



Crowdsourcing i traducció/localització: una amenaça o una oportunitat?

Cristian Capdevila Fernández
Traductor, postgrau en Traducció i Tecnologies de la UOC

RESUM

El crowdsourcing ha canviat el món de la traducció als darrers anys. En alguns projectes ja no hi participen professionals, sinó usuaris i aficionats, i en molts casos sense compensació econòmica. La qualitat, però, no té perquè estar compromesa, ni tampoc la professió del traductor/localitzador en un futur pròxim.

Paraules clau: crowdsourcing, traducció col·laborativa, Q&A, flux de treball, ONG, llengües minoritàries, futur de la traducció.

RESUMEN (*Crowdsourcing y traducción/localización: ¿una amenaza o una oportunidad?*)

El crowdsourcing ha cambiado el mundo de la traducción en los últimos años. En algunos proyectos ya no participan profesionales, sino usuarios y aficionados, y en muchos casos sin compensación económica. La calidad, sin embargo, no tiene por qué estar comprometida, ni tampoco la profesión del traductor / localizador en un futuro próximo.

Palabras clave: crowdsourcing, traducción colaborativa, Q&A, flujo de trabajo, ONG, lenguas minoritarias, futuro de la traducción

ABSTRACT (*Crowdsourcing and translation/localization: Threat or opportunity?*)

Crowdsourcing has changed the translation world in recent years. Professional translators are not involved in some projects anymore, as some users and amateurs participate without getting economic compensation. However, neither the overall quality nor the profession of the translator/localizer in the near future are necessarily in danger.

Keywords: crowdsourcing, collaborative translation, Q&A, workflow, NGO, minority languages, future of translation



1. Introducció

En els darrers anys ha aparegut un nou terme al món de la traducció que ha suposat una petita revolució en aquesta professió: el crowdsourcing. Es tracta d'un canvi de paradigma en com es duu a terme la traducció, ja que no és el traductor l'actor principal del procés, sinó "la massa" o, com s'anomena en anglès, the crowd: un grup de persones no professionals que tenen la motivació suficient per traduir sense esperar un benefici a canvi de la seva feina o bé cobrant quantitats molt inferiors a les habituals en la professió. Aquest tipus de traducció comunitària comporta que moltes persones no professionals siguin les que tradueixin o localitzin un producte determinat.

No cal dir que la irrupció d'aquest fenomen ha anat acompanyat de molt de recel i desconfiança per part dels traductors professionals, els quals han vist com la seva feina corre perill de ser desplaçada per aficionats que tot i ser bilingües no tenen formació com a traductors. A aquesta recança derivada del crowdsourcing hi hem de sumar, a més, la que ha anat sorgint arran de les millores cada cop més visibles en la traducció automàtica (TA) i de l'intrusisme laboral que sempre ha estat més o menys present en la nostra professió, a més de les condicions laborals precàries imposades per alguns proveïdors de serveis lingüístics (LSP).

Davant d'aquest escenari que a primera vista ens pot semblar tan poc encoratjador, cal preguntar-se si el crowdsourcing representa realment una amenaça a la professió o si es tracta només d'una evolució de la mateixa. Cal recordar que quan aparegué la TA i les eines de traducció assistida (TAO) com les memòries de traducció (TM), molts traductors les consideraren enemigues de la seva professió, ja que molts van témer que aquestes novetats comportarien menys treball per a ells i pitjors condicions laborals. Finalment, s'ha vist que en comptes de significar un decreixement en el flux de feina aquestes noves eines han comportat un augment en el nivell de productivitat i qualitat. Si bé la por a la novetat és inherent a moltes professions, és important analitzar les seves millores en comptes de rebutjar-les per se.

Com sempre que sorgeix una novetat que pot canviar l'estatus quo d'una professió, trobem arguments a favor i en contra del crowdsourcing, a més d'algunes preguntes que cal analitzar abans d'extreure'n conclusions precipitades:

- Quines motivacions pot tenir una empresa o organització per emprar el crowdsourcing de forma exclusiva o parcial a les seves necessitats de traducció/localització?
- Es pot assegurar que la qualitat del producte final serà l'adequada si fem servir persones sense coneixements lingüístics o traductològics?
- Què fa que hi ha hagi gent disposada a traduir/localitzar de franc o per molt pocs diners?
- Quines amençades i oportunitats representa el crowdsourcing en la professió del traductor/localitzador?

El que està clar és que tota professió, sigui quina sigui, experimenta canvis més o menys dramàtics al llarg dels anys. Reciclar-se i estar atents a aquestes novetats esdevé, doncs, qüestió de supervivència.

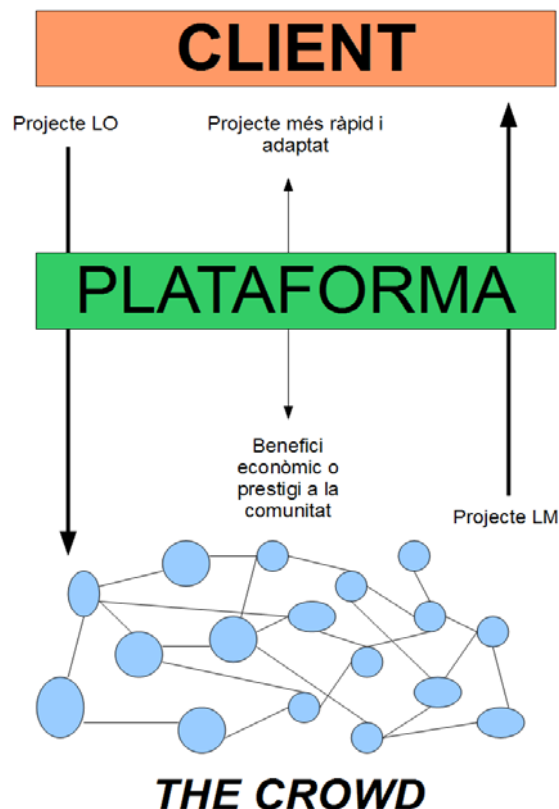
2. Com s'organitza el flux de treball del crowdsourcing?

La millor manera d'entendre com funciona el *crowdsourcing* i de posar de manifest les diferències amb la traducció/localització tal com la coneixem actualment és observant la forma en què es treballa. Si apliquem la mateixa lògica que Raymond (1999), trobem que l'activitat tradicional d'aquesta professió segueix el model de construcció d'una catedral: un



procés jerarquitzat on tothom té clar quina és la seva feina. Els traductors realitzen les tasques encarregades pel client o pel proveïdor de serveis lingüístics pel qual treballen i reben, a canvi, una compensació econòmica per la seva feina. Si els procediments estan clars i hi ha una bona comunicació entre totes les parts s'assoleix l'èxit i el projecte acaba traduït/localitzat.

Ara bé: com funciona un entorn col·laboratiu com el *crowdsourcing*? Seguint l'exemple anterior, a aquest procés el podem comparar amb un basar, on hi ha molta gent implicada sense un control tan estricte i on les veus i les opinions dels membres es comparteixen lliurement i, en ocasions, sense una organització fixa. Aquesta forma de treballar és comparable a la dels desenvolupadors d'aplicacions de codi obert. En l'esquema següent podem veure com entre el client i *the crowd* o massa de traductors/localitzadors aficionats existeix una plataforma creada pel client on es treballa de forma conjunta tant el procés de traducció/localització com la gestió del projecte. D'aquesta manera, el client envia el projecte en llengua original i els usuaris, tot interactuant entre ells i la plataforma, reenvien el projecte en llengua meta al client. Els beneficis per a la massa poden ser econòmics o de prestigi dins de la seva comunitat, mentre que el client obté a canvi un producte traduït/localitzat de forma més ràpida que al flux tradicional i més adaptat a les seves necessitats, ja que hi han col·laborat experts temàtics o bé usuaris que mitjançant la plataforma de treball escollida modelen el producte al seu gust.



En un model de traducció *crowdsourcing*, el treball s'estructura de manera que abans d'entregar-se a la comunitat de traductors/localitzadors, el gestor o l'encarregat projectes pot fer servir memòries de traducció o la traducció automàtica per pretraduir-lo abans d'enviar-lo a la plataforma on es traduirà. S'incorporen, però, dues característiques:



- El grup de traductors/localitzadors treballen de forma asíncrona.
- Un grup d'experts temàtics que no són traductors/localitzadors poden assessorar a aquests últims sobre temes rellevants del document sobre el que treballen i responen els dubtes que puguin sorgir durant el procés.
- El procés de control de qualitat es duu a terme a la vegada que es tradueixen/localitzen els segments, ja que diferents usuaris poden estar treballant en un mateix segment. En aquests casos es pot fer servir, per exemple, un sistema de votació perquè siguin els propis usuaris qui decideixen quina opció és la més adequada.

Per il·lustrar el que acabem de dir, podem fixar-nos en dos exemples:

El cas de Mechanical Turk

Es tracta d'una plataforma en línia que permet contractar traductors/localitzadors aficionats i pagar-los petites quantitats. Aquest sistema es fa servir, sobretot, per contractar personal que dugui a terme tasques que van més enllà de la TA, conegudes com HIT (*Human Intelligence Tasks*). De vegades, el preu que es paga per paraula o segment localitzat pot ser tan baix com 1 cèntim de dòlar nord-americà. Aquests tipus de plataformes faciliten l'accés al mercat de llengües minoritàries on no és ni fàcil ni barat trobar personal qualificat, i a més afavoreix el temps de resposta per realitzar les tasques, ja que varies persones poden treballar en paral·lel al mateix projecte durant qualsevol moment del dia.

Un dels majors inconvenients d'aquesta plataforma és que no permet saber quin tipus d'usuari la fa servir, amb el que es corre el risc de que no estigui qualificat per fer la tasca correctament i que la qualitat se'n senti.

El cas de Facebook

Aquesta coneguda xarxa social va començar a fer servir el *crowdsourcing* l'any 2007 per localitzar a diferents llengües el seu producte i ha aconseguit mobilitzar fins 30.000 usuaris per dur aquest projecte a terme (O'Hagan, 2009). Aquests usuaris van haver de traduir diferents segments del web mitjançant una aplicació anomenada *Facebook Translations*. D'aquesta manera s'aconseguia que els propis usuaris d'aquest servei fossin els que en realitzessin una localització adaptada a les seves necessitats. Una manera de motivar als usuaris era creant taulers virtuals on es podien veure els participants més actius en termes de qualitat i quantitat de material localitzat. O'Hagan també desmenteix que l'empresa ho fes per estalviar-se professionals, ja que el cost de desenvolupar una plataforma de localització col·laborativa tampoc no era menyspreable. Vegem, doncs, que no només era important el factor econòmic, sinó també el temporal: un equip de professionals no hagués pogut localitzar, testejar i gestionar un volum superior a les cent mil paraules en poques hores, sinó que haurien trigat mesos (Sargent, 2008). Gràcies a aquesta aplicació de localització els usuaris podien valorar la feina d'altres i votar quines propostes de localització eren les més adequades perquè apareguessin a la versió final. D'aquesta manera s'aprofitava al màxim la intel·ligència col·lectiva dels usuaris de Facebook i es garantien resultats de qualitat que satisfessin als usuaris finals del producte d'una forma més democràtica que no pas l'únic criteri d'un sol traductor/localitzador que imposés la seva versió a comunitats de milions de persones. Tot i així, la documentació de caire legal on els usuaris no tenien tanta motivació per traduir i que requeria un tractament especial fou enviada a professionals per assegurar-ne resultats de qualitat (Sargent, 2008).

3. Crowdsourcing, ong i llengües minoritàries

El *crowdsourcing*, a més de servir de plataforma perquè les empreses trobin una manera innovadora i eficaç de gestionar el procés de traducció/localització que els permeti obrir-se a nous mercats, pot ser també una eina molt valuosa per organitzacions sense ànim de lucre siguin governamentals o no. El fet de connectar a traductors/localitzadors voluntaris pot ser



molt útil per organitzacions que no treballin amb professionals de forma habitual o bé que per circumstàncies puntuals no puguin permetre-s'ho. Aquesta manera de treballar pot donar, fins i tot, resultats molt positius i acabar salvant vides en cas de donar-se una situació d'emergència com la que descriu Munro (2010), autor que desenvolupà una plataforma *crowdsourcing* després del terratrèmol d'Haití l'any 2010 per facilitar la traducció de missatges de text mòbils del crioll haitià a l'anglès. Aquests SMS contenien informació sobre ferits i punts d'abastiment, i es calcula que es varen salvar centenars de vides gràcies a aquest sistema.

Pel que fa a les llengües minoritàries, el *crowdsourcing* també pot ser de gran ajut per fer-les present a Internet i assegurar-ne la seva existència en la difusió del coneixement a la xarxa. El contingut disponible a Internet no ha parat de créixer des de la seva creació i actualment es pot trobar, sobretot, en anglès i en el grup de llengües anomenades FIGS (francès, italià, alemany i castellà, segons les sigles en anglès) tot i que n'hi ha d'altres com el xinès, el japonès i el coreà que també han estat molt presents al mercat de la localització de programari i de pàgines web. A més d'aquestes llengües *majoritàries* que són les que més solen reportar beneficis econòmics en aquest mercat, hi ha tot un seguit de llengües *minoritàries* que, ja sigui per tenir un nombre menor de parlants o per no reportar grans beneficis econòmics al mercat, poden ser deixades de banda. Tot i així, la expansió d'idees i coneixements que propicia Internet fa que existeixi la voluntat de que tot el contingut sigui accessible al major nombre de persones possible. La TA no assoleix encara resultats del tot acceptables pel que respecta a la qualitat de la traducció/localització de contingut web, i contractar professionals no està sempre a l'abast de molts llocs web per la quantitat de contingut que tenen i pel preu que haurien de pagar (Grunwald, 2011).

Aquests dos motius (mancances de la TA i preu de la traducció/localització de grans quantitats de contingut per professionals) han propiciat que molts autors i usuaris vegin en el *crowdsourcing* una solució per apropar aquest contingut a tota mena d'usuaris a un preu reduït. Amb un seguit d'eines que s'incorporarien als navegadors d'Internet, els usuaris que ho desitgessin podrien traduir/localitzar gratuïtament contingut a aquestes llengües per fer-les més accessibles. Es tracta, doncs, d'una bona oportunitat de tenir contingut a Internet disponible a moltes llengües sense que el seu pes econòmic sigui un factor determinant en la seva presència.

4. Crowdsourcing i control de qualitat

Actualment, la majoria de sistemes de TA es basen en sistemes SMT (*Statistical Machine Translation*) que fan servir una gran quantitat de corpus paral·lels que funcionen a nivell de frase. Aquests corpus, però, són en ocasions molt difícils de trobar en el cas de llengües minoritàries, i fer-los traduir per professionals seria massa car. En aquests casos, el *crowdsourcing* pot ser una molt bona eina per reduir aquest cos, però perquè la TA sigui correcta s'haurà d'assegurar un nivell de QA que sigui acceptable. Ambati et al. (2010), en la creació del seu marc d'ACT (*Active Crowd Translation*), un sistema que combina TA i *crowdsourcing* per millorar els sistemes de SMT, desenvolupen diferents algorismes per assegurar i calcular el nivell de qualitat de les traduccions i dels traductors de forma automàtica i fiable i per conèixer quina de les alternatives proposades pels traductors/localitzadors aficionats és la més correcta.

Ara bé, com podem assegurar la QA dels traductors/localitzadors aficionats, que en aquest cas treballen mitjançant la plataforma Mechanical Turk d'Amazon? Segons Ambati et al. (2010) i Zaidan & Callison-Burch (2011) es poden adoptar les següents mesures:

- Treballar amb plataformes que descriminin els candidats segons les seves qualificacions, mitjançant una prova de traducció, segons país de residència o anys que ha après la LO o la LM; o bé que s'acceptin només aquells que hagin demostrat la seva vàlua en projectes anteriors.



- Explicar clarament els objectius de la tasca als participants ha demostrat ser útil per descoratjar traduccions de mala qualitat.
- Recompensar als participants ja sigui econòmicament, amb mèrits o mitjançant l'ús de serveis gratuïts com és el cas de Duolingo, on es pot aprendre una llengua sense pagar res mentre es fan petites tasques de traducció.
- Fer que el segment a traduir/localitzar sigui una imatge i no un text per evitar que facin servir serveis de TA. També pot ajudar recol·lectar prèviament els resultats de la TA amb els segments a traduir per ajudar a identificar qui fa trampes i la fa servir.
- Fer que diferents traductors tradueixin/localitzin els mateix segment i que votin aquell que creguin que és millor.
- Desenvolupar algoritmes que calculin el nivell de fiabilitat de les traduccions i dels traductors. També es poden comparar alguns dels segments traduïts pels aficionats amb versions traduïdes per professionals, de manera que com més s'assembla a la segona, major serà la qualitat. És demostrable que aquells que fan bones traduccions ho continuaran fent al futur, i al contrari també (Zaidan & Callison-Burch, 2011).
- Wasala et al. (2012) proposen que hi hagi professionals que revisin els segments en cas d'haver rebut un nombre similar de vots, així com que els traductors/localitzadors aficionats estiguin obligats a inscriure's per evitar que facin un mal ús de l'aplicació. Consideren també útil crear un model tipus *wiki* perquè els usuaris puguin discutir sobre aspectes que concerneixen la qualitat final.

5. Conclusions

Si bé és cert que hi ha tasques que les poden fer aquests usuaris *amateurs*, la pràctica de la professió ens ensenya que per traduir correctament no només és necessari parlar dos o més idiomes, sinó que cal tenir coneixements sobre el que es tradueix per fer un text meta que tingui sentit. A més, cal dominar la llengua materna bé per fer textos correctes i amb un registre i vocabulari que s'adaptin als de l'original. Considero, doncs, que el *crowdsourcing* pot ajudar a traduir/localitzar grans quantitats de contingut del tipus UGC (*User Generated Content*) o plataformes en què els usuaris finals formen part d'una comunitat molt compromesa amb la qualitat final del producte. Parlo, fonamentalment, de producte que es desenvolupen en entorns molt col·laboratius, com ara Facebook o en el cas de programes i plataformes de codi obert que tenen un model de producció poc jerarquitzat. El *crowdsourcing* pot ser també de gran ajuda per localitzar contingut de webs que no tenen recursos per pagar la traducció professional i per ONG. Aquests productes solen generar la motivació necessària que cal per dur a terme tota aquesta feina sense pagar diners (o molt pocs) a les persones que els localitzen: desig de tenir prestigi a les comunitats on formen part, ganes de millorar el món o de resoldre problemes diversos que afecten l'usuari d'una manera o d'una altra, etc.

El gran error dels traductors i localitzadors professionals serà, doncs, ignorar o menysprear el fenomen del *crowdsourcing* en comptes de trobar la manera d'encabir la seva professió en aquest nou escenari, ja sigui com a guardians de la qualitat final de la traducció o aportant serveis de valor afegit com ara assessorament lingüístic. És difícil preveure una situació en un món cada cop més globalitzat on els traductors o les persones amb grans coneixements lingüístics no siguin necessaris. Un model col·laboratiu de dur a terme la traducció o la localització canviarà necessàriament i renovarà la professió del traductor. En paraules de García (2010), si tots aquests canvis tan positius per tothom representen la fi de la professió del traductor/localitzador tal com es coneix avui dia, però la transforma i l'adapta als nous temps, *so be it*.



Bibliografia

- Ambati, Vamshi, Stephan Vogel, y Jaime Carbonell (2010). *Active Learning and Crowd-Sourcing for Machine Translation*. Pittsburgh: Carnegie Mellon University.
- B. Sargent, Benjamin (2008). Community Translation Lifts Facebook to Top of Social Networking World.
<<http://www.common-senseadvisory.com/Default.aspx?Contenttype=ArticleDetAD&tabID=63&Aid=525&moduleid=391>> Data de la darrera actualització: 14.08.2008.
Data de consulta de la pàgina: 12.10.2012
- Beninato, Renato S; DePalma, Donald (2008). "Collaborative Translation". *MultiLingual Annual Resource Directory*, 2008.
- García, Ignacio (2010). "The proper place of professionals (and non-professionals and machines) in the web Translation". *Revista Tradumàtica*, 8.
- Grunwald, David (2011). Website translation using post-edited machine translation and crowdsourcing.
<<http://www.w3.org/International/multilingualweb/pisa/slides/grunwald.pdf>> Data de la darrera actualització: 31.03.2011. Data de consulta de la pàgina: 28.04.2012
- Munro, Robert (2010). *Haiti emergency response: the power of crowdsourcing and SMS*. Stanford: Stanford University.
- O'Hagan, Minako (2009). "Evolution of User-generated Translation: Fansubs, Translation Hacking and Crowdsourcing". *The Journal of Internasionalisation and Localisation*. Volum I, 2009.
- Raymond, Eric S (1999). *The Cathedral & the Bazaar*. Sebastopol: O'Reilly.
- Wasala, Asanka; Schäler, Reinhard; Weerasinghe, Ruvan i Exton, Chris (2012). "Collaboratively Building Language Resources while Localising the Web". *Proceedings of the 3rd Workshop on the People's Web Meets NLP, ACL 2012*. Jeju, Corea del Sud: Association for Computational Linguistics.
- Zaidan, Omar F; Callison-Burch, Chris (2011). "Crowdsourcing Translation: Professional Quality from Non-Professionals". *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*. Portland: Association for Computational Linguistics.