

# Verstaanbaarheid en de erkenning van streektalen

Wilbert Heeringa en John Nerbonne<sup>1</sup>

## Inleiding

In 1992 werd door het Comité van Ministers van de Raad van Europa een verdrag vastgesteld dat als doel heeft om regionale talen en talen van minderheden binnen de Europese Unie te beschermen en het gebruik daarvan te bevorderen. Dit verdrag trad in 1998 in werking en staat bekend als het *Europees Handvest voor regionale talen of talen van minderheden*.

In deze bijdrage focussen we op ‘regionale talen’. Het handvest omschrijft ‘regionale talen of talen van minderheden’ als talen die:

- (1) van oudsher worden gebruikt in een bepaald gebied van een staat door onderdanen van die staat die een numerieke minderheid vormen ten opzichte van de overige bevolking van die staat;
- (2) verschillen van de officiële taal/talen van die staat;
- (3) geen dialecten van de officiële taal/talen van de staat of talen van migranten zijn.

Het handvest is in Nederland op 1 maart 1998 in werking getreden. Het verdrag is niet ondertekend door België, onder andere omdat dit de taalpolitieke evenwichten in België zou kunnen verstoren.

Het handvest bestaat uit drie delen. Het eerste deel bestaat uit diverse algemene bepalingen. Deel II bestaat uit algemene doelstellingen en beginselen die van

---

1 Wilbert Heeringa, Fryske Akademy, Leeuwarden.  
John Nerbonne, DFG, Words, Bones, Genes and Tools Center, Tübingen, Duitsland..

toepassing zijn op alle regionale talen en talen van minderheden die volgens het Handvest erkend worden. Het Nedersaksisch en het Limburgs werden in respectievelijk 1995 en 1997 erkend volgens deel II. Dat betekent dat de overheid het gebruik van deze talen moet faciliteren en aanmoedigen. Een verzoek tot erkenning van het Zeeuws werd door de overheid in 2005 afgewezen. In de praktijk functioneert het Zeeuws echter op dezelfde manier als het Nedersaksisch en het Limburgs.

Deel III bestaat uit 68 uitgewerkte maatregelen ter bescherming en bevordering van regionale talen of talen van minderheden, waarvan er minstens 35 gekozen moeten worden. Het Fries is erkend volgens deel III. Dat betekent dat het Fries een verplicht vak is op basisscholen, gebruikt mag worden in de rechtbank en in contacten met overheidsinstanties en publieke diensten.

Het is duidelijk dat zowel de erkenning volgens deel II als volgens deel III politieke en administratieve consequenties met zich meebrengt, en daarom moeten er duidelijke criteria zijn volgens welke bepaald wordt of een groep van lokale niet-standaardvariëteiten erkend kan worden als een regionale taal of als een minderheidstaal. Volgens het handvest moeten erkende talen verschillend zijn van de officiële talen binnen de betreffende staat en mogen ze geen dialecten zijn van de officiële taal/talen binnen de staat. De vraag is echter hoe men objectief kan bepalen of variëteiten (voldoende of beduidend) verschillen van de officiële taal, en hoe men objectief kan bepalen of variëteiten al dan niet dialecten zijn van de officiële taal. En verder is het ook de vraag hoe bepaald wordt of een groep van lokale niet-standaardvariëteiten erkend wordt volgens deel II of deel III van het Europees handvest.

Erkenning is een politieke handeling en moet daarom ook op haar politieke merites beoordeeld worden. Voor de erkenning moet er bijvoorbeeld voldoende draagvlak onder de bevolking zijn. Wie de erkenning van een groep lokale dialecten als regionale taal nastreeft, moet kunnen laten zien dat daarvoor een brede en aanhoudende belangstelling onder de bevolking bestaat, bijvoorbeeld in de vorm van dialectverenigingen, toneeloptredens, voordrachten, publicaties, uitzendingen en activiteiten in de sociale media. Ons inziens wordt dat alles in de procedure voor erkenning terecht in rekening gebracht.

Vaak wordt gezegd dat taalvariëteiten op basis van hun onderlinge verstaanbaarheid vergeleken moeten worden (Gooskens et al. 2015). We zouden daarom graag willen weten in hoeverre de resultaten van het erkenningsbeleid zoals

dat tot nu toe gevoerd is, een basis heeft in de onderlinge verstaanbaarheid van de variëteiten. Tot zover wij weten is dat nog nooit systematisch onderzocht, misschien ook omdat het ondoenlijk is om de onderlinge verstaanbaarheid van honderden lokale variëteiten in een taal- of dialectgebied te meten op basis van een statistisch voldoende groot aantal sprekers voor elk paar van (naburige) variëteiten.

Als oplossing voor dat probleem maken wij gebruik van een meetmethode uit de glottometrie – een deelgebied van de kwantitatieve taalkunde – die in onze benadering fungeert als een proxy voor verstaanbaarheid. Verstaanbaarheids-experimenten waaraan honderden sprekers moeten meedoen, zijn daarbij niet nodig. In plaats daarvan wordt een meetmethode gebruikt die verstaanbaarheid voorspelt op basis van steekproefsgewijs verkregen taal- en dialectmateriaal (Van Heuven et al. 2015).

De procedure die we in deze bijdrage schetsen, bestaat uit twee stappen. In de eerste stap bepalen we welke gebieden in het taal- of dialectlandschap onderscheiden worden, en tot welke van deze gebieden de officiële talen gerekend moeten worden. Alle gebieden die geen van beide officiële talen omvatten, kunnen erkend worden als regionale taal. Echter, als een groep een officiële taal omvat, laat dit zien dat de variëteiten binnen die groep te weinig verschillen van die officiële taal, en kan alleen die groep niet als regionale taal erkend worden. Daarmee doen we recht aan de eis van het Europese handvest dat een regionale taal of een minderheidstaal verschillend moet zijn van de officiële taal.

In de tweede stap kijken we op welk niveau dialectgebieden erkend kunnen worden. We presenteren een index van taalkundige onafhankelijkheid die daarbij hulp kan bieden.

De stappen zoals zojuist genoemd, worden hieronder nader uitgewerkt. In paragraaf 2 beschrijven we de gegevensverzameling die we voor onze metingen gebruiken. Paragraaf 3 legt uit hoe afstanden gemeten worden tussen de variëteiten (lokale dialecten en officiële talen). Op basis van de afstanden vinden we gebieden. Dit wordt uitgelegd in paragraaf 4. In paragraaf 5 bepalen we voor elk gebied de index van taalkundige onafhankelijkheid. We eindigen met enkele conclusies in paragraaf 6.

## Gegevensverzameling

De gegevensbron van ons onderzoek is de *Reeks Nederlandse Dialectatlassen* (RND). De 16 delen waaruit deze atlasreeks bestaat, verschenen tussen 1925 en 1982 onder redactie van E. Blancquaert en W. Péé. In de atlasen vinden we voor 1956 dialecten in Nederland, Vlaanderen, het graafschap Bentheim en Noordoost-Luik vertalingen en transcripties van 139 zinnen. We maakten een selectie van 361 dialectplaatsen die over het algemeen een regelmatig net vormen. Voor Tjalleberd, Donkerbroek en Appelscha zijn er zowel een Friese als een Nedersaksische variant. De selectie omvat ook 33 Friese variëteiten.

Uit de 139 zinnen selecteerden we min of meer willekeurig 125 woorden. De reden waarom we de RND gebruikten, is dat bij gebruik daarvan variatie in drie taalkundige niveaus betrokken kan worden. Een woord kan lexicaal variëren, variëren in klankcomponenten en morfologisch variëren.

De officiële taal die het meest verwant is aan de dialecten in Nederland en Vlaanderen zal meestal het Standaardnederlands zijn, en misschien voor sommige van de meest oostelijke dialecten het Standaardduits. Van beide officiële talen hebben we daarom transcripties nodig. Voor de transcriptie van het Standaardnederlands vormde het *Tekstboekje* van Blancquaert (1939) de basis. Voor de Duitse transcripties gebruikten we met enkele aanpassingen het *Wörterbuch der deutschen Aussprache* (1969). Voor meer details over de data: zie Heeringa (2001). Met 361 lokale dialecten en twee officiële talen hebben we in totaal 363 variëteiten.

## Het meten van taalkundige afstanden

Gooskens & Swarte (2017) rapporteren over een onderzoek waarin zij de verstaanbaarheid tussen vijf Germaanse talen meten door middel van onder andere een gesproken cloze-test. In deze test beluisterden proefpersonen een tekst, waarin twaalf woorden vervangen waren door een beep. De weggelaten woorden werden bovenaan het scherm weergegeven. Voor elk weggelaten woord kregen proefpersonen 30 seconden de tijd om het juiste woord te vinden. De score van de test is gelijk aan het aantal correct gekozen woorden. De metingen proberen de auteurs vervolgens te verklaren door middel van een stepwise regressiemodel met daarin lexicale afstanden, orthografische woordstamafstanden, affix-afstanden, fonetisch/fonologische afstanden en syntactische afstanden (gemeten op drie manieren) als predictoren. Wanneer zij focussen op de proefpersonen die

(nauwelijks) bekend waren met de taal waarvoor zij de *cloze test* deden, blijken alleen de lexicale afstanden en de orthografische stamafstanden in het model over te blijven. De verstaanbaarheidsscores van de taalparen Deens/Zweeds en Zweeds/Deens blijken uitschieters te zijn, en wanneer beide taalparen weggelaten worden, blijven de fonetisch/fonologische afstanden over als enige voorspeller in het model. Dit onderzoek laat zien dat vooral de uitspraak van woorden (inclusief variatie in lexis en klankcomponenten) de verstaanbaarheid van Germaanse talen bepaalt. We gebruiken daarom de Levenshtein-afstand voor het meten van uitspraakafstanden als proxy voor verstaanbaarheidsmetingen.

In 1995 gebruikte Kessler de Levenshtein-afstand voor het meten van taalkundige afstanden tussen Ierse dialecten. De Levenshtein-afstand is gelijk aan de minimale kosten die nodig zijn om de ene realisatie te veranderen in de andere. In het eenvoudigste geval zijn drie operaties mogelijk: een segment toevoegen, verwijderen of vervangen door een ander element. We laten dit zien aan de hand van een voorbeeld. In het dialect van Amsterdam wordt *konijn* uitgesproken als [kɔnɛ:n], en in het dialect van Zwollekerspel als [kni:nə]. We krijgen dan de volgende oplijning:

1	2	3	4	5	6
k	ə	n	ɛ:	n	
k		n	i:	n	ə
	1		1		1

*Tabel 1: Oplijning van de Amsterdamse en Zwollekerspelse varianten voor het woord konijn*

We zien drie operaties: verwijdering van [ə], vervanging van [ɛ:] door [i:] en toevoeging van [ə]. De oplijning bestaat uit zes posities.

Onze versie van het Levenshtein-algoritme houdt rekening met syllabificatie: de [j] en de [w] mogen met een klinker (of andere medeklinker) corresponderen, de [i] en de [v] mogen met een medeklinker (of andere klinker) corresponderen, de [ə] mag met een sonorante medeklinker (of andere klinker) corresponderen, alle andere klinkers mogen alleen corresponderen met klinkers, en alle andere medeklinkers mogen alleen corresponderen met medeklinkers. Bovendien gebruiken we graduele gewichten waardoor vervanging van bijvoorbeeld een [i] door een [ɪ] minder zwaar weegt dan vervanging van een [i] door een [a]. Voor meer details: zie Heeringa (2004) en – meer beknopt – Gooskens & Heeringa (2004).

Stel: we berekenen nu de afstand tussen Middelstum en Ommen op basis van zes woorden. De berekening ziet er dan als volgt uit:

	Middelstum	Ommen	Levenshtein afstand	Lengte oplijning
<i>schip</i>	sxip	sxip	0	4
<i>pet</i>	pet	petə	1	4
<i>roepen</i>	rəupm	ərupm	2	6
<i>springen</i>	sprɪŋ	sprɪŋkt	2	7
<i>kelder</i>	kɛlər	kɛldər	1	6
<i>huis</i>	hus	hys	1	3
			7	30

Tabel 2: Afstand tussen Middelstum en Ommen op basis van de woorden *schip*, *pet*, *roepen*, *springen*, *kelder* en *huis*

De afstand tussen Middelstum en Ommen is nu gelijk aan  $7/30 = 0.23$  of 23%. De afstandsmetingen in deze bijdrage zijn gebaseerd op maximaal 125 woorden.

## Bepalen van gebieden

De 361 lokale dialecten vormen een steekproef uit een groot gebied waarin sommige plaatsen (vrijwel) hetzelfde dialect hebben, en andere plaatsen onderling heel verschillend zijn. Op basis van de verschillen in uitspraak willen we kijken welke variëteiten zó vergelijkbaar zijn dat men ze tot hetzelfde dialectgebied zou moeten reken. We zoeken groepen binnen de verzameling van 361 variëteiten en de twee officiële talen.

We schreven hierboven dat we voor elk lokaal dialect transcripties hebben van 125 verschillende woorden. Voor het vinden van de groepen gebruiken we nu een procedure die voor het eerst toegepast werd door Heeringa & Hinskens (2013) en uitvoerig beschreven is door Heeringa (2017). In deze procedure wordt in elk van 100 rondes 125 keer aselect een woord gekozen uit deze set van 125 woorden. We krijgen dan 100 steekproeven. Elk van deze 100 steekproeven heeft 125 woorden, maar omdat ze aselect gekozen zijn, is het waarschijnlijk dat in (bijna) alle 100 steekproeven sommige woorden meerdere keren voorkomen, en andere woorden helemaal niet voorkomen. Voor elke steekproef worden twee stappen uitgevoerd.

In de eerste stap berekenen we de afstanden tussen de variëteiten op basis van de 125 woorden uit de steekproef. Op basis van die afstanden genereren we een indeling in groepen (of: clusters) door middel van clusteranalyse. Het resultaat is een structuur die bestaat uit twee clusters. Die clusters bestaan weer uit subclusters, de subclusters uit subsubclusters, enz. Deze hiërarchische structuur kan weergegeven worden als een boom, waarbij de bladeren de variëteiten zijn. De takken in de bomen (of: dendrogrammen) die door middel van clusteranalyse gemaakt zijn, representeren de afstanden tussen dialecten en clusters. Voor meer details: zie Jain & Dubes (1988).

Er bestaan verschillende clustermethoden. Wij gebruikten een methode die bekend staat als *single-link* clusteranalyse. De keuze om een cluster te fuseren met een ander cluster op een hoger niveau in de hiërarchie wordt uitsluitend bepaald op basis van de afstand tussen hun meest gelijkende variëteiten.<sup>2</sup> Deze methode lijkt daarmee op de pijltjesmethode. Deze methode werd in 1886 door P. Willems voor het eerst toegepast op Nederfrankische dialecten in het gebied tussen Duinkerken en het Rijnland. Later werd vrijwel dezelfde methode onder andere toepast door W. Rensink (1955) en Jo Daan (Daan & Blok 1969) voor Nederland. Volgens deze methode worden plaatsen met dialecten die volgens de sprekers (vrijwel) hetzelfde zijn door pijlen met elkaar verbonden. Witte stroken ontstaan daar waar plaatsen niet met elkaar verbonden zijn. Als variëteiten A en B door de sprekers als hetzelfde worden beoordeeld, en variëteit C wordt wel als hetzelfde beoordeeld als variëteit A, maar niet hetzelfde als variëteit B, wordt variëteit C toch tot dezelfde groep gerekend als die waartoe de variëteiten A en B behoren.

Het resultaat van de vorige stap is een hiërarchische boomstructuur. In de tweede stap vereenvoudigen we deze structuur tot een 'platte' verdeling in groepen. Beginnen we bij de wortel van de 'boom', dan vinden we een verdeling in twee groepen. Maar vervolgens zien we dat een van die twee groepen ook weer verdeeld is in twee groepen, waarmee we een verdeling in drie groepen hebben. Als we zover doordringen in de structuur dat we de bladeren bereiken, vinden we een verdeling in 363 groepen. Op deze manier zijn er 362 niveaus, met op het hoogste niveau een verdeling in twee groepen en op het laagste niveau een verdeling in 363 groepen. Voor elk niveau bereken we nu hoe goed de hiërarchische structuur tot op dat niveau de afstanden zoals gemeten in de vorige stap weer spiegelt. In een grafiek zetten we de variantie die door het betreffende deel van

---

2 Maar zie Prokić & Nerbonne (2008) voor een andere keuze, die hier echter onbelangrijk is.

de hiërarchische structuur verklaard wordt uit tegen de 362 niveaus. Het zal dan blijken dat een verdeling in maar een paar groepen al veel variantie verklaart, en dat bij toename van de groepen de toename van de verklaarde variantie afneemt. In de curve zien we echter ook een knik of ‘elleboog’. Deze markeert het aantal natuurlijke clusters (Aldenderfer and Blashfield, 1984). Na dit knikpunt is de toename van de verklaarde variantie ineens veel kleiner dan daarvoor. De procedure die dit knikpunt automatisch vindt, wordt beschreven in Heeringa (2017). Als we voor elk van de 100 steekproeven de natuurlijke groepen bepaald hebben, krijgen we 100 verdelingen. Voor elk paar variëteiten dat gevormd kan worden, tellen we nu het aantal keren dat beide variëteiten in dezelfde natuurlijk groep zitten. Dat kan variëren van 0 (nooit) tot 100 (altijd). Als dit 95 keer of vaker is, markeren we de variëteiten als ‘verbonden’. Op die manier ontstaan netwerken. Als bijvoorbeeld variëteiten *A* en *B* verbonden zijn en de variëteiten *C* en *D* zijn verbonden, maar de variëteiten *A* en *B* zijn niet verbonden met de variëteiten *C* en *D*, hebben we hier twee netwerken oftewel twee groepen. Variëteiten die met geen enkele andere variëteit verbonden zijn, blijven ongeklasseerd.

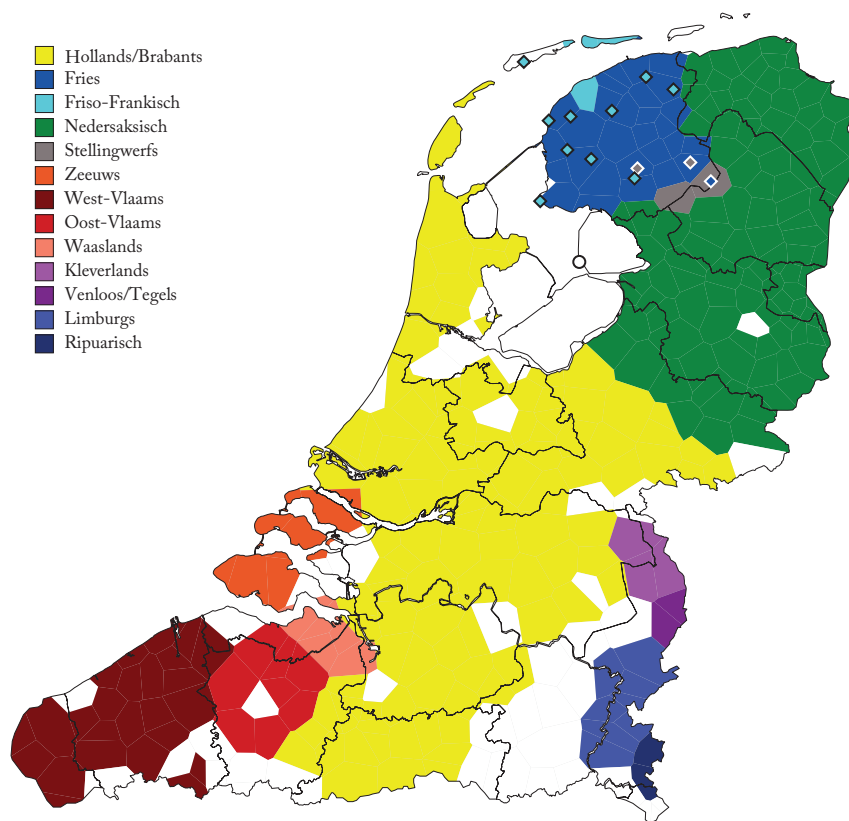
De resultaten worden getoond in Figuur 1. De procedure vond 13 netwerken die 13 groepen representeren. De groep ‘Hollands/Brabants’ omvat ook het Standaardnederlands. Volgens onze procedure worden dialecten in deze groep niet erkend als een regionale taal naast het Standaardnederlands omdat ze daaraan te sterk verwant zijn. Het Zeeuws daarentegen dat tot nu toe nog niet erkend is als regionale taal, zou volgens onze analyse wel erkend kunnen worden. Het Limburgs dat wel erkend is, zou volgens onze analyse opgedeeld moeten worden in vier kleinere regionale talen: Kleverlands, Venloos/Tegels, Limburgs en Ripuarisch. Er is geen enkele groep die het Standaardduits omvat. Dat betekent dat er in Nederland en Vlaanderen geen Duitse dialecten gesproken worden.

### Het meten van taalkundige onafhankelijkheid

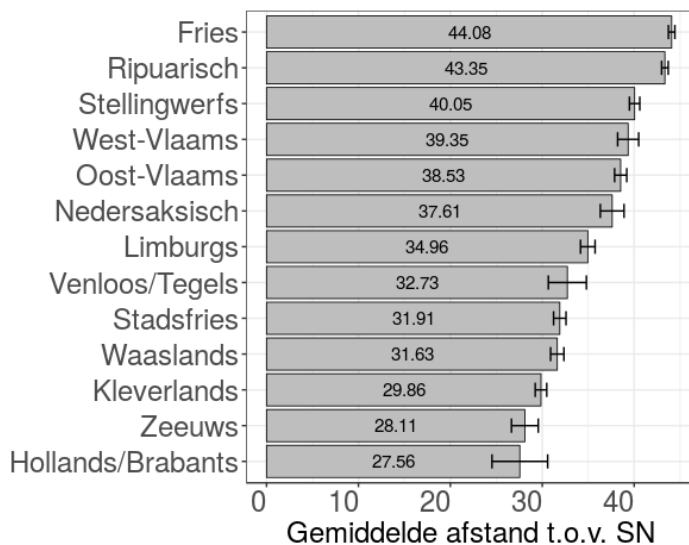
Omdat het Europese handvest twee verschillende graden van erkenning kent, namelijk erkenning volgens deel II en erkenning volgens deel III, hebben we ook een index van taalkundige onafhankelijkheid ontwikkeld die kan helpen om hier een keuze te maken. De berekening van deze index bestaat uit drie stappen. In de eerste stap meten we voor elke groep in Figuur 1 de gemiddelde afstand van de lokale dialecten (dus zonder eventuele officiële talen) tot het Standaardnederlands. Figuur 2 toont de resultaten. Het Fries heeft met 44,08% de grootste afstand, onmiddellijk gevolgd door het Ripuarisch (43,35%). Het Hollands/Brabants heeft de kleinste afstand (27,56%).



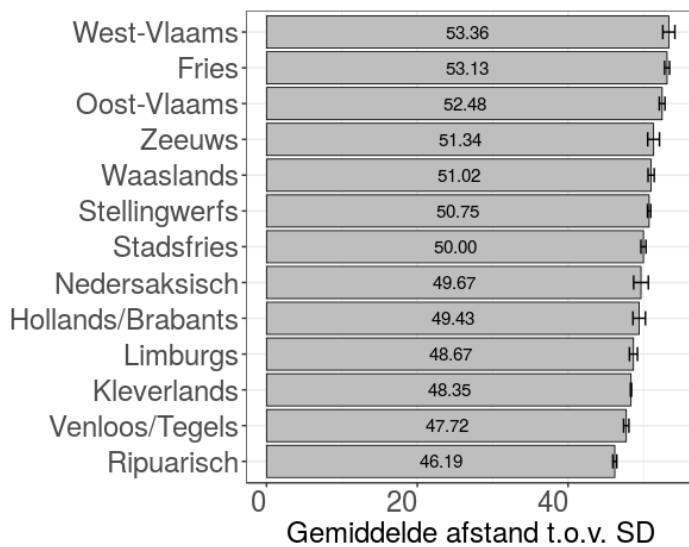
Omdat we benieuwd zijn of er dialecten zijn die meer verwant zijn aan het Standaardduits dan aan het Standaardnederlands, hebben we ook de gemiddelde afstanden tot het Standaardduits berekend (zie Figuur 3). Het Ripuarisch heeft de kleinste afstand tot het Standaardduits (46,19%), maar deze afstand is iets groter dan de afstand van het Ripuarisch tot het Standaardnederlands (43,35%). Voor alle groepen geldt dat de gemiddelde afstand tot het Standaardnederlands kleiner is dan de gemiddelde afstand tot het Standaardduits.



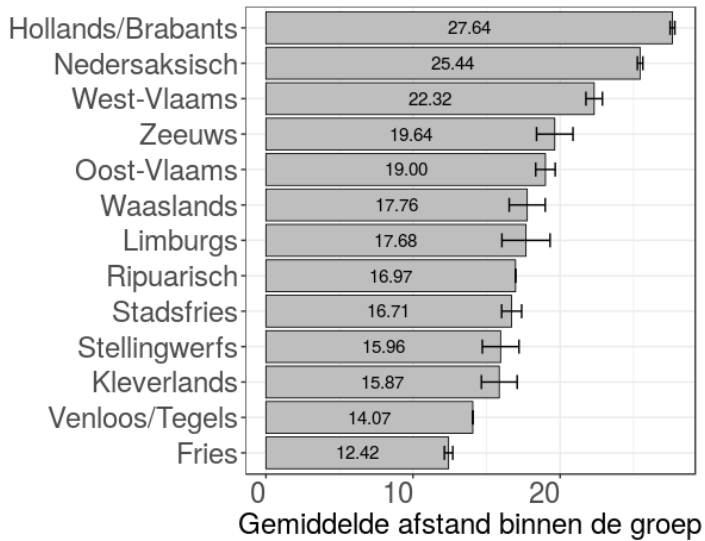
*Figuur 1: Taal- en dialectgebieden in Nederland, Vlaanderen, het graafschap Bentheim en Noordoost-Luik. Het Hollands/Brabants omvat ook het Standaardnederlands. De turquoise diamanten representeren Friso-Frankische dialecteilanden, de grijze diamanten representeren de Stellingwerfse variëteiten in (van west naar oost) Tjalleberd en Donkerbroek, de blauwe diamant representeert de Friese variëteit in Appelscha.*



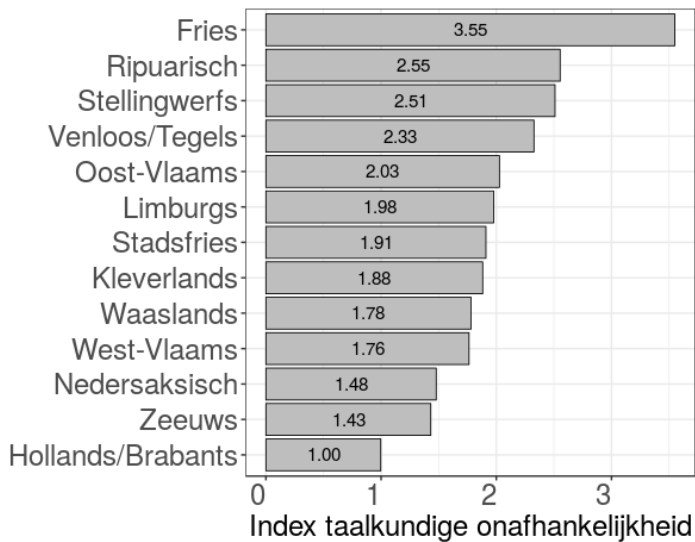
Figuur 2: Gemiddelde afstanden ten opzichte van het Standaardnederlands



Figuur 3: Gemiddelde afstanden ten opzichte van het Standaardduits



*Figuur 4: Gemiddelde afstand tussen de lokale dialecten per groep*



*Figuur 5: Index van taalkundige onafhankelijkheid berekend als de gemiddelde afstand ten opzichte van Standaardnederlands gedeeld door de gemiddelde binnengroepsafstand*

In de tweede stap willen we kijken hoe homogeen elk gebied is. Voor elk gebied bepalen we de gemiddelde taalkundige afstand tussen de lokale dialecten binnen dat gebied. Hoe homogener het gebied is, hoe kleiner deze afstand zal zijn. In Figuur 4 vinden we de resultaten. Het Fries is het meest homogeen met een gemiddelde afstand van 12.42%, en het Hollands/Brabants is het meest heterogeen, de gemiddelde afstand voor dit gebied is 27.64%. Het valt op dat de twee grootste gebieden – Hollands/Brabants en Nedersaksisch – het meest heterogeen zijn.

In de laatste stap berekenen we de index van taalkundige onafhankelijkheid. Deze wordt berekend door de gemiddelde afstand tot het Standaardnederlands te delen door de gemiddelde afstand tussen de lokale dialecten binnen de groep. We krijgen dan een index die aangeeft hoeveel keer groter de afstand van een gebied tot het Standaardnederlands is ten opzichte van de gemiddelde afstand binnen dat gebied. Hoe hoger de index, hoe hoger de mate van erkenning als regionale taal zou moeten zijn.

In Figuur 5 zien we dat het Fries de grootste index heeft (3.55), en het Hollands/Brabants de kleinste index (1.00). Tussen deze twee extremen in vormen de andere gebieden een soort middengroep, variërend van het Ripuarisch (2.55) tot het Zeeuws (1.43). De grafiek ondersteunt de besluiten volgens welke het Fries de hoogste erkenning kreeg, namelijk volgens deel III, en het Nedersaksisch en het Limburgs een lagere erkenning, namelijk volgens deel II. Echter, zoals we al eerder opmerkten, valt het Limburgs in onze analyse uiteen in vier kleinere gebieden.

## Conclusies en discussie

Om te voorkomen dat onze opzet verkeerd begrepen wordt, benadrukken we nogmaals dat we de rol van politieke en sociale factoren in de erkenning van regionale talen niet willen ondermijnen of doorkruisen. De officiële status brengt immers ook politieke en sociale gevolgen met zich mee. Deze factoren moeten daarom meteen vanaf het begin van de procedure meegewogen worden, en niet pas in later stadium of zelfs na erkenning aan het licht komen. Echter, niemand wil dat variëteiten die minimaal verschillen van de standaardtaal erkend worden als regionale taal, of dat twee groepen van variëteiten die erg vergelijkbaar zijn allebei een officiële status krijgen. De begrippen ‘minimaal verschillend’ en ‘erg vergelijkbaar’ kan men in de kwantitatieve taalkunde operationaliseren, en het was de bedoeling van ons stuk om aan te tonen hoe dit in zijn werk gaat.

De resultaten in deze bijdrage ondersteunen gedeeltelijk de besluiten zoals die tot nu toe genomen zijn inzake de erkenning van regionale talen. Ook als we alléén letten op verschillen in de uitspraak van woorden verdient het Fries een erkenning op het hoogste niveau, en het Nedersaksisch en het Limburgs op een lager niveau. Wel zien we in onze analyse dat het Limburgs niet als complete provincie wordt erkend, maar uiteenvalt in vier onderscheiden gebieden die elk apart erkend zouden moeten worden.

Volgens onze resultaten zouden onder andere ook het Stellingwerfs, het Friso-Frankisch, het Zeeuws, het West-Vlaams en het Oost-Vlaams erkend kunnen worden. Voor de laatste twee groepen is dat echter niet mogelijk, omdat België het Europees handvest niet ondertekend heeft.

De methodiek die we gebruikten, is in meerdere opzichten voor verbetering vatbaar. Dat betreft allereerst de data die we hebben gebruikt. We hebben de RND gebruikt omdat we op basis van dit materiaal de mate van lexicale variatie, variatie in klankcomponenten en morfologische variatie in één maat kunnen samenvatten. Maar het verschil tussen palatale [ç] en de velaire [x] komt in de RND transcripties niet uit de verf. Verder is het materiaal van de RND wellicht al verouderd en is het opgenomen in een groot tijdsinterval, namelijk van 1921 tot en met 1975.

We hebben in ons onderzoek de Nederlandse variant van het Standaardnederlands gebruikt. Echter, met betrekking tot de Vlaamse variëteiten zou het goed zijn om ook de Belgische variant van het Standaardnederlands in het onderzoek op te nemen. Waar we hier bovendien niet naar gekeken hebben, is syntaxis en tonaliteit. Met name tonen en het onderscheid tussen stoottoon en sleeptoon spelen een belangrijke rol in de Limburgse en Ripuarische dialecten, en daarmee onderscheiden zij zich ook van de andere dialecten.

Ten slotte merken we op dat we de analyse in dit onderzoek het liefst zouden herhalen op een gegevensverzameling die alle taal- en dialectgebieden in het West-Germaanse continuüm omvat. Bijvoorbeeld, het Nedersaksisch en het Ripuarisch vinden we ook in Duitsland. Het is de vraag of voor die gebieden het Nederlandse deel en het Duitse deel elk apart erkend moeten worden, indien historisch gezien beide delen één geheel vormen. Als we het politieke onderscheid buiten beschouwing laten, zal gekeken moeten worden of de Nederlands/Duitse staatsgrens zich inmiddels ontwikkeld heeft tot een taalkundige grens, of dat gebieden aan weerszijden van de staatsgrens nog steeds een eenheid vormen

en als zodanig als één regionale taal erkend kunnen worden. Dat is op dit moment echter niet goed te bepalen. Een hedendaagse gegevensverzameling met een voldoende dicht netwerk van plaatsen en consistente transcripties die het complete West-Germaanse gebied dekt en variatie voor alle taalkundige niveaus laat zien bestaat er – tot zover wij weten – nog niet.

## Referentielijst

- Aldenderfer, M. & R. Blashfield (1984). *Cluster Analysis*. Newbury Park (CA): Sage.
- Blancquaert, E. (1939). *Tekstboekje*, 2e druk (Nederlandse Fonoplaten van Blancquaert en van der Plaetse, Eerste Reeks). Antwerpen: De Sikkel.
- Blancquaert, E. & W. Pée, (red.) (1925-1982). *Reeks Nederlands(ch)e dialectatlassen*. Antwerpen: De Sikkel.
- Daan, J. & D. P. Blok (1969). Van Randstad tot Landrand; toelichting bij de kaart. *Dialecten en Naamkunde Vol. XXXVII (Bijdragen en mededelingen der Dialectencommissie van de Koninklijke Nederlandse Akademie van Wetenschappen te Amsterdam)*. Amsterdam: Noord-Hollandsche Uitgevers Maatschappij.
- De Caluwe, Johan (ed.) (2002). *Taalvariatie en taalbeleid: bijdragen aan het taalbeleid in Nederland en Vlaanderen*. Garant, 112-113/127
- Gooskens, Charlotte & Femke Swarte (2017). Linguistic and extra-linguistic predictors of mutual intelligibility between Germanic languages. *Nordic Journal of Linguistics* 40(2), 123–147.
- Gooskens, C. & W. Heeringa (2004), The Position of Frisian in the Germanic Language Area. In: D. Gilbers, M. Schreuder & N. Knevel (red.). *On the Boundaries of Phonology and Phonetics*, in honour of Dr. Tjeerd de Graaf, 61-87.
- Gooskens, C., A. Schüppert & N. Haug Hilton (2015). Special issue: Experimental approaches to the mutual intelligibility between closely related languages. *Linguistics* 53(2).
- Heeringa, W. (2001). De selectie en digitalisatie van dialecten en woorden uit de Reeks Nederlandse Dialectatlassen. *TABU; bulletin voor taalwetenschap*, 31, 61-103.
- Heeringa, W. (2004). *Measuring dialect pronunciation differences using Levenshtein distance*. Proefschrift Rijksuniversiteit Groningen, Groningen.
- Heeringa, W. & F. Hinskens (2014). Convergence between dialect varieties and dialect groups in the Dutch language area. In: B. Szmrecsanyi & B. Wälchli (red.). *Aggregating dialectology, typology, and register analysis; linguistic varia-*

- tion in text and speech, Linguae et Litterae* 28. Berlin and Boston: De Gruyter, 26-52 en 452-453.
- Heeringa, W. (2017). Finding Dialect Areas by Means of Bootstrap Clustering. In: M. Wieling, M. Kroon, G. van Noord & G. Bouma (red.). *From Semantics to Dialectometry; Festschrift in honor of John Nerbonne. Tributes* 32. London UK: College Publications, 127-135.
- Jain, A. & R. C. Dubes (1988). *Algorithms for clustering data*. Prentice Hall, Englewood Cliffs, N.J.
- Kessler, B. (1995). Computational dialectology in Irish Gaelic. In: *Proceedings of the 7th conference of the European chapter of the association for computational linguistics*. Dublin, EACL, 60-67.
- Prokić, J. & J. Nerbonne (2008). Recognising Groups among Dialects. *International Journal of Humanities and Arts Computing* 2 (1-2), 153-172.
- Rensink, W. G. (1955). Dialectindeling naar opgaven van medewerkers. *Mededelingen der Centrale Commissie voor Onderzoek van het Nederlandse Volkseigen* 7, 20-23.
- Van Heuven, V., C. Gooskens & R. van Bezooijen (2015). Introducing Micrela: Predicting mutual intelligibility between related languages in Europe. In: J. Navracscics & S. Batyi (red.), *First and second language: interdisciplinary approaches*. Studies in Psycholinguistics, Vol. 6. Tinta konyvkiado, 127-145.
- Wörterbuch der deutschen Aussprache (1969). *Wörterbuch der deutschen Aussprache*. München: Max Hueber Verlag.