

 CIÈNCIA

# Lingüistes catalans ajuden a resoldre un cas d'extorsió

► Un jutge considera prova suficient una anàlisi del llenguatge usat en uns 'e-mails'

► Nous sistemes informàtics permeten detectar plagis i identificar anònims

MICHELE CATANZARO  
BARCELONA

Per escriure un missatge sense ser reconegut no és suficient redactar-lo a l'ordinador amb un nom fictici. «El llenguatge ens delata», resumeix Maria Teresa Turell. Aquesta lingüista va ajudar a identificar l'autor d'una extorsió analitzant el text dels seus correus electrònics. Per primera vegada a Espanya, un jutjat de Tarragona va considerar aquesta atribució d'autoria per mètodes lingüístics com a prova concludent. El cas va tenir lloc el 2003, però la publicació científica que detalla els mètodes usats és de final del 2010.

L'exemple de Tarragona és una prova dels nous poders de la lingüística. Gràcies al processament de grans extensions de textos per ordinador, els investigadors són capaços de detectar patrons universals de la llengua i senyals especials que caracteritzen cada autor. Detectar plagis o identificar autors desconeguts són algunes de les aplicacions possibles.

**EXTORSIÓ** // L'estiu del 2003, responsables d'una empresa catalana van rebre cinc faxos i quatre correus electrònics firmats per un treballador. Els rumors sobre una retallada de plantilla van ser una possible causa d'aquests missatges. Tot i així, un dels correus sollicitava diners a canvi de no revelar informacions. Arran d'aquesta extorsió, l'empresa va acomiadar el firmant, però aquest va presentar una demanda per acomiadament improcedent: admetia ser autor dels faxos, però no dels correus. Suposadament, algú hauria intervingut el seu ordinador.

Llavors, l'empresa va demanar una perícia del ForensicLab, el laboratori de lingüística forense de l'Institut Universitari de Lingüística Aplicada (IULA) de la UPF, coordinat per Turell. «El nostre treball va ser comparar els textos i atribuir l'autoria dels correus al mateix autor dels faxos, amb alta fiabilitat», prossegueix.

Una de les evidències més clares va ser l'ús de l'expressió *el cual*. «En les bases de dades de referència de textos en espanyol, el pronom relatiu *que* apareix en el 98% dels casos, mentre que *el cual*, i les altres formes de relatiu compost apareixen només en el 2% restant –explica Turell–. Per contra, l'autor els utilitzava respectivament en el 52% i el 48% dels casos». A posteriori, la investigadora va



JONATHAN GREVSEN

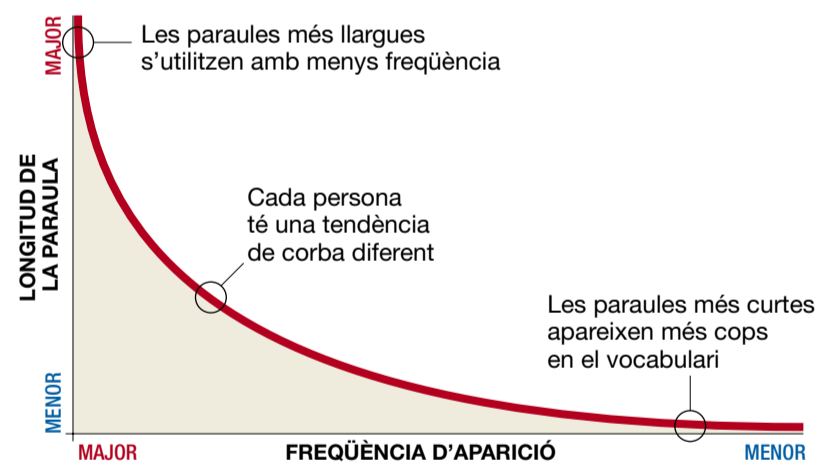
► Maria Teresa Turell, lingüista forense de la UPF, a la biblioteca del Col·legi d'Advocats de Barcelona

## Llengües i sons dels animals podrien compartir característiques

► Entre els humans hi ha comunitats de parla que identifiquen una professió, gènere, classe social o edat. «A més, cada parlant té el seu *idiolecte*, la seva selecció individual de les possibilitats que ofereix la llengua», explica Maria Teresa Turell. Mentre aquesta investigadora persegueix aquest estil únic i irrepetible, altres busquen l'altre extrem: lleis universals de la llengua que es podrien estendre fins i tot als sistemes de comunicació no humans. «Un exemple és la llei de brevetat, que afirma que en totes les llengües les paraules més freqüents tenen menys lletres que les més rares», explica Ramon Ferrer Cancho, investigador en lingüística computacional de la Universitat Politècnica de Catalunya.

► Altres s'han descobert gràcies al processament per ordinador. L'any passat, per exemple, es va arribar a la conclusió que les paraules més freqüents també tendeixen a ser irregulars, com els verbs «ser» i «haver». Ferrer ha estudiat l'aplicació d'una d'aquestes lleis als xiulets de comunicació entre dofins: la freqüència dels diferents tipus de xiulet sembla distribuir-se de manera semblant a les paraules humanes.

## LLEI DE BREVETAT



Font: UPC

EL PERIÓDICO

saber que aquest havia viscut en el passat a Catalunya i probablement estava influït per l'ús dels relatius del català. Però el més important és que en els correus es trobava la mateixa proporció que en els faxos. «La perícia també va analitzar l'estil, l'estructura del text, l'ús de frases subordinades, la freqüència de cada paraula, la densitat i la riquesa del lèxic, els errors, etcètera», puntualitza Turell. Finalment, el jutge va considerar concludent aquest conjunt de proves, juntament amb una perícia informàtica.

«Una prova d'aquest tipus és vàlida només quan hi ha una mostra suficient de textos. Això no passa sempre en el context forense, en el qual els textos solen ser curts, de vegades escrits sota estrès i sense l'objectiu de perdurar», observa Turell.

«La lingüística forense se serveix dels avenços dels sistemes

computacionals», afirma Turell. Per mitjà dels ordinadors es poden fer estadístiques sobre un corpus de textos de referència i comparar-les amb un cas concret en estudi, per detectar com els autors difereixen entre si i de la mitjana.

**PLAGI** // «Una de les aplicacions és la detecció de plagis», explica Paolo Rosso, investigador de la Universitat Politècnica de València i col·laborador de Turell. Amb sistemes informàtics es pot escanejar internet per detectar còpies en treballs universitaris, articles o obres literàries. Rosso ha dissenyat un mètode per identificar plagis basats a traduir textos aliens.

«La lingüística forense està activa des de fa 50 anys. Malgrat tot, només ara comença a obrir-se camí en els tribunals espanyols», observa Turell. ≡