

Rompiendo la Ley de Zipf

Breaking Zipf's Law

Alfonso Enrique Ramírez Sanabria

Director, Grupo de Investigación en Catálisis, Profesor Titular, Departamento de Química, Universidad del Cauca, Popayán, Colombia.



Para citaciones: A. Ramírez Sanabria, "Rompiendo la Ley de Zipf", *Ing-Nova*, vol. 3, no. 2, pp. 181-183, Jul. 2024.
<https://doi.org/10.32997/rin-2024-4949>

Recibido: 1 de junio de 2024
Revisado: 10 de junio de 2024
Aprobado: 10 de junio de 2024

Autor de correspondencia:
Alfonso Enrique Ramírez Sanabria
aramirez@unicauca.edu.co

Editor: Miguel Ángel Mueses. Universidad de Cartagena-Colombia.

Copyright: © 2024 A. Ramírez Sanabria. Este es una editorial de acceso abierto, distribuido bajo los términos de la licencia <https://creativecommons.org/licenses/by-nc-nd/4.0/> la cual permite el uso sin restricciones, distribución y reproducción en cualquier medio, siempre y cuando que el original, el autor y la fuente sean acreditados.



Tópico

En un artículo anterior, hacíamos una apología al proceso de escritura mostrando como realizar, más a menudo, esta actividad; puede servir como un elixir para la eterna juventud [1]. Sin importar sobre lo que escribamos, tal vez si sea más importante la forma en como lo hacemos y deberíamos, al escribir, romper, más a menudo, la regla de Zipt. Me explico:

La llamada ley de Zipf, fue formulada en la década de 1940 por George Kingsley Zipf, lingüista de la Universidad de Harvard, es una ley empírica según la cual, en una determinada lengua la frecuencia de aparición de distintas palabras sigue una distribución que puede aproximarse por la siguiente ecuación:

$$P_n \approx 1/n^a \quad (1)$$

donde P_n representa la frecuencia de la n -ésima palabra más frecuente y el exponente a es un número real positivo, en general ligeramente superior a 1 [2].

¿Qué significa todo esto? Podemos describirlo de dos maneras. La manera "complicada", que la ecuación lo que muestra es que hay un patrón en la repetición de palabras, en donde la frecuencia de una palabra es inversamente proporcional a su rango estadístico y este depende del número de palabras diferentes. La manera "práctica", que las palabras más utilizadas tienen un patrón de aparición, que es el siguiente: La palabra más utilizada, que ocuparía el primer puesto en el ranking, sería el doble de veces más utilizada que la segunda más utilizada, el triple de veces que la tercera, el cuádruple de veces que la cuarta, y así sucesivamente [3].

La ley de Zipf se cumple para la mayoría de las lenguas, ver Figura 1.[4]

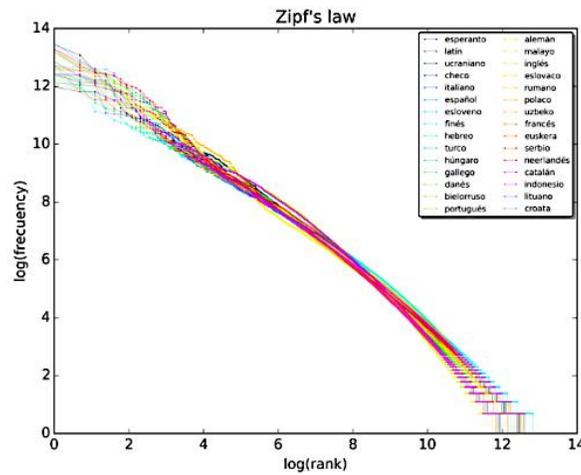


Figura 1. Frecuencia de la Ley de Zipf en diferentes lenguas.

Si examinamos la versión original de "El origen de las especies" escrita por Charles Darwin (en inglés en 1859) vemos que la palabra más utilizada en el primer capítulo es "the", con una aparición de cerca de 1.050, mientras que la segunda es "and", apareciendo cerca de 400 veces, y la tercera es "to" apareciendo unas 300. Aunque no de forma exacta, se puede ver que la segunda palabra aparece la mitad de las veces que la primera ($1050/400 = 2.6$) y la tercera un tercio ($1050/300 = 3.5$).

Estas palabras tienen algo en común, son palabras que por si solas carecen de significado, pero son las que se encargan de darle estructura gramatical a la frase. Al decir que carecen de significado queremos indicar que, si se dice una frase en la que no hay ningún sustantivo, adjetivo, verbo ni adverbio, la frase carece de sentido. Por ejemplo:

... y en una ... de al ... de

En cambio, si sustituimos los puntos por palabras con significado, podemos tener una frase como la siguiente.

Miguel y Ana tienen **en** su casa **una** mesita **de** color marrón **al** lado **de** su cama.

Este fenómeno ocurre en otras lenguas y se hace interesante pensar, en cómo el cerebro humano procesa el lenguaje [3]. ¿Cómo lo procesa? Uhhmm, tal vez Robert Sapolsky pueda ayudarnos. En su libro *Compórtate*, Sapolsky afirma que los estímulos inconscientes con palabras influyen en nuestro comportamiento, que las palabras cambian inconscientemente pensamientos y sentimientos, que las palabras son poderosas y que pueden salvar, curar, inspirar, devastar, desalentar y matar [5]. Un ejemplo se da en juego experimental económico "El dilema del prisionero" en el que los participantes deciden cooperar o competir en varias coyunturas [5]. El comportamiento se "altera" por "etiquetas situacionales": por ejemplo, si llamas al juego "el juego de Wall Street" la gente se vuelve menos cooperativa, pero si lo llamas "el juego de la comunidad" ocurre lo opuesto. Algo parecido sucede si los sujetos leen una lista de palabras aparentemente al azar antes de jugar: Si incrustamos en esa lista palabras cálidas y amables -ayuda, armonía, justicia, mutuo- se fomenta la cooperación, mientras que palabras como -rango, poder, fiereza, desconsideración- fomentan lo opuesto. Una persona es un "terrorista" para unos y para otros en un "luchador por la libertad" [5].

Otro ejemplo, es el de las señales lingüísticas en asuntos vitales y sus consecuencias. Para una misma intensidad, a los huracanes que se les ha dado arbitrariamente un nombre femenino, matan a más personas que aquellos que reciben un nombre masculino ¿por qué? La gente toma inconscientemente los nombres masculinos de los huracanes más seriamente y se tienen más probabilidades de obedecer las órdenes de evacuación [7].

Daniel Kahneman y Amos Tversky, ganadores del premio nobel en economía en 2002 [8], demostraron que la estructuración de las palabras altera la toma de decisiones, por ejemplo, si a la gente, incluidos los médicos, se le dice: "El fármaco tiene un 95% de índice de supervivencia" es más probable que la gente, apruebe su uso, en cambio sí se les dice: "El fármaco tiene un índice de mortalidad del 5%", hay mucha menos probabilidad de que sea aprobado.

Conclusión

Empezamos este escrito recordando uno anterior en donde se invitaba a realizar más a menudo el proceso de escritura, ahora en este, invito a que rompamos más a menudo la ley de Zipf y que en esos escritos tengamos con mayor frecuencia las palabras correctas para hacer de nuestro mundo, un mundo mejor; más cooperativo, más bondadoso, más solidario y por qué no, más en equilibrio. Que empleemos, las 10 palabras más usadas de nuestro idioma¹ para aportar a una sociedad mejor.

Referencias

- [1] A.E. Ramírez. 2023. *Innovación y Ciencia*. Volumen XXXIII (1). El secreto de la vida eterna.
- [2] M.A. Montemurro. 2001. *Physica A: Statistical Mechanics and its Applications* 300: 567-578. Beyond the Zipf-Mandelbrot law in quantitative linguistics.
- [3] Nahum Montagud Rubio. 2020. Ley de Zipf: un curioso fenómeno social y matemático. Tomado de: <https://psicologiyamente.com/cultura/ley-zipf>. Consultado el 5 de mayo de 2024.
- [4] Ley de Zipf. Tomado de: https://es.m.wikipedia.org/wiki/Ley_de_Zipf. Consultado el 29 de diciembre de 2023.
- [5] R. Sapolsky. *Compórtate. La biología que hay detrás de nuestros mejores y peores comportamientos*. 2017. Capitán Swing Libros, S.L., Madrid.
- [6] El dilema del prisionero. Tomado de: https://es.wikipedia.org/wiki/Dilema_del_prisionero. Consultado el 21 de febrero de 2024.
- [7] K. Jung et al., 2014. *Proceedings of the National Academy of Sciences*. 111, 8782. Female hurricanes are deadlier than male hurricanes.
- [8] The nobel prize. Tomado de: <https://www.nobelprize.org/prizes/economic-sciences/2002/kahneman/facts/>. Consultado el 5 de mayo de 2024.

¹ Las diez palabras más usadas en el idioma español son: 'que', 'de', 'no', 'a', 'la', 'el', 'es', 'y', 'en' y 'lo'.