

Gran resonancia internacional por un estudio científico del IFISC

Coche, auto, carro, concho y movi o cómo investigadores del IFISC descubren la existencia de dos superdialectos del español en Twitter

El primer estudio de dialectos en redes sociales revela dos grandes superdialectos: uno compartido en grandes ciudades españolas y americanas y otro en zonas rurales.

Publicaciones internacionales como [Newsweek](#), [New Scientist](#), [MIT Technology Review](#), o [Popular Science](#), entre muchas otras, se han hecho eco del trabajo “[Crowdsourcing Dialect Characterization through Twitter](#)” llevado a cabo por investigadores del Instituto de Física Interdisciplinaria y Sistemas Complejos, IFISC (CSIC-UIB) y de la Universidad de Toulon (Francia) sobre dialectos del español en Twitter.

Un dialecto es una forma particular de lenguaje limitado a una región específica o a un grupo social. Tradicionalmente, se han estudiado a través de cuestionarios o entrevistas en grupos limitados de individuos o analizando el lenguaje de los medios de comunicación. Ambos métodos se ven limitados, en el primer caso por la elección de lugares y personas, y en el segundo, por el uso de estándares lingüísticos que no reflejan el uso cotidiano de la lengua.

Pero el reciente aumento de las herramientas sociales en línea ha dado lugar a una avalancha sin precedentes de datos generados por millones de personas distribuidas geográficamente que tienden a usar formas coloquiales y no se sienten obligadas a utilizar normas lingüísticas estándar. Esto, combinado con el uso generalizado de teléfonos inteligentes con GPS, ofrece una oportunidad única de observar cómo se emplean los idiomas en distintos puntos del planeta.

Los investigadores Bruno Gonçalves de la Universidad de Toulon (Francia) y David Sánchez del Instituto de Física Interdisciplinaria y Sistemas Complejos, IFISC (CSIC-UIB) en España, han utilizado una gran base de datos de los tweets geolocalizados para estudiar las variedades dialectales del español. El estudio [Crowdsourcing Dialect Characterization through Twitter](#), aporta una nueva manera de estudiar los dialectos a escala mundial utilizando mensajes publicados en Twitter. Los resultados revelan una sorpresa importante sobre la forma en que los dialectos se distribuyen en todo el mundo y ofrecen una fascinante instantánea de su evolución bajo varias nuevas influencias, como los mecanismos globales de comunicación tipo Twitter.

Gonçalves y Sánchez han recogido 50 millones de tweets geolocalizados escritos en español durante dos años. La mayoría de ellos se ubicaron en España, Hispanoamérica y Estados Unidos, aunque también se hallaron resultados relevantes en las principales ciudades Latinoamericanas y del Este de Europa, seguramente debido a la emigración y al turismo.

Para poder determinar con exactitud cuáles son los principales variedades locales de español, se usó una lista de conceptos y expresiones seleccionadas a partir de un estudio exhaustivo de las variantes léxicas en ciudades de habla hispana. Se seleccionó, a continuación un subconjunto de conceptos para minimizar posibles ambigüedades

semánticas. Entonces, buscaron estos tuits para variaciones de palabras que son indicativos de dialectos específicos. Por ejemplo, la palabra para ‘coche’ en español puede ser *auto*, *automóvil*, *carro*, *hire*, *concho*, o *movi*, dependiendo de la zona dialectal, mientras que las variaciones en el caso de ‘ordenador’ incluyen *computador*, *computadora*, *microcomputador*, *microcomputadora*, *ordenador*, *PC*, y así sucesivamente.

Los tuits que usaban esas formas diferentes se ubicaron luego en el mundo, generándose un mapa con su distribución geográfica. Este mapa muestra claramente cómo diferentes palabras son de uso común en ciertas partes del planeta. Sin embargo, también observaron los entornos en los que se utilizaron las palabras, ya sea en grandes ciudades o en zonas rurales.

Los investigadores descubrieron algo inesperado: los dialectos del español se dividían claramente en dos superdialectos. El primero de ellos, una especie de variedad internacional del español, usado casi de manera exclusiva en las principales ciudades españolas y americanas; y el segundo, en las zonas rurales. Para Gonçalves y Sánchez el primer caso se explicaría debido a la homogeneización creciente de la lengua causada por distintos mecanismos de nivelización (educación, medios de comunicación, Twitter...).

En el segundo caso, en las zonas rurales de España y América, se detectaron tres variedades diferentes que corresponderían a un dialecto utilizado en España, otro presente en amplias zonas de Hispanoamérica y un tercero exclusivo del Cono Sur. Esta división es compatible con los estudios lingüísticos tradicionales que atribuyen el uso distinto de la lengua a los patrones de asentamiento de la administración colonial española. En primer lugar se ocuparon los territorios de México, Perú y el Caribe y mucho más tarde llegaron al Cono Sur. Esta herencia cultural todavía es observable en las bases estudiadas y los investigadores explican que merece ser analizada con detalle en futuros trabajos.

Los investigadores afirman que los resultados arrojados por este trabajo demuestran que “*son relevantes para entender empíricamente cómo se usan las lenguas en la vida real a través de regiones geográficas muy diferentes. Creemos que nuestro trabajo abre un nuevo arco de posibilidades para nuevas aplicaciones en estudios lingüísticos computacionales, un campo lleno de grandes oportunidades*”. Este es una pequeña muestra de lo que se puede hacer. No cuesta nada imaginar análisis mucho más profundos que señalan el camino hacia nuevos avances en los estudios sociolingüísticos (bilingüismo, variedades criollas). “*Nuestro trabajo se basa en un enfoque sincrónico de la lengua. Sin embargo, las posibilidades presentadas por la combinación a gran escala de redes sociales en línea con dispositivos geolocalizados asequibles a la mayoría de población son tan notables que nos podría permitir observar, por primera vez, cómo surgen diferencias y cómo se desarrollan en el tiempo*”.

Por último la elección del español como objeto de estudio se debe, explican los investigadores, al hecho de que no sólo es uno de los más hablado en el mundo sino que tiene la ventaja añadida de encontrarse distribuido espacialmente a través de varios continentes. Otros idiomas con más hablantes nativos como el mandarín o suprarregionales, como el inglés, tienen dificultades añadidas. En el primer caso, la limitada disponibilidad local de Twitter y el segundo requiere un análisis lexicográfico más cuidadoso.