

Gran ressonància internacional per un estudi científic de l'IFISC

Coche, auto, carro, concho i movi o com investigadors de l'IFISC descobreixen l'existència de dos superdialectes de l'espanyol a Twitter

El primer estudi de dialectes en xarxes socials revela dos grans superdialectes: un compartit a grans ciutats espanyoles i americanes i un altre a zones rurals.

Publicacions internacionals com [Newsweek](#), [New Scientist](#), [MIT Technology Review](#), o [Popular Science](#), entre moltes altres, s'han fet ressò del treball "Crowdsourcing Dialect Characterization through Twitter" dut a terme per investigadors de l'Institut de Física Interdisciplinària i Sistemes Complexos, IFISC (CSIC-UIB) i de la Universitat de Toulon (França) sobre dialectes de l'espanyol a Twitter.

Un dialecte és una forma particular de llenguatge limitada a una regió específica o a un grup social. Tradicionalment, s'han estudiat a través de qüestionaris o entrevistes en grups limitats d'individus o analitzant el llenguatge dels mitjans de comunicació. Tots dos mètodes es veuen limitats, en el primer cas, per l'elecció de llocs i persones, i en el segon, per l'ús d'estàndards lingüístics que no reflecteixen l'ús quotidià de la llengua.

Però el recent augment de les eines socials en línia ha donat lloc a un allau de dades sense precedents generades per milions de persones distribuïdes geogràficament que tendeixen a usar formes col·loquials i no se senten obligades a utilitzar normes lingüístiques estàndard. Això, combinat amb l'ús generalitzat de telèfons intel·ligents amb GPS, ofereix una oportunitat única d'observar com s'empren els idiomes en diferents punts del planeta.

Els investigadors Bruno Gonçalves de la Universitat de Toulon (França) i David Sánchez de l'Institut de Física Interdisciplinària i Sistemes Complexos, IFISC (CSIC-UIB) a Espanya, han utilitzat una gran base de dades dels tuits geolocalitzats per estudiar les varietats dialectals de l'espanyol. L'estudi Crowdsourcing Dialect Characterization through Twitter, aporta una nova manera d'estudiar els dialectes a escala mundial utilitzant missatges publicats a Twitter. Els resultats revelen una sorpresa important sobre la forma en què els dialectes es distribueixen a tot el món i ofereixen una fascinant instantània de la seva evolució sota diverses noves influències, com els mecanismes globals de comunicació tipus Twitter.

Gonçalves i Sánchez han recollit 50 milions de tuits geolocalitzats escrits en espanyol durant dos anys. La majoria d'ells es van situar a Espanya, Hispanoamèrica i Estats Units, encara que també es varen trobar resultats rellevants a les principals ciutats Llatinoamericanes i de l'Est d'Europa, segurament a causa de l'emigració i del turisme.

Per poder determinar amb exactitud quines són les principals varietats locals d'espanyol, es va usar una llista de conceptes i expressions seleccionades a partir d'un estudi exhaustiu de les variants lèxiques en ciutats de parla hispana. Es va seleccionar, a continuació un subconjunt de conceptes per minimitzar possibles ambigüitats semàntiques. Llavors, els investigadors triaren els tuits per a variacions de paraules que són indicatius de dialectes específics. Per exemple, la paraula per a *coche* en espanyol pot ser *auto*, *automóvil*, *carro*, *hire*, *concho*, o *movi*, depenent de la zona dialectal, mentre que les variacions en el cas de *ordenador* inclouen *computador*, *computadora*, *microcomputador*, *microcomputadora*, *ordenador*, *PC*, i així successivament.

Els tuïts que usaven aquestes formes diferents es van situar després al món, generant-se un mapa amb la seva distribució geogràfica. Aquest mapa mostra clarament com diferents paraules són d'ús comú en certes parts del planeta. No obstant això, també van observar els entorns en els quals es van utilitzar les paraules, ja sigui en grans ciutats o en zones rurals. Els investigadors van descobrir una cosa inesperada: els dialectes de l'espanyol es dividien clarament en dos superdialectes. El primer d'ells, una espècie de varietat internacional de l'espanyol, usat gairebé de manera exclusiva a les principals ciutats espanyoles i americanes; i el segon, a les zones rurals. Per a Gonçalves i Sánchez el primer cas s'explicaria a causa de l'homogeneïtzació creixent de la llengua causada per diferents mecanismes de anivellament (educació, mitjans de comunicació, Twitter...).

En el segon cas, a les zones rurals d'Espanya i Amèrica, es van detectar tres varietats diferents que correspondrien a un dialecte utilitzat a Espanya, un altre present en àmplies zones d'Hispanoamèrica i un tercer exclusiu del Con Sud. Aquesta divisió és compatible amb els estudis lingüístics tradicionals que atribueixen l'ús diferent de la llengua als patrons d'assentament de l'administració colonial espanyola. En primer lloc es van ocupar els territoris de Mèxic, Perú i el Carib i molt més tard van arribar al Con Sud. Aquesta herència cultural encara és observable a les bases estudiades i els investigadors expliquen que mereix ser analitzada amb detall en futurs treballs.

Els investigadors afirmen que els resultats llançats per aquest treball demostren que *“són rellevants per entendre empíricament com s'usen les llengües en la vida real a través de regions geogràfiques molt diferents. Creiem que el nostre treball obre un nou arc de possibilitats per a noves aplicacions en estudis lingüístics computacionals, un camp ple de grans oportunitats”*. Aquest estudi és una petita mostra del que es pot fer. No costa gens imaginar anàlisis molt més profundes que assenyalen el camí cap a nous avanços en els estudis sociolingüístics (bilingüisme, varietats criolles). *“El nostre treball es basa en un enfocament sincrònic de la llengua. No obstant això, les possibilitats presentades per la combinació a gran escala de xarxes socials en línia amb dispositius geolocalitzats, assequibles a la majoria de població, són tan notables que ens podria permetre observar, per primera vegada, com sorgeixen diferències i com es desenvolupen en el temps”*.

Finalment l'elecció de l'espanyol com a objecte d'estudi es deu, expliquen els investigadors, al fet que no només és un dels més parlats al món sinó que té l'avantatge afegit de trobar-se distribuït espacialment a través de diversos continents. Altres idiomes amb més parlants nadius com el mandarí o suprarregionals, com l'anglès, tenen dificultats afegides. En el primer cas, la limitada disponibilitat local de Twitter i en el segon, es requereix una estudi lexicogràfic més acurat.

Més informació: Rosa Campomar. Comunicació i Cultura Científica IFISC. Tel.: 971 25 97 19. E-mail: rosac@ifisc.uib-csic.es