Spreek2Schrijf

CLST en Telecats doen samen een haalbaarheidsstudie naar de mogelijkheid om de gesproken taal in de plenaire bijeenkomsten van de Tweede Kamer, automatisch om te zetten in schrijftaal zoals die nu in de Handelingen wordt gebruikt. Zij doen dit in opdracht van de Dienst Verslag en Redactie van het Nederlandse Parlement, de DVR.

Arjan van Hessen,
Telecats en HMI Universiteit Twente,
Henk van den Heuvel,
Maarten van Gompel
CLST, Radboud Universiteit

Spreek- versus Schrijftaal

Spraakherkenning is in de afgelopen drie jaar heel veel beter geworden. De belangrijkste oorzaak hiervan is het gebruik van een techniek uit de AI, genaamd Recurrent Neural Networks (RNN). In een aantal herkenningsexperimenten die we recent gedaan hebben, bleek een foutpercentage van minder dan 5% haalbaar, waarbij moet worden opgemerkt dat het om heel duidelijke, goed gearticuleerde, correct gesproken spraak ging. Maar toch… dit is iets dat een aantal jaren geleden beslist niet mogelijk was.

Deze goede spraakherkenning roept ook weer nieuwe vragen op. Kun je ook automatisch de sprekerswisselingen bepalen en dus herkennen wie wanneer spreekt? En kun je de herkende tekst omzetten in een beter leesbare tekst, die uit grammaticaal correcte zinnen bestaat in plaats van onafgeronde zinnen met uhm, af- en onderbrekingen en haperingen, die zo karakteristiek zijn voor gesproken taal? Wij mensen spreken immers anders dan we schrijven. Een gesproken zin als "ik ben gisteren…. o nee, dat was eergisteren en ook samen met Piet, naar de, eh het huis geweest", is voor de meeste luisteraars volstrekt duidelijk. Ook geschreven begrijp je wel wat er bedoeld werd, maar welgevormd is het niet.

Daarover gaat het project Spreek2schrijf. Zou het niet mogelijk zijn om de gesproken tekst om te zetten in geschreven tekst, waarbij de opeenvolgende woorden worden omgezet in zinnen met een hoofdletter en een punt, zonder de inhoud van de gesproken tekst geweld aan te doen?

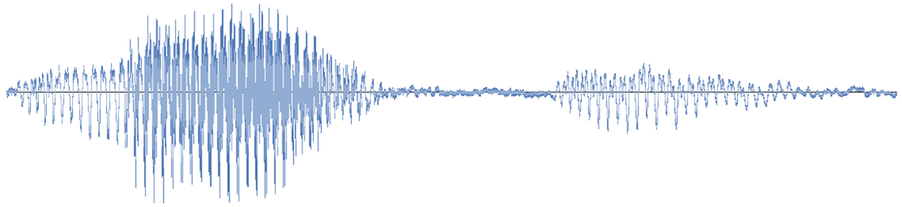
Dit is geen makkelijke klus omdat gesproken tekst vaak interne verbeteringen kent ("ik ben, o nee wij zijn…"). Maar ook een niet-perfect resultaat zou de mensen die van de gesproken tekst een leesbare geschreven tekst moeten maken, al enorm kunnen helpen.

Doel

Het doel van het Spreek2Schrijf-project (S2S) is uit te zoeken, in hoeverre de letterlijke, schriftelijke weergave van hetgeen er in de plenaire sessies van het Nederlandse Parlement gezegd is (bron), omgezet kan worden in schrijftaal (doel), die de manier van schrijven zoals die nu gebruikelijk is, zoveel mogelijk benadert.

Omdat automatische spraakherkenning (ASR) gebruikt zal worden om de spraak in tekst om te zetten, moeten we ervan uitgaan dat het bronmateriaal fouten zal bevatten. Een bijkomend doel van het onderzoek is dan ook om te zien in hoeverre de door de spraakherkenner gegenereerde teksten in schrijftaal overgezet kunnen worden. Hiermee zal menselijke correctie van die schrijftaal een stuk efficiënter gedaan kunnen worden.

© jank1000
www.fotosearch.com
Stock Photography



Data

Sinds september 2014 worden alle vergaderingen in de plenaire zaal van de Tweede Kamer volledig opgenomen. De door de DVR gemaakte teksten (de zogeheten Handelingen) worden automatisch opgelijnd. Hierdoor kunnen de video's van de plenaire debatten automatisch worden ondertiteld en kan er op de gesproken woorden worden gezocht. Deze spraakopnamen worden op dit moment niet door de automatische spraakherkenner gehaald. Om voldoende materiaal voor het S2S-project te krijgen, zullen zoveel mogelijk debatten alsnog door de herkenner gehaald moeten worden.

Er zijn ongeveer 1 miljoen gesproken woorden nodig, naast de parallelle teksten die door de DVR zijn geproduceerd. Een korte beschouwing van de huidige ondertitelbestanden laat zien dat er tussen de 9K en 10K per uur gesproken worden. Dat houdt in dat we minimaal 100 uur AV-materiaal door de herkenner moeten halen om op de benodigde 1M woorden te komen.

Vertaalmodule

De gesproken en de geschreven teksten worden in S2S beschouwd als twee verschillende talen. Het ligt dan ook voor de hand om, net als bij het vertalen van Nederlands naar Italiaans, automatische vertaalsoftware in te zetten. In S2S gaan we hiervoor gebruikmaken van Moses[1]: een statistisch vertaalsysteem. De vergaarde parallelle teksten, met enerzijds de uitvoer van de spraakherkenner en anderzijds de formele teksten van de DVR, dienen hier als leermateriaal waaruit een statistisch vertaalmodel geleerd wordt. Dit model bestaat uit frasen (woorden of zinssneden) uit de spreektaal, gepaard met frasen uit de schrijftaal die als vertaling gezien kunnen worden. Daarbij wordt voor elk paar een bepaalde waarschijnlijkheid berekend. Het model wordt uiteindelijk ingezet om nieuwe teksten in spreektaal om te zetten naar schrijftaal. Voor andere aanpakken van machinaal vertalen verwijzen we graag naar de artikelen van Lieve Macken (p. 7) en Sander Wubben (p. 22) in dit DIXIT-nummer.

Twee strategieën

Correctie van de herkenning
Het is niet zo dat de door de politici gesproken en door de DVR geschreven teksten geheel verschillen. In de meeste gevallen volgt de DVR de gesproken tekst. Wellicht kunnen via een betrouwbaarheidsscore in de oplijnroutine die plekken waar spraak en schrift afwijken, automatisch gedetecteerd worden. Als dat lukt dan hoeft alleen de tekst die afwijkt herschreven te worden. Als dat werkt, hebben we 'perfecte' herkenning (=spreektaal) waarop het spreek2schrijftaal algoritme (S2S-module) losgelaten kan worden.

Correctie van de schrijftaal
Een andere optie is om de imperfecte spraakherkenningsresultaten met het Spreek2schrijftaal algoritme om te zetten in (imperfecte) schrijftaal en die schrijftaal vervolgens te corrigeren. Het is zelfs mogelijk dat consequente fouten in de ASR-output door de S2S-module automatisch gecorrigeerd worden (als dezelfde fouten in het trainmateriaal van de S2S-module voorkwamen).

In het S2S-onderzoek zullen we beide opties in samenspraak met de DVR bekijken (wat werkt het best?).

Conclusie

S2S is een mooi voorbeeld van publiek-private samenwerking van twee NOTaS deelnemers: het CLST en Telecats. Als blijkt dat de gekozen aanpak inderdaad leidt tot een verbetering van het werkproces van de DVR (het maken van de Handelingen op basis van wat er gezegd werd in de Kamer), dan biedt de gekozen aanpak enorm veel mogelijkheden voor andere projecten waarbij gesproken spraak in een verslag moet worden omgezet. Maar het is waarschijnlijk niet zo dat de software 1-op-1 ingezet zal kunnen worden. De spraak in de Tweede Kamer is nu eenmaal anders dan die in een willekeurige gemeenteraadsvergadering, of bestuursvergadering van een hondenclub. Wel zal hopelijk blijken dat het idee om dit te zien als een vertaalprobleem tot een succesvolle oplossing kan leiden.

1. http://www.statmt.org/moses/