

DICCIONARIO, SOCIEDAD Y COMPUTACIÓN

*Jorge Antonio Leoni de León**

RESUMEN

En este artículo abordamos uno de los cambios en el tejido social que nos obliga a poner en perspectiva nuestra labor científica. Internet revela una increíble cantidad de usos lingüísticos. Este hecho conlleva serios retos para los lingüistas: ¿Cómo representar? ¿Cómo debemos tratar con estos datos global? La lexicografía siempre ha lidiado con la diferencia entre conocer las palabras y comprenderlas. La lexicografía contemporánea asume una cierta visión de la lógica interna del lenguaje que se interroga por la lexicografía como una ciencia que intenta al mismo tiempo describir, explicar y reproducir los fenómenos que estudia.

Palabras clave: sociedad y computación, representatividad en lingüística, pertinencia teórica, lexicografía, adecuación explicativa, capacidad explicativa.

ABSTRACT

Here we treat one of those changes in the social weave that forces us to have a different way of thinking. Internet reveals an amazing amount of linguistic uses. This fact brings serious problems for linguists: What is representative? How we should deal with this global data? Lexicography has always dealt with words, with the difference between know them and understand them. Contemporary Lexicography thinks about the internal logic of language. It questions itself because Lexicography as a science intends to describe and explain in order to reproduce the phenomena.

Key Words: society and computation, representativeness in linguistics, theoretical pertinence, lexicography, explanation adequacy, explanation capacity.

1. Introducción

La computación modifica la sociedad y, al mismo tiempo, también nuestra visión sobre nosotros mismos se transforma. Esta afirmación peca de poco original cuando Google¹ ya es un verbo inglés² y la ausencia de una página personal en Facebook³, para muchos, equivale a no existir; es más, no es raro que los intercambios en Facebook y Hi5⁴ sustituyan a una vida social « off line ».

Sin embargo, no es nuestra intención referirnos a este tipo de cambios, cuyo valor en la evolución de la especie todavía está lejos de ser aprehendido; no, realmente, lo que pretendemos es abordar aquellas alteraciones en el tejido social que nos obligan a pensar diferente simplemente porque nuestras “evidencias” son diferentes

o también aquellas que nos permiten retomar antiguos conceptos porque, ahora sí, disponemos de los recursos⁵ para ponerlas en práctica.

2. En torno a “tejido social” como concepto

Sigamos a Saussure (1972, originalmente publicado en 1916) por un momento y admitamos la dimensión social del lenguaje en detrimento de sus aspectos psicológicos (Pennebaker, Mehl y Niederhoffer 2003, por ejemplo). Esto, a algunos, les puede parecer una evidencia, pero a la luz de los dos elementos mencionados anteriormente, sociedad y computación, ahora podemos darle una nueva interpretación a un término mencionado, muy rápidamente, al inicio

* Profesor de la Escuela de Filología, Lingüística y Literatura, Universidad de Costa Rica.
Recepción: 15/03/11. Aceptación: 02/05/11.

de esta charla: tejido social. En efecto, siendo saussureanos, podemos concebir el tejido social como un hecho lingüístico, más que sociológico. Llevando este concepto un poco más lejos tenemos que admitir dos cosas: primero, nos interesa la manifestación de ese tejido en un momento determinado de su existencia, el actual y, segundo, esas manifestaciones nos interesan en tanto expresiones de una colectividad.

3. Los datos y el saber

Ahora bien, una ojeada al web nos revelará, muy rápidamente, una gran variedad de usos lingüísticos, sobre los cuales, podemos lanzar la hipótesis de que una cantidad no despreciable es representativa de muchos hablantes: nunca las Ciencias del Lenguaje habían estado confrontadas a tal cantidad de datos en tan poco tiempo, nunca habíamos tenido tanta necesidad de catalogar tales usos, ni habíamos estado tan abrumados por la tarea de levantar un inventario lingüístico representativo por simple que sea. Pero: ¿Qué es lo representativo? Estadísticamente lo representativo sería una muestra mínima de una población a partir de la cual sea posible extraer generalizaciones. Sin embargo en Lingüística, tradicionalmente lo representativo se limita al lingüista o a un grupo de “consejeros lingüísticos”, generalmente muy reducido (muchas veces sólo es una persona). Lo representativo continúa siendo un problema constante en Lingüística y la actual dispersión de las comunidades no hace más que complicar su definición: cuando Saussure postula la realidad social de la lengua (Saussure, 1972: 160), es decir, su imposibilidad de existir fuera de un grupo social, probablemente tenía en mente la idea de grupos homogéneos de hablantes nativos, más bien cercanos unos a otros. Pero ahora, en la época del *inglés global*⁶ todos pertenecemos a múltiples comunidades lingüísticas, lo queramos o no, por lo que un frente se abre en nuestra disciplina: los datos globales. ¿Debemos filtrarlos? ¿Debemos analizarlos en su conjunto? Cualquiera que sea la respuesta, ésta siempre será una que necesariamente implique el abordaje de los datos de alguna manera: su negación o su

corrección no podrán ser considerados como actitudes científicas.

Pero una cosa es *saber algo* y otra *interiorizarlo*. Una de las primeras disciplinas lingüísticas en comprenderlo fue la Lexicografía. No es un azar que al comenzar las experiencias computacionales con el lenguaje, en la década de 1950, estas tuvieran una fuerte vocación léxica. Con cierto romanticismo podríamos ver aquí un paralelismo con nuestros primeros intentos por aprender otra lengua, los cuales pasan primero por memorizar palabras, sin desarrollar aún la intuición de cómo organizarlas con sentido. Sea como sea, esos primeros intentos iniciaron una lenta revolución: significaron el paso de una Lexicografía erudita, en la que se privilegiaba el conocimiento del lexicógrafo y se soslayaba el lenguaje en su realidad, es decir, en su dimensión social. Esta revolución no significó la desaparición del erudito, pero sí implicó el paso de una concepción etimológica de la ciencia (como “conocimiento”, del latín *scientia*) a una visión sistemática del conocimiento. Digámoslo de otra forma, lo que sabemos acerca de las “palabras”, como excusas anecdóticas, como *entidades*, pasa a segundo plano cuando lo que buscamos son los principios o reglas que las rigen. Este cambio no fue producto de una reflexión erudita sobre el lenguaje, fue el resultado de la necesidad: las nuevas herramientas automatizadas exigían nuevos métodos. En efecto, de repente ya no bastó con plasmar lo que el científico creía, sino que se debía recuperar (o extraer) la palabra en su contexto, en sociedad, a través de los libros, de los periódicos, de las revistas y más recientemente incluso a través de los mensajes telefónicos de texto, es decir, en el uso real del lenguaje. Podemos dejar para otro momento las condiciones políticas y financieras que están detrás de todo esto, concentrándonos más bien en la evolución de esta disciplina y su influencia en otras áreas.

4. La lengua: ¿asunto de estado o asunto de lógica?

Si bien es cierto que poco a poco, en los últimos cincuenta años, pasamos de describir un

estado « perfecto » de la lengua, para imponer o sugerir maneras de expresarnos en ella, a otro en el que nos interesa más bien encontrar la lógica del lenguaje, no ha habido, y muy probablemente no veremos nunca, la desaparición de la manera tradicional de trabajar en Lexicografía: todavía el esquema del erudito sigue siendo utilizado, pero éste se encuentra en un área vecina a la Ciencia, no dentro de ella. Tenemos un ejemplo de esto en la Real Academia (Real Academia Española, 2006), la cual describe un estado de lengua considerado como ejemplar, generalmente lo más cercano posible a la lengua literaria (en su mayor sentido de institucionalización), a partir del cual se establece un modelo ideal para todos los hablantes, actitud resumida en el célebre lema « Limpia, fija y da esplendor ». Esto no significa, de ninguna manera, que la Real Academia Española (2006) no aproveche los recursos científicos, sino que sus objetivos originales fueron desviados en favor sus propias metas institucionales.

Pero, ¿en qué consiste esa científicidad de la lexicografía contemporánea? En primer lugar, se trata de un abandono de la subjetividad, ya no van a interesarnos nuestros recuerdos o creencias sobre “cómo decimos algo”, sino más bien, buscaremos fuentes en el mundo real. Esto es, crearemos un corpus. Luego, el valor ontológico de los términos estará dado por su existencia y su contexto en el corpus y no por las preferencias del lexicógrafo: su peso está, entonces, determinado por el lugar que ocupa en el tejido social, que ya definimos más arriba. Encontramos aquí un terreno fértil para la Metalexigrafía: ¿Cómo se construye un corpus? ¿Cómo se alimenta? ¿Cómo se escogen los ejemplos? ¿Cómo se determinan los sentidos? La respuesta a cada una de estas preguntas puede tomarnos meses enteros de deliberaciones. Bástenos entonces con mencionarlas.

5. Ser y representación

Queda, sin embargo, una pregunta a la que todo lingüista y, en particular, todo lexicógrafo deben encontrar respuesta: ¿Cómo representamos la palabra?⁷ Aunque el espacio no nos permite abordar esta interrogante en todas

sus dimensiones, sí podemos, cuando menos, señalar los rumbos que hay que recorrer para poder satisfacerla. En primer lugar es necesario definir el concepto de “palabra”, tomando en consideración sus aspectos prosódicos, morfológicos, sintácticos y semánticos. Las siguientes expresiones manifiestan tratamientos diferentes según el nivel del lenguaje desde el que estemos hablando: “Dámelo”, es una sola unidad prosódica y morfológica, pero sintáctica y semánticamente se trata de tres unidades distintas; “ha venido”, es una unidad semántica, pero no necesariamente se trata de una sola unidad prosódica y sintáctica; “colgó las tenis” es un único valor semántico, pero sintácticamente consiste en tres sintagmas⁸.

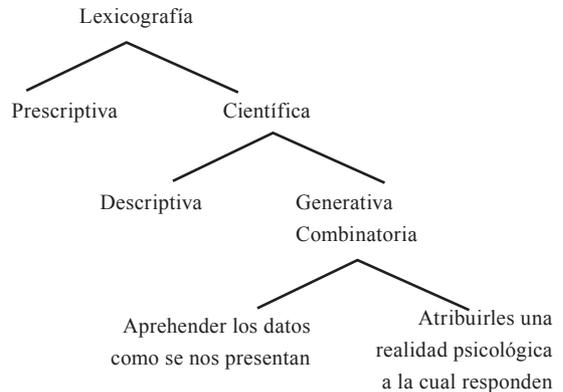
Sin embargo, hasta ahora sólo hemos señalado el problema desde la capacidad meramente denotativa de las “palabras”. ¿Qué hay sobre su comportamiento? Veamos por un momento los problemas que surgen cuando queremos aprehender las palabras. Así, por ejemplo, podemos considerar las palabras según las pronunciamos, las construimos o las organizamos. También podemos pensar en cuándo las utilizamos (valores geográficos y sociolectales, por ejemplo) o en qué niveles de realización⁹ tienen. También hemos de superar un problema: ¿Cómo reconocer (automáticamente) las unidades pluriverbales? Este problema es particularmente difícil si tomamos en cuenta la complejidad sintáctica que caracteriza a muchas de esas construcciones. Una vez reconocidas las unidades volvemos a nuestra pregunta anterior: ¿Cómo representarlas? En este contexto, este aspecto constituye un verdadero reto para la Lexicografía, porque ya no se trata de describir el significado de una entrada léxica, sino que además tenemos que disponer de un mecanismo formal que permita la reproducción de dicho comportamiento.

Pero una serie de problemas persiste, como si no fuera suficiente con los señalados hasta ahora. ¿Nuestra representación responde a una motivación social o psicológica? No han faltado quienes, los más por ahora, den una respuesta probabilística: no hay representación, sino probabilidades. Sin embargo, el problema

científico persiste. Ahora bien, aquí es a dónde queríamos llegar: Esta interrogante sobre la naturaleza de las palabras y su representación tiene varias aristas, que no podemos ignorar. Ninguna de ellas debe necesariamente primar sobre las otras, sin embargo, en nuestra busca debemos saber en qué ámbito nos encontramos. En este punto conviene recordar lo señalado al principio: “pensar diferente porque nuestras evidencias son diferentes”. ¿Qué tan diferentes pueden ser nuestras evidencias? A priori, no es importante, por cuanto ya no será pertinente, desde el punto de vista de los datos, lo que queramos ver, pero sí va a ser relevante que no imponamos a la realidad nuestra visión: somos nosotros los que debemos aprender a “escuchar” las evidencias y confrontar continuamente nuestro modelo conceptual, no para juzgar las evidencias, sino para perfeccionar nuestros métodos, nuestras propuestas teóricas. No obstante, un nuevo reto heurístico aparece, el cual nos brinda pistas sobre cómo representar las palabras en su contexto: ¿Cómo interpretar esas evidencias? ¿Cómo valorar los datos?

Planteada esta última interrogante, varias respuestas se perfilan, todas sintetizadas en la Figura 1. En primer lugar, podemos considerar la lengua como un modelo al que todo hablante debe aspirar. Estamos en una perspectiva prescriptiva, que descartamos, porque lo que buscamos es conocimiento sistematizado, lo que equivale a la diferencia entre saber dónde están los planetas y cómo se desplazan, con respecto a dónde nos gustaría que estén y cómo, entonces, debemos hacer prueba de creatividad para justificar sus posiciones. En segundo lugar, si nuestro interés no es imponer o sugerir una forma de expresión, sino más bien saber “cómo funcionan”, entonces nos encontraremos en el ámbito científico. Sin embargo, las formas de hacer Ciencia son muchas, podemos simplemente pretender describir un fenómeno, pero también podemos tener como objetivo la posibilidad de reproducirlo.

La reproducción de un fenómeno lingüístico se entiende en un sentido formal con vistas a su modelización por medio de un sistema de procesamiento del lenguaje natural



que posibilite la comprobación de la certeza del método y la plausibilidad de la representación. Aquí aún son posibles dos posiciones: podemos considerar los fenómenos como el resultado de un sistema lingüístico interiorizado, psicológico, o bien es posible estimar que debemos abordar el lenguaje a partir de su realización: el habla. Aquí los caminos se bifurcan y las representaciones que proponamos se van a distanciar en función de la posición adoptada.

6. Advertencias

Todo esto no significa la muerte del erudito (y, por extensión, del filólogo), sino que sólo anuncia un cambio en su manera de realizar su trabajo: ahora también debe comunicar sus resultados a toda una comunidad menos especializada, ávida de informaciones compactas utilizables inmediatamente. Por otra parte, un movimiento contrario se perfila sobre el que es necesario estar atentos: la ausencia de un dato en la web no implica su inexistencia, por lo que el acto de ignorarlo por dicho motivo no está *teóricamente justificado* (otra historia es si se justifica *metodológicamente*).

En síntesis, en tanto lexicógrafos y lingüistas, tomamos decisiones teóricas y metodológicas con serias consecuencias en nuestro trabajo científico, las cuales debemos asumir conscientemente desde el inicio de la investigación, con plena lucidez en cuanto a la nueva complejidad de las estructuras sociales, muchas veces sometidas a la influencia de

mundos virtuales que apenas comenzamos a aprehender.

Notas

1. Dirección: <http://www.google.com> .
2. Por ejemplo, en la edición en línea del diccionario Merriam Webster (<http://www.merriam-webster.com/dictionary/google>).
3. Dirección: <http://www.facebook.com> .
4. Dirección: <http://www.hi5.com> .
5. Y cuando decimos recursos, nos referimos a las técnicas intelectuales más que a las herramientas materiales propiamente dichas.
6. A veces llamado globish por la lengua artificial homónima (<http://www.globish.com/>).
7. Para una discusión sobre este tema, recomendamos leer Leoni de León (2008).
8. Leoni de León, Schwab y Wehrli (2008); Wehrli (2007); Nerima, Seretan y Wehrli (2006); Wehrli(2004), abordan varios aspectos de los problemas que plantean las locuciones y las colocaciones.
9. Leoni de León (2008).

Referencias

Leoni de León, Jorge Antonio. 2008. « *Modèle d'analyse lexico-syntaxique des locutions espagnoles* ». Tesis doct. Ginebra, Suiza: Université de Genève. URL: <http://www.unige.ch/cyberdocuments/theses2008/LeonideLeonJA/meta.html>.

Leoni de León, Jorge Antonio, Sandra Schwab y Éric Wehrli. 2008. « *Análisis sintáctico profundo del español: un ejemplo del procesamiento de secuencias idiomáticas* ». En: *Procesamiento del*

Lenguaje Natural. Ed. por Paloma Martínez, Fernández, Dolores Cuadra Fernández y F. Javier Calle Gómez. 41. Sociedad Española para el Procesamiento del Lenguaje Natural, Departamento de Informática, Universidad de Jaén. Jaén, págs. 37-44. URL: <http://www.sepln.org/revistaSEPLN/revista/41/secl-art5.pdf>.

Nerima, Luka, Violeta Seretan y Éric Wehrli. 2006. « *Le problème des collocations en TAL* ». En: *Nouveaux cahiers de linguistique française* 27, págs. 95-115. ISSN: 1661-8246. URL: <http://www.latl.unige.ch/personal/vseretan/>.

Pennebaker, James W., Matthias R. Mehl y Kate G. Niederhoffer. 2003. « *Psychological Aspects of Natural Language Use: Our Words, Our Selves* ». En: *Annual Review of Psychology* 54. 1, págs. 547-577.

Real Academia Española. 2006. La Real Academia Española. Sitio web. [Página web: <http://www.rae.es/>. Madrid.

Saussure, Ferdinand de. 1972. *Cours de linguistique générale*. Paris: Payot.

Wehrli, Éric. 2004. « *Un modèle multilingue d'analyse syntaxique* ». En: *Structures et discours : Melanges offerts à Eddy Roulet*. Ed. por Antoine Auchlin y col. Langue et pratiques discursives. Canada: Éditions Nota bene, págs. 311-332.

_____. 2007. « *Fips, A "Deep" Linguistic Multilingual Parser* ». En: *ACL 2007 Workshop on Deep Linguistic Processing*. Prague, Czech Republic: Association for Computational Linguistics, págs. 120-127. URL: <http://www.aclweb.org/anthology/W/W07/W07-1216>.

