

## METODOLOGÍA PARA EL PROCESAMIENTO DE LAS PREGUNTAS DE LÉXICO

Fernando CASTAÑOS  
Julia Isabel FLORES

### *Obtención de los índices de disponibilidad léxica*

1. El objetivo inicial del procesamiento de las preguntas de asociación libre —preguntas 14 a 27 de la sección del cuestionario de representaciones sociales y pregunta 37 en el cuestionario general para la palabra “*Constitución*”— es obtener un vector de dos posiciones, con las frecuencias de mención en primero y en segundo lugar. Ejemplo: “*renovación*” (5, 16).

La palabra *renovación* apareció 5 veces en primer lugar y 16 veces en segundo lugar.

2. El objetivo final es obtener un índice de disponibilidad que se calcula con base en dichas frecuencias.

3. Es necesario someter las respuestas a un proceso de normalización, que incluye su captura en minúsculas y sin acentos. Se consideran como variantes de la misma palabra, y se cuentan como la misma palabra:

*a)* Las formas escritas con distinta ortografía o distinta tipografía (incluyendo mayúsculas y minúsculas) que representan la misma palabra. Ejemplo: *Democracia=democrasia*.

*b)* Las variaciones de género y número que corresponden a una misma raíz, siempre y cuando se conserve la clase de palabras (sustantivo o adjetivo). Ejemplo: *nuevo=nueva=nuevos* (novedad no se considera variante de la misma palabra, puesto que es sustantivo y las anteriores eran adjetivos).

*c)* En el caso de verbos, se consideran la misma palabra todas las variantes de modo, tiempo y persona que correspondan a la misma raíz. Ejemplo: *renovar=renovó=renovaron*.

*d)* En el caso de que la respuesta sea una frase, registrar el núcleo de la misma. Eliminar artículos, preposiciones, conjunciones. Ejemplo: “*para progresar*”, registrar “*progresar*”.

4. Si hay frases u oraciones muy complejas emplear un código que indique que no se pudo registrar. Ejemplo: “*la nueva situación que empieza*”.

Una vez que se normalizó la ortografía, las palabras fueron agrupadas por familias léxicas y se obtuvieron sus frecuencias. Se ponderaron las cifras por medio de un procedimiento estadístico que otorga una mayor ponderación a las menciones en la primera posición que a las menciones en segunda o tercera posiciones.

Para construirlo se reelaboró la fórmula de López Chávez y Strassburger (1987)<sup>51</sup> modificada por Fernando Castaños<sup>52</sup> para obtener un “índice de disponibilidad” que, en principio, puede variar entre 0 y 1. Para proporcionar una idea del significado del índice pueden considerarse

<sup>51</sup> Véase López Chávez, Juan y Strassburger, Carlos, “Otro cálculo del índice de disponibilidad léxica”, ponencia presentada en el coloquio Lingüística Computacional, organizado por la Asociación Mexicana de Lingüística Aplicada y El Colegio de México, 1987.

<sup>52</sup> Véase una reelaboración del índice de disponibilidad léxica en: Castaños, Fernando, “Comunicación, sujetos, lenguaje”, *Los mexicanos de los noventa: una encuesta nacional de actitudes y valores*, México, IISUNAM, 1996.

algunos ejemplos. Si un descriptor no fuera seleccionado por ningún encuestado, tendría un índice de 0; si fuera la primera opción de todos los encuestados tendría un índice de 1; si fuera la segunda opción de todos los encuestados tendría un índice de 0.42; y si fuera la primera opción de la mitad de los encuestados su índice sería de 0.5; pero si fuera seleccionado por el 25% de los encuestados como primera opción y el 25% como segunda opción, tendría un índice de 0.36.

La fórmula utilizada es:

$$D = \sum_{i=1}^n e^{-c \frac{(i-1)}{(n-1)} \frac{f_i}{I}}$$

En donde  $D$  significa “disponibilidad”. El índice  $i$  es el número de la posición en que ocurre la respuesta analizada (la palabra asociada o el descriptor elegido). El índice  $n$  es la máxima posición alcanzada; en nuestro caso este número es igual a 3. El coeficiente  $c$  es un factor de dispersión; López Chávez y Strassburger recomiendan igualarlo a 2.3. El numerador  $f_i$  es la frecuencia absoluta de la respuesta analizada.  $I$  es el número de informantes que respondieron la pregunta.