



Het Nederlands en Taal en Spraaaktechnologie

Amsterdam 18 Oktober 2012

Jan Odijk



Universiteit Utrecht

Overzicht

- **META-NET Studie**
- Het Nederlands in de META-NET studie
- Het Nederlands in Nuance Spraaktechnologie
- De ontwikkeling van spraaktechnologie
- Conclusies



META-NET Studie

- Digitale ondersteuning voor 30 van de 80 talen gesproken in Europa
- Door META-NET
 - Europees excellentienetwerk (60 onderzoekscentra, 34 landen)
 - <http://www.meta-net.eu>
- Digitale ondersteuning door taal- en spraaktechnologie
 - Technologie en toepassingen
 - Spraakherkenning, spraaksynthese, grammaticale analyse, semantisch analyse, tekstgeneratie, automatisch vertalen
 - Data en kennisbanken
 - Tekstcorpora, spraakcorpora, parallelle corpora, lexicale taalbronnen (bijv. woordenboeken), grammatica's

META  NET



META-NET Studie

- Vergelijking tussen de talen op vier gebieden
 - Automatisch vertalen
 - Spraakinteractie
 - Tekstanalyse
 - Beschikbaarheid van taalbronnen

META  NET



META-NET Studie: Resultaten

- Beschreven in 'taalwitboeken' <http://www.meta-net.eu/whitepapers>
 - Digitaal en in papieren versie beschikbaar
- Belangrijkste resultaten
 - 21 van de 30 talen: ondersteuning is 'niet-bestaand' of 'zwak'
 - 4 daarvan: laagste score op alle gebieden (IJslands, Lets, Litouws en Maltees)
 - Als er niets gebeurt dreigt de digitale uitsterving van deze talen!
 - Geen enkele taal heeft 'excellente' ondersteuning
 - Alleen voor Engels is er 'goede' ondersteuning

META  NET



META-NET Studie: Quotes

- *Voor de meerderheid van de Europese talen is er een ernstig gebrek aan taalbronnen en enkele worden bijna geheel verwaarloosd. In deze zin zijn vele van onze talen nog niet toekomstbestendig* (prof.dr. Hans Uszkoreit, DFKI, Berlijn)
- *Er zijn dramatische verschillen in taaltechnologische ondersteuning tussen de verschillende Europese talen en technologische gebieden. Het gat tussen de 'grote' en de 'kleine' talen blijft maar groter worden. We moeten ervoor zorgen dat we alle kleinere talen en talen met te weinig taalbronnen uitrusten met de noodzakelijke basistechnologie, anders zijn deze talen gedoemd digitaal uit te sterven* (dr. Georg Rehm, DFKI, Berlijn)

META  NET



META-NET Studie: Oproep

- Oproep voor een gezamenlijk programma van nationale overheden en Europa, van kennisinstellingen en industrie om
 - te zorgen voor behoud van de meertaligheid in Europa ook in het digitale tijdperk
 - De taal- en spraaktechnologie op een hoger plan te tillen
 - De Europese meertaligheid tot een economische sterkte te maken
- Daartoe is ook een [Strategische Onderzoeksagenda](#) opgesteld
 - <http://www.meta-net.eu/sra>
 - Wordt in november officieel gepubliceerd

META  NET



Overzicht

- META-NET Studie
- **Het Nederlands in de META-NET studie**
- Het Nederlands in Nuance Spraaktechnologie
- De ontwikkeling van spraaktechnologie
- Conclusies



META-NET Studie: Nederlands

- [Het Nederlands in het Digitale Tijdperk](#)
- Nederlands doet het niet zo slecht
 - In dezelfde klasse als Frans en Spaans
 - Behalve voor automatisch vertalen
- ‘gematigd optimistisch’



Nederlands

- Hoe komt het dat het Nederlands het zo goed doet?
 - Door bewuste taalpolitiek om het Nederlands te versterken in de digitale informatiemaatschappij
 - Gez. Nederlands-Vlaams programma STEVIN (2004-2011)
 - Voor het creëren van Nederlandse taalbronnen
 - Voor strategisch onderzoek naar Nederlandstalige taal- en spraaktechnologie
 - Nauwe samenwerking tussen kennisinstellingen en industrie
 - Nauwe samenwerking tussen Nederland en Vlaanderen
 - <http://taalunieversum.org/taal/technologie/stevin/>



Nederlands

- STEVIN programma is beëindigd
 - Heeft excellente evaluaties door internationale en nationale experts
 - Heeft grote hoeveelheden nieuwe data en technologie opgeleverd
 - <http://taalunieversum.org/taal/technologie/stevin/etalage/>
 - http://taalunieversum.org/taal/technologie/stevin/documenten/stevin_resultatenboek.pdf
 - Spyns & Odijk (2012) *Essential Speech and Language Technology for Dutch: Results by the STEVIN-programme*. Springer Verlag. ISBN 978-3-642-30909-0. [Preprint](#)
 - levert daarmee een uitstekende basis voor geavanceerd onderzoek aan de hand van de Nederlandse taal
 - Waarmee de toekomst van het Nederlands in de digitale informatiemaatschappij bestendig kan worden



Nederlands

- Er moet nu een vervolgprogramma komen
 - om deze resultaten optimaal te benutten
 - Om de digitale ondersteuning van het Nederlands te blijven waarborgen
 - Er is hiervoor een concreet voorstel dat compatibel is met de META-NET Strategische Onderzoeksagenda
 - Rapport (Nederlandse Taalunie)
 - [Presentatie](#)



Overzicht

- META-NET Studie
- Het Nederlands in de META-NET studie
- **Het Nederlands in Nuance Spraaktechnologie**
- De ontwikkeling van spraaktechnologie
- Conclusies



Dragon NaturallySpeaking

- Bestaat in ca 8 talen
- En Nederlands zit daarbij!
 - Hoe komt dat?
- Andere spraaktechnologie is er voor veel meer talen:
 - TTS 40+
 - Vocon ASR 40+
 - Mobile Dictation: 35
- Waarom dit verschil?



Talen in spraaktechnologie

- TTS
 - Relatief kleine investering
 - Bedrijven kunnen dat zelf opbrengen
 - → veel talen afgedekt
- ASR (netwerk, desktop, mobile)
 - Grote investering
 - Bedrijven kunnen dat niet voor alle talen zelf opbrengen
 - Samenwerking tussen verschillende bedrijven
 - Gedeeltelijk ondersteund door overheidsgeld
 - SpeechDat familie van projecten 1995-2005
 - → veel talen afgedekt



Talen in spraaktechnologie

- Desktop Dicteersystemen
 - Zeer grote investering
 - Er is niet samengewerkt door bedrijven
 - De technologie is wel samengekomen in 1 bedrijf
 - Investerings waren grotendeels in dezelfde talen
 - Minder goede afdekking van talen
 - Waarom wel Nederlands?
 - Oa De spraaktechnologie van Nuance
 - Gedeeltelijk Vlaamse oorsprong (L&H)
 - Gedeeltelijk Nederlandse oorsprong (Philips)



Spraaktechnologie

- Volledige taalafdekking kan niet van 1 bedrijf komen
 - Ondanks moderne methodes van datacollectie via de cloud (Bijv. Nuance mobiel dicteren)
- Additionele bronnen zijn nodig



Overzicht

- META-NET Studie
- Het Nederlands in de META-NET studie
- Het Nederlands in Nuance Spraaktechnologie
- **De ontwikkeling van spraaktechnologie**
- Conclusies



De ontwikkeling van Spraaktechnologie

- De gebruikte technologie
 - Voor het eerst toegepast in de jaren 70
 - Is goed en biedt nog steeds mogelijkheden tot vooruitgang in performantie
 - Wordt tegenwoordig ook voor automatisch vertalen gebruikt
- Maar de vooruitgang is incrementeel / evolutionair; vooral gedreven door
 - Meer data
 - Meer rekenkracht (op PCs, in de cloud)
- Volgens velen zijn er inherente beperkingen aan de technologie
 - Nieuwe paradigma's moeten onderzocht worden
 - Dat kan alleen indien er ruimte is voor nieuwe onderzoeksprogramma's



Conclusies

- Er moet een gezamenlijk inspanning komen om de meertaligheid van Europa ook in het digitale tijdperk te behouden
- Voor het Nederlands moeten de eerder ingezette programma's nu een vervolg krijgen (anders kan het Nederlands ook in gevaar komen)
- De technologie moet verder verbeteren
 - niet alleen incrementeel
 - maar ook naar een hoger niveau getild worden
- De META-NET Strategische Agenda en de geformuleerde plannen voor Nederland en Vlaanderen vormen een goede basis voor onderzoeksprogramma's die dit concreet kunnen realiseren



META-NET SRA

