

## **Noves tecnologies i llengua**

**L'autora repassa la complexa interrelació entre noves tecnologies i llengua: la importància de disposar d'aplicacions en diferents llengües i les dificultats que això representa. Finalment, se centra en les aplicacions que tenen com a objecte mateix la llengua, que esdevindran claus en el futur.**

Amb el títol de noves tecnologies s'agrupen un seguit d'aplicacions i serveis en àmbits molt diferents que tenen com a nucli comú les comunicacions i la informació, i com a eina principal la informàtica, que mitjançant el desenvolupament de programes dóna sentit a tota la informació que es transmet per diferents canals. La gran majoria d'aquestes aplicacions i serveis està pensada per ser usada per persones i, per tant, ha de comptar amb la llengua dels usuaris. La llengua està present en quasi totes aquestes aplicacions tecnològiques i per poder-nos ser útils, còmodes i entenedores, els usuaris demanem que se'ns adrecin i ens permetin d'adreçar-nos tot usant la nostra llengua.

Per comprendre l'abast i la complexitat de la relació entre llengua i noves tecnologies farem primer una mirada endarrere per recordar com n'estan de relacionades. Després incidirem en les dificultats per poder disposar d'aplicacions en diferents llengües i, finalment, ens referirem concretament al problema que suposen els programes que tenen com a objecte mateix la llengua. Aquests programes, com veurem, tenen encara capacitats limitades però, donada la seva importància, esdevindran clau en moltes de les aplicacions futures a les quals ens referirem en l'últim apartat.

Aquells que han començat a relacionar-se amb les noves tecnologies en els últims anys podrien pensar que els ordinadors treballen amb independència total de les llengües dels seus usuaris i que disposar de programes en una llengua concreta és solament una qüestió de traduir certs materials, però no és ben bé així. Comparat amb fa 15 anys, avui en dia no es fan evidents les dificultats de la relació tan directa que hi ha entre llengua i aplicacions informatitzades: no sembla que hi hagi problemes per escriure en català en la major part d'aplicacions que treballen amb text, tot i que no totes les aplicacions tinguin els textos adreçats als usuaris en català. Però alguns altres recordaran que durant un temps els teclats no tenien accents, perquè tampoc hi havia una codificació<sup>1</sup> informàtica dels caràcters accentuats que permetés processar-los, amb la qual cosa tampoc no hi podia haver programes que els tinguessin en compte.

Poder escriure caràcters accentuats en qualsevol ordinador no hauria estat possible si els desenvolupadors d'aplicacions no haguessin arribat a la conclusió que era imprescindible permetre introduir, veure, imprimir i processar informació expressada en llengües que tenien caràcters que no corresponien als de l'anglès.

**Codificació de  
caràcters gràfics**

Tot fent memòria cal dir que el primer problema va ser arribar a una codificació estàndard dels caràcters gràfics, en aquell moment de l'anglès. L'ASCII (American Standard Code for Information Interchange) va ser la primera codificació estàndard dels caràcters i símbols gràfics; és a dir, una representació única usada per tots els ordinadors i programes. La creació de l'estàndard ISO-8859 per a les llengües alfabètiques el 1987 va ser el reconeixement internacional que les aplicacions informàtiques havien de pensar-se per a usuaris amb diferents llengües i, per tant, amb diferents caràcters gràfics. I no va ser fins al 1991 que no es va arribar a la representació UNICODE2 com a codificació estàndard que assigna una representació única a tots els caràcters i símbols per escriure totes les llengües del món. La implantació encara és deficitària, tot i que els teclats s'adapten a les necessitats gràfiques de les diferents llengües, molts programes i aplicacions encara no l'usen o ho fan solament per a algunes llengües. Encara ara els accents no són determinants en els criteris de cerca en moltes bases de dades, per exemple; el programa no els sap interpretar. El problema de la seva implantació pot estar en la ignorància o la manca de previsions dels desenvolupadors pel que fa a les necessitats dels usuaris potencials, però també en el cost de la transició que pot suposar reescriure una bona part del codi de l'aplicació.

A més a més, a banda de les aplicacions informàtiques s'han anat ampliant els components de maquinari amb els quals podem entrar-hi o veure-hi dades. En els darrers anys s'han anat succeint les aparicions de nous dispositius: els escàners, les taules digitalitzadores, el llapis òptic, el micròfon, el telèfon mòbil, són eines per entrar dades en aplicacions informàtiques. I pel que fa a la sortida, la pantalla de l'ordinador i la impressora deixen pas a la pantalla de la televisió, del telèfon mòbil. Cada un d'aquests nous dispositius ha de reconèixer amb els caràcters de totes les llengües dels usuaris potencials i treballar-hi.

### **Interfícies: la comunicació amb els sistemes informàtics**

La gran majoria d'aplicacions informàtiques estan adreçades a usuaris que han de donar instruccions per tal que els ordinadors facin tasques determinades. S'anomenen *interfícies d'usuari* els programes que gestionen aquesta comunicació entre les persones i els ordinadors. Durant molts anys, els esforços dels informàtics han anat dirigits al desenvolupament d'interfícies gràfiques (GUI: Graphical User Interface) que facilitessin la comunicació amb unes màquines que per la seva simplicitat de funcionament es feien complexes a les persones. L'ús d'icones semblava que garantia la comprensió de l'usuari i facilitava l'ús de funcions cada cop més complexes. El canvi de les primeres interfícies de comandament del DOS, on calia aprendre's els comandaments en anglès (per exemple, *delete*, per esborrar un fitxer) o l'ús de les tecles de funcions, a les atractives pantalles gràfiques amb icones que van suposar les innovacions dels Apple primer, i que va popularitzar el Windows de Microsoft després, van suposar un canvi substantiu.

La popularització dels ordinadors gràcies a aquestes interfícies gràfiques que en facilitaven l'ús va fer que la informàtica entrés en les lleis del mercat més tradicionals, en el qual els proveïdors, per guanyar valor afegit, miren d'acostar-se a les necessitats dels usuaris: també les lingüístiques. A l'adequació lingüística i cultural dels programes informàtics se li va posar el nom de *localització*: adaptació del programa a característiques locals. En la majoria de casos la localització significava molt més que no solament la traducció dels textos dels menús, diàlegs i de la documentació d'ajuda. La part textual estava, i encara està per a molts, entre les línies del codi informàtic del programa, i s'intercala a les instruccions particulars de la interfície gràfica. Per exemple, el codi per obrir a la pantalla una finestra en

la qual apareix un text determinat conté instruccions precises de les dimensions de la finestra per tal que hi càpiga el text en qüestió; la dificultat de la localització en aquests casos, en traduir el text, és que l'espai que la traducció ocupa pot ser més gran o més petit, així que cal redefinir les indicacions de les dimensions de la finestra, o redactar una versió del text que tingui la mateixa longitud que el text en la llengua origen. Així, en augmentar les capacitats i les funcions del programa, augmenta també la quantitat de feina necessària per fer la versió en una altra llengua. Per fer-nos-en una idea: el Windows 3.1, de l'any 1992, tenia 3 milions de línies de codi. El Windows XP, la darrera novetat, en té 50 milions.

Cal dir, però, que la magnitud de la tasca no ens ha de fer sinó valorar l'estratègia comercial dels proveïdors de sistemes operatius que han anat comercialitzant versions locals del seu programari, tot i haver estat poc apreciades pel públic català: Apple, IBM, Sun Microsystems han disposat o disposen de versions locals en català dels seus sistemes operatius. Aquestes empreses, entre d'altres, han estat les que, conscients del problema de recursos que la disponibilitat de versions locals suposava, han estat líders en el disseny i desenvolupament de mètodes i estàndards que minimitzen els recursos necessaris per fer diferents versions locals. Aquesta iniciativa ha estat anomenada *internacionalització* dels programes,<sup>3</sup> i els objectius són aconseguir que:

- Un executable al qual se li subministren dades locals (moneda i manera d'expressar-la, per exemple) pugui funcionar adequadament per a qualsevol versió local.
- Els elements de text, com ara els missatges i els noms dels components de la interfície gràfica, no estiguin dintre del codi. Han d'estar emmagatzemats fora del codi font i han de ser recuperats dinàmicament.
- El suport per a noves llengües sigui possible sense necessitat de recompilar el codi font.
- El desenvolupament de versions locals es pugui fer en un període curt de temps.

La localització o el desenvolupament de versions locals esdevé llavors el procés de traduir i adaptar a les dades locals els elements de la interfície d'usuari que són, ara sí, externs a l'aplicació mateixa.

La globalització del mercat, amb possibles usuaris arreu, és el factor determinant perquè les empreses del sector informàtic dediquin més interès i esforços a dissenyar les aplicacions amb aquests requeriments. La internacionalització del programari esdevé una de les seves prioritats tot i que això suposi reescriure gran part del codi. Tampoc es poden permetre altres estratègies si tenim en compte que l'element decisiu ha estat l'obertura del mercat xinès. Ara bé, com ja hem vist abans amb l'adopció d'UNICODE, la implantació serà lenta per la naturalesa mateixa de l'exercici.

A Microsoft estan tot just ara començant a aplicar estratègies que els permetin cobrir les demandes d'aquest mercat cada cop més ampli i més exigent: internacionalitzar el seus programes, és a dir, reescriure el codi informàtic de les seves aplicacions més conegudes, i desenvolupar un sistema de traducció automàtica per poder donar suport documental a Internet a desenvolupadors i usuaris de tot el món. Amb el que anomenen Multilingual User Interface o MUI, l'usuari pot triar una llengua (i altres característiques locals) entre diverses per tal de veure la versió local de la interfície d'usuari, els menús, els diàlegs i les ajudes. L'anomenat Language Interface Packages ofereix la possibilitat d'instal·lar tots aquests components per a una llengua específica a sobre del nucli comú del sistema operatiu i

la interfície d'usuari en una altra llengua. Aquesta última és la solució que s'ha adoptat per a la versió de Windows XP en català,<sup>4</sup> on les necessitats dels usuaris no estan tan lligades a triar una llengua entre diverses, com pot succeir en organitzacions d'àmbit internacional, com a disposar d'una interfície d'usuari en català.

Però, pel que fa a la disponibilitat de la versió local catalana d'aplicacions d'ús popular, val a dir que més rellevant que la política comercial de les empreses d'informàtica ha estat la implicació i motivació del col·lectiu voluntari català. La voluntat de persones sense cap altre interès que disposar de programes en català ha fet possible la disponibilitat de diferents programes amb interfícies localitzades, com ara les darreres aplicacions adaptades per Softcatalà.<sup>5</sup>

L'OpenOffice (per a Windows, Linux i Solaris) és la versió de lliure distribució del paquet ofimàtic StarOffice, propietat de Sun Microsystems. Es tracta d'un producte molt complet que conté, de forma integrada, un processador de textos, un gestor de fulls de càlcul, diverses eines per a la creació i edició de presentacions, dibuixos vectorials, gràfics i fórmules, i diverses utilitats de conversió de fitxers (entre altres, els formats de Microsoft Office) i un corrector ortogràfic per al català integrat. Un altre processador de textos en versió catalana, i també de distribució lliure, és l'AbiWord (per a Windows i Unix). El Gimp és un programa de manipulació d'imatges de lliure distribució que permet fer manipulació de fotografies, composició i creació d'imatges. El Navegador 4.7 (per a Windows, Unix i Mac) és la traducció al català del conegut Netscape Communicator 4.7; permet navegar per Internet, enviar i rebre correu, accedir als grups de notícies, i crear pàgines web, tot en català. També ha estat localitzat el navegador Mozilla 1.2.1, amb la qual cosa el col·lectiu Softcatalà es decanta definitivament per les iniciatives de codi lliure. El Mozilla és també una eina d'ús d'Internet que permet explorar els continguts, enviar i rebre correu electrònic, consultar i participar en els fòrums de notícies i crear pàgines web.

### **Programes que tenen com a objecte la llengua**

Però donar suport a les llengües dels diferents usuaris va més enllà de localitzar la interfície d'usuari. Cal diferenciar entre la disponibilitat d'aplicacions amb interfície d'usuari i documentació en una llengua determinada i la d'aplicacions en les quals el material lingüístic és l'objecte de les funcions del programa, com ara les de cerca de mots, correcció ortogràfica, la cerca d'informació mitjançant paraules, etc.

Un àmbit en el qual la llengua és especialment important és Internet. La informació a Internet és, per definició, accessible per a tothom, però per les seves dimensions el problema rau a trobar-la. És per aquest motiu que des dels seus inicis les aplicacions per a Internet han esmerçat molts esforços a ajudar els usuaris a trobar-la, bé ordenant-la manualment en directoris sota arbres d'etiquetes o noms que defineixen el contingut de les pàgines web que s'esmenten, bé desenvolupant motors de cerca que permeten trobar ràpidament en el text les paraules clau que serveixen com a criteri de selecció. Els directoris i els cercadors són, així, les dues grans aplicacions a Internet. Cal dir que els directoris en català han estat entre els pioners: Vilaweb, nosaltres.com, cercat.com, han estat subministrant informació als usuaris catalans. Darrerament cercadors i directoris d'àmbit internacional han incorporat el català com a llengua de les seves interfícies (Google va tenir primer una adaptació al català de les seves pàgines feta per Softcatalà) i ara donen ple suport a la cerca en pàgines identificades com a escrites en català. Avui comptem amb els següents cercadors i directoris que donen suport al català: Lycos (l'última incorporació), Yahoo en català, Google, Alltheweb. És



important esmentar que els cercadors a Internet basen la seva activitat en els motors de cerca i indexadors i que hi ha 6 grans motors dels quals solament dos donen suport a tots els caràcters del català: Google Search Engine i Inktomi.

Pel que fa als correctors ortogràfics, aquests han de tenir no ja els caràcters sinó llistes de les paraules de cada llengua per poder comparar i detectar les errades. Tanmateix, aquesta informació particular de cada llengua en forma de diccionari de paraules també és necessària per fer òptims els resultats d'altres aplicacions, com el reconeixement òptic de caràcters<sup>6</sup> (OCR: Optical Character Recognition) i el reconeixement d'escriptura a mà,<sup>7</sup> que incorporen dispositius com les agendes electròniques (PDA: Personal Digital Assistant), un dels dispositius electrònics més venuts darrerament.<sup>8</sup>

Pel que fa als correctors ortogràfics, el mercat al nostre país té uns productes líders clars que són els de Microsoft, que ha incorporat els recursos lingüístics necessaris per donar suport en la correcció ortogràfica a 48 llengües en l'última versió d'Office XP (2002), per cert, amb una nova versió del corrector ortogràfic per al català. Però altres productes que donen suport al català en la correcció ortogràfica són els productes Adobe (FrameMaker, PageMaker i Acrobat), els paquets ofimàtics d'IBM i Lotus, així com els d'OpenOffice i StarOffice de Sun.

La disponibilitat de correctors gramaticals i d'estil és menor pel que fa a nombre de llengües i a productes que l'incorporen (ara per ara, en català solament existeix el producte WordCorrect).<sup>9</sup> En aquestes aplicacions els coneixements específics de la llengua comprenen no solament les paraules, sinó també la possibilitat de combinacions de paraules en una llengua perquè el que detecten són errors de sintaxi: errors de concordança, d'inserció de mots en context, etcètera, i per tant han de conèixer també la gramàtica i disposar d'eines que analitzin el text d'acord amb aquesta. Com més coneixement gramatical de la llengua en qüestió tenen, millors resultats ofereixen.

Però el paradigma dels programes informàtics que requereixen informació específica i detallada sobre les llengües i sobre les relacions entre les llengües són, sens dubte, els traductors automàtics. Podem classificar l'ús de les eines de traducció automàtica (TA) segons les necessitats que poden satisfer. Bàsicament, la traducció automàtica ha estat, des dels seus inicis, una solució a una gran demanda: traduir de forma barata i ràpida. Avui en dia, però, es considera més que un substitut, una eina per millorar la productivitat dels traductors, i els seus usuaris pertanyen bàsicament a grans organitzacions, amb gran capacitat d'inversió i amb demandes de traducció molt altes. Exemples d'usuaris que han inclòs la traducció automàtica a la seva metodologia de producció de documentació són, a més a més de la Generalitat de Catalunya, l'empresa de desenvolupament de programari SAP, Daimler Chrysler, Microsoft, la PanAmerican Health Organisation (PAHO), la US Intelligence Community, els serveis de traducció de la Comissió de la Unió Europea o l'agència de traduccions CLS Corporate Language Services AG. Alguns d'aquests usuaris han publicat dades sobre l'increment de productivitat pel fet d'usar traducció automàtica: entre un 30% i un 40% de reducció del temps de treball (fins a arribar a un 60% en combinació amb altres eines, com memòries de traducció).<sup>10</sup> La percepció dels usuaris dels sistemes de TA el pot resumir la declaració dels traductors de la Unió Europea<sup>11</sup> en parlar del traductor Systran, en el qual han estat implicats durant 25 anys:

The wide and growing use of MT within the European Commission can be seen as reflecting the enormous potential demand for the technology

## **Traducció automàtica**

in a professional environment; but should not be taken as a measure of the satisfaction of the Commission's Officials with the linguistic output of the system. Much remains to be done, and enhancement of the 18 existing language pairs is being pursued.

Les eines de traducció automàtica es consideren una solució específica per a continguts que, per les seves característiques de caducitat o de volum, no poden traduir-se manualment donat l'alt cost econòmic i de recursos humans que això implicaria, i per les característiques dels destinataris, es fa necessària una solució multilingüe. Un exemple clar és la proposta en l'estudi Preliminary Study for the Implementation of a Portal for the EU Administration,<sup>12</sup> que inclou la possibilitat d'usar traducció automàtica en certs nivells de la informació.

A més a més, els sistemes de traducció automàtica han esdevingut eines conegudes i usades pel gran públic en l'àmbit d'Internet. Des dels anys 90 ençà han proliferat les eines que ofereixen una traducció de poca qualitat que és, malgrat tot, interessant per al públic que vol accedir al contingut de planes web, fòrums, correus electrònics i documents accessibles en llengües que desconeix. La qualitat de la traducció no és tan important com la velocitat d'obtenir una versió intel·ligible. Les eines de traducció automàtica s'incorporen ara a portals, i pàgines web de mitjans, com un servei afegit. Exemples d'aquests són Lycos, Terra, elmundo.es, elpais.es, Altavista, AOL, Compuserve i Google, entre d'altres.

Hi ha indicadors<sup>13</sup> que apunten, doncs, a un ús limitat però en augment dels traductors automàtics. Ara bé, tot i que existeix una gran varietat de productes, lligada moltes vegades a les llengües disponibles, podem parlar de 6 grans productes per a llengües europees:

- iTranslator<sup>14</sup>
- Systran<sup>15</sup>
- Reverso<sup>16</sup>
- Compendium<sup>17</sup>
- IBM Websphere Translation Server<sup>18</sup>
- Logos<sup>19</sup>

S'han fet pocs estudis de comparació metodològicament correctes, i menys encara que s'hagin fet públics. L'últim que coneixem publicat és Elisabeth Maier i Anthony Clarke (2001):<sup>20</sup> *Evaluation of machine translation systems at CLS Corporate Language Services AG (U)*, Proceedings of VIII MT Summit, 2001.

Com a aplicacions d'ús personal o com una eina més a Internet, la traducció automàtica és un producte comercial i útil. Pel que fa al suport del català, actualment existeixen cinc sistemes que tradueixen del castellà al català i un traductor anglès-català en tots dos sentits. Els sistemes són: Castellà—>Català: Ara<sup>21</sup> i SALT;<sup>22</sup> Català<—>Castellà: Incyta-GenCat,<sup>23</sup> InterNostrum,<sup>24</sup> AutomaticTrans<sup>25</sup> i Compendium,<sup>26</sup> empresa que també disposa d'un sistema Català<—>Anglès.

## **Reconeixement i síntesi de veu**

Avui dia ja existeix la possibilitat d'usar la veu per comunicar-se amb sistemes informatitzats. Aquestes aplicacions es serveixen de tecnologies de reconeixement i síntesi de veu que també requereixen informació específica de la llengua en la qual han de permetre als usuaris dialogar amb sistemes informatitzats i a hores d'ara tenen cada cop més èxit en aplicacions d'atenció i informació telefònica. Tanmateix, les previsions<sup>27</sup> són una presència cada cop més important com una manera de facilitar l'accés a la informació i als serveis informatitzats disponibles amb dispositius de mà, els telèfons mòbils i les agendes digitals, que

per les seves dimensions fan incòmode l'ús del teclat, i els ordinadors del cotxe, on també se'n pot fer ús.

El nombre de subministradors i integradors d'aquestes tecnologies és cada cop més nombrós i si fa 5 anys no hi havia cap empresa que comercialitzés aquests productes per al català, ara disposem de 4 proveïdors: ScanSoft,<sup>28</sup> que ha comprat la tecnologia desenvolupada per Philips Speech Processing; Atlas,<sup>29</sup> una empresa catalana que comercialitza la tecnologia desenvolupada pel grup de processament de la parla de la Universitat Politècnica de Catalunya, amb qui es va desenvolupar el sistema Attemps d'informació meteorològica;<sup>30</sup> Telefónica I+D, que ha desenvolupat aplicacions per a les diferents llengües de l'Estat com la de l'Agència Tributària,<sup>31</sup> aplicacions de banca telefònica, i serveis a mòbils entre d'altres, i l'empresa italiana Loquendo, que també té reconeixement i síntesi de veu per al català<sup>32</sup> en un catàleg de 21 llengües disponibles.

Les tecnologies que sortiran al mercat en els propers anys estan ara als laboratoris de recerca aplicada i, deixant a banda sorpreses, podem tenir ja un tast del que vindrà a partir de les línies de desenvolupament d'aplicacions que s'estan seguint. Tant els laboratoris d'empreses com els acadèmics es concentren a millorar principalment dues línies de productes. Una, de la qual ja hem parlat, són les interfícies d'usuari. L'altra és la gestió de l'enorme quantitat d'informació posada a l'abast dels usuaris per fer-la valuosa. Ara per ara són línies de treball separades, però tendeixen a convergir en una de les aplicacions més desitjades: l'assistent personal.

Repassarem breument les dues línies esmentades tot remarcant l'important paper que la llengua hi té reservat. Des del punt de vista de la tecnologia, és important garantir la internacionalització de tots aquests programes que esmentarem, però pel que hem anat veient fins ara, assolir un bon nivell de funcionament per a cada una implica la necessitat de tenir components dedicats a emmagatzemar i processar les dades lingüístiques de cada una de les llengües per a les quals es vol que l'aplicació funcioni. Ja no és una qüestió de disposar d'una versió local del programa, sinó que tota l'aplicació no podrà funcionar per a una llengua si no es desenvolupen components específicament dedicats.

Les interfícies d'usuari han de deixar de ser sistemes gràfics lligats a una pantalla per permetre la interacció entre qualsevol tipus de servei automatitzat i l'usuari. Les possibilitats que ofereix la xarxa per al comerç electrònic juntament amb la varietat de formes d'accedir-hi (telèfon, televisió, ordinadors, quioscs, etc.), fan que les interfícies de veu siguin una de les línies de desenvolupament que compta amb més recursos actualment.

La comunicació interpersonal independentment de la llengua és, també, una de les fites tecnològiques, amb projectes com el TC-STAR, on participa el grup TALP de la Universitat Politècnica de Catalunya i on es considera el català com a llengua d'aplicació, que treballa en el desenvolupament de recursos i tecnologia per a la traducció automàtica de llengua oral. El reconeixement i la síntesi de la parla han d'anar un pas més enllà per fer que es puguin posar en correspondència diferents llengües i per això han de resoldre la identificació del continguts dels missatges, tot implicant el processament de les dades lingüístiques.

També amb reconeixement de veu i processament del contingut es volen desenvolupar aplicacions de seguretat. Els darrers esdeveniments mundials han fet incrementar els fons destinats a desenvolupament de sistemes que espion les

### **El futur de les noves tecnologies en relació amb la llengua**

### **Processament de la veu**

converses siguin telefòniques o per Internet (correus electrònics, fòrums, etc.) a partir de la identificació de paraules, com fan els cercadors.

## **Processament d'informació**

Internet és ara una gran font d'informació a la qual podem accedir des de qualsevol punt, una informació interrelacionada per la qual podem navegar, i una informació que en gran part és textual i, per tant, expressada en diferents llengües.

El primer problema a resoldre és la sobrecàrrega d'informació. Cal fer tota aquesta informació operativa per a l'usuari. Les tecnologies de recuperació i gestió d'informació tenen ja aplicacions comercials de primera generació que poden processar grans quantitats de material textual per tal de, per exemple, classificar documents de forma automàtica (*automatic categorization*), derivar resums de documents (*automatic summarization*), extreure informació rellevant (*data mining*) i extreure informació sobre les activitats dels usuaris de la informació per tal de crear les dades i poder fer arribar automàticament la informació pertinent per a cada usuari. Les aplicacions del futur usaran aquestes tecnologies per a tasques com la resposta automàtica (*question answering*) a correus electrònics o preguntes fetes per telèfon, entenent que l'usuari podrà fer arribar preguntes escrites o parlades i que un sistema automatitzat trobarà la resposta adequada.

Pel que fa a la xarxa, estrictament, les línies de recerca que s'estan plantejant és fer que a més a més de tenir la informació relacionada amb vincles, diguem-ne estàtics, es tracta de fer programes, a partir de la identificació de les accions més comunes dels usuaris, que s'encarreguin de dur-les a terme automàticament. Si molts humans estan fent les mateixes accions, per què no es poden fer programes que les facin automàticament? Hi ha molts exemples pràctics. Per posar-ne un parell: un autor d'una pàgina web ha de fer els vincles a diccionaris per facilitar a l'usuari la comprensió de paraules no gaire conegudes, o bé els mateixos usuaris han de marcar la paraula amb el cursor, obrir la pàgina d'un diccionari o base de dades terminològiques i enganxar la paraula en qüestió per obtenir-ne una traducció. Una acció que fem tantes vegades podria sens dubte estar inclosa en el nostre navegador com una funció més. Assenyalem la paraula i demanem amb una funció extra que ens trobi la definició als diccionaris a la xarxa, o als nostres preferits, o als diccionaris que prefereixen la major part d'usuaris que han consultat sobre la mateixa paraula. O també podríem tenir una funció per trobar-ne més referències, o buscar una altra pàgina web del mateix tema sense necessitat d'anar al cercador i repetir-la per fer la cerca. I, a més a més, voldríem la definició correcta atenent al context on l'hem trobada.

Les línies de recerca aplicada estan treballant en sistemes per fer possibles aquestes funcions tenint en compte el context de cada acció: qui fa la demanda actual, si algú l'ha fet ja i pot treure informació d'accions similars anteriors emmagatzemant la informació sobre les decisions preses, potser a partir de l'opinió explícita d'altres usuaris sobre una pàgina web. Les possibilitats són moltes però és necessari afegir informació sobre la informació mateixa (són les metadades) per fer que es puguin desenvolupar eines que puguin fer coses amb totes les dades que ens poden subministrar Internet i l'ús que en fan els usuaris. La informació ha de tenir un sentit definit, un significat definit. Per això es diu que s'ha de fer un «web semàntic»: que tingui sentit per a les eines, per desenvolupar eines. Les metadades s'usaran per afegir valor als serveis: vendre informació específica a determinats subscriptors, redirigir el tràfic, descobrir informació que pot ser venuda, etc.

La major part d'aquesta informació sobre la qual volen obtenir metadades és en forma de text, i ha de ser així per tal que els humans la puguem llegir i crear amb



naturalitat. Una possibilitat per crear dades sobre les dades és processar el text per obtenir informació sobre el que diu. Però per tal de processar-lo necessitem eines i recursos específics per a cada llengua i aquí és on la disponibilitat d'aquestes eines pot posar en joc l'ús d'una llengua a la xarxa. Si per escriure en una llengua determinada no podem tenir les mateixes funcionalitats que si n'haguéssim triat una altra, estarem discriminant els continguts per la llengua en què estan escrits.

## Notes

1. La codificació o assignació d'una representació permet la traducció al codi binari que és, en últim terme, el que processa l'ordinador.
2. <<http://www.unicode.org>>.
3. Abreviat a vegades I18N, que fa referència a les 18 lletres que van des de la I a la N en la paraula anglesa *internacionalization*.
4. L'anomenat per Microsoft Catalan-LIP (Language Interface Pack) està disponible de forma gratuïta a: <<http://www.microsoft.com/spain/windowsxp/catalan/default.asp>> i s'ha d'instal·lar sobre la versió castellana de Windows XP Professional.
5. Podeu consultar la llista sencera a <[www.softcatala.org](http://www.softcatala.org)>. Aquests programes han estat distribuïts en un CD-ROM pel diari *Avui* i se'n poden trobar còpies als indrets esmentats a: <<http://cultura.gencat.net/llengcat/informat/preguntes.htm#open>>.
6. IRIS Readiris Multilingual OCR – MAC, PRO), FineReader 6.0 Professional for Windows i OmniPage PRO de l'empresa ScanSoft, són programes de reconeixement òptic de caràcters que publiquen el reconeixement específic del català.
7. PenReader (Paragon Software) és un programa de reconeixement d'escriptura manual en agendes electròniques (PDA) o ordinadors de butxaca (WinCE/Pocket PC, Linux, Windows(9x/NT/2000/ME)). És l'únic que hem trobat que esmenti el català entre les llengües que inclou.
8. El 2001 se'n van vendre més de 10 milions d'unitats segons la consultora IDC.
9. La versió professional del verificador WordCorrect és compatible amb Windows XP, i es pot adquirir a les llibreries de la Generalitat i en establiments especialitzats.
10. Brundage, J. (2001) *Machine Translation, Evolution not Revolution*, Proceedings of the VIII MT Summit, 2001.
11. Angeliki Petrīts, Francine Braun-Chen, Cameron Ross, Rosemarie Sauer, Angelo Torquati, Alain Reichling: *The Commission's MT System: Today and Tomorrow*. Proceedings of VIII MT Summit, 2001.
12. <<http://europa.eu.int/ISPO/ida/export/files/en/730.pdf>>.
13. L'estudi *Language Translation: World Market Forecasts, Industry Drivers & eSolutions*, d'Allied Business Intelligence Inc., parla d'un mercat mundial de traducció d'uns 5,7 bilions de dòlars, dels quals 117 seria el volum del mercat de la traducció automàtica. L'estudi de Lionbridge Technologies Inc. del 2001, *When to use Machine Translation*, també fa una predicció d'un ús més extens d'eines de traducció automàtica per als propers 3-5 anys, amb independència de millores en la qualitat de la traducció.
14. <[http://itranslator.bowneglobal.com/BGSX/BGSXeng\\_us-EntryPage.htm](http://itranslator.bowneglobal.com/BGSX/BGSXeng_us-EntryPage.htm)>.
15. <<http://www.systransoft.com>>.
16. <<http://www.softissimo.com>>.
17. <<http://www.comprendium.biz>>.
18. <<http://http://www-3.ibm.com/software/pervasive/tech/demos/translations.html>>.
19. <<http://www.logos.it>>.
20. Elisabeth Maier i Anthony Clarke (2001): *Evaluation of machine translation systems at CLS Corporate Language Services AG (U)*, Proceedings of VIII MT Summit, 2001.
21. <<http://www.ara-autotrad.com>>.
22. <[http://www.cult.gva.es/DGOIEPL/SALT/salt\\_programes.htm](http://www.cult.gva.es/DGOIEPL/SALT/salt_programes.htm)>.
23. Disponible a la intranet corporativa de la Generalitat de Catalunya.
24. <<http://www.internostrum.com>>.
25. <<http://www.automatichtrans.es>>.
26. <<http://www.comprendium.biz>>.
27. La consultora The Kelsey Group va publicar un informe l'octubre passat en què afirmava que almenys la meitat dels usuaris de telefonia mòbil usaran algun tipus de portal de veu l'any 2005: <<http://www.kelseygroup.com>>.
28. <<http://www.scansoft.com/philips>>.
29. <<http://www.atlas-cti.com>>.
30. El telèfon d'accés del servei Attemps és el 906 789 987.
31. El telèfon d'accés del servei d'informació de l'Agència Tributaria és el 901 121 224.
32. El desenvolupament de la veu catalana de Loquendo ha comptat amb la col·laboració d'experts de la Universitat Autònoma de Barcelona: <<http://www.loquendo.com>>.