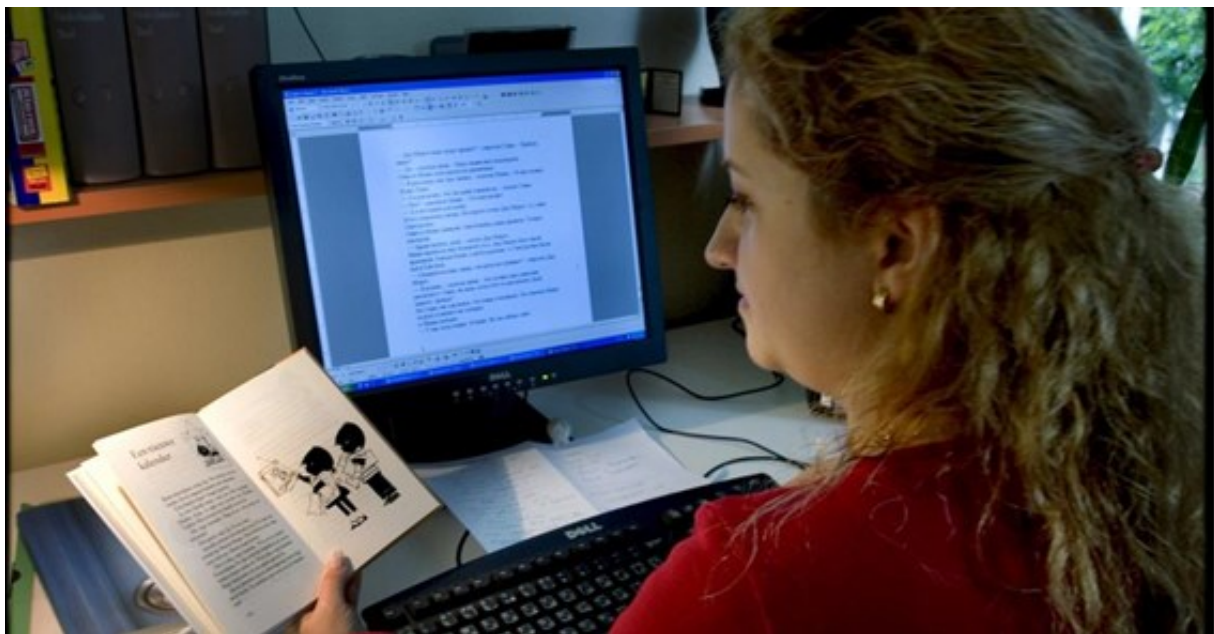


COMPUTERVERTALING IS REKENWERK

Zoeken naar de perfecte match

06/11/2013 | Sam Delbeke

De universele vertaler uit Star Trek, die alle talen uit het universum ogenblikkelijk vertaalt, is nog niet voor morgen. Maar met de vertaalsoftware van vandaag kom je toch al een heel eind. Met dank aan de statistiek.



Bijna alle vertalers werken vandaag met vertaalsoftware. pn

De computer is een breinloze doos die eentjes en nulletjes leest, maar voor het overige niks snapt van de wereld om zich heen. Lexicale dubbelzinnigheid, verschillende woordvolgorde, idiomatische uitdrukkingen: voor ons is het relatief makkelijk te behappen, maar de computer krijgt zulke nuances nooit op afdoende wijze in bits-en-bytesreeksen gegoten. Daarom zal een computer nooit betrouwbare vertalingen afleveren.

Dat was in een niet eens zo ver verleden de consensus binnen de vertaalwereld. Maar intussen is de taal- en spraaktechnologie blijven evolueren. Vertalers die een eenvoudige tekst snel willen verwerken, kunnen hun vertaling hardop uitspreken en die door spraakherkenningssoftware laten uittikken, wat flink wat tijdwinst betekent. En met professionele vertaalsoftware kun je vandaag, afhankelijk van het type tekst en de talencombinatie, in bepaalde gevallen ronduit uitstekende resultaten boeken.

‘De huidige vertaalgeheugens worden door veel professionals gebruikt en vaak als zeer nuttig ervaren’, zegt hoogleraar Veronique Hoste van de vakgroep Taaltechnologie van de UGent. ‘In de sector zijn er trouwens nog maar weinig vertalers die niet met vertaalsoftware werken. Veel vertaalbureaus verplichten hun vertalers de facto om met bepaalde vertaalsoftware aan de slag te gaan. Ze betalen de vertaler daarbij alleen voor zijn “nieuwe” vertaalwerk, dat ze nog niet in het eigen vertaalgeheugen hebben.’

‘Zo’n vertaalgeheugen moet je zien als een reusachtige databank van parallelle, originele en vertaalde, teksten. Het programma vergelijkt op grote schaal een nieuw te vertalen zin met het geheugen van originele en vertaalde teksten. Met de *perfect matches* (de perfect overeenstemmende vertalingen) en de *partial matches* (gedeeltelijk overeenstemmende vertalingen) die de vertaalsoftware aanreikt, kom je als vertaler al een heel eind. Dat geldt uiteraard vooral, zo niet uitsluitend, voor teksten als handleidingen, procedures of jaarverslagen, waarin heel vaak dezelfde of soortgelijke termen en zinnen gebruikt worden.’

Hybride vertaalmachine

Maar ook volautomatische vertaling verovert steeds meer terrein. ‘Het systeem van op statistiek gebaseerde machinevertaling maakt gebruik van kolossale databases van door professionals vertaalde teksten en trekt daar statistische conclusies uit: wat zijn de mogelijke vertalingen van een bepaald woord of een bepaald stukje zin? En wat is daarvan, gezien de context, de waarschijnlijkste? Daarbij geldt: *no data like more data*. Hoe groter de databank van vaktermen en zinsconstructies waarvoor goede vertalingen bestaan, hoe beter de vertaling zal zijn.’

‘Uit de feedback van de vertalers van de Europese instellingen blijkt dat de kwaliteit van de automatische vertaling bijzonder sterk afhankelijk is van het onderwerp, de talencombinatie en de kwaliteit van de brontekst’, zegt Floris van Nierop, vertaler bij de Europese Commissie. ‘Vertalingen van Romaanse of Germaanse talen onderling zijn over het algemeen goed tot heel goed. De vertaalcombinatie Engels-Zweeds, bijvoorbeeld, levert vaak zelfs ronduit uitstekende resultaten op. Voor bepaalde andere vertaalcombinaties, bijvoorbeeld van het Engels naar een agglutinerende taal (een ‘plaktaal’, waarbij woordvorming plaatsvindt door het aaneenrijgen van kleinere woordelementen), zijn de resultaten slecht tot heel slecht.’

‘Persoonlijk denk ik dat we op middellange termijn nog veel progressie zullen maken met hybride vertaalmachines. Daarbij wordt de statistische machinevertaling, waarbij de betekenis meestal goed is maar de grammatica te wensen overlaat, gecombineerd met op grammaticale regels gebaseerde vertaalsoftware, waarbij de vertaling doorgaans grammaticaler is, maar weleens inhoudelijk tekortschiet.’

Vertalers kunnen hun vertaling hardop uitspreken en die door spraakherkenningssoftware laten uittikken, wat flink wat tijdwinst betekent

