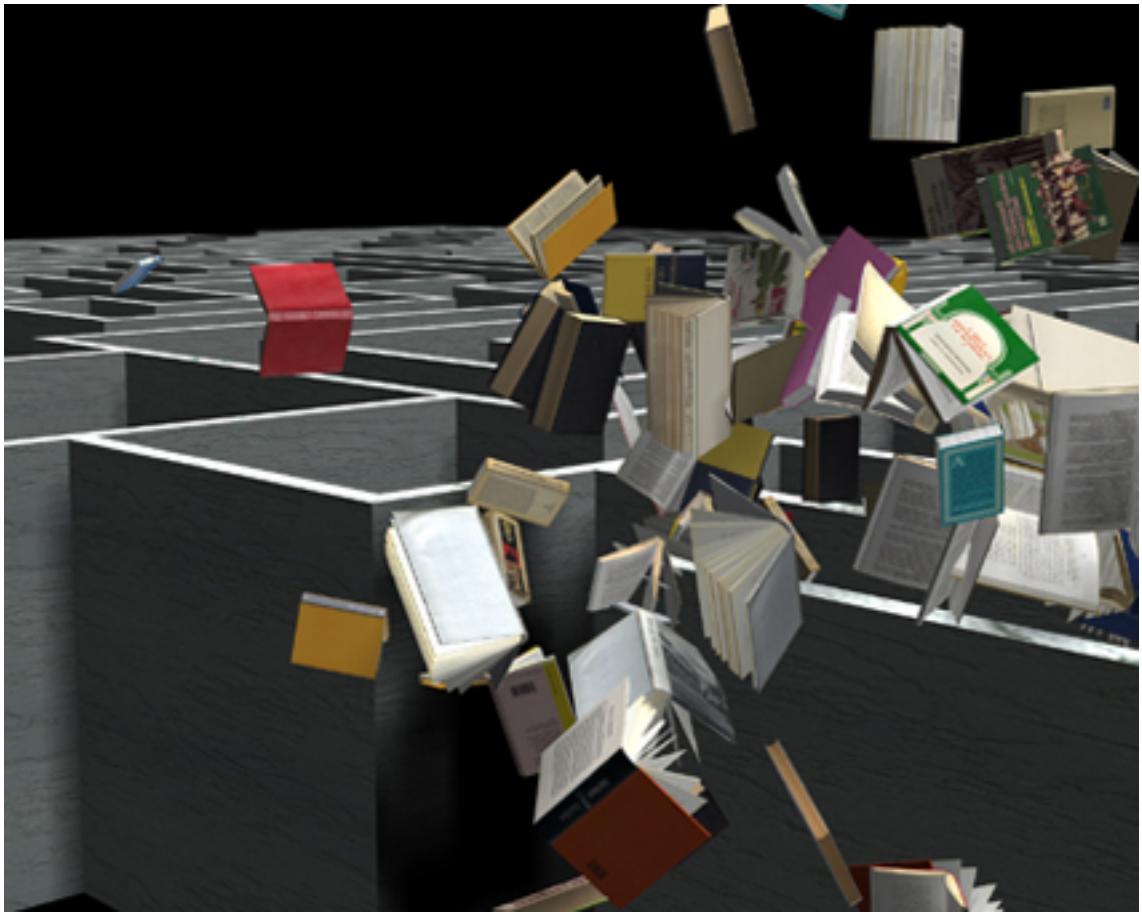
Eduardo Altmann (Instituto Max Planck para la Física de Sistemas Complejos) y sus colegas Giampaolo Cristadoro
y
Mirko Degli Esposti
(Universidad de Bolonia) han estudiado con la ayuda de
métodos estadísticos
cómo las letras y las palabras se correlacionan con la trama de un libro.



Alicia Martín, “Poliglotas”, 2003 (http://www.galica.it/ing/artista.asp?id=33)

Según sus investigaciones, lo que hace que una palabra sea **clave** dentro un texto no es el hecho de que aparezca con mucha frecuencia, sino que aparezca en abundancia en determinados lugares del texto.

Estos científicos han estudiado ***matemáticamente*** las propiedades semánticas de los textos, traduciendo para ello diez libros -como la edición de inglesa de
Guerra y Paz
de León Tolstoi- escritos en inglés a diversos códigos.

Por ejemplo, uno de los experimentos realizados ha sido el de trasladar las letras de un texto a una secuencia binaria, reemplazando las vocales por 1 y las consonantes por 0. Usando otras funciones matemáticas, han examinado diferentes niveles del texto -vocales, letras, palabras completas, etc.- traducidas a diferentes códigos. De este modo, han podido identificar patrones repetitivos dentro del texto visto como un todo.

Los científicos han comprobado -entre otros muchos aspectos- que la frecuencia de aparición de determinadas palabras no tiene necesariamente que ver con su importancia en el texto -por ejemplo, los artículos no transmiten información relevante-; es necesario ver en que lugar se encuentran. Además, este análisis estadístico -que intenta estudiar la relevancia de determinadas palabras en el argumento de un relato- no depende del idioma del escrito: lo importante es la propia historia y no las normas específicas del lenguaje.

Estos resultados podrían utilizarse en el futuro para mejorar los motores de búsqueda en Internet, y también podrían ayudar a analizar los textos en búsqueda de posibles plagios.

Visto en las Noticias del Max Planck Institute for the Physics of Complex Systems

Más información:

- Eduardo G. Altmann, Giampaolo Cristadoro and Mirko Degli Esposti, *On the origin of long-range correlations in texts,*
PNAS 2012 109 (29) 11582-11587; doi:10.1073/pnas.1117723109 [
abstract
y
pdf completo
]

- Eduardo G Altmann, Janet B Pierrehumbert and Adilson E Motter, *Beyond Word Frequency: Bursts, Lulls, and Scaling in the Temporal Distributions of Words,* PLoS ONE (2009) vol 4, Issue 11, doi: 10.1371/journal.pone.0007678 [abstract]

- *Decodificando con ayuda de la estadística* , Matematicalia, Noticias, 18 de agosto de 2009
- *La huella literaria* , Matematicalia, Noticias, 10 de diciembre de 2009

Artículo publicado en el blog de la Facultad de Ciencia y Tecnología (ZTF-FCT) de la Universidad del País Vasco ztfnews.wordpress.com