Siguiente](#) | [Hoja final](#) | [Final](#)

Encontrado(s) 42 registro(s)

obra	zona	contexto izquierdo	izquierda	expr. clave	derecha	contexto derecho
15	Chile		Doblen	luego	a	la izquierda".
17	Argentina		Y	luego	así	:
19	España	el hombre es fugo y la mujer estopas y		luego	ipues	pasan las cosas!
15	Chile			Luego	bruscamente	disparé contra un árbol.
15	Chile	Y que toda la vida se esquivaron		luego	como	de mutuo acuerdo.
17	Argentina	Reflexiono unos instantes y		luego	con	menos entusiasmo, le doy una explicación menos técnica, conservando algunas geodésicas, pero haciendo intervenir aviadores y disparos de revólver.
18	México			Luego	condenaron	la puerta, hasta que él se seca;
15	Chile		Más tarde	luego	de	haber cerrado todas sus puertas, se extenderá sobre el lecho, volverá la cara contra la pared y recién entonces se echará a sufrir.
17	Argentina			Luego	descubrió	los satélites de Júpiter, que si bien constituirán otra prueba de esa hipótesis eran filosóficamente absurdos:
17	Argentina	Eddington necesitaba que los astrónomos y los físicos le dieran ese número obtenido con telescopios y balanzas,	para	luego	artificiar	la física.
17	Argentina	las ideas manixtas sobre el átomo son equivocadas		luego	el	socialismo es una ingenuidad.
15	Chile	Atisbó un suspiro		luego	el	crujir del lecho bajo el peso del cuerpo de Antonio.
15	Chile	Vertientes subterráneas la arrastraron		luego	en	su carrera bajo inmensas bóvedas de bosques petrificados.
15	Chile			Luego	enrociste	de golpe y llevándole los puños a la boca temblaste de pies a cabeza.
17	Argentina	En el buen tiempo viejo, un señor trabajaba un año en un escritorio, y haciendo cálculos,		luego	enviaba	un telegrama a un observatorio:
17	Argentina		Desde	luego	es	necesario que haya un paisaje permanente para que la repetición se pueda realizar.
17	Argentina	Son las características del sello que el hombre aplica sobre la naturaleza y	que	luego	ha	rescatado a través de los siglos ?
17	Argentina	Lo cierto es que el movimiento fue apoyado por los grandes financieros alemanes y extranjeros,	que	luego	haya	formado autonomía propia, es también cierto; pero eso sucede siempre en la historia;
15	Chile			Luego	llueve	nuevamente.
17	Argentina		Desde	luego	los	pobres de espíritu fueron, a partir de entonces, sus más encarnizados defensores y se mofan de los nuevos Mayer que aparecen por ahí.

[Mostrar todos](#) | [BUSCAR](#) | [Exportar\[XLS\]](#) | [Logout](#)
[Primera hoja](#) | [Anterior](#) | [Siguiente](#) | [Hoja final](#) | [Arriba](#)

Otros proyectos relacionados con COVARES => VARILEX PHP-Atlas VARILEX (Hiroto Ueda)

El proyecto COVARES ha sido subvencionado por el programa Open Research Center de la Universidad Sophia de Tokio y el Programa Báltazar Gracián del Ministerio de Cultura de España
 Copyright © Antonio Ruiz Tinoco

Fig. 2

Corpus de variación del español -- COVARES --

Found 1 record(s)

obra núm.	título	año publicación	país	lengua original	género	autor	año de nacimiento
15	La amortajada	1,941	Chile	español	cuento	María Luisa Bombal	1,910

Fig. 3

Una de las características prácticas de COVARES es que el resultado obtenido de las búsquedas se puede guardar en formato Excel, de modo que los datos se pueden volver a reprocesar o reordenar en otras aplicaciones a las que se esté acostumbrado.

También informamos que durante el *IV Simposio Open Research Center* de la Universidad Sophia de Tokio, celebrado el 9 de diciembre de 2005, Antonio Ruiz

Tinoco hizo una presentación de estos proyectos.

En el congreso de ALFAL, Toshihiro Takagaki hizo una presentación de un nuevo proyecto sobre variación gramatical, que ya ha recogido datos principalmente en España y próximamente se empezará a recoger datos en varios países hispanohablantes. Esperamos la colaboración de los miembros del Equipo Varilex.

A continuación, se presentaron las novedades de Varilex como se hizo en Methods XII, y pasamos a la discusión del futuro del proyecto entre los asistentes.

Los coordinadores del proyecto Varilex esperamos terminar el proyecto, si es posible, para poder presentarlo en el próximo congreso de ALFAL. Creemos que la mejor forma sería editar un volumen con los principales trabajos revisados y un CD-ROM con los datos en formato fácilmente utilizable. También estamos considerando la posibilidad de reproducir toda la base de datos en un CD-ROM. Ya que la base de datos se administra en un servidor Linux, se podría reproducir toda la base de datos completa con los programas necesarios en un CD-ROM de modo que funcionara como una réplica de la base de datos original que se podría utilizar de la misma forma que la que se encuentra en línea. Los detalles técnicos de esta posibilidad serían análogos a los del sistema KNOPPIX², que consideramos la alternativa más viable en estos momentos.

Sin embargo, independientemente de la forma que tome la presentación final y la distribución de los datos, está claro que hay que terminar de reunir la mayor parte de datos de una manera que resulte útil para su posterior estudio. Entre las tareas que queda por hacer creemos que es de la mayor importancia las relacionadas con los datos en sí:

- a) Fijar el número definitivo de ciudades hispanohablantes que se deben incluir en el proyecto. Ya que el proyecto Varilex trata la variante urbana, necesitamos reunir datos de bastantes ciudades hispanohablantes, bastantes de ellas con más de un millón de habitantes y todavía no representadas en los cuestionarios. Para ello, lo ideal sería contactar con nuevos colaboradores dispuestos a llevar a cabo las encuestas en un período relativamente breve de tiempo. Alternativamente se podrían llevar a cabo las encuestas por medio de Internet, pero creemos que puede plantear algunos problemas de metodología. Por ejemplo, las siguientes ciudades con más de medio millón de habitantes no tienen representación todavía en los datos de Varilex.
Argentina: Córdoba, Rosario, Mendoza, La Plata, Mar del Plata y San Miguel de Tucumán
España: Bilbao

² <http://www.knoppix.org/>

México: Puebla, Ciudad Juárez, Tijuana, León, Mexicali, Culiacán, Acapulco de Juárez, Tampico, Mérida, Chihuahua, San Luis Potosí, Toluca, Querétaro, Morelia, Hermosillo, Saltillo, Orizaba, Torreón, Villahermosa, Durango, Veracruz

Colombia: Cali, Barranquilla, Cartagena, Cúcuta, Bucaramanga, Pereira

Venezuela: Maracaibo, Maracay, Barquisimeto, Ciudad Guayana, Petare, Ciudad Losada

Ecuador: Guayaquil

Perú: Callao, Chiclayo, Trujillo, Piura, Iquitos

Bolivia: Santa Cruz de la Sierra, Cochabamba

Chile: Valparaíso, Rancagua, Puente Alto

- b) En las ciudades en las que ya hay colaboradores locales, todavía falta llevar a cabo algunas de las encuestas.
- c) Considerar elevar el número mínimo de informantes por cada una de las encuestas. Hasta el momento, el número mínimo es de cuatro informantes por ciudad. En caso de obtener una ocho informantes por cada localidad los datos serían más fiables así como los cálculos de cuantificación.
- d) Corregir los errores de las encuestas ya realizadas. Hay que tener en cuenta todos los tipos de errores, ya sean simples de transcripción de las encuestas escritas a mano a la base de datos, como los de tipo conceptual.
- e) Enriquecer los datos con comentarios sobre los usos de los diferentes términos.
- f) Ni que decir tiene que también es necesario mejorar la programación de la base de datos y el interfaz en Internet, así como los relacionados con la generación automática de mapas que incluyan datos estadísticos también. Sería ideal preparar un servidor GIS para mostrar los datos, pero creemos que todavía es muy prematuro para poder llevar implementarlo.

Por todo lo comentado anteriormente, quedamos a la espera de los comentarios de los miembros del Equipo Varilex y sobre todo de la inestimable colaboración, sin la cual sería imposible llevar a buen fin el proyecto.