

SOMUIB Noticias

Dirección de la Estrategia de la Comunicación y Promoción Institucional

Descubren la existencia de dos superdialectos del español en Twitter

El primer estudio de dialectos en redes sociales hecho por los investigadores del IFISC (CSIC-UIB) revela dos grandes superdialectos: uno de compartido en las grandes ciudades españolas y americanas y otro propio de 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 los investigadores del Instituto de Física Interdisciplinar y Sistemas Complejos, IFISC (CSIC-UIB), y de la Universidad de Tolón (Francia) sobre dialectos del español en Twitter.

Un dialecto es una forma particular de lenguaje limitada en una región específica o en 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 un luz de datos sin precedentes generados por millones de personas distribuidas geográficamente que tienden a usar formas coloquiales y no se sienten obligadas a utilizar normas lingüísticas estándar. Ello, combinado con el uso generalizado de teléfonos inteligentes con GPS, ofrece una oportunidad única de observar cómo se emplean los idiomas en diferentes puntos del planeta.

Los investigadores Bruno Gonçalves de la Universidad de Tolón (Francia) y David Sánchez, del Instituto de Física Interdisciplinar y Sistemas Complejos, IFISC (CSIC-UIB), en España, han utilizado una gran base de datos de los twits 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 la que los dialectos se distribuyen en todo el mundo y ofrecen una fascinante instantánea de su evolución bajo diversas nuevas influencias, como los mecanismos globales de comunicación tipo Twitter.

Gonçalves y Sánchez han recogido 50 millones de twits geolocalizados escritos en español durante dos años. La mayoría se situaron en España, Hispanoamérica y los Estados Unidos, aunque también se encontraron resultados relevantes en las principales ciudades latinoamericanas y del este de Europa, seguramente a causa de la emigración y del turismo.

Para poder determinar con exactitud cuáles son las principales variedades locales del español, se usó una lista de conceptos y expresiones seleccionados 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, los investigadores escogieron los twits 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 twits que usaban estas formas diferentes se situaron después en el mundo, y se generó un mapa con su distribución geográfica. Este mapa muestra claramente cómo diferentes palabras son de uso común en ciertas partes del planeta. No obstante, también observaron los entornos en los que se utilizaron las palabras, ya sea en grandes ciudades o en zonas rurales.

Los investigadores descubrieron una cosa inesperada: los dialectos del español se dividieron claramente en dos superdialectos. El primero, una especie de variedad internacional del español, usado prácticamente 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 a causa de la homogeneización creciente de la lengua causada por diferentes mecanismos de nivelación (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 diferente de la lengua en 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 se llegó al cono sur. Esta herencia cultural aun es observable en las bases estudiadas, y los investigadores explican que merece ser analizada al detalle en futuros trabajos.



Galería fotográfica

Los investigadores afirman que los resultados lanzados 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 estudio 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. No obstante, las posibilidades presentadas por la combinación a gran escala de redes sociales en línea con dispositivos geolocalizados, asequibles en la mayoría de población, son tan notables que nos podrían permitir observar, por primera vez, cómo surgen diferencias y cómo se desarrollan en el tiempo*».

Finalmente, la elección del español como objeto de estudio es debida, explican los investigadores, al hecho que no solo es de los idiomas más hablados del mundo sino que tiene la ventaja añadida de encontrarse distribuido espacialmente a través de diversos continentes. Otros idiomas con más parlantes nativos, como el mandarín, o suprarregionales, como el inglés, tienen dificultades añadidas. En el primer caso, la limitada disponibilidad local de Twitter y en el segundo, se requiere un estudio lexicográfico más cuidadoso.

Documentos relacionados

- Resum de premsa (925 KB)

Fecha de publicació: 02/09/2014

« Vuelve atrás - Archivo de noticias

Breus

Universitat: accés per als més grans de 25, 40 i 45 anys (fins dia 28/11/2014)

Taller: Les pràctiques en empreses com al primer esgraó de la teva vida professional (19/11/2014)

Concurs d'imatge Què és per a mi la UIB? (fins dia 28/11/2014)

Tweets

Seguir

 **UIB** @UIBuniversitat 2h
Avui a les 17.30 h debatrem sobre el fenomen de l'allotjament privat. bit.ly/1ufGAHI #SomUIB
Abrir

 **Bel Llodrà** @belll07 3h
Interessant cartell sobre #bigdata avui a @UIBuniversitat uib.cat/agenda/12-11-2...
Retwitteado por UIB
Abrir

 **UIB** 3h
Twittear a @UIBuniversitat