

Investigaciones sobre la variación léxica del español: Proyectos y resultados de 1992 a 2007

Hiroto Ueda, Universidad de Tokio
Antonio Ruiz Tinoco, Universidad Sofía, Tokio

En esta ocasión nos permitimos revisar los trabajos hasta el presente realizados, junto con algunas publicaciones, y estudios basados en nuestros datos léxicos del español. Este año es, después de quince años desde el inicio del proyecto, el último de la primera fase de nuestras encuestas, con miras hacia el futuro que consistirá en desarrollar nuevos métodos de investigación. Hablaremos, por lo tanto de dos temas; resultados anteriores y proyectos para el futuro.

1. Proyecto¹

Nuestro proyecto, después de un año de preparación en 1992, empezó con ocasión del *Décimo Congreso de la Asociación de Lingüística y Filología de América Latina (ALFAL)* celebrado en Veracruz, México, en 1993. Allí explicamos nuestra intención y el resultado de un estudio piloto. Desde entonces hasta el presente, hemos venido investigando nuestro tema: variación léxica del español.

1. 1. Colaboradores

Este proyecto nace de un anhelo por conocer la situación actual del léxico español de todo el mundo. Como todos sabemos, el español es una lengua multinacional, distribuida en una gran área del mundo y que se está desarrollando con dinamismo. Los problemas de espacio, --más de veinte países con sus capitales y grandes ciudades importantes-- y tiempo --el constante cambio lingüístico y el nacimiento de nuevos términos--, no podrían ser solucionados por un solo investigador sin una buena comunicación exterior, y en esta época, sin informatización. Dado que no existe un estudio exhaustivo e informatizado del léxico urbano actual de las distintas regiones hispanohablantes, proponemos formar una red internacional de investigación, con los siguientes objetivos:

- Formar una red internacional de investigación sobre el léxico español moderno-urbano. No se trata sólo de elaborar un estudio sobre algún tema específico, sino buscar una forma de cooperación internacional para la investigación lingüística.
- Reunir datos del español moderno (con especial atención a la variante urbana) para después analizarlos con métodos multivariados y cartografiarlos en múltiples mapas con las grandes ciudades hispanohablantes.

¹ Esta sección es una versión reducida de nuestro trabajo de 2003 (Ávila, Samper y Ueda, Ruiz Tinoco): 160-168.

- Distribuir los datos obtenidos a todos los participantes de este proyecto. Los datos pueden ser usados por los participantes, ya sea individualmente o en equipo, como en sus estudios o publicaciones.
- Establecer un Centro como banco de datos en la Universidad de Tokio. Las cuestiones léxicas, al ser recibidas en el Centro y reunidas en cierta cantidad, son distribuidas a los participantes. Una vez recogidos los materiales, el Centro se encarga de imprimirlor en forma de libretos y enviárselos a los miembros.

El deseo de los coordinadores del proyecto es crear un banco de datos léxicos que reúna los materiales siempre actualizados en todo momento. No se trata de hacer una obra lingüística con datos más o menos estables, sino formar un archivo de léxico distintivo del que podamos sacar informaciones útiles para cada investigador y ofrecer a los estudios interesados los datos que necesiten obtener para realizar sus trabajos dialectológicos, lexicográficos, lexicológicos, taxonómicos, históricos, etc.

Este proyecto tiene una característica muy peculiar: el mismo colaborador puede ser informante, observador, investigador, revisor de los datos y planificador de futuros estudios. Como informante tiene que contestar cuestionarios enviados por el centro de investigación (en Tokio). Se espera que contesten las preguntas sobre el léxico en un corto plazo de acuerdo con las indicaciones expresadas en el formulario mismo. Para no cargar con mucho trabajo a los participantes, se trata de responder cada año a unas doscientas preguntas de respuesta selectiva.

Si el colaborador tiene algún comentario sobre el uso de las palabras, se convierte en observador en ese momento y anota al margen sus apreciaciones sobre el uso de la palabra en cuestión. Todas ellas son muy útiles y se archivan en el inventario léxico central.

Cada año, los colaboradores reciben los resultados de la recolección de datos de los lugares investigados y los pueden usar individualmente o en colaboración con otros. Sería interesante que se compararan nuestros materiales con los reunidos en otras investigaciones. Los aspectos de interés que surjan de esa comparación pueden ser compartidos y discutidos con otros colaboradores del proyecto.

1. 2. Encuestas

La investigación propuesta tiene dos dimensiones básicas: la lingüística (unidades léxicas) y las geográficas (localidades). Ninguna de ellas es de carácter cerrado. Así, en nuestra primera aproximación hemos trabajado con 25 puntos de encuesta (Parcela A de nuestro gráfico), pero este número puede ampliarse de forma ilimitada (Parcela -B):

Palabras	Localidades			
	L-1	L-2	L-3 ... L-25	L-26 ... L-n
p-1		Parcela -A		Parcela -B
p-2				
p-3				
...				
p-219				
p-220		Parcela -C		Parcela -D
p-221				
p-222				
...				
p-n				

Del mismo modo, por el momento contamos con el análisis de 219 unidades léxicas, que serán recogidas también en aquellas ciudades que se vayan incorporando al proyecto. Pero esto no quiere decir que no se pueda ampliar el número de formas. De hecho nuestro modelo de investigación prevé la incorporación progresiva de nuevas unidades (Parcelas C y D):

1.3. Estado actual y prospectivas

Hasta el presente hemos realizado once encuestas (de A a K) referentes a 1490 conceptos:

Encuestas anuales

Año	Encuesta	Ciudades	Conceptos	Publicación
1994	A	44	152	Varilex 3
1995	B	39	163	Varilex 4
1996	C	38	193	Varilex 5
1997	D	45	199	Varilex 6
1998	A,E	44	E: 41	Varilex 7
1999	F	61	187	Varilex 8
2000	B,C,G	35	G: 13	Varilex 9
2001	H	31	176	Varilex 10
2002	D,I	27	I: 17	Varilex 11
2003	J	25	104	Varilex 12
2004	K	26	145	Varilex 13
2005	F	24	100	Varilex 14

El proyecto marcha continuamente, investigando más ciudades y estudiando más conceptos del léxico variable. Nuestros objetivos inmediatos son los siguientes:

- Establecer más contactos con las ciudades importantes del continente

americano y Filipinas. Los países representados sólo por una ciudad los deberíamos estudiar por lo menos con un punto más de investigación. Son El Salvador, Costa Rica, Panamá, Bolivia, Paraguay y Uruguay.

- Ampliar el número de conceptos hasta llegar a la cantidad de 2.000.
- Elaborar páginas interactivas en Internet de modo que el investigador pueda ir acumulando y corrigiendo los datos. Esto nos ayudará a establecer más contactos y, al mismo tiempo, publicar simultáneamente los resultados de las investigaciones en tiempo real.

Con todo lo expuesto hasta aquí, creemos haber presentado unos conceptos nuevos de investigación lingüística: crear un archivo libre, abierto, relacional, flexible en continua renovación. El proyecto nace de la necesidad de seguir el ritmo acelerado de los cambios lingüísticos, tanto diacrónicos como diatópicos, y hacer una aportación al desarrollo de la metodología lingüística en alguna de sus ramas. Proponemos poner en marcha un método de investigación flexible y eficaz. La investigación lingüística actual necesita proyectos de alcance nacional e internacional y mucho más cuando se trata de estudiar una lengua con la proyección que tiene el español. Sin una colaboración internacional poco se podrá avanzar. Por otro lado, un proyecto como el que ahora presentamos nace con la finalidad de ser útil a todos los que tienen algún interés por los hechos lingüísticos. El proyecto y los materiales que genere serán obra de todos los que colaboren y todos han de beneficiarse de las ventajas que el trabajo en equipo puede ofrecer.

2. Estudios

Los estudios relacionados con el Proyecto Varilex pueden ser clasificados en los seis grupos siguientes: (1) Metodología, (2) Datos de distribución geográfica, (3) Comentarios, (4) Temas lingüísticos, (5) Temas dialectales y (6) Aplicación a la didáctica al español.

2. 1. Método de investigación

A principios de los años 1990, publicamos en varias ocasiones artículos sobre el método, programas, colaboradores y algunos resultados de las encuestas piloto: Equipo de investigación Varilex (1993), Takagaki (1993), Ueda (1994, 1996a, 1999, 2003)².

Con respecto a las técnicas de estadística aplicadas a los datos léxicos, destacamos la elaboración de la matriz de coeficientes de correlación, el análisis de cluster y nuestro propio método de patronización. En esta ocasión nos permitimos presentar el último, por no ser conocido en nuestro ámbito de investigadores de la variación lingüística³.

Ciertamente es interesante observar la distribución geográfica de las palabras revisándolas una por una. Sería tarea del lexicógrafo si se interesara en

² Para los detalles véase Ávila et al. (2003). Del último trabajo hemos reproducido una parte de la misma obra en la sección anterior. En cuanto a las técnicas de cartografía y bases de datos: Aoto (2004), Ruiz Tinoco (1999, 2000, 2001). Hemos presentado en la sección anterior su parte esencial.

³ Es una reproducción parcial de la sección 3 de nuestro trabajo (2005). Para los detalles de procesamiento, véase Ueda (2006).

etiquetar las características locales del léxico en cuestión. La geografía lingüística, por otra parte, trata el léxico como conjunto para observar el panorama que se presenta

en el universo de la matriz (vocablos por localidades), tal como se presenta en la Figura 2.1. A partir de la matriz bidimensional, podemos elaborar una imagen nítida de la distribución.

Supongamos que estamos ante una distribución como la presentada en la parte izquierda de la Figura 2.2.

Fig.2.1 Estado prepatronizado

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	A	B	C	D	E	F			A	B	C	D	E	F	
2	a			+		+			d		+				
3	b				+	+			e		+	+			
4	c			+	+	+			g		+	+	+		
5	d		+						h		+	+			
6	e		+	+					a		+	+	+		
7	f			+			+	+	b		+	+			
8	g		+	+	+				c		+	+	+		
9	h		+	+					k			+	+		
10	i			+	+	+		+	j				+	+	+
11	j				+	+	+		f				+	+	
12	k				+	+			i						+

Fig. 2.2 Patronización

	A	B	C	D	E	F
d		+				
e		+	+			
g		+	+	+		
h		+	+			
a		+		+	+	
b			+	+		
c				+	+	+
k					+	+
j						+
f						+
i						

En los dos gráficos, uno a la izquierda y otro a la derecha, podemos observar que las posiciones de las reacciones (+) no varían antes y después de la patronización. Por ejemplo en la línea-a, tenemos reacciones en las columnas B y D, tanto en el gráfico de la izquierda como en el de la derecha. Lo mismo puede observarse para otras líneas (de b a k). Lo que ha cambiado es la posición de las líneas, de $a-b-c-d-e-f-g-h-i-j-k$ a $d-e-g-h-a-b-c-k-j-f-i$, por lo cual se obtiene una distribución más nítida en forma de línea diagonal desde la parte izquierda superior a la derecha inferior. Por medio de esta operación, nos es posible interpretar la distribución de los objetos (a, b, c, \dots, k) agrupados según la semejanza de la elección de los atributos (A, B, C, \dots, F). Remitimos la explicación matemática sobre la patronización a nuestros trabajos anteriores, y ahora procedamos directamente a analizar los datos tratados en este estudio. Compárense las dos figuras: Fig-2.3 y Fig-2.4⁴. La primera (Fig-2.3) es la visión total de la distribución prepatronizada. La Fig-4, por otro lado, es el resultado del proceso de patronización. Podemos confirmar que se ha conseguido, *grosso modo*, la patronización diagonalizada.

2.2. Distribución geográfica

A partir del año 1993 hemos venido publicando los resultados tanto de forma impresa como en forma de páginas web con respecto a los datos de distribución

⁴ Los datos son del léxico de blasfemia en español, tratado en la encuesta D (1997).

geográfica⁵. Los informes impresos también se encuentran en el mismo sitio en formato PDF. Lo principal de estos datos son los cuadros bidimensionales constituidos por formas lingüísticas en un eje y localidades en otro. En la primera parte se encuentran la definición y ejemplo de uso y formas que se han preguntado:

Ejemplo: El jefe no se atreve a llamarle la atención, porque (_____).

- (1) agarrarse a buenas aldabas; (2) estar apitulado; (3) estar bien agarrado;
- (4) estar enchufado; (5) tener buenas agarraderas; (6) tener buenas aldabas;
- (7) tener enchufe; (8) tener la manija; (9) tener palanca.

A continuación mostramos el cuadro bidimensional correspondiente:

	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	3	3	3	3	3	3															
1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7													
E	E	E	E	E	E	E	E	E	E	E	C	R	P	E	E	M	M	M	E	H	H	N	P	C	V	B	C	C	P	A	A								
S	S	S	S	S	S	S	S	S	S	S	S	U	D	R	U	X	X	X	L	O	O	I	N	O	E	O	H	H	A	R	R								
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-								
C	S	S	Z	B	G	M	V	S	M	A	P	S	S	D	N	N	M	A	M	S	T	N	L	P	M	C	P	A	C	S	S								
O	C	T	L	A	A	D	A	E	L	L	U	A	C	T	O	Y	O	O	G	E	R	S	E	A	A	E	B	A	R	O	S	A	J	U					
R	O	D	M	R	R	L	D	L	V	G	M	R	L	U	I	R	K	R	N	S	D	A	G	C	O	N	D	O	Z	I	N	U	L	N					
																																					Total		
0	No	se	me	ocurre.	-	-	-	-	-	-	-	1	-	-	-	1	-	-	2	-	-	2	-	-	1	-	-	-	-	-	-	-	-	7					
2	estar	apitulado	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	8						
3	estar	bien	agarr	-	-	-	-	-	-	-	-	1	-	2	-	1	3	1	-	-	1	1	1	1	-	-	-	-	-	-	-	-	14						
4	estar	enchufado	3	3	-	2	4	2	1	-	1	4	4	2	2	1	-	1	-	-	-	-	-	2	-	-	-	-	-	-	-	-	32						
5	tener	buenas	aga	1	-	3	-	1	-	2	-	1	-	2	-	1	1	-	1	-	1	-	-	-	-	-	-	-	-	-	-	15							
6	tener	buenas	ald	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1							
7	tener	enchufe	4	3	3	3	2	1	3	2	4	3	4	3	3	-	1	1	-	-	-	-	-	-	-	-	-	-	-	-	44								
8	tener	la	manija	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	4						
9	tener	palanca	-	-	-	-	-	-	-	-	1	-	3	1	4	4	2	4	3	7	3	4	-	4	3	4	4	4	-	-	1323								
	Total	8	6	3	8	6	3	5	3	7	7	8	8	5	8	4	4	4	10	4	5	3	9	3	4	3	6	4	5	5	6	3	4	4	2	5	3	4	189

Finalmente, ofrecemos otras formas obtenidas en las encuestas: *estar acomodado* AR-BUE (1); *estar agarrado* AR-BUE (1); *estar apadrinado* EL-SSA (1); *estar bancado* AR-BUE (1); *estar bien apañados* AR-SJN (1); *estar bien parado* MX-MEX MX-MRD HO-NAC (3), etc.

2.3. Comentarios

En los informes incluimos comentarios sobre la variedad de la localidad específica observada en el informe del año anterior⁶. Nuestro colaborador de Argentina, Francisco Fernández, desde 1994 ha contribuido con sus comentarios sobre la variedad de Salta (Argentina), de la que vamos a citar una parte:

“Antes de abocarme al estudio de los términos y compuestos que muestran total o parcial unanimidad, como también dispersión en las acepciones, echaré una mirada general en todas aquellas que son dignas de observar en

⁵ Equipo de investigación Varilex (1993-2006). La dirección del la página web es: <http://gamp.c.u-tokyo.ac.jp/~ueda/varilex>.

⁶ Fernández (1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2003, 2004); Nakajima (1999, 2000), Segura (1999) y Perelló (2002).

algunas de sus características, en atención a los usos que, en mi calidad de usuario y observador calificado (como lo expliqué arriba), he estimado conveniente destacar. Haré esto teniendo en cuenta el orden de distribución de palabras, tal como viene en la publicación.

D002 – ANNOY (CAUSAR MOLESTIA) Los usos más efectivos detectados (según mi criterio, por supuesto) son los siguientes, teniendo en cuenta: a) espontaneidad y total informalidad; b) espontaneidad con una cierta formalidad; c) semiformalidad; d) formalidad; e) alternativas para alguna de las categorizaciones anteriores; f) fuerte preferencia. Estos parámetros serán válidos para todo el análisis. En caso de aplicarse sólo a Salta, puedo consignarlo como (S); si fueran compartidos por San Juan (SySJ) o por Buenos Aires (SyBA); si lo fueran por las tres ciudades (Sy2). e) *Hinchar, hinchar las pelotas, las bolas, joder* (Sy2).”

2. 4. Temas lingüísticos

Con respecto a los temas lingüísticos, contamos con estudios de Ueda (1995a) sobre fonética y fonología, Koike (2001) sobre fraseología, Nishimura (2002) sobre diminutivos, Aoto (2003) sobre las perifrasis verbales de gerundio, Ueda (1996b, 2000, 2005) sobre el léxico y Ruiz Tinoco (2004) sobre el parámetro sexo. Reproducimos la introducción del trabajo de Koike (2001:77-78), donde se explica la importancia de los estudios variacionistas de expresiones fraseológicas:

“Dada su extensión geográfica, el español presenta una importante variación fraseológica. Por variantes fraseológicas entendemos aquellas unidades fraseológicas que, manteniendo el mismo significado, son parcialmente idénticas en su estructura y en sus componentes. Son variantes, pues, *hacer la pelota, hacer la rosca, hacer la barba y hacer la pata*, pero *chupar las medias* no es variante de *hacer la pelota*, sino una locución sinónima. La lingüística hispánica apenas ha prestado atención a la variación fraseológica, debido fundamentalmente a la falta de datos fraseológicos que cubran toda el área hispana y con los que se puedan hacer investigaciones sobre el tema.”

A esta explicación sigue su comentario sobre la falta de estudios realizados hasta el momento.

“Los escasos trabajos sobre los fraseologismos de determinados países no han sido suficientes para obtener una visión global de la variación fraseológica del mundo hispánico. A través de la red del proyecto Varilex se llevó a cabo en 1999 la encuesta sobre la variación fraseológica del español en el mundo. El material de dicha encuesta ha sido elaborado a partir de 186 conceptos. El número total de unidades léxicas y fraseológicas recopiladas por dicha encuesta es de seis mil seiscientas diez, lo que pone de manifiesto que existe un promedio de 35 expresiones para cada concepto. Se trata de la primera encuesta de mayor envergadura sobre la variación fraseológica del español. Hasta dicha fecha la fraseología del español no disponía de datos estadísticos que abarcaran toda el área hispana, lo que impedía, como hemos dicho antes, que la investigación se basara en datos fieles a la realidad lingüística.”

2. 5. Temas dialectales

Al lado de los estudios panhispánicos, se ha presentado un estudio específico de un área determinada, siempre con una visión amplia para situar dicha área dentro del panorama de la variación general. Aoto (2005) se dedica a la variedad de Paraguay. En su estudio se han presentado más de 40 ítems de variación, los cuales se dividen en dos tipos: el que tiene sólo variantes locales, y el otro que presenta la coocurrencia de las variantes locales y la forma estándar del español peninsular. El análisis de cluster ha permitido establecer el tráfico vial como el factor geográfico más destacado. Se resalta igualmente el factor social, que produce un contraste entre las regiones urbana y marginal. Por otra parte, Ueda (en prensa) trata de la variedad del léxico de Guinea Ecuatorial con los datos reunidos tanto *in situ*, como en París con la ayuda de los residentes ecuatoguineanos.

2. 6. Aplicación

Nuestro estudio de variación léxica, a pesar de que pertenece al campo de la geografía lingüística, no aspira a contribuir solamente en el enriquecimiento de nuestro conocimiento del léxico, sino que también pensamos en su aplicación a la didáctica del español como segundo idioma. Ueda (2003) presenta algunas reflexiones sobre el tema de la aplicación del estudio dialectológico a la enseñanza del español. Uno de los resultados concretos puede ofrecerse en forma de diccionario. Rubio et al. (1992, 2004), Takagaki (2003) y Ueda et al. (2006) han utilizado los datos recogidos por Varilex. Otra aplicación posible es ofrecer nuestros datos y propuestas a los medios de comunicación, los cuales se enfrentan diariamente a los problemas causados por la existencia de distintas formas para el mismo concepto (Ueda 1996c).

3. Base de datos y cartografía automática

Los datos obtenidos de las sucesivas encuestas se van almacenando en nuestra base de datos tanto con el simple objetivo de ordenar los datos mismos como la facilidad de cualquier tipo de análisis cuantitativo. Además, por medio de un servidor web los datos están a disposición de cualquier persona con acceso a Internet. A continuación exponemos brevemente las características técnicas de la base de datos, y algunos usos actuales.

3.1 Entorno LAMP⁷,

Además de la página web⁸ del coordinador general del proyecto, Hiroto Ueda, utilizamos otro servidor experimentalmente⁹ utilizando el entorno *LAMP*, es decir, *Linux*¹⁰ como sistema operativo; *Apache*¹¹, como servidor web; *MySQL*¹², como administración de base de datos y *PHP*¹³, como lenguaje de programación.

⁷ [http://en.wikipedia.org/wiki/LAMP_\(software_bundle\)](http://en.wikipedia.org/wiki/LAMP_(software_bundle))

⁸ <http://gamp.c.u-tokyo.ac.jp/~ueda/varilex/index.html>

⁹ <http://lingua.cc.sophia.ac.jp/varilex/php-atlas/lista3.php>

¹⁰ <http://www.linux.org/>

¹¹ <http://www.apache.org/>

¹² <http://www.mysql.com>

¹³ <http://www.php.net>

Utilizando tanto hardware como software relativamente poco complejo podemos preparar un sistema multiplataforma de manera que cualquier usuario conectado a Internet con cualquiera de los sistemas operativos más frecuentes, como Windows¹⁴, Mac OS¹⁵ o Unix¹⁶, y con cualquier navegador, como Firefox¹⁷, Internet Explorer¹⁸ o Netscape¹⁹ pueda acceder sin problemas a la base de datos. Una ventaja de este entorno es que el software empleado principalmente es de licencia GNU/GPL²⁰ que resulta gratuito y se utiliza ampliamente por lo que es fácil encontrar manuales o expertos a los que poder consultar en caso de necesidad.

Este entorno se complementa con una serie de librerías²¹, ya sea con funciones gráficas o para generar documentos en formato PDF, etc.

3.2 Estructura de la base de datos

La base de datos de Varilex consta de cuatro tablas principales relacionadas entre sí como se muestra en la Fig. 1. Es decir, *conceptos*, *respuestas*, *ciudades* e *informantes*. En estas tablas se van introduciendo los datos correspondientes a los detalles de las encuestas, contestaciones obtenidas, características de los informantes, etc. para después por medio de interfaces adecuadas poder obtener el resultado de las consultas en línea. Las interfaces se preparan básicamente en lenguaje PHP y el comando SELECT de MySQL. Aunque resulta muy útil, el usuario no tiene necesidad de conocer los detalles técnicos de estos lenguajes ya que las interfaces se pueden utilizar simplemente rellenando algunas ventanas y haciendo clic en los lugares indicados de forma bastante intuitiva. Hay otras interfaces más complejas²² para usuarios conocedores de MySQL. Evidentemente, esta estructura es mejorable o ampliable en el futuro.

¹⁴ <http://www.microsoft.com/windows/default.mspx>

¹⁵ <http://www.apple.com/macosx/leopard/index.html>

¹⁶ <http://www.unix.org/>

¹⁷ <http://www.firefox2.com/en/>

¹⁸ <http://www.microsoft.com/windows/downloads/ie/getitnow.mspx>

¹⁹ <http://home.netscape.com/>

²⁰ <http://www.gnu.org/>

²¹ Una librería es un conjunto de funciones de un lenguaje de programación. Con el uso de las funciones que se incluyen en las librerías podemos crear algoritmos ya comprobados y no partir de cero en la creación de los programas.

²² Los administradores utilizamos también la conocida interfaz *PhpMyAdmin* (http://www.phpmyadmin.net/home_page/index.php), con funciones más complejas y para el mantenimiento en sí de la base de datos.

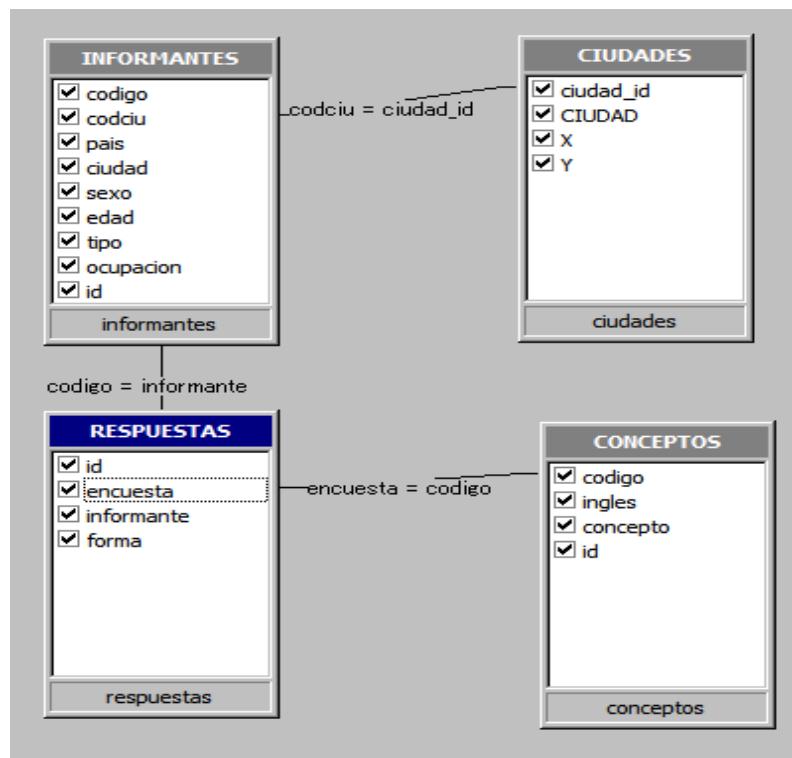


Fig. 3.1 tablas de la base de dato

3.3 Métodos de búsqueda

El programa de administración de base de datos MySQL permite una gran cantidad de funciones en una interfaz como la que se muestra en la Fig. 3.2. A modo de ejemplo, para calcular la matriz bidimensional para un concepto (A001) y varias ciudades (Buenos Aires, México y Madrid) se puede utilizar la siguiente fórmula:

```

SELECT respuesta,
SUM( IF(ciudad = 'BUE', 1, 0) ) AS 'Buenos Aires',
SUM( IF(ciudad = 'MEX', 1, 0) ) AS 'México',
SUM( IF(ciudad = 'MAD', 1, 0) ) AS Madrid,
COUNT( * ) AS total
FROM contestaciones
WHERE orden = "A001"
AND ciudad IN ('BUE', 'MEX', 'MAD')
GROUP BY respuesta
ORDER BY total DESC

```

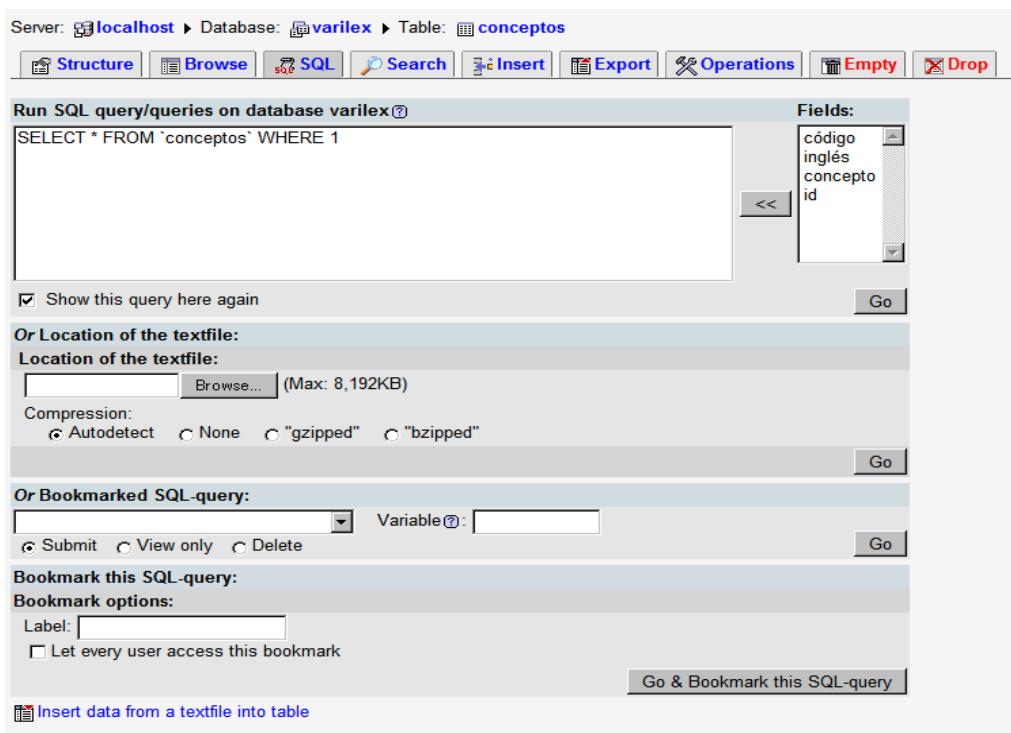


Fig. 3.2 Interfaz tipo phpMyAdmin

Sin embargo, estamos trabajando en la preparación de interfaces tales que no sea necesario utilizar directamente ningún tipo de comando, por lo menos para las funciones más frecuentes, como se puede ver en las Fig. 3.3 y 3.4.

Por ejemplo, en la ventana “Forma” de la Fig. 3.4 se puede escoger “contiene”, “empieza por” y “termina por”. Normalmente es suficiente elegir “contiene”. Es decir, lo que buscamos es una forma léxica que contenga la secuencia de caracteres que nos interesa. Dejaremos “empieza por” y “termina por” para casos muy especiales en que nos interese que la expresión completa empiece o termine por tales caracteres.

Si sabemos de antemano que existe un término determinado en la base de datos podemos escribirlo completo directamente. Por ejemplo, “saco”. De esta manera obtenemos, además de “saco” (en el sentido de JACKET, A001), “saco de traje”, saco (A002, CARDIGAN), etc. y también “labasaco” y “mandar a tomar viento (fresco) por saco” porque contienen la secuencia “saco”.

La secuencia no tiene que ser una palabra completa, podemos hacer búsquedas por una secuencia cualquiera. Por ejemplo, con “manc” encontraremos *mancuernillas*, *mancornas*, etc.

Los espacios se tienen en cuenta en este tipo de búsquedas por secuencia de caracteres. Así, podemos introducir “ito ” (observar el espacio tras la secuencia *ito*), y obtendremos muchos términos que, bien sea al final de algún término compuesto o a mediados contengan tal secuencia. Entre los ejemplos encontrados estarán: “cuernito de mantequilla”, “pancito de viena”, y también otros como

“cámara de circuito cerrado” o “valer un negro con pito y todo” que evidentemente contienen la secuencia “ito”.

Tenemos dos tipos de comodín, que son los que se utilizan regularmente en el lenguaje de MySQL. En primer lugar, podemos usar “_” que equivale a un carácter cualquiera. Así, por ejemplo, “ch_m_a” nos va a dar como resultado *chompa*, *chompa abierta*, y también *chomba*, *chumpa*, *chamba*, *chimba* y hasta *chimpancé*. El otro comodín es “%”, que se puede usar para varios caracteres. Si buscamos “a%z%n” vamos a encontrar (si existe) una expresión donde existe una “a”, después una secuencia indeterminada de caracteres, una “z”, otra secuencia indeterminada de caracteres, una “z” y tal vez más caracteres a la derecha. Si hacemos la prueba, encontraremos términos como *azulón*, *calzones*, *lápiz de carbonillo*, etc.

BÚSQUEDAS (* para "todos")	
Forma	contiene <input type="text"/>
Serie en la Encuesta	*
Categoría	*
Campo semántico	*
Referencia (inglés)	*
Definición según encuesta	contiene <input type="text"/>
<input type="button" value="Volver"/> <input type="button" value="Buscar"/> <input type="button" value="Limpiar"/>	

Fig. 3.3 Interfaz simple de búsqueda

El resultado de las búsquedas a través del interfaz mostrado en la Fig. 3 se muestra en una forma similar al mostrado en la Fig. 4. Este tipo de página tiene la doble función de mostrar el resultado de una búsqueda y al mismo tiempo servir de interfaz simple para otras búsquedas relacionadas. Si hacemos clic en “saco” obtendremos los resultados de las respuestas que contengan esta secuencia de caracteres. Si hacemos clic en A001 obtendremos las respuestas correspondientes a A001. Si hacemos clic en JACKET obtendremos las mismas respuestas que con A001.

Forma	Serie	Categoría	Campo	Ref. (inglés)	Definición en Encuesta	mapa
americana	A001	sus	prenda	JACKET	Prenda de vestir masculina, que forma con el chaleco y los pantalones el traje completo. No es de paño con botones dorados.	
saco	A001	sus	prenda	JACKET	Prenda de vestir masculina, que forma con el chaleco y los pantalones el traje completo. No es de paño con botones dorados.	
vestón	A001	sus	prenda	JACKET	Prenda de vestir masculina, que forma con el chaleco y los pantalones el traje completo. No es de paño con botones dorados.	
chaleco	A001	sus	prenda	JACKET	Prenda de vestir masculina, que forma con el chaleco y los pantalones el traje completo. No es de paño con botones dorados.	
gabán	A001	sus	prenda	JACKET	Prenda de vestir masculina, que forma con el chaleco y los pantalones el traje completo. No es de paño con botones dorados.	
saco de traje	A001	sus	prenda	JACKET	Prenda de vestir masculina, que forma con el chaleco y los pantalones el traje completo. No es de paño con botones dorados.	
caser	A001	sus	prenda	JACKET	Prenda de vestir masculina, que forma con el chaleco y los pantalones el traje completo. No es de paño con botones dorados.	
traje	A001	sus	prenda	JACKET	Prenda de vestir masculina, que forma con el chaleco y los pantalones el traje completo. No es de paño con botones dorados.	
chaqueta de punto	A002	sus	prenda	CARDIGAN	Prenda de lanas, con botones.	
suelter	A002	sus	prenda	CARDIGAN	Prenda de lanas, con botones.	
chaqueta	A002	sus	prenda	CARDIGAN	Prenda de lanas, con botones.	
cárdigan	A002	sus	prenda	CARDIGAN	Prenda de lanas, con botones.	
rebeca	A002	sus	prenda	CARDIGAN	Prenda de lanas, con botones.	
chompe	A002	sus	prenda	CARDIGAN	Prenda de lanas, con botones.	
chompa abierta	A002	sus	prenda	CARDIGAN	Prenda de lanas, con botones.	

Fig. 3.4 Interfaz simple de búsqueda con enlace a cartografía automática

3.4 Cartografía automática

Si pinchamos en el ícono de mapa a la derecha en la Fig. 4 podemos obtener el mapa de distribución de uso de la forma en cuestión en formato PDF A4 y que se puede descargar. En la Fig. 5 se muestra un ejemplo de mapa de distribución del término “saco”. Queremos resaltar que el mapa obtenido no está almacenado en el servidor, sino que se genera dinámicamente en el momento de pinchar el ícono con los datos disponibles en ese momento en la base de datos. De esta manera, podemos ofrecer un sistema de cartografía automática sin necesidad de esperar a que el proyecto haya terminado. Este sistema tiene además la ventaja de que los datos se pueden editar, corregir, añadir, etc. sobre la marcha al contrario que los atlas editados en gruesos volúmenes de papel.



Fig. 3.5 mapa de distribución de “saco”

Obviamente el uso de la tecnología SIG²³ sería la continuación natural de este sistema de cartografía automática en Internet aprovechando los datos obtenidos en las encuestas junto con otros datos de diversa índole: concentraciones urbanas, clima, altura, ingresos, zonas bilingües o multilingües, etc. Sin embargo, teniendo en cuenta las características de Varilex, ya que los puntos de encuestas están separados por grandes distancias generalmente, creemos que este tipo de cartografía simple como la que acabamos de mostrar puede ser de gran utilidad.

Las bases de datos nos ayudan no solamente a ordenarlos sino también a descubrir relaciones que de otra manera pasarían desapercibidas. Creemos que las bases de datos relacionales son de gran utilidad para el lingüista, tanto si trabaja individualmente como si lo hace en equipo.

Las técnicas recientes de lenguajes de programación especialmente concebidos para ser usados en Internet funcionan de la parte del servidor, por lo que no es necesario que el usuario tenga que instalar periódicamente versiones de los programas ni introducir datos de nuevo. Es suficiente que estas operaciones se hagan una sola vez en el servidor y desde ese mismo momento queda a disposición de todos los usuarios que se conecten con navegadores normales.

²³ GIS en inglés: Geographic Information System

4. Final

Han pasado quince años desde que en 1992 empezamos a preparar los estudios del léxico variable del español. Durante este período hemos experimentado desarrollos de las teorías y prácticas de la investigación. Hemos aplicado técnicas de investigación del campo de la tradición europea de la geografía lingüística y técnicas de estadística moderna para observar el complejo mundo de la variación lingüística. Últimamente nos debemos al desarrollo de la informática tanto en los aspectos concernientes a su capacidad como en los de la posibilidad de procesar tal cantidad de datos.

Por otra parte debemos subrayar la importancia de la colaboración. Nuestro equipo cuenta con más de 80 colaboradores locales que anualmente nos envían los cuestionarios llenos. Sin ellos nada hubiera sido posible. Nuestro proyecto no se trata, sin embargo, de formar un grupo cerrado de investigadores. Para ampliar nuestro ámbito de investigación pensamos abrir nuevas páginas web para poder reunir datos léxicos y procesar automáticamente en forma de base de datos con dos interfaces para los usuarios: encuestas en web y base de datos, análisis estadísticos y cartografía automática coordinados en tiempo real con las encuestas. Con este nuevo sistema de investigación, sería posible conocer la realidad de la lengua en la forma más actualizada al servicio de los interesados, profesores, estudiantes, lexicógrafos, lingüistas y especialistas de campos relacionados.

Referencias citadas:

- Aulds, Charles. 2000. *Linux Apache Web Server Administration (Linux Library)*, Sybex.
- Aoto, Seiichi. 2003. “Acerca de la variación de las perifrasis verbales de gerundio”, *Varilex*, 11, pp. 20-28.
- _____. 2004. “Aplicación de LMS en la cartografía electrónica de Varilex”, *Varilex*, 12, pp. 6-19.
- _____. 2005. “Estudio sobre la variación léxica del español en Paraguay”, *Varilex*, 13, pp. 6-27.
- Ávila, Raúl, José Antonio Samper, Hiroto Ueda, Antonio Ruiz Tinoco. 2003. *Pautas y pistas en el análisis del léxico hispano(americano)*. Iberoamericana Vervuert.
- Bowman, Judith S.; Emerson, Sandra L; Darnovsky, Marcy: *The Practical Sql Handbook : Using Structured Query Language*, Addison-Wesley Pub, 1996
- Converse, Tim; Park, Joyce. 2000. *PHP4 Bible*. IDG Books Worldwide.
- Dubois, Paul; Widenius Monty: *MySQL*, New Riders Publishing, 1999
- Equipo de investigación Varilex. 1993. *VARILEX 1. Variación Léxica del Español del Mundo. 1. Mapas y estadísticas*. Proyecto Internacional de Investigación. Español del Mundo.
- _____. 1994d. *VARILEX 2. Variación Léxica del Español del Mundo. 2. Índice de palabras*. Proyecto Internacional de Investigación. Español del Mundo.
- _____. 1995. *VARILEX 3. Variación Léxica del Español del Mundo. 3. Distribución de palabras*, A. Proyecto Internacional de Investigación. Español del Mundo..
- _____. 1996a. *VARILEX 4. Variación Léxica del Español del Mundo. 4. Distribución de palabras*, B. Proyecto Internacional de Investigación. Español del Mundo.
- _____. 1997d. *VARILEX 5. Variación Léxica del Español del Mundo. 5. Distribución de palabras*, C. Proyecto Internacional de Investigación. Español

del Mundo.

- _____. 1998e. VARILEX 6. *Variación Léxica del Español del Mundo. 6. Distribución de palabras*, D. Proyecto Internacional de Investigación. Español del Mundo.
- _____. 1999a. VARILEX 7. *Variación Léxica del Español del Mundo. Distribución de palabras*, E. Proyecto Internacional de Investigación. Español del Mundo.
- _____. 2000i. VARILEX 8. *Variación Léxica del Español del Mundo. Distribución de palabras*, F. Proyecto Internacional de Investigación. Español del Mundo.
- _____. 2001e. VARILEX 9. Variación Léxica del Español del Mundo. Distribución de palabras, B, C, G. Proyecto Internacional de Investigación. Español del Mundo.
- _____. 2002d. VARILEX 10. *Variación Léxica del Español del Mundo. Distribución de palabras*, H. Proyecto Internacional de Investigación. Español del Mundo.
- _____. 2003d. VARILEX 11. *Variación Léxica del Español del Mundo. Distribución de palabras*, I. Proyecto Internacional de Investigación. Español del Mundo.
- _____. 2004d. VARILEX 12. *Variación Léxica del Español del Mundo. Distribución de palabras*, I. Proyecto Internacional de Investigación. Español del Mundo.
- _____. 2005. VARILEX 13. *Variación Léxica del Español del Mundo. Distribución de palabras*, K. Proyecto Internacional de Investigación. Español del Mundo.
- _____. 2006. VARILEX 14. *Variación Léxica del Español del Mundo. Distribución de palabras*, F. Proyecto Internacional de Investigación. Español del Mundo.

Gerken, Till and Rastchiller, Tobias. 2000. *Web Application Development with PHP*. New Riders.

- Fernández, Francisco. 1994. "Comentarios", Varilex, 2, pp.41-53.
- _____. 1995. "Comentarios críticos sobre Varilex 2", Varilex, 3, pp.19-29.
- _____. 1996. "Comentarios críticos sobre Varilex 3", Varilex, 4, pp.9-25.
- _____. 1997. "Comentarios críticos sobre Varilex 4", Varilex, 5, pp.28-40.
- _____. 1998. "Comentarios críticos sobre Varilex 5", Varilex, 6, pp.19-36.
- _____. 1999. "Comentarios críticos sobre Varilex 6", Varilex, 7, pp.9-42.
- _____. 2000. "Comentarios críticos sobre Varilex 7", Varilex, 8, pp.10-38.
- _____. 2001. "Comentarios críticos sobre Varilex 8", Varilex, 9, pp.35-76.
- _____. 2003. "Comentarios críticos sobre Varilex 10", Varilex, 11, pp.29-38.
- _____. 2004. "Comentarios críticos sobre Varilex 11", Varilex, 12, pp.20-35.
- Koike, Kazumi. 2001. "Variación fraseológica del español", Varilex, 9, pp. 77-92.
- Meloni, Julie C. 2000. *PHP Fast & Easy Web Development*, Prima Publishing.
- Nakajima, Sayaka. 1999. "Comentarios críticos sobre Varilex 6", Varilex, 7, pp.43-45.
- _____. 2000. "Comentarios críticos sobre Varilex 7", Varilex, 8, pp.39-47.
- Nishimura, Kimiyo. 2002. *Diminutivos españoles. Morfología, funciones semánticas y análisis diatópicos y diastráticos*. Tesis doctoral, en japonés (Universidad de Estudios Extranjeros de Osaka).
- Perelló, Juan Luis. 2002. "Comentario de los resultados para Chile", Varilex 10, pp. 36-52.
- Rubio, Carlos y Hiroto Ueda (eds.) 2004. *Nuevo diccionario español-japonés*, Tokyo, Kenkyusha.
- _____. 1992. *Diccionario Crown japonés-español*, Tokyo, Sanseido
- Ruiz Tinoco, Antonio. 1999. "El Proyecto VARILEX en Internet. Base de datos compartida de variación léxica", Varilex, 7, pp. 50-60.
- _____. 2000. "Manual de lenguaje SQL aplicado a la base de datos de Varilex",

- Varilex*, 8, pp. 48-70.
- _____. 2001. "Cartografía automática en Internet", *Bulletin of the Faculty of Foreign Studies*, 36, Sophia University, pp. 77-93.
- _____. 2004. "Léxico variable en los países hispanohablantes según el sexo", *Diccionario, léxico y cultura* (eds. Josefina Prado Aragónés y María Victoria Galloso Camacho), Universidad de Huelva, pp. 189-198.
- _____. 2005 "Nuevo interfaz del Proyecto PHP-Atlas de cartografía automática", *Varilex*, 13, pp. 28-30.
- _____. 2006 "Atlas lingüísticos automáticos y bases de datos dialectales", Jornadas sobre Métodos Informáticos en el Tratamiento de las Lenguas Ibéricas, 2006/7/1-2, Centro de Estudios Hispánicos, Universidad Sofía, Tokio.
- Segura, Joaquín. 1999. "Breve comentario", *Varilex*, 7, pp. 46-47.
- Takagaki, Toshihiro. 1993. "Hacia la descripción del español contemporáneo de las grandes ciudades del mundo hispánico", *Lingüística Hispánica*, (Círculo Lingüístico de Kansai, Japón), 16, pp.65-86.
- _____. (ed.) 2003. *Diccionario de bolsillo español-japonés, japonés-español*, Tokyo, Shogakukan.
- _____, Masami Miyamoto, Noritaka Fukushima, Hiroto Ueda y Antonio Ruiz Tinoco 2004e. *Encuesta sobre Problemas Sintácticos de la Lengua Española*.
- Ueda, Hiroto. 1994. "Banco de datos léxicos del español: un proyecto internacional de investigación", *Verba, Anuario Galego de Filoloxía*, 21, pp.379-416.
- _____. 1995a. "Variación fonológica en el léxico urbano español", *Lingüística Hispánica*, 18, pp.105-125.
- _____. 1995b. "Zonificación del español. Palabras y cosas de la vida urbana", *Lingüística*, 7, pp.43-86.
- _____. 1996a. "Variación léxica del español urbano. Distribución: Vestuario y equipo", *Revista del Departamento de Lenguas Extranjeras. Facultad de Artes y Ciencias de la Universidad de Tokio*, 43/4, pp.99-144.
- _____. 1996b. "Estudio de la variación léxica del español. Métodos de investigación", *Homenaje al profesor Makoto Hara*, Universidad de Estudios Extranjeros de Tokio, pp.341-376.
- _____. 1996c. "Análisis demolingüístico del léxico variable español", *Lingüística Hispánica*, vol. 19, pp.63-98.
- _____. 1999. "Dialectología urbana en equipo y utilización de Internet", José Antonio Samper Padilla y Magolia Troya Déniz (eds.), *Actas del XI Congreso Internacional de la Asociación de Lingüística y Filología de la América Latina*, pp.1299-1310.
- _____. 2000. "Distribución de las palabras variables en España y en América. Léxico del transporte", *Estudios de lingüística hispánica. Homenaje a María Vaquero*, Universidad de Puerto Rico, pp.637-655.
- _____. 2003a. "Estudio de la variación léxica del español en el mundo. El método intensivo y el extensivo", *Lengua, variación y contexto. Estudios dedicados a Humberto López Morales*, pp.895-905.
- _____. 2003b. "Tipos del léxico y enseñanza del español como segundo idioma", *Cuadernos Hispanoamericanos*, vol. 631. *Homenaje al Prof. Manuel Alvar*, pp. 27-33.
- _____. 2005. "Léxico de la blasfemia: Análisis por patronización", Josefina Prado Aragónés y María Victoria Galloso Camacho (eds.) *Diccionario, Léxico y*

- Cultura* (Universidad de Huelva, España), pp. 233-245.
- _____. 2006. "Patternized configuration of geolinguistic data: Application to Spanish lexical variation", *Topics in Dialectal Variation*, Markku Filppula, Juhani Klemola, Marjatta Palander, Esa Penttila (eds.) Joensu University (Finland), pp. 163-187.
- _____. (en prensa) "Zonificación múltiple de las ciudades hispanohablantes según el léxico urbano moderno - Análisis cluster y análisis de componentes principales", *Jornadas sobre Métodos Informáticos en el Tratamiento de las Lenguas Ibéricas*, 2006/7/1-2, Centro de Estudios Hispánicos, Universidad Sofía, Tokio, 147p.
- _____. (en prensa) "Un estudio comparativo del léxico de Guinea Ecuatorial: Resultado de encuestas y clasificación", *Congreso Internacional de Hispanistas en Asia*, Centro Cultural Español de Malabo.
- _____ y Carlos Rubio (eds.), 2006d. *Puerta al español. Diccionario español-japonés*. Tokyo, Kenkyuusha.
- Yarger, Randy Jay; Reese, George & King, Tim: *MySQL & mSQL*, O'Reilly & Associates, 1999.

Fig. 3. Antes de patronizar

Fig. 4. Despu s de patronizar