

LA COMPUTADORA EN LINGÜÍSTICA: EL CASO DEL
DICCIONARIO DEL ESPAÑOL DE MEXICO

*Luis Fernando Lara,
Isabel García H,
María Pozzi,
Roberto Ham,
Javier Becerra,
Paulette Levy*
EL COLEGIO DE MEXICO

1. *Introducción.*

La investigación para elaborar el *Diccionario del español de México* comenzó en 1973, con la ayuda de un fideicomiso creado por el gobierno federal a nombre de El Colegio de México y el Fondo de Cultura Económica.

El plan de elaboración de la investigación y de la redacción de la obra, su desarrollo metodológico y técnico, y la redacción final de las entradas del diccionario ha sido resultado de un dilatado trabajo interdisciplinario de lingüistas, matemáticos, especialistas en estadística y computación y jóvenes escritores mexicanos.

Dos han sido las principales etapas del trabajo: durante los primeros cuatro años, los planes de investigación, el sistema de programación de la computadora, la recolección de materiales y el procesamiento de datos; a partir del quinto año de trabajo, la explotación de los resultados y la redacción de entradas para el diccionario.

El objeto de este informe es describir, de manera breve, la aplicación de la computadora en cada una de las etapas de la elaboración del diccionario. Una exposición exhaustiva de varios aspectos de este trabajo se puede encontrar en el libro *Investigaciones lingüísticas en lexicografía*, publicado por El Colegio de México a principios de 1980.

2. *El problema.*

Durante siglos la lexicografía ha sido una tarea que consume tiempo, esfuerzo y hasta la vida de muchos hombres dedicados a registrar, ordenar y explicar los usos y los significados de millones de palabras emitidas por una comunidad lingüística. Baste recordar a Sir James Murray, quien a pesar de haber dedicado toda su vida al famoso diccionario de Oxford, murió antes de verlo terminado, pues la obra se llevó 96 años. Los talleres lexicográficos dan siempre la misma impresión: interminables filas de archiveros y ficheros, abigarradas bibliotecas en donde lo mismo se encuentra la más famosa obra literaria que el más modesto manual de recetas de cocina; redactores sumidos hoy en la investigación de un término de la navegación marítima, mañana en una expresión de la boca del pueblo.

Construir un diccionario implica: a) reunir el mayor número posible de palabras utilizadas por una comunidad de hablantes; b) certificar el uso extendido de cada palabra; c) documentar sus varias acepciones, la clase de hablantes que las utilizan, la región, el nivel de lengua, etc.

La tradición lexicográfica ha consistido siempre en una refundición de los vocablos registrados en diccionarios anteriores para hacer diccionarios nuevos; un lexicógrafo toma palabras de un diccionario y las vuelve a colocar en otro nuevo, quizá con algunas modificaciones. En el caso de México esta técnica nos llevaría al fracaso, pues nunca ha habido un diccionario mexicano que registre palabras efectivamente usadas en México y, por lo tanto, copiar el diccionario de la Academia junto con el Diccionario de Mejicanismos de Santa María, no aseguraría ni que las voces reunidas se usaran efectivamente en México, ni que las que se dejaran de reunir no se usaran efectivamente en México. Así que había que registrar vocablos en una forma diferente de la tradicional en lexicografía. Esto convierte un problema lexicográfico en un serio problema lingüístico.

3. *El corpus y su estadística.*

Se puede pensar que el vocabulario del español mexicano es la unión de los vocabularios conocidos por todos los mexicanos, pero como es imposible consultar a cada uno de nosotros y además pedir que emita una lista de las palabras que conoce, pues nadie tiene en su conciencia inmediata todo el vocabulario que usa o que entiende, es necesario recolectar emisiones lingüísticas en forma de conversaciones grabadas, libros, artículos periodísticos y de revista, etc. Se forma así una muestra del español usado por los mexicanos.

La muestra intenta cubrir las variedades regionales de la República, las diferencias debidas a las posiciones socio-culturales de los hablantes, los distintos usos de la lengua en situaciones profesionales o sociales, los dos sexos y las varias edades de los habitantes de México. Así se compuso un *Corpus del español mexicano contemporáneo* (CEMC) con muestras de habla que representan todas esas distinciones.

Elaborar un corpus de esta clase implica más que una colección arbitraria y aleatoria de textos, pues no todos importan lo mismo. Para dar un ejemplo, basta pensar que si se considera socialmente que los escritores son quienes mejor utilizan el español, es necesario darle una importancia correlativa a la narrativa contemporánea mexicana, en tanto que si las fotonovelas son textos muy leídos en México -aunque su calidad no nos guste a mucha gente- es también necesario que aparezcan representadas en la muestra. Se hace entonces imprescindible asignar a la formación del corpus una hipótesis sociolingüística que determine los distintos pesos cuantitativos de cada género de textos reunido.

Como el *Diccionario del español de México* tiene por meta presentar a sus usuarios el vocabulario estándar que se utiliza en la comunicación general del país, la ponderación de los géneros del corpus se establece a partir de la llamada "lengua culta" (de la literatura, el texto científico y el periodismo) y decrece hasta géneros menores, como el caló del hampa o el lenguaje de

la onda. En el cuadro 1 se puede ver la composición del CEMC y el número de ocurrencias de palabras para cada género.

Tal procedimiento no ha sido usual en trabajos anteriores de estadística lingüística y, por eso mismo, plantea problemas especiales al cálculo: normalmente lo que interesa del estudio estadístico de una palabra es saber a) qué tan frecuentemente se usa, b) qué tan extendido esté su uso. Lo primero se resuelve con un recuento de su frecuencia de aparición en el corpus; lo segundo con una medida de la dispersión o repartición del vocablo entre varios tipos de texto. Pero puesto que los géneros de texto no comprenden el mismo número de ocurrencias, sino que este número ha sido determinado por la ponderación sociolingüística antes mencionada, ambas medidas no dan una descripción adecuada de las dos cuestiones. Para resolver este problema optamos por la introducción de dos cálculos más: el primero, llamado "frecuencia corregida" (KF en el cuadro 2) establece una relación entre frecuencia, dispersión y el tamaño relativo del género; el segundo traduce esa relación en una medida entre cero y uno, que indica al redactor cuando un vocablo es usual y frecuente en español mexicano (C en el cuadro 2).

La utilización de estos resultados permite dar una buena respuesta al problema inicial de la lexicografía, tal como se expuso en el párrafo 2. El CEMC ofrece la posibilidad de reconocer un gran número de palabras (aproximadamente 68,000) y a la vez de certificar la extensión del uso de un importante porcentaje de ellas (por lo menos el del 75% de las voces más usadas en México).

4. *El analizador gramatical.*

Pero reunir los dos millones de ocurrencias que forman el CEMC es una tarea imposible para un trabajo manual; solamente el advenimiento de la computadora ha hecho posible que esta clase de investigaciones puedan efectuarse. La computadora lee los textos, separa las palabras, identifica sus flexiones y categorías gramati-

cales, agrupa formas distintas de una misma palabra, las cuenta, establece la localización de cada palabra en el corpus y hace los cálculos antes citados.

Para hacer este trabajo había dos posibilidades: o bien que una persona leyera todos los textos del corpus y marcara las categorías gramaticales de cada palabra (una pre-edición), para que el trabajo de computación sólo consistiera en la identificación de cadenas alfabéticas, o bien se elaboraba un sistema de análisis automático de los textos. Optamos por la segunda posibilidad, que ahorra el tiempo que llevaría la pre-edición de los textos y a la vez aprovecha las ventajas que ofrece la computadora.

Un texto de nuestro corpus está formado por uno o varios párrafos entresacados aleatoriamente de un libro, una revista o una cinta magnética. La unidad "párrafo" es una unidad de sentido necesaria para obtener de ella el análisis semántico de las palabras que contenga. El texto consta de tantos párrafos como hagan falta para reunir 2,000 palabras gráficas. La unidad de sentido es a la vez una garantía de que el texto bajo análisis conserva todas las peculiaridades de la lengua natural. En el uso diario del español hay distintos tipos de oraciones, significados ambiguos, repeticiones, indicaciones anafóricas, etc. Nuestra investigación consistió en elaborar un sistema de análisis morfológico y rudimentariamente sintáctico, que pudiera determinar automáticamente la morfología de las palabras analizadas y su categoría gramatical; a partir de ese análisis, el reconocimiento y recuento estadístico de los vocablos y la producción de listas de contextos para el análisis semántico se vuelve posible.

En términos teóricos, el sistema de análisis automático es una gramática distribucional, en que el reconocimiento de cada segmento del texto se hace por relaciones de dependencia de unos vocablos con otros; es decir, se determina la palabra y su categoría gramatical según las características de las palabras antecedentes y de las palabras consecuentes. Llevaría más páginas que las dedicadas a este artículo para hacer una explicación detalla-

da del analizador gramatical del DEM; baste un ejemplo para dar una idea de su funcionamiento.

El analizador está compuesto por dos listas de paradigmas gramaticales, que contienen artículos, preposiciones, conjunciones, adverbios, pronombres, etc. Para cada una de estas palabras se han establecido dos índices: uno sobre su categoría gramatical y otro sobre la clase de palabras que pueden seguirlas (por ejemplo: a un artículo determinando puede seguir un nominal); cuando una misma forma puede tener dos categorías (por ejemplo /a es un artículo o un pronombre), cada forma se ha listado en diferente archivo. El analizador tiene además una tabla de arborescencias para el análisis de las terminaciones verbales, y cuatro tablas más de relaciones de precedencia, postcedencia y distribución homográfica de terminaciones verbales y no-verbales. En total, cerca de 1000 registros previos. A base de ellos se reconoce el resto de las palabras, fundamentalmente nominales, verbos, adverbios, demostrativos y casos de homografía.

El proceso de análisis consiste, a grandes rasgos, en comenzar buscando si la primera palabra gráfica del texto forma parte de los archivos de paradigmas cerrados (llamados "diccionario" y "decisiones"); si eso se comprueba, los indicadores asociados a cada palabra de los archivos permiten utilizar las tablas de dependencia y así prever la clase de palabras que pueden continuar; una vez rectificada la palabra en cuestión, el procedimiento continúa hasta terminar el texto. El mismo papel de apoyo lo tiene el reconocimiento de las terminaciones verbales. Un ejemplo se encuentra en el cuadro 3; un ejemplo del análisis terminado en el 4.

Una pregunta importante que uno se plantea cuando se ha dado cuenta de la enorme gama de posibilidades de construcción sintáctica del español es sobre el rendimiento del sistema. Como punto de referencia para ello baste decir que, en otros sistemas que tienen la misma finalidad para el análisis del inglés -ya que no hay ningún otro sistema para el español que permitiera una comparación más adecuada- el rendimiento del análisis no pasa

del 67%. En nuestro caso el tope ha sido 68% en el análisis de textos literarios y el mínimo de 54% en el análisis de textos periodísticos.

Ello significa que hay un trabajo manual posterior, que corrige las fallas del análisis automático. Con todo, el tiempo ganado gracias al analizador automático, la sistematicidad del análisis y los varios juegos de clasificación que permite la computadora hacen del trabajo un éxito.

5. *La producción de concordancias.*

Además de los resultados estadísticos explicados en el párrafo 3, que dieron lugar a nuestro diccionario estadístico, del análisis gramatical se obtienen listas en que, para cada vocablo, se ofrecen tres líneas de contexto junto con su localización en el corpus. Estas listas, técnicamente llamadas *concordancias*, son la materia prima del posterior análisis semántico que se refleja en la redacción del diccionario. Un ejemplo se da en el cuadro 5. Por nuestra experiencia, hemos visto que tres líneas de contexto son suficientes para la mayor parte de los análisis; en caso de que haya todavía alguna ambigüedad, siempre podemos regresar al texto completo.

6. *Un programa especial: la "Horquilla".*

La obtención de concordancias plantea también problemas. El número de ocurrencias de un vocablo en el corpus sigue una curva asintótica según la relación entre cantidad de palabras del corpus y cantidad de palabras distintas que se obtienen de él. Esto quiere decir que hay vocablos como *la*, que aparece 87,827 veces, para luego comenzar a dechecer drásticamente la relación. Ya para el centésimo vocablo más frecuente, *trabajo*, la frecuencia es de 1,542 apariciones. Hay que imaginar qué le pasaría a un redactor que tenga que revisar todas las concordancias de las palabras más frecuentes. Analizar 200 concordancias le lleva a uno un día de trabajo. Así, salvo que se quisiera uno convertir en personaje de Borges, es preferible intentar un método que reduzca el número de

concordancias por analizar sin que elimine información importante. Este intento dio lugar a nuestro programa de la "Horquilla". Consiste en un reconocimiento de patrones morfo-sintácticos obtenidos del análisis anterior y que permite, para cada patrón que se repita idénticamente, registrar su aparición para luego seleccionar, de entre ellos, tres ejemplos. En el cuadro 6 hay un ejemplo. Así, del patrón: "Edad de + verbo", que aparece seis veces, hay una reducción a los tres contextos citados: *edad de ser*, *edad de comprender* y *edad de tener*. En esa forma el redactor ve disminuido su trabajo, hasta ahora, en cerca del 30 al 40%; recibe además una lista ordenada de concordancias y tiene datos sobre construcciones sintácticas usuales, clichés, recciones de verbos y preposiciones, etc.

En esta forma se cumple la tercera de las condiciones lexicográficas expuestas en el párrafo 1. La información obtenida de la explotación del corpus elimina la subjetividad del lexicógrafo en la selección de datos, sistematiza regularmente los vocablos y sus análisis gramaticales, indica la extensión y las peculiaridades del uso de un vocablo en México, ahorra trabajo de clasificación al redactor y permite utilizar el material cuantas veces sea necesario en arreglos diversos, tanto para fines lexicográficos como para estudios lingüísticos de otros órdenes.

7. *Control de la redacción.*

Pero aunque la subjetividad del redactor ha dejado de intervenir en la obtención de los datos, no deja de aparecer -y se la asume metodológicamente- en el análisis semántico y en la presentación del artículo lexicográfico. Como se sabe, apenas despuntan hoy en día algunos experimentos computacionales con la semántica y sus resultados no pasan de ser curiosos aunque inútiles para su aplicación lingüística. El redactor es un intérprete del significado de una palabra y sus esfuerzos por adecuar su análisis a la realidad presente en los datos no dejan de dar paso a sus valoraciones personales, su criterio, su gusto y, hay que recalcarlo, a su ignorancia. Así, un

lexicógrafo puede optar por seleccionaran ejemplo de una obra determinada en vez de otra; consultar cierto libro especializado en vez de otro; juzgar los mejores datos de un diccionario en vez de los que ofrece otro, etc. Por eso interesa tener, a lo largo del trabajo de análisis semántico y redacción, una serie de informes sobre las fuentes consultadas para redactar un artículo, sobre la proveniencia de la entrada misma (que puede ser distinta al CEMC, si se trata de la recomendación de alguno de los asesores científicos del diccionario), sobre los autores de los ejemplos incluidos, etc. Además se requiere un informe sobre vocablos que exigen una referencia cruzada en letras del diccionario que todavía no se redactan.

Para ello hemos -creado un sistema interactivo de control de la redacción que indica: quién definió y quienes revisaron la redacción de una entrada, qué palabras implican referencias cruzadas, cuáles fueron las fuentes de la entrada, de la definición y de los ejemplos, qué palabras parecen ser mexican ismos, cómo se ha resuelto una entrada en relación con los resultados estadísticos y cuál es la relación de entradas y subentradas. Este sistema, que próximamente pondremos en uso diario, permite saber qué redactor escribió cuáles palabras, en qué artículos se ha utilizado la misma fuente de consulta, etc., etc.

La utilidad del sistema de control de la redacción no solamente se ve ahora, sino que da la base para posteriores correcciones y adiciones al diccionario. Convierte nuestro trabajo en un banco de datos lexicográficos muy útil para el futuro. Un ejemplo de la hoja de datos para este sistema aparece en el cuadro 7.

8. *Fotocomposición.*

Por último, explotamos una capacidad más de la computación electrónica. Si el DEM tendrá 30,000 entradas y cada una de ellas contiene varias acepciones, ejemplos, subentradas, construcciones (véase el cuadro 8) y referencias cruzadas, el trabajo de imprenta se convierte en un serio problema de revisión de pruebas y manejo tipo-

gráfico. Según lexicógrafos experimentados del extranjero, los procesos de revisión de pruebas de un diccionario consumen una mitad más del tiempo total empleado en escribir la obra. Hemos pensado que, si cada redactor se ocupa de preparar la hoja de redacción de cada artículo escrito por él, en tal forma que ésta pueda pasar por un lector óptico, se elimina una cantidad importante de errores, se reduce el tiempo de revisión y se aprovecha la cinta producida por el lector óptico para alimentar la computadora con el diccionario terminado, abriendo así la posibilidad de conservar la obra "en máquina" y poderla someter a correcciones en el momento en que se desee. Todo esto nos lo facilita el sistema de fotocomposición, de cuyos resultados ofrecemos una prueba en el cuadro 8.

9. *La máquina.*

No está de más afirmar que hemos tenido suerte en nuestro trabajo con la computadora; se ha debido, fundamentalmente, a los buenos consejos que recibimos, al inicio y ahora, de Enrique Calderón, a la colaboración prestada por el Centro de Procesamiento y Evaluación Dr. Arturo Rosenblueth, de la SEP, y al entendimiento que hemos logrado construir en nuestro equipo interdisciplinario.

Todos los sistemas se han preparado en la máquina Univac 1106 del Centro mencionado; su programación se ha hecho en la versión de ALGOL 60 de la Universidad de Noruega.

Hemos preferido presentar este resumen de nuestro trabajo con muy pocas referencias técnicas, para hacerlo más accesible a un público amplio; no obstante, estamos dispuestos a pasar a mayores detalles con cualquier persona interesada en ellos.

10. *Conclusión.*

Las posibilidades de trabajo que abre la computación para la lingüística son enormes. No sólo permite manejar grandes cuerpos de datos y obtener de ellos variadas composiciones de clasificación, sino que además obliga a

uno a establecer con mayor rigor y sistematicidad sus cálculos y modelos. Ofrece, además, la oportunidad de extrapolar de ella ideas, sugerencias y procesos de análisis que ayudan a mejorar nuestra comprensión del único lenguaje que subierte nuestros conocimientos: la lengua de todos los días.

CUADRO 2

Num.	VOCABLO	CAT. GRA.	KF	F	C	Núm.	VOCABLO	CAT. GRA.	KF	F	C
1	LA	ART	85919.13	87827	0.9780	51	VEZ	S	3152.41	3239	0.9730
2	EL	ART	78942.85	80845	0.9806	52	NOS	PRON	3146.58	3389	0.9277
3	DE	PREP	63088.75	63541	0.9941	53	AÑO	S	3102.49	3294	0.9412
4	Y	CONJ	59255.11	59476	0.9962	54	SABER	V	3085.00	3441	0.8954
5	QUE	PRON	58364.00	58595	0.9960	55	SIN	PREP	2999.75	3188	0.9403
6	EN	PREP	48228.00	49299	0.9780	56	HASTA	PREP	2979.90	3034	0.9820
7	A	PREP	44569.77	44809	0.9946	57	QUERER	V	2916.40	3546	0.8204
8	SE	PRON	33442.49	33703	0.9922	58	DEBER	V	2893.21	3199	0.9033
9	NO	ADV	28670.65	31468	0.9101	59	TODO	PRON	2892.46	3137	0.9212
10	SER	V	24428.99	24918	0.9802	60	AQUI	ADV	2818.17	4033	0.6953
11	UN	ART	20561.20	20646	0.9958	61	UNO	PRON	2655.45	3097	0.8558
12	POR	PREP	19694.88	19862	0.9915	62	DIA	S	2623.75	2785	0.9414
13	CON	PREP	18660.50	18775	0.9938	63	ESO	PRON	2526.39	3352	0.7509
14	SU	ADJ	17233.52	17823	0.9665	64	QUÉ	PRON	2522.64	3176	0.7919
15	UNA	ART	15597.41	15692	0.9939	65	ELLA	PRON	2410.23	2612	0.9219
16	HABER	V	14373.69	14619	0.9830	66	SOBRE	PREP	2410.01	2692	0.8941
17	PARA	PREP	13951.97	14164	0.9849	67	BIEN	ADV	2361.08	2499	0.9442
18	AL	CONTR	11028.71	11207	0.9839	68	LLEGAR	V	2348.04	2442	0.9611
19	ESTAR	V	10940.95	11673	0.9366	69	MAS	ADJ	2341.78	2465	0.9494
20	COMO	ADV	10753.78	10818	0.9940	70	DONDE	ADV	2274.15	2320	0.9800
21	TENER	V	9402.55	9818	0.9572	71	ENTRE	PREP	2268.93	2457	0.9226
22	LE	PRON	919213	10539	0.8707	72	NI	CONJ	2249.99	2391	0.9404
23	HACER	V	8607.80	8826	0.9750	73	OTRA	ADJ	2242.32	2265	0.9899
24	YA	ADV	8238.52	9815	0.8376	74	ENTONCES	ADV	2229.34	2960	0.7503
25	O	CONJ	7989.34	8368	0.9542	75	ESA	ADJ	2227.29	2333	0.9542
26	PERO	CONJ	7706.24	8343	0.9228	76	LLEVAR	V	2185.06	2297	0.9507
27	DECIR	V	7612.18	8881	0.8555	77	PONER	V	2129.97	2319	0.9176
28	QUE	CONJ	7565.24	7978	0.9477	78	PARTE	S	2116.71	2241	0.9439
29	LO	ART	7335.91	7378	0.9942	79	TE	PRON	2048.26	3392	0.5993
30	ME	PRON	6928.90	10427	0.6607	80	TIEMPO	S	2047.34	2068	0.9899
31	MAS	ADV	6877.20	6964	0.9874	81	DOS	S	1995.51	2019	0.9882
32	PODER	V	6200.40	6350	0.9762	82	DESPUES	ADV	1988.68	2035	0.9770
33	ESTE	ADJ	6180.97	7142	0.8639	83	DEJAR	V	1982.57	2098	0.9444
34	IR	V	5621.48	7047	0.7954	84	DESDE	PREP	1886.32	1910	0.9875
35	LO	PRON	5549.14	6118	0.9060	85	HOMBRE	S ^	1877.77	1989	0.9434
36	Sí	ADV	4976.58	8656	0.5701	86	ESE	ADJ	1869.75	1980	0.9437
37	VER	V	4891.83	5566	0.8775	87	CADA	ADJ	1826.01	1884	0.9689
38	DAR	V	4876.03	5045	0.9661	88	VENIR	V	1786.61	2094	0.8515
39	CUANDO	ADV	4751.19	4907	0.9679	89	QUEDAR	V	1786.50	1862	0.9590
40	MUY	ADV	4503.17	4913	0.9156	90	AHORA	ADV	1785.13	1920	0.9290
41	YO	PRON	4390.14	7021	0.6210	91	ESTO	PRON	1758.29	1811	0.9706
42	PORQUE	CONJ	4202.55	5079	0.8255	92	PASAR	V	1756.64	1906	0.9207
43	EL	PRON	4101.76	4510	0.9085	93	NADA	PRON	1722.46	2165	0.7933
44	MI	ADJ	4046.84	5744	0.7012	94	SIEMPRE	ADV	1610.75	1667	0.9659
45	PUES	CONJ	4005.46	5918	0.6731	95	VIDA	S	1560.79	1697	0.9188
46	LA	PRON	3927.33	4259	0.9212	96	CASA	S	1546.17	1813	0.8512
47	ASI	ADV	3739.74	4157	0.8985	97	SÓLO	ADV	1531.92	1713	0.8931
48	ESTA	ADJ	3530.77	3735	0.9447	98	TOMAR	V	1505.79	1536	0.9801
49	TODO	ADJ	3428.97	3455	0.9924	99	FORMA	S	1501.64	1707	0.8783
50	TAMBIEN	ADV	3388.18	3498	0.9682	100	TRABAJO	S	1495.21	1542	0.9693

CUADRO 3

Ilustración 3

DICCIONARIO

	Catgra	Ingra	Frec	Catgra	Ingra	Frec
de	4	13		cual	1	6	+1
ellas	5	3	+1				
cual	5	6					
si	3	1	+1				
*fuesen	9	30	1				

DECISIONES

.....tirando de ellas cual si fuesen toros bravos.....
 5 1 3 9

3

13-----12 de ellas

6

3-----23

ellas cual

3

6-----12 cual si

30

3-----13 si fuesen

CUADRO 4

ANALISIS.GRAMATICAL

@000194@000007003@DE.PRONTO SE OYO/UN DISPARO, EL PERRO LANZO/ UN GEMIDO SORDO Y NO
 @000195@000007004@LADRO/MA/S.
 @000196@000013014@DEMETRIO@DESPERTO/SOBRESALTADO, VADEO/EL RI/O Y TOMO/LA VERTIENTE
 @000197@000013015@OPUESTA DEL CA+O/N. COMO HORMIGA ARRIERA ASCENDIO/LA GRESTERIA,
 @000198@000013016@CRISPADAS LAS MANOS EN LAS PE+AS Y RAMAZONES, CRISPADAS LAS PLANTAS
 @000199@000013017@SOBRE LAS GUIJAS DE LA VEREDA.
 @000200@000020007@LOS FEDERALES GRITABAN A LOS ENEMIGOS, QUE, OCULTOS, QUIETOS Y

DE	INGRA: 1	CATGRA:PREPOSIC	FRECUENCIA: 1
PRONTO	INGRA: 1	CATGRA:ADVERBIO	FRECUENCIA: 1
SE	INGRA: 3	CATGRA:PRONOMBR	FRECUENCIA: 2
OY-	INGRA: 3	CATGRA:VERBO	FRECUENCIA: 1
UN	INGRA: 2	CATGRA:ARTICULO	FRECUENCIA: 1
DISPARO	INGRA: 3	CATGRA:NOMINAL	FRECUENCIA: 1
EL	INGRA: 5	CATGRA:ARTICULO	FRECUENCIA: 1
PERRO	INGRA: 3	CATGRA:NOMINAL	FRECUENCIA: 1
LANZ-	INGRA: 30	CATGRA:VERBO	FRECUENCIA: 1
UN	INGRA: 2	CATGRA:ARTICULO	FRECUENCIA: 2
GEMIDO	INGRA: 31	CATGRA:NOMINAL	FRECUENCIA: 1
SORDO	INGRA: 0	CATGRA:AMBIGUA	FRECUENCIA: 0
Y	INGRA: 1	CATGRA:CONJUNCI	FRECUENCIA: 1
NO	INGRA: 7	CATGRA:ADVERBIO	FRECUENCIA: 1
LADR-	INGRA: 30	CATGRA:VERBO	FRECUENCIA: 1
MA/S	INGRA: 23	CATGRA:AMBIGUA	FRECUENCIA: 0
DESOERT-	INGRA: 30	CATGRA:VERBO	FRECUENCIA: 1
SOBRES.ALTADO	INGRA: 0	CATGRA:AMBIGUA	FRECUENCIA: 0
VADE-	INGRA: 30	CATGRA:VERBO	FRECUENCIA: 1
EL	INGRA: 3	CATGRA:ARTICULO	FRECUENCIA: 2
RI/O	INGRA: 31	CATGRA:NOMINAL	FRECUENCIA: 1
Y	INGRA: 1	CATGRA:CONJUNCI	FRECUENCIA: 2
TOM-	INGRA: 30	CATGRA:VERBO	FRECUENCIA: 1
LA	INGRA: 4	CATGRA:AMBIGUA	FRECUENCIA: 0
VERTIENTE	INGRA: 0	CATGRA:AMBIGUA	FRECUENCIA: 0
OPUESTA	INGRA: 0	CATGRA:AMBIGUA	FRECUENCIA: 0

CUADRO 5

APELLIDO	SITUACION/ PARFJA.		
1.- 03023024 03023024	MI NOMBRE ES NICOMACOOI MI APELLIDO FLOO EN MITAD DEL INVIERNO CUANDO CALLADA A	ACQ. HACTA A LA MEDIANOCHE/ 1/2AS. CON GRAN DESOLACION	identificación después del nombre
APELLIDO			casta, familia
2.- 044075007 044075009	VOLUNTAD AVANZAS Y TROPIZAS Y SIN RECIBO FUERZA DEL APELLIDO. EL ULTIMO DE LA EST TODAS LAS MUJERES PERDIDO Y EN TODOS LOS	POSICIONES QUIJANA POR LA Y Y EL QUE TERMINO/ EN MUJERES CLASIFICADO. NO/ NO	nombre, identificación
APELLIDO			identificación, padre
3.- 134030047 134030049	DESATA MAJES MAYORES QUE LOS QUE SE PRON PORQUE DAFIPLANTE A LA EMPRESA DE PAIT UNA SUGERENCIA SANDRA QUIZA TROPICIFIA	TA SONTEP/ SIHO ADEMAS NO (PNO HAY EN EL APELLIDO DE PAHOOPAI) SE DESAPAREJA	identificación, padre
APELLIDO			del marido, mayúscula
4.- 102013010 102013014 102013016	E/L CON DOS PALABRAS: FREYA LONITON. MI APELLIDO COMIENZA QUE PODIA DE-TROPICIA AL TAN BIEI PUNTAO A LA MUJER QUE LO HABIA	TOSO/ PORQUE ERA UN DDE/ ME PUNCA TUVO Y QUE HIZANO COMO JOPELO OMARO Y	del marido, mayúscula
APELLIDO			nombre, legal
5.- 494110024 494110026	SEFORA DE SEMIPLUMAO. SOALO TUA/ DE MAYI CORRESPONIA AL APELLIDO DEL MARIDO SEFOR DELO PIZOO SEFORA DE DELA TROBADI.	ELA LA PARTICULA QUE DE DEL CALSO/ SEFORA DE	nombre, legal
APELLIDO			nombre
6.- 603030100 603030102	DE HO ESTAR CASADO.../ SEFORA LE CALIDAD LOS SIN NO ESTAN RECONOCIDOS COMO TALSO. MI HABLAR DE L/ DAFITE FORTIFICIA/ TAN IMPORT	MPAP/ MI SIONTERA A VECES. APELLIDO... ESTO/ SIH DE EN UN HIJO DESDE QUE	nombre
APELLIDO			representación
7.- 672077290 672077296	PAULO/ HEDIENTE A LA DIFERENCIA DE CLAS/ HUCIA IMPORTANCIA. POR OTRA PARTE, EL AP/ COSA. SALVO/ COMO ME OCURRIA A MI/ QUE	AQUELLO NO DEBIA TENER UN SHOREO NO ERA GRAN TIENRA CUBIERTO DE OPO EN	representación
APELLIDO			representación, casta
8.- 672077292 672077294 672077296	AQUELLA MISMA TAMBE ESTUBE A PHITO DE DEF RICOS Y PORES. PHO/ MI PADRE ES RINITASI TIENE OTRO QUE CASI ESTREMECE. NUNCA TE A	FI/ HEDIEF HARER SHOREO Y DETRAS DE ESE APELLIDO TIRA/ COMO MARIDO NI/0, A	representación, casta
APELLIDO			representación, casta
9.- 672077294 672077296	TIENE OTRO QUE CASI ESTREMECE. NUNCA TE A CLARO/ NO SE LO DICE.	TIRA/ COMO MARIDO NI/0, A IPERE TU FALTA DE FORTUNA/.	representación, casta
APELLIDO			representación, casta
10.- 745003020 745003021	ENTONCES ARTUROO ME DICE. LO QUE SI/ ME / SU PRIMER NOMBRE, PERO ME ACURRO DE SU / SA/ NOMEZO... BUENO/ PUES A/I ME TICHES/ C	ROO DICE. NO ME ACUERDO DE L/0/ SE APELLINA PE/MEZO AL DIRECTORIO Y PE/MEZO	representación, casta
APELLIDO			identificación
11.- 829003094 829003097	ERA UNO DE IRIA TIENDA MUY GRANDE/ VIVI/A ESTE... /ME QUE/? NO ME ACURRO/ ME SU AP/ TIENDA TAMBIEN GRANDE. AGUI/ ENTRARA TOD	NO VICENTFO/ ESTE... QUE... DO/ E/STE TENI/A OTRA A AZU/CAR DE LA HACIENDA	identificación
APELLIDO			casta, pertenecer
12.- 917103040 917103041 917103042	HIJO. PA POCO ME CORRIO/ CUANDO ME FUI A QUINCE DI/AS Y YO ESTABA MA/S CHICO. IPOI ISI E/L ES CAPAZ DE NEGARME COMO SU HIJO	IMLCOOZ... Y ESO QUE FUERON /O ME QUITE/ SI APELLIDO! LO NIEGO COMO PADREI/	casta, pertenecer
APELLIDO			casta, pertenecer
13.- 917103041 917103042	HIJO. PA POCO ME CORRIO/ CUANDO ME FUI A QUINCE DI/AS Y YO ESTABA MA/S CHICO. IPOI ISI E/L ES CAPAZ DE NEGARME COMO SU HIJO	IMLCOOZ... Y ESO QUE FUERON /O ME QUITE/ SI APELLIDO! LO NIEGO COMO PADREI/	casta, pertenecer

CUADRO 6

EDAD DE + ART.	FAC 2
910006 73	1.- PUES LA...LA MU/SICA PORQUE MI PADRE ES MU/SICO, Y...YO ME PUSE A ESTUDIAR DE LA EDAD DE UNOS...DOCE AÑOS, ME PUSE A ESTUDIAR LA MUSICA Y EMPEOL/ A TOCAR EN LA ORQUESTA, COMO DE LA EDAD DE CATORCE
EDAD DE + PR FREC= 1 891005043	MADRE SE...SE NOS...TUVIERON UN...TUVIERON UN DISGUSTO Y YO POS 1.- COMO...EMPECE/ A BOLEAR DESDE LA EDAD DE SEIS A+OS, DE UNOS SEIS A+OS, ENTONCES YO BOLEABA ALLA/ EN LA FUNDICIO/N DE A/VALOS, QUE ES MI
EDAD DE + AV FREC=2 757016129	LO QUE LE ...POR LO...CON LO QUE TIENE CADA UNO...EL ESTADO CIVIL, 1.- GUARDAN, LA EDAD, DE DO/NDE SON, DO/NDE HAN VIVIDO, TODO ESO VIENE EN LOS, EN...LA COLUMNNA, EN...LA, LA...EN LA 7C0/M0 SE LLAMA LA...?,
EDAD DE +VB FREC+6 124068058	UNA ADOLESCENTE, DESCUBRE MIENTRAS JUEGA CON UNA TORTUGA DE PELUCHE 1.- LLAMADA JACKO EL DESTRIPIADOR, QUE YA ESTA/EN EDAD DE SER PRESA ERO/TICA, PERO TEREÑC100 REGRESA A LA CIUDAD, A LA CORRUPCIO/N Y AL
601006125	MI/A SE SE PORTE MAL, HABRA/ MUERTO PARA MI/ Y YA NUNCA QUERRE/ SABER 2.- MA/S DE ELLA'. LAS CHICAS YA ESTARAN EN EDAD DE COMPRENDER LO QUE AQUEL 'PORTARSE MAL' SIGNIFICABA Y LO ENTENDIERON A LA PERFECCIO/N. LA
812005125	AMIGA'. Y POR ESO MISMO HASTA LES HE PERMITIDO QUE VENGA, AL MENOS 3.- LUPEO NO ESTA/ EN EDAD DE TENER NOVIO, SIN EMBARGO YA LE PERMITI/QUE VENGA, COMO AMIGO.
EDAD DE + AJ FREC=5 492027048	CAMPO ACTUAL DE LA PINTURA MEXICANA, GENERACIONES NO TANTO 1.- IDENTIFICABLES POR LA EDAD DE SUS MIEMBROS CUANTO POR SU PERTENENCIA A ACTITUDES O INTERESES ARTI/STICOS. TOMANDO EN CUENTA SO/LO LA PINTURA
3440096036	DESARROLLO DE LA IMAGINACIO/N CREADORAYALGUNAS HABILIDADES PARA OPERAR 2.- CON HERRAMIENTAS SENCILLAS; ES, ADEMA/S, ADECUADO A LA EDAD DE MI HIJO. INSTRUCCIONES. LA PRESENTE ESCALA CONTIENE OCHO ASPECTOS ESENCIALES EN
602352182	PODRI/A ESTAR LIBERANDO O EMITIENDO LAMISMA CANTIDAD DE ENERGI/A POR 3.- UN PERI/ODO DEL ORDEN DE TRES MIL SEISCIENTOS MILLONES DE A+OS. LA EDAD DE NUESTRO SOL, CALCULADA POR DOVERSO/SME/TODOS, NOS DA CINCO O SEIS...
EDAD DE + PN FREC=1	

COMPUTADORA Y LINGUISTICA

DEM	HRD 1/2	Artículo No.	Documento LFG	Concordancias. <u>X</u>
	Fecha 15.5.80	Redactor EMS	Fecha documento 4.7.78	Dts. estadísticos <u>X</u> Dts. bibliográficos <u>Cuadro 7</u>

Entrada	x	Sub-entrada	GI-3
ATACAR		ATACADO	GI-13

Tipo	No. orden.	Total	% total	KF	C
ATAC-	5954	93	0.00492	92.3A3639A	3.9919546
ATACA-	5955	8	0.00042	7.7122034	3.9589905
ATACABA-	5956	1	0.00005	0.2208075	D.1117520
ATACA	5960	3	0.00016	1-9692493	D.6085286
ATACARA	5961	1	0.00005	0.6564164	3.6083286
ATA0UE-	5980	2	0.00011	1.3128328	3.6083286
AT'C.	6010	1	0.00005	0.6564164	3.6083286
Vocablo					
ATACAB					

Fuentes pronunciación:	Fuentes entrada: CEMC
-------------------------------	------------------------------

Fuentes definición: Ii DUF; EMS; 2 W3, EMS; 3 PR, EMS; 4 W3, WMS; 5 CEMC, PLCT; 6 EMS; 7 VELO; 8 MAS, DTM; II IVSON, VELO, RVPM; 2 EMS; III i DRM; DESAEH

Fuentes ejemplos Ii 338; 2 EMS; 3 279; 4 241; 5 PLCT; 6EMS; 7 VELO, II i EMS, EMS; 2 EMS, EMS

Diccionario (1) GRT.

atacar y tr II Actuar con violencia en contra de algo o alguien para vencerlo, destruirlo o causarle algún daño, especialmente usando la fuerza o las armas; agredir "El Coronel Braulio Vargas *atacó* al invasor la noche del 22 de enero" 2 Combatir verbal o moralmente algo o a alguien: criticar "*Han atacado* muy duramente su tesis sobre la propiedad privada" 3 Empezar una acción enérgica para solucionar o vencer problemas o dificultades "Deben *atacarse* los grandes problemas nacionales: empleo, salud, habitación, entre otros" 4 Perturbar el funcionamiento normal de algo o alguien, cualquier agente que tenga la propiedad de ejercer, particularmente sobre él, efectos nocivos "Se trabaja en el control y prevención de las enfermedades que *atacan*, al cañado" 5 Ejercer una acción química una sustancia sobre otra que resulta transformada o corroida "La humedad *ataca* al fierro" 6 *atacarie a uno aleo o alguien Popul* Producir en uno una reacción desfavorable, de molestia, de hastio "M/e *ataca* cuando empieza ron sus sermones" 7 *atacarie a alguien un pulmón Popul* Darle a alguien un cigarro "Atócame un pulmón, ¿no?" 8 *Más* Comenzar a ejecutar una composición musical con decisión y energía 11 *Popul* 1 Atiborrar a alguien, especialmente de comida "Lo *atacaba* de pan para que luego no le pidiera de comer" *atacrse* "Del diario se *ataca* de sopes y fritangas" 2 *atacar de risa a alguien* Provocarle mucha risa "Lo *atacaron* de risa con lo que le contamos" *Atacarse de (la) risa* Darle a uno mucha risa, reírle compulsivamente "Es una película en que *fe atacas de la nsa*" III *Hip* Retener la cabalgadura trabándole las nendas en la cabeza de la silla o conduciéndola con las riendas muy tirantes, de tal modo que arque el cuello hacia abajo, con el hocico cerca del pecho atacado 11 pp de *atacar* 2 adj Que padece o está enfermo de algo "Lo individuo *atacado* de hidrofobia" *Atacado del mal Rural* Enfermo de rabia II adj *Popul* Drogado; con adición a la drogas "Ese chavo ha sido siempre muy *atacado* y va a acabar mal" III adj *Popul* Muy ajustado, pegado al cuerpo, apretado "Traía unos pantalones bien *atacados*"

ataque sm II Acrío que se dirige contra aleo o alguien para vencerlo o destruirlo "El combate comenzó con un *ataque* contra un pueblo de Corea del Sur" *lanzarse al ataque* "El equipo de San Luis Potosí se *lanzó al ataque* apoyado por su público, que llenó el estadio" 2 Agresión (verba) o moral: crítica "Debe evitarse que haya el menor *ataque* a su reputación" 3 Medida que se toma para solucionar o enfrentar un problema a una dificultad "Es necesario el *ataque* a esa conducta sexual para determinar sus causas" *plan de ataque* Conjunto de actividades organizadas que persiguen un fin determinado: estrategia "\ oy a reconquistarlo a romo dé lugar: Raúl tiene que ser mio y empezaré mi *plan de ataque*" 3 *Mus* Decisión y energía con que un circulante comienza una frase musical "Tiene un buen *ataque* con el arco del violín III Perturbación repentina del funcionamiento normal de cualquier órgano o sistema: crisis: "*ataque cardiaco, Ataque de apendmitis, de epilepsia, etc.* 2 Alteración repentina de la conducta normal *Ataque de risa* a XLFYU) bc Y UHei Y Popul Perturbarse o enojarse por algo que ha sucedido y que no se esperaba "LY] UXLFY UHei Y cuando vea que le quemaron su vestido nuevo"

atacador sm v *engallador*atación sm v *atración* "Se dio tal *atación* de mota que no se repuso en varios días"

abuodar v intr I Existir, darse o tener algo en gran cantidad "Región fértil y próspera en que *abundaban* por dondequiera los arroyos y aguajes" *Abundar en* "La cercana Sierra Madre *abunda* aún hoy en animales de caza como tigres, tigrillos, osos, etc." 2 Entenderse en consideraciones y razonamientos: ampliar, agregar *Abundar en* *Abundar sobre* "Y *abundó sobre* el mismo tema para que no hubiera duda alguna"

abundamiento *a mayor abundamiento* Además: por si fuera poco; con mayor razón "Los problemas de la guerra empiezan cuando el hacha y el cuchillo se substituyen por el arco y la flecha o, a *mayor abundamiento*, por el rifle y el cañón"

abundancia f I Gran cantidad de algo *En abundancia* "El pueblo tenía agu a *en abundancia*" 2 Prosperidad económica; bienestar "Nos hacemos la ilusión de que los buenos tiempos de la paz y la *abundancia* habían vuelto a nuestra casa"

abundante adj m y f Que se da en forma cuantiosa, en gran número o cantidad "*abundante en* "Zonas de aguas profundas, *abundantes en* especies marinas"

aceite sm I Líquido graso, combustible, no soluble en agua, de origen animal, vegetal, mineral o sintético. Se emplea, de acuerdo con sus tipos, como lubricante para máquinas y motores (el que se obtiene por refinación del petróleo); en la preparación de alimentos (*aceite de oli* a. *aceite de cártamo, aceite de ajonjolí*, etc.); en medicina (principalmente como tónico y purgante: *aceite de hCgado de bacalao, aceite de nano*, etc.); en la fabricación de cosméticos (*aceite de almendras, aceite de tortuga*, etc.); en la producción de pinturas (*aceite de linaza, aceite dq soya*); *ser como el agua y el aceite* Ser, dos personas o cosas, totalmente diferentes; no llevarse bien 2 v *oleo* aceite* ro 1 adj y s Que se refiere al aceite "Industria *aceitera*" 2 & Recipiente que sirve para contener o aplicar aceite aceitoso adj Que tiene aceite, generalmente en exceso, o consistencia de aceite; grasoso

Nota: estos artículos son de prueba tipográfica; no son los definitivos.

PUBLICACIONES SOBRE EL DICCIONARIO DEL ESPAÑOL DE MEXICO.

Informaciones generales en la Gaceta del FCE:

- LARA, L.F., "Sobre la justificación de un diccionario de la lengua española hablada en México".
19 (julio de 1972) 1-6
"Sobre el Diccionario del Español de México".
30 (junio de 1974), 10-14.
"El Diccionario del Español de México (estado actual del proyecto)". 68(agosto de 1976), 2-A.

Sobre la aplicación de la computadora en lexicografía:

- LARA, L.F.; GARCIA Hidalgo, I. "El uso de la computadora electrónica en la elaboración del Diccionario del Español de México" en: Reunión de trabajo sobre la aplicación de las computadoras en el área de las ciencias sociales. INAH/CIMAS. México, mayo de 1974. *Cuadernos de trabajo del INAH*.

- GARCIA Hidalgo, I., "Sobre la generación del primer Diccionario del Español de México" en: *Boletín Informativo de la Dirección General de Planación Educativa*, SEP. agosto de 1975, pp. 62-79.

- LARA, L.F., "On lexicographical computing: some aspects of the work for a Mexican Spanish Dictionary" en: *Bulletin of the Association for Literary and Linguistic Computing*. 4, 2 (1976), 1-8.

- LARA, L.F., "Méthode en lexicographie: valeur et modalité du dictionnaire de machine: en *Cahier de Lexicologie*. 29 (1976), 103-128.

GARCIA Hidalgo, I. "A computational system for the Dictionary of Mexican Spanish", de próxima publicación en las actas de la *3rd. International Conference on Computers and the Humanities*, Waterloo. Ontario, Canadá.

"La formalización del analizador gramatical del DEM" en: *Investigaciones lingüísticas en lexicografía*. El Colegio de México, (Jornadas 89), 1979. pp. 85-155.

SOLER, Ma. Angeles "Problemas en la elaboración de un diccionario de máquina" en: *Actas del IV Congreso de ALFAL*, Lima, Perú, 1978.

POZZI, M. BECERRA, J. RANGEL, J. "A method to reduce large number of concordances" en: *8th. International Conference on Computational Linguistics*, Tokio, 1980.

LARA, L.F. HAM Chande, R. "Base estadística del Diccionario del Español de México", *NRFH*, 23 (1974), republicado en: *Investigaciones lingüísticas en lexicografía*. El Colegio de México, (Jornadas 89), 1979.

HAM, R. "Del 1 al 100 en lexicografía" en: *Investigaciones lingüísticas en lexicografía*. El Colegio de México, (Jornadas 89), 1979, PP. 41-83.

Sobre gramática y semántica;

VALADEZ, C.D. "El papel del nombre sustantivo en los diccionarios de lengua" en: *Actas del IV Congreso de ALFAL*.

GAVALDON, L. "Para el análisis lexicográfico de las locuciones", en *Actas del IV Congreso de ALFAL*.

LARA, L.F. "Del análisis semántico en lexicografía" en: *Investigaciones lingüísticas en lexicografía*. El Colegio de México, (Jornadas 89), 1979, pp. 157-266.

Sobre lexicografía:

LARA, L.F. Regional dictionaries: "A lexicographical proposal for the Third World" en: *Actes du III Congres de Linguistique Appliquée*, Montreal, 1979.

Recibido en enero de 1981