

Aplicații informatice de asistare a traducătorilor

Sanda Cherata
Universitatea Babeș-Bolyai Cluj-Napoca
Facultatea de Litere

1. Introducere

Rolul traducătorilor în societatea modernă este în continuă creștere. Dacă în trecut se ocupau cu precădere cu traduceri literare și de documente legale, în momentul de față ei sunt confrunțați cu sarcini mult mai complexe.

În procesul globalizării la care asistăm în prezent, transferurile internaționale (atât în ce privește importurile, cât și exporturile) generează, pe lângă schimbul de produse, procese, tehnologii etc. și schimb de documentație. Din ce în ce mai mult, produsele vândute într-o altă țară decât cea de origine sunt **localizate**, adică adaptate la specificul țării importatoare; aceasta nu se rezumă numai la traducerea și adaptarea documentației în limba țării primitoare, ci implică și aspectele culturale și de conținut, adică adaptarea aspectelor non-textuale ale produsului (culori, simboluri, unități de măsură etc.). **LISA** (The Localization Industries Standard Association) definește localizarea ca fiind procesul de modificare a produselor și/sau serviciilor în scopul de a răspunde diferențelor existente în diferite piețe. Un exemplu edificator este localizarea aplicațiilor informatice (sisteme-program, pagini web etc.). Localizarea lor înseamnă, pe lângă traducerea documentației (manuale, ghiduri de utilizare etc.), traducerea interfeței cu utilizatorul în limba țării importatoare (meniuri, casete de dialog, ajutor etc.), adaptarea setului de caractere, a unităților de măsură etc.

Dinamica schimburilor care au loc în prezent a generat o adevărată industrie a localizării. Aceasta cuprinde trei aspecte principale: vânzarea produselor, oferta de servicii de localizare și dezvoltarea instrumentelor care servesc localizării. Industria localizării este o parte a ceea ce constituie în momentul de față industria limbilor.

Industria limbilor cuprinde¹:

- a) scrierea documentelor tehnice (prin opoziție cu scrierile literare);
- b) managementul documentelor (care include utilizarea procesoarelor de texte, indexarea automată, regăsirea automată a informației etc.);
- c) managementul terminologiei;
- d) localizarea (care include traducerea);
- e) publicarea (care include, pe lângă publicarea tradițională pe suport hârtie, și publicarea pe suport electronic și pe Internet).

Cele cinci aspecte ale industriei limbilor nu sunt independente, ci se întrepătrund. De exemplu, managementul terminologiei multilingve joacă un rol foarte important în documentele domeniilor de specialitate. Dacă într-o traducere literară varietatea de stil este de dorit, atunci când e vorba de un termen dintr-un limbaj specializat, acesta trebuie tradus consecvent cu termenul echivalent corespunzător. *Publicarea* documentelor implică existența unui sistem de *management al documentelor*, care, la rândul lui, să fie aplicat în procesul de *localizare*, astfel încât aceeași terminologie să fie utilizată de la producerea inițială a documentului (în limba sursă) în toate traducerile documentului.

¹ Cf. Melby 2000

Mai mult decât atât, pentru produsele și tehnologiile scoase pe piață apar îmbunătățiri și modificări (în cazul sistemelor informatice, versiuni noi), iar documentația aferentă trebuie să reflecte această dinamică. Rezultă că apar documente cu conținut înrudit, care trebuie traduse într-un timp cât mai scurt. În aceste situații, e de dorit ca tot ceea ce a fost tradus pentru o versiune anterioară a unui document să poată fi folosit la traducerea versiunii actualizate; acest lucru se impune nu numai din rațiuni de creștere a eficienței muncii traducătorului, ci și pentru a asigura consecvența terminologiei din documentele respective.

Fiind vorba, deci, de o adevărată industrie a limbilor, se impune ca ea să fie însoțită și de tehnologiile corespunzătoare. Au apărut instrumente informatice (sisteme-program complexe) care asistă traducătorul pe tot parcursul procesului de traducere, astfel încât, în momentul de față, se vorbește de **traducerea asistată de calculator** (în engleză, **Computer-Assisted Translation – CAT**).

2. Categoriile de instrumentele pentru asistarea traducătorilor

Traducerea asistată de calculator este o procesul în care traducătorul uman traduce textele folosind instrumente informatice concepute să-l ajute și să-i faciliteze activitatea.

Termenul de traducere asistată de calculator este destul de larg, deoarece instrumentele informatice de asistare a traducătorului, pe care le implică, sunt dintre cele mai variate. Astfel, ele cuprind verificatoare ortografice și gramaticale, sisteme de management al terminologiei, dicționare uni- sau multilingve în format electronic (pe CD sau consultabile on-line), baze de date terminologice uni- sau multilingve (pe CD sau consultabile pe Internet), instrumente de indexare automată a documentelor și de regăsire a informației, concordanțe (care permit găsirea tuturor ocurențelor unui cuvânt dat, într-un text sau un corpus de texte, împreună cu contextele lor de utilizare), memorii de traducere (care memorează segmentele dintr-un text scris într-o limbă sursă împreună cu traducerea segmentelor respective într-o limbă țintă și reutilizarea acestor traduceri), sistemele de traducere automată.

Aceste instrumente pot fi grupate în trei categorii mari:

- a) aplicații de traducere automată [en: *Machine Translation (MT)*]
- b) sisteme de management al terminologiei [en: *Terminology Management Systems*]
- c) aplicații de asistare a traducătorului uman [en: *Machine Assisted Human Translation (MAHT)*]

Când s-a pus problema utilizării calculatoarelor în procesul de traducere, primele încercări care s-au făcut (în anii '60) au fost acelea de a crea programe care să realizeze traducerea automată a unui text din limba sursă în limba țintă. Desigur, la vremea respectivă nici echipamentele de calcul, nici sistemele de operare și limbajele de programare, nici teoriile lingvistice nu erau suficient de dezvoltate pentru a realiza o aplicație de asemenea complexitate, astfel încât aceste tentative au eșuat. S-a dovedit că sarcina realizării unor sisteme de traducere automată este mult prea complexă pentru a produce aplicații eficiente. Din păcate, nici astăzi, când există astfel de sisteme, nu se poate spune că ele răspund pe deplin cerințelor traducătorilor.

De aceea, s-a căutat să se implementeze aplicații de complexitate mai redusă, dar care să ajute efectiv traducătorul să realizeze o traducere de înaltă calitate și cu o eficiență maximă. Astfel, s-a ajuns la crearea aplicațiilor de asistare a traducătorului uman, care sunt folosite din ce în ce mai mult de traducători. Acesta

este motivul pentru care în prezentare ne vom referi pe larg la sistemele din această categorie.

2.1 Sisteme de traducere automată

Sistemele de traducere automată sunt aplicații complexe în care rolul central în procesul de traducere revine sistemului de calcul. Textul sursă este tradus automat în limba țintă de către sistem.

Procesul de traducere poate fi schițat în felul următor:

- Se face o analiză morfo-lexicală, sintactică și semantică a textului sursă, obținându-se o reprezentare internă a fiecărei propoziții sau fraze, de obicei sub formă de arbore sintactic, cu noduri care conțin informații de natură lexicală, sintactică și semantică; acest proces se numește *analiză*.
- Se face apoi un *transfer* de la forma de reprezentare internă a textului sursă în forma de reprezentare internă corespunzătoare limbii țintă.
- Ultima etapă este realizarea substituției lexicale, în care se face înlocuirea (în noua reprezentare internă obținută) a cuvintelor sau expresiilor cu cele corespunzătoare limbii țintă și generarea textului în limba țintă; acest proces este numit *generare* sau *sinteză*.

Procesul de traducere descris astfel pare simplu, dar în realitate necesită teorii lingvistice care să permită descrierea gramaticii limbilor, sisteme de calcul cu resurse suficient de puternice și elaborarea aplicațiilor care să implementeze algoritmi de analiză, transfer și generare. Pentru realizarea aplicațiilor de traducere automată trebuie să conlucreze lingviști, informaticieni și matematicieni.

Măsura în care este necesară intervenția umană în timpul procesului de traducere zisă „automată” variază de la un sistem la altul.

Astfel, există sisteme (puține, e drept, și care conțin multe restricții) care realizează traducere automată completă [en: *Fully Automatic Machine Translation*], fără nici o intervenție a traducătorului uman. Aceste aplicații sunt dedicate, de obicei, traducerii unor texte dintr-un domeniu specializat și foarte bine delimitat, în care lexicul este mai mult sau mai puțin restrâns, dar fără ambiguități, iar construcțiile sintactice sunt relativ simple. Un exemplu de aplicații din această categorie sunt sistemele de traducere automată a buletinelor meteorologice (cum este **METEO** utilizat de **Canadian Meteorological Office**² pentru traducerea buletinelor meteorologice între cele două limbi oficiale ale Canadei, engleza și franceza).

Majoritatea sistemelor de traducere automată necesită, însă, intervenția utilizatorului fie numai înainte și după traducerea automată, fie și în timpul acesteia. Aplicațiile de acest gen (care sunt cele mai răspândite) se numesc **aplicații de traducere automată asistată de utilizator** [en: *human-aided machine translation*). De obicei, asupra textului în limba sursă trebuie efectuată o prelucrare prealabilă a traducerii (o pre-editare) pentru a se asigura consistența lexicală și terminologică, în scopul de a se evita ambiguitățile, și pentru a se utiliza acele construcții sintactice care se pretează mai bine la prelucrarea automată. După obținerea textului tradus este necesară revizuirea traducerii și post-editarea textului pentru a se obține textul tradus în limba țintă. Unele sisteme din această categorie necesită intervenția interactivă a traducătorului uman, acestuia cerându-i-se, în timpul procesului de traducere, să selecteze un anumit termen sau cuvânt în caz de polisemie, să rezolve problemele stilistice sau să selecteze, dintre mai multe traduceri posibile ale unei sintagme sau fraze, pe cea corectă.

Există în momentul de față sisteme de traducere automată și traducere asistată de utilizator, unele dintre ele destul de performante: **METEO**, **Systran**, **PROJECTMT**,

² Cf. Dennett 1995

MuST, PeTra și altele. Din păcate, așa cum am mai menționat, ele nu sunt folosite pe scară foarte largă.

2.2 Aplicații de management terminologic

În cadrul unui proiect de traducere este absolut necesară gestiunea terminologiei. Se știe că un **termen** desemnează, într-o limbă dată, un **concept** dintr-un domeniu de cunoștințe. Termenii fac parte dintr-un limbaj specializat. De aceea, indiferent de natura textului pe care îl traduce, de câte ori întâlnește un termen, traducătorul trebuie să-l trateze ca atare. Mai mult decât atât, atunci când se traduc documente cu conținut înrudit, termenii trebuie utilizați cu consecvență. Managementul terminologic în procesul de traducere implică faptul că traducătorul trebuie să consulte dicționare de specialitate, baze de date terminologice și chiar să-și creeze propriile glosare și baze de date. În sprijinul acestor cerințe vin dicționarele și glosarele în format electronic, pe CD sau consultabile pe Internet, și diferite instrumente informatice pe care traducătorul le are la dispoziție, integrate în procesoarele de texte sau apelabile ca aplicații separate.

Trebuie știut că informațiile dintr-o bază de date terminologică diferă de cele dintr-un dicționar (general sau specializat). O intrare dintr-o bază terminologică conține informațiile referitoare la *un singur concept*; pentru fiecare termen se specifică, pe lângă informații gramaticale (clasa lexico-gramaticală, genul pentru substantive etc.), și informații de natură terminologică: definiția – cu sursă și, eventual, notă – contexte de utilizare, registrul termenului, statutul termenului, relațiile lui cu termenii care desemnează concepte super- sau subordonate, informații administrative etc.

Un traducător individual își poate crea propriile glosare și terminologii, fie folosind un sistem general de baze de date (**Access** sau **FoxPro**, de exemplu), fie un sistem dedicat bazelor de date terminologice (de exemplu, **MultiTerm**). În acest caz, el își poate defini propria structură a bazei terminologice. Într-un proiect de traducere, însă, la care participă mai mulți traducători și în cadrul căruia se traduc mai multe documente, este de dorit ca managementul terminologic să se facă în mod unitar, iar informațiile terminologice să poată fi îmbogățite și consultate de toată echipa. În acest scop trebuie utilizate mijloace informatice de tratare a bazelor de date terminologice fie elaborate de specialiști informaticieni, fie pachete de programe dedicate acestui scop (de exemplu, **LogiTerm**, **MultiTerm**, **Termex**).

Trebuie să menționăm că în unele sisteme integrate de asistare a traducătorilor există componente de **extragere a termenilor** dintr-un corpus [en: *term extraction*]. Acestea sunt foarte utile celor care se ocupă cu gestiunea terminologiei. Instrumentele din această categorie identifică și propun ca termeni-candidați sintagmele ocurente în corpus și oferă mijloace ușor de utilizat pentru ca informațiile terminologice să fie introduse într-o bază de date specificată. Utilizatorul decide care dintre candidați constituie termeni și îi poate introduce în baza de date, împreună cu informații utile (definiție, contexte de utilizare, sinonime și forme prescurtate etc.). Instrumentele din această categorie cresc foarte mult eficiența muncii de gestiune terminologică, mai ales în cazul corpusurilor mari, deoarece prin extragerea asistată de calculator a termenilor informațiile terminologice pot fi obținute rapid și pot fi puse la dispoziția echipei de traducători.

2.3 Aplicații de asistare a traducătorului uman

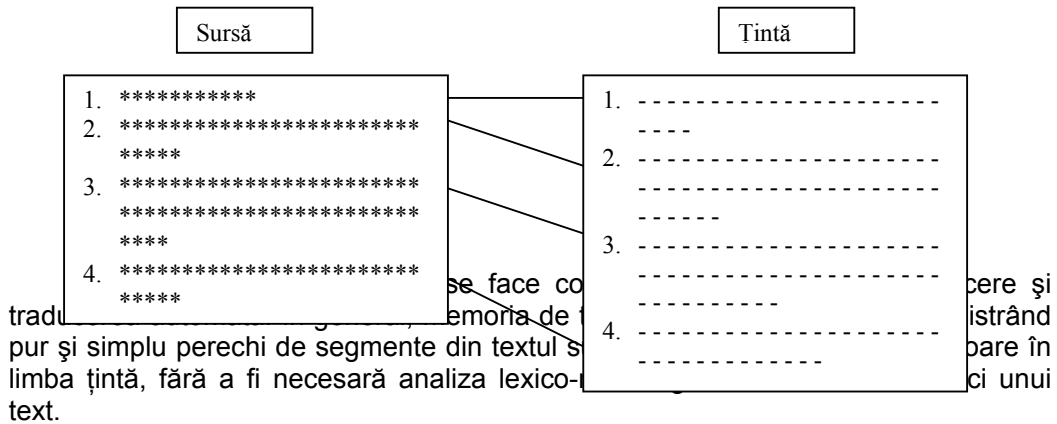
Deoarece, așa cum am arătat, traducerea automată nu a reușit să răspundă exigențelor traducătorilor, s-au creat aplicații mai puțin sofisticate care să-i asiste pe aceștia pe tot parcursul procesului de traducere.

Aplicațiile cele mai importante din această categorie sunt cele care se bazează pe principiul **memoriilor de traducere**.

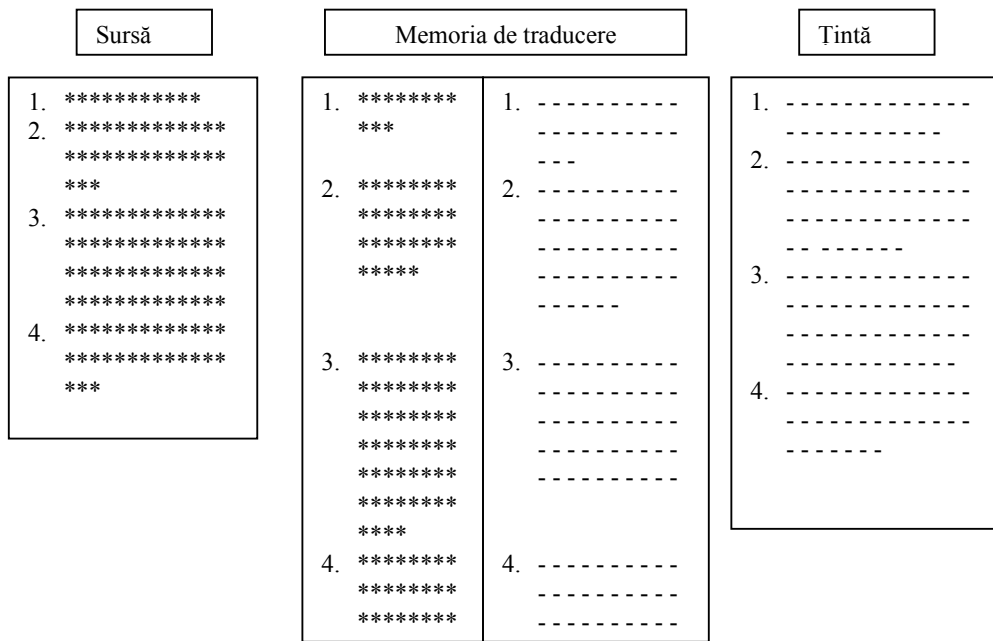
În esență, o memorie de traducere este o bază de date formată din perechi de segmente de text; o astfel de pereche este formată dintr-o unitate de text în limba sursă și unitatea corespunzătoare tradusă în limba țintă. Memoria de traducere este creată fie de un traducător uman, fie prin aplicații informatice de aliniere a textelor. Ulterior, dacă se folosește memoria de traducere la traducerea unui document cu conținut înrudit, pentru toate segmentele identice [*exact matching*] sau care sunt foarte apropiate [*fuzzy matching*] cu cele găsite în memorie se furnizează traducerea existentă în memorie, pe care utilizatorul o poate accepta sau modifica după dorință. Un segment poate fi constituit dintr-o propoziție sau frază, dintr-un paragraf sau poate fi stabilit de utilizator.

Expert Advisory Group on Language Engineering Standards (EAGLES) definește memoria de traducere ca fiind „o arhivă de texte multilingve care conține texte (segmentate, aliniate, analizate și clasificate) și care permite memorarea și regăsirea segmentelor de texte multilingve aliniate, conform unor criterii de căutare variate”.

Două texte **aliniate** se obțin luând segmentele corespunzătoare dintr-un text scris într-o limbă sursă și traducerea acelui text într-o limbă țintă:



O memorie de traducere poate fi schițată grafic astfel:



Pe lângă perechile de segmente de texte, într-o memorie de traducere se memorează și așa-numite **atribute interne**, cum sunt: data creării, numele utilizatorului sau al creatorului, clientul, identificatorul proiectului, domeniul căruia îi aparține documentul etc. Aceste atribute pot fi folosite ulterior la filtrarea segmentelor traduse sau a grupelor de documente.

De multe ori, estimarea efortului depus pentru realizarea unei traduceri nu este o sarcină ușoară; un sistem bun de memorii de traducere oferă și posibilitatea de a **analiza** un text nou, pentru a vedea măsura în care conține fragmente repetabile sau pentru a-l compara cu documente deja traduse și a lua decizia dacă să se folosească o memorie de traducere sau nu.

După natura textelor pe care le procesează, astfel de aplicații se pot clasifica în:

- a) aplicații orientate pe texte (sau paragrafe), care prelucrează documente create cu editoare de texte (Word, WordPerfect etc.) în format document sau format text îmbogățit (Rich Text Format);
- b) aplicații de localizare, care prelucrează texte scrise într-un limbaj cu marcaje (XML, HTML etc.).

În general, un sistem de memorii de traducere constă dintr-un **procesor de texte**, o fereastră de afișare a conținutului **memoriei de traducere** și, opțional, dintr-o **aplicație de gestiune terminologică**.

Desigur, memoria de traducere este un instrument util oricărui traducător. Chiar dacă efectuează o traducere individuală, un astfel de sistem îi permite să gestioneze terminologia, să găsească traduceri fragmentelor de text din textul sursă și să reutilizeze memoria de traducere și baza terminologică pentru lucrări ulterioare.

Sunt însă situații³ când utilizarea memoriilor de traducere crește substanțial eficiența activității:

- a) modificări ale textului de tradus făcute de client chiar în timpul procesului de traducere;
- b) revizii ale textului sursă făcute de client, care constituie obiectul unui nou proiect de traducere;
- c) „reciclarea” unei munci deja efectuate – în cazul când translatorul primește spre traducere un text foarte asemănător cu un text deja tradus.

Categoriile de documente care se pretează la utilizarea memoriilor de traducere sunt cele care au un conținut repetitiv. În Web 1998 se arată că majoritatea textelor de tradus aparțin uneia din următoarele domenii:

- Corespondență
- Jurnalism/Comunicare
- Afaceri/Comerț
- Marketing
- Publicitate
- Administrație
- Juridic/Legal
- Științific
- Tehnic
- Cultură
- Literatură

unde prin ➤ s-au marcat domeniile care se pretează cel mai bine la utilizarea memoriilor de traducere. Tipurile de texte pentru care memoriile de traducere cresc spectaculos eficiența activității sunt:

³ Cf. Web 1998

- brevete (Juridic/Legal)
- contracte (Juridic/Legal, Afaceri/Comerț)
- manuale de utilizare (Tehnic)
- rapoarte (Afaceri/Comerț)

De asemenea, utilizarea memoriei de traducere este mai eficientă când se lucrează în echipă.

Dintre cele mai cunoscute sisteme de memorii de traducere sunt: **Deja Vu** (Atril); **MetaTaxis**; **MultiTrans TransCorpora** (MultiCorpora); **ProMemoria**; **Sakhr Computer Aided Translation System**; **SDLX** (SDL International); **TransIt** (Star); **Translator's WorkBench** (Trados); **Translation Manager** (IBM); **TransSearch** (University of Montreal); **Trans Suite 2000**; **WordFast**.

3. Concluzii

În momentul de față traducătorii profesioniști sunt puși în fața unor exigențe din ce în ce mai mari. Pentru a răspunde cerințelor de profesionalism în procesul de traducere, ei sunt puși în situația de a utiliza toate mijloacele informatice care le pot veni în ajutor.

Traducătorii trebuie să cunoască, în primul rând, gama de instrumentele de asistare a traducătorilor (în continuă creștere și diversificare), apoi, pentru fiecare dintre acestea, care sunt situațiile în care utilizarea lor conduce la maxima eficiență și, în consecință, să recurgă la acele instrumente care sunt cele mai adecvate proiectului concret de traducere.

Referințe bibliografice

*** The Free Dictionary by Farlex: <http://encyclopedia.thefreedictionary.com/>

*** **Trados Tm 6.5 Training Guide**

TRADOS www.Trados.com

Dennett, Gerald 1995 **Translation Memory: Concept, products, impact and prospects**, Project Report, MSc ISE Part-Time Year 2, South Bank University, School of Electrical, Electronic and Information Engineering, London, 1995

Melby, Alan K. & Sue Ellen Wright. 1999. **Leveraging Terminological Data for Use in Conjunction with Lexicographical Resources**. (A paper delivered at the TKE-99 conference). Via Internet: <http://www.ttt.org/TKE-99.pdf>

Melby, Alan K. 2000 "**Globalization and the Need for International Language Standards**." Via Internet: http://www.ttt.org/g11n_xml.zip

Webb, Lynn E. 1998 **Advantages And Disadvantages Of Translation Memory: A Cost/Benefit Analysis**, Monterey Institute of International Studies, Monterey, California, 1998-2001, <http://www.webbsnet.com/translation/thesis.html>