

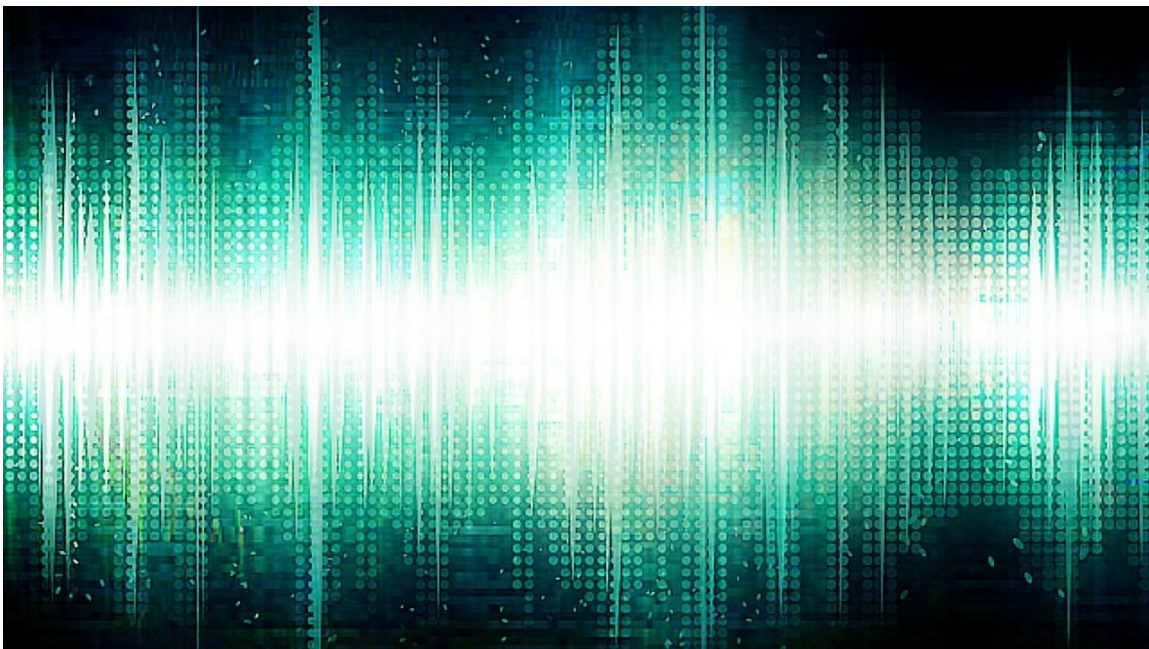
Notas de prensa

18.07.2017

LAS LEYES UNIVERSALES DEL HABLA

Científicos de diferentes organismos de investigación, entre los que se encuentra la UPM, proponen un método que permite medir patrones estadísticos del habla en señales acústicas sin necesidad de tener acceso a la transcripción del mensaje.

Los lingüistas conocen desde hace mucho tiempo la existencia de leyes universales en el lenguaje escrito, patrones en los textos independientes de la lengua. En un trabajo publicado recientemente en Scientific Reports en el que han participado investigadores de las universidades [Politécnica de Madrid](#) (UPM) y [Politécnica de Cataluña](#) (UPC), [Queen Mary de Londres](#) y [Telefónica I D de Barcelona](#), se han encontrado por primera vez leyes equivalentes en la voz, en la señal acústica de conversaciones en multitud de lenguas. La sencillez del método de análisis que se propone abre la posibilidad de estudiar el lenguaje animal desde una nueva y prometedora perspectiva.



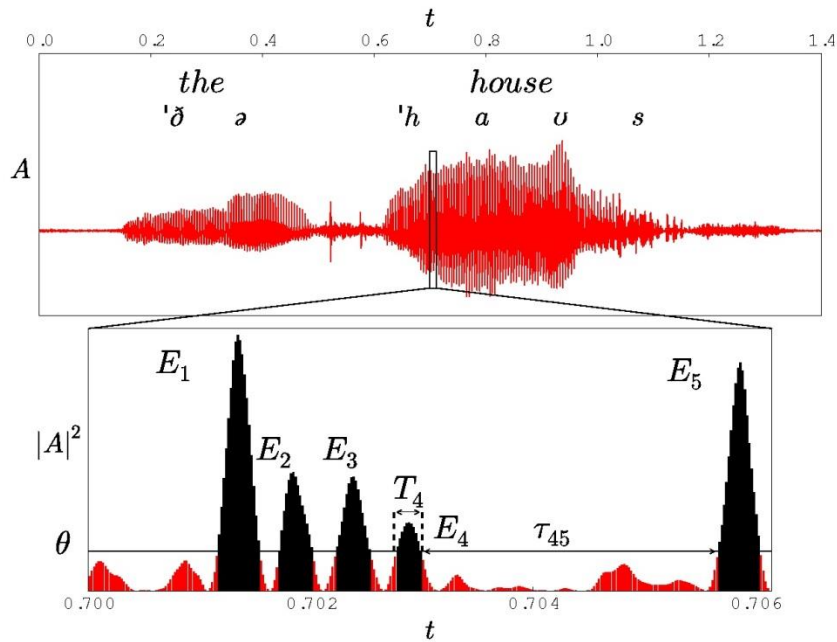
Fuente: pixabay

Cuando en 1848 Marx y Engels publicaron en alemán un tratado de apenas cuatro capítulos, no imaginaron el impacto político y social que alcanzarían sus ideas a lo largo del siguiente siglo. Sin embargo, lo que seguro no imaginaron ni de lejos es que ese ensayo pudiera tener tantas cosas en común con otro rival escrito 72 años antes en inglés por Adam Smith. Sus puntos en común, por supuesto, no fueron ideológicos, sino lingüístico cuantitativos.

La lingüística cuantitativa es la disciplina científica que pretende comprender las regularidades y patrones en la estructura del lenguaje y la comunicación en general. Si tomamos un texto escrito suficientemente largo descubriremos que la frecuencia de aparición de las palabras sigue una distribución determinada casi independiente del idioma. Por ejemplo, la segunda palabra más frecuente aparecerá aproximadamente la mitad de veces que la primera, la tercera más frecuente aparecerá una tercera parte respecto de la primera, etc. Y todo ello independientemente de que el texto haya sido escrito en español o bengalí. Esto se conoce como Ley de Zipf.

Lo interesante de esta regularidad es que es universal y común a todas las lenguas. La ley de Zipf aparecerá tanto en “El origen de la riqueza de las naciones” como en el “Manifiesto Comunista”, aunque obviamente las palabras sobre las que aplica serán distintas. Existen otras regularidades o leyes lingüísticas semejantes y bien conocidas como por ejemplo: que el vocabulario de un texto crezca alométricamente con la longitud del mismo (Ley de Heaps), o la tendencia de las palabras más frecuentes a ser más cortas (Ley de Brevedad), etc. Estos patrones se atribuyen a que el lenguaje se ha ido moldeando a lo largo de un proceso evolutivo para ser eficiente, preciso y facilitar su aprendizaje.

Los métodos empleados por la lingüística cuantitativa a la comunicación escrita o a las transcripciones no pueden aplicarse fácilmente al habla, donde solo se dispone de una señal sonora. Al hablar, por ejemplo, apenas creamos “espacios” entre palabras, que solemos encadenar casi sin solución de continuidad. De hecho, una misma palabra tiene duraciones totalmente distintas dependiendo como se diga (exclamación, interrogación, etc) y sin embargo estará escrita siempre igual. Si se desea aplicar las mismas técnicas estándar al lenguaje no escrito se requieren transcripciones, y resulta imposible para lenguas ágrafas, comunicación animal o comunicación extraterrestre, por ejemplo.



Esquema del método de los umbrales empleado por los investigadores en el trabajo. Convierten una señal en una serie de paquetes de energía con cierta duración separados por silencios. En la parte superior aparece una señal de audio y en la inferior un zoom de ella después de elevarla al cuadrado. Se aplica un umbral arbitrario (línea horizontal) de forma que todo lo que quede por encima serán eventos de voz (negro) y lo que quede debajo silencios (rojo). El resultado final es análogo a disponer de una serie de palabras (paquetes de energía) separadas por espacios (silencios).

En un estudio reciente, en el que han participado los investigadores de la Universidad Politécnica de Madrid Iván González y Bartolo Luque, se propone un método para estudiar las leyes lingüísticas directamente sobre la señal de voz sin necesidad de una transcripción. Partiendo de conversaciones grabadas, se aborda la señal acústica, sin conocimiento del contenido semántico, segmentándola y transformándola en una sucesión de eventos de voz separados por silencios. La idea es que estos eventos microscópicos (tanto energética como temporalmente) de voz son análogos a las palabras de los textos escritos y los silencios se corresponderían a algo así como los espacios. De esta forma, los investigadores han sido capaces de aplicar los métodos tradicionales de la lingüística cuantitativa y recuperar las leyes y patrones lingüísticos tal y como aparecen en textos escritos. El análisis, que trabaja a escalas de tiempo (centésimas de segundo) mucho menores que las de interés para la lingüística tradicional, muestra resultados equivalentes en conversaciones grabadas en 16 lenguas distintas, denotando su universalidad.

Como señala Bartolo Luque, “la importancia de este estudio es que sienta las bases de una explicación microscópica de las leyes lingüísticas y abre la posibilidad de investigar la comunicación animal o incluso de posibles señales extraterrestres desde la perspectiva de los sistemas críticos auto-organizados”.

Torre, I. G., Luque, B., Lacasa, L., Luque, J., & Hernández-Fernández, A. (2017). Emergence of linguistic laws in human voice. *Scientific reports*, 7. <https://www.nature.com/articles/srep43862>