

De computer als scherprechter

Is de automatische taaltoets van Verdonk haalbaar?

Helmer Strik - Radboud Universiteit Nijmegen

Half maart veegden enkele deskundigen de vloer aan met de automatische taaltoets die minister Verdonk deze zomer wil invoeren voor immigranten. Wat houdt die toets precies in? En is er echt zo veel mis mee?

Op 1 juli dit jaar wil minister Verdonk van Vreemdelingenzaken en Integratie een inburgeringsexamen invoeren voor immigranten uit niet-EU-landen. Dat examen bestaat naast een toets over de Nederlandse samenleving uit een toets voor mondelinge taalvaardigheid. Beide toetsen moeten afgelegd worden in het land van herkomst, bij de ambassade of het consulaat. Het examen zal worden afgenomen via een telefoon die in verbinding staat met een centrale computer, en de antwoorden zullen automatisch verwerkt worden met behulp van spraakherkenningstechnologie. Het gehele proces zal trouwens automatisch verlopen – de computer velt zelfs het eindoordeel: voldoende of onvoldoende. Reclameren over dit eindoordeel is niet mogelijk. Deelname aan de toets gaat ongeveer €350,- kosten. Een voorwaarde voor toelating tot Nederland is dat beide toetsen met goed gevolg afgelegd zijn.

In maart dit jaar ontstond veel commotie toen enkele deskundigen hun twijfels uitten over de kwaliteit en de haalbaarheid van de automatische taaltoets. Hebben zij gelijk?

Zenuwachtig

De toets bestaat uit een aantal onderdelen. Allereerst krijgen de kandidaten via de telefoon een aantal zinnestjes te horen die ze moeten nazeggen. Een voorbeeld van zo'n zinnestje is: 'Het is niet helemaal gegaan zoals we hadden verwacht.' Ten tweede moet de kandidaat enkele eenvoudige vragen beantwoorden, zoals: 'Als je thee zet, gebruik je dan heet water of koud water?' Ten slotte moet de geëxamineerde van een aantal woorden het tegengestelde woord verzinnen: bij 'wit' moet 'zwart' worden gezegd, bij 'hoog' 'laag', etc.

De computer controleert alle antwoorden, zowel dat wat nagezegd is als de antwoorden op de eenvoudige vragen. De software moet dus bepalen of de mondelinge taalvaardigheid voldoet aan het vereiste niveau, het zogenoemde A1-minniveau, dat omschreven is door de commissie-Franssen (zie het kader 'Het niveau van de toets'). Maar hoe goed kan de computer dat?

Allereerst zou je je kunnen afvragen of een zo laag niveau als het voorgestelde A1-min wel toetsbaar is door een computer. In het rapport van de commissie-Franssen wordt de spraak op A1-minniveau omschreven als: "De uitspraak is vaak onbegrijpelijk en wordt sterk beïnvloed door de moedertaal. Een spreker op dit niveau kan slechts met zeer grote moeite begrepen worden." Als die spraak voor een mens al problematisch is, hoe zou een computer het dan beter kunnen doen?

Daar komt bij dat in de na te zeggen uitingen van de computer veel woorden en constructies zitten die de kandidaten nog nooit eerder gehoord hebben, in ieder geval niet als ze alleen in eigen land les hebben gehad op niveau A1-min. Vaak zullen ze niet weten hoe ze daar adequaat op moeten reageren. Bedenk hierbij dat het gaat om mensen die moeten proberen om in hun eigen land Nederlands te leren, veel geld moeten betalen om aan de toets te kunnen deelnemen, in een voor hen meestal vreemde omgeving (consulaat of ambassade) een telefoon in de hand gedrukt zullen krijgen, en dan vervolgens, vaak zenuwachtig, via de telefoon moeten gaan praten met een computer in een heel ander deel van de wereld.

Verfzaak

Een andere vraag is of deze taaltoets wel test wat hij moet testen. Eigenlijk test hij alleen een reactie op dingen die de computer zegt. Wat je eigenlijk zou willen testen, is of iemand zichzelf kan redden in onze maatschappij. Wezenlijke kenmerken van een normale conversatie ontbreken in de toets. Stel, je gaat naar een verfzaak, en de verkoper vraagt ergens in het gesprek: 'Wit?' Dan is er een enorme variatie in wat je kunt antwoorden. Zeg je bijvoorbeeld 'Wit? Nee, ik bedoel eh, z-z-zwart ja', dan zal dit geen enkel probleem opleveren, zelfs als je het woord *zwart* met een zwaar accent uitspreekt, bijvoorbeeld als 'z-z-zwert' of 'sjwoart'. Zelfs het aan elkaar plakken van twee woorden ('z-z-zwer-tja') komt in alledaagse spraak veel voor. De automatische spraakherkenner van de taaltoets zal het in veel van deze gevallen moeilijk hebben. En wat erger is: in de automatische taaltoets ontbreken vele mogelijkheden van een normale conversatie. Je kunt niet zeggen 'Wat zegt u?', of 'Wit?' Het welslagen van een gewoon gesprek hangt voor een groot deel af van het gebruik van deze technieken, en die kunnen bij de taaltoets niet gebruikt worden.

Treinreizen

De deskundigen betwijfelen of de huidige spraaktechnologie ingezet kan worden voor de taaltoets. Hoe zit dat?

Een spraakherkenner zet gesproken woorden om in geschreven tekst. Ofschoon spraakherkenners – net als mensen – af en toe fouten maken, kan de huidige technologie heel goed gebruikt worden voor het maken van nuttige toepassingen. De tekst die u hier leest, heb ik bijvoorbeeld gedictreed in plaats van getypt. Dicteerprogramma's moeten echter wel uitgebreid getraind worden op het woordgebruik en de uitspraak van één gebruiker. Dan nog zijn ze bepaald niet foutloos, maar fouten kunnen snel gecorrigeerd worden en het programma leert in de loop van de tijd bij. Dicteerprogramma's zijn een voorbeeld van 'sprekerafhankelijke' spraakherkenning.

Sprekerónafhankelijke spraakherkenning (geschikt voor een willekeurige gebruiker) is al een stuk moeilijker. Het aantal herkenfouten neemt hierbij snel toe. Daarom wordt deze technologie meestal alleen gebruikt voor beperkte domeinen. In Nederland bestaan er bijvoorbeeld toepassingen waarbij via de telefoon informatie opgevraagd kan worden over treinreizen, beurskoersen, nieuws en files.

Voorspellen

In de taaltoets wordt sprekerónafhankelijke spraakherkenning gebruikt. Dat lijkt op het eerste gezicht misschien wel te kunnen, want de goede antwoorden liggen voor een groot deel vast. Zinnetjes moeten worden nagezegd, of er is maar een beperkte variatie mogelijk. Als je het tegengestelde van *wit* moet noemen, moet het woord *zwart* in je antwoord zitten. De spraakherkenningstechniek is meestal behoorlijk in staat om in een willekeurige zin het woord *zwart* te herkennen (de zogeheten wordspottingstechniek). Toch zullen uitspraakvarianten (de al eerder genoemde zware accenten) voor problemen zorgen, en verder zijn er ook praktische problemen. Stel, de reactie op 'Wit' is: 'Zwart? Nee, vuil', dan zal de computer dit antwoord ten onrechte goed rekenen.

Als een spraakherkenner op basis van wordspotting niet goed genoeg werkt, kun je ook proberen het hele antwoord te laten herkennen. Dan krijg je echter het probleem dat je moet aangeven wat mensen kunnen antwoorden op de vraag naar het tegengestelde van het woord *wit*. Soms antwoorden mensen niet alleen met het woord *zwart*, maar met een uiting als 'Eh, eh, zwart.' Je kunt de spraakherkenner in beperkte mate leren om dit soort loze toevoegingen toe te staan, maar om dat goed te kunnen doen, moet je van tevoren goed kunnen inschatten wat de mensen zullen gaan zeggen tegen de computer in de taaltoets. En dat is heel moeilijk te

voorspellen. Dit is al moeilijk met autochtonen, maar zeker met allochtonen met een zo grote diversiteit in achtergronden.

Spraakgebrek

Behalve mensen met zware accenten (die op dit taalniveau toegestaan zijn) krijgen ook mensen met een spraakgebrek, bijvoorbeeld stotteraars, het moeilijk met de taaltoets. Verder zullen alle geluiden die de computer binnenkrijgt van invloed zijn op wat herkend wordt, bijvoorbeeld echo's, kraak- en andere stoorsignalen op de telefoonlijn, muziek, pratende mensen, slaande deuren en andere achtergrondgeluiden.

Ook de verdere omstandigheden waaronder de toets wordt afgelegd kunnen beslissend zijn. Krijgen de mensen vooraf duidelijke instructies? Kan het personeel op de ambassade of consulaat helpen als dat nodig is, of doen zij niet meer dan het aangeven van de telefoon? Dit is belangrijk omdat er vaak iets mis blijkt te gaan als mensen voor de eerste keer via de telefoon met een computer praten.

Er zijn dus nogal wat factoren die ertoe kunnen leiden dat de spraakherkenner veel fouten maakt. Toch zijn de resultaten van de spraakherkenner de enige basis voor de computer om tot een eindoordeel te komen: voldoende of onvoldoende, toegelaten of niet. Als ik een afgewezen kandidaat was, zou ik wel graag willen weten welke fouten ik precies gemaakt heb, en waar de computer mij misschien verkeerd verstaan heeft.

Hachelijk

Het onderliggende spraakherkenningssysteem voor de taaltoets heeft in het verleden zijn waarde wel bewezen, maar het is voor iets heel anders ontwikkeld. Het systeem (PhonePass genaamd) is in staat om een indicatie te geven van je taalniveau, van A1 (beginner) tot C2 (zeer gevorderd). Dat gaat heel aardig. Ben je een spreker van niveau B1, dan zal het oordeel van het systeem ook in de buurt van B1 liggen. Ben je A1, dan zal het systeem je heus niet als C2 karakteriseren. Toch is de taaltoets fundamenteel anders: het systeem moet een scherpe grens trekken tussen net boven of net onder niveau A1-min. Je kunt het een beetje vergelijken met leeftijden schatten. Het is niet zo moeilijk om vast te stellen dat iemand ongeveer veertig is, of drie, maar lastiger wordt het om van kinderen tussen de vijf en de tien jaar te zeggen wie er ouder dan zeven is. De taaltoets gebruikt een systeem voor een globale schatting om een scherpe grens te trekken. Dat is uitermate hachelijk, en het moet inderdaad, zoals de minister aankondigt, geëvalueerd worden – iets wat momenteel gebeurt. In mei verschijnen hierover twee rapporten.

Ofschoon het bij de huidige stand van zaken in de spraaktechnologie zeker mogelijk is om goede en nuttige toepassingen te maken, is het twijfelachtig of die technologie al geschikt is voor een dergelijke automatische taaltoets. Vele deskundigen hebben in de media hun twijfels geuit. De minister, haar woordvoerders en mensen die bij de ontwikkeling van deze toets betrokken zijn, verweren zich tegen de kritiek door te beweren dat wetenschappelijk onderzoek aangetoond heeft dat de toets betrouwbaar is. Maar tot dusver is dat onderzoek niet gepubliceerd. De twijfels die er bestaan, kunnen alleen weggenomen worden door een geheel onafhankelijk onderzoek.

Belangrijk hierbij is wel dat onafhankelijke instanties het complete onderzoek doen (data verzamelen, analyseren en resultaten publiceren), dat dit gebeurt voor een representatieve steekproef, en dat het systeem gebruikt wordt op dezelfde manier waarop het later ook gebruikt zal worden. Als een gedeelte van dit onderzoek, bijvoorbeeld het verzamelen van de data, wordt uitgevoerd door belanghebbenden, dan heeft dit onderzoek weinig waarde. Ik ben benieuwd naar de resultaten.

[Noot:]

De Tweede Kamer heeft 5 april ingestemd met de taaltoets voor immigranten, maar trekt zijn steun in zodra een van de in mei te verschijnen rapporten daartoe aanleiding geeft.

xxxxxxxxxxxxxxxxxxxx begin kader xxxxxxxxxxxxxxxxxxxxxxx

Het niveau van de toets

In veel landen wordt gebruikgemaakt van het zogenoemde Europees Raamwerk voor Moderne Vreemde Talen, waarmee gemakkelijk bepaald kan worden op welk niveau iemand een vreemde taal beheerst. In dat systeem worden zes taalniveaus onderscheiden: A1, A2, B1, B2, C1 en C2, waarbij dat laatste niveau geldt voor zeer vaardige gebruikers van de nieuwe taal en A1 voor mensen die alleen de basis van de taal beheersen. Zo kan iemand zich bij A1 voorstellen in de nieuwe taal. Verder kan hij iemand vragen naar zijn woonplaats, naar kennissen of naar bezittingen, en hij kan – als hij een geduldige, hulpvaardige gesprekspartner heeft – zulke vragen ook beantwoorden.

Toch is dat A1-niveau nog te hoog voor de immigranten die de taaltoets moeten doen. Een commissie van deskundigen, de commissie-Franssen, heeft daarom voorgesteld om het niveau A1-min te hanteren, en de minister heeft dit voorstel overgenomen. Dit A1-minniveau is door de commissie-Franssen zelf afgeleid van het A1-niveau, en wordt omschreven als:

- Men kan slechts een beperkt aantal vertrouwde woorden en basiszinnen begrijpen die betrekking hebben op de directe, persoonlijke levenssfeer en op de allereerste levensbehoeften; en alleen in direct contact met Nederlandssprekenden die gewend zijn zich aan te passen.
- Men kan zich slechts in zeer beperkte mate uitdrukken, eigenlijk alleen met behulp van losse woorden en standaardformuleringen ('formulaic speech'), op een gering aantal terreinen die verband houden met de directe, persoonlijke levenssfeer.

[noot:]

Bron: *Inburgering getoetst. Advies over het niveau van het inburgeringsexamen in het buitenland* (rapport van de commissie-Franssen). Te vinden op www.justitie.nl/Images/inburgering_getoetst_tcm35-46016.pdf.

xxxxxxxxxxxxxxxxxxxx einde kader xxxxxxxxxxxxxxxxxxxxxxx