

En de beste gratis online vertaalmachine is ...

Gent, augustus 2014

Ja, de gratis vertaalmachine van Google is de bekendste. Maar is het ook de beste? En als je de kwaliteit niet vertrouwt, kun je daar dan als leek zelf iets aan doen?

Of is Bing dan toch beter? Klopt het dat je (nog) beter Spaans krijgt als de vertaalmachine je Nederlandse zinnen eerst in het Engels voorvertaalt en dan uit het Engels in het Spaans vertaalt? Helpt het als je je zinnen zelf voorvertaalt in het Frans, ook als je Frans niet helemaal top is?

Yanah Van Geerteruy studeert meertalige communicatie aan de Universiteit Gent. Tijdens haar zomerstage nam ze twee weken de tijd om het uit te pluizen. Wat ze vond, is niet wereldschokkend maar wel verrassend. Haar onderzoekje suggereert minstens dat meer en grondiger onderzoek nuttig kan zijn.

Ola María?

Het is weer zomer. De warme temperaturen en de volle terrasjes doen je denken aan je heerlijke vakantie vorig jaar in Spanje. Barcelona. Antoni Gaudí. En de Montjuïc. Om dan nog te zwijgen over María, de gastvrouw bij wie je logeerde. Als je terugdenkt aan hoe zij jou verwendde, krijg je meteen weer een warm gevoel. Dan besluit je haar een e-mail te sturen om te zeggen dat je haar deze zomer graag opnieuw een bezoekje zou brengen. Er is echter één probleem. Je Spaans is niet al te best, en zoals iedereen weet, hebben de Spanjaarden niet veel kaas gegeten van het Engels. Hoe pak je dit aan?

Gratis online vertaalmachines de oplossing?

Iedereen kan zich in de situatie inleven: je hebt iets te vertellen aan iemand die jouw moedertaal niet spreekt en die ook geen kennis heeft van het Engels, in zulke situaties meestal de voertaal. Zijn online vertaalmachines de oplossing? Vertaalmachines genoeg, maar wat zijn ze waard? Google Translate is de bekendste, maar is het ook de beste? Omdat niemand graag een mal figuur slaat, leek het me interessant om een klein onderzoek op te zetten: welke online vertaalmachine levert de meest correcte vertaling?

Creatief met vertaalmachines

Met welke ingrepen kun je die machinevertaling eventueel zelf nog verbeteren? Concreet: helpt het als je je oorspronkelijke input eerst vertaalt in een taal die couranter is dan je moedertaal (zoals het Engels couranter is dan het Nederlands)? En helpt het als je je oorspronkelijke input eerst vertaalt in een andere (eventueel

courantere) taal die al tot dezelfde taalfamilie als je doeltaal (bv. Frans en Spaans) behoort?

En: doe je dat “voorvertalen” beter zelf – ook als je Engels of je Frans niet top zijn – of kun je ook daarvoor beter de vertaalmachine inzetten? Dat is wat ik allemaal wilde weten.

Voor mijn onderzoek heb ik de vertaalmachines Google Translate, Bing Translator en Intertran geselecteerd. Ik koos verder voor Nederlandse input en Spaanse output en voor Engels en Frans als bijkomende talen voor de ‘voorvertaling’.

Om de kwaliteit van de vertalingen objectief te kunnen vergelijken, heb ik met een internationaal gerenommeerd puntensysteem gewerkt. Verder in dit artikel bespreek ik uitvoerig hoe ik te werk ben gegaan.

En de beste is ...

En de beste vertaalmachine is ...?

De uitkomst van mijn onderzoek naar de beste gratis online vertaalmachine voor de talencombinatie Nederlands-Spaans is toch enigszins verrassend.

Dit zijn de vier belangrijkste resultaten.

1. Vertaalmachine Bing Translator geeft de meest correcte output.
2. Werken met Bing Translator via het Engels naar het Spaans levert - zowel bij een menselijke als bij een machinale voorvertaling - een beter resultaat op dan rechtstreeks vanuit het Nederlands.
3. Werken via het Frans, een taal die tot dezelfde taalfamilie als het Spaans behoort, levert geen beter resultaat op.
4. **Het beste resultaat in het Spaans krijg je met Bing Translator én als jijzelf – dus niet de machine - de oorspronkelijke Nederlandse input voorvertaalt naar het Engels – zelfs als je Engels niet top is.**

Vertalen en vertaalmachines: achter de schermen

Een tekst vertalen, wat is dat eigenlijk? In essentie: een tekst in de ene taal omzetten in een tekst in een andere taal. De eerste tekst noemen we de brontekst, de tweede de doeltekst.

Hoewel er veel verschillende visies bestaan over wat een goede vertaling is, gaan we er in dit artikel van uit dat een geslaagde vertaling de inhoud van de brontekst correct weergeeft in de doeltekst. Traditioneel doet de vertaler daarvoor vooral een beroep op zijn talenkennis, eventueel in combinatie met woordenboeken. Deze menselijke manier van vertalen kost echter veel tijd en geld. Daarom probeert men het vertalen de laatste jaren steeds meer te automatiseren. Dit betekent niet dat de

menselijke vertaler overbodig is geworden. Dat zegt ook Lynne Bowker, professor aan de School of Translation and Interpretation in Ottawa, in haar boek *Computer-aided Translation Technology*:

“Automatisch vertalen is het proces waarbij een computer de primaire verantwoordelijkheid voor de vertaling van een tekst heeft. Een mens kan helpen bij het proces door middel van taken zoals pre- of post-editing, maar het is de computer in plaats van de mens die een daadwerkelijke ontwerpvertaling produceert.” (eigen vertaling)

Mens versus machine

Volgens professor Bowker doet een computer bij automatisch vertalen dus weliswaar het grootste werk, maar de menselijke vertaler kan nog altijd ingrijpen. Dat ingrijpen kan men dan echter niet langer vertalen noemen, maar eerder reviseren. De vertaler wordt hier dus eigenlijk revisor.

We kunnen dus om te beginnen twee types van automatisch vertalen onderscheiden:

- de volledig automatische machinevertaling (Fully Automatic Machine Translation ofte FAMT)
- de door de mens ondersteunde machinevertaling (Human Aided Machine Translation ofte HAMT)

Zoals professor Bowker aangeeft, kan de mens die de machinevertaling ondersteunt - de revisor dus - dat op twee manieren doen: hij kan de brontekst bewerken en hij kan de doelttekst bewerken.

Pre-editeren

In het eerste geval spreekt men van pre-editeren: de brontekst wordt zodanig aangepast dat de vertaalmachine minder moeite heeft met het omzetten. De onderzoekers Chris Callison-Burch en Philipp Koehn kwamen tot de vaststelling dat vertaalmachines vooral moeite hebben met woordvolgorde, dubbelzinnigheid op het vlak van structuur en woordenschat, gebruik van voornaamwoorden, en gebruik van de tijden. Het zijn dus vooral deze problemen waar de revisor bij het pre-editeren op moet letten.

Postediteren

In het tweede geval bewerkt de revisor niet de brontekst maar de doelttekst. Dan spreekt men van postediteren. Dit betekent dat de revisor de doelttekst grondig nakijkt. Hij moet dan opnieuw speciale aandacht hebben voor Callison-Burch en Koehns voornoemde problemen.

Regels versus statistiek

Vervolgens is het interessant om te weten dat vertaalmachines volgens verschillende principes kunnen werken.

Ofwel is een vertaalmachine regelgebaseerd, wat inhoudt dat ze kennis heeft van de grammaticale regels van bron- en doeltaal. Een regelgebaseerde vertaalmachine is in staat om de verschillende grammaticale functies die woorden kunnen aannemen (bv. onderwerp en gezegde) te onderscheiden.

Ofwel werkt een vertaalmachine volgens statistische methodes. Het uitgangspunt is dan dat elke zin of elk deel van een zin in één taal een mogelijke vertaling is van een zin of een deel van een zin in een andere taal. Tegenwoordig bestaan er ook hybride vertaalmachines die de regelgebaseerde en de statistische aanpak combineren.

Ten slotte zijn er naast de reeds besproken volledig automatische machinevertaling (FAMT) en de machinevertaling die door de mens wordt ondersteund (HAMT) ook systemen ontwikkeld waarbij de computer de vertaler helpt bij het vertaalproces. Omgekeerd dus. We spreken dan van Machine Aided Human Translation (MAHT) of Computer-aided Translation (CAT). De brontekst wordt in dit geval vertaald door een menselijke vertaler, die een computer gebruikt om bijvoorbeeld de spelling te optimaliseren of de consistentie tussen verschillende vertalingen te bevorderen.

Over mijn onderzoek

Zoals gezegd zijn er vrij veel online vertaalmachines. Ik moest dus een selectie maken. Daarbij heb ik vooral gelet op de gratis beschikbaarheid en de bekendheid bij het brede publiek.

Selecteren

Uiteindelijk selecteerde ik de vertaalmachines Google Translate (statistische vertaalmachine), Bing translator (statistische vertaalmachine) en Intertran (regelgebaseerde vertaalmachine). Google Translate en Bing Translator liggen voor de hand. Google Translate is de meest gebruikte online vertaalmachine. Bing Translator is ontwikkeld door Microsoft, dat we kennen van Microsoft Office en de populaire tekstverwerker Word, waaraan Bing Translator is gekoppeld.

Hergebruiken

Mijn onderzoeksmethode heb ik zo eenvoudig mogelijk gehouden, zodat belangstellenden ze gemakkelijk kunnen hergebruiken om andere vertaalmachines te vergelijken of andere talencombinaties te evalueren.

Minicorpus

Om te beginnen heb ik een minicorpus samengesteld van negen Nederlandse zinnen die als input ter vertaling moesten dienen. Daarbij heb ik rekening gehouden met de typische problemen die Nederlandstaligen ondervinden als ze Spaans willen schrijven of spreken. Deze problemen lijken vooral betrekking te hebben op

woordenschat en grammatica. Wat woordenschat betreft heb ik onderzocht welke vertaalmachine het beste overweg kan met collocaties (kleine, vaste woordgroepen), realia (feiten of zaken die uniek zijn voor een land, taal of cultuur), het geslacht van een bepaald woord en de keuze van het adequate woord. Wat grammatica betreft heb ik onderzocht welke vertaalmachine het beste resultaat geeft als het gaat om congruentie (bv. overeenkomst tussen onderwerp en werkwoord), het gebruik van ser en estar, het gebruik van de wijzen (bv. indicatief, conjunctief), het gebruik van de tijden en het gebruik van voorzetsels. De uitkomst van mijn onderzoek laat dus tevens zien welke problemen het best worden aangepakt door welke vertaalmachine.

Objectief evalueren

Vervolgens heb ik de output van de verschillende machines vergeleken. Om een objectieve vergelijking te kunnen maken, heb ik het internationaal gerenommeerde puntensysteem van SAE International gebruikt. Dit puntensysteem houdt niet alleen rekening met het aantal fouten in een vertaling, maar ook met de zwaarte van die fouten. Zo krijgt elke soort fout een precies aantal punten. Vervolgens moet men al deze punten optellen en delen door het aantal woorden in de brontekst. De vertaling die de laagste score heeft, heeft dus de beste kwaliteit. Een volledige analyse van de Spaanse vertalingen is in de [eerste annex](#) bij dit artikel te raadplegen.

De uitkomst van mijn onderzoek laat dus tevens zien welke fouten het meest/minst gemaakt worden door welke vertaalmachine.

De kwaliteit van machinevertaling zelf verbeteren?

Hoe kun je als gebruiker zelf de output van een vertaalmachine verbeteren? Om die vraag te beantwoorden heb ik onderzocht of een voorvertaling van de oorspronkelijke Nederlandse input naar een andere inputtaal een gunstig effect heeft op de uiteindelijke Spaanse output. Concreet heb ik de effecten onderzocht van voorvertaling in het Engels en voorvertaling in het Frans.

Beter met Engels?

Mijn keuze voor het Engels is gebaseerd op de aanname dat vertaalmachines voor een belangrijke voertaal als het Engels een grotere databank hebben om gegevens uit op te halen, en dat ze bijgevolg correctere vertalingen zouden moeten kunnen leveren.

Of beter met Frans?

Voor het Frans heb ik gekozen omdat die taal tot dezelfde taalfamilie behoort als het Spaans. Mijn redenering is dat verwante talen meer grammaticale overeenkomsten moeten hebben, waardoor vertaalmachines correctere vertalingen zouden moeten kunnen leveren.

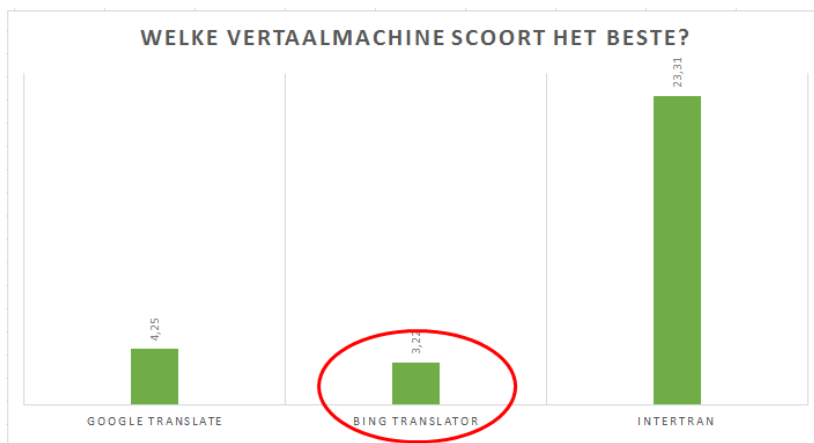
Zelf vertalen of met de machine?

Bij de Engelse en Franse input heb ik in mijn onderzoek ook een onderscheid gemaakt tussen enerzijds input vertaald door een vertaalmachine, en anderzijds input vertaald door een menselijke vertaler.

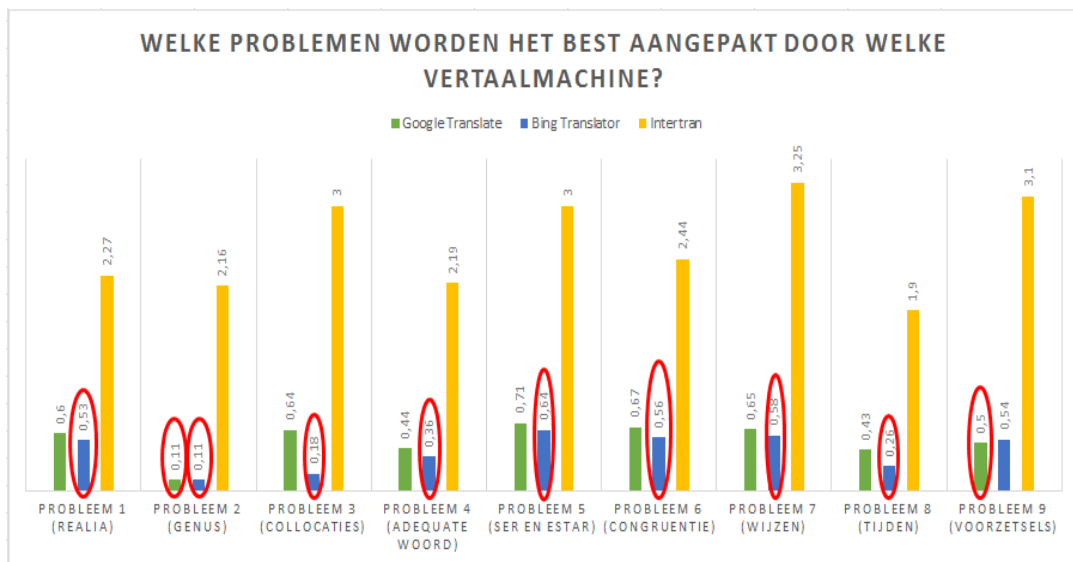
In dit artikel wil ik bruikbare tips geven om betere machinevertalingen te verkrijgen. Ik heb daarbij de gemiddelde gebruiker voor ogen: de doorsnee Nederlandstalige met een gemiddelde kennis van Engels en Frans. Omdat ikzelf zo'n gemiddelde gebruiker ben (en geen professionele vertaler), heb ik er zeer bewust voor gekozen om de zinnen zelf in het Engels en het Frans te vertalen in plaats van een beroep te doen op een professionele vertaler. Die zou wel beter Engels en beter Frans hebben geproduceerd maar hij kan niet model staan voor de gemiddelde gebruiker voor wie deze tips bedoeld zijn. Een volledige analyse van dit onderzoek naar mogelijke verbeter tips is in de [tweede annex](#) bij dit artikel te raadplegen.

En de beste is ...

Bing Translator produceerde het meest correcte Spaans bij een Nederlandse input.

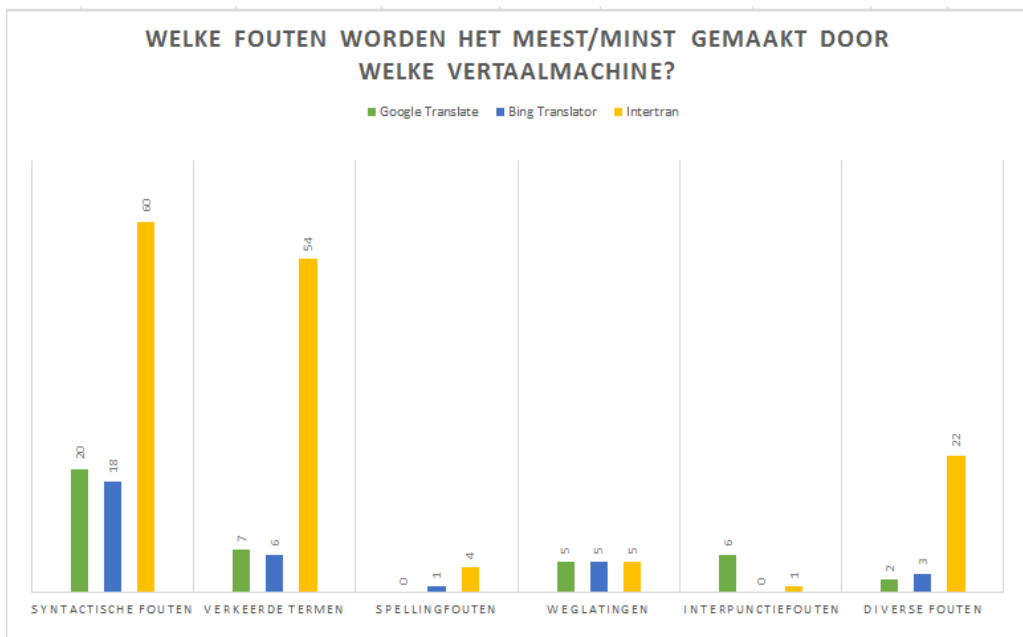


Als we ook in detail bekijken hoe de vertaalmachines scoren op de specifieke grammaticale en woordenschatproblemen die ik voor de Nederlandse inputzinnen selecteerde, blijkt dat het opnieuw Bing Translator is die al deze problemen het meest correct vertaalt. Nochtans zou je kunnen verwachten dat de grammaticale problemen beter vertaald zouden worden door Intertran, aangezien deze vertaalmachine regelgebaseerd is.



Bekijken we de problemen ten slotte afzonderlijk, dan kunnen we besluiten dat:

- syntactische fouten, fouten door verkeerde termen en interpunctiefouten die Nederlandstaligen typisch maken in het Spaans, het minst worden gemaakt door Bing Translator,
- typische spellingfouten en andere diverse typische fouten bij Nederlandstaligen het minst worden gemaakt door Google Translate,
- foute weglatingen in dezelfde mate voorkomen bij de drie vertaalmachines.



Voor de Spaanse vertaling van Nederlandse zinnen komt dus Bing Translator als beste gratis online vertaalmachine uit ons onderzoek.

Beter met Engels

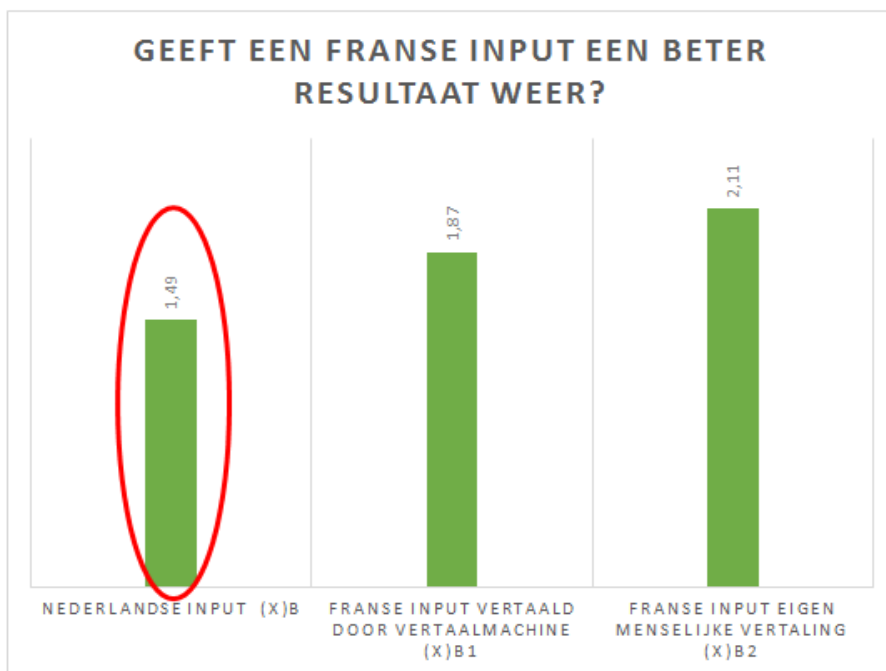
Ben je als doorsnee gebruiker bereid een inspanning te doen met het oog op een betere kwaliteit, dan geeft een voorvertaling naar het Engels een beter resultaat dan de oorspronkelijke Nederlandse input.

Frans helpt niet

Bovendien geeft de voorvertaling naar het Engels ook een correctere output dan een voorvertaling naar het Frans. Dit is op zich wel bijzonder, aangezien het Frans tot dezelfde taalfamilie als het Spaans behoort. Men zou dus bij deze laatste talencombinatie een beter resultaat kunnen verwachten.

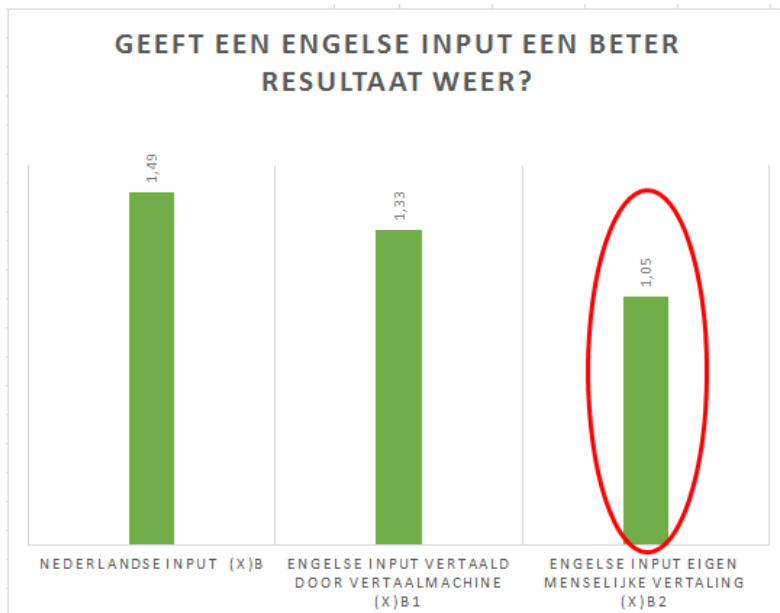
Frans helpt niet (2)

Nog opmerkelijk bij de Franse voorvertaling is dat de Franse voorvertaling door de vertaalmachine een beter resultaat in het Spaans opleverde dan mijn eigen Franse voorvertaling. Aangezien ik zoals gezegd geen professionele vertaler ben, bevat mijn Franse vertaling ongetwijfeld nog grammaticale fouten. Toch verwachtte ik dat mijn menselijke voorvertaling uiteindelijk een betere Spaanse vertaling zou opleveren dan een machinale voorvertaling.



Beter met Engels (2)

Wat betreft de Engelse input, scoorde mijn eigen voorvertaling wél beter dan de machinale voorvertaling.



Eindconclusie

Bing Translator geeft de beste Spaanse vertaling uit het Nederlands. De beste manier om met Bing Translator een nog betere vertaling te verkrijgen, is door je Nederlandse brontekst zelf in het Engels voor te vertalen – ook als je Engels niet top is.

Meer weten? Neem gerust contact op met Yanah Van Geerteruy via redactie@detaalsector.be