cultuur

Vertaalapps worden massaal gebruikt, maar kunnen we daar wel op vertrouwen?

Elk uur 6 miljard woorden vertalen

ILLUSTRATIE TJARKO VAN DER POL

De computer kent steeds meer woorden

«THE COMPUTER HAS MORE AND MORE WORDS» Zo vertaalt Google Translate de kop bij dit stuk. Moeten vertalers voor hun baan vrezen, nu vertaaltechnologieën steeds beter worden?

Door onze redacteur **Liza van Lonkhuyzen**

Illustratie **Tjarko van der Pol**

Vergelijk zelf

We kozen een fragment uit de roman *The Circle* van Dave Eggers. Vergelijk de vertalingen:

Originele tekst:
She pushed open the heavy door. The front hall was as long as a parade, as tall as a cathedral. There were offices everywhere above, four floors high on either side, every wall made of glass. Briefly dizzy, she looked downward, and in the immaculate glossy floor, she saw her own face reflected, looking worried. She shaped her mouth into a smile, feeling a presence behind her. „You must be Mae."

Vertaling Google Translate:
Ze duwde de zware deur open. De voorzaal was zo lang als een parade, zo lang als een kathedraal. Er waren kantoren overal boven, vier verdiepingen hoog aan weerszijden, elke muur van glas. Kortom duizelig keek ze naar beneden, en in de onberispelijke glanzende vloer zag ze haar eigen gezicht weerspiegelen, en maakte zich zorgen. Ze vormde haar mond in een glimlach, voelde een aanwezigheid achter haar. „Je moet Mae zijn."

In de Nederlandse vertaling van het boek:
Ze duwde de zware deur open. De grote hal was zo lang als een promenade, zo hoog als een kathedraal. Bovenin zag ze overal kantoren, aan beide kanten vier verdiepingen hoog, alle wanden waren van glas. Ze werd even duizelig, keek naar beneden en zag in de smetteloos glanzende vloer haar eigen gezicht weerspiegeld met een zorgelijk trekje erop. Ze krulde haar lippen tot een glimlach en voelde dat er iemand achter haar stond. „Jij bent vast Mae."

Vertaalster Adine Rooyackers: „Google mist nuances: 'parade' betekent in dit fragment uit het boek 'een bouwwerk waar je overheen kan lopen'. Parade wordt in het Nederlands meer gebruikt voor een optocht. Google mist ook de nuance tussen long en tall – in deze context betekenen ze 'lang' en vervolgens 'hoog'."

In de ochtend vertaalt ze een gesprek bij uitkeringsinstantie UWV. 's Middags helpt ze een Poolse verdachte in een rechtszaak. Tussendoor is ze in de weer met juridische aktes. De Pools-Nederlandse vertaalster Anna Roslon heeft het druk.

Maar komt er een einde aan deze stroom werk? Steeds meer grote techbedrijven begeven zich op de vertaalmarkt. Daarbij worden grote stappen gezet, mede dankzij kunstmatige intelligentie. Computers leren vertalen door simpelweg heel veel menselijke vertalingen te bestuderen en daarvan zelfstandig te leren. Google Translate, de meestgebruikte vertaaldienst, is daar ver mee. Google wil vertalen met zelflerende computersystemen nog dit jaar voor elke taal hebben geïmplementeerd. Moet Roslon zich zorgen maken?

Vertalers beginnen de hete adem van techreuzen te voelen, zegt Antal van den Bosch, hoogleraar taal- en spraaktechnologie aan de Radboud Universiteit. Als voorbeeld wijst hij naar de handleiding van een nieuwe tv. Wie die openslaat, leest geen mooie zinnen in. Dikke kans dat de instructies door een computer zijn vertaald naar tien verschillende talen. „Nog net met een menselijke correctieronde", zegt Van den Bosch.

Veel vertalers als Roslon hebben een of meer specialismen, bijvoorbeeld voor de autobranche, de politie of de medische wereld. „In mijn werk komt veel jargon uit de rechtspraak langs", zegt Roslon. „Een ondertoezichtstelling wordt een OTS genoemd. Daarbij kan een dienst als Google Translate nog niet helpen."

Toch sluipt Google Translate gaandeweg het territorium van Roslon binnen. Polen gebruiken het om brieven van instanties, gemeenten, UWV of sms'jes te lezen. „Als ze dat doen, raad ik ze aan om simpele zinnen te gebruiken, anders is Translate niet betrouwbaar." Ze merkt wel dat de woordenschat van de dienst de afgelopen jaren enorm vooruit gegaan. „Zelf gebruik ik het graag als een soort woordenboek, voor suggesties."

„Laatst had een klant een app op zijn telefoon waarmee hij tekst kon inscannen en vertalen", zegt Janske van Dun-Vrijsen, vertaalster Chinees-Nederlands. „Zo kon hij een beeld krijgen van de inhoud van de tekst, maar het was geen goede vertaling."

Vertalers waarschuwen dat de context van de alinea of de schrijfwijze verloren kan gaan met vertaaltechnologie. Ook cultuur heeft zijn eigen taalcontext. Van Dun-Vrijsen merkt dat vaak in haar werk. „In het Chinees zijn mensen minder direct, zelfs in zakelijke e-mails", zegt ze. „Ze wringen zich in allerlei bochten om geen 'nee' te hoeven zeggen." Andere omgangsvormen zijn politiek geladen. „Als ik documenten vertaal voor de Chinese ambassade, schrijf ik Hongkong als een expliciete regio van de Volksrepubliek China."

Komma's op verkeerde plaats

Het kan in sommige omstandigheden zelfs levensgevaarlijk zijn om op technologieën, of niet gescreende vertalers te vertrouwen, zegt Mirjam van Dootingh, interim-manager van het Nederlands Genootschap van Tolken en Vertalers. „Als je bijsluiters vertaalt voor geneesmiddelen en je zet komma's op de verkeerde plaats, kan dat dodelijke gevolgen hebben. Hetzelfde geldt voor de chemische industrie."

Toch wordt vertaalwerk steeds meer door machines gedaan. Vertalers gebruiken bijvoorbeeld al jaren automatische '*translation memory*'. Dat is een soort database, die op basis van eerdere ingevoerde tekst suggesties doet. „De software levert in feite half werk", zegt Antal van den Bosch. „Terwijl je typt, krijg je tekstsuggesties. Werk gaat op die manier sneller." Veel vertaalbureaus stellen het gebruik van *translation memory* zelfs als eis.

Een groeiend deel van de markt bestaat daarnaast uit post-editing, waarbij vertalers automatisch vertaalde teksten alleen nog corrigeren. „Veel vertalers zijn daar geen fan van", zegt Mirjam van Dootingh. „Het eindresultaat is vaak minder goed."

Ook binnen *translation memory* wordt geïnnoveerd, onder meer door *machine learning*. Als vertalers een bepaalde vertaalconstructie vaak aanpassen, wordt dat als suggestie in het programma voor alle vertalers doorgevoerd.

Bij technisch vertaalwerk wordt technologie prominenter, maar voor literaire vertalers is er weinig aan de hand. „Bij verhalende vertalingen heb je een dubbele taak", zegt Van den Bosch. „Je moet ook de sfeer van de originele auteur weten vast te houden. Vertalers nemen dan ook meer vrijheid. Een dienst als Google Translate faalt hier nog enorm in."

Blijft er genoeg werk over voor vertalers? Van Dootingh blijft optimistisch. „Gelukkig neemt de informatiestroom exponentieel toe en daarmee ook de hoeveelheid teksten die naar een andere taal omgezet moeten worden."

Ook de Pools-Nederlandse vertaalster Roslon ligt nog niet wakker van de technologiereuzen. „Ik vind het wel spannend, maar vooral in de positieve zin. Mijn vak is altijd al in beweging." In haar beginjaren als vertaalster was er bijvoorbeeld geen software die Slavische tekens kon produceren. Bijzondere tekens zette ze erbij met een zwarte pen. „Evolutie is van alle tijden, je moet gewoon meegaan in de ontwikkeling."

Daar is Van Dun-Vrijsen het mee eens. „Als ondernemer moet je blijven kijken hoe je toegevoegde waarde kan leveren."

Wat de vertaalapps wel en (nog) niet kunnen

«WHAT THE TRANSLATION APPS DO AND CAN NOT (YET)»

Computersystemen vertalen steeds beter, maar de tekst ook echt begrijpen kunnen ze (nog) niet.

Met meer dan 140 miljard te vertalen woorden per dag heeft Google Translate het druk. En inmiddels is de vertaaldienst de half miljard maandelijkse gebruikers gepasseerd, aldus *The New York Times*. En het is lang niet de enige techgigant die zich begeeft op vertaalgebied. Facebook vertaalt met behulp van kunstmatige intelligentie - systemen die zelf kunnen leren en redeneren - buitenlandse statusupdates voor je. Het Chinese Baidu bouwt een systeem dat beter menselijke taal kan begrijpen, zodat computers namens bedrijven klanten te woord kunnen staan als chatbots. Ook Microsoft en Amazon werken aan vertaaldiensten.

Vertaaltechnologieën zetten grote stappen door *machine learning*, waarbij computers uit zichzelf leren. Zo wil Google nog dit jaar in alle talen overgaan naar de zogeheten '*long short term memory*'-techniek. Dit 'lang kortetermijngeheugen' kan tegen het einde van de zin het begin nog onthouden. Zo wordt er niet meer woord voor woord vertaald, maar kijkt de software naar de zin als geheel. Volgens het techbedrijf dringt het fouten met 60 procent terug.

Het zijn spannende tijden, zegt Lieve Macken, taaltechnoloog aan de opleiding Taal- en Letterkunde in Gent. Ze wijst op de stap naar zinsvertalingen die Google in april zette voor het Engels naar Nederlands. „De Nederlandse taal bestaat uit veel lastige werkwoorden. Als je zegt: 'Ik bel hem morgen na de lunch terug', dan is het moeilijk voor een vertaalsysteem om te snappen dat 'terug' nog bij 'bellen' hoort. Google snapt dat nu en schrijft: *'I'll call him back tomorrow after lunch.'"*

Het kon al, maar nu nog beter

Grote bedrijven werken aan zelflerende technieken als automatisch vertalen, automatische spraakherkenning en tekst-naar-spraaksystemen. „Allemaal technologieën die al lang bestonden", zegt Antal van den Bosch, hoogleraar taal- en spraaktechnologie aan de Radboud Universiteit. „Met meer data als voorbeeldmateriaal worden ze steeds beter. Door de komst van technologiereuzen, snelle netwerken en meer opslagcapaciteit van data worden schaalproblemen overwonnen."

Toch zijn er ook nadelen aan dit soort vertaaldiensten. „Men vergeet weleens dat je feitelijk de teksten aan het bedrijf uitlevert", zegt Macken. „Misschien wil je dat helemaal niet vanwege de privacy. Het is een van de redenen waarom de Europese Commissie haar eigen vertaalsysteem heeft gebouwd."

Andere nadelen zijn er ook. Zo maakt Google nog steeds veel rare fouten in woordkeuze. Macken: „Wat automatische vertaalsystemen nog altijd niet kunnen - al claimen ze vaak van wel - is een zin écht begrijpen zoals wij taal begrijpen. Wie weet verandert ook dat."

Liza van Lonkhuyzen

