

REPETITOR

op weg naar geautomatiseerde uitspraakcorrectie

ALIED BLOM EN HENK VAN DEN HEUVEL

Voor het groepsonderwijs Nederlands als tweede taal vormt correctie op de uitspraak een lastig probleem. Al die cursisten met hun eigen uitspraakproblemen: hoe kun je die allemaal in een plenaire behandeling tot hun recht laten komen? Bovendien is correctie alleen niet genoeg, maar is ook intensieve oefening geboden, en dat valt groepsgewijs moeilijk te realiseren. Onder de naam Dutch-CAPT is door de Radboud Universiteit Nijmegen een programma ontwikkeld dat voorziet in automatische correctie van specifieke aspecten van de uitspraak. In samenwerking met de Radboud Universiteit is door de Technische Universiteit Delft een vervolgproject opgestart onder de naam REPETITOR. Dit project bevindt zich in de ontwikkelingsfase; in een later stadium volgen toetsing en evaluatie. Hieronder een vergelijking tussen doelstelling, opzet en didactische inhoud van Dutch-CAPT en REPETITOR.

Geautomatiseerde uitspraakcorrectie

Geautomatiseerde correctie op de uitspraak, wat een voordelen kan dat opleveren! Het verbeteren van de uitspraak is

dan niet langer uitsluitend de taak van de docent, maar wordt gedeeltelijk overgenomen door een computerprogramma. In de groepsles blijft dan meer tijd over voor wat men het meest zinvol in groepsverband kan doen: praten. Bovendien krijgen de cursisten de correctie die toegesneden is op hun specifieke uitspraakproblemen, en kunnen ze op die speciale probleempunten zo lang en zo vaak ze willen gaan oefenen, buiten gehoorsafstand van meer bedreven medecursisten. In de cursussen Nederlands volgens de Delftse methode vormt praten een zeer belangrijk element, zodat automatische correctie op de uitspraak ook binnen deze methode een uitkomst zou zijn. Maar is het technisch mogelijk? Is de spraaktechnologie al zo ver?

Dutch-CAPT: het Nijmeegse project.

Onder de naam Dutch-CAPT (Computer Assisted Pronunciation Training) is aan de Radboud Universiteit Nijmegen in 2007 een programma ontwikkeld (<http://lands.let.ru.nl/~strik/research/Dutch-CAPT/index.html>) waarin spraaktechnologie ingezet wordt om leerders van het Nederlands te

helpen hun uitspraak te verbeteren. Het basisprincipe is dat de uitingen van de leerders gecontroleerd worden op een aantal potentiële uitspraakfouten, op grond waarvan een goed/foutmelding afgegeven wordt, en de gebruiker het nog eens kan proberen.

Voor de selectie van potentiële uitspraakfouten is uitgegaan van een verzameling nagesproken en spontaan ingesproken uitingen van NT2-leerders, voor het merendeel van halfgevoerd tot gevorderd niveau, met in totaal 14 moedertalen. Hierin is een tiental hardnekkige, frequente uitspraakfouten geselecteerd, die duidelijk waarneembaar en potentieel storend zijn voor de communicatie. Vervolgens zijn akoestisch-fonetische kenmerken gezocht op grond waarvan deze fouten te onderscheiden zijn van de correcte uitspraak of uitspraakvarianten van de desbetreffende klanken. Dit is voor het merendeel van de geselecteerde klanken gelukt. De gebruikers die in een van de geselecteerde klanken een fout maakten, kregen inderdaad de juiste feedback, en effectiviteits-onderzoek wijst uit dat er uiteindelijk vooruitgang geboekt werd op de uitspraak van die klanken (Neri, 2007).

REPETITOR: het Delftse project.

In 2008 is aan de Technische Universiteit Delft in samenwerking met de Radboud Universiteit Nijmegen het project REPETITOR van start gegaan (Blom & Van den Heuvel, 2008). REPETITOR is gebaseerd op het Nijmeegse Dutch-CAPT project, maar werkt in tegenstelling tot het Nijmeegse programma uitsluitend met nagesproken tekst. Een stukje tekst verschijnt zin voor zin op het scherm, de gebruiker beluistert de voorbeelduitspraak en spreekt de zin vervolgens in. De feedback bestaat uit

een simpele foutmelding in de vorm van een onderstreping van de fout uitgesproken letter(s); daarna krijgt de gebruiker de gelegenheid om de eigen productie te beluisteren en die te vergelijken met de voorbeelduitspraak. Gescherpt door deze combinatie van visuele en auditieve feedback gaat de gebruiker een of meer nieuwe pogingen wagen.

Ook voor dit project is een dataverzameling aangelegd van in totaal 1972 zinnen, ingesproken door 44 voor het merendeel beginnende sprekers, met in totaal 21 moedertalen. De criteria op grond waarvan deze trainingsdata bijeengebracht zijn, waren anders dan bij het Nijmeegse project, omdat doelgroep en didactische inhoud van REPETITOR verschillen van die van Dutch-CAPT. We gaan nu eerst in op deze twee aspecten.

Beginners of (half)gevoerden?

Wie hebben de meeste behoefte aan zo'n programma, beginners, halfgevoerden of gevorderden? Het Nijmeegse programma Dutch-CAPT zet in op gevorderden. Een van de redenen hiervoor is, dat veel beginnersfouten na verloop van tijd vanzelf verdwijnen, ook zonder (al of niet geautomatiseerde) feedback. Automatische correctie van hardnekkige fouten zoals die zich bij meer gevorderden voordoen, lijkt dus in een grotere behoefte te voorzien. Toch hebben wij voor REPETITOR beginners als doelgroep gekozen. De reden hiervoor is dat wij dit programma willen integreren in het leer materiaal van de Delftse methode. In deze methode wordt vanaf het begin veel aan praten spreken gedaan, zodat vanaf dat eerste begin een redelijke uitspraak essentieel is. Door de uitspraakcorrectie gedeeltelijk buiten de les plaats te laten plaats vinden, houden we vanaf het begin meer tijd over voor het leerzaamste onderdeel van elke

taalles, namelijk het oefenen van het praten. En natuurlijk is een geautomatiseerde uitspraak-oefenmodus voor degenen die de taal via zelfstudie leren en geen cursus volgen, de enige mogelijkheid om feedback te krijgen op de uitspraak.

Spontane taal of nagesproken zinnen?

Op het eerste gezicht lijkt het aantrekkelijk om het programma spontane taal te laten beoordelen, omdat je dan de vinger kunt leggen op fouten die zich ook voordoen bij natuurlijke interactie. Dutch-CAPT werkt dan ook met dialoogjes, waarin machine en gebruiker afwisselend een beurt nemen. Maar een gegeven bij spontane spraak is dat sprekers geneigd zijn om moeilijke klanken te vermijden, terwijl het voor het effectiviteitsonderzoek naar Dutch-CAPT noodzakelijk was dat de gebruikers alle tien klanken waar het programma op getraind is, in voldoende mate zouden gebruiken. Dit probleem werd opgelost door de gebruikers te laten kiezen uit drie voorgedrukte zinnen, waarin die geselecteerde klanken allemaal voorkomen. De notie 'spontaan' wordt hierdoor wel ingeperkt, en ook doen de dialoogjes niet altijd erg natuurlijk aan. Maar een veel groter nadeel van deze opzet is dat de omvang van het gesproken materiaal noodzakelijkerwijs beperkt blijft. Het is al een hele toer om een dialoogje met drie beurtwisselingen in elkaar te zetten, en dan komt de helft van de spreekstof ook nog voor rekening van de machine.

REPETITOR werkt daarom met nagesproken zinnen. De gebruiker beluistert een voorbeeldzin, desgewenst meerdere malen, en spreekt die zo goed mogelijk na. Hiervoor zijn de eerste tien teksten van het eerste deel van de Delftse Methode gebruikt, wat ons niet minder dan 1972 zinnen opleverde. Werken met het Delftse

lesmateriaal heeft voor ons het extra voordeel dat wij door middel van feedback op de uitspraak extra sturing kunnen geven aan het leergedrag van onze cursisten. Wij dragen hun immers op om, vóórdat ze naar de les komen, het tekstmateriaal grondig te beluisteren en na te spreken. Maar pas in de les blijkt of ze deze opdracht goed uitgevoerd hebben, als ze hun bijdrage moeten leveren aan de conversatie. Als ze al tijdens de voorbereiding feedback op hun uitspraak krijgen, weten ze meteen of ze goed bezig zijn. Slecht scoren op de uitspraak doordringt ze er in een veel eerder stadium van dat ze niet goed, of in elk geval niet effectief, aan het luisteren zijn. Geautomatiseerde uitspraakcorrectie kan cursisten helpen de juiste studiemethode te kiezen wat betreft het luisteren. Hieronder meer over de foutenselectie waar REPETITOR op gebaseerd is.

Uitspraakcorrectie: een complexe aangelegenheid

Zowel Dutch-CAPT als REPETITOR werken met een beperkt aantal potentiële fouten waarop de ingesproken zinnen gecorrigeerd worden. Een team van docenten NT2 aan de TU Delft is de verzameling trainingsdata gaan beluisteren met het oog op een verantwoorde foutenselectie. Wat wij deden lijkt natuurlijk op de manier waarop wij in de les de uitspraak van onze cursisten corrigeren: ook daar maken we keuzes, leggen we niet op elke slak zout, corrigeren we niet elke afwijking van de norm waaruit blijkt dat de gebruiker een andere moedertaal heeft dan het Nederlands. Alleen: wij maakten onze selectie buiten de les, in de rust van onze studeerkamer, met de mogelijkheid tot herhaald afluisteren. Zo werden wij ons ervan bewust hoe lastig het eigenlijk is om iemands uitspraak op het moment

zelf, in de les dus, te beoordelen. Hoe vaak moesten wij niet een fragment meer dan eens beluisteren voordat wij konden beslissen of er werkelijk sprake was van een fout, en zo ja, waar de fout hem dan precies in zat? Zei de spreker luk in plaats van leuk of oek in plaats van ook, of hadden we te doen met heel kort uitgevallen *eu's* en *oo's*? Hoorden wij een *e-tje* voor de *s* in *boodschappen*, of speelde de stemhebbend uitgesproken *d* ons parten?

Bovendien: wat een verscheidenheid aan fouten waaruit de docent moet kiezen! Is het ingevoegde *e-tje* in *boódeschappen* het belangrijkste, moeten we beginnen met de stemloosheid van de *d*, of gaan we eerst de klemtoon aanpakken? Hoe maakt de docent zo gauw uit welke fout van voldoende algemeen belang is om plenair te gaan corrigeren en welke fouten hij voorlopig maar even laat zitten? Hoe kan een docent dit onder de tijdsdruk en in de hectiek van een groepsles zo snel uitmaken? Een goed werkende geautomatiseerde uitspraakcorrectie die docenten deze last gedeeltelijk van de schouders kan nemen: niet alleen handig en tijdsbesparend, maar, zo beseften wij, een absolute noodzaak.

Maximering van het aantal fouten per zin?

Een vraag die hierbij rijst is of REPETITOR misschien zo opgezet moet worden dat er een maximum is aan het aantal foutmeldingen per zin. Worden beginners niet al te zeer ontmoedigd als ze meerdere fouten in één zin aangemeld krijgen? Hoe bepalen ze aan welke van al die fouten ze eerst gaan werken? Is de feedback dan nog wel effectief?

In Dutch-CAPT wordt om deze reden het aantal foutmeldingen per zin inderdaad beperkt tot drie. Maar zachte heelmesters, weet u wel? De beginnende leerder die niet

elke fout aangemeld krijgt, kan verkeerde conclusies trekken: 'Aan m'n *r* moet ik nog wat sleutelen, maar met m'n *g* zit het gelukkig wel goed!'. Bovendien verwachten vooral beginners niet anders dan dat ze veel fouten maken. Bij een beperking van het aantal foutmeldingen zouden ze het vertrouwen in het instrument kunnen verliezen. Om deze redenen is bij REPETITOR het aantal foutmeldingen niet gemaximeerd, en het is aan de gebruikers zelf om te kiezen welke fout(en) ze het eerst gaan aanpakken: de eerste, de moeilijkste, of juist de makkelijkste? Op één punt echter tonen ook wij compassie met de gebruiker. In tweede en volgende pogingen van één zin worden geen nieuwe fouten gemeld, maar wordt de uiting alleen op die punten beoordeeld waarop bij de eerste poging een fout was geconstateerd.

Foutenselectie: hardnekkige, communicatie-belemmerende of storende fouten?

Zoals hierboven gesteld was bij Dutch-CAPT de hardnekkigheid van fouten een belangrijk criterium. Veel fouten verdwijnen immers na verloop van tijd vanzelf, ook zonder correctie. De letter *e* bijvoorbeeld, die op drie manieren uitgesproken kan worden, is in het begin een frequente foutenbron, maar met de toename van de woordenschat doen veel van deze fouten zich niet meer voor. Toch zullen veel docenten in hun beginnerslessen het daar niet altijd op aan laten komen. Als een beginnende cursist het woord *regering* als *régering* uitspreekt, zullen zij dit waarschijnlijk toch direct corrigeren, om in een vroeg stadium af te rekenen met dit soort storende fouten. En zo zijn wij ook bij onze foutenselectie werk gegaan: REPETITOR is voor beginners bedoeld, en detecteert met name typische beginnersfouten.

Een ander criterium dat bij Dutch-CAPT een rol heeft gespeeld is begrijpelijkheid. Speciaal die fouten werden geselecteerd die de uiting onbegrijpelijk maken en zo de communicatie belemmeren. Net zoals docenten NT2 er in hun lessen waarschijnlijk naar streven om vooral op begrijpelijkheid te corrigeren, en niet op perfectie. Maar handelen ze altijd hiernaar? Laten zij *gemàchelech* in plaats van *gemakkelijk* passeren, met de *k* als *ch* gerealiseerd, alleen omdat het zo volkomen duidelijk is wat de spreker bedoelt? Zien zij *stad* uitgesproken met een stemhebbende *dentaal* op het eind door de vingers, omdat die fout de communicatie niet belemmert? Of zetten zij dit soort storende fouten toch meteen even recht?

Wij zijn bij onze selectie van uitspraakfouten voor REPETITOR te werk gegaan zoals wij ook in onze lespraktijk te werk gaan. Vier docenten NT2 aan de TU Delft hebben in voortdurend overleg met elkaar die fouten geselecteerd in de verzameling trainingsdata die wij in onze lessen als storend zouden ervaren. Fouten waarvan wij zouden vinden dat we die als docent niet kunnen laten passeren omdat dat onze eerte na is. Het op deze manier beluisteren van onze trainingsdata leverde ons het grote aantal van tegen de honderd verschillende uitspraakfouten op.

Daaronder bevinden zich dus heel wat fouten die vanzelf verdwenen zouden zijn in latere stadia van het leerproces. Ook zijn afwijkingen in de uitspraak op onze foutenlijst komen te staan, die geen probleem vormen voor de communicatie. Een voorbeeld zijn de talloze gevallen van spellinguitspraak in onze data. Zo wordt de slotlettergreep van *negentig* vaak met een korte *i* uitgesproken, en die van *eigenlijk* met een *ij*, ook al is in de voorbeelduitspraak in beide gevallen de korte 'stomme *e*' hoorbaar. En ook al horen de insprekers 't of 'n, met

een 'stomme *e*', ze zeggen toch vaak het en *een*, en in plaats van de gereduceerde vormen *me*, *je*, *we*, *ze* wordt hardnekkig *mij*, *jij*, *wij* en *zij* ingesproken. Natuurlijk begrijpt iedereen de bedoeling van deze sprekers zonder enige moeite, maar men zal het erover eens zijn dat een uiting met een overdaad aan 'volle' vormen en gevallen van spellinguitspraak al gauw onnatuurlijk klinkt. En omdat wij onze cursisten in een vroeg stadium willen dwingen tot kritisch luisteren naar het uitspraakvoorbeeld, hebben wij deze afwijkingen op de lijst van potentiële uitspraakfouten gezet.

De door REPETITOR te detecteren uitspraakfouten

De grootste groep op de aldus verkregen lijst uitspraakfouten vormen de deleties: *Nederlan* in plaats van *Nederland*, *hij ga* voor *hij gaat*, *mij voor mijn*, *groeten* in plaats van *groenten*, enzovoort. De tweede groep vormen de inserties, zoals *wordent*, *Encheland/Enkeland*, *boodeschappen*. Hierna komen de substituties, zoals *maggelig* voor *makkelijk*, *klant* voor *krant*, *proeg* voor *vroeg*, *boiten* voor *buiten*, *noe* voor *nu*, en *jij wij* en *zij* voor *je*, *we*, *ze*. Een geval apart vormen gevallen van metathesis, zoals *drop* voor *dorp*, *strek* voor *sterk*, *kegocht* voor *gekocht*, en zoals te verwachten zaten er ook klemtoonfouten in onze data: *boodschappen*, *règering*.

Niet alle afwijkingen in de uitspraak op deze lijst zijn met de huidige stand van de spraaktechnologie ook inderdaad detecteerbaar. Zo zijn inserties moeilijker te detecteren dan deleties, omdat – voor de spraakherkenner in de computer – de aanwezigheid van wat niet verwacht wordt, lastiger vast te stellen is dan de afwezigheid van iets wat wel verwacht wordt. Bij deleties bepaalt de akoestisch-fonetische afstand tussen gedeleerde klank en de context de

detecteerbaarheid, en evenzo gaat het bij substituties om de akoestisch-fonetische afstand tussen de twee verwisselde klanken. Detectie van klemtoonfouten vergt een andere techniek dan de overige typen,

terwijl ook de frequentie van dit type fouten te laag was: derhalve zal REPETITOR klemtoonfouten ongemoeid moeten laten. De lijst met fouten waar REPETITOR uiteindelijk op getraind is, ziet er als volgt uit.

1. deleties

- deletie van *t* (*Nederlan*, *hij ga*)
- deletie van *h* ('*t*', '*m*' in plaats van *het*, *hem*)
- deletie van *s* (*vakantie*)

2. inserties

- insertie van *t* (*wordent*, *ik gaat*)
- insertie van *ch* of *k* in de *ng*-klank (*Encheland*, *jonchen*, *Enkeland*, *jonken*)
- insertie van *h* voor de 'stomme *e*' (*het*, *hem* in plaats van '*t*', '*m*')

3. substituties

- substitutie van de 'stomme *e*' door *ee*, *ie*, *ij*, *è* (*jee*, *wee*, *mie*, *die*, *makkelijk*, *jij*, *het* (in plaats van *je*, '*t*'))
- substitutie van *ui* door *ou*, *oe*, *ò*, *oj* (*hous*, *out*, *roemte*, *oet Spanje*, *ot Spanje*, *sòker*, *ojt*, *bojten*)
- substitutie van *ie* door *aj*, 'stomme *e*' (*draj*, *precajs*, *famelie*, *de iplv die*)
- substitutie van *è* door *aj*, *aa* (*we haben lajs*)
- substitutie van *k* door *ch* (*machelech*, *ich*, *ooch*, *gechocht*)
- substitutie van *ch* door *k* (*we kaan*, *kroen*, *onkeveer*, *skool*)
- substitutie van *p* door *f* (*frettig*, *van flan*, *ik hef*)
- substitutie van *aa* door *à* (*gan*, *nar huis*)
- substitutie van *ù* door *oe* (*boes*, *koersoos*)
- substitutie van *uu* door *oe* (*noe*, *drie oer*)
- substitutie van *eu* door *oo* (*in de koken*)
- substitutie van *v* door *p* of *b* (*proeg*, *toepal*, *bandaag*, *bakantie*)
- substitutie van *b* door *w* (*weginnen*, *hewwen*)

Op dit moment zijn wij een prototype van REPETITOR aan het testen. Met dit prototype moeten wij in staat zijn te bepalen hoe vaak de computer een uitspraakfout niet vindt en hoe vaak hij ten onrechte een uitspraakfout meldt. We kunnen dan nagaan of dit in een acceptabele verhouding staat tot het aantal keren dat de computer ede juiste beslissing neemt. Vervolgens stellen wij een gebruiksversie van REPETITOR ter beschikking aan onze cursisten, waarna het onderzoek naar de effectiviteit van geautomatiseerde uitspraakcorrectie van start kan gaan.

LITERATUUR

Blom, A., & Heuvel, H. van den (2008). REPETITOR: een uitspraakcoefenprogramma in ontwikkeling. *DIXIT Tijdschrift voor Taal- en Spraaktechnologie* 5,2, 30

Neri, A. (2007). *The pedagogical effectiveness of ASR-based Computer Assisted Pronunciation Training*. Academisch Proefschrift. Nijmegen: Radboud Universiteit Nijmegen.

ALIED BLOM is sinds medio jaren tachtig docent Nederlands voor Buitenlanders aan de TU Delft. E-mail: <alied.blom@tudelft.nl>.

HENK VAN DEN HEUVEL is verbonden aan het CLST (Centre for Language and Speech Technology), Radboud Universiteit Nijmegen. E-mail: <H.vandenHeuvel@let.ru.nl>.

SIGNALEMENT

Elsen, A. (2009) *Testing for Autonomy. Three exploratory case studies of the language testing practices of degree-one teachers of English in the fourth forms of upper secondary education in the Netherlands*. Academisch Proefschrift. Nijmegen: Radboud Universiteit Nijmegen.

De afgelopen decennia werden docenten in moderne vreemde talen in Nederland geconfronteerd met drie ingrijpende vernieuwingen: een taaldidactische, te weten communicatief taalonderwijs; een pedagogische en wel zelfstandig leren en een organisatorische in de vorm van de Tweede Fase. Dateert de eerste vernieuwing al uit de tachtiger jaren van de vorige eeuw, het is vooral de gelijktijdige invoering van Zelfstandig Leren en de Tweede Fase in de negentiger jaren van diezelfde eeuw die een grote impact heeft gehad op docenten moderne vreemde talen in het algemeen. Daarover gaat het proefschrift van Adri Elsen die bij drie ervaren eerstegraadsdocenten Engels, twee mannen en een vrouw, heeft onderzocht hoe hun opvattingen over en ervaringen met de invoering van die vernieuwingen in hun onderwijs waren. De focus van het proefschrift ligt daarbij op de impact van die vernieuwingen op de toetsing- en evaluatiepraktijk van deze docenten.

De onderzoeksvragen waren: (1) Wat vinden eerstegraads docenten Engels van actief en zelfstandig leren, communicatief taalonderwijs en de rol van beoordeling en evaluatie? (2) In hoeverre vindt men de beliefs van de respondenten terug in hun beoordelings- en evaluatiepraktijk? (3) Welke aanbevelingen kunnen worden gedaan voor het stimuleren van leeromgevingen waarin toetsen een positief terugslag-effect hebben op de wijzen

waarop leerlingen leren communiceren in het Engels?

Het proefschrift is in uitstekend en heel leesbaar Engels geschreven en beslaat twaalf hoofdstukken. De eerste twee hoofdstukken leiden het onderwerp en de methode van onderzoek in en de hoofdstukken drie, vier en vijf bevatten achtereenvolgens een uitgebreide bespreking van theorie en praktijk van zelfstandig leren, communicatief taalonderwijs en toetsing en evaluatie. In hoofdstuk zes wordt de context van innovatie in Nederland besproken. Dan volgen vier datahoofdstukken op basis van interviews, observaties en analyse van documenten, vooral door de docenten gebruikte en/of door hen zelf vervaardigde toetsen. Eerst een verhalend hoofdstuk, waarin de drie docenten worden voorgesteld en vertellen over hun ervaringen, opvattingen en praktijk, en dan achtereenvolgens meervoudige case-studies van de drie docenten. In hoofdstuk elf worden de drie docenten met elkaar vergeleken voor wat betreft hun opvattingen en praktijk en het proefschrift wordt afgesloten met een conclusie- en discussiehoofdstuk.

Het proefschrift biedt een interessante kijk op de worsteling met bovengenoemde vernieuwingen van deze drie (en met hen, vele andere) docenten moderne vreemde talen in de afgelopen decennia in Nederland. Ik schrijf met opzet worsteling, omdat Zelfstandig Leren en de Tweede Fase niet alleen gelijktijdig, maar ook nog top-down werden ingevoerd, zonder rekening te houden met wat er op de werkvloer gebeurt. Problemen daarbij waren en zijn, aldus Elsen, dat betrouwbare, valide of zelfs aannemelijke onderzoeksgegevens over hoe deze vernieuwingen effectief te implementeren ontbreken, dat de vernieu-