

Diseñan un algoritmo para calcular la riqueza léxica individual

Investigadores de la Universidad de Málaga hallan un método para estimar de forma fiable la cantidad y calidad del vocabulario de los hablantes. Los resultados confirman que en los individuos con educación superior y universitaria se concentra el grueso de resultados de mayor riqueza léxica, en concreto, un 40% más que en el de personas no escolarizadas o con pocos años de formación.

SINC

4/4/2017 13:53 CEST



La causa de tener un mayor o menor índice de riqueza léxica radica principalmente en el acceso a la educación formal y a la cultura, aunque la profesión y el rol desempeñado en esta parecen también ser clave. / UCIENCIA

La variedad de vocablos que utilizamos al expresarnos, ya sea de forma oral o escrita, es resultado de una amplia gama de condicionantes sociales, culturales y económicos. Tradicionalmente el estudio de la densidad léxica ha estado orientado a los textos escritos, y se toman en cuenta variables como la estructura del discurso, la reiteración de términos o el uso de sinónimos. Sin embargo, existe otro tipo de perspectiva aplicada por los

lingüistas que, junto a los anteriores factores, pone el foco también en la expresión oral y en cómo el contexto sociocultural del propio sujeto influye de un modo u otro en los resultados.

Es el caso del estudio presentado por investigadores del grupo Vernáculo Urbano Malagueño (VUM) de la Universidad de Málaga que han hallado un método para estimar lo que denominan el tamaño virtual del vocabulario de un individuo. Para ello han utilizado un algoritmo de optimización que garantiza resultados “muy fiables”.

En concreto, para el trabajo se ha utilizado un corpus lingüístico del español hablado en la ciudad de Málaga. Un material oral que fue obtenido tras más de 80 horas de grabación en los que, además, se trató de registrar mediante cuestionario las variables como la edad, el género, la ocupación profesional, el nivel educativo y la clase social de cada participante.

En este sentido, como señala el doctor Antonio M. Ávila Muñoz, responsable de la investigación, “gracias al *corpus* del que partimos se pueden detectar diferencias en la comunidad estudiada y los modelos comunicativos que tenemos presentes los hablantes durante la interacción comunicativa”.

El artículo, publicado en la revista *Spanish in Context*, recoge el diseño de un modelo para el cálculo del número de vocablos utilizados por los hablantes compatible con el concepto tradicional de densidad léxica, usado con más frecuencia para medir la riqueza de los textos escritos.

El algoritmo permite conocer la diversidad léxica usada en el intercambio lingüístico y determinar cuándo la densidad del léxico aumenta y cuándo decrece en la conversación

“A diferencia de los procedimientos tradicionales, nuestra propuesta permitió acceder al tamaño virtual del vocabulario individual sin necesidad de recortar de un modo artificial el corpus de datos. El modelo se adapta con facilidad a las características propias de la dinámica de la conversación”, explica el experto.

El factor sociocultural

El algoritmo permite conocer la diversidad léxica usada en el intercambio lingüístico y determinar cuándo la densidad del léxico aumenta y cuándo decrece en la conversación. Entre los factores a tener en cuenta, los más importantes a juicio de los investigadores son el contacto con los medios de comunicación, la ocupación profesional, la educación y la clase social. “La relación que tiene el individuo con los medios de comunicación, por ejemplo, influye en gran medida en su integración con la cultura general y los valores y normas de la comunidad”, asegura el investigador.

Esta conclusión queda patente con datos que reflejan la importancia del contacto de las personas con el código escrito y su reflejo en el habla. Los hablantes con mayor contacto con la prensa escrita obtuvieron un 3,1% más de puntuación en textos de léxico más avanzado que los participantes del grupo con un contacto más restringido a la televisión.

Por otro lado, en el plano laboral los resultados recogen la evidencia de que el mayor porcentaje (el 53%) de los textos más complejos fueron los registrados por empresarios y personas con profesiones técnicas, cargos medios y funcionarios. En cuanto a la educación, la tendencia es similar a los anteriores apartados. Además, se confirma que en los individuos con educación superior y universitaria se concentra el grueso de resultados de mayor riqueza léxica, en concreto, un 40% más que en el de personas no escolarizadas o con pocos años de formación.

Por su parte, la comparativa realizada por los investigadores en función de la clase social señala que de los registros extraídos de la clase media baja (la más alta estudiada) un 68% eran de mayor densidad léxica, frente a los aportados por la clase baja, cuyo porcentaje no superó el 27%.

Según los expertos, lo que también queda demostrado es que los diferentes estratos sociales usan distintos lexicones (conocimiento del vocabulario completo de determinada comunidad de habla), al menos desde el punto de vista cuantitativo; de hecho, en la muestra analizada el léxico fluctuó entre las 862 y las 3.055 palabras de media según el grupo social al que pertenecían los hablantes.

Tras analizar los datos obtenidos, Ávila Muñoz sostiene que “la causa de tener un mayor o menor índice de riqueza léxica radica principalmente en el acceso a la educación formal y a la cultura, aunque la profesión y el rol desempeñado en esta parecen también ser clave”. En este sentido, “lo que también es cierto, es el hecho de que tener más o menos léxico no parece influir en la eficacia de la comunicación, ya que el individuo suele adaptarlo a sus necesidades y objetivos”, añade.

Ávila sostiene que las conclusiones y datos extraídos del trabajo se incorporan al debate sobre el lenguaje y con ello pueden servir para abrir nuevas perspectivas a futuras investigaciones. “Estoy seguro que para la sociedad en general los resultados se pueden explotar de manera útil en ámbitos como los medios de comunicación, donde los profesionales pueden disponer de un modelo riguroso de representación de la riqueza léxica individual; en la educación, tanto en la enseñanza de la lengua materna como de una segunda lengua; así como otros campos como la política y la planificación lingüística, tan importantes en una comunidad de habla como la andaluza en la que conviven variedades lingüísticas en contacto que generan actitudes y creencias tanto dentro como fuera de nuestra región”.

Referencia bibliográfica:

Antonio Manuel Ávila Muñoz (2016). "Can speakers' virtual lexical richness be calculated? Individual and social determining factors" *Spanish in Context*, 13:2, 285-307. Disponible en línea:
<http://dx.doi.org/10.1075/sic.13.2.06avi> ORCID -
<http://orcid.org/0000-0002-5239-2670>

Derechos: **Creative Commons**

TAGS

RIQUEZA LÉXICA | LÉXICO | SOCIOLINGÜÍSTICA | VOCABULARIO |

Creative Commons 4.0

Puedes copiar, difundir y transformar los contenidos de SINC. [Lee las condiciones de nuestra licencia](#)

