

Eines informàtiques per al corrector: com se'n pot treure el màxim profit

Lluís de Yzaguirre

Cap del Laboratori de Tecnologies Lingüístiques
Institut Universitari de Lingüística Aplicada (Universitat Pompeu Fabra)

1. Presentació

Les pàgines que segueixen parlaran de tecnologia aplicada a la correcció lingüística, fonamentalment de recursos relacionats amb el tractament de textos. Qui en parla no és un corrector, detall que podria fer minvar l'interès de qui ho llegeixi; ho compensa, i és bo que se sàpiga, que l'autor té àmplia experiència sobre aquests recursos pel fet d'haver contribuït, en diferents mesures, des de l'inici dels anys vuitanta, a la producció de diversos verificadors ortogràfics en català (ADHOC del Programa d'Informàtica Educativa, LocoSpell de LocoScript, PageMaker...), tenir una dilatada experiència docent i haver participat en diversos projectes que d'una manera o altra, com es veurà, tenen relació amb el tema (projecte de creació assistida de documents administratius universitaris del Servei de Llengua Catalana de la Universitat de Barcelona, projectes de lematització i desambiguació a l'Institut Universitari de Lingüística Aplicada de la Universitat Pompeu Fabra). A tot aquest capital s'hi han d'afegir les consultes fetes entre persones que utilitzen algunes de les eines que es presentaran, amb el resultat que sumem experiències en les diverses fases de generació dels recursos que ens interessa analitzar.

2. Control de la correcció segons les etapes de producció

Diuen que la medicina tradicional xinesa entenia que un metge només havia de cobrar si els seus pacients estaven sans; doncs, el professional més reeixit era aquell que no havia de guarir mai cap pacient per tal com la seva tasca preventiva i educativa els evitava les malalties.

Amb la mateixa idea, aquí no ens limitarem a proposar els recursos que la persona que corregeix té a la seva disposició per a *guarir* els textos, sinó també els que pot usar per prevenir els errors, seus o de les persones a qui ha de corregir els escrits. Sota aquest punt de vista, considerarem que la correcció en un organisme o institució té tres etapes de producció: preventiva, interactiva i pal·liativa.

En l'etapa preventiva, la persona que corregeix pot incidir en els recursos que hi ha a disposició de cadascuna de les persones que li faran arribar textos per corregir, en l'adequació d'aquests recursos a les necessitats productives (llengües, terminologia de l'àmbit professional, onomàstica...), especialment si l'empresa treballa amb aplicacions *groupware* (de treball cooperatiu), i en la familiarització del personal amb aquests recursos.

En l'etapa interactiva (o consultiva), la persona que corregeix ha de proporcionar al personal de l'empresa vies àgils i resolutives d'evacuar consultes sobre la correcció dels textos que produeixen.

Finalment, en l'etapa pal·liativa trobarem els recursos que el mateix professional de la correcció farà servir per esmenar els textos que li arribin en suport electrònic. En aquesta etapa ens preocuparà no només explotar els recursos tecnològics amb la màxima eficiència, sinó també implantar unes pautes d'utilització que generin informació qualitativa i quantitativa sobre els errors que han requerit la seva intervenció per tal de retroalimentar les etapes preventiva i interactiva.

2.1. *Etapa preventiva*

En aquesta etapa, la intervenció tindrà dues línies d'acció: la configuració de les eines lingüístiques i la mediatització de l'ús del tractament de textos (TT).

En la primera línia, hem d'intentar que els usuaris tinguin eines adequades i, sobretot, que les facin servir; hem de conèixer-ne bé les limitacions (doncs, la tasca serà més fàcil si les eines són les mateixes per a tot el personal). Determinades pautes d'intervenció seran impracticables si cada ordinador està instal·lat de qualsevol manera: cal formalitzar per a cada ordinador les diverses carpetes auxiliars del TT, com les que contenen els diccionaris personals o les plantilles. Sense aquesta uniformitat no serà possible implantar una política sistemàtica d'actualitzacions de la configuració o de les eines lingüístiques.

Qualsevol propòsit de superar les limitacions de les eines lingüístiques (de correcció, però també de consulta, com el diccionari de sinònims) requereix haver avaluat el

diccionari general (cobertura de lèxic general, de tecnicismes, cobertura enciclopèdica i onomàstica, suggeriments, sensibilitat geogràfica...), les opcions del diccionari d'usuari, els mòduls de partició de mots, de sinonímia i, en cas que n'hi hagi, l'operativitat de la correcció gramatical. Avaluarem en l'apartat següent els principals verificadors i correctors, que és evident que tenen importància tant en aquesta etapa com en la pal·liativa.

La segona línia, mediatització de l'ús del tractament de textos, cerca un principi bàsic de la història de la humanitat: la mandra com a motor del canvi. Hem d'aconseguir que a la nostra clientela li costi menys fer les coses ben fetes que no pas fer-les malament. Per exemple, si cada vegada que comencen a escriure el nom de l'empresa, després de tres pulsacions el TT els ofereix el nom complet per tal que es limitin a prémer la tecla de retorn (↵), probablement aconseguirem estalviar-nos una correcció freqüent.

Si cerquem a la xarxa Internet una cosa tan imprevisible com «Pompeo Fabra», ens sortiran noranta pàgines que tenen aquesta cadena de caràcters (figura 1), i en trobarem dinou que tenen «Pompeu Favra» (figura 2). En canvi, si la nostra clientela es troba que, en el moment d'haver escrit «Univ» se li desplega el nom complet (figura 3), probablement ens estalviarem col·leccionar errors com els que segueixen:

FIGURA 1

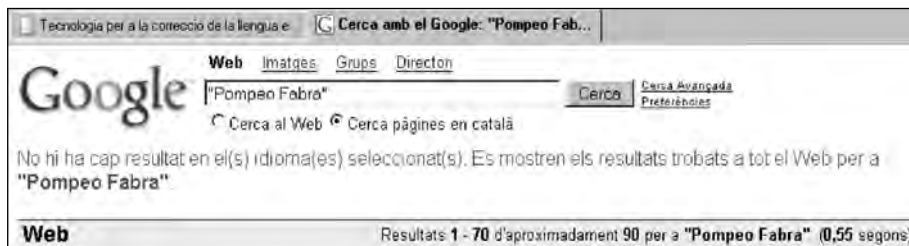


FIGURA 2

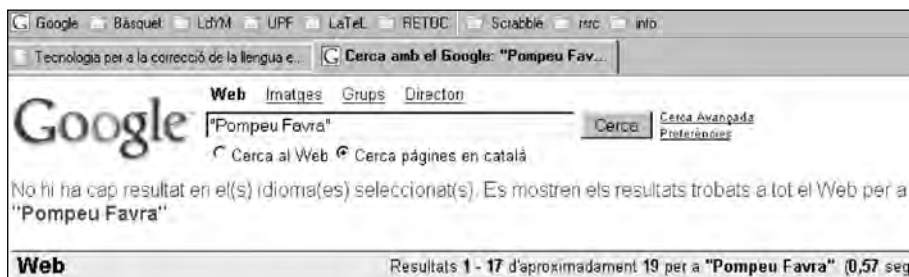
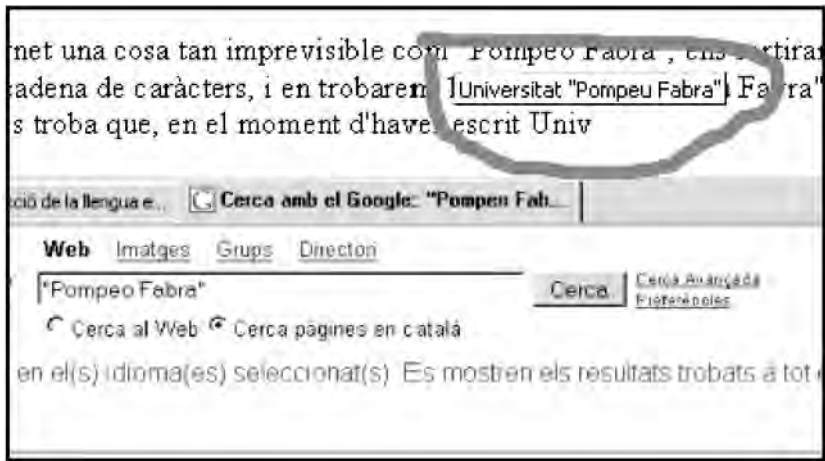


FIGURA 3



Sembla fàcilment argüible que no ens podem equivocar en allò que no fem, de manera que, si la nostra clientela no arriba a escriure «Pompeu Fabra» perquè el TT ja li proposa el nom complet de la universitat, no haurem de corregir ni «Pompeu» ni «Favra».

La llista de recursos a la nostra disposició inclou les plantilles (que poden incorporar macros per al farciment assistit de continguts, caixes de text protegides, predefinició dels idiomes en el cas dels documents multilingües); macros per assegurar l'activació de les opcions de configuració que calgui, per resoldre errors freqüents, per farcir de mots el diccionari d'usuari; l'autotext o l'autocorrecció... Per a qui no estigui familiaritzat amb termes com *plantilles*, *autotext*, *autocorrecció*, *macros*, etc., un glossari al final intenta facilitar una iniciació suficient.

Estic convençut que s'entén molt millor el que es pot esperar d'aquests recursos si es veuen en acció que no a partir d'una presentació teòrica. Per aquest motiu, entre els materials que he preparat per a aquest seminari (que es troben a l'URL <http://terminotica.upf.es/tallerCo/>) he inclòs algunes macros per a la correcció en sentit preventiu (taula 1):

TAULA 1

Verifica mots	Aquesta macro permet eliminar d'un document tots els mots que el verificador ortogràfic accepta. Ajuda a alimentar el diccionari d'usuari sense inflar-lo supèrfluament.
Neteja l'autocorrecció dels valors de fàbrica	Serveix per eliminar de l'autocorrecció els valors predefinitos. Com que la llista és molt llarga, s'ha de dividir en dues macros.
Autocorregeix les sigles i les abreviatures	Serveix per introduir en el Word unes determinades sigles o abreviatures, com ara UPF+, que es desplegaran gràcies a l'autocorrecció.
Afegeix nous valors a l'autocorrecció	Serveix per afegir nous valors per a l'autocorrecció. En aquest exemple apareixen bàsicament els pocs valors inicials de la versió 97 de l'MS Word que es poden aprofitar en català.
Afegeix nous valors a l'autotext	Introdueix uns valors d'autotext (text complet que el TT suggereix a mesura que l'usuari el va introduint).
Marca dubtosos	Destaca certes dificultats acolorint mots susceptibles d'error.
Rectifica guions	Adequa els mots verificats erròniament o no segons les normes anteriors del guionet, de manera que s'ajustin a la norma actual.
Marca manlleus	Passa a cursiva els manlleus acceptables. A més d'influir en l'aspecte del document, orientarà l'usuari en el moment de fer la verificació del document.
Amplia el diccionari personal	Afegeix una llista de mots al diccionari personal.
Prepara la macro precedent	Genera, a partir d'una llista, el codi de la macro que amplia el diccionari personal.
Elimina apòstrofs	Neteja del diccionari personal qualsevol addició errònia d'un element apostrofat a l'inici o a la fi del mot.

Un fet que cal destacar és que, tot i que no siguin idèntiques les solucions concretes que haurien d'habilitar les persones responsables de la correcció en dues seccions o delegacions diferents de la mateixa empresa o en departaments diferents de l'Administració, el noranta per cent de l'esforç d'implantar una política de correcció preventiva es pot desenvolupar conjuntament: cal esmentar en aquest sentit l'experiència DAU (*Documents administratius universitaris*) del Servei de Llengua Catalana de la Universitat de Barcelona (SLCUB), que va aplicar-se a les secretaries de totes les seves facultats amb les especificitats necessàries però a partir d'un esforç coordinat des del SLCUB.

2.1.1. Verificadors i correctors

En català tenim un bona colla de recursos en aquesta categoria, però ens limitarem a avaluar, per raons pràctiques, els que s'utilitzen prioritàriament a la plataforma Windows, majoritària a l'Administració, l'ensenyament i la majoria dels sectors productius. Són quatre: els recursos incorporats al TT de Microsoft (Word), dues eines comercials que interactuen amb l'MS Word (WordCorrect i Maxigramar) i els recursos incorporats al TT de l'OpenOffice. Aquesta darrera *suite* pren protagonisme per mor del pacte de govern del tripartit:

3. Liderar la modernització de les administracions públiques catalanes i implementar plenament l'Administració Oberta de Catalunya transparent, en xarxa, moderna, propera i integrada al servei de la ciutadania.

[...]

3.5. Fomentar la producció de programari lliure i la corresponent documentació d'ús en català.

3.6. Utilitzar per part de la Generalitat de Catalunya, els organismes autònoms públics, les empreses públiques i les de capital mixt sota control majoritari de la Generalitat de Catalunya preferentment en els seus sistemes i equipament informàtic programari lliure en català.

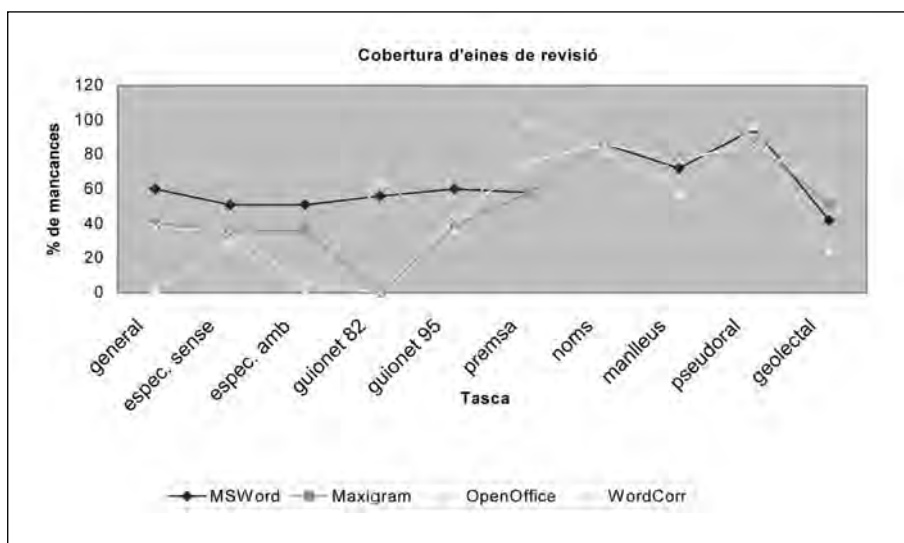
Per materialitzar aquest punt programàtic sense detriment del català, cal incloure el producte líder en *software* lliure, el paquet ofimàtic OpenOffice.

He establert un conjunt de proves més centrades en la verificació que no pas en la correcció gramatical. Això es deu al fet que, tant per la meua pròpia estimació

com per consens en les observacions que he recollit dels usuaris, la part de la correcció gramatical és notablement menys satisfactòria i provoca més feina que no pas n'estalvia, cosa que implica que a un usuari insegur li crea més dubtes que els que li resol, i a un que no dubta li fa més nosa que servei. Entre aquestes proves hi he inclòs també una mesura de la variació dialectal admissible en estàndard (de fet, tots els tractaments de textos i programes que inclouen verificació, com els sistemes de traducció assistida, solen recollir avui en dia les variants regionals de les llengües: castellà de l'Argentina, anglès de Jamaica, francès de Suïssa, alemany de Liechtenstein...), quan encara no sabíem que el nostre Govern ha incorporat el programa valencià de traducció castellà-català Salt 3.0 a les eines que utilitza l'Administració autonòmica.

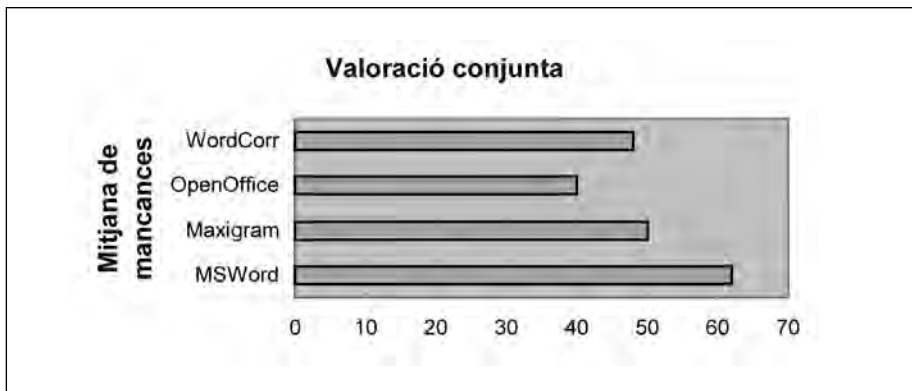
La dificultat d'objectivar el potencial de cadascuna d'aquestes eines rau en diferències relatives a qüestions com ara com validen els mots, com gestionen els diccionaris d'usuari, quants en permeten, quines estratègies segueixen a l'hora de fer suggeriments o fins a quin punt permeten ser manipulades per automatismes com les macros amb vista a mediatitzar-ne el funcionament segons els interessos de l'empresa. La figura 4 resumeix les proves que he fet i que no tinc ocasió de detallar; només en vull destacar que les diferències són en qüestions bones de resoldre.

FIGURA 4



Establint una ponderació combinada dels diversos comportaments avaluats que prioritza els aspectes relacionats amb la controlabilitat de les eines lingüístiques per part dels responsables dels serveis lingüístics, obtinc la gràfica de deficiències següent (figura 5), que deixa l'OpenOffice en la millor posició i l'MS Word en la pitjor, amb diferències importants però que es podrien compensar quasi totalment amb segons quina política de promoció d'aquests recursos.

FIGURA 5



De l'observació d'ambdues gràfiques se'n desprèn que les dues aplicacions comercials que es venen com a complements de l'MS Word tenen característiques quasi idèntiques.

2.2. Etapa interactiva

Aquesta etapa del procés correspon al moment en què la clientela està produint els documents que després ens arribaran per corregir. Tot el que fem per tal que puguin resoldre els seus dubtes satisfactoriament acabarà estalviant-nos feina. Per exemple, si amb un sol clic en una icona de la barra de tasques inferior del Windows poden accedir per la Xarxa a un servei, com un diccionari o una llista QS ('qüestions sovintejades' = FAQ o PMF), a la llarga ho notarem.

En l'exemple següent, la icona  del Microsoft Explorer ens durà amb un sol clic a la seu del *Gran diccionari de la llengua catalana* (GDLC):

FIGURA 6



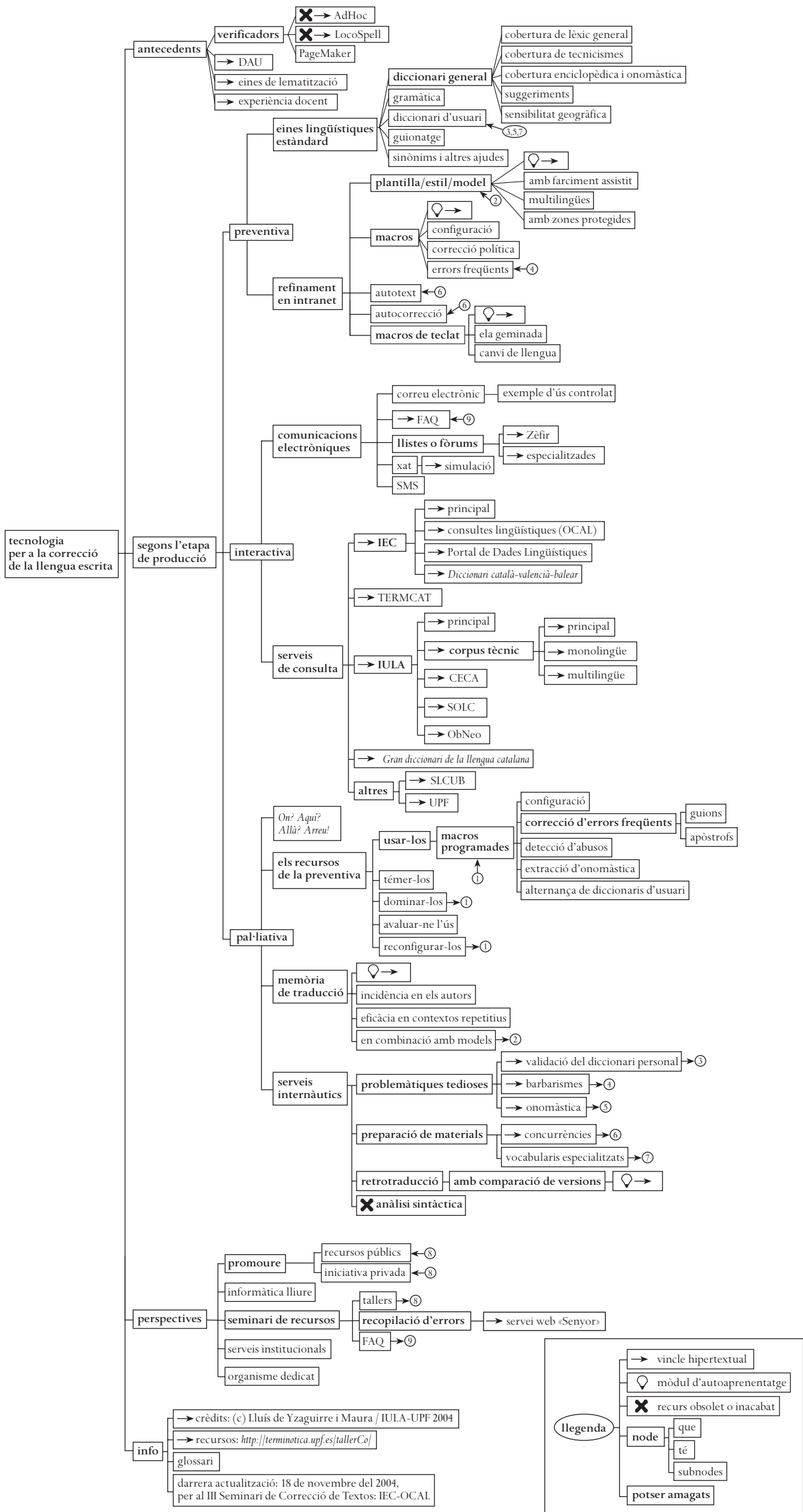
Els recursos per a l'etapa interactiva es poden subdividir en dos grups: serveis de consulta i canals de comunicació internàutica o electrònica.

Entre aquests darrers trobem el correu personal, les llistes QS (sistemes de consulta sobre preguntes freqüents relatives a una temàtica concreta), les llistes de correu (com Zèfir) (algunes, molt especialitzades), els fòrums, els canals de xat i fins i tot els SMS (pensant en professionals que tenen un dubte anant pel món: *v. gr.* una unitat mòbil d'un mitjà de comunicació que connectarà en directe d'ací a deu minuts...).

Pel que fa als serveis de consulta, podem esmentar-ne una bona pila, començant pels de l'IEC (força heterogenis: l'Oficina de Correcció i Assessorament Lingüístics, el Portal de Dades Lingüístiques i el *Diccionari català-valencià-balear*), el TERMCAT, l'IULA-UPF (el Corpus Tècnic Multilingüe, el corpus CECA, el servidor SOLC i l'Observatori de Neologia), el servei ja esmentat del GDLC i els serveis lingüístics de les universitats de tot el domini lingüístic.

Un concepte com el de *interactivitat* implica més d'un agent: per part nostra, l'aportació ha de ser la d'implantar en els ordinadors de la clientela procediments com més àgils millor d'accedir a fonts d'informació per resoldre dubtes i fer-los conèixer. Però això podria resultar insuficient si l'altra part no en sent la necessitat. Avanço en aquest punt que per practicar amb coherència una correcció preventiva cal que l'etapa paliativa inclogui l'estudi de les deficiències o desviacions de cada *pacient*, almenys de les més transcendents. De la mateixa manera que el metge que no es limiti a guarir ha de proposar canvis en la dieta, l'aire condicionat, les fibres de la roba, etc., qui practiqui la correcció preventiva ha de promoure entre la seva clientela la coneixença de tots aquests canals de resolució de dubtes.

FIGURA 7



llenda

- vincle hipertextual
- mòdul d'autoaprenentatge
- recurs obsolet o inacabat
- node
 - que
 - té
 - subnodes
- potser amagats

2.3. *Etapa pal·liativa*

En aquesta etapa, la intervenció de la persona que corregeix consisteix a pal·liar els problemes dels textos, però la seva feina no acaba ací, car l'estudi de com i on ha calgut intervenir ha de realimentar l'etapa preventiva.

Evidentment, la persona que corregeix farà servir en aquesta etapa almenys els mateixos recursos que utilitzi la seva clientela. Caldrà que els conegui per entendre sempre com li arriben els textos, ateses les limitacions de les eines de correcció.

A més d'emprar aquests recursos, la persona que corregeix farà bé d'ajudar-se amb altres recursos que li pot oferir la tecnologia actual:

- les macros li poden oferir vies per completar la supervisió dels textos (detecció i/o correcció d'errors freqüents o de vicis estilístics segons la procedència del document, qüestions no resoltes per les eines de correcció, com neologia o onomàstica...) per augmentar el seu potencial de treball (alternança àgil entre diccionaris d'usuari segons la temàtica o el departament dels documents...);

- la memòria de traducció li pot servir per crear una base de dades textual, que en el cas de documents repetitius (informes, certificats...), pot ser útil en la tasca de correcció (especialment en documents basats en alguna matriu, com ara les plantilles); en la resta de casos, aquesta base de dades pot ser de gran ajuda per a qualsevol estudi sobre les necessitats d'intervenció preventiva: formació dels usuaris, divulgació d'orientacions lingüístiques, auditoria de recursos (revisió de diccionaris d'usuari, d'autotext o d'autocorrecció...);

- determinats ciberserveis poden ajudar en algunes tasques especialment tedioses o per a les quals les eines estàndard resultin limitades: revisió dels diccionaris d'usuari, de barbarismes, d'onomàstica, extracció de concurrències...

Tot aquest plantejament, en un entorn professional, sigui institucional, sigui empresarial, implica, per coherència, que la coneixença adquirida sobre les mancances de la clientela s'utilitzi no solament per millorar l'eficàcia dels recursos preventius sinó també en plans de formació lingüística, per incidir en els processos de selecció de personal i en els de promoció interna. Normalment això no serà competència directa de qui corregeixi (excepte, potser, en serveis lingüístics amb un únic membre), però serà bo que la formació interna, si n'hi ha, tingui en compte les mancances que els serveis de correcció detectin.

3. Perspectives

Les noves tecnologies poden ser una eina de dinamització de qualsevol llengua que en tingui, en el mateix grau en què pot ser un obstacle a la seva pervivència no tenir-ne. Els professionals de la correcció poden treure un gran benefici d'anar assumint les innovacions en aquest àmbit i, en el cas de les llengües no lucratives, de promoure aquestes innovacions.

Algunes de les idees que s'han presentat en aquest article demanen recursos humans no negligibles; això, però, no ha de ser un obstacle si es volen tirar endavant. Se'n podria fer una promoció institucional si es trobés el suport adequat:

- podria ser una bona tasca per a l'Institut Català de les Indústries Culturals;
- alternativament, justificaria donar continuïtat al Centre de Referència en Enginyeria Lingüística;

- Política Lingüística podria assumir-ho si se suposa que només cal impulsar-ho per tal que després les empreses productores de tecnologia lingüística hi donin continuïtat;

- finalment, l'IEC podria muntar el seu propi laboratori de tecnologies lingüístiques, que, de passada, permetria a la Secció Filològica preveure les implicacions tecnològiques d'algunes de les seves decisions (com la modificació de les normes relatives al guionet el 1995).

Si no tenen suport institucional, diverses associacions professionals o de defensa de la llengua en podrien fer una promoció social, amb mecanismes cooperatius com els de *sourceforge.org* o *softcatala.org*, *weblogs* o blocs o bé QS.

Les vies apuntades suara, al marge dels canals comercials, no gaire propicis a les llengües no lucratives, s'haurien de concretar implementant al màxim les ajudes en aplicacions informàtiques lliures com OpenOffice i dinamitzant seminaris de recursos entre professionals de la correcció lingüística.

4. Agraïments

Aquesta comunicació és tributària d'unes quantes persones: en primer lloc, haig d'agrair la instigació de Teresa Cabré i Josep Maria Mestres; en segon lloc, vaig consultar fructíivament i fruitosa els meus col·legues Joan Costa i Josep Maria Vallhonrat; tercerament, cal que manifesti gratitud a l'empresa que produeix Maxigramar per les facilitats que em va donar per avaluar el seu producte; finalment, gràcies als bons ofi-

cis del seu mestre de cerimònies, en Màrius Serra, la comunitat de Verbàlia ens va proporcionar textos creatius per als bancs de proves.

5. Referències

- DAU: Documents administratius universitaris* (1991). Barcelona: Universitat de Barcelona. Servei de Llengua Catalana. [Les macros són accessibles a <http://traductica.upf.es/modular2000/DAU/DAU.htm>]
- HALVORSON, Michael; KINATA, Chris (1997). *Microsoft Word 97: Visual Basic paso a paso*. Madrid: McGraw-Hill.
- RUBIO, Salvador C. (2000). *StarOffice 5.1*. Barcelona: Marcombo Boixareu Editores. [NB: StarOffice fou el precursor de l'OpenOffice]

Els materials preparats com a demostració de les possibilitats tecnològiques en el camp de la correcció i el mapa conceptual dels continguts reproduït entre les pàgines 112 i 113 (figura 7) es troben a <http://terminotica.upf.es/tallerCo/>.

6. Glossari

- autocorrecció** *f.* Procediment per substituir automàticament text erroni o estenogràfic pel que li correspongui.
- autotext** *m.* Text predictiu que s'ofereix després d'unes poques pulsacions i que es completa pitjant la tecla de retorn.
- ciberservei** *m.* Servidor internàutic que permet automatitzar algun tipus de consulta o correcció.
- concurrència** *f.* Grup fixat de mots (en anglès, *collocation*).
- estil** *m.* Equivalent a plantilla en AmiPro i WordPro.
- macro** *f.* Conjunt d'instruccions informàtiques que permet automatitzar un procés com si fos l'usuari qui el realitzés cada vegada que s'executa la macro.
- model** *m.* Equivalent a plantilla en entorns Apple.
- paquet integrat** *m.* Equivalent a *suite*.
- plantilla** *f.* Document a partir del qual se'n crea un altre copiant-ne totes les opcions d'usuari: tipografia, idioma preconfigurat, marges, continguts fixos (en un formulari, *v. gr.*), macros...
- suite ofimàtica** *f.* Conjunt dels programes més habituals (tractament de textos, full de càlcul, agenda, navegador...).