



La postedició de continguts en publicacions diàries

Ignasi Navarro

Incyta Multilanguage (www.incyta.com)

distribuïdora oficial de tecnologia Lucy Software

(www.lucysoftware.com)

ignasi_navarro@incyta.com

RESUM

A Espanya hi ha actualment experiències interessants en l'ús de la traducció automàtica en mitjans de comunicació. L'article tracta sobre l'impacte de la tecnologia de traducció automàtica en aquests projectes, la integració d'aquesta en la cadena productiva del diari i la irrupció del posteditor en el procés d'edició.

Paraules clau: postedició, traducció automàtica, premsa escrita, premsa digital, premsa multilingüe, rendibilitat, edició, entorn de postedició

RESUMEN (*La postedición de contenidos en publicaciones diarias*)

En España existen actualmente interesantes experiencias en el uso de la traducción automática en medios de comunicación. El artículo aborda el impacto de la tecnología de traducción automática en estos proyectos, su integración en la cadena productiva del diario y la irrupción del poseedor en el proceso de edición.

Palabras clave: posesión, traducción automática, prensa escrita, prensa digital, prensa multilingüe, rentabilidad, edición, entorno de posesión

ABSTRACT (*Content post-editing in daily publications*)

Some interesting experiences could be gathered in Spain with regard to the use of MT systems in the media. This article addresses the impact of MT technology in bilingual press publications, its integration within the printing process and the emergence of the posteditor in the edition process.

Keywords: post-editing, machine translation, written press, digital press, multilingual press, profitability, editing, post-editing environment.



Introducció

Hem de remuntar-nos 15 anys per trobar la primera referència d'un mitjà escrit diari que utilitzi tecnologia de traducció automàtica (TA) en el procés d'edició. Em refereixo al diari Segre de Lleida, pioner a Espanya en l'ús de tecnologia lingüística per fer l'edició en català. Poc després, altres mitjans de premsa escrita, com El Periódico de Catalunya, van seguir el camí obert pel Segre i van decidir també llançar-se a l'aventura de la doble edició. Gràcies a això, la normalització de l'ús del català en la premsa escrita era més a prop. Finalment, fa encara no un any i mig, un dels degans de la premsa escrita europea, La Vanguardia, amb una antiguitat de 130 anys, va apostar també per la doble edició dels continguts en paper. Aquests tres diaris constitueixen un clar exemple que l'ús de la traducció automàtica pot ajudar a dur a terme projectes que, d'entrada, podrien semblar impossibles.

Durant aquests 15 anys l'ús de la traducció automàtica s'ha estès per altres àmbits amb resultats molt satisfactoris: Administració pública, portals web, editorials, indústria, entitats financeres, etc. En totes aquestes àrees hi ha múltiples processos que incorporen la traducció automàtica per optimitzar el temps i, per tant, els costos dels procediments de gestió de continguts multilingües. Ningú no qüestiona actualment que en determinades combinacions lingüístiques la traducció automàtica ofereix uns resultats espectaculars, i n'és evident l'aplicabilitat a qualsevol àmbit.

Posteditar, un estalvi inqüestionable

En primer lloc, convé definir amb el màxim rigor possible en què consisteix el procés de postedició. Encara no hi ha un consens ple sobre l'abast d'aquesta tasca, però segurament tots acceptaríem la definició com 'el procés de correcció que es fa sobre un text traduït automàticament a fi d'aconseguir una traducció correcta i fidel a l'original'. Es tracta, doncs, d'un tipus de correcció (potser més lèxica, sintàctica i estilística que ortogràfica), però també forma part d'un procés de traducció, per la qual cosa resulta indispensable disposar, en tot moment, de la referència del text original. La postedició no és assimilable a una simple correcció d'un text, però tampoc no podem entendre-la com un tipus de traducció; està a cavall de tots dos processos.

L'esforç que suposa la postedició d'un document dependrà molt del parell de llengües font i destinació, i de la qualitat de l'original –aquests dos factors tenen una relació directa amb la qualitat final de la traducció automàtica. Però hi ha altres aspectes que també influeixen en l'esforç de postedició:

- La finalitat de la traducció: un text que es publicarà requereix una postedició molt més acurada que un altre que ha de servir d'esborrany o d'informació d'ús intern.
- La naturalesa del text: no és el mateix traduir automàticament una unitat informativa completa i posteditar-la que fer el mateix amb segments descontextualitzats, sense cap referent que ens ajudi a resoldre ambigüitats.
- L'entorn de postedició utilitzat: és bàsic disposar del màxim confort, potència i seguretat per poder fer les tasques de la manera més ràpida i segura possible.

Sobre l'abast lingüístic de la tasca de postedició es pot convenir que hi ha dos nivells més o menys consensuats entre els professionals de la traducció/postedició; almenys així ho recullen les *Machine Translation Postediting Guidelines* de TAUS (*think tank* sobre innovació i plataforma de serveis compartits per al món de la traducció):

"L'esforç en la postedició està determinat per dos criteris principals:

1. La qualitat del resultat 'en brut' de la traducció automàtica.
2. La qualitat final esperada del contingut.



Per aconseguir una qualitat similar a la que s'assoleix amb la 'traducció i revisió humana d'alta qualitat' (també coneguda com a 'qualitat per a publicació'), es recomana generalment una postedició completa. Si es busca una qualitat inferior, sovint descrita com a 'prou bona' o 'apta per als propòsits', es recomana en general una postedició ràpida. Tanmateix, és possible que la postedició ràpida del resultat d'una traducció automàtica força pobre no proporcionï un text que compleixi amb els estàndards de qualitat per a la publicació. En canvi, si el resultat 'en brut' de la traducció automàtica és de bona qualitat, probablement només es requereix una postedició ràpida i no una de completa per obtenir una qualitat per a la publicació."

I en què consisteix el treball de postedició?

Posteditar, una tasca enganyosa

Malgrat el que hem exposat anteriorment, els professionals de la correcció/traducció tenen alguns recels a assumir el paper de posteditors.

Fa alguns anys, a les redaccions dels diaris els especialistes en llengua es dedicaven a corregir i revisar els textos que els arribaven. De vegades eren originals; d'altres, segones revisions. Però sempre es tractava només de text. Aquests professionals eren coneguts com a *correctors*, un espècimen situat entre la redacció i els tallers, dos mons molt pròxims però gairebé antagònics.

Amb l'arribada dels ordinadors la feina va canviar força. Ja no es dedicaven només a buscar faltes o a esmenar errors. Els avenços tècnics i tecnològics els permetien tenir la maqueta davant, amb el text inclòs. Això els obligava a ajustar el text, cosa de què no s'havien hagut de preocupar mai abans. Però no era només una qüestió de correcció i adequació a la mida del quadre de text, sinó que també era necessari controlar el peu de foto a la imatge que il·lustrava l'article. I els correctors van passar a ser, per art de l'evolució, editors: més feina, més variada i més responsabilitat.

I aleshores van aparèixer les màquines traductores. I l'editor va guanyar un prefix i va passar a ser posteditor. A la tasca que ja realitzava se li va afegir la cerca de falsos amics, de traduccions literals, canvis de gèneres, concordances impossibles, etc.

Parlem ara de rendibilitat

No resulta fàcil precisar la xifra d'estalvi que suposa l'ús de la traducció automàtica en un procés productiu, sobretot tenint en compte que hi ha processos en què la traducció forma part d'una cadena de tasques consumidores de recursos. En moltes d'aquestes cadenes o processos la traducció pròpiament dita és probablement la menys costosa de les tasques (en comparació amb l'edició, la maquetació, la publicació, la impressió, etc.). No obstant això, si en fem una comparativa respecte a paràmetres estàndards de traducció, podem afirmar que fer servir TA i posteditar el resultat obtingut (corregir-lo fins a aconseguir que la qualitat sigui assimilable a la d'un procés convencional) pot arribar a ser entre tres i set vegades més ràpid que el procés d'una traducció manual. Aquesta rendibilitat pot ser molt més alta en processos en què la traducció automàtica ofereix una qualitat suficient que, sumada a una postedició bàsica, ens porta a obtenir textos lingüísticament acceptables de manera ràpida (o molt més ràpida que un procediment manual convencional).

Aquest crec que és el punt més important: l'ús de la traducció automàtica (i, sens dubte, d'altres tecnologies) possibilita la realització de projectes que per un mitjà convencional serien inassumibles en termes econòmics.

Als elements que anteriorment hem descrit com a inherents al procés de postedició cal afegir altres factors que també influeixen en la productivitat del procés global. Aquests factors tenen



a veure més amb aspectes tecnològics del projecte en qüestió que amb el mateix procés de postedició. Alguns d'aquests elements són el motor de traducció utilitzat, el nivell de personalització implementat en el motor, la possibilitat de reutilització de text ja posteditat, les eines de millora del motor de traducció, l'entorn de postedició de què disposa el corrector, etc.

Portant a l'extrem màxim els paràmetres de rendibilitat, coincidiríem que el procés de postedició s'hauria d'assimilar (en temps i en costos) al d'un procés de lectura pausada del text.

Posteditar: és el mateix en qualsevol llengua?

Pel que fa a experiències en combinacions lingüístiques entre llengües no llatines cal indicar que els resultats obtinguts de la traducció automàtica, encara que acceptables en alguns casos, no són tan determinants (o productius) per assolir eficiències tan explícites com en les combinacions lingüístiques entre llengües llatines. És a dir, la qualitat de traducció que proporciona el motor de TA en combinacions entre llengües llunyanes (castellà, català o gallec amb anglès o alemany, per exemple) és més dispar i, per tant, es requereix una postedició més profunda. Probablement en aquestes altres combinacions lingüístiques no han existit ni les necessitats de volum de traducció ni els condicionants sociopolítics que sí que s'han donat en projectes o àmbits amb altres parells de llengües com el castellà-català o el castellà-gallec i que han resultat determinants per al desenvolupament i la maduració d'aquesta tecnologia. Sens dubte, actualment el repte és aconseguir que combinacions de traducció entre l'anglès i el castellà o el català puguin arribar a oferir rendibilitats similars a les que s'obté entre el castellà i el català o el castellà i el gallec.

Prensa multilingüe

Com s'ha esmentat al principi, a Espanya, concretament a Catalunya, hi ha una alta concentració de mitjans de comunicació que es dirigeixen als lectors en dues llengües. A les experiències de *Segre*, *El Periódico de Catalunya* i l'última, de *La Vanguardia*, caldria afegir-ne moltes altres de diaris digitals que ofereixen als lectors diverses versions lingüístiques (eldebat, e-notícies, elperiodico.com, etc.) i l'oferta d'agències de notícies que també publiquen les informacions en diferents llengües (Agència Efe, Europa Press o l'Agència Catalana de Notícies, per exemple).

Afegir una llengua al procés productiu d'un diari en paper, d'una agència de notícies o bé d'un mitjà digital no és tasca fàcil ni, molt menys, immediata. Si bé la tecnologia pot ajudar molt en la generació del contingut en la llengua de destinació, no hem d'oblidar que l'elaboració i la publicació de continguts periodístics es caracteritzen per ser un procés en què la immediatesa (sobretot en els mitjans digitals) té un paper fonamental. Per tant, sigui quin sigui el procés que s'utilitzi, el simple fet de "transferir" o "traslladar" una notícia a la llengua de destinació introdueix un retard en la publicació del contingut que provoca una pèrdua notòria de l'agilitat i, per tant, de la competitivitat, que ha de tenir un mitjà de comunicació. És, doncs, un objectiu primordial minimitzar aquest retard en la publicació del contingut, i per fer-ho el procés d'edició haurà de considerar tots els elements possibles que li permetin retallar aquest desfasament.

Hi ha diverses maneres d'entendre el procés de publicació multilingüe:

Procés asimètric. Una de les llengües és considerada llengua origen i, com a tal, el procés de publicació li dona prioritat.

Edició llengua A - Publicació llengua A - Traducció a llengua B - Publicació llengua B

Aquest procés implica un retard en la publicació de la llengua B. La demora dependrà molt de la naturalesa de la notícia (tipus, longitud, etc.), de com estigui integrat el procés de traducció a l'entorn d'edició i, finalment, de les eines de què disposi el



posteditor per corregir la notícia. Aquests aspectes són clau per aconseguir que el retard en la publicació de la llengua B sigui el mínim possible.

Procés simètric. Les dues llengües tenen la mateixa prioritats i, per tant, no es publica res fins que totes dues versions lingüístiques estan disponibles.

Edició llengua A - Traducció a llengua B - Publicació simultània llengua A+B

Aquesta opció retarda la publicació de la llengua A però garanteix que totes dues versions arribin al lector alhora. Com en el cas anterior, la magnitud d'aquest retard dependrà de la naturalesa de la notícia, de la integració del procés de traducció en l'entorn editorial i, finalment, dels recursos que tingui a l'abast el posteditor.

Paper: el coll d'ampolla de l'hora del tancament

El procés productiu d'un diari en paper obliga, pel que fa a l'elaboració de continguts, a disposar dels articles que el formen abans de l'hora de tancament de l'edició. En el cas de diaris bilingües, els continguts de totes dues edicions han d'estar disponibles abans d'aquest moment si el procés d'impressió es fa en paral·lel. Tots dos continguts arriben alhora a la rotativa i, per tant, el procés d'impressió i de distribució s'executa en paral·lel per a totes dues edicions. Això implica que s'ha de tenir en compte que el desfasament de generar la notícia en la llengua B és un temps que influeix en la producció de tots els articles de l'exemplar. Per tant, cal dissenyar molt bé el procés i considerar tots els aspectes organitzatius i tècnics rellevants per aconseguir que aquest temps sigui el mínim possible i, d'aquesta manera, minimitzar-ne l'impacte. Així, el retard introduït per l'edició bilingüe té un efecte sumatori en tots els articles (com ja s'ha explicat anteriorment), però el tancament de l'edició pot provocar cues i col·lapses en el procés de postedició, que poden portar a no tenir disponible l'edició de la segona llengua a temps.

Per evitar, o minimitzar, aquest risc, és fonamental no infravalorar cap dels aspectes organitzatius que intervenen en el procés:

- Definir correctament la unitat de postedició: article o pàgina.
- Dimensionar adequadament els equips de correctors per poder fer front, amb garanties, als pics de postedició que es produiran en l'estona prèvia a l'hora de tancament de l'edició.
- Flexibilitzar els recursos de correctors per donar una cobertura adequada a necessitats informatives especials: esdeveniments esportius, eleccions, esdeveniments rellevants, etc.

També hi ha aspectes tècnics que cal considerar de manera adequada:

- Integrar el procés de traducció automàtica dins de l'entorn editorial del mitjà. La traducció automàtica s'ha de desenvolupar de manera transparent i de la manera més ràpida possible sense haver de sortir de l'entorn editorial.
- Assegurar que el motor de traducció automàtica disposa dels ajustos lèxics i lingüístics necessaris per aconseguir que la traducció automàtica sigui com més correcta millor i que, alhora, s'aproximi als criteris del llibre d'estil del diari.
- Facilitar al posteditor eines que li permetin desenvolupar la tasca de la manera més ràpida possible, però també amb la màxima seguretat i consistència. Amb aquest objectiu, és molt important tenir un entorn de postedició en què es pugui disposar del text original, accés a eines en línia per fer consultes lèxiques i també del llibre d'estil, a banda de marques de colors sobre el text que facilitin la ràpida detecció de possibles problemes de sobretraducció o infratraducció que s'hagin produït en la fase de TA.



- Disposar, també finalment, de mecanismes ràpids i senzills que permetin una retroalimentació contínua del sistema de TA per detectar errors repetitius i continuar millorant la qualitat de la traducció.

D'aquesta manera, incidint en *tots* els elements organitzatius i tècnics del projecte, podrem garantir que el temps de postedició, que intervé en *tots* els articles de *totes* les edicions de *tots* els dies de publicació del diari pugui minimitzar-se de manera efectiva. Un error de plantejament o la infravaloració d'algun d'aquests elements incidirà, irremeiablement, en una pèrdua de temps en la correcció.

Digital: immediatesa i rapidesa

En el cas de les edicions digitals el problema és sensiblement diferent. Com que no hi intervenen ni tasques d'impressió ni de distribució, el procés de publicació és molt menys complex. En aquests mitjans, en canvi, el desfasament entre la publicació de la notícia en una llengua i una altra pot suposar que el lector visualitzi una jerarquia lingüística difícil d'entendre.

El procés de postedició del contingut és similar al que es fa en els mitjans en paper. Aquí, com en el paper, resultarà fonamental integrar-hi la TA, disposar d'un motor personalitzable (i convenientment personalitzat) i tenir un bon entorn de postedició. No obstant això, els aspectes organitzatius sí que són una mica diferents:

- El procés de postedició s'ha de desenvolupar durant una cobertura horària àmplia i continuada (gairebé 24 hores, 7 dies a la setmana).
- L'equip de correctors s'ha de dimensionar per poder assumir càrregues i pics informatius en diferents moments del dia i també poder atendre possibles esdeveniments extraordinaris com accidents, tragèdies, etc.
- Les notícies en els mitjans digitals són dinàmiques. Una mateixa notícia pot anar evolucionant al llarg del dia. Aquest aspecte és molt important, ja que ens obliga a haver de gestionar els diferents continguts amb les successives versions en totes dues llengües.

El mitjà digital ens permet alguns enfocaments diferents en el procés de postedició. En la postedició convencional, seqüencial i completa de la notícia o article abans de la publicació (mètode que actualment s'utilitza tant en els mitjans en paper com digitals) es plantegen altres mètodes interessants:

- Publicació de notícies traduïdes automàticament sense posteditar (o amb postedició limitada). Aquest mètode va ser utilitzat per l'ACN amb resultats molt satisfactoris. Les notícies originals en català eren traduïdes automàticament al castellà i publicades tal com sortien del sistema. Un procés de postedició limitat en esforç i en cobertura horària feia correccions a posteriori sobre els errors més greus. D'aquesta manera s'aconseguia la traducció de tots els continguts en una qualitat prou bona i alhora es garantia la detecció i l'eliminació dels errors més greus. Sens dubte aquest tipus d'experiències només és possible entre combinacions lingüístiques pròximes i una vegada s'ha personalitzat el sistema de traducció automàtica de manera molt precisa.
- Postedició limitada i dirigida. De manera similar es podria abordar la postedició de continguts limitant-la a criteris, com per exemple la gravetat d'un determinat error o bé el nombre de visites d'aquest contingut. Enfocant adequadament els esforços de postedició segons un determinat criteri (error, nivell d'importància de la notícia, visites) i en combinació amb un entorn de postedició potent que ens permeti recuperar totes les ocurrencies d'un determinat error i procedir a la correcció de manera ràpida, podem aconseguir rendiments molt superiors als que s'obtenen



actualment en els procediments seqüencials i complets de revisió. Hi ha, això sí, el risc de publicar errors mentre no es detectin i es corregeixin. Sens dubte, en aquest tipus d'enfocaments convé consensuar un equilibri entre els recursos de postedició i el risc de publicar continguts amb alguns errors. Una vegada més, un ajustament correcte del motor de traducció en aspectes com el tractament de noms propis i topònims pot minimitzar força el risc de publicar errors greus.

Conclusió

Segons el que publiquen Guus Extra i Durk Gorter a *Multilingual Europe: Facts and Policies*, "el català figura entre les deu llengües més traduïdes del món", tot i que és la 88a llengua del món en nombre de parlants. Probablement deu ser gràcies als sistemes de traducció automàtica.

Aquest fet demostra l'eficàcia de les tecnologies de TA. Sense l'existència d'aquestes tecnologies serien impossibles projectes com l'edició bilingüe de premsa escrita i digital i altres projectes que impliquen la traducció de grans volums d'informació en poc espai de temps.

No obstant això, la introducció d'aquest tipus de tecnologies requereix un procés de postedició minuciós i precís, realitzat per professionals, que són els qui finalment aporten la qualitat final a la traducció.

L'extensió de tasques de postedició en diverses àrees requereix l'existència de professionals adequadament formats i coneixedors de les virtuts i limitacions dels motors de traducció automàtica. La formació d'aquestes persones constitueix un repte per a les nostres institucions universitàries.

Agraïments

Agraeixo en gran manera les aportacions dels meus companys Marc Ràfols i Pere Herrero (responsables del projecte de postedició de *La Vanguardia* en català) i d'Esther Barjuan i Volker Zenner, sense els comentaris dels quals la publicació d'aquest article no hagués estat possible.