

La importància de les memòries de traducció en els processos traductius automatitzats

Assessorament
i Terminologia

L'automatització de la traducció

La traducció és una activitat molt antiga relacionada amb la necessitat de comunicació entre les comunitats i els individus: avui dia, però, aquesta necessitat pren dimensions incommensurables tant pel volum de textos que es tradueixen, com per les combinacions lingüístiques que hi prenen part. Des d'aquest punt de vista, no és d'estranyar que, en l'era de les tecnologies de la informació i la comunicació, s'hagin invertit recursos de tot tipus per automatitzar el procés de traducció.

Podem dividir els sistemes d'automatització de la traducció en dos grans grups en funció del paper de la intervenció humana: d'una banda la traducció automàtica i de l'altra la traducció assistida.

Els sistemes anomenats *de traducció automàtica* proporcionen una traducció generada a partir d'uns reculls lexicals i d'unes regles gramaticals extretes de la simplificació dels llenguatges en presència, i la qualitat del text resultant varia en funció de com el sistema és capaç de resoldre les ambigüitats pròpies del llenguatge natural: lèxiques, morfològiques, sintàctiques i semàntiques. Fent ús dels diccionaris i de les gramàtiques corresponents, el sistema analitza l'oració que es vol traduir i transferenceix l'estructura sintàctica amb el lèxic adient a la llengua per generar a continuació l'oració traduïda. En qualsevol dels casos, la qualitat de la traducció d'aquests sistemes ha de venir avalada per un traductor o corrector, el qual ha de preparar el text abans d'enviar-lo a traduir, tot eliminant les possibles ambigüitats i dificultats lingüístiques, operació que es coneix com a *preedició*; i posteriorment, ha de revisar la proposta de traducció, tot corregint els possibles errors de traducció, tasca que és coneguda com a *postedició*.

Val a dir que el resultat de les traduccions obtingudes amb els sistemes de traducció automàtica varia en funció del grau de similitud entre les llengües de treball, en aquest sentit no és el mateix, quant a resultats, la combinació català → castellà que català → anglès.

Els sistemes considerats de traducció assistida són coneguts també com a *sistemes de*

gestió de memòries de traducció (SGMT) i es fonamenten en la recuperació d'unitats de traducció traduïdes en processos anteriors. Estem parlant, doncs, de bases de dades o col·leccions de documents que contenen unitats de traducció constituïdes per oracions en la llengua origen amb les seves corresponents traduccions. Al seu torn aquestes unitats de traducció poden venir acompanyades per informació relativa a les preferències lingüístiques dels clients, als dominis temàtics o d'altres dades considerades rellevants per a la tasca del traductor.

La unitat de traducció bàsica és coneguda com a *segment*, i consta d'una cadena de paraules delimitada pels següents signes de puntuació: punt i coma, punt i seguit, punt i a part, dos punts, signe d'interrogació, signe d'admiració o bé una tabulació tipogràfica.

Procés de treball amb memòries de traducció

El procés de treball amb memòries de traducció es basa en la cerca de cadascun dels segments que conformen el text original i la seva substitució automàtica quan troba una unitat de traducció existent dins la memòria. El sistema detecta també segments amb un grau elevat de similitud però no idèntics que poden ser presentats al traductor per tal que aquest determini el seu possible aprofitament tot esmenant la diferència existent amb la traducció adequada. En qualsevol cas el traductor haurà de traduir tots i cadascun dels segments que el sistema no ha pogut traduir. Alguns sistemes utilitzen la percentualitat per indicar si els segments trobats són idèntics (100%) o similars (especificant en aquest cas el percentatge de similitud <100%); d'altres opten per una codificació cromàtica per indicar aquests tipus d'informació (verd: idèntica, etc). Un cop traduït el text tots els segments que el conformen passaran a formar part de la memòria de traducció per a usos posteriors. Des d'aquest punt de vista, doncs, com més es fan servir les memòries de traducció, més gran és l'índex potencial de segments traduïbles en projectes posteriors.

Autor

Ramon Piqué
Grup Tradumàtica
Departament de Traducció
i d'Interpretació
Universitat Autònoma de Barcelona

L'article exposa les principals estratègies per a l'automatització del procés traductiu tot destacant el funcionament de les memòries de traducció. Es fa una introducció als aspectes més rellevants d'aquests sistemes pel que fa a la gestió terminològica, el manteniment de memòries, els formats i els filtres de conversió, la gestió de projectes i l'intercanvi de memòries entre sistemes. Per acabar es presenta un esquema del procés de treball que integra els sistemes de traducció automàtica i assistida.

En tant que es tracta de reciclar traduccions, el procés de treball amb memòries de traducció és més productiu quan treballem amb textos amb un índex de repetició elevat i quan hem de vetllar de manera especial per la seva coherència terminològica. D'entrada, doncs, estem parlant de textos especialitzats, com ho són els científics, els tècnics, els jurídics i els econòmico-comercials, encara que algunes característiques del treball amb memòries fan que aquesta estratègia de reaprofitament de traduccions pugui ser aplicada a d'altres àmbits textuais, com ara els textos d'informació periodística.

De la mateixa manera que es pot recuperar un segment d'una base de dades o d'una col·lecció de documents, també alguns sistemes tenen la possibilitat de cercar automàticament els termes concrets que apareixen en els segments amb l'objectiu de facilitar l'elaboració de noves traduccions. En aquest sentit, tots els sistemes de gestió de memòries de traducció incorporen una eina de gestió terminològica; alguns fins i tot amb funcions avançades de substitució automàtica dels termes que ofereixen propostes de traducció a partir dels termes o porcions de segment trobats: els resultats són força apreciables quan es tracta de llengües properes.

Tipus de sistemes de gestió de memòries de traducció

Podem classificar els SGMT en funció de la forma d'emmagatzemar i recuperar la informació. D'una banda tenim els que consisteixen en sistemes de gestió de bases de dades, i que han estat dissenyats seguint les pautes d'aquesta mena d'aplicacions informàtiques. Per tant, cada unitat de traducció es correspon a un registre i cada segment es correspon a un camp, i també són camps aquelles dades que es puguin afegir referents als clients o als àmbits temàtics de cadascuna de les unitats de traducció.

D'altra banda tenim els sistemes que consisteixen en col·leccions de parells de documents, un en llengua origen i el seu corresponent en llengua destí, que estableixen relacions d'equivalència entre cadascun dels diferents segments que conformen els parells d'arxius. En certa manera estem parlant de corpus textuais bilingües i/o multilingües, i a l'hora de configurar el procés de treball, haurem de seleccionar quina col·lecció de documents, arxius o

carpetes volem fer servir com a memòria de traducció per traduir un nou projecte.

L'editor: l'espai de treball del traductor

Tal com hem apuntat a l'hora de descriure el flux de treball amb les memòries de traducció, una tasca amb què ens trobarem habitualment és la de modificar i/o traduir els segments que no han trobat equivalent en les memòries. En definitiva, estem parlant d'una àrea de treball on hem de poder realitzar aquelles tasques d'edició pròpies del tractament de textos: escriure, esborrar, moure, copiar, fer macros, fer autotext, autocorrecció, correcció ortogràfica, etc. Òbviament, tots els sistemes de gestió de memòries de traducció disposen d'aquest espai de treball; en alguns casos el sistema fa servir directament un editor de textos estàndard; en d'altres casos el sistema disposa d'un editor propi que incorpora moltes de les tasques que normalment podem fer amb un processador de textos.

El manteniment de les memòries de traducció

Un dels aspectes importants del treball amb les memòries de traducció és el seu manteniment, i en aquest sentit entenem la possibilitat de poder modificar una memòria de traducció, ja sia per esmenar errors com per invertir les llengües de treball, fer substitucions globals, fusions de dues o més memòries de traducció, etc.

Tots els sistemes incorporen funcions per realitzar aquestes tasques de manteniment, d'altra banda indispensables per assegurar el bon aprofitament de les memòries de traducció.

En aquest apartat, però, és important destacar la possibilitat, que ofereixen alguns sistemes, de realitzar consultes per veure exemples de com ha estat traduït un terme específic en segments anteriors.

La gestió de la terminologia

Probablement és en la gestió de la terminologia on la informàtica ha esdevingut capdavantera en l'assistència a la tasca del traductor. El treball amb memòries de traducció aprofita, i fins i tot millora, aquest tipus d'assistència ja que, a més de facilitar la cerca de termes, i la possibilitat d'afegir

entrades o de modificar-les durant el procés de traducció, molts sistemes incorporen la possibilitat de substituir de manera automàtica els termes del segment que es vol traduir per les traduccions trobades dins les bases de dades terminològiques, etc.

Alguns sistemes tenen funcions, manllevades dels sistemes de gestió de corpus lingüístics, els quals ofereixen la possibilitat de crear llistes de paraules extretes dels textos que han de ser traduïts i ordenades per la freqüència d'aparició. El resultat és una radiografia molt bona dels termes més utilitzats en el text i, consegüentment, del grau de complexitat del text objecte de traducció.

Aquestes llistes de paraules, al seu torn, serveixen de base per a la confecció de glossaris, amb els corresponents equivalents en la llengua destí, els quals proporcionen una ajuda inestimable a l'hora de traduir els segments no trobats al si de la memòria de traducció.

Aprofitament de les traduccions anteriors

Una de les primeres demandes associades a l'increment de l'ús de les memòries de traducció és la de poder aprofitar aquells textos traduïts abans de l'existència dels sistemes de gestió de memòries de traducció. Tractant-se com es tracta d'un procés de treball basat en l'aprofitament de traduccions, per què no s'haurien de poder aprofitar els textos fets abans de la incorporació de les memòries de traducció? Evidentment, això té molt de sentit en el cas de les empreses de traducció, dels serveis de traducció d'empreses, o bé en el cas de molts traductors autònoms.

Això vol dir disposar d'eines que tenen per objectiu la creació de memòries de traducció, que hauran de poder ser utilitzades en els processos de traducció ulteriors tal com hem descrit anteriorment. Crear memòries a partir de textos existents implica identificar i associar cadascun dels segments de la llengua original amb els seus corresponents de la llengua destí. Aquestes eines són anomenades *alinearadors*, en tant que la seva feina consisteix a alinear els segments tenint en compte els delimitadors de puntuació que hem descrit. I és en aquest fet, en la segmentació per signes de puntuació, on es troba el principal escull d'aquestes eines, en tant que a l'hora de traduir no és inusual que es produeixin variacions a l'hora de situar, eli-

minar o introduir aquests elements gramaticals separadors i més encara tractant-se com es tracta de llengües diferents amb normes de puntuació probablement no coincidents.

Ara com ara, el procés d'alineació és un treball força automàtic, però que requereix el concurs d'un traductor o corrector perquè vetlli per eliminar les falses equivalències derivades principalment dels canvis de puntuació.

Formats i filtres

Tot document de text electrònic consta de dues parts perfectament identificables: el text pròpiament dit i el format electrònic. El format electrònic té a veure amb l'aplicació informàtica amb què ha estat creat el document, i normalment cada format porta associada una extensió d'arxiu específica.

<i>Format</i>	<i>Extensió</i>
Word	.doc
Plantilla word	.dot
Adobe	.pdf
Rich Text File	.rtf
Text/ANSI	.txt
Text/ANSI	.asc
Word Perfect	.wpd
Page Maker	.pmd
Web	.htm
Web	.html
MSWorks	.wps

Els documents de text que traduïm, doncs, poden estar guardats en qualsevol dels formats apuntats més amunt, i fins i tot amb formats força menys comuns. Ara com ara, però, la majoria de documents estan fets amb Microsoft Word, que a la pràctica està funcionant com a estàndard per a aquest tipus de documents. Tanmateix en determinats àmbits de treball la producció de documentació es realitza amb programes d'autoedició com ara Frame Maker o Page Maker. I també cada cop més el format HTML és utilitzat per a un nombre important de tasques vinculades a Internet.

No és estrany, doncs, que els traductors es trobin amb encàrrecs de traducció amb documents guardats en formats poc comuns. En aquest sentit, convé conèixer com resolen aquesta qüestió els sistemes de gestió de memòries de traducció. D'en-

trada convé distingir entre els sistemes que utilitzen com a editor un processador de textos estàndard, com ara el Word, i els que incorporen el seu propi editor.

Els primers necessitaran nous editors dissenyats expressament per tractar els formats diferents de l'editor estàndard. Així trobem que Trados Workbench, que té el Word com a editor estàndard, per treballar amb documents HTML incorpora un editor a part, anomenat Tag Editor.

Els segons són aquells sistemes que requereixen que el text que es es vol traduir sigui convertit al format propi. Normalment aquesta conversió es produeix en el que en aquests sistemes s'anomena *procés d'importació*, procés que realitza també la segmentació del text. La importació correcta dels documents implica que el sistema en qüestió disposi del filtre corresponent al format en què està guardat el document que volem traduir. Com que les versions d'un mateix programa poden originar arxius amb formats diferents en relació amb les seves pròpies versions anteriors, aquests sistemes necessiten actualitzar els filtres d'importació quan apareixen nous formats derivats de noves versions o de noves aplicacions.

El valor afegit d'aquests sistemes rau, però, justament en el mode de treball basat en la conversió de formats, ja que estalvia al traductor haver de conèixer l'aplicació amb què s'havia guardat el document que ha de traduir.

La gestió de projectes

Parlar de gestió de projectes en el món de la traducció és parlar d'encàrrecs de traducció voluminosos que requereixen el concurs de diversos traductors, i en conseqüència tenen la necessitat de coordinar la seva feina. Aquesta voluminositat dels projectes pot venir definida pel nombre de paraules del projecte, però també pel nombre de combinacions de les llengües de treball. Un escenari possible des d'aquest punt de vista pot comptar amb professionals de la traducció, autònoms i/o en plantilla, amb la missió de traduir cadascun d'ells una part del projecte. Un escenari d'aquests, doncs, requereix tenir en compte la tasca de repartiment de la feina entre els traductors, també la de vigilar perquè la col·lectivització de la traducció no generi inconsistències de caire terminològic i estilístic, i evidentment també perquè un cop traduït tot l'encàrrec con-

servi la seva coherència plena, tant dels del punt de vista lingüístic com des del punt de vista del format tipogràfic.

Si estem parlant d'un procés de traducció digitalitzat que fa servir memòries de traducció la gestió de projectes suposa l'enviament de les memòries i de les bases de dades terminològiques als traductors, suposa la conversió de formats quan el sistema així ho requereixi, i suposa també en molts casos resoldre tots els problemes traductològics que es derivin del projecte en qüestió, especialment els que l'afecten globalment. Un cop traduïdes cadascuna de les parts, la persona encarregada de la gestió del projecte té per missió, a més de vetllar pels aspectes de coherència abans esmentats, l'alimentació de les memòries de traducció i de les bases de dades terminològiques generals a partir de les que han anat generant els traductors que han participat en en projecte. Cal no oblidar que un dels eixos d'aquests sistemes és la retroalimentació (reaprofitament) per a usos posteriors, tal com hem anat assenyalant repetidament.

L'intercanvi de memòries

Existeixen en el mercat una dotzena llarga de sistemes de gestió de memòries de traducció. Cadascun té les seves particularitats i en general tots responen, si fa no fa, a l'esquema de funcionament descrit anteriorment. Alguns funcionen en sistemes operatius específics, d'altres van néixer de la dinàmica professional traductora de l'empresa que els comercialitza, i uns tercers han estat desenvolupats per traductors que els ofereixen a baix preu als professionals que ho desitgin.

Val a dir, però, que malgrat aquest nombre aparentment elevat de sistemes existents n'hi ha tres o quatre que són els més estesos en el sector de la traducció i que són per tant els més requerits en el currículum del professional de la traducció. A la pràctica, un traductor pot acabar treballant amb dos o tres d'aquests sistemes, depenent del client o de l'empresa de traducció per a la qual faci feina. Una conseqüència directa de la diversitat de sistemes és l'acumulació de memòries de traducció fetes amb formats diferents i la consegüent dificultat perquè les memòries siguin aprofitables per a tots els sistemes disponibles.

Actualment aquesta dificultat pot ser resolta gràcies al format TMX (Translation

Memory eXchange), un format basat en el metallenguatge XML que permet marcar els elements bàsics d'una memòria i que està pensat justament per facilitar l'intercanvi entre memòries de traducció. Val a dir que avui dia la majoria dels sistemes existents al mercat disposen de filtres d'importació i exportació a TMX.

Les memòries de traducció a través de la xarxa

En una perspectiva de futur, si tenim en compte la lògica de funcionament de les memòries de traducció, basada en la cerca i el reaprofitament de cadenes de caràcters, juntament amb les capacitats tecnològiques actuals de la xarxa, que cada cop permeten una major velocitat d'accés i de gestió, i tenint en compte les aportacions orientades a la millora de la recuperació i l'aprofitament de les memòries com ara el llenguatge d'etiquetes TMX, no és d'estranyar que comenci a haver-hi propostes d'accés a memòries de traducció ubicades en servidors accessibles via Internet.

I encara en aquesta línia, alguns autors dibuixen un escenari on Internet esdevindrà un dipòsit obert de traduccions accessible amb cercadors, tipus Google, especialitzats en la consulta de corpus multilingües.

En aquest sentit només cal pensar, per exemple, que molts dels corpus lingüístics paral·lels que podem trobar a la xarxa, convenientment marcats amb un llenguatge que segueix les especificacions de l'XML com ho és el TMX, són potencialment memòries de traducció formades per col·leccions de textos, de manera similar com ho són els sistemes de gestió de memòries de traducció basats en parells d'arxius marcats.

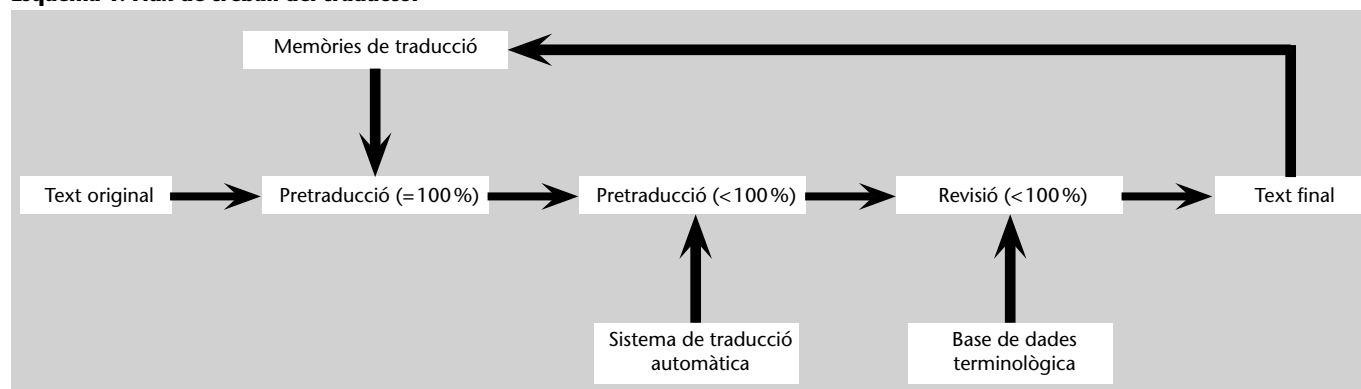
La integració de la traducció automàtica amb la traducció assistida

Seguint la idea de futur apuntada en el paràgraf anterior és pertinent de situar en l'escenari la combinació de les dues estratègies de mecanització apuntades al començament d'aquest article: la traducció automàtica i la traducció assistida. Una visió desapassionada dels límits de la tecnologia permet veure que no són dues propostes excloents, ans al contrari, són perfectament integrables, i que aquesta integració ofereix d'entrada uns resultats a tenir en compte. Cal recordar, també, que la similitud de les llengües de treball és un factor a tenir en compte, juntament amb les característiques del text i el grau d'aprofitament d'una memòria de traducció específica, a l'hora de decidir el pes d'aquestes dues categories d'eines dins el procés de traducció que volem aplicar.

L'esquema de funcionament segueix essent el que hem apuntat anteriorment quan descrivíem les memòries de traducció: els segments idèntics són reemplaçats automàticament, i els segments que no han estat trobats a la memòria de traducció i que, per tant, haurien de ser traduïts de bell nou pels traductors, són traduïts mitjançant un sistema de traducció automàtica. El resultat és que tots els segments, absolutament tots, tenen una proposta de traducció, i el traductor es limita a repassar i validar tots aquells segments que han estat generats pel sistema de traducció automàtica.

L'esquema 1 il·lustra el flux de treball del traductor desgranat al llarg del text: (1) cadascun dels segments del text original és cercat dins la memòria de traducció i reemplaçat si escau automàticament per la seva traducció; en aquest esquema, aquest

Esquema 1. Flux de treball del traductor



procés està identificat com a pretraducció =100%; (2) els segments que no han estat trobats a la memòria de traducció són els que seran traduïts pel sistema de traducció automàtica; en aquesta fase el traductor pot optar per aprofitar els segments trobats a la memòria de traducció que no són plenament idèntics (<100%); (3) en la tasca de revisió i de validació, el traductor té també accés a la les eines de suport terminològic.

En darrer lloc, només cal apuntar que els segments que han de ser corregits i validats són detectables de manera ràpida i còmoda per la majoria d'aquests sistemes, ja sia per mitjà d'una identificació cromàtica o per d'altres mitjans d'identificació visuals.

Bibliografia d'ampliació

ABAITUA, Joseba (2001). «Memorias de traducción en TMX compartidas por Internet». *Revista Tradumàtica*, número 0, *Memòries de traducció*.

<<http://www.fti.uab.es/tradumatica/revista/num0/index.htm>>

AUSTERMÜHL, Frank (2001). *Electronic tools for translators*. Manchester: St. Jerome.

ISABELLE, Pierre; CHURCH, Kenneth W. (ed.) (1996) *New Tools for Human Translators. Machine Translation*, vol. 12 (1-2), Pittsburgh.

MORENO-TORRES, Ignacio (1999). «Técnicas informáticas para la traducción». A: José Manuel BLECUA [et al.] (ed.) *Filología e informática*. Barcelona: Universitat de Barcelona. Pàgines 117-130.

PIQUÉ, Ramon; CASTELLANOS, Carles (2002). «Els avantatges i servituds de la traducció assistida per ordinador». A: J. MALLAFRÉ (ed.) *Actes de les II Jornades per a la Cooperació en l'Estandardització Lingüística*. Barcelona: Institut d'Estudis Catalans. Pàgines 143-156.

SÁNCHEZ-GIJÓN, Pilar [et al.] (2002). «La tradumática: las tecnologías de la información y las comunicaciones y la traducción». A: Silvia GAMERO y Amparo ALCINA (ed.) *La traducción científico-técnica y la terminología en la sociedad de la información*. Castelló: Servei de Publicacions. Universitat Jaume I. (Estudis sobre la Traducció; 10)

SÁNCHEZ-GIJÓN, Pilar (2001). «Catalèg de memòries de traducció». *Revista Tradu-*

màtica, número 0, *Memòries de traducció*. <<http://www.fti.uab.es/tradumatica/revista/num0/index.htm>>

VALERO, Carmen; CRUZ, Isabel de la (ed.) (2001). *Traducción y nuevas tecnologías. Herramientas auxiliares del traductor. Encuentros en torno a la traducción 4*. Madrid: Servicio de Publicaciones, Universidad de Alcalá de Henares.

VILARNAU, Joan (2001). «La tradumática. Delimitació conceptual i importància en el procés de formació del traductor». Treball de recerca. Bellaterra: Universitat Autònoma de Barcelona. Departament de Filologia Catalana.