

Dit is

# Autonomata Too

Henk van den Heuvel

**Met spraakherkenning kun je alle kanten uit. Dat geldt zeker voor navigatiesystemen. Navigatiesystemen worden gekenmerkt door het gebruik van namen, en meer algemeen door het gebruik van Points of Interest (POIs). Namen leveren voor automatische spraakherkenning aparte problemen op. Autonomata Too zal oplossingen voor deze problemen onderzoeken en een demonstrator bouwen waarin deze oplossingen kunnen worden getoond en geëvalueerd.**

## Uitdaging

Het letter per letter ingeven van de bestemming in een navigatiesysteem is niet echt comfortabel te noemen, zelfs indien die bestemming meestal reeds na het intikken van de eerste drie of vier letters op het scherm verschijnt. Het kunnen inspreken van de bestemming is een vooruitgang, tenminste, als die ingesproken bestemmingen voldoende betrouwbaar herkend worden. Automatische spraakherkenning (ASH) van namen is om een aantal redenen lastig. Vaak is er een andere relatie tussen schrift en klank dan bij gewone woorden, zodat de gangbare verklankingsregels niet opgaan. De spellingwijze van namen is soms verouderd (zoals in *Blaauw*, *Aelderts*), en er zijn namen van buitenlandse herkomst (*Korlaelçi Demirgökçen*, *Gillygooley Road*). En soms verbaas je je alleen maar (vergelijk de uitspraak van het woord *vulpen* met die van de naam in *Gebr. Van Vulpen*). Het komt bovendien voor dat de spreker de juiste uitspraak van de naam niet kent, en een alternatieve uitspraak produceert die we een uitspraakvariatie noemen. In elk geval vereist een juiste herkenning dat de werkelijk gebruikte uitspraak in het lexicon van de ASH-machine voorkomt.

## Autonomata.

Het voorloperproject Autonomata leverde resources op die doelgericht onderzoek naar de ASH van namen toelaten. Ten behoeve van de letter-naar-klank-omzetting (kortweg g2p-omzetting) van namen, zijn goede g2p-omzetters voor persoonsnamen (voornamen, familienamen) en toponiemen (straatnamen, plaatsnamen) gemaakt. En er is een transcriptietoolbox ontwikkeld waarmee men uitgaand van eigen trainingslexica nauwkeurige g2p-omzetters voor andere naamsoorten (bv. POIs) kan bouwen.

Om de uitspraakvariëaties in kaart te brengen is een corpus van voorgelezen namen gebouwd. Daarin lezen 240 sprekers elk 250 namen voor. Naast Nederlandse en Vlaamse namen, zitten er ook namen van Franse, Engelse, Turkse en Marokkaans herkomst bij. Naast Nederlanders en Vlamingen, werden er ook anderstalige sprekers gerekruteerd die het Nederlands wel tot op zekere hoogte beheersen. Alle naamuitspraken zijn fonetisch getranscribeerd.<sup>1</sup>

## Autonomata Too

Autonomata Too betekent: Autonomata Transfer Of Output. In dit project gaan we de resources van Autonomata inzetten om een demoherkenner voor POIs te maken.

In onze toepassing leggen we ons toe op de ASH van overnachtingsadressen en eetgelegenheden in twee grote steden (vermoedelijk Amsterdam en Antwerpen).

De Autonomata transcriptietoolbox zal worden ingezet voor de bouw van een g2p-omzetter voor POIs en voor het leren modelleren van uitspraakvariëaties zoals die te vinden zijn in het Autonomata namencorpus. De bedoeling is om een zo goed mogelijk lexicon van de ASH-machine te krijgen. Aangezien namen vaak delen van vreemde herkomst bevatten zal ook onderzocht worden hoe de ASH verder te verbeteren valt door ook een (minimaal) aantal akoestische modellen voor buitenlandse fonemen te gebruiken.

---

<sup>1</sup> De resources van Autonomata zijn via de TST-centrale te verkrijgen (zie [www.tst.inl.nl](http://www.tst.inl.nl))

De uiteindelijke demo zal worden getest op een dataset die apart voor dit project zal worden opgenomen en getranscribeerd.

### **Uitvoerders**

De projectuitvoerders zijn de universiteiten van Nijmegen (CLST), Gent (ELIS), en Utrecht (UiL-OTS) tezamen met de bedrijven Nuance en TeleAtlas.

Het consortium verheugt zich erop om dit project in het kader van STEVIN uit te voeren. We werken aan resultaten waar velen van u alle kanten mee uit kunnen.

Autonomata is een project uit de derde ronde.

Het project loopt tot eind februari 2010. Zie:

<http://lands.let.ru.nl/projects/AutonomataToo/>

Henk van den Heuvel is directeur van CLST en projectleider van Autonomata-Too