

## Integració de la TA en el procés professional de la traducció

Toni Badia

GLiCom, Barcelona Media – Centre d'Innovació i Departament de Traducció i Filologia, Universitat Pompeu Fabra

### 1. Introducció

Considerem TA qualsevol sistema automàtic que, donat un text en una llengua, en proporcioni una versió en una altra. I això independentment de la tècnica concreta que faci servir, de la qualitat del resultat, o de la finalitat amb què ha estat establert... A la vegada, és clar que no té sentit incorporar a un procés professional de traducció un sistema de TA (STA, d'ara endavant) amb poc vocabulari i amb poca finor en la selecció lèxica en traduir, que hagi estat pensat per ser usat en el context d'internet (amb poc vocabulari, i per a textos breus). Més aviat ens cal pensar en sistemes que s'adaptin bé al flux de treball en què s'han d'inserir, que tinguin una qualitat mínima en el tractament del vocabulari, i que puguin ser millorats amb el treball quotidià en el seu context professional. En aquest article entrem a valorar aquestes qüestions. Analitzem primer què entenem per STA. A continuació observem que la manera de treballar a què s'han acostumat molts traductors amb les memòries de traducció pot ser reproduïda perfectament amb un STA, amb la condició que tingui unes característiques determinades, que analitzem. Finalment, fem esment de dos aspectes complementaris de la traducció amb STA: l'ús del llenguatge controlat en la producció de textos, i la relació de la producció de textos multilingües amb la traducció.

### 2. Distincions prèvies

Els STA es poden observar (i classificar) des de diverses perspectives: des de la finalitat de la TA, les llengües involucrades en el sistema, les tècniques emprades... Fem doncs, a l'inici d'aquest article, unes petites distincions que ens ajudaran a entendre de què parlem. Sovint, [1] es distingeix entre traducció automàtica d'alta qualitat, traducció automàtica assistida per traductors o editors, i traducció humana assistida per ordinador. En la primera opció, el STA seria capaç de produir traduccions d'una qualitat semblant a la de les traduccions produïdes pels traductors humans.[2] Aquesta opció és clarament irrealista; per tant, l'alternativa es troba entre sistemes que ajuden a traduir i sistemes que necessiten traductors per finalitzar la tasca; ara bé, aquesta distinció, que sembla molt clara de bon antuvi, no resulta operativa si mirem productes i situacions concretes. Perquè, des del punt de vista de l'usuari, realment quina diferència hi ha entre un STA inserit en un programa d'edició de traduccions i una memòria de traduccions en un entorn de traducció? Podem pensar que en el primer cas el STA produeix traduccions 'artificials', mentre que la memòria de traduccions proposa sempre traduccions 'humanes'...; però això no és ben bé així, ja que el STA pot ser dels que es basen extensivament en dades, i llavors ja no és tan clar que les traduccions proposades són 'artificials' sempre. I per altra banda, l'entorn de treball pot ser el mateix; i llavors les diferències queden reduïdes al tipus de correccions que cal fer (i finalment les diferències acabaran depenent totalment de la memòria de traduccions consultada, o del STA emprat).

Una distinció bàsica en parlar de TA és la que parteix de la finalitat amb què s'usa un STA, que reflecteix l'ús de la traducció en general. Hutchins (2003) distingeix

entre tres usos dels STA: per disseminar textos, per assimilar el contingut de textos, i per intercomunicació immediata entre parlants de llengües diferents. Naturalment, en un entorn professional tots tres tipus de traducció hi poden tenir un lloc; no obstant, la majoria de les traduccions que es fan en el context professional respon al paradigma de traducció per disseminar textos, siguin informes, manuals, catàlegs, o documents comercials, administratius o jurídics.

Per altra banda, la majoria de les classificacions dels STA es basen en aspectes tècnics: si operen amb regles o amb regularitats estadístiques extretes de grans col·leccions de textos, i totes les subdivisions possibles.[3] Aquí prescindim d'aquestes distincions; [4] i considerem STA tot aquell sistema que no recorre directament a les dades tal com són emmagatzemades, sinó que les manipula d'alguna manera per oferir més propostes de traducció que les simplement emmagatzemades. Així, una memòria de traducció que no només recuperi segments traduïts, sinó que els recomposi d'alguna manera (és a dir, seguint algun criteri quantitatiu o d'altre mena), serà un STA; mentre que un producte que només ofereixi substitució lèxica dels mots no ho serà.

### **3. Les memòries de traducció en els entorns de traducció**

Repassem breument el treball en un entorn de traducció amb una memòria de traduccions ja iniciada. Després de crear el projecte, importar els documents i establir els paràmetres amb què es vol treballar, el procés de traducció pròpiament dit té les fases següents. Al principi, en col·locar el cursor en la primera oració del document, es produeix una consulta automàtica a la memòria de traducció; el sistema retorna les unitats de traducció que tenen l'oració original idèntica a la que es vol traduir o semblant (en un grau predeterminat en establir els paràmetres de treball); aquestes unitats de traducció són proposades al traductor de forma ordenada segons el grau de similitud que tingui la part original amb la frase que es vol traduir (el nombre de propostes ofertes normalment també és parametrizable). Amb aquestes dades, el traductor cerca la millor proposta de traducció, la valida si la considera correcta, o la modifica si considera que no ofereix una traducció adequada. En aquest procés, el traductor compta amb una sèrie de recursos addicionals (principalment, accés a diccionaris, al document sencer que s'està traduint i a tota la memòria de traducció –que actua de corpus de referència per a l'encàrrec concret de traducció). A continuació, el procés es reinicia amb l'oració següent. Alternativament, el sistema pot fer totes les tasques automàtiques seguides, sense esperar la validació o modificació feta pel traductor. En aquest cas, el traductor treballa al final amb les propostes de traducció ofertes per totes les oracions del text. Sovint, aquest procés és complementat per un treball inicial de determinació de la terminologia o d'altre vocabulari específic, o d'estudi dels aspectes complexos del document des de la perspectiva de la traducció.

Així doncs, la feina del traductor que treballa en un entorn de traducció equipat amb una memòria de traducció consisteix bàsicament en recollir l'output del sistema i modificar-lo, si és el cas, perquè sigui una traducció adequada. Aquesta revisió té dos aspectes, que tot i ser visibles sobretot si es treballa de forma interactiva són conceptualment distingibles en general. Per una banda, hi ha una validació o modificació oració per oració, que té en compte el context oracional simplement; i per l'altra, hi ha una revisió de la totalitat del text, per garantir-ne la cohesió i la consistència.

### **4. La incorporació d'un STA en un entorn de traducció**

Doncs bé, aquest procés pot ser clarament reproduït en treballar amb un STA. Per això només cal que el STA estigui integrat en un entorn de traducció adequat, de

manera que el traductor es trobi assistit en la tasca de traduir de la mateixa manera que s'hi troba quan treballa amb memòries de traducció. Llavors el STA té una funció equivalent a la de la part automatitzada del treball amb memòries de traducció: proporciona al traductor una proposta de traducció per al text original, perquè la revisi, validi o modifiqui segons convingui. Per tant, si el STA està incorporat en un entorn de traducció, es pot donar el cas que el qui hi treballi no noti cap diferència formal si el sistema li proporciona traduccions que provenen d'una memòria de traducció o d'un STA.

D'aquesta manera, la seqüència de treball amb un STA no difereix gaire de la que se segueix quan es treballa amb una memòria de traducció. Ara bé, hi ha diferències quant al tipus de modificacions que s'esperen del traductor, si l'input li prové d'una memòria de traducció o d'un STA. En el primer cas, el punt de partida serà sempre correcte lingüísticament, tot i que probablement allunyat del contingut que cal transmetre amb la traducció (depenent del grau de semblança requerit en parametritzar la consulta a la memòria de traducció). En el segon cas, el punt de partida conté molt sovint inadequacions o errors de traducció, provocats per deficiències del STA. Per tant, hi ha diferències importants en el tipus de modificacions que cal que faci el traductor i, en general, en els aspectes en què s'ha de fixar. Així, un corrector entrenat a corregir errors humans de redacció pot tenir inicialment dificultats en corregir els resultats d'un STA, ja que hi apareixeran errors força diferents; o un traductor habituat a treballar amb memòries de traducció pot necessitar un període d'entrenament per tenir un rendiment satisfactori en treballar amb un STA. També és clar que, com que hi ha diferències importants entre els STA, els hàbits adquirits en treballar amb un poden no ser adequats per treballar amb un altre.

Per tant, no sembla que hi hagi d'haver incompatibilitats importants en l'ús d'un STA en un procés professional de traducció. Ara bé, perquè la integració sigui beneficiosa, cal que el STA estigui inserit en un entorn de traducció que permeti a l'usuari traductor treballar amb la mateixa comoditat amb què treballa en els entorns de traducció amb memòries de traducció.

## **5. Característiques d'un STA per ser inserit en un entorn de traducció**

Les característiques del STA condicionen la bona inserció en un entorn professional de traducció. En general, l'adaptació al tipus de llenguatge propi dels encàrrecs de traducció, la interacció amb memòries de traducció i les facilitats de modificació del STA segons l'ús que se'n fa, són essencials per a l'ús natural d'un STA. L'adaptació del STA al tipus de llenguatge afecta tant el lèxic com les estructures lingüístiques del tipus de text amb què s'ha de treballar. Si els encàrrecs de traducció es refereixen a un sol tipus de text (p. ex., es treballa per a un sol client: traducció de manuals de manteniment d'automòbils, d'un diari...), s'ha de produir una adaptació inicial, consistent en la introducció del lèxic i de les estructures adequades.[5] Si en canvi els encàrrecs de traducció es refereixen a diversos tipus de textos, aquesta adaptació inicial ha d'incloure la capacitat de gestió de diversos diccionaris i, idealment, la possibilitat de modificació del tractament de les estructures lingüístiques.[6]

La interacció amb memòries de traducció és més una característica de l'entorn de traducció en el qual s'insereix el STA, que no pas del STA mateix. Parlem d'interacció entre un STA i una memòria de traducció quan l'entorn de traducció pot proporcionar al traductor tant el resultat de la cerca en la base de dades de traduccions com la traducció proporcionada pel STA.[7] Una qüestió bàsica en aquesta interacció és com els dos sistemes segmenten el text; dit altrament, amb quina unitat de traducció operen. Normalment, tant les memòries de traducció com

els STA segmenten el text en oracions. [8] Ara bé, no hi ha dubte que una interacció estreta entre els dos tipus de sistemes només es pot aconseguir si se segmenta en unitats inferiors, perquè l'aprofitament mutu de les dues estratègies es maximitza si els segments són menors que l'oració: els errors del STA disminueixen en traduir sintagmes nominals i la presència de segments menors que l'oració en una base de dades és major que la de les oracions.

També és essencial poder modificar el STA per adaptar-lo a les necessitats pròpies dels textos que s'hi hagin de traduir. Encara que es tracti d'un STA creat específicament per a una tasca determinada (p. ex., la traducció d'un diari), cal que pugui ser adaptat als resultats de l'experiència d'usar-lo i als canvis que es van produint al llarg dels anys en els textos que cal traduir i en el llenguatge en general. [9] Idealment, un STA hauria de poder ser adaptat a les necessitats que van sorgint a mesura que s'hi va treballant, de manera que anés aprenent de les estructures lingüístiques i vocabulari que van apareixent quan s'hi treballa. [10]

És clar, doncs, que per poder ser integrat en un flux professional de traducció un STA ha de tenir unes característiques determinades, que per altra banda no es troben habitualment entre els STA disponibles al mercat. Això explica, en gran part, el poc ús que es fa dels STA en l'àmbit de la traducció professional per difondre. En trobem alguns exemples, en camps tècnics o de la premsa, però són relativament pocs, comparat amb l'increment notabilíssim en l'ús de les memòries de traducció i dels entorns de traducció en què aquestes estan inserides.

Un bon exemple d'introducció d'un STA en un procés professional de traducció el trobem amb la producció diària de la versió catalana del El Periódico. [11] D'aquesta experiència, ens n'interessa de remarcar aquí la integració de la traducció en el procés general de producció del diari. De fet, aquest va ser un dels requeriments imprescindibles que hi havia a l'hora de plantejar-se de traduir automàticament el diari: si la integració no era total, no era possible complir amb els terminis per tenir el diari als quioscos a l'hora. I com a conseqüència d'aquest requeriment, van haver d'optar per no utilitzar cap dels STA disponibles en aquell moment per traduir del castellà al català.

## **6. El control del llenguatge dels textos que s'han de traduir**

La majoria d'exemples de STA introduïts en fluxos professionals de traducció els trobem en el camp de la traducció tècnica. En aquest àmbit, es donen les circumstàncies favorables que permeten que un STA convenientment integrat en el flux de traducció pugui ser usat amb avantatges clars. Habitualment, en els textos tècnics el vocabulari emprat és relativament limitat i, sovint, menys ambigu que en el llenguatge general, les estructures lingüístiques que hi apareixen són també bastant limitades, i l'ús d'estructures relativament simples i repetides no perjudica la qualitat dels textos que es tradueixen. Aquestes característiques de molts textos tècnics han propiciat una estratègia específica per augmentar el rendiment dels STA en aquest tipus de tasques. Es tracta la conversió dels textos que s'han de traduir en el que es coneix com un llenguatge controlat.

Parlem de llenguatge controlat quan aconseguim que els textos que s'han de traduir tenen unes característiques determinades, quan són el resultat d'un procés de control del seu vocabulari i estructures sintàctiques. Per entendre'ns, podem relacionar el concepte de llenguatge controlat amb l'aplicació sistemàtica del llibre d'estil que moltes empreses i institucions tenen en relació als documents que produeixen. En una guia d'estil hi podem trobar indicacions del tipus "d'aquest objecte en direm X", "evitarem expressions del tipus de Y", o "en cas de dues opcions X i Y optarem sempre per Y". Si s'apliquen les tècniques pròpies dels

correctors gramaticals a aquest objectiu, s'obtenen programes automàtics de control de les característiques lingüístiques dels textos, tant de vocabulari com de sintaxi i d'estil. Doncs bé, un text controlat adequadament pot ser traduït més fàcilment que no un en què hi poden aparèixer qualsevol mot o estructura lingüística. Sobretot si el control s'efectua tenint en compte les possibilitats reals dels STA, podem ajudar a què aquests programes tinguin molts menys entrebancs a l'hora de traduir.[12]

Si això es fa bé, el mateix procés que garanteix que un text sigui conforme a les característiques lingüístiques que se li volen donar, el prepara per ser traduït per mitjans automàtics amb la millor qualitat possible.

De fet, el procés de control del llenguatge d'un text es correspon gairebé exactament amb el procés de preparació del text abans de ser traduït, que es preveu habitualment en els manuals d'utilització dels STA. En aquest context cal remarcar que la preparació del text per ser traduït és essencial per a la millora dels resultats del procés automàtic i, per tant, per a la reducció del temps de correcció o postedició. Si, addicionalment, la preedició es pot fer de forma automàtica o semiautomàtica, com és el cas de processos amb llenguatge controlat, el temps global esmerçat en la traducció es redueix considerablement.

## **7. La traducció i la producció de textos multilingües**

Finalment, aquestes darreres consideracions ens porten a una altra observació important. És diferent enfrontar la traducció com a activitat aïllada, o com un pas més en el procés de produir textos multilingües. En el primer cas, el procés de traducció s'ha d'adaptar enterament al text que li arriba, el qual no ha estat escrit tenint en compte el seu futur processament automàtic per un STA. En aquesta situació, el grau d'independència del responsable de la traducció és molt baix; ha de retornar una traducció raonable del text que li arriba, i per tant s'ha de mantenir fidel al màxim a l'estructura i el llenguatge del text.

En canvi, la situació és radicalment diferent si l'autor del text original és també el responsable de la traducció. En aquest cas, l'autor pot prendre decisions que, sense perjudicar el contingut del text original, faciliten la tasca de traduir automàticament. I aquesta és la funció del procés previ de control del llenguatge de què parlàvem.

En la producció de textos tècnics el concepte de text original és poc pertinent, mentre que resulta essencial que el text resultant sigui igualment comunicatiu en cada una de les llengües. Per això, la manera com s'obté el primer text té relativament poca importància. Per altra banda, s'obté una clara reducció de costos de traducció si el text està escrit en llenguatge controlat i es pot traduir sense massa dificultats. Ho veurem clar en comparar les estratègies de dos productors d'automòbils respecte de la producció dels manuals de manteniment dels seus models. El primer té la producció de manuals centralitzada, de manera que tant la primera versió del manual com les seves traduccions són produïdes per la mateixa unitat de treball. En segon, en canvi, responsabilitza de la traducció dels manuals els centres distribuïdors d'automòbils a cada país. Clarament la primera estratègia és més adequada, perquè garanteix que en el moment de redactar la primera versió dels textos ja es té en compte que hauran de ser traduïts i, de la mateixa manera, permet de tenir en compte que poden ser traduïts per STA.

## **8. Conclusió**

En aquest article hem reflexionat sobre la integració dels STA en el procés

professional de la traducció. La varietat de STA que hi ha al mercat i el fet que majoritàriament són capses tancades que no es poden manipular ni adaptar gaire fan que la seva utilització en entorns professionals sigui escassa. Hem examinat les característiques que ha de tenir un STA per a ser usat satisfactòriament per a la traducció professional, i hem observat algunes pràctiques en aquest sentit. En definitiva, l'article ha mostrat que la traducció mitjançant un STA és possible, però només en unes condicions determinades.

### **Bibliografia citada**

Hutchins, J. (2003). "Commercial systems: the state of the art", a Somers, H. (ed.) (2003), *Computers and translation: a translator's guide*, Amsterdam: John Benjamins.

Hutchins, J. & H. Somers (1992). *An introduction to machine translation*. London: Academic Press.

Trujillo, A. (1999). *Translation Engines: Techniques for Machine Translation*. London: Springer Verlag.

### **NOTES**

1 - Vegeu, p.ex., Hutchins i Somers, 1992. [[tornar](#)]

2 - Sense obviar el fet que és considerablement difícil de mesurar la qualitat de les traduccions, tant si són resultat d'un procés automàtic, com del treball d'un traductor. [[tornar](#)]

3 - Vegeu Hutchins i Somers (1992) o, més actual, Trujillo (1999) per a detalls sobre les subclasses de STA més habituals. Vegeu, també, altres articles d'aquest mateix número de Tradumàtica: sobretot els de Melero, Farwell i Oliver. [[tornar](#)]

4 - Així com d'altres, essencialment les que tenen a veure amb les llengües amb què treballa un STA. Aquestes qüestions són pràcticament irrellevants en el moment d'inserir un STA en un entorn professional, ja que el que interessa realment és el nivell de qualitat assolit pel STA en els parells de llengües en consideració. [[tornar](#)]

5 - En els sistemes de base estadística, això suposa un nou entrenament del sistema a partir de textos típics de l'encàrrec de traducció corresponent; i en els sistemes simbòlics, la modificació de les regles de manera que contemplin les estructures pròpies d'aquests textos. [[tornar](#)]

6 - En sistemes estadístics, això suposa un entrenament diferent del sistema per a cada classe de textos; i en els simbòlics, l'organització de les regles en paquets diferents de manera que contemplin les estructures pròpies de cada classe de text. [[tornar](#)]

7 - Hi ha dues possibilitats en aquesta interacció: per una banda, l'entorn de traducció pot proporcionar a la vegada el resultat de la consulta a la base de dades de traduccions i la traducció que retorna el STA; i per l'altra, pot activar el STA només quan la base de dades de traduccions no retorna cap proposta de traducció. En el primer cas, el traductor sempre treballa amb els dos resultats, mentre que en el segon en uns casos treballa amb propostes provinents de la memòria de traducció i en d'altres amb traduccions fetes pel STA. [[tornar](#)]

8 - Per als STA això és així, perquè l'oració és la unitat lingüística natural; i per a les memòries de traducció, perquè els criteris de segmentació en oracions solen ser independents de llengua (és a dir, no demanen tractaments específics per a cada llengua). [[tornar](#)]

9 - Ens referim als canvis referents al vocabulari, però també a expressions o girs que esdevenen obsolets i a d'altres que s'incorporen a la llengua; a canvis en el llibre d'estil de l'empresa en qüestió, segons el qual estructures permeses en un

moment deixen de ser-ho en un altre, i viceversa; a canvis motivats per variacions importants en la temàtica (p. ex., a causa d'un canvi tecnològic important).

[[tornar](#)]

10 - Aquesta capacitat d'aprenentatge (i fins i tot la seva possibilitat) és radicalment diferent segons l'estructura tècnica interna del STA. Els STA estadístics (o, més generalment, construïts sobre dades reals) poden construir-se de manera que es vagin adaptant a les característiques dels diferents textos que s'hi tradueixen. Els STA basats en regles, en canvi, resulten molt més costosos d'adaptar a noves situacions, ja que s'hi ha de construir explícitament, i normalment de forma manual, les entrades lèxiques del vocabulari i les regles gramaticals adequades. [[tornar](#)]

11 - Vegeu, en aquest mateix volum, l'article de Ricart Fité sobre la implantació del STA en la producció diària de El Periódico. [[tornar](#)]

12 - En aquest sentit és il·lustratiu de llegir les recomanacions que es consignen en el manual d'ús del STA Systran en el parell anglès – espanyol: s'hi fan recomanacions com "evita l'ús de més dos noms compostos", "sempre que puguis posa l'article en els sintagmes nominals", "evita les oracions de relatiu sense pronom", etc. Com es pot veure, són recomanacions encaminades a evitar estructures en les quals el STA pot no produir la bona traducció. [[tornar](#)]

Desembre 2006