

SNELHEID VS. KWALITEIT: SCHUILT ER WEL WINST IN HET GEBRUIK VAN AUTOMATISCHE VERTAALSYSTEMEN?

Joke Daems, Lieve Macken & Sonia Vandepitte

De toenemende globalisatie zorgt voor een snel evoluerende vertaalsector. Zo steeg het vertaalvolume van het DGT van de Europese Commissie op acht jaar tijd met maar liefst een half miljoen pagina's (European Commission, 2005; European Commission, 2013). Het stijgende volume te vertalen materiaal leidt op zijn beurt tot een groeiend tekort aan vertalers. In een poging dit tekort op te vangen, probeert men onder andere gebruik te maken van automatische vertaalsystemen en de output ervan te verbeteren (of post-editen): dat proces zou immers sneller zijn dan het gebruikelijke manueel vertalen. De voorbije jaren is de kwaliteit van automatische vertaalsystemen er immers sterk op vooruitgegaan met de komst van statistische systemen. In 2009 al stelde het DGT het project MT@EC (European Commission) voor, dat vertaaldiensten zou leveren op basis van zulke statistische automatische vertaalsystemen. Nochtans bleek uit de cijfers van 2012 dat slechts 0,14% van de vertaaltaken van het DGT bestaat uit post-editing. Ook uit onze eigen bevraging bij zeventien masterstudenten vertalen (Engels) blijkt dat automatische vertaalsystemen nog lang niet ingeburgerd zijn: slechts 22% zegt vaak gebruik te maken van een automatisch vertaalsysteem. Hoewel de helft van de studenten gelooft dat post-editing sneller is dan manueel vertalen, lijken de studenten niet overtuigd te zijn van de kwaliteit: 67% gelooft dat manuele vertalingen van betere kwaliteit zijn, de overige 33% denkt dat er geen verschil in kwaliteit is.

Maar hoe zit het nu echt? Is post-editing werkelijk sneller dan manueel vertalen en zo ja, gaat dit dan niet ten koste van de kwaliteit? Het zijn onder andere deze vragen waarop we binnen LT³ (de afdeling Taaltechnologie van de vakgroep Vertalen, Tolken en Communicatie aan de Universiteit Gent) door middel van het ROBOT-

project (LT³, 2012) een antwoord trachten te vinden. Hoewel eerder onderzoek erop wees dat post-editing bijvoorbeeld binnen softwarelokalisatie tot snellere vertalingen kan leiden (Guerberof, 2009; Plitt & Masselot, 2010; Tatsumi, 2010), is er weinig onderzoek over het gebruik van post-editing bij algemene teksttypes terug te vinden. Het is dan ook het doel van ons project om een beter inzicht te krijgen in beide vertaalprocessen en -producten, in beide vertaalrichtingen, zowel voor studenten als professionele vertalers.

Binnen het ROBOT-project voerden we reeds twee voorstudies uit waarin we studenten een tekst lieten post-editen en een tekst lieten vertalen. In het eerste experiment werkten we met krantenartikelen, in het tweede met technische teksten. De resultaten van de tijdsanalyse waren behoorlijk eenduidig: post-editing was in beide studies voor iedere tekst gemiddeld sneller dan manueel vertalen. De kwaliteitsanalyse vormde echter een net iets ingewikkelder probleem. Bestaande metrieken voor kwaliteitsanalyse - zoals SAE J2450 (2001), LISA (2011) en EN-15038 (2006) - zijn weliswaar nuttig voor een snelle kwaliteitscheck binnen een bedrijf of vertaalbureau, voor een gedetailleerde analyse bleken ze minder bruikbaar. Vertrekkend van bestaande categorieën besloten we dan ook een eigen methodologie voor de analyse van vertaalkwaliteit uit te werken. We splitsten de analyse op in twee aspecten: aanvaardbaarheid (acceptability) en adequaatheid (adequacy), naar de verdeling van Toury (1995). De aanvaardbaarheid van een tekst wordt bepaald door de mate waarin de regels van de doeltaal en de doeltekst gerespecteerd worden: grammatica, lexicon, spelling, stijl, cohesie. De adequaatheid van een tekst wordt bepaald door de mate waarin de boodschap van de brontekst werd overgebracht in de doeltekst: tegenstellingen, woordverschuivingen, verkeerde woordbetekenis, toevoegingen, weglatingen,...

Bij het toepassen van onze methodologie op de vertalingen van de studenten, werd één ding meteen duidelijk: de analyse van vertaalkwaliteit is een bijzonder complexe taak, niet alleen

kwalitatief maar ook kwantitatief. Hoewel er twee annotatoren/evaluatoren waren, werd meer dan de helft van de problemen maar door één van beide aangeduid. Dit had enerzijds te maken met subjectiviteit (sommige personen zijn strenger dan andere¹), anderzijds ook met complexiteit (er zijn heel veel problemen waar je zo overheen leest). Na een consolidatiefase was de overeenkomst heel wat groter, en op basis van die overeenkomst konden we de verschillende vertaalfouten en -problemen in detail analyseren.

Op hoog niveau kunnen we alvast stellen dat bij studenten post-editing van algemene teksten zeker niet tot een slechtere kwaliteit leidt dan manueel vertalen, hoewel de eerstgenoemde vertaalmethode duidelijk sneller is. Bij manueel vertalen zijn adequaatheidsproblemen couranter dan bij post-editing, bij post-editing zijn aanvaardbaarheidsproblemen dan weer couranter. Opvallende verschillen vinden we ook wanneer we naar de meest voorkomende categorieën kijken bij beide vertaalmethodes: zo vormen woordesambigeringsproblemen bijna 10% van alle post-editingfouten, in vergelijking met slechts 5% van alle problemen bij manueel vertalen. Weglating is dan weer een probleem dat vooral bij manueel vertalen courant blijkt te zijn (7% van alle manuele vertaalfouten in vergelijking met 4% van alle post-editing fouten). Vooral opmerkelijk is het feit dat de problemen die na het post-editen bleven staan vaak problemen waren waarvan we hadden verwacht dat ze een post-editor meteen zouden opvallen:

Zo werd het begin van de zin '*Claims of apparent cure by novel treatment strategies or even by unconventional medicine or 'faith healing' should be seen in an appropriate context*' vertaald als: '*Vorderingen van schijnbare genezing...*' of bleven er zinnen staan met congruentieproblemen: '*Simulaties worden direct in de interne*

¹ Er is echter wel een correlatie tussen de beoordelingen van strenge en minder strenge evaluatoren ($r=0.89$, $n=38$, $p<0.001$ voor aanvaardbaarheid en $r=0.70$, $n=38$, $p<0.001$ voor adequaatheid).

indeling van de editor opgenomen en kan later als gebruikelijk frames worden bewerkt'.

De hoeveelheid fouten die opvallend vaak terugkeren bij het post-editen van automatisch gegenereerde vertalingen zou op enkele manieren gereduceerd kunnen worden. Zo zouden post-editors getraind kunnen worden om deze specifieke fouten te herkennen of zou het automatische vertaalsysteem kunnen leren van de aanpassingen die een vertaler maakt, zodat steeds terugkerende problemen (zoals incongruentie) niet langer over het hoofd gezien worden.

Binnen LT³ zijn er een aantal projecten die kunnen bijdragen aan de verbetering van de kwaliteit van automatische vertaalsystemen en post-editing (zie <http://www.lt3.ugent.be/en/projects/> voor meer informatie over ieder project). Zo werkte Els Lefever in het project ParaSense (*Parallel Corpora for Word Sense Disambiguation*) rond woordbetekenisdesambiguering.

Aangezien desambigueringsproblemen zo'n 10 procent uitmaakten van alle gemaakte post-editingproblemen uit onze voorstudie, zou onderzoek naar woordbetekenisdesambiguering zeker de kwaliteit van automatische vertaalsystemen en/of post-editing kunnen verbeteren. Ook binnen het ROBOT-project voeren we verder onderzoek naar het vertaal- en post-editingproces. Aan de hand van tools zoals eye-tracking (het registreren van oogbewegingen) en toetsregistratie zullen we onderzoeken hoe het post-editingproces verschilt van het reguliere vertaalproces. Een van de mogelijke toetsregistratietools is Inputlog (www.inputlog.be), waarvoor LT³ in het project Inputlog++ een extensie ontwikkelde waarmee procesdata verrijkt kunnen worden met taalkundige informatie (Macken et al., 2012). Daarnaast zal ook het SBO-project SCATE een grote impact hebben op de interactie tussen vertaler (of post-editor) en output van een automatisch vertaalsysteem. Binnen het SCATE-project wordt een vertaalomgeving ontwikkeld waarbinnen de samenwerking tussen mens en computer geoptimaliseerd wordt. Er

wordt onder andere gekeken naar typische post-editingproblemen en post-editinginspanning. Op basis van die gegevens is het de bedoeling om de post-editinginspanning a priori te bepalen, zodat de vertaalomgeving enkel een automatische vertaalsuggestie aanbiedt als die suggestie ook een meerwaarde (tijdsinstroom zonder grotere cognitieve belasting) met zich meebrengt voor de vertaler.

Referenties

EN 15038. (2006). *Translation services - Service requirements*.

European Commission. (2005). *Translation in the Commission: where do we stand eight months after the enlargement?* MEMO/05/10. [Online] http://europa.eu/rapid/press-release_MEMO-05-10_en.htm#fn1

European Commission. (2009). *Machine Translation Service MT@EC*. [Online] http://ec.europa.eu/isa/actions/02-interoperability-architecture/2-8action_en.htm

European Commission. (2013). *Translation in figures - 2012*. [Online] http://ec.europa.eu/dgs/translation/whoweare/translation_figures_en.pdf

Guerberof, A. (2009). *Productivity and quality in MT post-editing*. Paper presented at the MT Summit XII -Workshop: Beyond Translation Memories: New Tools for Translators MT, Ottawa, Ontario, Canada.

Localization Industry Standards Association. LISA QA Model 3.1. Geraadpleegd via www.lisa.org/LISA-QA-Model-3-1.124.0.html

LT³. (2012). *Robot*. [Online] <http://www.lt3.ugent.be/en/projects/robot/>

Macken, L., Hoste, V., Leijten, M., & Van Waes, L. (2012). *From keystrokes to annotated process data: Enriching the output of*

Inputlog with linguistic information. Paper presented at LREC 2012, Istanbul.

Plitt, M., & Masselot, F. (2010). A Productivity Test of Statistical Machine Translation Post-Editing in a Typical Localisation Context. *The Prague Bulletin of Mathematical Linguistics*, 93, 7-16.

SAE J2450. (2001). *Quality Metric for Language Translation*. [Online] www.apex-translations.com/documents/sae_j2450.pdf

Tatsumi, M. (2010). *Post-Editing Machine Translated Text in a Commercial Setting: Observation and Statistical Analysis*. Dublin: Dublin City University.

Toury, G. (1995). The Nature and Role of Norms in Translation. In Toury, G., *Descriptive Translation Studies and Beyond* (pp.53-69). Amsterdam/Philadelphia: John Benjamins.