

## **Algunas dificultades iniciales en el tratamiento electrónico del léxico periodístico**

DUDAS DE UN AUXILIAR DE INVESTIGACIÓN

EDNA R. IBARRONDO

En el Instituto de Investigaciones Sociales de la Universidad Nacional Autónoma de México se ha iniciado una investigación sobre la prensa nacional, su lenguaje y la adecuación sociopolítica y pedagógica del mismo a la realidad de México, bajo la dirección de la investigadora Regina Jiménez de Ottalengo. En esta pesquisa se maneja el léxico periodístico dentro de un marco suministrado por la visión analítica del “proceso de comunicación”. Son elementos de dicho proceso: el emisor, el receptor y el mensaje, los cuales se vinculan entre sí, y se sitúan en cierto contexto en el cual se realiza la comunicación. El mensaje se trasmite del emisor al receptor por diversos “medios de difusión” entre los que destaca la prensa, que desempeña un papel muy importante en la sociedad. Este medio se vale de la palabra (más particularmente, de la palabra escrita) para cumplir su cometido y servir de enlace entre el emisor y el receptor. Es por esto por lo que, una investigación como aquella a la que nos referimos, cuyo fin último es sociológico, tiene que tener un arranque lingüístico y, más específicamente, ha de tener un punto de partida lexicológico.

En esta segunda mitad del siglo xx, los medios de difusión han aumentado en número, variedad y accesibilidad y —como indica la investigadora citada— por ello mismo “podría pensarse que hoy todos los grupos sociales pueden tener acceso —en mayor o menor grado— a esos medios de difusión”. Sin embargo —como también señala— “es indudable que uno de los grandes problemas sociales de la actualidad es la falta o la insuficiencia de la comunicación, a pesar del incremento y de la irradiación cada vez mayores de los medios de difusión”.<sup>1</sup> Esto nos lleva a preguntar si estos medios de difusión están cumpliendo o no con los fines que con ellos se espera alcanzar.

Para responder a esa pregunta, se necesita realizar un estudio que presenta varias facetas y que, por lo mismo, impone que se realicen

aquellos trabajos interdisciplinarios que permitan ver todos sus aspectos de manera integral: conexas y no aislada. Por eso mismo, se ha pensado que la sociología, la psicología, la lingüística y la sicolingüística son algunas de las disciplinas que deben intervenir en esa pesquisa.<sup>2</sup>

Por otra parte, debido a la extensión de un estudio como éste, se ha pensado en llevarlo a cabo por etapas. De éstas, la primera consiste en: elegir un medio de difusión (la prensa) limitando —por razones prácticas— el ámbito de observación; así, en un primer sondeo se eligió un solo diario de circulación nacional (“Excélsior”, de México) y se concentró la atención únicamente en la población lectora del Distrito Federal mexicano. Con tal propósito, se ha pensado:

1) Determinar qué circulación tiene y ha tenido el diario seleccionado y por cuáles razones económicas y sociales se puede explicar esa circulación suya en el presente y en el pasado.

2) Precisar qué estratos estadísticos de población cubre el periódico: a) en forma directa (suscriptores, compradores habituales...) y b) indirecta (lectores que no lo compran, pero que lo leen en hemerotecas, sitios públicos, etcétera).

3) Describir el lenguaje que emplea ese diario:

- 3.1 en general,
- 3.2 por secciones,
- 3.3 según redactores,

con referencia a:

- 3.01 su vocabulario,
- 3.02 su morfología y sintaxis,
- 3.03 su semántica.

4) Encontrar cuál es el lenguaje que utilizan los lectores de ese diario:

- 4.1 en general,
- 4.2 por estratos sociales,
- 4.3 por capas ocupacionales,

en referencia a:

- 4.01 su vocabulario,
- 4.02 su morfosintaxis y
- 4.03 en términos semánticos.

5) Descubrir los rasgos peculiares que presentan:

- 5.1 el lenguaje del diario, y
- 5.2 el lenguaje del público lector,

en contraste con:

- 5.01 los rasgos generales de uso real del castellano (y, en particular, de su empleo en México),
- 5.02 con las que se consideran como normas de corrección por las autoridades del idioma, para

5.03 establecer un contraste entre el lenguaje del diario y el del público.

6) Determinar una fórmula o una serie de fórmulas apropiadas para medir la efectividad comunicativa del diario (o sea, para determinar si se le entiende, si no se le entiende, si se le entiende a medias, o si se le malentiende).

7) Determinar la ineficacia relativa del diario frente a un público potencial, constituido: a) por el resto de la población que no lee (en general) y b) por un público lector que, aunque lee, no lo prefiere a otros diarios (en especial), a fin de establecer las imputaciones causales (sociolingüísticas) pertinentes.

8) Determinar la política que en materia lingüística podría adoptar ese diario para ampliar su circulación tanto entre el público en general como específicamente frente a ciertas capas socioculturales, ocupacionales, políticas, económicas (o de cualquier otro tipo) de la población mexicana.

9) Determinar la política general capaz de aumentar su circulación y efectividad en cuanto medio de comunicación.<sup>3</sup>

Para poder cubrir los apartados 3, 5 y 8, se hace necesario analizar el léxico del periódico y, en cuanto el volumen de éste es muy considerable, lo más pertinente es realizar ese análisis por medios electrónicos.

En efecto, si bien —en apariencia— no es complicado analizar un periódico, en la realidad esta tarea no es sencilla y sí es tardada.

Hay, así, muchos puntos que no se han considerado y que necesitan precisarse antes del recuento, así como otros que se refieren a si éste se debe hacer en forma manual o de manera mecánica. Entre los primeros se encuentran varios problemas de definición de las unidades por contar, el tratamiento de los homónimos, etcétera,<sup>4</sup> a los que no nos referiremos concretamente en esta nota. Respecto de los segundos, hay que decir que, si bien existen técnicas ya establecidas para utilizar la computadora en tareas lexicológicas, es a las dificultades *más inmediatas*, de quien ni se mueve en un medio de especialistas en cómputos ni en otro de especialistas en lenguas, a lo que queremos dedicar esta modestísima nota.

La labor que se emprende no se ha intentado o casi no se ha intentado en México, no ya en relación con el léxico de un periódico sino que ni siquiera en lo que se refiere a la elaboración de un diccionario de frecuencias o léxico estadístico del español tal como se emplea en el país pues, a pesar de que Antonio Carrillo Flores, durante su periodo como Director anunció que el Fondo de Cultura Económica emprendería una empresa como ésta —meritísima— no hay indicios de que la misma esté en vías de realizarse. En cambio, sí se ha informado, más recientemente, que esa tarea la está realizando ya El Colegio de México y que espera tenerla terminada en los próximos años.

Por lo menos hasta ahora, no tenemos conocimiento de otros esfuerzos de este tipo (y, en su caso, agradeceríamos se nos sacara de nuestra ignorancia). Ello nos coloca en un campo virgen, lleno de dudas, interrogantes, riesgos y problemas. Es eso lo que nos ha hecho pensar que convenía escribir una breve nota indicando cuáles han sido los tropiezos que ya hemos tenido al intentar hacer este tipo de análisis y de recuento, así como cuáles han sido las soluciones provisionales que les hemos dado a esos problemas, y esto con dos fines: 1o.) con el de que los más capaces juzguen de ellos y nos sugieran modos de evitar los unos y mejorar las otras y 2o.) con el de que, por otro lado, las personas que piensen realizar estudios similares tengan, así, una modesta experiencia previa de la que partir.

### *Algunas dificultades de carácter técnico*

Para poder hacer uso de la computadora, hay que idear la forma de vaciar la información léxico-periodística en tarjetas. Por desgracia, como lo más común en nuestro medio es que perforen números y no palabras, nos encontramos con un doble problema, que proviene —en gran parte— de las rutinas de trabajo desarrolladas por la programación originaria, y de la falta de colaboración entre diversos especialistas. Estos son: por un lado, el de la falta de costumbre que tiene el personal de perforación, en cuanto a llevar a cabo este tipo de labor; por el otro, el que resulta del manejo del periódico —fuente de información— ya que para el estudio hay que copiar el texto de los artículos que en él aparecen.

Para resolver estas dificultades, se elaboró lo que llamamos “formato de perforación”, mediante el que intentamos facilitar el manejo de los artículos. Para elaborar dicho formato: 1o.) tomamos en cuenta los datos que sirven de control; 2o.) con fines analíticos, distinguimos el título del texto del artículo, y 3o.) dedicamos varias de las columnas de cada tarjeta a fines de comprobación. Así en caso de que ocurra algún contrat tiempo, se podrán volver a ordenar fácilmente las tarjetas perforadas.

El formato consta de las tres páginas que describiremos en seguida:

#### PRIMERA PÁGINA DEL FORMATO:

Esta primera página viene a ser la tarjeta de control. Para que sirva a este fin, se consignan en ella las columnas que se dedican a cada uno de los apartados, como se indica a continuación, y de conformidad con los datos que se requieren:

- 1) *título de la investigación*; abreviado, no se perfora;
- 2) *número del artículo*; previa numeración, se le dan las columnas de la primera a la cuarta (1-4);

3) *características del artículo* (incluidas aquellas que consideramos importantes):

- a) para la fácil identificación del artículo;
- b) para poder reconstruir el periódico (tomando en cuenta la colocación que tiene el principio del artículo, dentro del periódico),  
y
- c) para realizar el análisis lingüístico.

Con este fin, se consignan:

- 3.1 el número de la tarjeta que se perfora, al cual se le dan las columnas 5-8;
- 3.2 la fecha en la que apareció el artículo, a la que se le asignan las columnas 9-14;
- 3.3 el tipo de fuente oficial (el cual interesa especialmente para el análisis sociolingüístico). Entre las fuentes, consideramos las siguientes:

3.31 agencias noticiosas:

nacionales,  
extranjeras,

3.32 Latin Reuter

3.33 fuentes gubernativas:

nacionales;  
extranjeras,

3.34 corresponsales:

en el país,  
en el extranjero;

3.35 corresponsales de agencias extranjeras o de otro periódico;

3.36 reporteros;

3.37 público;

3.38 reporteros;

3.39 otros.

A éstas se les asignan las columnas 15-16.

3.4 las agencias, en particular, pues la información sobre ellas constituye un dato de gran interés para el análisis propiamente sociolingüístico (como lo ha demostrado Jiménez de Ottalengo al relacionarlas con la tesis de la dependencia).

A éstas se les destinan las columnas 17-21.

Las agencias identificadas en el caso son:

3.41 *Associated Press* (A. P.),

3.42 *France Press* (A.F.P.),

3.43 *United Press International* (U.P.I.)

3.44 *International News Service* (I.N.S.)

3.45 *Reuter*

3.46 *Latin-Reuter*

3.5 El lugar y la materia a la que se refiere cada artículo, según la página en que se encuentra, y su lema, a ellos se les dan las columnas 22-23. Clasificamos como significativos los siguientes lugares del periódico:

- 3.51 primera plana,
- 3.52 segunda y tercera plana,
- 3.53 plana editorial,
- 3.54 plana cultural,
- 3.55 plana cinematográfica,
- 3.56 plana de radio y televisión,
- 3.57 información policiaca,
- 3.58 plana de provincia,
- 3.59 plana deportiva.

- 3.591 resto de la "Sección A",
- 3.592 resto de la "Sección B",
- 3.593 resto de la "Sección C",
- 3.594 resto de la "Sección D",
- 3.595 "Diorama de la Cultura" (que es un suplemento hebdomadario).
- 3.596 Magazine.

3.6 la clasificación; a ésta se le dan las columnas 24-25.

3.7 la subclasificación, que sirve para anotar el tipo de noticia que contiene el artículo, y a la que se le da la columna 26. Esta subclasificación obedece a criterios sociológicos, y pretende tipificar el contenido del mensaje.

SEGUNDA PÁGINA DEL FORMATO

La segunda página contiene:

- 1) título de la investigación; abreviado, no se perfora;
- 2) número del artículo; se le dan las columnas 1-4;
- 3) número de la tarjeta que se está perforando; se le dan las columnas 5-8. (En esta página, se empieza a numerar a partir del 0001).

El resto de la página se deja en blanco, para pegar el título del artículo.

TERCERA PÁGINA DEL FORMATO

Esta tercera página contiene:

- 1) título de la investigación; abreviado, no se perfora;
- 2) número del artículo que se está perforando; se le dan las columnas 1-4;

- 3) número de la tarjeta que se está perforando; se le dan las columnas 5-8 (se inicia la numeración en la tarjeta 0010).

El resto de la página se deja para pegar el artículo. En ocasiones un artículo suele requerir más de una "página tres".

Un segundo problema de carácter técnico se debe a que la perforadora: 1) no cuenta con letras minúsculas, 2) no tiene acentos y 3) carece de varias letras (por ejemplo, la ñ). Eso nos llevó a dar un código de signos-guía, que llamamos "Código de Perforación", el cual nos permite tener "la información" tal como aparece en el periódico, puesto que nos interesa llevar a cabo un análisis de tal índole que el mismo nos obliga a considerar importante cualquier signo (incluso los de admiración, interrogación, etcétera), así como el que algunas palabras tengan inicial mayúscula, u otras estén escritas todas con mayúsculas, así como otros muchos detalles similares (pequeño, pero *no insignificantes*), en particular en un medio que como el periodístico no utiliza el lenguaje en términos escuetamente racionales, sino que lo matiza, le da coloraciones emocionales y que con éstos y otros medios promueve sus propios fines de propaganda, de adoctrinación o de enseñanza (o sea, sus finalidades políticas y educativas).

El "Código de perforación" es una lista de los símbolos que se van a usar para sustituir los que faltan en la perforadora, así como los que la computadora considera como de carácter inválido. Para su elaboración, se consultó al personal del Departamento de Programación del Instituto de Investigaciones Sociales de la Universidad Nacional Autónoma de México, a cuyos miembros (en particular a Romoaldo Vitela y María de la Luz Guzmán) les agradecemos su orientación y su consejo.

Así, por medio de una estrecha comunicación entre el o los investigadores sociolingüistas y el personal de programación, se han logrado conciliar los intereses teóricos de la investigación con las posibilidades y las limitaciones de la perforación y el tratamiento y/o el procesamiento de los datos, superando así, los primeros obstáculos de este género de pesquisa.

En concreto, para evitar confusiones, se anota cuál es el signo que aparece en el artículo; se perfora y aparece en "listado". Se empieza por consignar los signos de mayor frecuencia; se sigue con los que la tienen menor, y se termina al agotar todas las posibilidades que se encuentran realizadas en el texto. Después se solucionan los casos haciendo uso del ampersand (&) para indicar signos o anotaciones especiales.

#### *Un problema psicológico de repercusión técnica*

Un tercer problema es de carácter psicológico, pero se refleja en lo técnico. Consiste en el rechazo que inicialmente manifiesta el personal perforador hacia el manejo de este tipo de información. Esto se refleja en retrasos y comisión de errores, lo cual depende de que, por lo general

—como ya dijimos— la “información”, que se vacía en tarjetas se encuentra previamente codificada; esto es, que lo más frecuente es que se perforen números y no letras mientras que, en el caso de un análisis lexicológico como el que nos interesa, se trata de palabras que deben transcribirse tal como aparecen escritas (en los artículos periodísticos, en este caso).

Ese rechazo se debe, más que nada, a que falta la costumbre de manejar este tipo de “información” y, para evitarlo hay que motivar al personal, informándolo de los fines y necesidades de la investigación; 1o.) explicándole, en forma general, cuáles son los propósitos de la misma, y 2o.) dándole tiempo para que se familiarice con el trabajo.

Del problema psicológico deriva uno técnico que ya señalamos, y que consiste en la comisión de errores, lo cual es muy comprensible. En cuanto el personal que opera la computadora es humano, psicológicamente es vulnerable y, por ello, comete errores.

A los errores cometidos los consideramos un problema técnico, ya que su solución es de este orden; pero, su prevención y eliminación dependen de procedimientos psicológicos importantes en la colaboración científica interdisciplinaria, y un ejemplo de ello es la motivación a la que ya aludimos.

Para poder controlar los errores cometidos y corregirlos, procedemos a listar los artículos que ya están perforados. Para ello se procede por etapas. Estas son:

Primera: la de hacer un listado de cierto número de artículos.

Segunda: la de verificar los mismos, confrontando los listados con los artículos (de los formatos), anotando cuáles fueron los errores cometidos, para corregirlos en el listado mediante el empleo de un lápiz rojo.

Tercera: la de corregir las tarjetas perforadas, para eliminar errores.

### *Sugestiones prácticas para emprender con éxito el tratamiento electrónico del léxico periodístico*

Con la experiencia obtenida, podemos decir que un trabajo de esta índole revela la necesidad de contar con un equipo de trabajo: a) convenientemente motivado, b) bien entrenado, c) homogéneo, con el que se pueda contar de principio a fin de la pesquisa. También sugerimos:

1. Planear el trabajo con anticipación, particularmente en términos de costo y de tiempo, tomando en consideración:

- 1.1 las posibles descomposturas de la máquina,
- 1.2 la falta de tarjetas,
- 1.3 la incapacitación temporal del personal por enfermedad y
- 1.4 las interrupciones debidas a los días festivos.

2. Elaborar un manual escrito (o proceder en alguna otra forma que



se crea conveniente) para consignar las observaciones pertinentes (código, formato, etcétera).

3. Llevar el control.
  - 3.1 del trabajo diario realizado,
  - 3.2 de la asistencia del personal,
  - 3.3 de su trabajo individual (ya sea diario o semanal),
  - 3.4 de las tarjetas que se perforen por día.
4. Dar a cada perforista una parte del trabajo, para que se responsabilice de ella en forma individual.
5. Mantener una continua comunicación entre el investigador y el equipo de programación.

En una labor tan larga e ingrata como ésta, los mencionados son apenas los primeros pasos de la pesquisa; pero si éstos no se intentan y si no se trata de que sean firmes (mediante el sometimiento de notas tan modestas como la nuestra, a la consulta y comentario de la comunidad científica de la que formamos parte) correremos el riesgo de no emprender nunca la marcha o de tropezar tarde o temprano en una senda que, a largo plazo es sociológica, pero que comienza por ser lingüística (lexicológica) y, que en sus términos más escuetos, es una labor modesta de recuento de las palabras empleadas por uno de los grandes difusores de nuestras sociedades contemporáneas.<sup>5</sup>

<sup>1</sup> Regina Jiménez de Ottalengo: *La Prensa Nacional, su Lenguaje y su Adecuación Sociopolítica y Pedagógica a la realidad de México*. Algunas consideraciones generales. Versión mimeografiada. I.I.S., México, 1974.

<sup>2</sup> Regina Jiménez de Ottalengo: "El Periódico como medio de comunicación colectiva y su Estudio Interdisciplinario". *Revista Mexicana de Sociología*. Año 1973. Vol. 3. Núm. XXXV.

<sup>3</sup> Vide R. Jiménez de Ottalengo: "El Periódico..." ya citado.

<sup>4</sup> Algunas de las formas de resolver estos problemas se encuentran, por ejemplo en Alphonse Juilland y E. Chang Rodríguez: *Frequency Dictionary of Spanish Words*, Mouton & Co. London-The Hague-Paris, 1964. Parte introductoria.

<sup>5</sup> Para nuestros recuentos pensamos inspirarnos en un estudio aparentemente lejano, pero que creemos nos será útil en cuanto es uno de los pocos en este campo: Barker H. J., Hamdani and K. M. S. Dihlavi. *An Urdu News paper Count*, publicado en 1970, que es un estudio estadístico de las frecuencias con que aparecen las palabras en un periódico de India, publicado en lengua urdu.