

# La gestió del multilingüisme a Internet i a les noves tecnologies de la informació i la comunicació

Miquel Strubell Trueta

PID\_00169662



Universitat Oberta  
de Catalunya

[www.uoc.edu](http://www.uoc.edu)



# Índex

<b>1. La gestió del multilingüisme a Internet.....</b>	<b>5</b>
1.1. Tria de la versió lingüística als portals web .....	5
1.2. Establiment de criteris d'ús lingüístic en espais multilingües .....	10
1.2.1. <i>Yahoo! Respuestas</i> .....	10
1.2.2. <i>Catalunya és Espanya. Cataluña es España. Catalonia is Spain</i> .....	11
1.3. Orientacions per al disseny de llocs web multilingües .....	12
<b>2. La gestió del multilingüisme en els programes informàtics..</b>	<b>13</b>
2.1. Microsoft .....	13
2.2. Softcatalà .....	14
2.3. Ubuntu .....	14
2.4. Orientacions per a l'elaboració de programari bilingüe .....	15
2.5. Altres aplicacions .....	15
<b>3. Eines per a la gestió de la diversitat lingüística en els textos</b>	<b>16</b>
3.1. Eines integrals de gestió del multilingüisme .....	17
3.2. Els traductors automàtics .....	18
3.3. La traducció assistida per ordinador .....	28
3.4. Les tecnologies de la parla .....	29
<b>Resum.....</b>	<b>31</b>



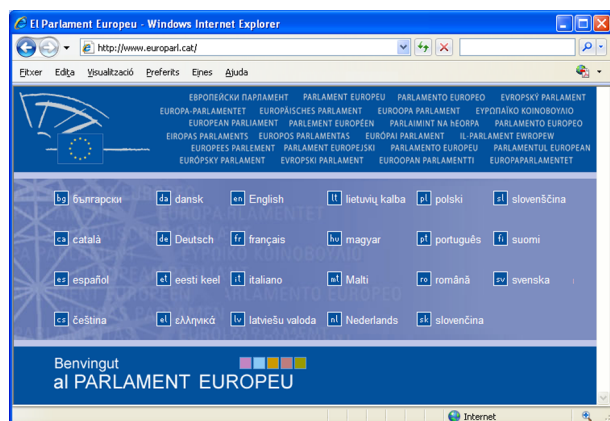
## 1. La gestió del multilingüisme a Internet

És interessant veure com els diversos espais web negocien l'ús de més d'una llengua. Ho poden fer oferint diferents espais monolingües. I quan són espais interactius i compartits per usuaris de diferents llengües, ho poden fer establint criteris de convivència lingüística més o menys neutrals. En aquest apartat veurem algunes de les diverses solucions que hi ha.

### 1.1. Tria de la versió lingüística als portals web

Quan visitem per primer cop una pàgina web que descobrim (o sospitem) que disposa de diferents versions lingüístiques, com gestiona el dissenyador del portal la manera de poder triar la llengua?

En navegar per la xarxa, sovint ens trobem amb un portal web que no ens deixa continuar si no triem, primer, una opció lingüística. El portal del Parlament Europeu n'és un bon exemple.



#### Web del Parlament Europeu

Encara que ho sembli, el lloc web reproduït en la imatge no és l'oficial. Si en voleu saber més, d'aquesta versió que inclou el català, seguiu-ne el vincle.

La segona opció és que hi hagi una versió lingüística que apareix per defecte (generalment en funció de dues dades contingudes en el nostre ordinador: l'adreça IP i l'opció de preferència lingüística que hàgim introduït en el navegador) i una o més versions a escollir mitjançant un botó a la primera pàgina.

En aquest context, és pertinent la informació que ha posat a disposició dels internautes catalanoparlants la pàgina *Navega en català*, que explica així la importància del tema:

Independentment de la llengua en què tingues els menús del teu navegador, és també molt important la llengua que defineixes com a preferida per a aquesta aplicació. Tot i que aparentment no ho sembli, la llengua del navegador que utilitzes per a moure't per Internet dóna a conèixer la teua preferència lingüística.

Per què es té en compte la llengua que es predetermina en el navegador?

- En un web multilingüe et sortirà per defecte la pàgina en català quan el sistema reconeix el català com la llengua del teu navegador.
- Si descarregues un programa multilingüe, aquest es baixa en la llengua que tens configurada en el navegador. Per tant, disposaràs més fàcilment de programari en català si el teu navegador està correctament configurat.
- Els administradors de webs, a més, fan servir la llengua que detecten com a predefinida en els navegadors com una estimació dels visitants/parlants de la llengua. Si per exemple les visites amb navegadors amb el català com a llengua preferida no suposa un percentatge elevat, poden decidir deixar d'oferir planes en català. Així, que les empreses multinacionals tinguin webs específiques en català pel seu mercat catalanoparlant també depèn del fet que el català siga la teua llengua predefinida.



Font: [http://www.navegaencatala.cat/media/comu/titol\\_navega.jpg](http://www.navegaencatala.cat/media/comu/titol_navega.jpg)

A continuació veurem com una gran empresa com Google permet personalitzar lingüísticament el seu portal, mitjançant les galetes corresponents. Reproduïm aquesta explicació escrita pel Google Team en resposta a una queixa perquè l'opció per defecte que ens surt (pel fet de tenir l'adreça IP de l'ordinador a l'Estat espanyol) és en castellà:

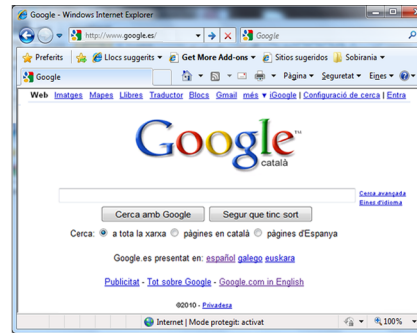
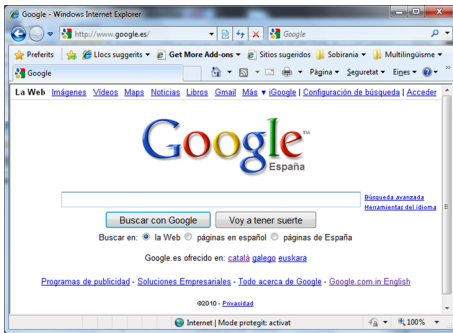
If you'd like to use *www.google.es* in Catalan, click the 'Catala' link. As long as you have cookies enabled in your browser, you'll only have to do this once; after that, your browser will remember that you prefer Catalan, and show you the page in this language. Because this is a local international site, rather than *www.google.es*, the logo shows 'Espanya' beneath the regular Google logo. This doesn't reflect the language the page is appearing in, only its location.

If you prefer to use *www.google.es* in Catalan, click the 'Google.com in English' link. This will redirect you to Google.com. Again, as long as you have cookies enabled, you won't need to do this more than once. To set this page to display in Catalan, click the 'Preferences' link<sup>1</sup>. Select Catalan from the drop-down menu for 'Interface Language'. Be sure to click 'Save Preferences' before leaving the page.

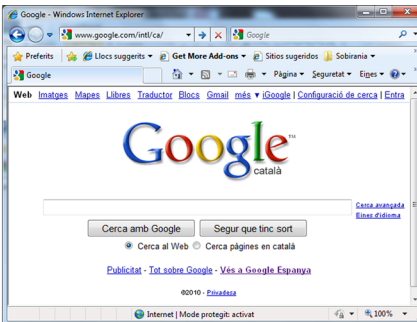
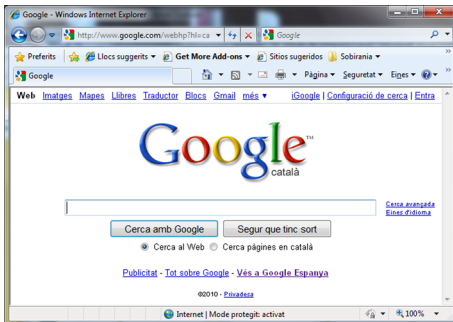
If you prefer not to use cookies at all, you may want to set a bookmark for *www.google.com/intl/ca/* to access *www.google.es* in Catalan.

<sup>(1)</sup>De fet, actualment cal clicar a *Language tools*. L'elecció de llengua no es fa a partir d'un menú desplegable (*drop-down menu*), sinó que es tria en una llista fixa.

Podem reproduir les pantalles corresponents a les quatre situacions: la que surt, per defecte, en castellà; la que surt un cop triada l'opció *català*; la que surt passant per *Google.com in English* i, després de triar *Language tools*, optant per *Catalan*, i finalment la pàgina que Google facilita als interessats que no volen activar les galetes (*www.google.com/intl/ca/*). Vegem-ne els passos:



Pàgina web per defecte a l'Estat espanyol (*Google España*) i pàgina web resultant (*Google Espanya*)



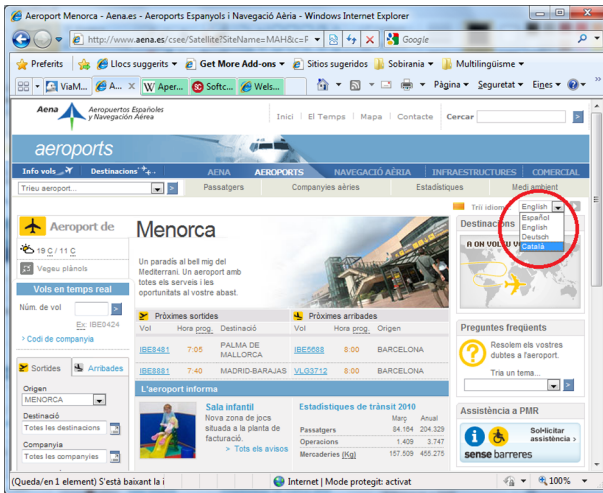
Pàgina web resultant de passar primer per les opcions lingüístiques a la versió en anglès (*Google català*) i opció de qui no vol activar les galetes (*www.google.com/intl/ca/*)

Veiem, doncs, que les solucions "per defecte" solen estalviar feina als usuaris de les llengües estatals.

Hi ha casos en què no hi ha galeta i, per tant, no es conserva la tria de llengua. Així, cada vegada que l'usuari retorna a la pàgina web d'*El Periódico de Catalunya*, encara que en visites anteriors hagi optat per la versió en català, torna a aparèixer la versió en castellà.

Generalment, l'opció lingüística apareix a la part de dalt, a la dreta, del lloc web. Entre molts exemples podem citar el cas d'alguns ministeris espanyols, que en lloc d'identificar les llengües oficials pel nom, ho fan amb un acollidor *Benvingut...* La raó rau en la confusió que causa que una mateixa llengua tingui dues denominacions (tot i que es tracti de petites variacions morfosintàctiques o lèxiques). En comptes de posar *català* i *valencià*, o *català-valencià*, la solució de la benvinguda resulta encertada. Ara bé, generalment, en els llocs web ministerials només les pàgines de primer i segon nivell són disponibles en llengües altres que el castellà.

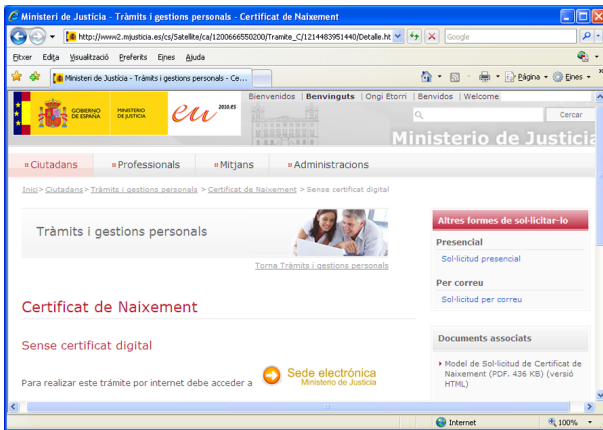
Són casos diferents organismes i empreses públiques com ara Aena (gestió d'aeroports) o Renfe (gestió de ferrocarrils). Aena ofereix al portal l'opció entre castellà i anglès; les altres llengües oficials a l'Estat espanyol només apareixen a les pàgines dels respectius aeroports. Es tracta, doncs, d'un model desequilibrat. A més, l'opció lingüística catalana no és en el lloc esperable, a la capçalera, sinó més avall.



Renfe segueix un criteri semblant. El seu lema, llegim al portal, és "Acortamos distancias. Acercamos personas". L'opció lingüística és a l'extrem oposat del que s'ha anat imposant com a norma. És a baix de tot, a l'esquerra. Ofereix pàgines en castellà (per defecte i sense que l'usuari pugui modificar aquest fet amb vista a visites posteriors), anglès, català, gallec i èuscar. Però així que l'usuari clica, dins la pàgina en català (o en anglès), al damunt d'un dels serveis oferts (compra de bitllets, horaris i preus de trens...), es torna a trobar amb una pàgina en castellà. Notem, a més, que el visitant anglòfon, per exemple, ha d'entendre l'expressió *seleccione su idioma* per tal de poder accedir a una pàgina en anglès...

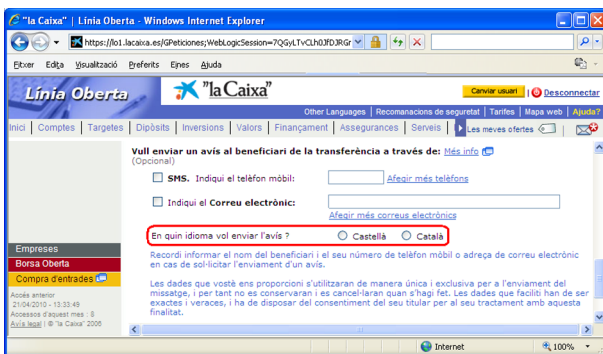
El cas de les certificacions del Registre Civil a partir del lloc web del Ministeri de Justícia és diferent per una altra raó. Malgrat que s'identifiqui com a *model de sol·licitud de certificat de naixement*, el formulari de petició del certificat és monolingüe, en castellà, i tampoc no admet que el sol·licitant indiqui en quina llengua vol que el certificat (de naixement, de matrimoni o de defunció) sigui expedit. El mateix s'esdevé per a altres certificats del Registre Civil. En aquest cas, el sistema digital de sol·licitud no respecta l'opció lingüística que els interessats poden exercir, segons la llei, en els registres civils situats en territoris on hi ha dues llengües oficials.





Font:  
[http://www2.mjusticia.es/cs/Satellite/ca/1200666550200/Tramite\\_C/1214483951440/Detalle.html?param1=4](http://www2.mjusticia.es/cs/Satellite/ca/1200666550200/Tramite_C/1214483951440/Detalle.html?param1=4)

En canvi, tot i que ja hem vist que el món socioeconòmic no té generalitzat l'ús del català en els seus llocs web, sí que dóna bons exemples de la possibilitat de triar la llengua per a serveis determinats (més enllà de la llengua de la pàgina consultada). Així, per exemple, la Caixa d'Estalvis i Pensions de Barcelona (La Caixa) permet, en el moment que un client ordena una transferència, triar la llengua en què l'efectuació de la transferència es comunicarà al destinatari.



En canvi, jo mateix mai no he sabut trobar la manera de demanar per mitjà del lloc web de l'Agència Tributària que se m'expedeixi en català (o valencià!) l'esborrany de la meua declaració anual de l'IRPF. La tria de la pàgina en català mena, indefectiblement, a una petició en què el formulari és en castellà i no hi ha possibilitat de fer constar la llengua en què el ciutadà vol rebre la documentació demanada.

Jaime, Rovira i Saval van demanar a algunes empreses que apliquen el multilingüisme a Internet com a element de qualitat que expliquessin la seva experiència i els resultats que n'han obtingut. Les tres autores, a l'hora de fer una presentació il·lustrativa de les diverses maneres de gestionar el multilingüisme envers els visitants als seus llocs web, en van triar quatre:

- **Grup Alimentari Guissona**, empresa catalana amb activitat en l'àmbit estatal, amb web en català i opcions de canvi d'idioma a l'espanyol i a l'anglès.

#### Referència bibliogràfica

Laura Jaime; Marta Rovira; Susanna Saval (2003). "Empreses i Internet: el multilingüisme a la xarxa com a element de qualitat". *Llengua i Ús* (núm. 26, pàg. 16-21).

- **Freixenet**, empresa catalana amb un mercat internacional importantíssim, amb web en català i opcions de canvi d'idioma a l'espanyol i a l'anglès.
- **Mercadona**, empresa valenciana amb mercat estatal, amb un web que permet la personalització automàtica de l'opció lingüística triada a la pàgina inicial.
- **Transportes Azkar**, empresa basca amb mercat estatal i europeu que ofereix al portal del seu web un ventall molt complet d'opcions lingüístiques.

## 1.2. Establiment de criteris d'ús lingüístic en espais multilingües

S'ha anat estenent l'ús del terme *etiqueta* (*netiquette* en anglès i *nétiquette* en francès). La *Wikipedia* en espanyol ens n'ofereix una definició i explica l'origen del mot:

*Netiquette* (o *netiqueta* en su versión castellanizada) es una palabra derivada del francés *étiquette* ('buena educación') y del inglés *net* ('red') o *network*. Es el conjunto de reglas que regulan el comportamiento de un usuario en un grupo de noticias (*newsgroup* en inglés), una lista de correo, un foro de discusiones o al usar el correo electrónico. Por extensión se utiliza también para referirse al conjunto de normas de comportamiento general en Internet. La *netiqueta* no es más que una adaptación de las reglas de etiqueta del mundo real a las tecnologías y el ambiente virtual. Aunque normalmente los lineamientos de etiqueta han evolucionado hasta llegar a formar incluso parte de las reglas de ciertos sistemas, es bastante común que las reglas de etiqueta se basen en un sistema de "honor", es decir que el infractor no recibe siquiera una reprimenda.

De la misma manera que existe un protocolo para los encuentros físicos con personas, la así llamada *netiquette* describe un protocolo que se debe utilizar al hacer "contacto" electrónico.

Hi ha estudis generals sobre l'etiqueta que inclouen l'ús del llenguatge, com per exemple el de Torres i Vilatarsana. Però el que farem és mirar alguns casos concrets d'espais web en què l'etiqueta parla de llengua, i en què l'aplicació de les normes és perjudicial per a la llengua catalana: *Yahoo! Respuestas* i l'espai a Facebook *Catalunya és Espanya. Cataluña es España. Catalonia is Spain*.

### 1.2.1. Yahoo! Respuestas

*Yahoo Questions!* ofereix espais per a preguntes en diferents llengües o, més ben dit, països: Alemanya, Argentina, Austràlia, Brasil, Canadà, Xina, Corea del Sud, EUA (amb una versió en anglès i una altra en espanyol), Espanya, Filipines, França, Hong Kong, Índia, Indonèsia, Itàlia, Japó, Malàisia, Mèxic, Nova Zelanda, Quebec, Regne Unit, Singapur, Tailàndia, Taiwan i Vietnam.

**YAHOO!** RESPUESTAS  
EN ESPAÑOL

#### Pàgina web

Qui vulgui obtenir més informació sobre l'**etiqueta** (encara que no parli de llengua!) pot consultar el document "La netiquette".

#### Referència bibliogràfica

M. Torres i Vilatarsana (2004). "Com es regula la comunicació interpersonal a Internet? La 'netiqueta'". *Llengua i Ús* (núm. 29, pàg. 37-42).

Si llegim les normes en aquesta pàgina, veiem que no hi ha cap referència a situacions multilingües. Però arran d'uns incidents que van consistir a eliminar intervencions fetes en llengua catalana, resulta que hi ha problemes per a l'ús del català; fins i tot, censura. Així, trobem que un usuari fa la pregunta següent: *¿Por qué censuran el catalán en yahoo preguntas?*

Entre les normes de *Yahoo! Respuestas* disponibles a la pàgina corresponent a l'Argentina, sembla que sí que n'hi ha una que fa referència a l'ús de les llengües, que s'aplica dins de l'Estat espanyol per a censurar (és a dir, eliminar) els textos que fan ús del català (o, ho hem de suposar, del gallec o el basc). És la norma número 19:

En Yahoo! Respuestas, no está permitido

[...]

19.- Publicar en un idioma incorrecto o publicar contenido incomprendible. El contenido escrito en español constituye una infracción de las presentes normas si aparece en cualquiera de los sitios de Yahoo! Respuestas en inglés. Por otro lado, el contenido escrito en inglés constituye una infracción de las presentes normas si aparece en cualquiera de los sitios de Yahoo! Respuestas en español.

És a dir, es prohibeix "publicar en un idioma incorrecto". La qüestió és quins són, i qui defineix, els "idiomas incorrectos". L'exemple que es dóna en la norma només fa referència al castellà, llengua que no es pot emprar en les seccions en anglès, i a l'anglès, llengua que no es pot emprar en les seccions en castellà. És especialment pertinent en el cas d'una versió de *Yahoo! Preguntas* en castellà, destinada concretament a la població hispana dels Estats Units. Però, de fet, s'ha aplicat a les intervencions en llengua catalana, per a suprimir-les..., i *Yahoo! Answers* no ha creat cap espai en català.

### **1.2.2. Catalunya és Espanya. Cataluña es España. Catalonia is Spain**

El segon espai afirma que no dóna suport ni comparteix la ideologia de cap partit polític nacional, regional o de cap altra mena. Malgrat el títol del grup de Facebook, també s'afirma que no dóna suport a cap tipus de nacionalisme radical, ni està a favor de l'ultranacionalisme, antigues dictadures o idees independentistes de nacionalismes regionals. És un grup de debat, diuen, per a les persones que tenen una filosofia moderada, són partidaris de l'enteniment i advoquen pel diàleg, la tolerància i el respecte mutu. El seu logo és prou eloqüent:



[http://profile.ak.fbcdn.net/object3/1890/47/n37383598277\\_244.jpg](http://profile.ak.fbcdn.net/object3/1890/47/n37383598277_244.jpg)

Mirem quines normes tenen per a regular l'ús de les llengües d'un espai que, en definitiva, té el títol en tres llengües:

NORMAS DE USO:

[...]

4. Nadie va a ser excluido por hablar cualquiera de los 2 idiomas oficiales en Catalunya. Muchos miembros de este grupo, entendemos los 2 idiomas, tal y como pasa en la calle (en referencia a Catalunya, obviamente), y por tanto, cada uno es libre de utilizar el idioma que quiera. También es cierto que en este grupo hay miembros del resto de España, quienes no entienden el catalán. Por tanto, se ruega, solamente SE RUEGA que aquella persona que quiera ser entendida por la totalidad de miembros de este grupo, debería tener en cuenta que todos conocemos el idioma español. A partir de aquí, si alguien quiere utilizar el idioma catalán, es muy libre de hacerlo, y lo entenderemos muchos, pero no todos. Por tanto, cada uno de nosotros tenemos el derecho, y la responsabilidad, de querer utilizar el idioma que queramos, desde el respeto, y con el ánimo de ser entendidos por los demás.

El text, sota l'aparença de tolerància i respecte, sembla pressionar els participants a fer servir només el castellà.

### 1.3. Orientacions per al disseny de llocs web multilingües

Torres (2006) parla de la traducció automàtica com a eina per als webs. Entre altres coses, presenta un servei interessant de la Generalitat de Catalunya, el lloc web *El multilingüisme als webs d'empresa*. El lloc web intenta convèncer l'empresari de les virtuts d'adaptar el web a les preferències del client, explica com s'adapta un web al públic objectiu de l'empresa, justifica que un web multilingüe és una bona inversió i afirma que és fàcil de fer. Ofereix models de pàgina de benvinguda, amb exemples reals. Al final de l'article, Torres recomana la introducció d'eines amb un sol clic que facilitin que el visitant pugui veure una pàgina web redactada en català en altres llengües, com ara l'anglès o el castellà.

## 2. La gestió del multilingüisme en els programes informàtics

A Internet hi ha eines que permeten la localització o la personalització de la interfície de diferents programes, de manera que l'usuari pugui tenir, davant seu, una versió en la llengua que vulgui i triï. Abans hem parlat del cas de Google, però n'hi ha moltes altres. La seva evolució ha estat i és diversa, tant en els ritmes d'introducció de diferents opcions lingüístiques com en les llengües mateixes que s'hi ofereixen.

### 2.1. Microsoft

Microsoft fa publicitat en el seu lloc web de la disponibilitat de diverses versions lingüístiques, en forma de *language packs*, dels productes englobats en el paquet Microsoft Office (Word, Excel, Outlook...) i també del navegador Explorer. Hi ha una versió en català:

"Si vols posar en català el teu ordinador et proposem que descarreguis els següents programes gratuïtament."

Cal dir que els *language packs* no solen arribar a oferir tot el contingut de cada programa en la llengua triada.

Com hem explicat, moltes llengües tenen una problemàtica molt superior que el català, pel que fa a l'escriptura, que pot fer servir sistemes fonètics o no, i fins i tot sistemes ideogràfics, com ara el coreà, el japonès i el xinès. Microsoft ha desenvolupat el que anomena *Input Method Editors*, és a dir, editors del mètode d'entrada, fent servir un teclat estàndard.

#### Microsoft Global Input Method Editors (IMEs)

Published: March 18, 1999

International communication keeps getting easier thanks to full-featured, high-performance IMEs. An IME is a program that allows computer users to enter complex characters and symbols, such as Japanese characters, using a standard keyboard. Microsoft is now offering two Global IMEs—Global IME 5.02 and Global IME for Office XP.

By launching any language version of Word 2000, Internet Explorer 5.01 or greater, or Outlook Express – then selecting Global IME 5.02 from the taskbar, you can write the language you choose.

And now, you can input Chinese, Japanese, and Korean text into any Microsoft Office XP applications, Web forms, and e-mail messages on any language version of Windows XP, Windows 2000, Windows Millennium, Windows 98, and Windows NT&reg; 4.0 with Service Pack 6 (SP6) or later. And it's easy. After you have installed the Global IME for Office XP, you just start your Office XP program, select the language from the Language bar, and you can type the language you choose, regardless of the language version of Office XP or Windows you are using.

#### Vegeu també

Hem parlat de la problemàtica que representen algunes escriptures en l'apartat "L'ús de diferents grafies a Internet" del mòdul "El multilingüisme a Internet".

## 2.2. Softcatalà

Una iniciativa d'un grup de voluntaris (semblant a WICCAC) és Softcatalà, que s'ha dedicat a adaptar programes informàtics populars a la llengua catalana. El moviment de **codi obert** (*open source* en anglès) ha simplificat enormement l'avenç del català (i d'altres llengües), mentre que les empreses més potents (i, concretament, Microsoft) negaven accés als programes i van blocar, durant anys, la disponibilitat de versions dels seus paquets informàtics en llengües altres que les comercialment més rendibles. Una gran part dels avenços informàtics dels darrers anys relacionats amb els idiomes s'ha fet amb codi obert, és a dir, sense que hi hagi prevalgut un objectiu lucratiu. No és gens estrany, doncs, que a vegades Softcatalà s'hagi de manifestar a favor de la generalització dels programes de codi obert i en contra de successius contractes milionaris de la Generalitat amb empreses com Microsoft o Yahoo!.

El ventall de programari "localitzat" (és a dir, traduït i adaptat per al mercat catalanoparlant) pels voluntaris de Softcatalà és espectacular, i dóna fe de la capacitat de producció que ofereix el treball en xarxa entre voluntaris. A més, el seu lloc web ofereix accés a un traductor automàtic, el que va desenvolupar de manera pionera l'interNOSTRUM.

## 2.3. Ubuntu

Un altre exemple d'aquestes iniciatives basades en el codi obert és l'Ubuntu, un sistema operatiu creat per un equip de programadors que treballen arreu del món i que conté un navegador, un paquet de programari d'oficina (equivalent al Microsoft Windows o a l'OpenOffice) i altres eines de comunicació.

En aquest marc trobem Rosetta (que no s'ha de confondre amb una altra iniciativa, Rosetta Stone), que és una plataforma per a la traducció d'aplicacions de codi obert creada per l'empresa Canonical. Dóna suport a qualsevol persona que vulgui traduir a la seva llengua una aplicació de codi obert. Com a tants llocs, aquí també hi ha catalans, fent versions catalanes (sovint en coordinació amb Softcatalà). El cap de l'equip, l'abril de 2010, era David Planella.

### La pedra de Rosetta

La pedra de Rosetta la van desenterrar l'any 1799 a Egipte soldats de l'exèrcit napoleònic. Com que era un text escrit de tres maneres diferents –una de les quals era el grec antic–, la pedra va permetre la descodificació, tres anys més tard, dels jeroglífics de l'època faraònica i la comprensió, per tant, dels textos.

### Codi obert

El terme *open source* o *codi obert* es refereix als productes de programari que estan disponibles gratuïtament i que són desenvolupats per comunitats concretes en línia.

### Vegeu també

En l'apartat "Els traductors automàtics" d'aquest mòdul parlarem de manera detallada d'aquest i d'altres traductors.

## 2.4. Orientacions per a l'elaboració de programari bilingüe

No podem acabar aquest apartat sense fer una breu referència al fet que Bwrdd yr Iaith Gymraeg (el Consell de la Llengua Gal·lesa) ha elaborat un conjunt interessant i útil d'orientacions pràctiques per a elaborar programes informàtics bilingües: *Bilingual Software Standards & Guidelines*.

## 2.5. Altres aplicacions

La llista seria inacabable, però donarem un exemple pràctic de com la diversitat lingüística pot no ser un entrebanc per a la cerca. Es tracta de la possibilitat que una base de dades (per exemple, però no necessàriament) accessible per mitjà d'Internet permeti "la compilació de textos mitjançant la cerca per paraules clau, que pot arribar a implicar l'accés a un servei web de traducció de les paraules clau a diferents llengües, per a aconseguir dades independentment de la llengua de consulta", en paraules de Bel i altres. Això no solament és una possibilitat: ha estat desenvolupat per la Biblioteca de la Universitat de Bozen, tal com expliquen Bernardi, Gobbetti i Siciliano. Es tracta d'un motor de cerca per a recuperar informació translingüística que fins i tot resol casos d'ambigüitat per polisèmia.

### Vegeu també

Tractarem la feina d'aquest Consell en el capítol "Nivell meso: País de Gal·les (Bwrdd yr Iaith Gymraeg)"

### Referències bibliogràfiques

Núria Bel; Santiago Bel; Sergio Espeja; Montserrat Marimon; Marta Villegas (2008). "El projecte CLARIN: una infraestructura de recerca científica per a les humanitats i les ciències socials". *Digitum* (núm. 10, pàg. 6). (També disponible en castellà: "El proyecto CLARIN: una infraestructura de investigación científica para las humanidades y las ciencias sociales", i en anglès: "The CLARIN project: a scientific research infrastructure for the humanities and social sciences".)

Rafaella Bernardi; Daniele Gobbetti; Luigi Siciliano (2009). *Multilingual Access to Library Catalogues: Word Sense Disambiguation via Classification Systems*. Ponència presentada a les International Conferences on Digital Libraries and the Semantic Web 2009 (ICSD 2009), 8 a 11 de setembre, University of Trento.

### 3. Eines per a la gestió de la diversitat lingüística en els textos

No cal dir que no totes les TIC (tecnologies de la informació i la comunicació) comencen i acaben a Internet. Hi ha moltes eines (algunes disponibles per mitjà d'Internet) que faciliten la gestió de múltiples llengües en altres àmbits i entorns.

Sense anar més lluny de les eines tecnològiques que permeten o faciliten les tasques de traducció, i de manera il·lustrativa, la Universitat Oberta de Catalunya ofereix formació de postgrau en matèria de traducció, entre la qual hi ha cursos d'especialització com ara els següents:

- Traducció assistida i gestió de projectes de traducció.
- Traducció i tecnologies: processos, eines i recursos.
- Traducció assistida per ordinador amb Trados.
- Localització de programari.
- Gestió de projectes de traducció.

Al llarg d'aquesta secció, i sense entrar en un gran detall tècnic, estudiarem no solament algunes d'aquestes eines tecnològiques de gestió de la diversitat lingüística en els textos, sinó també alguns productes fruit de l'ús d'aquestes eines. Un exemple ens l'ofereix un convertidor per a traduir automàticament els menús i les cartes de qualsevol bar, restaurant o casa a l'euscar. Es tracta d'una iniciativa del Patronat Municipal de l'Eusquera de Donostia (Sant Sebastià), portada a terme per l'empresa Dijitalidadea. S'han utilitzat diverses fonts per a crear-ne el corpus: el diccionari de l'apartat "Euskara platerean" i la col·lecció de menús que hi ha en el portal [www.merkatari.net](http://www.merkatari.net), els fons dels llocs web [gastronomiavasca.net](http://gastronomiavasca.net) i [sukaldea.com](http://sukaldea.com) de l'Escola d'Hostaleria de Leioa. S'hi han recollit 1.057 termes (més derivats).

La llengua catalana també té un servei equivalent, *Plats a la carta*, una aplicació que facilita al sector de la restauració l'elaboració de cartes i menús en català, i en permet també la traducció al castellà, el francès, l'anglès, l'italià i l'alemany. Es tracta d'un recurs multilingüe i gratuït, desenvolupat i gestionat per la Secretaria de Política Lingüística de la Generalitat de Catalunya en col·laboració amb el Termcat.



#### Pàgina web

Podeu trobar més informació sobre aquest traductor de menús a Berripapera.



### 3.1. Eines integrals de gestió del multilingüisme

En el mercat hi ha algunes eines que permeten a una organització gestionar la seva activitat en diferents llengües. El **Multitrans** n'és un exemple. Es tracta d'un paquet utilitzat per les Nacions Unides, pels governs del Canadà, França, Suïssa i Alemanya, i per molts clients comercials i industrials. El Multitrans, entre altres productes, ofereix els següents:

- Un servidor de gestió de terminologia que proveeix de terminologia el personal de les empreses, els traductors, els terminòlegs i els autors de textos.
- Un servidor *TextBase TM*, que permet l'accés a l'instant a una memòria de traducció.
- Eines de traducció que permeten treballar a equips de traductors amb diferents eines de memòria de traducció i amb documents de formats variats (com ara Word, WordPerfect, PowerPoint, HTML, XML, XLIFF i diversos formats DTP).
- Una eina de gestió de projectes, que facilita el seguiment de la qualitat i els costos de la traducció.

Un altre exemple és l'empresa **Lucy Software Ibérica**. L'empresa mare, Lucy Software and Services GmbH, té la seu principal a prop de Heidelberg, Alemanya. L'empresa ofereix tres tipus de serveis a empreses i administracions per a la gestió de les traduccions:

- **SAP-AT (Auditoria de traducció)**
  - Anàlisi del sistema per a obtenir una apreciació correcta dels volums de traducció, els pressupostos i els temps d'execució.
  - Avaluació del potencial per a l'optimització del sistema (reducció de temps i costos).
- **SAP-CE (Assessorament estratègic)**
  - Assessorament sobre estratègies de traducció i pràctiques recomanades.
  - Formació sobre les eines de traducció de SAP.
  - Resolució de problemes (per exemple, en l'àrea dels transports).
  - Desenvolupament d'eines específiques per a cada client que permeten optimitzar el procés de traducció.
  - Suport per a les actualitzacions.
- **SAP-SI (Servei integral)**
  - Gestió integral de tot el procés de traducció.
  - Anàlisi del sistema i configuració de l'entorn de traducció.
  - Gestió tècnica del projecte i resolució de problemes.
  - Selecció i gestió dels traductors.

- Realització de la traducció amb control de qualitat.

L'any 2009 una empresa pública basca, EJJIE, va encarregar a Lucy Software and Services GmbH l'elaboració d'una estratègia per a fer una versió en èuscar (a partir de la versió que hi havia en espanyol) de tot el sistema SAP de Lucy, amb vista a comercialitzar-lo en el mercat basc.

Abans de dir-se Lucy, Translendum SL va completar un projecte d'un traductor automàtic encarregat per la Xunta de Galícia. En una primera fase l'empresa va desenvolupar el programari per a la traducció automàtica d'espanyol a gallec. En una segona fase, el 2009, s'hi van incorporar tres combinacions noves: de gallec a espanyol, de gallec a anglès i d'anglès a gallec.

Hi ha altres productes que, sense ser pròpiament traductors, permeten treballar entre diferents llengües alhora. Per exemple, l'empresa Sail Labs ofereix el *Media Mining Indexer*, una eina que reconeix i indexa automàticament textos parlats en diverses llengües en temps real, com ara butlletins de notícies i arxius multimèdia. És capaç de tractar textos orals en àrab, anglès (tant del Regne Unit com dels Estats Units), francès, alemany, grec, noruec, rus i espanyol.

Un servei sofisticat l'ofereix el traductor automàtic Dixio (treballa amb una trentena de llengües, entre les quals hi ha el català).



L'empresa bilbaína Semantix Group SL ofereix una presentació audiovisual de les prestacions de l'eina Dixio que, a més de traduccions en línia entre català, anglès i espanyol, dona definicions i entrades de *Wikipedia* als termes triats.

### 3.2. Els traductors automàtics

Un dels camps en què més han evolucionat les eines que permeten afrontar els reptes que planteja la necessitat de comunicar-se en diferents llengües és el de la traducció automàtica.

Els traductors automàtics es comencen a desenvolupar a la dècada dels anys trenta del segle XX. Tanmateix, podem trobar una primera aproximació als principis teòrics per a l'automatització de la traducció cap a la meitat del segle XVII. En aquesta època, a la ciutat alemanya d'Espira, un monjo, Johannes Becher, va publicar un llibret en què descrivia com havia inventat un metallenguatge matemàtic l'objecte del qual era facilitar el significat de frases escrites en qualsevol llengua, tal com explica Freigang.

Freigang reproduïx part del memoràndum adreçat a la Rockefeller Foundation (1949) per Warren Weaver, en què fa una analogia famosa entre la traducció i la descodificació:

"I have a text in front of me which is written in Russian but I am going to pretend that it is really written in English and that it has been coded in some strange symbols. All I need to do is strip off the code in order to retrieve the information contained in the text."

La primera demostració d'un sistema de traducció automàtica que es pot considerar reeixida es va fer el 7 de gener de 1954 a la Universitat de Georgetown (Washington, D. C.). Es va muntar en un ordinador IBM un sistema que permetia traduir del rus a l'anglès, gràcies a un diccionari bilingüe amb prop de 250 termes. El sistema va ser capaç de traduir amb èxit un corpus experimental de 60 frases senzilles en rus, basant-se en una substitució paraula per paraula, complementada per unes regles addicionals per a la generació de l'ordenació correcta dels mots en les frases resultants en anglès.

El primer programa utilitzat a gran escala per clients molt importants –com ara la Comissió Europea, les Forces Aèries dels Estats Units o l'empresa Xerox– va ser el SYSTRAN, que encara funciona, per exemple en el servei Babelfish de traducció automàtica en línia, el qual veurem tot seguit. El SYSTRAN s'ha fet servir en els darrers anys per a treballar amb la memòria de traducció del cabal comunitari (*acquis communautaire*), és a dir, el conjunt de legislació europea vigent, incloent-hi tots els tractats, reglaments i directives adoptades per la Unió Europea, més les decisions del Tribunal de Justícia Europeu, en les 22 llengües oficials i de treball de les institucions de la Unió Europea.

A més a més, el SYSTRAN ofereix un servei de traducció automàtica alemany, anglès, castellà, coreà, francès, italià, japonès, neerlandès, portuguès, rus, suec i xinès.

## METAL

Patrick Little explica l'aplicació del METAL (Machine Evaluation and Translation of Natural Languages), un potent sistema desenvolupat per Siemens en una empresa concreta, en el treball següent:

**Patrick Little** (1990). "METAL - machine translation in practice". A: Catriona Picken (editora). *Translating and the Computer 11. Preparing for the Next Decade. Proceedings of a conference jointly sponsored by Aslib, The Association for Information Management, The Aslib Technical Translation Group, The Institute of Translation and Interpreting, 16-17 November 1989*. Londres: ASLIB.

## Referència bibliogràfica

**K. H. Freigang** (2001, octubre). "Automation of Translation: Past, Presence, and Future". *Revista Tradumàtica, Traducció i Tecnologies de la informació i la Comunicació* (núm. 0).

Una empresa catalana va desenvolupar un programa per a emmagatzemar i gestionar la memòria de traducció que va acumulant un traductor professional en solitari o en equip. El nom del producte, Déjà Vu, ja és prou gràfic. Hi ha productes més recents que encara fan més amigable (*user-friendly*) la interfície i altres elements, però això no treu una evidència: la memòria és buida fins que el professional no l'omple.

Té un interès especial per als usuaris el seguit d'eines automàtiques que permeten la traducció en línia de textos (relativament curts) i pàgines web. Un dels primers va ser el ja esmentat **Babelfish**, ofert des del cercador Altavista i, des de fa un temps, també per mitjà de Yahoo!.



Font: [http://espanol.babelfish.yahoo.com/translate\\_txt](http://espanol.babelfish.yahoo.com/translate_txt)

El servei Babelfish ofereix traduccions entre dotze de les llengües principals (alemany, anglès, castellà, coreà, francès, grec, italià, japonès, neerlandès, portuguès, rus i xinès), però no el català.

Per tal de trobar serveis d'aquest tipus que incloquin el català ens hem de desplaçar fins a Alacant, a la Universitat en què s'ha desenvolupat l'**interNOSTRUM**. Es tracta d'una eina creada per a fer traduccions entre el català i el castellà en línia, amb l'opció de donar prioritat a la varietat valenciana de la llengua catalana. L'interNOSTRUM el va dissenyar el Departament

#### Pàgina web

El portal lingüístic de la Corporació Catalana de Mitjans Audiovisuals (CCMA) ofereix una relació actualitzada d'una dotzena d'eines de traducció automàtica per a la llengua catalana.

de Llenguatges i Sistemes Informàtics de la Universitat (responsables, també, d'un altre gran projecte que treballa amb recursos lingüístics, la biblioteca virtual Cervantes i la seva correspondència pel que fa als textos i la literatura en llengua catalana, la biblioteca virtual Lull).

La pàgina web de la Generalitat Valenciana incorpora l'eina interNOSTRUM en el lloc web i una versió reduïda d'aquest producte és disponible a la pàgina web de Softcatalà.

Lucy Software Ibérica –una empresa que ja hem presentat– va adquirir un altre sistema de traducció automàtica, el **Comprendium**, que després de ser venut a **Translendum**, havia passat a mans de Sail Labs. Des del novembre de 2003, el Comprendium ofería un servei de traducció automàtica en línia en més de vint direccions de traducció, incloent-hi el català. El gener de 2004 hi incorporà un prototipus de traductor automàtic entre el català i el francès, el primer d'aquestes característiques disponible en línia.

Lucy Software Ibérica ven el programa de traducció automàtica i altres serveis per a la gestió multilingüe a indústries i administracions. A més, ofereix un servei en línia anomenat **Lucy LT KWIK Translator**. El traductor en línia cobreix algunes de les combinacions de nou llengües, incloent-hi el català.



A final del setembre de 2006 va entrar en funcionament el **servei de traducció automàtica de la Generalitat de Catalunya**, obert a la ciutadania per als parells de llengües català-castellà, català-anglès i català-francès, en ambdós sentits de traducció. Murtra explica els objectius del projecte, les característiques tècniques, funcionals i lingüístiques del servei de traducció, i l'impacte que va tenir durant els primers mesos de funcionament.

El traductor fa servir la tecnologia de Translendum, empresa de la qual acabem de parlar. Translendum SL, amb seu a Barcelona i dedicada al desenvolupament del sistema de traducció automàtica Comprendium, el 2009 va esdevenir Lucy Software Ibérica, filial participada al cent per cent de Lucy Software and Services GmbH.

El servei es basa en la plataforma de traducció automàtica de codi obert Apertium, l'origen de la qual explicarem tot seguit. Ofereix traduccions entre el català i el castellà, l'anglès i el francès des de l'inici. L'alemany s'hi introdueix el 2007, i el novembre de 2008 s'hi introdueix l'occità (i l'aranès), que es pot traduir entre el català per una banda i el castellà per l'altra.

#### Referència bibliogràfica

Pilar Murtra Bellpuig (2006). "El nou servei de traducció automàtica del web de la Generalitat de Catalunya". *Llengua i Ús* (núm. 37).

En aquest projecte la terminologia que aporta el Termcat és considerable: català ← → castellà, 20.204 i 20.086 registres respectivament; català ← → anglès, 19.940 i 19.114, i català ← → francès, 11.163 i 10.735.

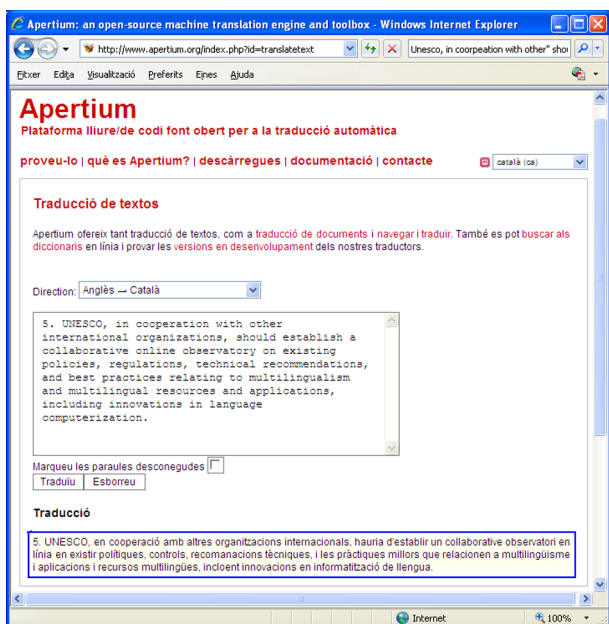
L'èxit d'aquest servei s'evidencia gràcies a les estadístiques d'utilització:

<b>Direcció de traducció</b>	<b>Paraules traduïdes (22/9/2006 – 14/12/2006)</b>	<b>Paraules traduïdes (2008)</b>
català > castellà	25.533.996	789.770.042
castellà > català	12.761.016	152.857.982
català > anglès	17.826.101	354.940.574
anglès > català	18.052.455	156.741.441
català > francès	2.932.238	311.265.407
francès > català	3.042.479	26.874.375
català > alemany	-	4.240.809
alemany > català	-	3.857.770

Fonts: Pilar Murtra Bellpuig (2006), "El nou servei de traducció automàtica del web de la Generalitat de Catalunya", *Llengua i Ús*, núm. 37. Generalitat de Catalunya (2009), *Informe de Política Lingüística 2008*

Pel que fa als primers mesos, es van traduir 80.140.000 paraules. Durant aquest temps hi va haver 782.000 peticions de servei, de les quals 710.000 van ser de traducció de textos curts, segons explica Murtra. Com veiem a la taula, el 2008 es van traduir gairebé 790 milions de paraules.

Ens acabem de referir a l'**Apertium**. Es tracta d'una eina de codi obert de traducció automàtica desenvolupada pel grup Transducens de la Universitat d'Alacant dins del projecte **Opentrad**, amb una subvenció del Ministeri d'Indústria, Turisme i Comerç espanyol i el suport de la Generalitat de Catalunya. Al principi l'Apertium només permetia la traducció entre llengües emparentades com l'espanyol i el català (interNOSTRUM) o l'espanyol i el portuguès (Universia). La versió 2.0 del motor de traducció permetia tractar, mitjançant una transferència estructural més potent, parells de llengües no tan relacionades, com ara l'anglès i el català. El 2007 se'n va publicar la versió 3.0, que permet l'ús de l'Unicode i així obre la porta a moltes altres llengües. Les dades lingüístiques desenvolupades per al projecte (amb la col·laboració de la Universitat de Vigo, la Universitat Politècnica de Catalunya i la Universitat Pompeu Fabra) li permeten traduir, en versió estable, el basc, el bretó, el català, l'anglès, l'esperanto, el francès, el gallec, l'islandès, l'occità, el portuguès, el romanès, l'espanyol i el gal·lès.



Dins els projectes pilot de la Universitat Oberta de Catalunya, l'Apertium serveix, a més, per a posar en funcionament aules (virtuals) bilingües o trilingües, la qual cosa fa possible la docència amb grups d'estudiants de llengües diferents. Es tracta de programari lliure.

### Referències bibliogràfiques

Per a més informació sobre les bases tecnològiques d'aquestes aules multilingües es pot consultar aquest document de la UOC.

I també l'article següent:

**Antoni Oliver; Cristina Borrell (2008).** "Com es treballa en una aula virtual multilingüe? Un entorn d'aprenentatge multilingüe amb Moodle i Apertium". *UOC papers, revista sobre la societat del coneixement* (núm. 6).

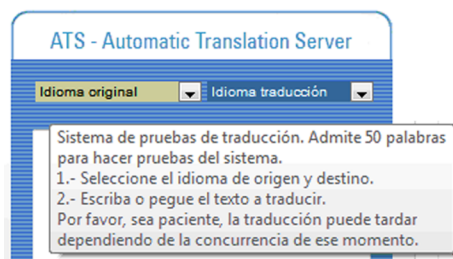
Funciona bé, l'Apertium? Comprovem-ho mitjançant una traducció feta utilitzant aquesta eina d'una descripció de les característiques que té:

Descripció extreta del material "La TA y el software de libre distribución" de la UOC	Traducció automàtica feta amb l'eina oferta en línia
<p>De entre todas las iniciativas de liberalización de la TA hay que destacar el proyecto de creación del motor de traducción del sistema Apertium. El sistema Apertium es un sistema de traducción automática de código abierto para lenguas bastante próximas entre sí. Concretamente, los pares de lenguas que ofrece son el español-catalán, el español-gallego y el español-portugués, entre otros. La primera versión apareció en julio de 2005. Posteriormente se han añadido los pares de lenguas catalán-francés, aranés-catalán e inglés-catalán.</p> <p>El sistema Apertium se basa en el sistema interNOSTRUM para el castellano-catalán, desarrollado por el grupo Transducens de la Universidad de Alicante, y en el traductor Uniersia para el castellano-portugués, desarrollado también por la Universidad de Alicante, ambos están disponibles en la red. También se basa en el castellano-gallego desarrollado en el consorcio OpenTrad, en el cual también se ha trabajado en un motor de traducción automática de código libre para el par castellano-vasco.</p> <p>Además del hecho de tener gratuitamente un traductor automático para los pares de lenguas señalados, de esta iniciativa destacamos, por una parte, la posibilidad de desarrollar a partir de un motor básico motores para otros pares de lenguas o motores mejorados para un par de lenguas ya en funcionamiento, y por otra parte, y no menos importante, la posibilidad de adaptar los recursos del motor de traducción para otras finalidades, algo muy difícil de poder hacer en programas propietarios. Por ejemplo, la lista de todas las formas conjugadas y flexionadas de las palabras contenidas en el léxico del catalán puede utilizarse para elaborar un corrector ortográfico de esta lengua.</p>	<p>D'entre totes les iniciatives de liberalització de la TA cal destacar el projecte de creació del motor de traducció del sistema Apertium. El sistema Apertium és un sistema de traducció automàtica de codi obert per a llengües bastant properes entre si. Concretament, els parells de llengües que ofereix són l'espanyol-català, l'espanyol-gallec i l'espanyol-portuguès, entre uns altres. La primera versió va aparèixer al juliol de 2005. Posteriorment s'han afegit els parells de llengües català-francès, aranès-català i anglès-català.</p> <p>El sistema Apertium es basa en el sistema interNOSTRUM per al castellà-català, desenvolupat pel grup Transducens de la Universitat d'Alacant, i en el traductor Uniersia per al castellà-portuguès, desenvolupat també per la Universitat d'Alacant, ambdós estan disponibles a la xarxa. També es basa en el castellà-gallec desenvolupat en el consorci Opentrad, en el qual també s'ha treballat en un motor de traducció automàtica de codi lliure per al parell castellà-basc.</p> <p>A més del fet de tenir gratuïtament un traductor automàtic per als parells de llengües assenyalats, d'aquesta iniciativa destaquem, d'una banda, la possibilitat de desenvolupar a partir d'un motor bàsic motors per a altres parells de llengües o motors millorats per a un parell de llengües ja en funcionament, i d'altra banda, i no menys important, la possibilitat d'adaptar els recursos del motor de traducció per a altres finalitats, alguna cosa molt difícil de poder fer en programes propietaris. Per exemple, la llista de totes les formes conjugades i flexionades de les paraules contingudes en el lèxic del català pot utilitzar-se per elaborar un corrector ortogràfic d'aquesta llengua.</p>

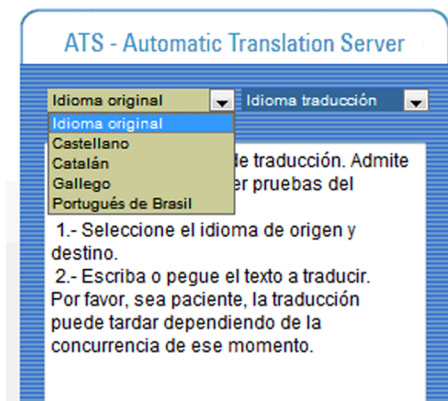
Actualment, l'Apertium ofereix 36 direccions de traducció entre un ventall considerable de llengües, que inclouen les dues variants de noruec, el bretó (cap al francès), l'islandès (cap a l'anglès), i les llengües peninsulars.

La llista d'altres eines és força llarga. L'**Opentrad**, per exemple, és un traductor automàtic de codi obert, basat en les tecnologies Apertium i Matxin (la segona, desenvolupada per a la traducció del castellà al basc). Ofereix les opcions català-castellà, gallec-castellà, basc-castellà i a la inversa.

Alguns productes ofereixen una combinació limitada de llengües, com per exemple el servei gratuït interactiu de traducció de textos d'espanyol a català (entre altres serveis lingüístics) **T6 Estàndard Lingüístic**, o el traductor automàtic castellà, català, gallec, portuguès (amb opció de portuguès del Brasil) **Automatic Translation Server (ATS)** de l'empresa barcelonina Automatic-Trans.







A més, des de 1997, l'**agència EFE** té integrada en el seu sistema de gestió de continguts la plataforma ATS, que permet traduir tot el material informatiu que genera (notícies, reportatges, entrevistes...) de la llengua original en la qual està redactat, al català i al portuguès.

Per altra banda, hem d'esmentar l'**Oficina del Español en la Sociedad de la Información** (OESI), un centre d'informació i documentació especialitzat en activitats de difusió i promoció de la tecnologia lingüística en espanyol, que pertany a l'Àrea Acadèmica de l'**Instituto Cervantes**. L'OESI ofereix un servei de traducció automàtica de textos curts i pàgines web (també desenvolupat per AutomaticTrans, SL) del castellà a l'anglès, el català, el gallec, el francès i el portuguès, i viceversa, (a més d'un prototipus de servei unidireccional, del castellà a l'èuscar).

La Generalitat Valenciana ofereix mitjançant la Conselleria d'Educació el **programa Salt**, un traductor que permet passar els textos del castellà al valencià, i viceversa, amb un mètode interactiu que minimitza els errors i, segons la pàgina web, garanteix una gran qualitat dels textos traduïts. De més a més, el Salt també és un corrector que, a part dels errors estrictament ortogràfics, detecta barbarismes, perífrasis incorrectes, locucions errònies, combinacions incorrectes de pronoms i faltes de concordança. Finalment, el Salt conté centenars de fitxes d'ajuda sobre lèxic, qüestions gramaticals, majúscules, signes de puntuació i tipus de lletra, entre altres. Hi ha dues versions del programa:

- **Salt 3.0**, que funciona de manera independent. S'ha de descarregar d'Internet i instal·lar-lo a l'ordinador de l'usuari.
- **Salt 4.0**, que no funciona de manera autònoma, sinó que s'instal·la sobre el programa Writer, l'editor de textos del paquet de l'OpenOffice, en la barra d'eines (ferramentes) del qual apareixeran les icones amb les funcions i les utilitats del programa. Així, l'usuari disposarà de totes les possibilitats d'un programa d'edició de textos com és el Writer. Té una versió per a Windows i una altra per a Linux.

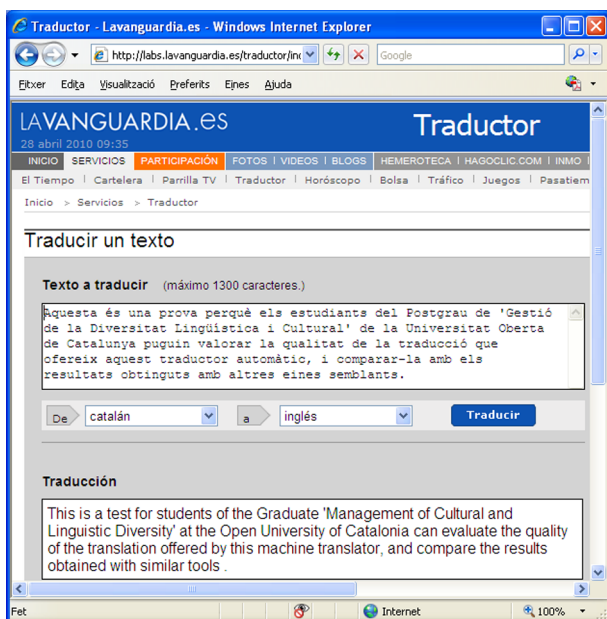
## OpenOffice

L'OpenOffice és el paquet de programari d'oficina de codi obert principal per al processament de textos, fulls de càlcul, presentacions, gràfics, bases de dades, etc. Està disponible en moltes llengües. Es pot descarregar i utilitzar completament de franc.



Finalment, ens referirem al traductor automàtic de **Google**. En les fases inicials va introduir onze llengües: l'alemany, l'anglès, l'àrab, el castellà, el coreà, el francès, l'italià, el japonès, el portuguès, el rus i el xinès. Es tracta d'una iniciativa ambiciosa posada en marxa fa pocs anys, i que des del 25 de setembre de 2008 inclou el català. Permet la intertraducció automàtica d'una trentena de llengües gràcies a un sistema desenvolupat per Franz-Josef Och (actualment cap del Departament de Traducció Automàtica de Google), guanyador d'un premi de traducció automàtica el 2003. El sistema es diu Traducció automàtica estadística. Es basa inicialment en la incorporació en un corpus multilingüe d'un gran nombre de documents de les Nacions Unides, en les sis llengües oficials: l'àrab, el xinès, l'anglès, el francès, el rus i l'espanyol. Aquest fons ha donat un corpus de 20 mil milions de mots. A partir d'aquest fons inicial es pot treballar amb llengües addicionals, si hom en té uns corpora bilingües de més d'un milió de paraules, més un corpus per a cada llengua de mil milions de mots. Uns models estadístics permeten, a continuació, la traducció entre aquestes llengües.

**La Vanguardia** ofereix un servei en línia, basat en la tecnologia de Google, que permet traduir qualsevol text que no depassi els 1.300 caràcters a 35 idiomes diferents, incloent-hi el català. També ofereix la possibilitat de traduir pàgines web si s'hi introdueix la URL corresponent, com ho fan molts altres traductors.



Per al lector interessat a aprofundir en aquests temes, recomanem el treball d'Oliver sobre la traducció automàtica a Internet. La classificació que presenta no es basa en la tecnologia de traducció automàtica que fan servir, sinó en la manera de presentar cada producte a Internet i altres característiques del producte:

- Mode d'accés: pàgina web / servei web
- Utilització: ús en línia / descàrrega
- Preu: comercial/gratuït
- Servei principal / servei de valor afegit

Aquestes són algunes de les eines. També hi ha, a Internet i també per a professionals amb altres formats, un gran nombre de diccionaris bilingües i multilingües (la terminologia és una gran branca de relació entre les TIC i el multilingüisme que aquí només esmentem de passada) que permeten la cerca de paraules concretes. Hi ha també algunes aplicacions que apareixen entre les eines de programes de tractament de textos que permeten obtenir de manera gairebé immediata traduccions d'aquesta mena.

També hi ha l'opció, en una pàgina web qualsevol, de poder incorporar-hi l'enllaç a un traductor automàtic que sense haver de copiar i enganxar, ni obrir pàgines de traductors específics, permet al visitant obtenir una versió de la mateixa pàgina en una altra llengua. Per exemple, el bloc d'*El Puzzle de la Fibromialgia / Síndrome Fatiga Crònica*. A la columna de la dreta té incorporat un enllaç que activa el traductor automàtic de la Generalitat de Catalunya, i permet la traducció de l'original català al castellà.



### Els correctors

No toca parlar, estrictament, dels correctors ortogràfics o gramaticals, perquè són específics per a una llengua, i no per a una situació de multilingüisme. Tot i així, és evident que la diversitat lingüística del nostre planeta passa per la incorporació d'eines com aquestes al nombre més gran possible de llengües.

Com a curiositat, val a dir que fins i tot la llengua criolla disposa d'un corrector ortogràfic. S'anomena Korektè Òtograf Kreyòl i al seu lloc web ens dóna la benvinguda així: *Byenvini sou sit wèb Korektè Òtograf Kreyòl la !.*



### Referència bibliogràfica

A. Oliver (2006). "La traducció automàtica a Internet". *Revista Tradumàtica: Traducció i Tecnologies de la informació i la Comunicació* (vol. 04).

### 3.3. La traducció assistida per ordinador

Hi ha una sèrie de productes oferts en més d'una llengua, però que són responsabilitat d'un traductor (o equip de traductors), que només poden fer la feina en el temps de què disposen gràcies a les eines de traducció assistida per ordinador que hi ha. En comentarem alguns.

#### Referència bibliogràfica

Els interessats en la història de la traducció automàtica i assistida (*machine translation*, en anglès) poden consultar el treball següent:

**Jonathan Slocum** (1985, gener-març). "A Survey of Machine Translation: Its History, Current Status, and Future Prospects". *Computational Linguistics* (vol. 11, núm. 1, pàg. 1-17).

El diari de Lleida, *Segre*, n'és un primer exemple. Gràcies als productes de Translendum (avui Lucy Software Ibèrica) aquest diari, fins l'any 1997 editat només en castellà, passa a treure dues edicions en paper cada dia, en català i en castellà, amb el suport econòmic de la Generalitat de Catalunya. A la mateixa època neixen les versions catalanes d'*El Periódico*, Europa Press o EFE, igualment amb traducció assistida.



Mirem una segona aplicació concreta, en un projecte empresarial, d'un d'aquests traductors. Es tracta del diari *El Periódico de Catalunya*.

L'edició en llengua catalana d'*El Periódico* és als quioscos des del 28 d'octubre de 1997 gràcies, en bona part, a l'ajuda d'un traductor automàtic. I diem en bona part perquè si bé és veritat que sense aquesta eina tecnològica el producte no es podria fer, no és menys cert que tampoc seria possible sense la participació diària d'un equip humà integrat per una quarantena llarga de lingüistes que deixen els textos resultants en condicions de ser publicats, tal com explica Fité.

**Referència bibliogràfica**

R. Fité (2006). "El Periódico, una experiència en traducció automàtica". *Revista Tradumàtica: traducció i tecnologies de la informació i la comunicació* (núm. 04).

Fité ens recorda el que dèiem al començament. És a dir, les TIC, en definitiva, són una ajuda a les persones:

"Hem de tenir molt present que la màquina ignora el coneixement del món, aquell coneixement que tenim les persones i que sense adonar-nos-en donem en general per sabut. Si no l'avisés [...] et pots trobar de cop i volta que el diari *La Vanguardia* s'ha convertit en *L'Avantguarda*, o que el president Zapatero es diu Sabater, i la seva vicepresidenta De l'Horta."

R. Fité (2006). "El Periódico, una experiència en traducció automàtica". *Revista Tradumàtica: traducció i tecnologies de la informació i la comunicació* (núm. 04).

L'article de Fité és molt útil perquè il·lustra com al llarg del temps la sofisticació de les regles que aplica el traductor automàtic va millorant el resultat global, però que també pot espatllar la traducció de frases que abans es traduïen bé.

### **Pàgina web**

A Internet hi ha una impressionant base de dades amb els textos de centenars de ponències i comunicacions sobre la temàtica, començant amb la conferència sobre traducció automàtica celebrada del 17 al 20 de juny de 1952 a l'Institut Tecnològic de Massachusetts (MIT), fins avui. S'hi pot accedir per mitjà del *Machine Translation Archive*.

Hi ha altres usos interessants i, fins i tot, curiosos. Per exemple, hi ha un web que permet veure una traducció automàtica (mitjançant el programa de Google) de la versió digital de *La Vanguardia*, diari editat des de sempre en castellà.

Els pronoms febles desapareixen d'aquests textos, i també hi ha curiositats com ara la traducció del mot *castellano* per *atalà*, a la notícia il·lustrada sobre el reglament del Senat.

## **3.4. Les tecnologies de la parla**

Tal com ens explica Llisterra, l'aplicació més coneguda de la síntesi de la parla és l'anomenada conversió de text en parla, ja que serveix per a transformar qualsevol text escrit en l'equivalent de la seva lectura en veu alta.

"[Algunes veus són] de persones de carn i ossos que han enregistrat prèviament els missatges que escoltem, de la mateixa manera que nosaltres gravem una salutació en el contestador automàtic del mòbil. D'altres, en canvi, han estat generades artificialment per un ordinador i representen el resultat del treball d'un equip de recerca dedicat al que s'anomenen, globalment, tecnologies de la parla. Els investigadors que treballen en aquest camp fan possible que arribem a mantenir un diàleg reeixit amb un sistema informàtic o que, quan conduïm, no ens calgui separar les mans del volant per a marcar, amb la veu, un número de telèfon."

Però també cobreix el cas invers, el reconeixement de la parla, que:

"[...] proporciona un mitjà d'entrada d'informació a un sistema informàtic sense necessitat de tenir ocupades les mans i la vista. Per això troba les aplicacions més comunes en el dictat automàtic de textos –que poden anar d'un article sencer a un SMS–, en el comandament vocal de programes informàtics o en accions com ara marcar amb la veu un número de telèfon o cercar un contacte en una agenda."

Joaquim Llisterra (2009). "Les tecnologies de la parla". *Llengua, Societat i Comunicació* (núm. 7, pàg. 11-12).

Aquestes tecnologies tenen un gran impacte econòmic, ja que poden cobrir –si més no, en part– serveis d'atenció al públic de manera automàtica.

L'Informe de política lingüística 2008 de la Generalitat de Catalunya explica en el capítol VI el suport de la Secretaria al Centre de Tecnologies i Aplicacions del Llenguatge i la Parla (TALP) de la Universitat Politècnica de Catalunya per a un projecte que inclou el següent:

"[...] l'estudi de tots els components d'un sistema de traducció de la parla que permeten obtenir un sistema obert, fiable i útil. Els components d'aquest sistema són els següents:

- Un sistema de reconeixement automàtic de la parla.
- Un sistema de traducció automàtica.
- Un sistema de síntesi de la parla.
- Una plataforma accessible que permeti la instal·lació d'un o de més d'un dels components esmentats."

Generalitat de Catalunya (2009). Informe de política lingüística 2008.

El compromís de suport de la Secretaria de Política Lingüística ascendia a un màxim d'1.500.000 euros (500.000 euros anuals), per als exercicis 2007, 2008 i 2009.

## Resum

En aquest mòdul hem vist els aspectes següents:

- Com es gestiona el multilingüisme a Internet i a les TIC.
- Com l'opció lingüística es gestiona a les pàgines web, amb alguns exemples dels criteris d'ús lingüístic en espais multilingües.
- Com es gestiona el multilingüisme fent servir programes informàtics. Els casos específics de Microsoft, Softcatalà i Ubuntu ens han il·lustrat diferents aspectes del procés. Hem vist com s'han elaborat orientacions per a l'elaboració de programari bilingüe.
- Algunes de les eines per a la gestió de la diversitat lingüística en els textos. Hi ha eines integrals de gestió del multilingüisme, generalment oferides per empreses especialitzades. I també hi ha eines per a la traducció, tant traductors automàtics com suports per a la traducció (assistida, doncs, per ordinador).
- Les tecnologies de la parla, que són útils en moltes aplicacions, incloent-hi els telèfons mòbils.

