

## **LOS INFORMANTES EN LAS ENCUESTAS DE VARIACIÓN LÉXICA**

Antonio Ruiz Tinoco

Universidad Sophia (Tokio)

a-ruiz@sophia.ac.jp

### **Abstract**

This research analyzes some of the characteristics of the informants of the Varilex Project on Spanish lexical variation. Our informants are middle class, male and female, young and old, and live in urban Spanish-speaking areas of the world. A database which is accessible by Internet was created so data can be easily analyzed and shared. Our data shows that there is not a major quantitative difference in the total variational lexicon used by males and females, both young and old. However, we found some differences in certain semantic fields. We have also found that, regardless of sex, age, or city, more than half of the informants gave us 1.2 answers as average for every question. The number of answers for the other informants increased gradually, reaching in some cases, more than 5 as average. This phenomenon can influence the methodology of the questionnaires and must be more thoroughly analyzed.

**Key words:** Spanish, lexical variation, informant, database, questionnaire

### **1. La elección de informantes**

La correcta elección de los informantes en los trabajos de campo de dialectología es uno de los principales factores a tener en cuenta para obtener datos que nos lleven a conclusiones aceptables. Los informantes que Chambers, J. K. & P. J. Trudgill (1980) llamaron NORM

(Non-mobile, Older, Rural, Males) han sido desde Jules Gilliéron, que también seleccionaba hombres mayores, de bajo nivel de educación y que vivieran en comunidades rurales, los informantes tradicionales en los estudios de dialectología. No se puede negar el valor de este tipo de informantes, como alega Orton (1962) cuando se trata de reconstruir formas antiguas, pero si nuestro objetivo es el estudio de la variación léxica preferimos que nuestros informantes vivan en ciudades, sean hombres y mujeres, mayores y jóvenes. Con frecuencia estos informantes suelen viajar a otras regiones y países.

Hay casos como el del quechua hablado en extensas zonas andinas, en los que si el objetivo es la búsqueda de las formas más antiguas, suele ser preferible que los informantes sean mujeres, ya que los hombres suelen viajar más, van a la escuela y tienen mayor contacto con el español. Por otra parte, según Eckert (2000) sobre el inglés de Detroit, las variaciones más marcadas suelen darse entre mujeres jóvenes. Así que no siempre el informante tradicional no es el más adecuado para todos los trabajos de campo. En Ruiz Tinoco (2005) hemos indicado algunas observaciones sobre las diferencias entre hombres y mujeres en la variación léxica.

En este estudio, nuestro objetivo es utilizar los datos del Proyecto VARILEX<sup>1</sup> de variación léxica, para observar y analizar algunas características de los informantes que han colaborado con las encuestas periódicas del proyecto de variación léxica.

## 2. Características del proyecto VARILEX

El proyecto VARILEX se creó para estudiar la variación léxica en las principales

---

<sup>1</sup> Además de la revista VARILEX, publicada anualmente, se puede consultar Avila, Raúl; Samper, José Antonio; Ueda, Hiroto *et al.* (2003), y la página oficial del proyecto en <http://gamp.c.u-tokyo.ac.jp/~ueda/VARILEX>.

ciudades de todo el mundo hispanohablante. Está coordinado por Hiroto Ueda de la Universidad de Tokio, Toshihiro Takagaki, de la Universidad de Estudios Extranjeros de Tokio y Antonio Ruiz Tinoco de la Universidad Sophia de Tokio. Comenzó el año 1993 en el X Congreso de ALFAL en Veracruz y se puede consultar la base de datos experimental en Internet desde 1999.<sup>2</sup>

Se realiza mediante encuestas enviadas por correo a los investigadores colaboradores locales, que deben escoger como mínimo cuatro informantes locales. La selección de los informantes se hace teniendo en cuenta que pertenezcan al estrato culto y a la clase socioeconómica alta o media y con residencia permanente en la ciudad. Deben ser dos mujeres y dos hombres, de más de 18 años hasta los 39, y desde los 40 hasta los 80. Los informantes no tienen un límite de tiempo para contestar a las encuestas. Una vez terminadas las encuestas, se envían a Tokio, donde se procesan los datos. También se publica anualmente la revista VARILEX, que contiene los resultados de tales encuestas así como algunos estudios y noticias.

Por medio de estas encuestas queremos estudiar el léxico variable que corresponde a cada ámbito conceptual. Por ejemplo, en el caso del concepto A188, “PEN”, usamos la definición “utensilio para escribir con tinta”. De esta manera, queremos incluir objetos que pueden ser incluso físicamente diferentes como son *estilógrafo, estilográfica, lápiz tinta, lapicera, lapicera de tinta, lapicera fuente, lapicero, lapicero de tinta líquida, lapicero de tinta mojada, pluma, pluma de fuente, pluma estilográfica, pluma fuente y plumilla*. En la figura 1 se muestra la parte de la encuesta correspondiente a este concepto.

Los *ámbitos conceptuales* que hemos estudiado se refieren principalmente a la vida normal urbana moderna con más posibilidades de variación, como pueden ser las prendas de

---

<sup>2</sup> <http://lingua.cc.sophia.ac.jp/varilex/php-atlas/lista3.php>

vestir, los instrumentos usados en la cocina, los muebles, las partes del automóvil, material de escritorio, etc.



**[A188] PEN: Utensilio para escribir con tinta.**

(1)estilógrafo, (2)estilográfica, (3)lápiz tinta, (4)lapicera,  
(5)lapicera de tinta, (6)lapicera fuente, (7)lapicero, (8)lapicero de  
tinta líquida, (9)lapicero de tinta mojada, (10)pluma, (11)pluma de  
fuente, (12)pluma estilográfica, (13)pluma fuente, (14)plumilla.

&Otro(s) \_\_\_\_\_, #No se me ocurre.

\$Comentario:

Figura 1. Cuestionario (parcial)

La estructura de la base de datos tiene cuatro tablas principales tal como se muestra en la figura 2.

La tabla *conceptos* contiene el campo *código* para ordenar los conceptos utilizados. Las encuestas están ordenadas por letras seguidas de un número. Así, por ejemplo, A001 se refiere al primer concepto estudiado en la primera encuesta realizada; *inglés*, como indica su propio nombre se utiliza para referirnos al ámbito conceptual de forma neutra; y *concepto*, que es la descripción verbal del ámbito conceptual en cuestión.

La tabla *respuestas* contiene los campos: *id*, número de identificación del dato, *encuesta*, código que se corresponde con *código* de la tabla *concepto*; *informante*, código del informante que relaciona esta tabla con la de *informantes* y *forma*, que contiene la respuesta dada por el informante.

La tabla *informantes* contiene los campos *código*, para referencia del informante, que coincide con el campo *informante* de la tabla de *respuestas*; *codciu*, código de la ciudad; *país*, *ciudad*, *sexo*, sexo del informante, 1 para hombre y 2 para mujer; *edad*; *tipo*, tipo de ocupación del informante, y *ocupación*. Por ejemplo, en *país*, AR se utiliza para Argentina,

BO para Bolivia, CH para Chile, etc.

La tabla *ciudades* contiene los campos *id*, que relaciona esta tabla con la de *informantes*, *CIUDAD*, que aunque no es necesario por estar incluida en *informantes* nos resulta más cómodo, y finalmente los campos *x* e *y*, que contienen la posición de la ciudad dentro del mapa blanco que usamos de base para la cartografía automática (Ruiz Tinoco, A. 2001).

Los datos de las encuestas se procesan en una base de datos en un servidor experimental de entorno LAMP.<sup>3</sup>

En este estudio utilizaremos parcialmente los datos de la primera<sup>4</sup> encuesta de VARILEX, serie A, en la que se estudiaron 152 ámbitos conceptuales, ordenados de A001 a A219.<sup>5</sup> Por ejemplo, se tratan los campos semánticos relacionados con los objetos que se encuentran normalmente en la casa, el vocabulario relacionado con el deporte, la música, la naturaleza, las prendas de vestir, la comida, etc. Una de las ventajas de usar una base de datos estándar es que podemos usar el lenguaje SQL<sup>6</sup> de bases de datos relacionales, que resulta de gran utilidad para nuestro objetivo, como iremos viendo. Los datos se pueden obtener en las páginas web del proyecto y en algunos casos indicaremos la fórmula utilizada. Para los detalles de uso de la base de datos del proyecto, ver Ruiz Tinoco, A. (2000).

---

<sup>3</sup> LAMP: Linux como sistema operativo; Apache como servidor web; MySQL para administrar la base de datos y PHP como lenguaje de programación.

<sup>4</sup> Equipo VARILEX (Coord. Ueda, H.; Takagaki, T.; Ruiz Tinoco, A.). *VARILEX, Variación léxica del español en el mundo*. Vols. 1-9.

<sup>5</sup> Para ver la lista completa de conceptos usar el comando siguiente:

```
SELECT * FROM conceptos WHERE código LIKE "A%";
```

<sup>6</sup> El lenguaje SQL para MySQL se puede consultar en <http://www.mysql.com> o en cualquiera de los muchos manuales existentes. Aconsejamos, por su claridad Dubois, Paul; Widenius Monty (1999).

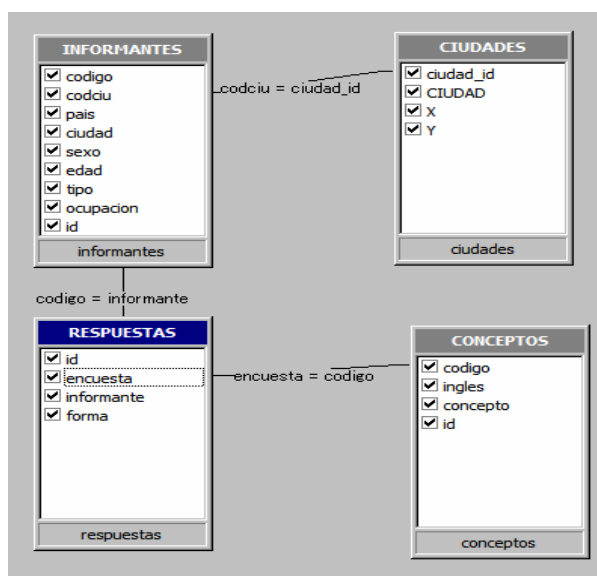


Figura 2. Estructura básica de la base de datos de VARILEX

En esta primera encuesta, realizada en 30 ciudades de 19 países, han colaborado 177 informantes con profesiones como psicólogos, amas de casa, docentes, abogados, etc. En general usamos el comando SELECT de SQL, y en este caso la siguiente fórmula<sup>7</sup> para ver la distribución de sexo y edad en los diferentes países.

```
SELECT país,
SUM( IF ( sexo = '1', 1, 0 ) ) AS hombres,
SUM( IF ( sexo = '2', 1, 0 ) ) AS mujeres,
SUM( IF ( edad > '39', 1, 0 ) ) AS mayores,
SUM( IF ( edad < '40', 1, 0 ) ) AS menores,
COUNT( * ) AS total
FROM informantes
GROUP BY país
ORDER BY total DESC;
```

El resultado se puede observar en la tabla 1. En España, ES, México, MX, y Argentina, AR, hay más informantes debido a que se han encuestado también más ciudades. En Estados

<sup>7</sup> Aunque el lenguaje SQL se utiliza frecuentemente en las bases de datos, echamos en falta la existencia de software especializado para el tratamiento de los datos de variación léxica o de dialectología en general.

Unidos, EU, sin embargo, los informantes son 4 hombres y 4 mujeres.

país	hombres	mujeres	mayores de 40	menores de 40	total
ES	28	28	29	27	56
MX	8	8	8	8	16
AR	8	8	8	8	16
CH	5	6	5	6	11
EU <sup>8</sup>	4	4	4	4	8
CU	4	4	4	4	8
VE	4	4	4	4	8
NI	4	4	4	4	8
PR <sup>9</sup>	4	4	4	4	8
UR	2	2	2	2	4
CO	2	2	2	2	4
EL	2	2	2	2	4
PE	2	2	2	2	4
PN <sup>10</sup>	2	2	2	2	4
RD	2	2	2	2	4
BO	2	2	2	2	4
CR	2	2	2	2	4
PA <sup>11</sup>	1	2	1	2	3
HO	2	1	1	2	3

Tabla 1. Distribución de los informantes por países, sexo y edad

### 3. Distribución general de las respuestas

En las secciones siguientes vamos a mencionar algunas variaciones generales que

<sup>8</sup> Estados Unidos. La primera encuesta en este país se realizó en Nueva Orleans.

<sup>9</sup> Puerto Rico.

<sup>10</sup> Panamá.

<sup>11</sup> Paraguay.

hemos observado en relación con la distribución del número de respuestas entre hombres y mujeres,<sup>12</sup> informantes mayores y menores de 40 años, y algunas diferencias en cuanto a la elección del léxico.

Aquí vamos a señalar el número de respuestas dado según el sexo y el país. Las cifras totales se pueden ver en las tablas 2 y 3. Aparentemente se dan algunas diferencias según el país. Por ejemplo, en España, Costa Rica, El Salvador, Estados Unidos, Perú, Panamá, Puerto Rico y Uruguay, el número de respuestas de los hombres es superior al de las mujeres. Sin embargo, en Argentina, Bolivia Cuba, México, República Dominicana las cifras son superiores para las mujeres. En los otros países encuestados no se aprecian diferencias claras.

	Hombres	Mujeres
ES	6.177	5.612
CR	401	301
EL	372	336
EU	1.400	801
PE	541	407
PN	453	368
PR	820	737
UR	427	368

Tabla 2

	Hombres	Mujeres
AR	1.363	1.846
BO	304	363
CU	842	1.090
MX	1.680	1.735
RD	304	406

Tabla 3

Debido a que la muestra por país es bastante pequeña, no creemos que estas diferencias se puedan interpretar en el sentido de que ciertos grupos tienen una disponibilidad léxica mayor que otros. Si observamos las tablas 4 y 5 de respuestas totales veremos que al ser la muestra mayor (88 hombres y 89 mujeres, 89 mayores de 40 años y 88 menos de 40), las cifras son más similares. Resulta interesante observar estas diferencias ya que nos da una idea

<sup>12</sup> Los datos de las diferencias por sexo utilizados en esta sección están tomados parcialmente de Ruiz Tinoco, A. (2005).



de las posibles variaciones con las que nos podemos encontrar entre los miembros de algunos grupos relativamente homogéneos. Además, resulta curioso que prácticamente no haya diferencias significativas en el número de respuestas de estos cuatro grupos. Sin embargo, como veremos más adelante, podemos ver algunas diferencias en la distribución del léxico según los campos semánticos.

Sexo	Total
Hombres (88)	18.620
Mujeres (89)	18.069

Tabla 4 Total de respuestas  
(hombres y mujeres)

Edad	Total
Mayores de 40 (89)	18.705
Menores de 40 (88)	17.984

Tabla 5 Total de respuestas  
(mayores y jóvenes)

### 3.1. Distribución de las respuestas según el sexo

Como muestra de la diferencia de distribución en la elección del léxico variable según el sexo, a continuación mostramos el léxico relacionado con el automóvil, correspondiente a los conceptos del A142 al A160. En primer lugar, obtenemos las frecuencias del léxico ordenado en sentido descendente.<sup>13</sup> En la tabla 5 se muestran solamente los términos de frecuencia total más alta para observar posibles diferencias.

El léxico variable más frecuente relacionado con el automóvil entre hombres en comparación con el de las mujeres, según la tabla 5, es *parabrisas, faros delanteros, faros,*

---

<sup>13</sup> SELECT respuesta,  
SUM( IF ( sexo = '1', 1, 0 ) ) AS hombres,  
SUM( IF ( sexo = '2', 1, 0 ) ) AS mujeres,  
COUNT( \* ) AS total  
FROM informantes, contestaciones  
WHERE contestaciones.informante = informantes.código  
AND orden BETWEEN "A142" AND "A160"  
GROUP BY respuesta  
ORDER BY total DESC.

*portaequipaje(s)*, *direcciona*, etc. En cambio, hay pocos vocablos en los que la frecuencia sea mayor entre las mujeres. Algunos ejemplos son *casa rodante*, *parrilla*, *luces o furgoneta*. La respuesta en la que se da la mayor diferencia en este campo léxico es NS (no sabe), mayor en las mujeres.

Podemos observar algunas diferencias llamativas. Por ejemplo, en el concepto A146, los hombres prefieren *parabrisas* frente a las mujeres (57/42), mientras que en la forma singular *parabrisa* no se advierte tal diferencia (25/26). Asimismo, en el concepto A145, los hombres prefieren *faros* (34/26) o *faros delanteros* (28/21). Resulta difícil de explicar, solamente con los datos disponibles, las diferencias entre, por ejemplo, Argentina y España. Las mujeres argentinas han ofrecido 193 términos frente a los 146 de los hombres. En España, al contrario, las mujeres 549 y los hombres, 612. Entre los países en los que las mujeres han respondido con más términos de este campo léxico también se encuentra Cuba (117/89), pero en el resto, o es pequeña la diferencia o son superadas por los hombres, y en algunos casos por un amplio margen, como es el caso de Estados Unidos (143/95), Perú (61/44) o Panamá (50/39).<sup>14</sup> Más adelante comentaremos sobre algunas características generales de los informantes como otra de las posibles causas de estas diferencias.

respuesta	hombres	mujeres	total
luces de freno	69	69	138
volante	66	62	128
capó	57	56	113
parabrisas	57	42	99
parachoques	39	40	79
placa	42	35	77

<sup>14</sup> Las diferencias de las cifras absolutas entre países se deben al diferente número de informantes en cada país por lo que lo que importa es la diferencia relativa dentro de cada grupo.

luces traseras	41	34	75
furgoneta	36	39	75
parrilla	33	39	72
matrícula	34	35	69
maletero	37	30	67
camioneta	35	29	64
guardabarro	31	32	63
casa rodante	28	34	62
faros	34	26	60
intermitente	31	29	60
Baca	29	30	59
NS	17	36	53
parabrisa	25	26	51
faros delanteros	28	21	49

Tabla 5. Frecuencias del léxico del automóvil (por sexo)

### 3.2. Frecuencia de léxico variable

En el Proyecto VARILEX se pregunta a los informantes si utilizan determinados vocablos que se les presentan explícitamente en las encuestas y al mismo tiempo se les pide que nos informen de otros que ellos mismos utilicen. Gracias a que los datos se van incluyendo en la base de datos relacional, podemos elaborar fácilmente todo tipo de listas de frecuencias, incluyendo condiciones de sexo, edad o cualquiera de los atributos de los informantes. Si hacemos la lista de vocablos<sup>15</sup> *más elegidos*, es decir, más conocidos por los

---

<sup>15</sup> Para la serie A, hombres, utilizaremos la siguiente fórmula:  
 SELECT respuesta,  
 SUM( IF ( sexo = '1', 1, 0 ) ) AS hombres  
 FROM informantes, contestaciones

informantes, con la condición del sexo solamente, podremos observar y comparar las preferencias léxicos de cada uno de los sexos. El resultado de los vocablos más frecuentes se puede ver en las tablas 6 y 7.

Las respuestas más frecuentes, NS (no sabe) y *coche* son las más frecuentes. En algunos vocablos podemos observar algunas diferencias, como *saco*, al contrario, ligeramente más frecuente entre hombres y *chaqueta* entre mujeres, entre las que *saco* tiene una frecuencia menor. También hay algunos vocablos entre los 20 más frecuentes que no aparecen en la lista del sexo opuesto, como es *calzoncillo*, *calzoncillos*, *volante* y *silenciador* entre hombres, y *pluma*, *basurero*, *encendedor* y *papel higiénico* entre las mujeres. Análogamente se pueden preparar fácilmente listas de frecuencia introduciendo las variables de país, ciudad, edad, etc.

### 3.3 Distribución de las respuestas según la edad

De forma análoga a como hemos procedido con los datos de hombres y mujeres podemos buscar diferencias según las edades de los informantes. Para todos los términos de la primera encuesta (serie A) obtenemos las respuestas más frecuentes para mayores y menores de 40 años que se muestran en las tablas 8 y 9.

Aunque no son demasiado llamativas, observamos algunas diferencias de frecuencia en los términos *llave* (93/78), *abrigo* (85/69), *bote* (84/70), *pluma* (80/66), etc. Las formas *bote* y *pluma* no aparecen entre los 20 vocablos más frecuentes para los menores de 40.

Para los vocablos relacionados con el automóvil, obtenemos las frecuencias mostradas en la tabla 10, donde se observan pequeñas diferencias de frecuencia. Por ejemplo,

---

```
WHERE contestaciones.informante = informantes.código
AND orden LIKE "A%"
GROUP BY respuesta
ORDER BY hombres DESC
```

los términos *volante, parabrisas, placa, furgoneta*, etc. son ligeramente más frecuentes entre los jóvenes que entre los mayores, y al contrario, *luces de freno, camioneta* o la opción NS (no sabe) son más frecuentes entre los mayores.

<b>respuesta</b>	<b>hombres</b>
NS	186
coche	127
ducha	107
saco	96
chaqueta	94
llave	90
suéter	85
banco	84
mochila	84
recepción	84
maleta	83
calzoncillo	82
armónica	81
abrigo	79
piscina	77
bote	76
ascensor	76
volante	75
calzoncillos	74
silenciador	74
<b>respuesta</b>	<b>mujeres</b>

Tabla 6. Léxico variable frecuente (hombres)

NS	203
coche	126
chaqueta	114
ducha	109
recepción	88
saco	87
banco	85
mochila	84
suéter	81
llave	81
piscina	80
maleta	80
bote	78
ascensor	78
pluma	77
abrigo	75
armónica	74
basurero	73
encendedor	73
papel higiénico	72

Tabla 7 Léxico variable frecuente (mujeres)

respuesta	mayores
NS	244
coche	122
ducha	110
chaqueta	109
llave	93
saco	91
recepción	87
abrigo	85
mochila	84
bote	84
maleta	83
banco	83
pluma	80
suéter	78
piscina	78
armónica	78
ascensor	77
basurero	75
calzoncillos	74
luces de freno	73

Tabla 8. Léxico variable frecuente (mayores)

respuesta	menores
NS	145
coche	131
ducha	106
chaqueta	99
saco	92
suéter	88
banco	86
recepción	85
mochila	84
calzoncillo	83
maleta	80
piscina	79
llave	78
ascensor	77
armónica	77
maletín	75
volante	75
poncho	74
biberón	73
encendedor	72

Tabla 9. Léxico variable frecuente (menores)

Si observamos los distintos geosinónimos para el concepto parabrisas (A146, cristal de la parte anterior del automóvil que defiende del viento a los pasajeros) obtenemos las frecuencias mostradas en la tabla 11. Aquí podemos apreciar que los tres primeros términos

más frecuentes, *parabrisas*, *parabrisa* y *vidrio delantero* son menos frecuentes entre los menores de 40 años.

<b>respuesta</b>	<b>mayores</b>	<b>menores</b>	<b>total</b>
luces de freno	73	65	138
volante	60	68	128
capó	54	59	113
parabrisas	45	54	99
parachoques	36	43	79
placa	34	43	77
luces traseras	34	41	75
furgoneta	32	43	75
parrilla	34	38	72
matrícula	34	35	69
maletero	31	36	67
camioneta	38	26	64
guardabarro	32	31	63
casa rodante	30	32	62
faros	29	31	60
intermitente	30	30	60
Baca	28	31	59
NS	34	19	53
parabrisa	24	27	51

Tabla 10. Frecuencias del léxico del automóvil (por edad)

Creemos que el vocabulario del automóvil ha ido variando poco a poco con el avance de la tecnología, por lo que los jóvenes no solamente conocen mejor el vocabulario sino que además lo renuevan. Otro caso de diferencia de frecuencias es el de los objetos de escritorio.

Por ejemplo, en la tabla 12 se muestran los datos sobre la *pluma estilográfica* (A188, utensilio para escribir con tinta).

<b>respuesta</b>	<b>mayores</b>	<b>menores</b>	<b>total</b>
parabrisas	45	54	99
parabrisa	24	27	51
vidrio delantero	4	16	20
crystal delantero	11	9	20
luna delantera	8	6	14
crystal	5	6	11
luna	5	2	7
vidrio	3	3	6
windshield	3	3	6
luneta	2	2	4
crystal de frente	3	1	4
NS	2	0	2
lunas delanteras	0	1	1
parabris	0	1	1
panorámico	0	1	1
crystal del presente	1	0	1

Tabla 11. Frecuencias de *parabrisas* para mayores y menores de 40 años



respuesta	mayores	menores	total
pluma	47	54	101
pluma fuente	18	21	39
pluma estilográfica	14	7	21
estilográfica	15	5	20
lapicera	8	9	17
lapicera fuente	9	6	15
pluma de fuente	10	4	14
lapicero	4	3	7
lapicera de tinta	2	4	6
estilógrafo	4	0	4

Tabla 12. Frecuencias de *pluma estilográfica* para mayores y menores de 40 años

Observamos que los términos *estilográfica*, *pluma de fuente* y *pluma estilográfica* son menos frecuentes entre los jóvenes que *pluma*, *pluma fuente*.

#### 4. Matriz bidimensional

Podemos obtener fácilmente la matriz bidimensional de cada uno de los ámbitos conceptuales para un microanálisis más minucioso. Como muestra, podemos obtener<sup>16</sup> la

<sup>16</sup> Ejemplo de obtención de matriz bidimensional. Para añadir puntos, ampliar la fórmula siguiente:

```
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,
SUM( IF ( ciudad = 'BAR', 1, 0 ) ) AS Barcelona,
COUNT( * ) AS total
FROM contestaciones
WHERE orden = "A016"
AND ciudad IN ( 'BUE', 'MEX', 'MAD', 'BAR' )
GROUP BY respuesta
ORDER BY total DESC;
```

matriz bidimensional de *pajarita*, (“pieza de adorno usada por el hombre, con dos extremos cruzados”) en la tabla 13, donde podemos observar las preferencias léxicas de las distintas ciudades.

respuesta	Buenos Aires	México	Madrid	Barcelona	total
pajarita	0	0	4	4	8
corbata de moño	0	6	0	0	6
moñito	3	0	0	0	3
moño	1	1	0	0	2
corbatín	0	0	0	1	1
corbata de lacito	0	1	0	0	1

Tabla 13. Matriz bidimensional

Si queremos añadir datos de sexo (M para mujeres y H para hombres), etc. de los informantes, la matriz resultante<sup>17</sup> será análoga a la tabla 14. Sin embargo, debido al pequeño tamaño de la muestra de informantes no podemos aventurar las preferencias hacia una forma determinada entre ninguno de los grupos de sexo o de edad. Evidentemente, también se pueden elaborar matrices conteniendo los atributos de edad.

<sup>17</sup> La fórmula será análoga a la siguiente:

```
SELECT  respuesta,
SUM( IF ( contestaciones.ciudad = 'BUE', 1, 0 ) AND IF ( sexo = '1', 1, 0 ) ) AS 'Baires H',
SUM( IF ( contestaciones.ciudad = 'BUE', 1, 0 ) AND IF ( sexo = '2', 1, 0 ) ) AS 'Baires M',
SUM( IF ( contestaciones.ciudad = 'MEX', 1, 0 ) AND IF ( sexo = '1', 1, 0 ) ) AS 'México H',
SUM( IF ( contestaciones.ciudad = 'MEX', 1, 0 ) AND IF ( sexo = '2', 1, 0 ) ) AS 'México M',
SUM( IF ( contestaciones.ciudad = 'MAD', 1, 0 ) AND IF ( sexo = '1', 1, 0 ) ) AS 'Madrid H',
SUM( IF ( contestaciones.ciudad = 'MAD', 1, 0 ) AND IF ( sexo = '2', 1, 0 ) ) AS 'Madrid M',
COUNT( * ) AS total
FROM    contestaciones, informantes
WHERE   orden = "A016"
AND     contestaciones.informante = informantes.código
AND     contestaciones.ciudad IN ( 'BUE', 'MEX', 'MAD' )
GROUP  BY respuesta
ORDER  BY total DESC;
```

<b>respuesta</b>	<b>Baires H</b>	<b>Baires M</b>	<b>México H</b>	<b>México M</b>	<b>Madrid H</b>	<b>Madrid M</b>	<b>total</b>
pajarita	0	0	0	0	2	2	8
corbata de moño	0	0	3	3	0	0	6
moñito	1	2	0	0	0	0	3
moño	1	0	1	0	0	0	2
corbata de lacito	0	0	0	1	0	0	1

Tabla 14. Matriz bidimensional con atención al sexo de los informantes

### **5. Distribución de los informantes según el número de respuestas**

Uno de los principales objetivos de las encuestas de variación léxica es el inventariado de geosinónimos. Ni que decir tiene que nos interesa conocer con la mayor aproximación posible los geosinónimos coexistentes en un lugar dado. Esa es una de las razones por las que pedimos a los informantes que añadan otros términos que no están incluidos en la misma encuesta. Como hemos dicho anteriormente, los informantes no tienen límite de tiempo para responder. Sin embargo, observamos un fenómeno curioso que creemos merece la pena observar ya que afecta a la cantidad de datos recogidos.

Según los datos de la tabla 15 y que se reflejan en la fig. 3 se puede observar que bastante más del 50% de los informantes da un 20% más de respuestas que las preguntas del cuestionario. Es decir, para las 152 preguntas de la serie A, más del 50% de los informantes da menos de 180 respuestas. Es decir, menos de 1.2 sinónimos por pregunta.

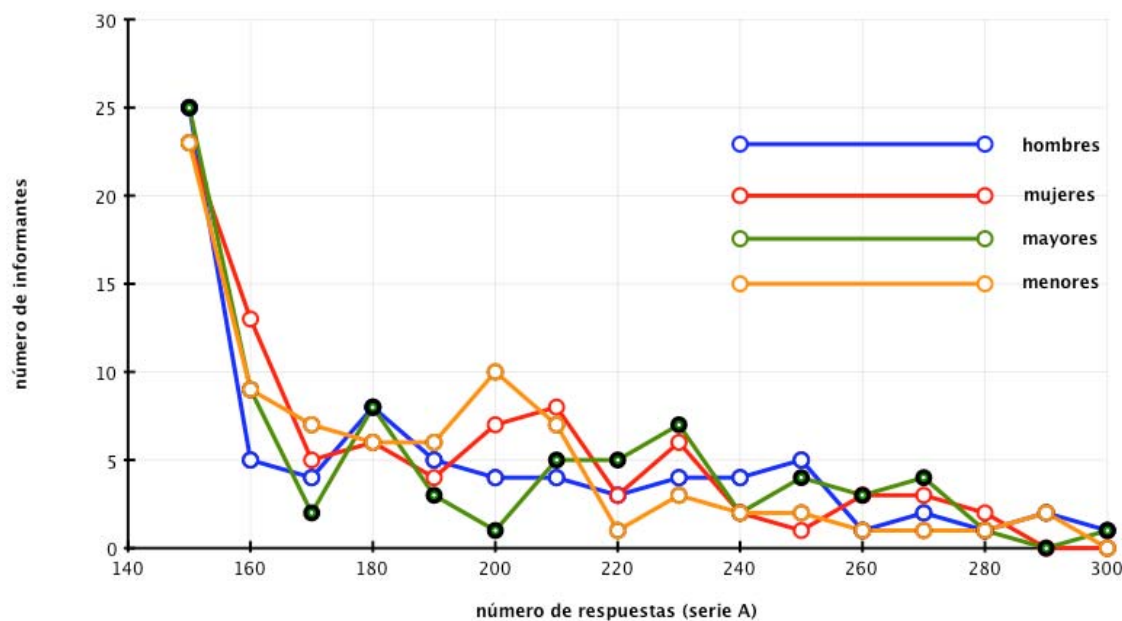
Número de respuestas	hombres	mujeres	mayores	menores
150	25	23	25	23
160	5	13	9	9
170	4	5	2	7
180	8	6	8	6
190	5	4	3	6
200	4	7	1	10
210	4	8	5	7
220	3	3	5	1
230	4	6	7	3
240	4	2	2	2
250	5	1	4	2
260	1	3	3	1
270	2	3	4	1
280	1	2	1	1
290	2	0	0	2
300	1	0	1	0

Tabla 15. Número de respuestas por grupos de informantes

Aunque en la figura 3 solamente se muestra hasta las 300 respuestas, 8 informantes han ofrecido más de 300, 4 más de 400 y un informante ha ofrecido 559 respuestas.

Estos datos sugieren que es posible considerar dos tipos de informantes en las encuestas de variación léxica. Un primer tipo, el más numeroso, de respuesta simple, que prácticamente se limita a escoger una respuesta por pregunta y ocasionalmente añade alguna más. Son más de la mitad de cada encuesta ofrecen hasta un máximo de 1.2 respuestas por pregunta. Un segundo tipo que considera más detenidamente las respuestas. Son bastante menos de la mitad y llegan a ofrecer 3 o 4 y algunos hasta 5 respuestas de media.

Creemos que es necesario observar más detalladamente las respuestas dadas por cada uno de estos tipos ya que puede afectar a la metodología de las encuestas.



## 6. Conclusiones

En este estudio hemos analizado algunas características de los informantes que han participado en las encuestas de variación léxica de Varilex. Hemos observado que, como era de esperar, que en vez de elegir informantes NORM, es preferible que haya variedad, ya que ofrecen distintos tipos de respuestas, tanto desde el punto de vista cuantitativo como del cualitativo.

Creemos que los datos de variación léxica se pueden estudiar muy cómodamente usando bases de datos en línea, no solamente por la posibilidad de analizar los datos con software ampliamente conocido, sino por la facilidad para trabajar en grupo.

En base a los datos de Varilex, no se observan grandes diferencias cuantitativas de variación léxica entre hombres y mujeres, o entre los informantes de las diferentes edades. Sin embargo, sí se observa que la distribución de las respuestas se ve afectada por el campo semántico del que se trate.

La mayoría de los informantes ofrece como máximo de 1.2 respuestas por pregunta y el resto va aumentando hasta llegar a 3 y 4 respuestas por pregunta, y excepcionalmente a más de 5. Creemos que este fenómeno puede afectar a la metodología de las encuestas y necesita un estudio más detallado.

## 7. Referencias

- ÁVILA, Raúl, José Antonio SAMPER, Hiroto UEDA, et al. (2003) *Pautas y pistas en el análisis del léxico hispano(americano)*, Madrid / Frankfurt: Iberoamericana Vervuert.
- CHAMBERS, J. K. & P. J. TRUDGILL (1980) *Dialectology*, Cambridge: Cambridge University Press.
- DUBOIS, Paul & Monty WIDENIUS (1999) *MySQL*, Indianapolis: New Riders Publishing.
- ECKERT, P. (2000) *Linguistic Variation as Social Practice*, Oxford: Blackwell.
- ORTON, H. (1962) *Survey of English dialects: introduction*. Leeds: E. J. Arnold.
- RUIZ TINOCO, Antonio (2000) *Manual de lenguaje SQL aplicado a la base de datos de Varilex*, en VARILEX 8, Tokio: Universidad de Tokio, Toshihiro Takagaki, Universidad de Estudios Extranjeros de Tokio y Universidad de Sophia.
- RUIZ TINOCO, Antonio (2001) “Cartografía automática en Internet”, en *Bulletin of the Faculty of Foreign Studies*, 36, Tokio: Sophia University, 77-93.

RUIZ TINOCO, Antonio (2005) “Léxico variable en los países hispanohablantes según el sexo”,  
en *Diccionario, Léxico y Cultura* (ed. Josefina Prado Aragonés y M<sup>a</sup> Victoria Galloso  
Camacho), Huelva: Universidad de Huelva Publicaciones, 189-198.