

El programa de traducció d'*El Periódico*

Ricard Fité

Coordinador lingüístic de l'edició catalana
d'*El Periódico de Catalunya*

L'edició en català d'*El Periódico de Catalunya* ha complert quatre anys, i el que podem dir un cop transcorregut aquest període és que el programa d'ajuda a la traducció que fem servir ens va la mar de bé, ja que ens permet fer el diari cada dia i tancar-lo a l'hora, sense que el procés de producció hagi variat gens ni mica: el diari s'escriu en castellà, passa per un programa de traducció i un grup d'editors revisa el text resultant per donar-li la forma definitiva. I si bé és evident que el programa ens va més bé que fa quatre anys gràcies al manteniment que en fem nosaltres mateixos, també ho és que no ha estalviat ni un sol lloc de treball. És a dir que el pas del temps ha desmentit la idea que la postedició seria cada cop menys rellevant i menys necessària. No és veritat. Com a traductors, ens estalviem de picar el text, i prou. No ens estalviem de revisar-lo atenent a la forma i al sentit. Per això som trenta-set lingüistes en plantilla, i dotze més per fer substitucions al llarg de l'any i sobretot a l'estiu. Fins i tot som un equip més nombrós encara que fa quatre anys, perquè fem més productes en català, els suplementes que no fèiem al començament, o sigui més pàgines.

Com és el programa

El programa d'*El Periódico* és tan i tan senzill, que acaba convertint-se en peculiar. Diuen els experts en lingüística computacional, i en aquesta taula n'hi ha uns quants, entre els quals no m'incloc, que existeix la traducció automàtica, la traducció assistida i la traductora d'*El Periódico*, que sembla que no és ni una cosa ni l'altra. Sigui com sigui, aquest programa que ens permet fer el diari cada dia funciona per comparació, mai per anàlisi. No té regles d'anàlisi gramatical ni regles lèxiques. I això és bo o és dolent?, us preguntareu. Doncs, finalment, i contra tots els pronòstics que al seu moment es van formular, entre els quals ara sí que incloc el meu, és més bo que no sembla. El programa és, efectivament, molt beneït: davant de cada paraula-estímul reacciona sempre amb una mateixa paraula-resposta, sempre la mateixa, sense tenir en compte el context. No li demaneu intel·ligència morfològica ni sintàctica ni semàntica, només

coneix l'ortografia i el lèxic que volem que conegui. Per no saber, no sap si una paraula és masculina o femenina, ni si és singular o plural, ni si és un substantiu o un adjectiu o una forma verbal o un adverb o una preposició. (Posats a dir la veritat, l'únic que sap és com funcionen l'apòstrof i les contraccions. Són les úniques regles que té, i encara, perquè sap apostrofar mentre la regla catalana és regular, ja m'enteneu: perquè tant li fa si una paraula masculina comença amb vocal, amb semiconsonant o amb *hac* aspirada: la regla arrasa i ho apostrofa tot. I una *essa* és una *essa*: no pretengueu que distingeixi si és líquida o no. I com que sap que el famós *la* no s'apostrofa davant les *is* i les *us*, doncs no ho farà mai, per més tòniques que siguin aquestes *is* i aquestes *us*.) Però, precisament gràcies a aquesta simplicitat, en qualsevol moment del procés de producció es pot tenir accés directe i immediat als motius pels quals el programa ens ha servit un sintagma incongruent. Un disseny tan simple potencia la transparència dels resultats, i per tant fa molt més detectables les males traduccions. El programa té una base de dades amb més de 400.000 paraules en castellà, a cada una de les quals li correspon una paraula en català, i només una. Així de senzill.

Si el programa acabés aquí, els editors tindriem bastant més feina, i potser hauríem de ser més dels que som i tot per poder enllestir les pàgines a l'hora convinguda. Però no acaba aquí. Al costat de l'esmentada base de dades de paraules, el programa disposa d'una base de dades de seqüències en castellà, a cada una de les quals li correspon una traducció catalana. Gairebé 100.000 seqüències. I gràcies a aquesta base de dades de seqüències, que actua abans que la base de dades de paraules, el programa resulta més fiable. Amb un exemple es veurà prou clar. Segons la base de dades de paraules, *dos tarjetas* són *dos targetes*, perquè hi té *dos* = *dos* i *tarjetas* = *targetes*. Segons això, estariem condemnats a trobar-nos sempre *dos targetes*. L'única manera d'aconseguir que el programa faci bé la traducció de *dos tarjetas* (almenys en català central, que distingim entre *dos* i *dues*) és introduint a la base de dades de seqüències el sintagma *dos tarjetas* = *dues targetes*, i així successivament amb tots els substantius femenins comptables del diccionari de la llengua: *dos mujeres* = *dues dones*, *dos camisas* = *dues camises*, *dos sillas* = *dues cadires*, *dos guitarras* = *dues guitarres*, etcètera, etcètera, etcètera. Tres vegades etcètera. Perquè es vegi la magnitud del problema. Aquesta possibilitat d'introduir seqüències també ens soluciona els errors que comet la regla de l'apòstrof, i d'aquesta manera el programa acaba fent bé els casos no regulars: *el iogurt*, *de Hudson*, *l'staff* i *l'illa*.

Com es manté el programa

Ara sí que el programa s'acaba aquí. I aquí en comença el manteniment. El fa un dels trenta-set lingüistes esmentats, que no és pas informàtic. Aquest és un altre dels avantatges de l'eina que tenim: la seva simplicitat tècnica fa que el manteniment no requereixi coneixements informàtics, sinó simplement lingüístics i, això sí, una gran dosi de lògica, intuïció i sentit comú. Per exemple, una de les feines de manteniment consisteix, des de fa quatre anys, a anar intro-

duint a la base de dades seqüències com les que acabo d'esmentar, amb criteris tan pragmàtics com els que es deriven de la mateixa productivitat diària. Seguint amb el mateix exemple de *dos/dues*, davant el descoratjador panorama que suposa haver d'entrar a la base de dades totes les seqüències possibles de *dues* + substantiu femení segons el contingut del diccionari de la llengua, i aquest només és un cas entre milers, hem optat per introduir-hi només les seqüències d'aquesta mena que han aparegut alguna vegada en els textos del diari, i aquest constitueix un dels criteris amb què de mica en mica hem anat adaptant el programa a les nostres necessitats: partint dels errors que hem hagut d'esmenar cada dia.

L'experiència ens ha demostrat que és possible millorar la qualitat del programa a còpia d'anar combinant la informació que donem a la base de dades de paraules amb la que donem a la base de dades de seqüències. Es tracta d'anar provant solucions, i si veiem que no ens van bé, podem fer marxa enrere immediatament. Agafem, per exemple, un terme com *sosa*, substantiu i adjectiu. Per la base de dades de paraules resollem l'adjectiu, i li diem que *sosa* = *insípida*. Per la base de dades de seqüències resollem el substantiu, i li diem que *sosa cáustica* = *sosa càustica*, *sosa dura* = *sosa dura* i *sosa comercial* = *sosa comercial*. Fins al dia que el programa ha de traduir el sintagma *sosa líquida*, o bé *sosa en malas condiciones*, en què *sosa* és substantiu, i la base de dades no té aquestes seqüències. Què fa aleshores el programa? Tracta el mot *sosa* com si fos l'adjectiu i li clava *insípida líquida* i *insípida en males condicions* respectivament. Doncs tornem-hi: li diem a la base de seqüències que *sosa líquida* = *sosa líquida*, i *sosa en malas condiciones* = *sosa en malas condiciones*, i un altre dia ja ho farà bé. El problema és sempre el mateix: no és possible preveure-ho ni resoldre-ho tot, però sí anar tapant forats. Aquestes seqüències, unes més útils que altres, són com granets de sorra que contribueixen modestament a fer el programa més eficaç.

Un altre exemple seria el mot *médico*, substantiu i adjectiu alhora, amb traducció diferent segons sigui la categoria gramatical. Doncs en aquest cas el que ens surt més a compte és que la base de paraules tingui l'equivalent com a adjectiu, *médico* = *mèdic*, i per la base de seqüències resollem el substantiu: *un médico* = *un metge*, *el médico* = *el metge*, *varios médicos* = *uns quants metges*, etcètera. En realitat, per la base de seqüències podem resoldre tot allò que tingui una traducció invariable, al marge de categories gramaticals. Així, si en la base de dades de paraules hi ha *parte* = *part* i *médico* = *mèdic*, per evitar l'automàtic *part mèdic* caldrà introduir a la base de seqüències *parte médico* = *comunicat mèdic*, en què *mèdic* és adjectiu. És qüestió d'anar fent segons el cas.

A això ens dediquem. I això no vol dir que estiguem esperant que arribarà un moment que el programa ja no cometrà errors. Algú s'ho pensava al començament. Però ja no s'ho pensa. La postedicció és i serà sempre imprescindible, perquè hi ha errors evitables i errors inevitables. *Dos targetes* és un error que podem evitar per sempre, amb la introducció de la seqüència corresponent. En canvi, *aliento*, *tanto*, *alcanzar*, *esas*, *cuyo*, *le* i *lo*, per exemple, són paraules que la màquina no ens traduirà mai. Fixeu-vos que no he dit *no ens traduirà mai bé*, sinó *no ens traduirà mai*, perquè ho hem decidit així. La senzillesa del programa ens permet triar entre donar

un equivalent a una paraula sabent que no sempre serà la traducció correcta, o bé no donar-li cap equivalent per poder tenir la certesa que la màquina «no ha traduït». En el primer cas, l'editor del text corre el risc que li passi per alt una traducció que atempti contra el sentit i la coherència de la frase. En el segon cas, sap que el programa «no ha actuat», és a dir que ha deixat la paraula tal com se l'ha trobada, sense traduir, i que segurament haurà d'actuar ell. I ho sap, formalment, perquè li apareix entre angles i de color vermell. L'experiència d'aquests quatre anys ens ha portat a preferir cada cop més aquesta segona opció (és a dir, que el programa no tradueixi), perquè fa disminuir considerablement el percentatge d'errors que li poden passar per alt a l'editor, i per tant fa augmentar en la mateixa proporció la qualitat final de la traducció.

No cal dir, d'altra banda, que portar a terme el manteniment del programa amb aquest criteri va en contra de la idea simplista, però molt estesa entre alguns fabricants de programes traductors, segons la qual la qualitat de l'eina és directament proporcional a la quantitat de paraules que és capaç de traduir, capacitat que expressen sense rubor en tants per cent: «Tradueix el noranta-tres per cent del text original!», et diuen cofois, recalcant la xifra. No descobreixo res de nou si dic que la quantitat d'entrades de què disposa la base de dades és un dels grans arguments, si no el principal, que esgrimeix algun fabricant per persuadir el comprador de la fiabilitat del seu producte. Però tampoc descobreixo res de nou si dic que la qualitat d'un programa de traducció, sigui de la mena que sigui, automàtic, assistit o de color de catxumbo, no es pot avaluar segons la quantitat de paraules que és capaç de traduir, sinó segons la quantitat de paraules que és capaç de traduir bé. L'adverbi *bé* és fonamental perquè tots plegats ens puguem entendre sense enganyar-nos. Per això el manteniment del programa consisteix, com és lògic, a donar d'alta paraules i seqüències que no tenia, però també a donar-ne de baixa quan descobrim que no genera traduccions fiables. I això últim passa més sovint que no sembla.

Altres aspectes del manteniment són:

- 1) Donar informació al programa sobre què ha de fer quan es troba una paraula en majúscula. D'entrada, el programa no distingeix una majúscula d'una minúscula. No dic que no ho pugui distingir. Simplement, dic que no ho fa. Una decisió que vam haver de prendre al començament va ser si ens estimàvem més que no traduís cap paraula en majúscula o que les traduís totes. Caixa o faixa. Què és millor: tot o res. La intuïció ens va portar a preferir que ho traduís tot, pensant en la gran quantitat de paraules en majúscula que s'han de traduir, des dels topònims amb forma tradicional catalana fins als noms propis transparents d'institucions públiques, sense oblidar les transcripcions dels simpàtics antropònims russos i àrabs, que no tenen la culpa, pobrets!, de no utilitzar l'alfabet llatí. Però la servitud ha consistit a haver de donar al programa la informació necessària perquè un cognom d'alfabet llatí *traduïble* quedi salvat de la diguem-ne transformació gràcies a la majúscula. És a dir, *punte* en minúscula serà *pont*, però en majúscula continuarà sent *Punte*. El problema és que aquesta informació l'hi hem hagut de donar cas per cas, al programa, després d'haver pres la decisió esmentada. I això és feina. Una feina interminable, d'altra banda, si pensem que tot cognom és susceptible de confondre's amb un nom comú. I si pensem, també, que el programa no distingeix entre llengües i processa tota

la informació com si fos castellà. Per aquest motiu, l'actor nord-americà *Lee Marvin* es converteix els primers dies en *Llegeix Marvin*. I un conte en català com *La cigala i la formiga*, citat així en l'edició del diari en castellà, es transforma en *L'escamarlà i la formiga*. I si et passa per alt, cosa no gens difícil si tenim en compte la versemblança de la traducció, original al marge, ja has begut oli davant l'estimat lector, t'has cobert de glòria i has acabat fent un pa com unes hòsties. És a dir: el ridícul més espantós. Qui pensi, doncs, que la postedició és estalviable, ja sap a què s'exposa. Perquè el programa, diria que cap programa, no té el coneixement del món que sí que tenim les persones.

2) Donar informació al programa sobre alternatives de traducció quan no volem que tradueixi res. L'editor té la possibilitat de visualitzar el diccionari de la traductora i comprovar si una paraula o una seqüència són a les bases de dades o no, i amb quin equivalent. En el cas que no vulguem que tradueixi res, aquest diccionari permet contenir informació sobre possibles alternatives, no precisament sinònimes, o bé informació que justifiqui per què no es vol que una determinada paraula o seqüència siguin traduïdes. Per exemple, *pincho* apareix no traduït, en vermell, i si volem saber per què, el diccionari del programa ens diu que, si som en una taverna basca, n'hem de dir *pintxo*, a la basca, i que genèricament n'hem de dir *broqueta*. De tal manera que en qualsevol moment del procés d'edició del text podem saber per quin motiu la màquina ha traduït o ha deixat de traduir el que sigui, i quines solucions concretes ens ofereix davant el cas que ens ocupa.

Influència del programa en la llengua d'*El Periódico*

No hi ha dubte que la utilització d'un programa de traducció contribueix a homogeneïtzar el model de llengua que es fa servir, molt més que no pas un llibre d'estil. *El Periódico* mateix n'és un bon exemple, si es compara el procés d'elaboració del diari en castellà i en català. En castellà hi ha centenars de mans diferents, entre redactors, corresponsals i col·laboradors, i els editors responsables d'unificar-ne l'estil han de tenir en principi més feina, encara que la disciplina del redactor en relació amb les normes del llibre d'estil de la casa sigui elevada. En canvi, en català hi ha d'entrada un sol redactor, el programa, i un equip de postedició que en modifica els resultats pel que fa a la qualitat i la fidelitat de la traducció, però que en principi accepta per defecte el text estilísticament uniforme que li arriba de manera automàtica.

Amb exemples potser es veu més clar. *Dijo* serà sempre, per defecte, *va dir*. I a més a més, ja volem que sigui així. Llevat, és clar, dels articles d'opinió que ens arriben originalment escrits en català, que també n'hi ha, encara que no gaires, als quals no apliquem les normes del llibre d'estil de la casa. O sigui, que el primer que trobi un *digué* en una pàgina d'informació d'aquest diari té un sopar pagat pel senyor Ricard Fité. Des d'aquest punt de vista, el programa és una eina útil per aplicar amb coherència una tria estilística. I qui diu una tria estilística diu qualsevol criteri que es consideri oportú d'aplicar. Si el programa té *atrae* = *atrau*, ja us podeu imaginar que la

forma *atreu* no sortirà mai, entre altres raons perquè el traductor editor ja sap que si el programa li diu això és que s'ha triat aquesta forma en detriment de l'altra i que ja es vol que sigui així. És a dir que tornem allà mateix: fins i tot en un terreny fins a cert punt poc rellevant com és el de les variants morfològiques, un programa de traducció unifica molt més que no pas una redacció de dos-cents periodistes amb un equip d'editors correctors al darrere que li revisi la feina.

Al costat d'exemples com aquest últim, d'unificació necessària des del punt de vista de la varietat estilística que es vol difondre, en què un programa de traducció pot contribuir positivament a consolidar usos estàndards, també n'hi ha molts altres en què la simplificació proposada pel programa no només no estandarditza, sinó que fa augmentar el perill d'empobriment del llenguatge o bé, cosa encara pitjor, porta a un ús imprecís i inadequat. Seria el cas del verb castellà *elegir*, que el programa tradueix invariablement per *elegir* (amb totes les formes verbals incloses, naturalment). Aquí l'experiència ens demostra, per una banda, que en castellà *elegir* té més rendiment que *escoger*, i per tant que té una freqüència d'aparició més alta que el paral·lel català *elegir*, quan a més a més el català pot optar per *triar* com a sinònim perfecte d'*escollir*. És evident que en aquests casos el programa no ens fa cap bé, i s'ha de confiar en l'experiència del posteditor per compensar la simplificació.

Sigui com sigui, com en tot en la vida, fa de mal generalitzar a partir d'uns quants exemples fins a determinar amb claredat i contundència si el programa de traducció ens facilita l'estandardització o no. En alguns aspectes ens la facilita, i en altres no és que ens la compliqui, simplement no és bo que la faci. Del que no hi ha dubte és que un programa ajuda a unificar criteris i grafies, en la mesura que sigui fàcilment modificable per l'usuari com aquest que estem descrivint. I en tot cas, ens ajuda a estandarditzar el que volem estandarditzar. Vull dir que aquest programa ens fa el que volem que faci, a diferència d'altres no modificables o difícilment modificables per l'usuari particular. Així, per posar un altre exemple, si en l'edició en castellà constatem una vacil·lació entre *el burka* i *la burka* i volem que en català surti publicat *el burca*, en masculí, que és el que diu que hem de fer el Cercaterm dels amics del TERMCAT, n'hi ha prou d'introduir a la base de dades de seqüències *la burka = el burca*, i l'editor s'estalvia el dubte, el comentari en veu alta i la intervenció, tot plegat un temps preciós, indispensable per tancar l'edició a l'hora.

En conclusió, com he dit al començament, tenim un programa fantàstic que ens serveix per fer la feina que hem de fer, i que ja fa quatre anys que fem: acabar un diari en català cada dia, mitja hora més tard que en castellà. Sense programa no hi hauria diari. Ni trenta-set llocs de treball. Per tant, no podem sinó expressar satisfacció per l'eina de què disposem, més enllà de com ha estat concebuda. A nosaltres ens funciona. I ja està. I pel que fa a la llengua, sembla que el diari es pot llegir i es llegeix amb normalitat i naturalitat, i que es ven. Doncs endavant, no? I que s'hi apunti *La Vanguardia*, si pot.