

# Speech group publications

## DE AUTOMATISCHE GENERATIE VAN FONEEMTRANSCRIPTIES EN SEGMENTATIES VOOR HET NEDERLANDS

In deze presentatie beschrijven we de automatische generatie van foneemtranscripties en de bijhorende foneem- en woordsegmentaties zoals die gemaakt worden voor het Vlaamse deel van het Corpus Gesproken Nederlands. Eerst beschrijven we de automatische generatie van een netwerk van alternatieve foneemtranscripties op basis van de orthografie. Uit dit netwerk selecteert de automatische spraakherkenner het akoestisch best passende pad. Vervolgens behandelen we de gebruikte algoritmes voor het maken van woord- en foneemsegmentaties. We besluiten met een gedetailleerde analyse van de verschillen tussen enerzijds de manueel geproduceerde foneemtranscripties en woordoplijningen en anderzijds de resultaten van het automatische proces. Deze evaluatie gebeurt op de uiteenlopende componenten die binnen het CGN aanwezig zijn: van voorgelezen spraak over spontane conversaties tot telefoonspraak.