



ACCÈSIT 1

Què *** és la lingüística?

per Rogelio Nazar

VI Premi Joan Oró a la Divulgació de la Recerca Científica

Convoca l'Associació Catalana de Comunicació Científica (ACCC) amb la col·laboració del Departament d'Innovació, Universitats i Empresa de la Generalitat de Catalunya

El major inconvenient que té dedicar-se a la lingüística possiblement sigui no saber què respondre quan et pregunten a què et dediques. És impossible respondre a aquesta pregunta sense dubtar. Això és incòmode sobretot quan estàs davant d'un funcionari de migracions, perquè sense passaport europeu no convé dubtar gaire. Ja em va passar una vegada a Alemanya. El salconduit que m'havia donat la Universitat deia que hi anava a estudiar *lingüística*, sense més especificacions, i la funcionària trobava inversemblant que algú pretengués estudiar lingüística a Alemanya sense parlar alemany. Crec que les expectatives d'aquesta funcionària van ser conseqüència de la important tradició filològica que van desenvolupar els nostres amics teutons.

Molt poca gent sap què és la lingüística. De fet, no hi ha gaires lingüistes que ho tinguin clar. Hi ha dues idees que semblen estar força esteses a l'imaginari social: un lingüista, o bé és algú que estudia moltes llengües, o bé és algú que es dedica a tenir cura de la puresa de la seva pròpia llengua i que castiga els que la utilitzen malament. Però, el que sí tenen clar molts lingüistes és que la seva tasca no té res a veure amb una cosa ni amb l'altra.

La lingüística és la ciència del llenguatge. Estudia les llengües, no per aprendre-les en si mateixes, sinó per comprendre la lògica que hi ha darrere seu. Aquesta lògica, aquest sistema, és el que entenem per llenguatge. Derivar lleis generals del llenguatge ens permetria, per suposat, comprovar el seu compliment en cada llengua, fins i tot en llengües que encara no existeixen. Quant a la qüestió de la correcció del llenguatge, és encara matèria de debat, i molts *vigilants de la llengua* es presenten com a lingüistes. Però el lingüista, com a científic, és neutral. El lingüista no pren partit sobre quina hauria de ser la forma correcta de parlar. Això seria com si un biòleg considerés que algunes espècies són correctes i altres són aberracions de la natura.

Està clar que a la pregunta sobre què és la lingüística no es respon amb els tipus de lingüística que hi ha. Però és així, hi ha tipus de lingüística molt diferents. Hi ha la fonologia, que s'especialitza en els sons de la llengua, i ensenya una cosa que va resultar xocant per a totes les ciències socials: el parlant actua obeint regles de les quals és totalment inconscient. La forma en què tota la vida hem pronunciat els sons de la nostra

VI Premi Joan Oró a la Divulgació de la Recerca Científica

Convoca l'Associació Catalana de Comunicació Científica (ACCC) amb la col·laboració del Departament d'Innovació, Universitats i Empresa de la Generalitat de Catalunya

llengua materna està regida per un sistema molt racional. Però aquesta raó no és la raó del parlant. Diguem que és un sistema racional malgrat el parlant. Tan perfectament racional com el funcionament d'un fetge. I després hi ha la morfologia: com es formen les paraules; la lexicologia –quines són aquestes paraules i com es relacionen entre elles; la sintaxi –com s'ajunten les paraules formant oracions gramaticalment correctes, i, la semàntica –la relació de les paraules amb els seus significats. La complexitat d'aquest últim tema es abismal, i es a això al que em dedico.

Més específicament, em dedico a estudiar les relacions entre les paraules i els conceptes. La hipòtesi de la qual parteixo és que podem analitzar els conceptes a partir de l'estudi dels contextos on apareixen les paraules que els designen. Això és cert fins a tal punt que es poden reconstruir els conceptes només estudiant com es combinen les paraules en grans col·leccions de textos. Per demostrar-ho, he fet un model computacional que es basa en una petita xarxa neuronal artificial en la qual cada neurona s'especialitza en una paraula. Aquest model aprèn amb textos d'Internet: cada vegada que veu que dues paraules apareixen juntes en una mateixa oració o en un mateix paràgraf, les neurones que estan especialitzades en aquestes paraules es connecten entre si o enforteixen la seva connexió si ja la tenien.

Amb aquest model d'aprenentatge és possible realitzar experiments en què un ordinador resol tasques que podríem considerar intel·lectuals amb una habilitat comparable a la d'un ésser humà. Una de les tasques més bàsica, dins de l'anàlisi conceptual, és saber quan una paraula o una seqüència de paraules d'un text fa referència a un concepte definit i poder determinar quines són algunes de les propietats bàsiques d'aquest concepte. Altres tasques poden ser determinar quan una mateixa paraula pot fer referència a més d'un concepte (la polisèmia) i així mateix saber quan dues o més paraules diferents fan referència al mateix concepte (la sinonímia).

Com és possible que un sistema així pugui donar resultats? Una explicació és que hi ha una llei general discursiva que diu que quan els autors introdueixen un concepte en un text tenen tendència a presentar algunes de les seves propietats més destacades. Això passa sobretot a l'inici del discurs, per demostrar que els principis entre escriptor i lector són compartits. Cada autor predicarà diferents informacions sobre els conceptes, però els autors que tractin sobre un mateix tema tendiran a partir d'una base comuna, i aquesta coincidència, pel seu efecte acumulatiu, pot ser apresada de manera estadística.

A aquesta alçada del text el lector s'estarà preguntant per què el títol. No sóc parlant nadiu del català i volia trobar una traducció per al títol que en castellà em sonava bé, però havia d'assegurar-me de no triar una cosa que sonés massa grollera. Per arribar a la traducció correcta vaig escriure "Què * és" a Google, restringint els resultats al català. Esperava trobar coses com *diable*, *punyetes* o *diantre*, però en canvi vaig trobar que *coi* era molt més freqüent. El meu diccionari no ajudava: una hamaca?, una referència al Comitè Olímpic Internacional? La Carme, col·lega catalana, m'aclareix que és un eufemisme de cony, però que es pot fer servir perquè ja ningú se'n recorda. L'incident,

VI Premi Joan Oró a la Divulgació de la Recerca Científica

Convoca l'Associació Catalana de Comunicació Científica (ACCC) amb la col·laboració del Departament d'Innovació, Universitats i Empresa de la Generalitat de Catalunya

però, em va fer deixar l'asterisc no només per decència, sinó perquè expressa de manera sintètica el que és una paraula: una paraula no és només una paraula, és també un context. I tant és així que de vegades no necessitem la paraula perquè el context la fa evident.

VI Premi Joan Oró a la Divulgació de la Recerca Científica

Convoca l'Associació Catalana de Comunicació Científica (ACCC) amb la col·laboració del Departament d'Innovació, Universitats i Empresa de la Generalitat de Catalunya