

## LA TRADUCCIÓ ASSISTIDA PER ORDINADOR VS. TRADUCCIÓ AUTOMÀTICA\*

MARIONA SABATÉ I CARROVÉ  
*Universitat de Lleida*

### 0 *Introducció*

La traducció va sorgir com una activitat necessària, espontània i bàsicament pragmàtica. A poc a poc va anar evolucionant fins arribar a aconseguir una base científica que ajudés a situar les diverses ideologies, tècniques, tipus i mètodes de traducció. Però la història de la traducció no té encara un final feliç; quan semblava que tot rutllava a la perfecció i que el matrimoni entre la lingüística i la traducció s'havia consumat satisfactoriament, apareix en escena l'anomenada traducció automàtica, una idea innovadora que pot posar en perill aquest perfecte enllaç.

La meua aportació a aquesta publicació té l'objectiu de fer una revalorització de la traducció automàtica de la mateixa manera que ho ha fet la traducció com a disciplina general. En primer lloc, faré una exploració de les seves perspectives de futur. Tot seguit analitzaré les raons del sentiment de desconfiança i escepticisme provinent de la traducció tradicional o 100% humana i, a partir d'aquí, faré una apologia de la traducció computeritzada com a terme que engloba tant la traducció automàtica (TA) com la traducció assistida per ordinador (TAO). L'objectiu final de la meua exposició és la de situar aquesta disciplina en el lloc que li pertany i es mereix en el món de la ciència i la lingüística.

### 1 *Perspectives de futur de la traducció computeritzada*

Tot i haver-hi un gran nombre de dates, autors i lingüistes que formen part de la història i descripció d'una gran quantitat de projectes i sistemes que s'han anat desenvolupant fins al moment actual, resulta més interessant analitzar les vessants per on es decanta la traducció computeritzada (TC) en l'actualitat i en el futur a fi de revaloritzar-la i establir-la com un fenomen que pot convertir-se en una realitat que la situï en un lloc prestigiós. Així doncs, s'aprecien dues tendències força clares: 1.1. El principi d'especialització i 1.2. Canvi d'orientació. Passem a analitzar-les.

#### 1.1 *Principi d'especialització (subllenguatges)*

La TC pot esdevenir un fet consolidat si es basa en el *principi d'especialització* amb la incorporació dels *sistemes de subllenguatges*.

\* La primera versió d'aquest article va ser presentada al I Taller de Ciències del Llenguatge que va tenir lloc a la Universitat Rovira i Virgili els dies 7 i 8 de juny de 1995.

La qualitat en la traducció computeritzada pot aconseguir-se de dues maneres: o bé amb la intervenció o assistència humana o bé minimitzant problemes mitjançant la imposició de restriccions quan al tipus de text per traduir i dissenyant sistemes per un tipus de text ben definit i específic. Així, es poden dissenyar sistemes que tradueixen un sol tipus de text, per exemple manuals d'aviació de l'anglès al francès (projecte TAUM-Aviation de la Universitat de Montreal des de 1976 al 1981). També es poden dissenyar sistemes per un camp temàtic concret on els textos estan escrits en un subllenguatge específic.

De fet, gairebé tots els subllenguatges tenen el seu propi argot, tant si són de ciències naturals (electró, ió, proteïnes), com de lingüística (fonema, pragmàtica) o afers militars (míssil, balística). La terminologia d'aquests subllenguatges inclou paraules del vocabulari comú amb significats especials (camp i massa en física, arbre i generar en lingüística, tanc i força en el llenguatge militar). Aquest fet, conegut ja des de la creació dels primers sistemes de TA, va provocar la creació de diccionaris especialitzats que els usuaris poden activar i accedir quan es tradueixen texts en un camp temàtic específic.

Si bé els subllenguatges no tenen gairebé mai unes característiques gramaticals concretes, certament hi ha uns estils que són més típics que d'altres. Per exemple, l'omissió d'articles definits és típica dels manuals d'instruccions en anglès, mentre que en manuals francesos s'utilitza l'infinitiu en comptes de l'imperatiu.

Els sistemes dissenyats per subllenguatges tenen els seus avantatges: terminologia ben definida, reducció d'homògrafs, concentració de problemes gramaticals típics d'un tipus de text i creació d'un sistema per una tasca específica amb un elevat grau d'èxit i acceptabilitat. Tot i així, en alguns subllenguatges poden presentar-se problemes. És relativament inusual trobar-se amb textos restringits a un tema específic. Així, un text mèdic pot incloure terminologia d'economia o física i un text de química pot incloure termes matemàtics i referir-se a sistemes biològics, etc. Un altre problema pot plantejar-se pel fet que un sistema dissenyat per un subllenguatge específic no pot fer-se fàcilment extensible a un altre. Una solució és la introducció de sistemes de *subllenguatges controlats* com ara el sistema TITUS, dissenyat per traduir documents de la indústria tèxtil en una de les quatre llengües anglès, francès, alemany i espanyol i que, en cas d'ambigüitat o homografia, deixa la decisió del significat més apropiat en mans de l'usuari.

De la mateixa manera que no es pot ser un bon traductor per a tots els camps de la ciència, no pot haver-hi un sistema de TC que serveixi per a tots els temes, tipus de text, etc. I aquí és on els sistemes específics o de subllenguatge poden ajudar a aconseguir qualitat en la traducció computeritzada.

## 1.2 *Canvi d'orientació: la TAO + la TA*

La qualitat i confiança en la TC pot també reestablir-se si, en comptes de concebre l'ordinador com a producte final, es pensa en aquest com a eina de treball que col·labora en la tasca traductora. La TA i la traducció feta per traductors no són les dues úniques alternatives sinó que han d'aplicar-se en una escala que indicaria el grau de combinació de les diferents formes d'automatització.

L'ordinador com a eina de traducció ha estat des de fa molt temps a l'abast. Pocs traductors utilitzen el medi escrit a mà. I són precisament aquestes eines que amenacen de canviar la imatge de la traducció d'un art i activitat artesanal a una

tècnica. No existeix, però, cap raó per témer que les eines informàtiques disminueixen la qualitat intel·lectual i artística de la traducció, així com tampoc no ho són l'ús de diccionaris i màquines d'escriure. Aquesta por sorgeix de la creença (provinent dels traductors de textos literaris) que la informàtica implica una pèrdua d'humanitat.

Tot i que els programes de tractament de text poden considerar-se eines informàtiques, no poden catalogar-se com a TAO. Com a mínim, un programa de TAO ha de tenir algun tipus d'ajut lingüístic de tipus informàtic, com ara un programa de correcció ortogràfica, gramatical o estilística. Cada vegada s'està estenent més l'ús de programes de referència en línia com ara diccionaris, tesausres, enciclopèdies i altres fonts d'informació general que els traductors poden consultar. Els discs de làser i compactes, capaços de contenir gran quantitat d'informació, poden integrar-se a entorns de programes de tractament de textos i pot accedir-se a aquests directament quan sorgeixi un problema de traducció. Assistim també a la comercialització de diccionaris bilingües en línia.

La integració de diversos recursos ha ocasionat el desenvolupament d'un concepte conegut amb el nom de taller de traducció (*translator's workbench*). Aquests sistemes incorporats als microordinadors permeten treballar amb pantalla dividida i múltiples finestres: una part de la pantalla és l'àrea de treball del text meta, amb tecles de funció per obrir finestres o subdividir la pantalla per consultar diccionaris en línia o altres fonts d'informació com ara traduccions anteriors en temes similars adreçades a un mateix client. Quan la informació pot transferir-se fàcilment d'una finestra a una altra s'anomena integració total. Els textos originals poden introduir-se directament a l'ordinador mitjançant el teclat, gravats des d'un disquet o transmesos mitjançant un enllaç telecomunicatiu via mòdem o introduït a l'ordinador mitjançant un lector òptic de caràcters. També permeten la creació de *glossaris orientats al text*, llistes de paraules que apareixen en un text determinat amb una traducció suggerida a la llengua meta; aquestes llistes poden estar per ordre alfabètic o per ordre d'aparició en l'original i les seves traduccions poden provenir del propi glossari del traductor o altres fonts en línia. Sovint es troben opcions alternatives de *consulta automàtica de termes*, consulta de bases de dades terminològiques internes o externes per paraules tècniques i especialitzades d'un text específic. Totes aquestes possibilitats d'escriptori de treball poden estalviar molt de temps als traductors, ja que s'ha calculat que els traductors tècnics inverteixen en les seves etapes inicials i fins arribar a familiaritzar-se amb el text en qüestió, un 60% del seu temps consultant diccionaris i llibres de referència.

El futur de la TC és una combinació de la TAO i la TA. A part d'aquestes ajudes terminològiques, existeixen programes que permeten la substitució lèxica d'una manera semi-automàtica i proporcionen una traducció aproximada. Són els anomenats *programes de pretraducció per ordinador*. Aquests programes substitueixen paraules i expressions del text original que tenen un sol equivalent en la llengua meta i deixen en versió original les paraules que tenen més d'una possibilitat de traducció en la llengua meta o que presenten ambigüitats com ara la traducció de l'anglès *use* a l'espanyol, que pot traduir-se com a *usar, utilizar, emplear*, etc), verbs (*do* i *make* en anglès), adjectius (*several, any* en anglès). Aquest tipus de pretraducció representa la mecanització d'una activitat amb la qual els traductors ja estan familiaritzats: la busca en un text de vocabulari tècnic que desconeixen i que necessita una recerca.

En resum, el restabliment de la confiança en la TC en el món de la lingüística es basa en la creació de programes d'ús especialitzat i que combinin la TAO i la TA. D'aquesta manera la TC adquirirà una base científica que la situarà en el digne lloc de prestigi dins el camp de la lingüística que es mereix.

## 2 *La traducció computeritzada: una pseudo-ciència?*

La traducció computeritzada està sotmesa a pressions provinents de tres fonts: el traductor 100% humà, el profà en la matèria i el mateix teòric de la traducció per ordinador. Aquest atac té la seva base en els dos màxims enemics de la ciència: (1) *la confusió* i (2) *la por*, dos factors que cal intentar eliminar "agafant el toro per les banyes" i analitzant la raó del seu perquè.

### 2.1 *La confusió*

La confusió pot desglossar-se en dues categories: terminològica i avaluativa.

#### 2.1.1 *Confusió terminològica*

Des dels primers esforços per elaborar un marc teòric on situar-hi la realitat de la TC fins al moment actual, s'ha creat i encunyat un extens corpus de termes i conceptes que, pel profà en la matèria, resulten ben semblants l'un de l'altre i, per l'analista de les diferents teories sobre la TC i lingüística, presenten ambigüitats i inconsistències que poden ser, fins i tot, contradictòries; de fet, una ullada per les diferents definicions i classificacions incloses en les teories sobre la TC fan evident el fet que s'utilitzen termes idèntics que cobreixen diferents referents o significats. Així, conceptes com traducció automàtica, traducció computeritzada, traducció assistida per ordinador, traducció humana assistida per ordinador, traducció automàtica assistida per humans, etc. tenen significats diferents per cada teòric. Per exemple, Lehrberger i Bourbeau (1988) fan la distinció entre traducció humana assistida per ordinador (THAO), traducció per ordinador assistida per humans (TOAH) i traducció per ordinador completament automatizada (TOCA). Per ells, la traducció humana assistida per ordinador significa traducció efectuada per un traductor amb poca ajuda de l'ordinador; en el cas de la traducció automàtica assistida per humans, el traductor dóna informació limitada per "omplir" la traducció automàtica i, després de subministrar aquesta informació, l'ordinador completa la traducció. Finalment, la traducció per ordinador completament automatizada no compta amb la intervenció humana fins al final de l'activitat traductora. Aquests conceptes són una evolució de la classificació establerta per Johnson i Whitelock (1987) on s'inclouia el concepte de traducció per ordinador completament automàtica d'alta qualitat, que s'anomenaria en l'actualitat traducció assistida per ordinador, o sigui, "una manera eficaç i productiva de compartir el que l'home i la màquina poden fer i que permet a la màquina augmentar la seva importància operativa". Posteriorment, Balkan (1992) distingeix entre traducció automàtica i traducció assistida per ordinador, referint-se a traducció automàtica com a "qualsevol sistema que efectua una traducció" i traducció assistida per ordinador "qualsevol eina de traducció computeritzada que ajuda a fer una traducció".

La manca d'homogeneïtat i l'encavalcament semàntic d'aquestes definicions i classificacions en el temps no contribueixen a clarificar idees i no fan més que augmentar la confusió sobre el món de la TC. Aquesta és una assignatura pendent

que cal aprovar amb bones notes si volem dotar la TC de credibilitat científica. La solució a llarg termini seria arribar a un consens terminològic. A curt termini, cal que els teòrics de la TC determinin exactament el significat dels termes elegits per tal de construir les seves teories. En un intent d'aconseguir-ho, i amb l'ànim de que serveixi de precedent, donaré definicions i establiré la classificació que considero més convenient per elaborar i exposar la meua teoria sobre l'anomenada "traducció automàtica".

En primer lloc, cal donar un nom genèric a tot aquest tipus de traducció feta amb ordinador o bé amb l'ajuda de l'ordinador; crec que el terme més adient és el de *traducció computeritzada*. A partir d'aquí podem establir una taxonomia, dient que existeixen en termes generals dos models de traducció computeritzada: la traducció per ordinador o automàtica (TA) i la traducció assistida per ordinador (TAO). Les diferències entre ambdues són diverses però la principal radica en el fet que la TA necessita molt poca assistència humana durant la fase de traducció mentre que la TAO deixa en mans del traductor el procés de traducció però col·labora de manera intel·ligent perquè aquest dugui a terme la seva activitat. D'aquesta se'n deriven unes quantes conseqüències, la més important de les quals és el producte final obtingut d'ambdues. Mentre que en la TA s'aconsegueix una equivalència amb el text original a nivell formal, de la TAO es desprèn una traducció lingüísticament més "completa" a tots els nivells: formal -sintàctic, morfològic-, semàntic i pragmàtic, ja que el traductor humà controla en tot moment el procés de traducció.

### 2.1.2 Confusió avaluativa

El segon element que contribueix a la confusió en el corpus de la traducció computeritzada és l'avaluació dels resultats. Donada la subjectivitat -l'enemic públic de la ciència- d'aquest factor, resulta difícil jutjar si una traducció és bona o dolenta. Però en quins paràmetres ens basem per dir que una traducció és bona o dolenta? Tenen o no aquests adjectius un elevat percentatge de subjectivitat? Certament que sí. Cal, doncs, delimitar amb objectivitat científica l'abast d'aquestes categories.

Abans d'iniciar una traducció, el traductor es veu sotmès a unes pressions i a partir d'aquestes establirà un ordre de prioritats en funció de les quals efectuarà la tasca de traducció. Així, una traducció serà bona o dolenta en tant que acompleixi les prioritats que s'ha marcat en un principi. En el cas de la TA, a més de les prioritats s'hi ajunten dos conceptes (Lehreberger i Bourbeau 1988): el model lingüístic introduït (glossaris, gramàtiques, bases de dades, etc) i el model computacional introduït per posar en pràctica el model lingüístic.

El cas de la TAO és diferent de la TA. Donat que el traductor sempre té l'última paraula en el procés de traducció, pot seguir amb més fidelitat i rigor lingüístic l'ordre de prioritats marcades, a més de donar consistència, qualitat i rapidesa al producte final.

## 2.2 La por

Tot seguit, analitzaré el segon màxim enemic de la TC: la por. Si, d'una banda, la confusió sorgeix de tres punts principals, els traductors 100% humans, el profà en la matèria i els mateixos traductors que utilitzen l'ordinador, la por té un sol focus causal: la comunitat traductora, en concret els traductors professionals que es guanyen la vida amb aquesta activitat. En general, l'ordinador ha estat vist des de sempre com una amenaça pels humans. És la lluita de l'home contra la màquina.

Aquest sentiment ancestral s'ha incrementat amb l'evolució qualitativa i quantitativa de la informàtica en el segle XX. D'objecte curiós ha passat a convertir-se en eina de treball indispensable. Ara bé, si l'ordinador ha arribat a ser una amenaça per al món laboral associat amb activitats repetitives i de memorització, no ho és encara en les activitats creatives i d'improvisació. La intel·ligència artificial (IA) és un fenomen que dista molt de ser un fet. Sense l'ajuda de l'home, la màquina és una capsa inerta, que no té la capacitat intrínseca de pensar, tot i que se li poden introduir unes pautes lingüístiques i matemàtiques a seguir. I aquesta és una realitat que encara dóna més força als que defenen els programes de taller de traducció com la millor manera d'efectuar una traducció competent.

### 3 *Apologia de la traducció computeritzada*

Des de les primeres versions de programes de TC s'ha establert un elevat corpus de classificacions i encunyat un gran nombre de definicions relacionades amb la TC. Aquest fet, lluny de ser concebut com "acientífic" o "poc científic", hauria de veure's com a positiu, ja que és un indicatiu molt clar de l'interès que aquesta disciplina ha despertat en tots els camps de la ciència com la matemàtica, la informàtica i sobretot la lingüística. Com a conseqüència d'aquesta varietat, la TC ha esdevingut una *ciència interdisciplinària* que engloba en el seu sí molts camps del saber. I això és un fet positiu perquè no fa més que confirmar la seva validesa i la de les altres disciplines científiques.

El problema sorgeix quan els detractors de la TC -traductors 100% humans, majoritàriament- qüestionen la seva qualitat lingüística i neguen la seva semblança amb el llenguatge natural. Aquestes afirmacions tenen el seu origen, novament, en la ignorància i, potser la por. Ignorància perquè, tot i que, no es pot negar que hi ha **ALGUNES** traduccions fetes per ordinador de poca qualitat, **NO TOTES** ho són. Segurament, els que critiquen la TC desconeixen els diversos tipus de TA i TAO existents i agafen exemples poc o gens representatius de programes de TC disponible en el mercat.

El que és un fet universal és que, de la mateixa manera que hi ha traduccions computeritzades "bones i dolentes", també hi ha traduccions 100% humanes bones i dolentes, sempre entenen bo i dolent segons acompleixin o no l'ordre de prioritats establert abans de començar el procés de traducció. Així, algunes traduccions no tenen com a prioritat màxima la seva lectura com a llenguatge natural o l'exactitud lingüística, sinó la comprensió per part del lector/client o la rapidesa en la producció de la traducció. No s'ha d'oblidar que la traducció, apart de ser un art, és un negoci i com a tal ha de satisfer uns terminis de lliurament i limitar-se a un pressupost. Aquests factors són fonamentals en la traducció i la TC permet una resposta afirmativa a aquests, mentre que la traducció 100% humana no pot sempre comprometre's o cenyir-se a les necessitats i paràmetres que marca el mercat laboral.

### 4 *Conclusió*

La varietat de classificacions, definicions i interpretacions que es troba en la lingüística computacional, lluny de crear confusió, por i desconfiança, no fan més que reafirmar aquesta disciplina en el camp de la lingüística general. De la mateixa

manera que no hi ha una única classificació, definició i interpretació en el camp de la traducció, també la traducció computeritzada disposa d'un gran ventall de taxonomies, mètodes classificatoris i terminologia.

La interdisciplinarietat de la lingüística computacional no hauria de representar un obstacle vers l'objectiu final de consagració d'aquesta disciplina dins el món de la lingüística sinó, ben al contrari, hauria de veure's com una realitat verificada des de diversos punts de vista. En aquest sentit, i per posar punt final a aquesta exposició, crec molt encertades les paraules d'Allen B. Tucker, qui ja preveia el futur de la traducció l'any 1984 en el seu article *Estratègies actuals en les investigacions i desenvolupament de la Traducció Automàtica*:

"La traducció automàtica és realment una de les matèries d'investigació més profundes i intrínsecament interdisciplinàries en la història de la recerca científica. I no s'arribarà a una solució eficaç fins que experts en diversos camps (lingüística, enginyeria de programes, intel·ligència artificial i psicologia) puguin fusionar la seva creativitat per tal d'aconseguir aquest objectiu comú."  
(Nirenburg, ed., 1987:41)

Mariona Sabaté i Carrové  
Dpt. d'Anglès i Lingüística  
Universitat de Lleida  
Apartat de Correus 471  
E-25080 Lleida  
e-mail: mariona@eup.udl.es

#### BIBLIOGRAFIA

- HEWSON, L.; MARTIN, J. (1991): **Redefining Translation, the Variational Approach**, London:Routledge.
- HUTCHINS, W.J.; SOMERS, H.L. (1992): **An Introduction to Machine Translation**, London:Academic Press.
- LEHRBERGER, J.; BOURBEAU, L. (1988): **Machine Translation**, Amsterdam/Philadelphia:John Benjamins.
- NEUBERT, A.; SHREVE, G. (1993): **Translation as Text**, Ohio, USA:Kent State University Press.
- NIRENBURG, S., ed. (1987): **Machine Translation: Theoretical and Methodological Issues**, Cambridge:Cambridge University Press.
- ROACH, P., ed. (1992): **Computing in Linguistics and Phonetics. Introductory Readings**, London:Academic Press Limited.
- RUIPÉREZ, G. (1995): **Enseñanza de Lenguas y Traducción con Ordenadores**, Madrid:Ediciones Pedagógicas.
- SAGER, J.C. (1994): **Language Engineering and Translation. Consequences of Automation**, vol. I, Amsterdam/Philadelphia:John Benjamins.

TIRKKONEN-CONDIT, S., ed. (1991): **Empirical Research in Translation and Intercultural Studies**, Tübingen:Gunter Narr.

### RESUM

Tant la traducció assistida per ordinador (TAO) com la traducció automàtica (TA) són considerades avui dia interdisciplinàries, i les associem amb camps com els estudis informàtics, la intel·ligència artificial i els llenguatges artificials, els subllenguatges, la pedagogia i l'aprenentatge de llengües estrangeres, entre d'altres. Aquesta interdisciplinarietat és l'origen de molts problemes, com és ara la consideració que té com a subdisciplina dins la comunitat lingüística. De la mateixa manera que la traducció no ha rebut fins fa ben poc el reconeixement i respecte de què altres professions han gaudit, la TAO i la TA estan encara patint les conseqüències d'aquesta actitud negativa. L'aportació que l'autora fa amb aquesta publicació té l'objectiu de recolzar aquestes dues disciplines i validar-les de la mateixa manera que la traducció ho ha fet mitjançant l'anàlisi de les tendències futures de la traducció computeritzada i veure com aquesta anàlisi pot fer veure aquesta disciplina en una nova perspectiva.

### SUMMARY

#### **Computer assisted translation vs. machine translation.**

The status of computer assisted translation (CAT) and machine translation (MT) is today considered interdisciplinary when related to computer studies, artificial intelligence and languages, sublanguages, pedagogy and foreign language learning among others. This interdisciplinarity is a source of many problems such as its consideration as a subdiscipline by the linguistic community. Since translation has not enjoyed the acknowledgement and respect that other professions have, CAT and MT are suffering now the consequences of such an attitude. My contribution to this publication is aimed at boosting MT and CAT just as translation has by analysing the future trends of computerised translation and how this analysis can show the discipline in a new light.