



**LEXIMANCER.**  
**Programa per a la generació**  
**automàtica de mapes**  
**conceptuals d'Andrew E. Smith**  
<http://www.leximancer.com>

**L**eximancer és un programa que pot interessar als terminòlegs perquè és una eina per a trobar associacions conceptuals entre els termes d'una disciplina. Aquest tema ha guanyat molta popularitat en els darrers anys, fins a transcendir l'àmbit de la lingüística. Podem dir que es tracta d'un moviment interdisciplinari que aglutina no tant investigadors en lingüística teòrica sinó més aviat especialistes en informàtica, en estadística i, fins i tot, en física.

En paraules d'Andrew Smith, autor del projecte, «Leximancer is founded in the observations of corpus linguistics, computational linguistics, and psycholinguistics that word co-occurrence statistics in natural language are rich source of information that correlates with certain aspects of human language learning, comprehension and performance» (Smith & Humphreys, 2006, «Evaluation of Unsupervised Semantic Mapping of Natural Language with Leximancer Concept Mapping». *Behavior Research Methods*, vol. 38, núm. 2, p. 265).

El que va produir el salt en l'increment de les propostes en el camp del que podem anomenar *generació automàtica de mapes conceptuals* va ser l'aplicació estadística. Tanmateix, la idea de la generació automàtica de representacions conceptuals per mitjà d'estadístiques de coocurrència lèxica no és nova, i en lingüística teòrica és on trobem els fonaments que van permetre desenvolupaments posteriors, els orígens dels quals es remunten, al seu torn, al distribuacionalisme americà.

Des de la dècada de 1970, en disciplines com la lingüística computacional o la intel·ligència artificial (IA), es va treballar en el camp d'anàlisi conceptual amb un enfocament diferent i, en general, amb poc èxit: el dels sistemes basats en regles. En termes d'IA, la idea d'una representació conceptual a partir de text estava lligada necessàriament a una anàlisi lingüística a tots els nivells: morfològic, sintàctic i semàntic, i l'anàlisi que-

dava sovint a mig camí per l'enorme dificultat d'aquestes tasques. La dificultat més gran és saber distingir quina informació és rellevant per tal de ser introduïda en un mapa conceptual, per això els sistemes experts d'IA necessiten una enorme quantitat de coneixement ontològic explícit prèviament instal·lat, la qual cosa limita el seu camp d'aplicació a llengües i a àmbits del coneixement específics.

De manera radicalment diferent, els treballs sobre anàlisi conceptual basats en estadístiques de coocurrència lèxica o les anàlisis de «semàntica distribuicional» poden aplicar-se sense coneixement previ. Les propostes en aquesta línia van anar evolucionant paral·lelament en psicologia cognitiva i en lingüística de corpus. En ambdós camps es va descobrir que l'associació lliure entre conceptes que fan les persones està correlacionada amb les estadístiques de coocurrència lèxica trobades en corpus, la qual cosa possibilita la detecció de l'associació conceptual entre termes d'una manera mecànica, amb independència de la llengua i amb algorismes molt més senzills que els que es basen en regles.

El que no és tan clar és com o amb quina finalitat es pot utilitzar la informació de les coocurrències lèxiques. En el cas de Leximancer, la idea fonamental és representar en un mapa de dues dimensions les paraules més importants (més freqüents) en un corpus agrupades en funció de les seves coocurrències. Els mots que acostumen a coocórrer en els mateixos segments de text (d'una extensió ajustable) tendeixen a aparèixer junts en el mapa, fet que reflecteix els diferents conceptes del corpus com si fossin regions que agrupen diferents paraules.

És possible utilitzar el programa durant un període de proves enviant una sol·licitud, i es pot adquirir per mitjà de llicències acadèmiques o professionals. El programa es pot descarregar o bé utilitzar com a aplicació web. Els mapes conceptuals que genera perme-

ten saber més o menys de què es parla en un corpus, però és clar que no hi ha una representació en forma de proposicions sinó de núvols de paraules. Sabem que una proposició és una sèrie de categories lligades per un predicat, perquè si només hi ha categories, no hi ha sentit. No cal dir que, en un núvol de paraules, tota l'argumentació racional dels textos es perd i, si el fil conductor i l'estructuració desapareixen, es corre el risc de tenir amb aquests mapes conceptuals una sensació d'estar entenent alguna cosa quan en realitat no entenem res, un efecte similar al que, de vegades, produeix el PowerPoint.

De totes maneres, del Leximancer hem de destacar la voluntat de reflectir la saviesa de masses: per a aquest programa la informació rellevant no és en cap text en particular, sinó en l'agregació de tots aquests. Aquest enfocament pot ser rellevant per exemple en estudis sobre implantació terminològica, per distingir entre

implantació *top-down* (el terme que s'imposa des d'organismes d'estandardització) i *bottom-up* (el terme que ja estan utilitzant els especialistes). Diversos investigadors en ciències socials, així com organismes públics i privats, utilitzen també el programa per a les anàlisis de l'opinió pública.

Tinc la sensació que la popularitat dels mapes conceptuals estadístics anirà en augment i l'expressió *xarxes de coocurrència* experimentarà un procés de «des-terminologització» similar al que va experimentar el terme *xarxes socials* després de romandre en l'àmbit de la sociologia des de la dècada de 1940.

**ROGELIO NAZAR**

Institut Universitari de Lingüística Aplicada  
Universitat Pompeu Fabra