

L'automatització de les traduccions al Departament de Justícia

Jordi Graells
Responsable dels serveis lingüístics
del Departament de Justícia de la Generalitat de Catalunya

1. Introducció

Al Departament de Justícia de la Generalitat de Catalunya ens trobem immersos en un procés d'automatització de les traduccions del nostre àmbit corporatiu. La mecanització d'aquest tipus de feines es fa amb estacions de treball ubicades en xarxa (sistema de traducció automàtica de Sail Labs) i amb d'altres situades localment en el PC (com el programa de traducció ARA¹ o el de traducció assistida amb memòries de traducció Déjà Vu).

En el cas concret d'aquesta intervenció, em centraré a parlar bàsicament del sistema de traducció automàtica de l'empresa Sail Labs amb el qual donem servei a determinats jutjats de Catalunya des de la xarxa informàtica judicial.

2. La traducció automàtica al Departament de Justícia

El 9 d'octubre de 2000 el Departament de Justícia va iniciar el Pla pilot de funcionament en català de quaranta oficines judicials a Catalunya amb l'objectiu de fer avançar l'ús del català en aquest àmbit. Calia demostrar que els jutjats podien treballar en català sense dilacions ni provocar indefensió a les parts per desconeixement de la llengua (les últimes dades sociolingüístiques mostren que un percentatge molt elevat dels residents a Catalunya entén el català).²

En el marc del Pla s'ofereixen eines als jutjats per treballar en català. D'entre aquests mitjans, destaca el sistema de traducció automàtica de Sail Labs, els drets de propietat del qual van ser

1. Jordi GRAELLS i Núria VIVES, «Els serveis lingüístics», *Administració, societat, llengua i Internet*, Barcelona, Escola d'Administració Pública de Catalunya, 2001, col·lecció «Papers de Recerca», núm. 7, p. 50-51.

2. Rosa LIZANDRA, «El pla pilot de funcionament en català de 40 oficines judicials», sessió de formació de la CILAJ (Comissió Interuniversitària de Llenguatge Administratiu i Jurídic) a la Universitat de Barcelona (maig de 2001).

adquirits per la Generalitat de Catalunya a l'empresa Incyta (actualment Sail Labs) per poder proveir d'aquest servei tots els seus departaments.³

El traductor automàtic ha acabat resultant el recurs més útil. Com la resta de programes de traducció automàtica, té en la qualitat final del text la seva principal limitació. Aquesta qualitat del text es pot incrementar al llarg de les diferents fases del procés de traducció automàtica:⁴

a) La preedició. El sistema treballa analitzant sintàcticament frase per frase. Per això, és important que el text original que s'hagi de traduir sigui acurat, sense errors de transcripció, amb frases curtes i ben estructurades.

b) La postedició. El lingüista edita el text que surt del traductor corregint errades i problemes d'interpretació del sistema.

L'entorn de la preedició i la postedició funciona sobre el processador de textos Microsoft Word.

3. Com funciona el sistema Sail Labs

El sistema funciona amb un servidor situat a la xarxa informàtica judicial. Les oficines judicials envien el document a traduir. Un cop traduït, abans de tornar-lo a l'oficina corresponent, el servidor el trameta a un lingüista perquè el corregeixi.

Aquest sistema de traducció es val d'informació lingüística d'alt nivell. A diferència dels sistemes anomenats *de traducció directa o de paraula per paraula*, es basa en l'anàlisi i la generació lingüística. Per traduir en cada parell de llengües, el sistema requereix tres mòduls:

a) El d'anàlisi. En el traductor castellà-català, és el diccionari monolingüe castellà que conté informació morfològica, sintàctica i semàntica de les paraules en castellà.

Per exemple: Si volem traduir al català la forma *Yo disfruto*, el sistema la reconeix com la primera persona del present d'indicatiu del verb *disfrutar*.

b) El de transferència (o bilingüe). S'hi emmagatzemen les traduccions de cada paraula, amb les condicions de selecció (si hi ha més d'una possibilitat de traducció) i les accions associades (transformacions morfosintàctiques necessàries). En l'exemple, identifica que la forma catalana de l'acció *disfrutar* és *gaudir*.

c) El de generació. En aquest cas, és el diccionari monolingüe de català i indica que la primera persona del singular del present d'indicatiu del verb *gaudir* és *jo gaudeixo*.

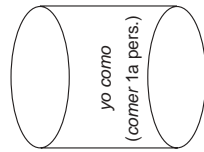
Si el procés falla en el mòdul d'anàlisi, el sistema ens dona com a desconeguda la forma *yo disfruto*. Si l'errada té lloc en el mòdul de transferència, només ens indicarà que la forma

3. DIRECCIÓ GENERAL DE POLÍTICA LINGÜÍSTICA, «3.2. La traducció automàtica i assistida», a: *Informe anual de política lingüística 2000 (XII Noves Tecnologies i Internet)* [en línia] <<http://cultura.gencat.es/llengcat/informe/cap12%2000.pdf>>.

4. Rosa PÉREZ i Judit RODÀ, «La traducció automàtica i l'experiència del BOE en català», *Llengua i Ús*, Direcció General de Política Lingüística, núm. 21 (segon quadrimestre de 2001).

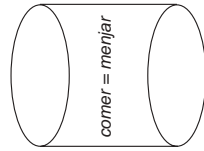
Sistema de traducció automàtica Sail LABS-INCYTA

ANÀLISI



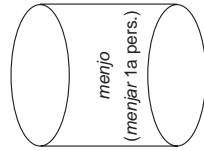
LÈXIC MONOLINGÜE CASTELLÀ
Si error → yo como com a forma desconeguda

TRANSFERÈNCIA



LÈXIC MONOLINGÜE CASTELLÀ
Si error → ens indica que la forma yo como té per infinitiu comer

GENERACIÓ



LÈXIC MONOLINGÜE CASTELLÀ
Si error → per a la forma yo como ens ofereix l'infinitiu menjar però sense conjuguar-lo

1. Analtzador morfològic (AM)

Per a cada paraula s'obtenen un conjunt d'interpretacions morfològiques (lema + informació morfològica deduïda de la forma).

2. Analtzador sintàctic (parser)

Agafen cada sortida de l'AM + informació del lèxic monolingüe + dades de la gramàtica d'anàlisi = arbre d'anàlisi sintàctica.

1. Condicions de selecció

Més d'una possibilitat de traducció.
P. ex. *calle*.

2. Accions associades

Canvis de gènere/ nombre o canvis estructurals.
P. ex. *dinero - diners*, *estornudar - fer un esternut*.

A partir de l'arbre d'anàlisi amb les paraules traduïdes al català i amb algunes modificacions estructurals ocasionades pel lèxic bilingüe, s'activen les tasques per generar la frase final en la llengua de destinació, en aquest cas el català.

yo disfruto té per infinitiu *disfrutar*. En canvi, si el problema es produeix al mòdul de generació del català, per a la forma *yo disfruto* ens oferirà la forma infinitiva *gaudir* sense fer-ne la conjugació.

L'avantatge d'aquest sistema de traducció és que els mòduls es poden tornar a aplicar per a la traducció d'altres parelles de llengües. Així, per exemple, el mòdul de generació del català del parell de llengües castellà-català es va poder usar per al parell de llengües anglès-català.

4. Fases de la traducció automàtica

4.1. Fase de la preedició

Com hem dit, correspon a la preparació del text original donant-li el màxim de qualitat en qüestions lingüístiques, gràfiques i textuals per optimitzar el procés de traducció.

La fase de la preedició la duen a terme directament les oficines judicials emissores del document. Per facilitar-los aquesta tasca, es col·loquen filtres lingüístics al servidor i s'elaboren macros d'Access i Word específiques que poleixin el text original en algunes qüestions bàsiques (com el format del document i la tipografia). També es recomana activar els verificadors ortogràfic i gramatical de l'idioma de partida.

4.2. Fase de la traducció automàtica

S'envia el document al servidor perquè el retorni traduït, ja sigui per mitjà d'un formulari HTML o mitjançant l'aplicació informàtica específica de l'àmbit (Temis 2).

El document resultant el rep sempre un lingüista per fer-ne la postedició.

4.3. Fase de la postedició

Com ja hem dit, l'entorn de la preedició i la postedició funcionen sobre el programa Word. La postedició disposa, a més, d'una barra d'eines específica. En aquesta barra d'eines, mitjançant el botó *Processa marques*, s'hi apliquen unes macros que visualitzem amb claus, barres i diferents colors per indicar-nos els casos d'homografia, de traducció de noms propis i, en general, on s'ha produït el problema en la traducció, si és en el mòdul d'anàlisi del castellà, en el de transcripció o bé en el de generació del català:

- enjudiciament{judici}, esmentat{citat}, testimoni{testimoniatge/testimoniança}
- decretamos (cast.) → vam decretar/decretem (cat.)
- L.E.C., D.T. → (en l'anàlisi del castellà no reconeix abreviacions amb punts)

Un cop posteditat el text traduït, s'han d'eliminar aquests senyals (colors, claus, etc.). En la postedició també es poden aplicar macros que corregeixen de manera automatitzada errors freqüents detectats en els textos traduïts pel sistema.

5. Una aproximació al model de llengua resultant

Com hem vist, la versió final del text català traduït amb aquest sistema dependrà principalment d'aquestes condicions:

- a) Grau de correcció del text original en castellà.
- b) Aptitud del sistema per traduir i interpretar el text castellà.
- c) Capacitat i temps disponible de la persona que el postedita.

El resultat de la traducció amb aquest sistema és sens dubte positiu. Els textos judicials en la versió catalana esdevenen correctes ortogràficament i gramaticalment. Ara bé, com que no han estat pensats originàriament en aquesta llengua, presenten en general problemes morfosintàctics:

- a) Pel que fa a la frase, l'ordre dels elements es presenta sovint alterat.
- b) Pel que fa al nom, hi ha dificultats per refer la relació anafòrica amb els elements ja apareguts en el text, alhora que s'usen nominalitzacions injustificades.
- c) Pel que fa al verb, s'empren construccions passives en comptes d'haver-hi una preferència per la veu activa (perifràstiques, falses passives pronominals, participis passius), construccions perifràstiques innecessàries (inclou, per exemple, les construccions castellanes que no funcionen en català *venir a + infinitiu* i *ser d'aplicació*), gerundis incorrectes (especificatius, finals i de posterioritat) i futurs d'obligació. Actualment l'empresa proveïdora Sail Labs estudia diverses possibles solucions per aplicar-les a les construccions passives i a les de futur d'obligació.

De les intervencions sobre altres qüestions textuais (puntuació i cohesió, estructuració del contingut, precisió i variació del llenguatge) i gràfiques (abreviacions, majúscules i minúscules, etc.), algunes es poden automatitzar, mentre que unes altres (com ara la puntuació o l'estructuració del text) han de ser totalment manuals.

Passem tot seguit a observar qüestions estrictament lingüístiques en aquest sistema de traducció.

5.1. En els diccionaris

El sistema conté terminologia jurídica i administrativa elaborada pel Centre de Terminologia TERMCAT, per la Comissió Assessora de Llenguatge Administratiu (CALA) i pel Departament de Justícia. Tot i amb això, de vegades el sistema no arriba a traduir correctament la paraula o expressió equivalent en català; alguns exemples:

- Tradueix *por falta de* (castellà) per *a falta de* (català).
- Tradueix sistemàticament el *don* (castellà) per *en* (català).
- Tradueix erròniament algunes locucions llatines (tradueix *res nullia* per *cap de bestiar nullia*), abreviacions (LEC), dates, etc.

En els noms propis de persona i topònims, podem trobar casos de traducció no desitjada i amb provada concordança com *Llabor Esquerra* per *Pepita Izquierdo* o *Jordi Amb oliva* N'hi hagi prou per *Jordi Baste Oliva*.

5.2. En l'anàlisi sintàctica

Hi ha errades del sistema causades normalment per una anàlisi incompleta o inexacta de la frase del text en castellà. En són exemples:

- Traducció de participis. El sistema tradueix de vegades el participi *pendiente* per *arracada*.
- Concordança errònia entre substantiu i adjectiu quan el substantiu és femení (tradueix *su capacidad* per *el seu capacitat*).
- Relatiu incorrecte. El sistema habitualment ofereix la forma del masculí singular (*les acusades els advocats del qual*).
- Reproducció de l'ordre dels elements de la frase en castellà.
- Calcs sintàctics castellans (*a la major brevetat possible, en tal intel·ligència*).
- Col·locació sempre de la preposició *a* davant de l'objecte directe en català (*informa als advocats*).
- Règims preposicionals erronis de verbs catalans (*recórrer [contra] la sentència, haurà de traslladar-se de la documentació* provinent de la perífrasi castellana *deberá darse traslado de*).
- El pronom castellà *le* només se'l fa correspondre amb el pronom *li* (objecte indirecte) i no pas també amb el d'objecte directe (*el*).
- Traducció errònia, incompleta o inexacta de formes verbals com les formes castellanen *carece de* (el tradueix per *té de*; aquesta traducció inadequada pot induir a errors greus en els documents jurídics), *se acuerda* (per *se'n recorda* en comptes de *s'acorda*), *fue* (només per *va anar* i no també per *va ser*), etc.
- Execució errònia en alguns casos de les contraccions *de+el* i *a+el* (*d'el procés*).
- Ús inadequat de les preposicions (*a Vïc, a vuit de..., documents que consten en actuacions*).

5.3. Altres casos

- Tradueix termes menys freqüents en l'àmbit judicial: *aixecar* per *alzar* en comptes d'*alçar* que és la que s'usa més.

— En alguns casos tradueix símbols i números erròniament, com e) per i) i el número romà VI per *vaig veure* (cast., *vi*).

Fins aquí hem enumerat algunes mancances lingüístiques que presenta el sistema i que de moment no s'han pogut esmenar. Malgrat aquestes mancances, volem destacar que es tracta d'un sistema apte per a les necessitats de traducció de la xarxa judicial. En un any, la versió del sistema Sail Labs s'ha anat adaptant a les especificitats de l'àmbit, gràcies a la col·laboració de l'empresa proveïdora amb els serveis lingüístics del Departament de Justícia.

6. Conclusions

1) Com acabem de dir, cal una estreta col·laboració entre el proveïdor tecnològic i l'usuari, en aquest cas, els serveis lingüístics de Justícia.

2) La qualitat dels textos traduïts pel sistema de traducció automàtica depèn, a més de la millora continuada del sistema, d'una bona execució de les fases de preedició i postedició i de les possibilitats de mecanitzar-les. S'ha d'invertir en formació per fer competitiu els recursos humans que s'hi dediquen.

3) El sistema tradueix des d'un sistema ubicat a la xarxa informàtica judicial. Aquesta circumstància permet establir un procés de traducció client-servidor i instal·lar directament filtres lingüístics al servidor.

4) Cal completar el procés d'automatització de les traduccions amb el reaprofitament dels textos resultants de la traducció automàtica per convertir-los en memòries de traducció, les quals filtrarien prèviament aquella part del text que ja s'hagués traduït anteriorment. Aquesta tècnica de traducció, anomenada traducció assistida amb memòries de traducció, també tradueix terminologia, fraseologia i altres expressions contingudes en la seva base de dades. Amb aquest plantejament es rendibilitzaria el cost de la traducció automatitzada, ja que només traduiríem automàticament la part del text realment necessària i situaríem la traducció automatitzada de documents en un flux de treball global.

5) El que caracteritza principalment el llenguatge jurídic és una morfosintaxi i una fraseologia específiques més que una terminologia pròpia. Per això, en aquest àmbit és tan important basar l'automatització de les traduccions en l'anàlisi morfosintàctica i en la fraseologia de l'àmbit.⁵

5. GRETEL (GRUP D'ESTUDIS DE TÈCNICA LEGISLATIVA), *La redacció de les lleis*, Barcelona, Escola d'Administració Pública de Catalunya, 1995, col·lecció «Manuals i Formularis», núm. 10.