

InfoMigjorn

Butlletí digital
sobre llengua
catalana

Expandint els límits dels verificadors lingüístics

Jaume Ortola i Font

jaumeortola@gmail.com

InfoMigjorn, núm. 1.285, 17 de juny de 2016

Popularment en diem «corrector ortogràfic», encara que potser és més precís dir-ne «verificador», com proposa el Termcat, que el defineix com el «programa informàtic que identifica errors ortogràfics, tipogràfics i d'altres, hi proposa alternatives i permet de substituir-los». És preferible «verificador» perquè el programa no corregeix els errors, sinó que només alerta de la seva presència. La correcció es deixa en mans de l'usuari.

Amb el reviscolament de la comunicació escrita que ens ha portat internet, aquests programes han esdevingut omnipresents. Primer els teníem en el processador de texts, però ara són a tot arreu: en el correu electrònic, en la missatgeria electrònica, en el navegador, en el xat..., i en tota classe de dispositius: en l'ordinador personal, en la tauleta, en el mòbil. És natural, doncs, que les aplicacions d'ajuda lingüística també s'hagin desenvolupat en sintonia amb aquesta expansió. Són cada vegada més usuals, per exemple, els sistemes d'autocompleció o autocorrecció, especialment en els mòbils i les tauletes.

La llengua catalana ha lluitat sempre per estar present en les noves tecnologies d'una manera adequada. L'any 1989 ja disposàvem del programa «ADHOC. Verificador ortogràfic de català» —es deia exactament així—, i entitats com Softcatalà, nascuda l'any 1998, han treballat intensament perquè el català tin-

gués un suport adequat en els àmbits de la informàtica i internet, especialment en matèria de diccionaris i correctors.

Els verificadors ortogràfics fan una funció relativament senzilla. El programa disposa d'una llista de paraules i comprova si cada paraula que escrivim és o no en la llista. Si no hi és, proposa una alternativa cercant paraules pròximes a la paraula que pot estar mal escrita (perquè li falta un accent, té una lletra canviada, etc.). Si el verificador ortogràfic és una mica més complet, pot dir-nos més coses, per exemple sobre apostrofació o pronoms febles: pot advertir-nos que **l'hidròlisi* no està ben escrit, però *l'última sí*; que *compra'n* és correcte, però **comprà'n* no. En canvi, no troba error en **la última* o en **se'n vas*. Aquest estat de la qüestió és el que trobem en el verificador ortogràfic de LibreOffice o de Microsoft Word, que en català no ofereix correcció gramatical com fa en altres llengües.

Aquestes limitacions haurien de ser conegudes, almenys intuïtivament, per qualsevol usuari amb uns coneixements mínims de llengua. Així i tot, per a deixar-ho més clar, algunes persones han intentat sistematitzar-les i explicar-les. Neus Nogué ho feia en l'article «[Avantatges i limitacions dels verificadors ortogràfics](#)». Des d'una altra perspectiva, Pep Bofarull proporciona [parells de paraules](#) que es poden confondre i poden provocar errors que un verificador convencional no detecta.

Doncs bé, aquest panorama ha començat a canviar recentment amb l'aparició d'eines de revisió lingüística molt més completes i sofisticades que els tradicionals verificadors ortogràfics. En el cas del català, un d'aquests programes és LanguageTool.¹ Es tracta d'un projecte de programari lliure, internacional i multilingüe, amb una comunitat d'usuaris i col·laboradors molt activa i compromesa, cosa que li dóna molta solvència tecnològica i li garanteix la continuïtat en el futur. LanguageTool funciona en més d'[una vintena de llengües](#). El grau de servei en les diferents llengües és certament molt desigual, però podem estar satisfets que el [català](#) sigui una de les llengües amb més desenvolupament. A part de la tradicional revisió ortogràfica, en català funcionen més de 2.500 regles que ajuden a fer una revisió lingüística que, si bé no és exhaustiva, va molt més enllà del que es podia fer fins ara.

¹ N'hi ha d'altres: el programari de propietat Maxigramar; el Corrector de la Universitat Pompeu Fabra, de codi lliure però intrombable (projecte abandonat); i el SALT de la Generalitat Valenciana, programari de propietat que ofería algunes funcions avançades de revisió (ara amb el desenvolupament aturat). Hi ha un mercat d'usuaris que requereixen aquests serveis, però no tots els models de desenvolupament i comercialització funcionen bé en la nostra situació sociolingüística i demogràfica.

És complicat donar una visió global de totes aquestes regles. Hi ha regles per a errors ortogràfics, barbarismes habituals, errors gramaticals de tota mena, errors tipogràfics, confusions típiques de picatge, errors semàntics segons el context, variants dialectals i diferents preferències estilístiques. La casuística de tot el que pot fer bé —o malament— un verificador lingüístic és tan inabastable com la llengua mateixa. Fer-ne una avaluació objectiva té moltes més complicacions de les que podria semblar a primera vista. Per a fer una prova i que tingui un cert aire d'objectivitat (que no sigui un examen en què l'alumne sap les preguntes abans de començar), he agafat com a guia l'article esmentat més amunt de Neus Nogué. Així podrem fer-nos una idea, encara que molt parcial, dels aspectes en què hem avançat i de les llacunes que no estan cobertes.

En la secció A de l'article, es parla de tots els errors que ja detecten els verificadors convencionals. LanguageTool (LT) fa tot això, evidentment, però a més dóna suggeriments en errors de lèxic habituals (apartat A.d).

És un gran aconteixement! [LT suggereix **esdeveniment**]

Haig d'anar a la peluqueria. [**perruqueria, barberia**]

Adreceu-vos a la guixeta de la dreta. [**taquilla, armariet**]

B.a) Errors de picatge que donen com a resultat paraules existents.

Això és pera la Júlia. [**per a**]

Les ametlles escullen al setembre. [**Error no detectat**]

No es pot descartar un altre atempta en el futur. [LT detecta que hi ha dos verbs mal connectats: pot... atempta]

B.b) Errors d'ortografia relacionats amb els accents diacrítics i casos similars.

Aquesta prova dona bons resultats. [**dóna**]

Això ja no és fa. [**es**]

No sap que vol. [**què**]

No ho facis, sinó vols. [**si no**]

No serè gaire exigent. [**seré**]

Van anar a París enlloc de Roma. [**Error no detectat**. LT interpreta «enlloc de Roma=en cap lloc de Roma», però sí que detecta molts dels errors referents a «enlloc/en lloc».]

B.c) Absència d'apòstrof.

Aprofita la oferta! [**l'oferta**]

No és la estructura més adequada. [**l'estructura**]

No és la única opció. [**l'única**]

B.d) Errors de concordança de gènere.

Aquesta resposta no és correcte. [LT detecta possible falta de concordança]

Haurem de pensar en un altre opció. [LT suggereix **altra**]

B.e) Qualsevol fenomen sintàctic.

Va entrevistar a tres jugadors. [**Error no detectat**. Però sí que es detecten problemes d'aquest estil en alguns verbs: afectar, atraure, perjudicar...]

A això no se'n va referir. [**Error no detectat**]

És perjudicial pel medi ambient. [**per al**]

L'arribada dels jugadors van provocar crits d'eufòria. [**Error no detectat**. Però sí que es detecta error en frases menys ambigües com ara: «L'arribada de l'equip van provocar eufòria».

Estic d'acord en que cal actuar. [**en el fet que, que...**]

És el primer autor en adonar-se'n. [**a**]

B.f) Altres fenòmens relacionats amb la combinació de diverses paraules.

Tal i com et vaig dir ahir... [**tal com**]

Pel que fa el segon cas... [**pel que fa al**]

Aquest anàlisi no és factible. [**aquesta anàlisi**]

B.g) Altres fenòmens relacionats amb les semblances entre paraules.

No s'hi sembla gens. [**Error no detectat**. Però en el pròxim «examen» LT ja ho farà bé.]

C) Paraules ben utilitzades que el verificador ortogràfic no té en el diccionari.

C.a) Composts i derivats ben formats.

És un llibre **interessantíssim**. [acceptat per LT]

De seguida van arribar cinc unitats **antiavalots**.

Un grup **autodefinit** com a radical.

Vivim en un món **hiperconnectat**.

Tenir-lo activat ens estalvia errors de **picatge**.

C.b) Neologismes o tecnicismes més o menys recents.

Treballarem en un entorn **wiki**.

C.c) Paraules que el diccionari normatiu ha acceptat fa poc.

Tinc *bastanta* gana. [LT dóna un avís: *bastanta* és acceptat pel DIEC, però no pel DNV, i té la marca de *popular* en alguns diccionaris.]

Busca-ho en el **blog** Retruc.

C.d) Variants morfològiques normatives de diferents variants de la llengua.

Va saludar la **meua** mare

No crec que **sigu** cert.



[LT accepta —requereix, de fet— aquestes formes quan es tria la variant adequada. El mateix passa amb l'accentuació (**café** o **cafe**).]

Amb aquest tast, esperem haver donat una idea de l'estat dels verificadors lingüístics en català. Visiteu les webs softcatala.org/corrector i languagetool.org per a conèixer les diferents maneres en què es pot usar el corrector LanguageTool. Cada vegada hi ha més possibilitats: corrector en la web, corrector per a LibreOffice, per a MS Word, per a Chrome, per a Android... També l'utilitzem per a fer correccions massives en la Viquipèdia o per a fer controls de qualitat en produccions editorials de gran volum.

LanguageTool està sempre en procés de millora continuat. L'objectiu és detectar el màxim nombre possible d'errors i minimitzar el nombre de falses alarmes (que podrien fer-lo massa molest). Creem regles noves i les refinem, afegim paraules al diccionari i hi introduïm correccions... Si voleu contribuir-hi amb les vostres observacions i les vostres idees, podeu fer-ho mitjançant el contacte en la pàgina softcatala.org/corrector.

Els verificadors lingüístics no fan miracles, com diu Neus Nogué en el seu article. Els més avançats fan molt més, però tampoc no són perfectes. A més a més, cal estar sempre amatents davant l'excés de confiança en què podem caure usant unes eines cada vegada més eficaces. En qualsevol cas, no hi ha dubte que aquestes eines són molt útils i que ajuden a millorar els textos de cada vegada més persones. Per això val la pena que les coneguem i les difonguem.

InfoMigjorn és un butlletí que distribueix missatges informatius relacionats amb la llengua catalana. Es publica dues vegades per setmana, els dimarts i els divendres. Elaborat i distribuït per la Xarxa de Mots. Edició: Isabel Casadevall

Subscripcions: <http://migjorn.cat/alta.php>. Adreça de contacte: infomigjorn@migjorn.cat.   @InfoMigjorn