



La traducció automàtica a Internet

Antoni Oliver

Universitat Oberta de Catalunya

1. Introducció

A Internet es poden trobar una gran quantitat de productes i serveis relacionats amb la traducció automàtica. Molts d'aquests productes poden ser d'utilitat per a un traductor. Els tipus de serveis que s'ofereixen poden ser diversos i proposem una classificació basada en 4 variables. La classificació que presentem no té res a veure amb la tecnologia de traducció automàtica que fan servir, sinó amb la manera de presentar aquest producte a Internet i altres característiques del producte.

Les quatre variables que proposem són:

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

Per mode d'accés al producte entenem la manera com l'usuari accedeix al producte o servei. La manera més freqüent és a través d'una plana web des d'on l'usuari pot fer una sèrie d'accions, com enviar a traduir una frase o document, descarregar un programa, etc. Existeix també una altra possibilitat i és que una aplicació, no l'usuari, accedeixi automàticament al servei que s'ofereix a Internet. Com a possibilitat d'utilització en diferenciem dues: exclusivament a través d'Internet, o bé la possibilitat de descarregar i fer servir el producte en el propi ordinador. Pel que fa referència al preu, distingirem els productes o serveis gratuïts dels comercials. I per últim, farem una distinció d'aquells casos en què la traducció automàtica és el servei principal d'aquells en què és un servei de valor afegit.

Qualsevol classificació que fem podrà tenir mancances, però aquesta que proposem ens permet presentar les idees principals i els productes i serveis de traducció automàtica més rellevants, posant un especial èmfasi als que s'ofereixen al nostre país.

2. Mode d'accés: plana web / servei web

El mode d'accés més habitual és el clàssic accés a través d'una plana web. El cas més habitual permet escriure o enganxar un fragment de text o bé carregar algun document, seleccionar el parell de llengües desitjat i en alguns casos algun

paràmetre més (com pot ser la especialitat o bé el fet de marcar d'una manera especial les paraules desconegudes) i realitzar la traducció. Típicament també permeten navegar per Internet traduint automàticament les pàgines que visitem. A la figura 1 presentem la interfície web del sistema de traducció automàtica OpenTrad (www.opentrad.com).



Figura 1. Interfície web del sistema de traducció automàtica OpenTrad

Aquesta forma d'accés a un sistema de traducció automàtica disponible a Internet és la més habitual. Requereix la intervenció d'un usuari humà, que es connecta a la plana web i fa les accions desitjades.

Una possibilitat semblant a aquesta i que també està pensada per què la faci servir un humà són les webs que ofereixen el servei de traducció automàtica de missatges de correu electrònic. Com a exemple podem mostrar la interfície de traducció d'emails del traductor PROMT (<http://www.online-translator.com/>), que podeu observar a la figura 2.

Figura 2. Interfície del sistema de traducció d'emails PROMT

L'aplicació ens permet escriure o enganxar un missatge, traduir-lo, i visualitzar-lo per poder-lo modificar si cal abans d'enviar-lo. També és possible enviar en el mateix missatge el text original i el traduït.

Existeix una altra possibilitat d'utilització d'un sistema de traducció automàtica a través d'Internet. En aquesta altra possibilitat ja no és un usuari humà, sinó un programa, qui accedeix al sistema de traducció automàtica. El sistema de traducció automàtica es presenta com a servei web. Un servei web (web service en anglès) és un programa dissenyat per a poder interaccionar amb un altre programa a través d'una xarxa, i més concretament a través d'Internet. Els avantatges dels serveis web són nombrosos:

- Programes escrits en diferents llenguatges de programació i funcionant sota diferents sistemes operatius poden comunicar-se entre ells.
- Com que funcionen sota protocol HTTP (el protocol d'Internet) poden aprofitar-se de tots els sistemes de seguretat existents, com per exemple els firewall, sense la necessitat de canviar les regles de filtrat.

L'arquitectura més habitual és la representada a la figura 3. Un servidor proporciona un cert servei i un programa que funciona al client sol·licita aquest servei.

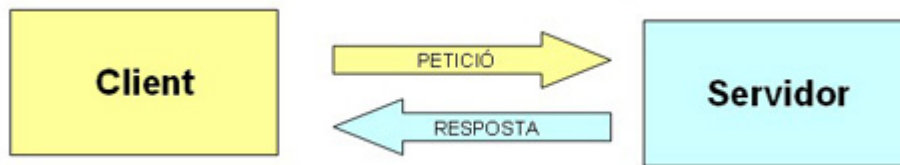


Figura 3. Arquitectura bàsica Client – Servidor a un servei web.

Existeixen molts serveis webs disponibles de manera pública. Podeu trobar una bona selecció d'aquests serveis i provar-los a la plana web: www.xmethods.com.

Un protocol molt emprat als serveis web és el protocol SOAP. Tant la petició del client com la resposta del servidor es realitzen mitjançant missatges en format XML. A les figures 4 i 5 es poden observar uns exemples de peticions i respostes SOAP.

```
<SOAP-ENV:Body>
<namespace2:translate_segment xmlns:namespace2=
"http://sindarin.upf.es/MTservice">
<c-gensym4 xsi:type="xsd:string">demo</c-
gensym4>
<c-gensym6 xsi:type="xsd:string">
Hello, good morning
</c-gensym6>
<c-gensym8 xsi:type="xsd:string">eng2spa</c-
gensym8>
</namespace2:translate_segment>
</SOAP-ENV:Body>
```

Figura 4. Un exemple de petició SOAP.

Els paràmetres estan marcats en negreta: usuari "demo", la frase a traduir "Hello, good morning" i el parell de llengües: "eng2spa".

```
<SOAP-ENV:Body>
<namespace2:translate_segmentResponse
xmlns:namespace2=
"http://sindarin.upf.es/MTservice">
<s-gensym8 xsi:type="SOAP-ENC:base64">
Hola, buenos días
</s-gensym8>
</namespace2:translate_segmentResponse>
</SOAP-ENV:Body>
```

Figura 5. Exemple de resposta SOAP. El servidor retorna la traducció "Hola, buenos días"

Aquesta tecnologia s'ha fet servir amb èxit al sistema de subtitulat automàtic e-Title (Melero-2006). El prototipus de sistema de traducció aplicat a aquest projecte feia servir el protocol SOAP tant per comunicar-se amb diversos sistemes de traducció automàtica, com per recuperar segments de memòries de traducció remotes. Antonopoulos (2004) també proposa una integració de tecnologies de traducció en un servei web fent servir el protocol SOAP.

Si el servei de traducció automàtica que desitgem utilitzar no disposa d'una interfície amb protocol SOAP o similar, l'accés automàtic des d'una altra aplicació serà més complicat. En alguns casos es pot accedir al servei mitjançant un POST que simula l'acció d'omplir el formulari i enviar la petició de traducció. A la figura 6 podem observar un fragment de codi en Perl que fa una petició al servei de traducció de Yahoo.

```
use HTTP::Request::Common qw(POST);
use LWP::UserAgent;

my $ua = LWP::UserAgent->new();
my $req = POST
'http://babelfish.yahoo.com/translate_txt',
[ text => "Hello, good morning", lp =>
"en_es", btnTrTxt=>"Translate"];
$content = $ua->request($req)->as_string;

print "$content\n";
```

Figura 6. Exemple de codi Perl que sol·licita la traducció de "Hello, good morning" de l'anglès al castellà.

El codi de la figura 6 sol·licita la traducció de "Hello, good morning" de l'anglès al castellà. La resposta del sistema s'emmagatzema a la variable \$content. Aquesta variable conté un codi html que si el visualitzem veurem la interfície habitual amb la traducció realitzada, com podem observar a la figura 7.

En español

Hola, buena mañana

Search the web with this text

Translate again  (Enter up to 150 words)

Hello, good morning

English to Spanish

Translate a web page 

http://

English to Spanish

Figura 7. Visualització (parcial) de la resposta de la petició via POST al servei de traducció automàtica de Yahoo

A la figura 8 es pot observar un fragment del codi html corresponent a la resposta del servei de traducció.

```

<h1>En espa&ntilde;ol</h1 >
<br clear="all">
<form action="http://search.yahoo.com/search" method=get >
<div id="result"><div style="padding:0.6em;">Hola, buena
mañana</div></div>
<input type="hidden" name="p" value="Hola, buena
mañana">
<input type="hidden" name="fr" value="bf-res">
<input type="hidden" name="ei" value="utf8">
<input class="inp_btn" type=submit name=search
value="Search the web with this text">
</form>
<hr noshade size="1" color="#FFF"><h2
style="float:right;margin-right:0.5em">( Enter up to 150
words )</h2>
<h1>Translate again</h1> &nbsp; 

```

Figura 8. Fragment de l'html de resposta del servei de traducció automàtica

La nostra aplicació haurà de recuperar el text de la traducció fent ús, per exemple, d'expressions regulars.

Cal tenir en compte que l'ús de programes que accedeixin automàticament a serveis que s'ofereixen mitjançant una web poden contravenir les normes d'ús del propi servei. Si es realitza un nombre excessiu de peticions és possible que el sistema de seguretat del servei rebutgi les peticions provinents de la nostra IP.

3. Utilització: ús online / descàrrega

La majoria de sistemes de traducció automàtica que s'ofereixen per Internet, almenys de forma gratuïta, permeten únicament la seva utilització on-line. Pocs són els sistemes que es poden descarregar i fer servir a la pròpia màquina.

- La utilització del sistema de traducció automàtica a la pròpia màquina ofereix diversos avantatges, entre els que es poden destacar:
- Una velocitat de traducció, en la majoria dels casos, molt més elevada.
- En la majoria de casos no hi haurà una limitació en el nombre de paraules o en la mida dels documents a traduir
- No hi haurà dependència en la disponibilitat o no de connexió a Internet
- En alguns casos serà possible personalitzar el sistema mitjançant l'addició de noves entrades lèxiques

Entre els sistemes comercials que permeten la descàrrega d'una versió de demostració de 30 dies podem destacar el sistema PROMP (<http://shop.e-promt.com/>). Hi ha també sistemes gratuïts que es poden descarregar. D'entre aquests sistemes en podem destacar dos:

- Salt (http://www.cult.gva.es/SALT/salt_programes_salt2.htm): es tracta d'un sistema català – castellà i castellà – català desenvolupat per la Generalitat Valenciana.
- Apertium – Openrad (<http://apertium.sourceforge.net/> i www.opentrad.org): es tracta d'un sistema de traducció automàtica gratuït, de lliure distribució i de codi obert.

El sistema Apertium – Openrad mereix una atenció especial. Aquest sistema ha estat desenvolupat en el marc del projecte PROFIT "Traducció automàtica de codi obert per a les llengües de l'estat espanyol". Es tracta d'un sistema de traducció automàtica de transferència sintàctica superficial desenvolupat inicialment per a la traducció entre llengües relacionades. La gran novetat d'aquest sistema és que s'ha alliberat sota una llicència de codi obert, cosa que significa que qualsevol persona amb els coneixements adequats pots ampliar el sistema de traducció i fins i tot crear nous sistemes per a noves llengües.

4. Preu: comercial / gratuït

Una altra distinció entre els diferents sistemes de traducció automàtica que podem trobar a Internet és la referent al preu, és a dir, si es tracta d'un servei o producte gratuït o comercial. Molts sistemes comercials ofereixen serveis gratuïts amb una limitació del nombre de paraules a traduir. D'aquesta manera poden

evitar un ús massiu del sistema per part de l'usuari.

Els sistemes de pagament, en la majoria dels casos, s'ofereixen com a programa per a instal·lar a un ordinador propi. Existeixen, però, alguns serveis de traducció automàtica de pagament per paraula traduïda. Un exemple d'aquesta possibilitat la constitueix Translendum (www.translendum.com). Aquesta empresa ofereix el sistema de traducció Comprendium com a producte o bé com a servei de traducció. També ofereix la possibilitat de traduir texts curts o planes web gratuïtament.

5. Servei principal / servei de valor afegit

Les principals empreses que desenvolupen sistemes de traducció automàtica tenen presència a Internet i ofereixen algun sistema per provar el seus productes. Però les empreses desenvolupadores no són les úniques que ofereixen traducció automàtica. Molts dels sistemes de traducció automàtica disponibles a Internet s'ofereixen com a valor afegit d'un altre tipus de servei. Entre aquests cal destacar les opcions de traducció automàtica que ofereixen els principals cercadors d'Internet:

- Google (<http://translate.google.com>): ofereix un total de 20 parells de traducció, entre ells el castellà-anglès i l'anglès-castellà)
- Yahoo (<http://babelfish.yahoo.com/>): ofereix un total de 38 parells de llengües, entre ells el castellà-anglès, l'anglès-castellà, el castellà-francès i el francès-castellà
- Altavista (<http://babelfish.altavista.com/>): ofereix un total de 34 parells de llengües, entre ells el castellà-anglès, l'anglès-castellà, el castellà-francès i el francès-castellà

Algunes institucions també posen a disposició del públic en general sistemes de traducció automàtica. Alguns casos a destacar són:

- Generalitat de Catalunya: a la seva plana web <http://traductor.gencat.cat/> ofereix la traducció automàtica de texts curts i planes web del català a l'anglès, castellà i francès i viceversa.
- Instituto Cervantes: a la seva plana web <http://oesi.cervantes.es/traduccion.jsp> permet traduir textos curts i planes web del castellà a l'anglès, català, francès i portuguès i viceversa.

6. Integració de la traducció automàtica oferta a Internet al flux de treball del traductor professional

La integració de la traducció automàtica dins del flux de treball d'un traductor presenta dues dificultats principals:

- La qualitat de la traducció automàtica
- La integració en el flux de treball del traductor

La qualitat de la traducció depèn molt del parell de llengües implicats. Per a llengües properes, com el català-castellà o català-francès, els resultats de la traducció automàtica acostumen a ser força acceptables. Per a un traductor professional la qualitat de la traducció es pot jutjar per un aspecte purament pràctic de temps de postedició de la traducció automàtica. Si es triga menys a

corregir els possibles errors de la traducció automàtica que a redactar la traducció des de zero, es pot considerar que el sistema presenta qualitat acceptable. Cal tenir present, però, que la correcció de traduccions automàtiques requereix una concentració molt alta i es fàcil descuidar-se d'algun detall.

Si considerem que un determinat sistema de traducció automàtica ofereix una qualitat suficient, queda per resoldre el segon problema. Com integrar la traducció automàtica en el nostre flux de treball? Considerarem aquí que el flux de treball inclou la utilització de sistemes de traducció assistida. Ens interessarà, doncs, poder combinar el nostre sistema de traducció assistida, amb el sistema de traducció automàtica. El resultat final desitjat serà que, dins de la nostra eina de traducció assistida, si el segment que estem traduint no es troba a la memòria de traducció amb una similitud suficient, se'ns presenti el resultat de la traducció automàtica. La integració de la traducció automàtica amb la traducció assistida no és una idea nova (Hutchins 2003), però la majoria d'eines encara no ofereixen solucions satisfactòries.

Ens podem trobar amb dos casos diferenciats: que l'eina disposi d'algun mecanisme per connectar-se amb aplicacions externes (com per exemple el WordFast) o que no disposi d'aquest mecanisme.

Si disposa del mecanisme per connectar-se amb aplicacions externes, caldrà programar una aplicació que es connecti amb un servei de traducció automàtica via SOAP o amb algun mecanisme tipus POST (vegeu l'apartat 2. Mode d'accés: plana web/servei web) i retorni la traducció automàtica. D'aquesta manera, quan no es trobi un segment adequat a la memòria de traducció, es farà la petició de traducció i es mostrarà el resultat. L'usuari podrà acceptar-lo i modificar-lo, o bé rebutjar-lo i fer la traducció manualment. El problema que pot tenir aquesta opció és que la resposta del servei de traducció automàtica pot trigar uns segons a completar la petició.

En el cas que el sistema de traducció assistida que fem servir no permeti fer una connexió amb una aplicació externa, caldrà plantejar una altra estratègia. Els sistemes TAO, quan passem a traduir un segment nou, fa la cerca a una o més memòries de traducció. La estratègia per poder incorporar traducció automàtica consistirà a crear una memòria de traducció a partir dels segments originals del projecte i la traducció automàtica d'aquests segments.

L'autor d'aquest article ha creat una petita eina gratuïta, anomenada, Stem-TMX Translator, que permet traduir, a partir d'alguns serveis gratuïts de traducció automàtica disponibles a Internet, fitxers TMX o de text tabulat. L'aplicació, a partir de l'original i de la traducció crearà una memòria de traducció que pot estar en format TMX o de text tabulat. Aquesta aplicació es pot descarregar de la plana web www.lsi.upc.es/~aoliver. Cal tenir en compte dos aspectes per la utilització d'aquesta eina. En primer lloc, cal recordar que l'eina fa servir serveis gratuïts i aquests poden canviar amb el temps, cosa que fa imprescindible anar comprovant l'existència de noves versions. En segon lloc, cal tenir present que l'ús dels sistemes poden contravenir les normes del propi servei i un ús abusiu (traduir un projecte enorme, per exemple) pot ocasionar la desconnexió del servei.

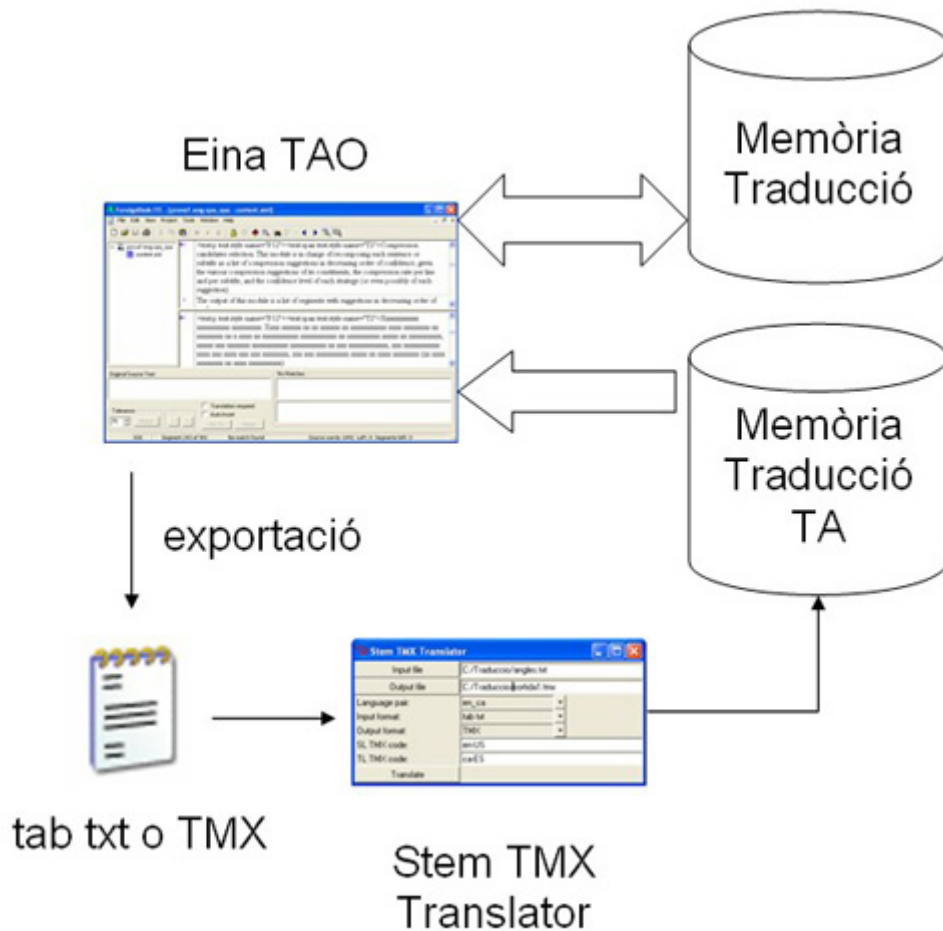


Figura 9. Flux de treball per crear memòries de traducció a partir de traduccions automàtiques fent servir l'eina Stem TMX Translator

A la figura 9 podem observar el flux de treball que cal seguir per crear una memòria de traducció a partir de traducció automàtica fent servir l'eina Stem TMX Translator. A continuació expliquem els passos amb més detall.

El primer que cal fer és crear el projecte de traducció amb tota normalitat. Podem assignar-li una o més memòries de traducció, si és que en tenim de disponibles.

Un cop creat el projecte de traducció hem d'aconseguir exportar el text del projecte en format text (amb un segment per línia) o en format TMX. La manera d'aconseguir això dependrà de l'eina que feu servir habitualment. Si no és possible exportar directament el projecte en algun d'aquests formats caldrà fer els següents passos. Cal pseudotraduir el projecte (si és que disposem d'aquesta opció) o bé copiar tots els segments originals a la traducció. Un cop fet això passarem tot el contingut del projecte a una memòria de traducció buida. Un cop creada la memòria de traducció d'aquesta manera l'exportarem a format TMX o bé a format de text separat per tabuladors.

Ara ja podem fer servir l'Stem TMX Translator. A la figura 10 podeu observar la interfície gràfica d'aquest programa. La seva utilització és molt senzilla. Només cal indicar el fitxer d'entrada (Input file) (que pot estar en format TMX i tab txt) i el

fitxer destí (Output file), on es guardarà la memòria de traducció generada a partir de traducció automàtica (que també es podrà generar o bé en format TMX o tab txt). Cal seleccionar el parell de llengües amb Language pair i també indicar els formats d'entrada i sortida. Per les operacions relacionades amb el format TMX caldrà indicar també els codis TMX de la llengua origen i destí (SL i TL TMX code). Un cop seleccionat tot això ja es pot pitjar el botó Translate i el procés de traducció s'iniciarà. Recordeu que cal estar connectat a Internet.

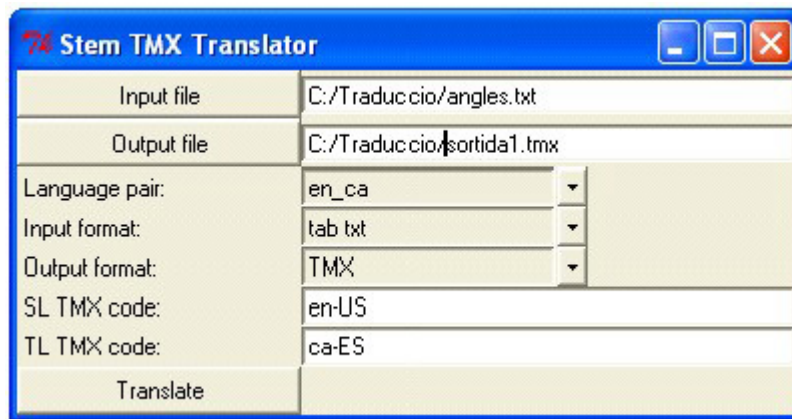


Figura 10. Interfície gràfica del programa Stem TMX Translator

7. Conclusions

A Internet s'ofereixen un gran ventall de productes i serveis de traducció automàtica. Alguns d'aquests productes poden ser interessants per a l'activitat d'un traductor professional. La integració d'aquests productes i serveis al flux de treball no sempre és senzilla. En aquest article hem presentat una classificació dels sistemes de traducció disponibles a Internet a partir d'uns paràmetres funcionals. També hem presentat una proposta per integrar aquests tipus d'eines al flux de treball.

Bibliografia

[Melero 2006] M. Melero, A. Oliver, T. Badia (2006) Automatic Multilingual Subtitling in the eTitle Project. To be published at the Proceedings of the ASLIB Conference "Translating and the Computer 28 Conference"

[Antonopoulos 2004] Antonopoulos, V.; Demiros, I.; Carayannis, G.; Piperidis, S., "Integrating translation technologies towards a powerful translation Web service," Cybernetics and Intelligent Systems, 2004 IEEE Conference on , vol.1, no.pp. 526- 531 vol.1, 1-3 Dec. 2004

[Hutchins 2003] Hutchins, J. (2003) Machine translation and computer-based translation tools: what's available and how it's used. University of Valladolid (Spain) – March 2003
<http://ourworld.compuserve.com/homepages/wjhutchins/Valladolid-2003.pdf>

Desembre 2006