

Wilbert Heeringa

## OVER DE INDELING VAN DE NEDERLANDSE STREEKTALLEN EEN NIEUWE METHODE GETOETST

### *Inleiding*

In november 2001 verscheen bij de uitgeverij Van Gorcum te Assen een boek met als titel *De indeling van de Nederlandse streektalen* geschreven door de gebroeders Cor en Geer Hoppenbrouwers (VIII + 212 blz.; ISBN 90-232-3731-5). Het was al lang bekend dat het boek uit zou komen, maar eindelijk is het er nu dan ook. De auteurs schrijven in de inleiding dat zij hebben gestreefd naar een optimale toegankelijkheid. We denken dat zij daar ruimschoots in geslaagd zijn. Het is een prettig leesbaar boek, en ook qua opzet overzichtelijk. In dit artikel willen we nader op het boek in gaan. De ondertitel van het boek luidt: *Dialecten van 156 steden en dorpen geklasseerd volgens de FFM*. Om de methodiek en de resultaten van de zogenaamde featurefrequentie-methode (FFM) te toetsen hebben we een kloon gebouwd van het computerprogramma 'Polyphon' dat door de gebroeders Hoppenbrouwers (H & H) ontwikkeld werd (blz. 14). In dit artikel gaan we in op de ontwikkeling van de kloon, en vergelijken onze resultaten met de resultaten van H & H.

In sectie 1 geven we een beschrijving van de featurefrequentiemethode. De basis voor de indeling van de Nedelands streektalen wordt gevormd door de gegevens uit de *Reeks Nederlandse Dialectatlassen* (RND) (Blancquaert en Pée, 1925-1982). We melden nu al dat er verschil is in de selectie van gegevens hieruit door H & H en door ons, maar bespreken dit nader in sectie 2. In sectie 3 bespreken we de featurematrix. Sectie 4 heeft de indeling van het Nederlandse taalgebied als onderwerp. In sectie 5 geven we resultaten zoals die door onze kloon worden gegenereerd en vergelijken die met de resultaten die H & H bespreken in de hoofdstukken 5 t/m 12 van hun boek. In sectie 6 geven we ten slotte onze conclusies.

## 1 De featurefrequentiemethode

### 1.1 De bepaling van frequenties

Ter inleiding op de uitleg van de featurefrequentiemethode (1.1) introduceren H & H de 'letterfrequentiemethode'. Per taal kunnen op basis van een gegeven tekst de frequenties van *letters* bepaald worden. Zo'n tekst vormt een steekproef, en omdat de steekproeven van de verschillende talen niet altijd even groot zijn, is het van belang de frequenties uit te drukken in percentages. De afstand tussen twee talen is nu gelijk aan de som van de verschillen van de letterfrequenties. H & H laten zien dat bij een dergelijke aanpak duidelijk uit de verf komt dat de afstand tussen het Afrikaans en het Nederlands veel kleiner is dan de afstand tussen het Afrikaans en het Samoaans en de afstand tussen het Nederlands en het Samoaans. Terecht wijzen H & H er echter op dat het verschil in spelling niet altijd een verschil in uitspraak betekent (bijv. *nog* vs. *noch*), en dat deze aanpak dus minder bruikbaar is.

Een nauwkeuriger aanpak bestaat uit het gebruik van fonetische teksten. In de inleiding van het boek (hoofdstuk 0) beschrijven H & H dat zij tien jaar geleden begonnen met een experiment op basis van de teksten uit de *The Principles of the International Phonetic Association* (1949), een brochure waarin voor 50 talen een fonetische transcriptie is opgenomen van een verhaaltje over de noordenwind en de zon. Daarbij werd voor iedere tekst de frequentie van *klanken* bepaald. Echter ook bij deze methode kunnen kanttekeningen geplaatst worden. Stel we hebben drie talen met de volgende klankpercentages:

	[e]	[ɪ]	[u]
taal 1	100%	0%	0%
taal 2	0%	100%	0%
taal 3	0%	0%	100%

Met de klankfrequentiemethode komt niet uit de verf dat de mate van verwantschap tussen taal 1 en taal 2 hoger is dan de mate van verwantschap tussen taal 1 en taal 3, en ook hoger dan de mate van verwantschap tussen taal 2 en taal 3. H & H schrijven in de inleiding dan ook dat zij al ras tot het

inzicht kwamen dat een meer verfijnde aanpak gewenst was. Daartoe ontwikkelden zij de featurefrequentiemethode (FFM).

Klanken kunnen beschreven worden door een reeks distinctieve kenmerken. De vocalen bijvoorbeeld kunnen verdeeld worden in klanken die voor, in het midden, of achter in de mond uitgesproken worden (beschreven door de kenmerken 'voor' en 'achter'), of in lage, centrale of hoge klanken (beschreven door het kenmerk 'laag'), of in klanken die met gespreide of geronde lippen uitgesproken worden (beschreven door het kenmerk 'rond'). Deze kenmerken worden vaak aangeduid met de engelse term *features*. Wanneer we nu een fonetische transcriptie hebben, kunnen we tellen hoeveel van de klanken bijvoorbeeld voor in mond worden uitgesproken, hoeveel lage klanken er zijn, of hoeveel klanken met geronde lippen worden uitgesproken. Met andere woorden: we bepalen de frequenties van de features. De featurefrequentiemethode bepaalt voor een gegeven transcriptie de frequenties van een vooraf vastgestelde reeks features. Het resultaat is een histogram. Op bladzijde 9 geven H & H het histogram voor het Kerkraads, en op bladzijde 10 het histogram voor het Gronings van Wagenborgen. De frequenties worden in deze beide histogrammen in percentages (gedeeld door 100) weergegeven. We nemen aan dat bij verdere berekeningen op basis van histogrammen dan ook steeds de percentages gebruikt worden en dat betekent dat het aantal klanken per tekst van een variëteit geen rol speelt, hoewel het aantal klanksegmenten wel als een dialect-specifieke eigenschap lijkt te mogen worden beschouwd. H & H melden zelf in hoofdstuk 2 (blz. 30) dat de tekst van het Gelders van Barneveld 4254 klanksegmenten bevat, terwijl de tekst van het Gronings van Bierum 3425 segmenten bevat. En H & H merken dan op: "Overigens, van de acht plaatsen onder aan de lijst werden er zeven opgenomen door een Groninger, landgenoten die in meerdere opzichten famous zijn om hun zuinigheid."

Het is nu mogelijk de mate van verschil tussen twee histogrammen te bepalen, maar ook kan de mate van overeenkomst bepaald worden. Als afstandsmaat noemen H & H de Manhattan-afstand. Deze is gedefinieerd als de som van de absolute verschillen. Als gelijkheidsmaat noemen H & H Pearson's product-moment correlatiecoëfficiënt. Heeft deze de waarde -1, dan is de mate van overeenkomst minimaal, en heeft deze de waarde +1, dan is de mate van overeenkomst maximaal.

We zagen dat de frequenties als percentages verwerkt worden, zodat het aantal klanken in de transcripties nooit een rol speelt. Zouden de frequenties gegeven zijn als absolute aantallen, dan speelt het aantal klanken in de transcriptie bij gebruik van de Manhattan-afstand een rol, maar bij gebruik van de Pearson's product-moment correlatiecoëfficiënt speelt het dan nog steeds geen rol.

In het vervolg van het boek hebben H & H steeds de Pearson's product-moment correlatiecoëfficiënt ( $r$ ) gebruikt. Om deze reden hebben we deze maat in onze kloon ook gebruikt, echter niet als gelijkheidsmaat, maar als afstandsmaat door berekening van  $1-r$ , waardoor 0 de betekenis heeft dat de overeenkomst maximaal, en 2 dat de overeenkomst minimaal is. Deze keuze maakten we met het oog op ons cluster-analyse programma dat een afstanden-matrix als invoer verwacht.

### *1.2 De classificatie van variëteiten*

Op basis van de afstands- of gelijkheidsmatrix kan een hiërarchische cluster-analyse uitgevoerd worden. Het doel daarvan is om belangrijke groepen (of clusters) in complexe data te identificeren. Wanneer bovendien een cluster subclusters kan bevatten (en een subcluster subsubclusters enz.), wordt daarmee een hiërarchie gegeven en spreekt men van 'hiërarchische cluster-analyse'. Het resultaat is een complete hiërarchische groepering van de variëteiten. Dit wordt gevisualiseerd in de vorm van een dendrogram, een boom waarin de bladeren overeenkomen met de variëteiten en de lengtes van de takken corresponderen met de afstanden.

H & H noemen als mogelijkheid voor de uitvoering van een cluster-analyse het gebruik van SPSS (een gangbaar statistisch computerprogramma). SPSS versie 8.0.0 (een al wat oudere versie uit 1997) biedt in totaal zeven verschillende clustermethoden. H & H schrijven dat alle in het boek getoonde dendrogrammen tot stand zijn gekomen "volgens de gangbare methode van 'Average Linkage (Between Groups)'" (blz. 56). Ook op blz. 12 wordt erop gewezen dat dit de meest gangbare clusteringmethode is. We denken dat dit een veilige keuze is, al zouden we ons kunnen voorstellen dat men ook op andere wijze tot een keuze zou kunnen komen, bijvoorbeeld door de bespreking van de verschillende methoden in Jain & Dubes (1988) te overwegen.

## 2 *Het gebruik van de RND als bron*

In hoofdstuk 2 bespreken H & H de RND als bron op basis waarvan de featurefrequentiemethode wordt toegepast. In de inleiding van dit hoofdstuk geven de schrijvers aan dat ze graag de uitspraak van J. Goossens onderschrijven dat de RND "voor het onderzoek verreweg het belangrijkste werkinstrument is geworden". Het is een citaat uit 1972, en in die tijd waren er inderdaad geen betere alternatieven. Echter op het moment dat het boek van H & H verscheen (2001), waren inmiddels meerdere delen van de *Fonologische Atlas van de Nederlandse Dialecten* (FAND) (1998-2000) verschenen. Deze atlas heeft als grote voordelen ten opzichte van de RND dat het tijdsinterval waarin de opnamen gemaakt werden veel kleiner is, en de transcripties veel consistentier zijn.

Als belangrijkste problemen waarmee men bij gebruik van de RND te maken krijgt noemen H & H de inhoud van de vragenlijst, en de aard en de betrouwbaarheid van de notaties. Beide problemen worden ook uitvoerig besproken in Heeringa (2001). De problemen betreffende verschillen in tijd, vragenlijsten en notatie bespreken we hieronder kort in respectievelijk de secties 2.1, 2.2 en 2.3. In 2.4 bespreken we de keuze van dialectplaatsen die H & H maakten, en de kleine selectie die wij zelf maakten. In 2.5 bespreken we de digitalisatie van de RND-teksten.

### 2.1 *Vershil in tijd*

Op bladzijde 18 wijzen H & H erop dat tussen de verschijning van deel 1 van de RND in 1925 en deel 14 in 1982 meer dan een halve eeuw tijdsverschil zit. Ze schrijven hier onder andere het volgende over: "De opnamen zijn begin jaren '20 gestart in Vlaanderen en in het midden van de jaren '70 in Oost-Nederland voltooid. Bij de geleidelijke verschuiving van dialect naar algemene taal, die in deze periode in versneld tempo plaats had, loopt Noord-Nederland duidelijk voor op Vlaanderen, zie Hoppenbrouwers (1990) en de daar genoemde literatuur."

H & H doen hier alleen een constatering, maar geven geen oplossing, vermoedelijk omdat die er ook niet is. Wel blijkt dat de tijdsverschillen tussen delen van gebieden die aan elkaar grenzen veel kleiner zijn, ofschoon

deze ook kunnen oplopen tot 30 jaar. De oudste opnamen in deel V (Zeeuwse eilanden) zijn gemaakt in 1932, terwijl de nieuwste opnamen in deel XI (Zuid-Holland en Utrecht) gemaakt zijn in 1962. We kunnen ons dus troosten met de hoop dat vertekening voornamelijk optreedt bij vergelijking van dialecten die geografisch ver van elkaar verwijderd zijn.

## 2.2 *Verschillende vragenlijsten*

Zoals H & H aangeven is niet voor ieder deel van de RND dezelfde vragenlijst gebruikt. In deel 6 (West-Vlaanderen en Fransch-Vlaanderen) wordt gemeld dat bij het interviewen bleek dat de zegslieden diverse woorden uit de vragenlijst niet kenden. Het probleem werd opgelost door ze te vervangen door de Franse equivalenten. In deel 15 (Friesland) zijn twee vragenlijsten afgedrukt, namelijk een Nederlandse en een Friese lijst. In de meeste plaatsen werd de Friese vragenlijst gebruikt. Echter niet alle woorden in de Friese lijst zijn semantisch gelijk aan die in de Nederlandse vragenlijst, bijvoorbeeld *vastenavond* (Nederlands) / *Sinteklaezjen* (Fries) (zin 15), *metselaar* (Nederlands) / *timmerman* (Fries). H & H noemen ook *kaarten* (Nederlands) / *damjen* (Fries (zin 65)). Op blz. 22/23 melden H & H het hiermee samenhangende probleem dat de genoteerde dialectvormen soms nogal vrij vertaald werden.

De FFM kan volgens H & H echter zonder bezwaar op het RND-materiaal toegepast worden. Op blz. 23 lezen we: "Dit bezwaar weegt minder zwaar als het gaat om een globale doorlichting van de klanksegmenten zoals in de FFM gebeurt. Voor de analyse op distinctieve kenmerken en de onderlinge vergelijking van de dialecten in kwestie zijn ze zeker geschikt."

Het bezwaar weegt dus "minder zwaar" bij gebruik van de FFM. Bij gebruik van de FFM is het materiaal wel degelijk bruikbaar. In Heeringa (2001) wordt echter voorgesteld om woorden die problematisch zijn als gevolg van verschillen in vragenlijsten, gewoon weg te laten.

## 2.3 *Verschillen in notatie*

H & H wijzen er ook op dat niet alle transcribenten eenzelfde wijze van noteren hadden, ondanks de grondige instructies van Blancquaert (sectie 2.2, blz. 23/24). We missen bij de bespreking echter een aantal punten die we

graag aan de orde hadden zien gesteld worden, punten die eveneens de onderlinge consistentie van transcripties van verschillende transcribenten betreffen.

In de eerste plaats betreft dit de notatie van de ee, oo en eu voor de r. Volgens het fonetisch overzicht in deel 13 zouden deze klanken moeten worden genoteerd als [r], [ʊ] en [ʌ]. Echter volgens het fonetisch overzicht in deel 16 zouden deze klanken moeten worden genoteerd als [e:], [o:] en [ø:].

Andere notatieproblemen betreffen woorden als 'zwemmen', 'spinnen' en 'springen' wanneer in de slotlettergreep 'en' de sjwa ontbreekt. Bijvoorbeeld voor Vlaanderen vinden we voor *zwemmen* (zin 42) onder andere de volgende transcriptie: [zwɛːmn] (Ronse), [zwæːm<sup>n</sup>] (Assenede), [zwæm̩:] (Damme). De vraag is of de uitgangen [mn], [m<sup>n</sup>] en [m̩:] verschillende uitspraken representeren. In het dialect van Zelzate vinden we [zwæːmn] (zwemmen) en [blʊːm:] (bloemen). In deel 12 wordt *bloemen* (zin 2) door Hol genoteerd als [blum̩] (Spankeren) en door Entjes als [bloːm̩] (Laren).

Iets vergelijkbaars vinden we bij woorden als 'geroepen', 'hebben', 'bakken' en 'geschlagen' (Duits). Voor Vlaanderen vinden we voor *geroepen* (zin 35) onder andere de volgende transcripties: [gəroːp̩n] (Ronse), [gəruːp<sup>n</sup>] (Bottelare), [g, ʔroːp̩n] (Damme), [g, əroːp<sup>m</sup>] (Oostkamp). In deel 12 wordt *dopen* (zin 137) door Hol genoteerd als [døːp̩n] (Spankeren) en door Entjes als [døːp̩m] (Groenlo).

De door ons genoemde problemen vonden we dus niet besproken in het boek van H & H. Wanneer echter voor deze problemen geen oplossing is toegepast, moet wel sprake zijn van enige vertekening in de resultaten. Onze kloon gebruikt de in Heeringa (2001) voorgestelde oplossingen.

Een ander probleem dat H & H noemen is de wijze waarop de invloed van assimilatie op de uitspraak genoteerd wordt (bladzijde 31). Daarbij lijkt het soms alsof de systematisch-fonologische interpretatie een rol speelt. Dit leidt dan bij de assimilatie van stem tot notaties als [nit ʏe:l] met onder de [v] het teken dat aangeeft dat de klank geheel of gedeeltelijk stemloos is. In dat geval hebben H & H de stemloze [v] vervangen door een [f] en als zodanig verwerkt. In 3.4 beschrijven we hoe onze kloon hiermee omgaat. Verder treedt assimilatie op in gevallen als [miŋ kɪnt]. Hierbij trekt de [k] die achter in de mond gearticuleerd wordt de voorafgaande nasaal naar zich toe. H & H spreken hierbij van "organische assimilatie". In verband hiermee hebben H

& H transcripties als [min kint] steeds gewijzigd in [mij kint]. Onze kloon laat de transcriptie echter ongewijzigd.

#### 2.4 De selectie van dialectplaatsen

Na ingegaan te zijn op beperkingen van de RND voor wat betreft de notatie, merken H & H ten aanzien van het gebruik van de FFM en het *aantal* plaatsen op (blz. 24): "Ongetwijfeld zal de uitkomst zijn begunstigd door het feit dat wij met een meer grofmazig net genoeg moesten nemen." Over de *keuze* van de 156 plaatsen schrijven H & H op blz 25: "Bij de keuze van de plaatsen hebben we gestreefd naar een redelijke spreiding over het taalgebied. We hebben daarbij uiteraard rekening gehouden met bekende dialectgrenzen door steeds plaatsen aan weerszijde van zo'n bekende grenslijn te kiezen. Ook opvallende dialecten en erkende probleemgevallen hebben we zoveel mogelijk bij onze klassering betrokken."

Hieraan zijn nog vier teksten toegevoegd, te weten een transcriptie van het Algemeen Nederlands door H & H, een RND-transcriptie van het Amsterdams in de Jordaan (naast die van Kattenburg, zie blz. 121 en 172-174), een aangepaste versie van de RND-tekst van het Zwols waarin de huig-R vervangen is door de tongpunt-r (zie blz. 93), en een aangepaste versie van de RND-tekst van het Schevenings waarin de schraap-R vervangen is door de tongpunt-r (zie blz. 160). Een precieze beschrijving van de totstandkoming van de transcriptie voor het Algemeen Nederlands konden we in het boek niet vinden. Op blz. 124 had daarover misschien iets gezegd kunnen worden.

Ten behoeve van onze kloon hebben we zeven variëteiten gekozen uit de lijst van 156 variëteiten van H & H. In H & H tonen de kaarten 3 t/m 7 de rangorde van de variëteiten vanuit respectievelijk het Gronings van Bierum, het Twents van Losser, het Fries van Grouw, het Nederlands, het Oost-Brabants van Rosmalen, het Zeeuws van Borssele, het Vlaams van Poeke en het Limburgs van Kerkrade. Omdat van het Nederlands geen tekst aanwezig is in de RND, hebben we ons beperkt tot de dialecten van Bierum, Losser, Grouw, Rosmalen, Borssele, Poeke en Kerkrade (zie Figuur 1). De rangorde voor elk van deze dialecten is niet alleen in de vorm van een kaart, maar ook in de vorm van een lijst gegeven. In zo'n lijst zijn ook de correlatiecoëfficiënten ten opzichte van het uitgangsdialect gegeven. Deze zeven lijs-

ten bieden dus de mogelijkheid om een gelijkheidsmatrix af te leiden die de onderlinge correlatiecoëfficiënten tussen de zeven dialecten bevat.



Figuur 1: *De zeven geselecteerde variëteiten uit de lijst van 156 variëteiten uit H & H.*

### 2.5 De digitalisatie van de teksten

Bij het digitaliseren van de teksten hebben we de volgende regels gehanteerd:

1. Woorden die in isolatie zijn opgevraagd en getranscribeerd zijn niet gedigitaliseerd, omdat deze relatief vaak ontbreken. Het betreft opsommingen van begrippen als bijvoorbeeld in zin 19 (spin, spinnewebben, ragebol), maar ook rijtjes werkwoordsvormen als bijvoorbeeld in zin 113 (bakken, ik bak, jij bakt, etc.).
2. Ook incomplete zinnen zijn gedigitaliseerd.

3. Soms worden in een RND-transcriptie meerdere varianten voor hetzelfde woord of voor dezelfde woordgroep gegeven. H & H schrijven dat zij zich in principe beperkt hebben tot het eerstgenoemde alternatief. Wij digitaliseerden eveneens alleen de eerste variant, tenzij aangegeven is welke variant de beste of de meest gangbare is.
4. Suprasegmentele of diacritische tekens tussen haakjes zijn weggelaten.
5. Klanken tussen haakjes worden bij digitalisatie genoteerd als klanken in superschrift (weergave waarbij de tekens iets kleiner en hoger worden gedrukt).
6. Klanken in superschrift en tussen haakjes zijn weggelaten.

In Tabel 3 op bladzijde 37 t/m 41 geven H & H de codes van de door hen geaccepteerde combinaties van een klank met één of meer suprasegmentele en/of diacritische tekens. Deze worden steeds gecodeerd met een combinatie van twee tekens die elk gekozen zijn uit de verzameling tekens die standaard op ieder toetsenbord te vinden zijn. Het coderingssysteem dat wij hebben gebruikt wordt beschreven in Heeringa (2001), en is ook te vinden op:

<http://www.let.rug.nl/~heeringa/dialectology/rnd/system>

In dit systeem kunnen alle mogelijke combinaties van een klank met één of meer suprasegmentele en/of diacritische tekens worden gecodeerd. De lengte van de codering van één combinatie hangt af van het aantal suprasegmentele en diacritische tekens. Met behulp van dit systeem duurt de codering van één tekst ongeveer twee uur.

Omdat we geen losse begrippen hebben gedigitaliseerd, maar alleen 'echte' zinnen, bleven in totaal 130 zinnen over. Zowel op bladzijde 9 als op bladzijde 17 wijzen H & H erop dat teksten gebruikt moeten worden van voldoende omvang. We hebben daarom voor elk van onze zeven variëteiten het aantal klanken bepaald. In het tabelletje hieronder geven de getallen links het aantal klanken waarbij een diftong als twee klanken is geteld, en de getallen rechts het aantal klanken waarbij een diftong als één klank is geteld.

Bierum:	3080	2962
Losser:	3005	2965
Grouw:	3267	3179

Rosmalen:	3392	3342
Borssele:	3082	3032
Poeke:	3328	3227
Kerkrade:	3380	3327

We hopen dat hiermee het aantal klanken per tekst groot genoeg is.

### 3 De featurematrix

Om de frequenties van features te kunnen bepalen, moeten de klanken in de transcripties van de RND in termen van features zijn gedefinieerd. De eerste stap die gezet moet worden is de keuze van de juiste features. H & H schrijven dat zij daarbij uitgegaan zijn van *The Sound Pattern of English* (SPE) van Chomsky & Halle (1968), een systeem met articulatorische kenmerken. Zij wijzen erop dat het een standaardwerk betreft (blz. 33) dat alom werd nagevolgd in studies en handboeken over moderne fonologie (blz. 34). Toch hadden we graag wat meer argumentatie gezien voor de keuze van dit uitgangspunt. Waarom werd bijvoorbeeld het door H & H zelf genoemde akoestische featuresysteem van Jakobson, Fant & Halle (1952) niet als basis gekozen, of andere meer recente systemen?

Het SPE-systeem is door H & H zodanig aangevuld dat met het gebruik daarvan zoveel mogelijk recht wordt gedaan aan het RND-materiaal. Daarmee ontstond een systeem met de volgende 21 features:

- |             |                 |
|-------------|-----------------|
| 1. Vocaal   | 11. Consonant   |
| 2. Voor     | 12. Anterieur   |
| 3. Achter   | 13. Coronaal    |
| 4. Rond     | 14. Posterieur  |
| 5. Laag     | 15. Laryngaal   |
| 6. Polair   | 16. Sonorant    |
| 7. Lang     | 17. Stemhebbend |
| 8. Perifeer | 18. Hoog        |
| 9. Diftong  | 19. Continuant  |
| 10. Nasaal  | 20. Lateraal    |
|             | 21. Syllabisch  |

De tabel waarin voor alle klanken de featurespecificaties gegeven zijn, kan in het boek van H & H gevonden worden op de bladzijden 37 t/m 41 als Tabel 3.

### 3.1 De vocalen

In deze sectie willen we onder meer het verband duidelijk maken tussen de IPA-klinkerdriehoek en de tabel van H & H. In navolging van Ladefoged (1975) gebruiken H & H het feature *perifeer*. Over perifere klinkers schrijven H & H: "Zowel in akoestisch als in articulatorisch opzicht bevinden deze onstabiele klinkers zich aan de buitenkant van de klinkerruimte".

Korte vocalen zijn altijd gespecificeerd als [-perifeer]. Een aantal vocalen wordt echter als [+perifeer] gespecificeerd wanneer zij halflang of lang zijn. In de tabel zien we dat de [e] en [ɪ], de [ø] en [ʏ], en de [o] en [ʊ] gelijk aan elkaar zijn indien ze kort zijn. Echter indien ze halflang of lang zijn, zijn [e], [ø] en [o] als [-perifeer], en de [ɪ], [ʏ] en [ʊ] als [+perifeer] gespecificeerd.

Uit de tabel kunnen we opmaken dat van de 28 vocalen in het IPA de volgende klanken als [+perifeer] gespecificeerd worden indien ze halflang of lang zijn:

	voor		midden		achter	
polair hoog	i	y			ɯ	u
hoog	e	ø			ɤ	o
laag						
polair laag	æ		a	ɶ	ɑ	ɒ

In dit tabelletje geldt dat een element links in een cel gespreid is, en een element rechts in een cel gerond. De [ɯ], [ɤ] en [ɶ] zijn door H & H niet gegeven omdat ze ook niet in de RND gegeven zijn. De [ɒ] is niet door H & H gegeven, hoewel deze wel in deel 1 van de RND gebruikt wordt. Verder kunnen we uit de tabel opmaken dat van de 28 vocalen in het IPA de volgende klanken altijd als [-perifeer] gespecificeerd zijn, ongeacht hun lengte:

	voor		midden		achter	
polair hoog			i	ɨ		
hoog	ɪ	ʏ	ə	ø		ʊ
laag	ɛ	œ	ɜ	ɞ	ʌ	ɔ
polair laag			ɚ			

Ook in dit tabelletje geldt dat een element links in een cel gespreid, en een element rechts in een cel gerond is. De *i*, *ɜ*, *ɚ* en de *ə* die volgens het IPA gedefinieerd is als een halfronde centrale klank die precies tussen hoog en laag inligt, zijn door H & H niet gegeven omdat ze ook niet in de RND gegeven zijn. De *ʏ* en *ə* worden zowel in de RND als in de tabel van H & H weergegeven als respectievelijk *ʌ* en *ə*. De *ø* en de *ɞ* worden in de tabel van H & H weergegeven als respectievelijk een *u* met een horizontaal streepje er dwars doorheen, en een *ɔ* met een horizontaal streepje er dwars doorheen. Hoewel de *ɨ*, *ø* en *ɞ* ook door H & H worden gegeven, hebben wij deze klanken in de RND niet gevonden.

In het SPE, bedoeld voor Engelse segmenten, kunnen precies drie hoogtegraden gedefinieerd worden met behulp van de features *high* en *low*. Voor de vocalen in Nederlandse dialecten is het noodzakelijk vier hoogtegraden te kunnen onderscheiden (zie blz. 35). In het systeem van H & H worden deze door middel van de features *laag* en *polair* als volgt gedefinieerd:

laag	polair	
close	-	+
close-mid	-	-
open-mid	+	-
open	+	+

Gevolg is dat het verschil tussen close en open-mid groter is dan het verschil tussen close en open. Eveneens is het verschil tussen open en close-mid groter dan het verschil tussen open en close.

In de tabel zijn bijna alle monoftongen gedefinieerd voor drie lengtes: kort, halflang en lang. Alleen de *ə*, *ɨ*, *ø* en *ɞ* zijn uitsluitend gedefinieerd voor de lengte kort. In onze kloon zijn *ə*, *ɨ*, *ø* en *ɞ* ook gedefinieerd voor de lengtes halflang en lang. In de tabel van H & H wordt echter geen onderscheid gemaakt tussen halflange en lange klanken, beide klanken zijn gede-

finieerd als +lang. Voor alle monoftongen (eventueel in verschillende lengte's) zijn zowel een halfnasale als een nasale versie gedefinieerd met respectievelijk \*nas en +nas.

### 3.2 Consonanten

In deze sectie willen we nagaan hoe de tabel van H & H gerelateerd is aan de IPA-medeklinkertabel. Daarbij willen we alle medeklinkers uit de IPA-tabel betrekken. Voor de gegeven klanken in de tabel van H & H laten we zien welk verband er met de IPA-tabel is, en voor de klanken die niet gegeven zijn laten we zien hoe het verband gelegd kan worden. Voor de IPA-kolomcategorieën geldt:

<i>IPA</i>	<i>ant</i>	<i>cor</i>	<i>post</i>	<i>lar</i>	<i>hoog</i>
bilabial	+	-	-	-	-
labiodental	+	-	-	-	-
dental	+	+	-	-	-
alveolar	+	+	-	-	-
postalveolar	-	+	-	-	+
retroflex	-	+	-	-	+
palatal	-	+	-	-	+
velar	-	-	+	-	+
uvular	-	-	+	-	+
pharyngeal	-	-	-	-	-
glottal	-	-	-	+	-

Alle approximanten zijn gespecificeerd als +hoog.

Voor de IPA-rijcategorieën geldt:

<i>IPA</i>	<i>nas</i>	<i>cons</i>	<i>son</i>	<i>cont</i>	<i>lat</i>
plosive	-	+	-	-	-
nasal	+	+	+	-	-
trill	-	+	+	+	-
tap or flap	-	+	+	+	-
fricative	-	+	-	+	-
lat. fric.	-	+	-	+	+
approximant	-	-	+	+	-
lat. appr.	-	+	+	+	+

De w (bilabiaal en approximant) is ook [+rond]. Het onderscheid stemhebbend/stemloos is door H & H op dezelfde manier gedefinieerd als in de IPA-medeklinkertabel. In de tabel van H & H worden alleen de klanken gegeven die ook in de RND voorkomen. In de tabel is de h uit de RND gedefinieerd als de fi in het IPA. Met de 21 features van H & H wordt voor alle consonanten uit de RND een unieke definitie gegeven. Dit lukt niet voor alle medeklinkers uit het IPA. H & H schrijven dan ook op bladzijde 9: "Met een kleine dertig kenmerken zouden we alle klanken kunnen karakteriseren die in natuurlijke talen voorkomen".

### 3.3 Diftongen

Op bladzijde 35 schrijven H & H dat zij reguliere diftongen beschouwen als segmentele eenheden. H & H maken onderscheid tussen sluitende diftongen en centrerende diftongen. Een sluitende diftong is een lange klinker waarbij sprake is van een verschuiving naar een niet-centrale positie in de klinkerruimte. Een centrerende diftong is een klinker waarbij juist sprake is van een verschuiving naar een centrale positie in de klinkerruimte, de sjwa. De waarden voor de vocaalfeatures worden afgeleid uit de feature-waarden van de vocaal die correspondeert met de beginpositie en de vocaal die correspondeert met de eindpositie.

In de RND is het tweede element vaak in superschrift genoteerd of met een kleinere letter weergegeven, mogelijk om aan te geven dat het eerste element meer bepalend is dan het tweede element. In de tabel van H & H zijn beide elementen van een tweeklank altijd op dezelfde manier weergegeven. We beschouwen nu twee opeenvolgende klanken in een transcriptie als een diftong als we de combinatie van die twee klanken vinden in de tabel. De notatie van het tweede element speelt hierbij geen rol, en wordt verder niet verwerkt.

Wanneer als tweede element van een sluitende diftong een j of een w is genoteerd, gaan we bij de bepaling van de featurewaarden voor de diftong uit van de featurewaarden van respectievelijk de i en de u. De vocaalfeatures van sluitende diftongen worden als volgt gedefinieerd:

voor:	gemiddelde van begin- en eindklank
achter:	gemiddelde van begin- en eindklank

rond:	gemiddelde van begin- en eindklank
laag:	waarde van beginklank
polair:	waarde van beginklank
lang:	altijd +lang
perifeer:	altijd +peri
diftong:	altijd +dift

In de tabel van H & H vinden we de volgende sluitende diftongen: [ɤy] (gedefinieerd als [œy]), [ɔi], [ɔ'i], [ei], [e'i], [øi] (gedefinieerd als [øy]), [øy], [ou], [o'u], [oj] (gedefinieerd als [ɔj]), [ɛi], [ɛ'i], [ɛj], [œi] (gedefinieerd als [œy]), [œy], [œ'y], [ɔu], [ɔ'u], [ɔw], [ɔi], [ɔ'i], [ɔj], [æi], [æ'i], [æj], [ay] (gedefinieerd als [œy]), [a'y] (gedefinieerd als [œy]), [ai] (gedefinieerd als [æi]), [au], [a'u], [aw], [aɪ], [a'i], [aj]. Hieraan voegden we de volgende tweeklanken toe: [ɔu], [ɔ'u], [ɔw], [ɔi], [ɔ'i], [ɔj].

Als een sluitende diftong genoteerd is als een combinatie klinker+klinker, zien we dat er zowel diftongen zijn waarvan het eerste element kort is, als diftongen waarvan het eerste element halflang is. Het eerste element is nooit lang. Als een sluitende diftong genoteerd is als een combinatie klinker+medeklinker, zien we dat het eerste element altijd kort is. Het eerste element is nooit halflang of lang. Op bladzijde 35 schrijven H & H dan ook: "Alleen sequentiële tweeklanken van het type [a:j], bestaande uit een lange klinker gevolgd door i of j zoals in fraai en mooi, worden bifonematisch geïnterpreteerd."

De vocaalfeatures van centrerende diftongen worden als volgt gedefinieerd:

voor:	gemiddelde van begin- en eindklank
achter:	gemiddelde van begin- en eindklank
rond:	gemiddelde van begin- en eindklank
laag:	waarde van beginklank
polair:	waarde van beginklank
lang:	altijd +lang
perifeer:	altijd -peri
diftong:	altijd -dift

In de tabel van H & H vinden we de volgende centrerende diftongen: [iɔ], [yɔ], [uɔ], [Iɔ], [Yɔ], [ʊɔ], [eɔ], [øɔ], [oɔ], [ɛɔ], [œɔ], [ɔɔ], [æɔ], [aɔ], [ɑɔ]. Hieraan voegden we de volgende tweeklank toe: [ʊɔ].

Als een centrerende diftong bestaat uit de combinatie klinker+medeklinker, zien we dat het eerste element altijd kort is. Het eerste element is nooit halflang of lang. Voor alle centrerende diftongen is ook een halfnasale en een nasale versie gedefinieerd.

### 3.4 Suprasegmentele en diacritische tekens

#### *Lengte*

In de RND worden klanken soms in superschrift genoteerd, met een kleinere letter weergegeven of tussen haakjes geplaatst. Ook kunnen klanken in superschrift tussen haakjes staan. H & H beschouwen tekens in superschrift, tekens tussen haakjes, en tekens die in superschrift tussen haakjes staan als hetzelfde (bladzijde 47). H & H schrijven dat bij enkele sporadisch voorkomende gevallen bij de Groningse en Oost- en West-Vlaamse dialecten de oneven gevallen als volle representanten werden overgenomen, en de even gevallen werden weggelaten. Jammer dat zij niet precies vermelden om welke klanken het gaat.

Ten aanzien van de [r], [g], [m], [n], [ɲ] en [h] kozen H & H een andere aanpak. De tabel werd uitgebreid met featurespecificaties voor de zwakke versies van deze klanken. Zowel de oneven als de even gevallen worden nu verwerkt waarbij de featurespecificaties van de zwakke versies gebruikt worden.

Bij de digitalisatie van de data voor onze kloon hebben we klanken die in superschrift of met een kleinere letter zijn weergegeven, en bovendien ook nog eens tussen haakjes zijn genoteerd, weggelaten. Vervolgens worden door onze kloon bij de bepaling van de frequenties de features van klanken die òf in superschrift, òf in een kleinere letter, òf tussen haakjes zijn genoteerd één keer geteld, en alle overige klanken twee keer. Het maakt daarbij niet uit of het monoftongen of consonanten betreft.

Halflange en lange klanken tellen ook steeds twee keer. Een mogelijkheid is om halflange klanken drie keer, en lange klanken vier keer te tellen. Voor de vocalen wordt lengte echter verwerkt met het feature *lang*, terwijl

bij perifere vocalen ook het feature *perifeer* een rol speelt (zie hierboven, 3.1). Lange consonanten worden beschouwd als syllabische klanken (zie hieronder).

#### *Stemhebbend/stemloos*

Op bladzijde 31 wijzen H & H erop dat het soms lijkt of de systematisch-fonologische interpretatie een rol speelt. Zoals in sectie 2.3 is vermeld leidt dit bij de assimilatie van stem tot notaties als [nit ye:l] met onder de v het teken dat aangeeft dat de klank geheel of gedeeltelijk stemloos is. In dat geval hebben H & H de stemloze v vervangen door een f en als zodanig verwerkt. Onze kloon gaat iets voorzichtiger te werk, en definieert een [v̥] als een [v] met \*stem (dus half stemhebbend en half stemloos). Een eventuele [f̥] (stemhebbende f) wordt gedefinieerd als een [f] met \*stem (dus half stemloos en half stemhebbend).

#### *Nasalisatie*

Op bladzijde 46 beschrijven H & H de verwerking van het verschijnsel nasaliteit. Het gaat dan om de voorkomens van de combinatie ((half)nasale) klinker + (kleine) nasale medeklinker. In de RND bestaat zowel voor 'half-nasaal' als voor 'nasaal' een diacritisch teken. Zoals we hierboven zagen kan ieder klank, en dus ook een nasale medeklinker, kleiner gedrukt of in superschrift worden weergegeven. De volgende combinaties hebben nu onze aandacht:

1. Niet nasale klinker + kleine nasaal
2. Niet nasale klinker + nasaal
3. Halfnasale klinker + niets
4. Halfnasale klinker + niet-nasale medeklinker
5. Halfnasale klinker + kleine nasale medeklinker
6. Halfnasale klinker + nasale medeklinker
7. Nasale klinker + niets
8. Nasale klinker + niet-nasale medeklinker
9. Nasale klinker + kleine nasale medeklinker
10. Nasale klinker + nasale medeklinker

Begrijpen we de uitleg van H & H goed, dan zijn de combinaties 1, 3 en 4 onmogelijk vanwege onderwaardering van de nasaliteit, en de combinaties 6, 9 en 10 zijn onmogelijk vanwege overwaardering van de nasaliteit. Alleen de combinaties 2, 5, 7 en 8 zijn mogelijk. De door H & H als onmogelijk beoordeelde combinaties moeten dus worden 'gecorrigeerd'. We zijn er echter niet van overtuigd dat deze combinaties onmogelijk zijn. Onze kloon gaat daarom uit van de notatie zoals die gegeven is, waarbij het door de RND gegeven onderscheid halflang/lang in rekening gebracht wordt.

### *Syllabiciteit*

H & H geven aan dat de aanduidingen voor vokalisering van de sonoranten (de nasalen, de l en de r) in hun transcriptie zijn overgenomen (blz. 31). Voor gevokaliseerde klanken worden de features *voK* en *syll* positief gemarkeerd. Op bladzijde 43 wijzen H & H erop dat soms naast of in plaats van het teken voor vokalisering ook het lengteteken (: ) is gegeven. In al deze gevallen worden de features *voK* en *syll* positief gemarkeerd.

Behalve in de gevallen waarin H & H een medeklinker als syllabisch verwerken, verwerkt onze kloon een medeklinker óók als syllabisch indien op grond van de daaraan voorafgaande medeklinker voorspeld kan worden dat deze syllabisch moet zijn, ook al is die in de transcriptie niet als syllabisch (vocalisch of lang) gemarkeerd. Wanneer bijvoorbeeld een lettergreep eindigt op [pm], moet de [m] wel syllabisch zijn.

### *3.5 Redundantie*

Op bladzijde 43 schrijven H & H: "Zo zal een tekst met veel coronalen en anterieuren ook veel consonanten bevatten, en een taal met relatief veel vocalen, zal daardoor relatief weinig ruimte hebben voor consonanten. Teneinde deze oneigenlijke bron van negatieve of positieve correlatie uit te sluiten, hebben we besloten de markeringen voor de kenmerken die op grond van de regels voorspeld worden, niet in de beschouwing te betrekken."

In de featurematrix (blz. 37-41) wordt de aanwezigheid van een kenmerk aangegeven door een '+', gedeeltelijke aanwezigheid door een '\*', en afwezigheid door een '-'. Echter wanneer aanwezigheid of afwezigheid voorspeld

kan worden op basis van één of meer andere kenmerken, wordt respectievelijk een '1' of een '0' genoteerd.

Onze kloon maakt gebruik van een tabel waarin alleen enen, sterretjes en nullen worden gebruikt. Enen en nullen worden dan op dezelfde manier verwerkt als de plussen en minnen in de tabel van H & H. Op deze specificaties worden vervolgens de door H & H gegeven redundantieregels toegepast. Wanneer op basis van een regel blijkt dat een '1' redundant is, wordt deze gewijzigd in een '0'.

H & H geven vijf regels, waarbij regel 1 en regel 2 elk onderverdeeld zijn in vier subregels. Omdat onze kloon geen verschil kent tussen '+' en '1', en tussen '-' en '0', en redundante waarden altijd door een '0' worden weergegeven, is het nodig om eerst de regels 2a t/m 2d uit te voeren, en daarna de regels 1a t/m 1d. Op die manier blijft het effect precies hetzelfde als bij de werkwijze van H & H.

De volgende regels geven aan dat de kenmerken [+voor] t/m [+pol] het kenmerk [+vok] voorspellen:

1a +voor	→	+vok
1b +acht	→	+vok
1c +rond	→	+vok
1d +laag	→	+vok
1e +pol	→	+vok

Ten behoeve van de sluitende diftongen voegden we hieraan toe:

1a *voor +dift	→	+vok
1b *acht +dift	→	+vok
1c *rond +dift	→	+vok
1d *laag +dift	→	+vok
1e *pol +dift	→	+vok

Regel 1e hebben we toegevoegd ten behoeve van klanken die gemarkeerd zijn als -voor -acht -rond -laag en +pol. Klanken die als zodanig zijn gemarkeerd, komen in de featurematrix van H & H niet voor, omdat de matrix alleen specificaties bevat voor de klanken die in de RND gebruikt worden,

maar onze kloon is zodanig opgezet dat eventueel alle IPA-klinkers verwerkt kunnen worden.

Ten behoeve van de centrerende diftongen voegden we hieraan toe:

- 1a \*voor -dift → \*vok
- 1b \*acht -dift → \*vok
- 1c \*rond -dift → \*vok
- 1d \*laag -dift → \*vok
- 1e \*pol -dift → \*vok

Klinkers zijn altijd gemarkeerd als +son +stem +cont +syll. Dit wordt door de volgende regels aangegeven:

- 2a +vok -cons → +son
- 2b +vok -cons → +stem
- 2c +vok -cons → +cont
- 2d +vok -cons → +syll

De conditie -cons hebben we toegevoegd. Syllabische medeklinkers zijn ook +vok, maar omdat ze ook +cons zijn, worden ze door deze extra conditie uitgesloten, wat blijkens de featurematrix van H & H ook het geval moet zijn.

Ten behoeve van de centrerende diftongen voegden we hieraan toe:

- 2a \*vok -cons → +son
- 2b \*vok -cons → +stem
- 2c \*vok -cons → +cont
- 2d \*vok -cons → +syll

H & H melden dat voor de tweeklanken het kenmerk [+lang] straffeloos kan worden weggehaald. Met 'tweeklanken' zullen hier alleen de sluitende diftongen bedoeld zijn. Centrerende diftongen zijn gespecificeerd als [-dift].

- 3 +dift → +lang

Bij achtermedeklinkers is het kenmerk [+hoog] overtollig.

4 +post → +hoog

Hieruit leidden we ook af:

4 \*post → +hoog

H & H schrijven dat voor de sonorante laryngaal, de keelklank [h] het kenmerk [+continuant] voorspelbaar is, en geven vervolgens als regel: +lar +son → +cont. Bij navraag blijkt dat als tweede conditie niet +son maar +stem bedoeld is. In de featurematrix is de [h] dan ook als [-son] en [+stem] gespecificeerd. De regel wordt nu:

5 +lar +stem → +cont

Uit de featurematrix blijkt dat voor syllabisch medeklinkers het kenmerk [+vok] redundant is. We hebben daarom de volgende regel toegevoegd:

6 +syll +cons → +vok

#### 4 *Indeling van het Nederlandse taalgebied*

##### 4.1 *Traditionele dialectologie*

H & H openen hoofdstuk 4 van hun boek met een mooi overzicht van indelingsmethoden en -kaarten uit de traditionele dialectologie (sectie 4.1). Zij beschrijven dat dialecten geklassificeerd werden 1. op basis van Germaanse volkstammen (Winkler, Jellinghaus, Te Winkel, Van Ginneken, Lecoutere/Grootaers), 2. op grond van isoglossen (Weijnen), 3. naar het oordeel van dialectsprekers (Daan), en 4. met behulp van de structuurgeografische methode (Goossens).

Bij een indeling op basis van volkstammen lijkt de dialectoloog vooral op grond van zijn gehoor en zijn intuïtie tot een indeling te komen. De isoglossenmethode lijkt op het eerste gezicht meer verantwoord, maar H & H wijzen wel op de vraag hoe zwaar een verschijnsel gewogen moet worden (blz. 54). Als voorlopig eindpunt (blz. 48) noemen H & H de pijltjesmethodekaart van Daan. Zij merken op dat daarbij ook gebruik gemaakt is van al bekende isoglossen. Jammer is dat voor het Vlaamse deel de pijltjesmethode

niet is gebruikt, maar voor de bepaling van de indeling van dit gebied "ter zake kundige taalgeografen geraadpleegd" zijn. H & H melden dat de kaart in verkleinde vorm te vinden is in "De Vries, Willemyns en Burger (1993)" (blz. 55), maar onder de Verwijzingen achterin het boek ontbreekt deze bron.

#### 4.2 Indeling volgens FFM

In sectie 4.2 geven en bespreken H & H de hoofdingeling op basis van de FFM. Deze wordt op bladzijde 58 in de vorm van een dendrogram weergegeven. De indeling is als volgt:

1. Limburgs
2. De rest
  - 2.1 Saksisch
  - 2.2 De rest
    - 2.2.1 Fries
    - 2.2.2 Hollands, Noord-Brabants
    - 2.2.3 Noord-Belgisch

H & H wijzen erop dat in deze indeling het Fries en het Saksisch als groepen duidelijk terug te vinden zijn, maar het Frankisch opgesplitst is in twee groepen die volgens dit dendrogram ver van elkaar afstaan, namelijk een groep met daarin het Hollands, Brabants, Zeeuws en Vlaams, en een groep met daarin de Limburgse dialecten.

De vijf meest significante groepen worden volgens het dendrogram als volgt onderverdeeld:

1. Nedersaksisch
  - 1.1 GRONINGS EN NOORD-DRENTS, MIDDEN-DRENTS EN WESTERWOLDS, TWENTS
  - 1.2 Zuid-Drents en Noord-Overijssels, Terrassen naar de Twentse kern
2. Fries
  - 2.1 Fries
    - 2.1.1 DE FRIESE DIALECTEN
    - 2.1.2 Stadfries, Kollumerlands, Bildts, Stellingwerfs

- 2.2 Veluwse overgangsdialecten
- 3. Hollands, Noord-Brabants
  - 3.1 Hollands
    - 3.1.1 Noord-Hollands
    - 3.1.2 Zuid-Hollands en Utrechts
  - 3.2 Noord-Brabants
    - 3.2.1 OOST-BRABANTS
    - 3.2.2 Dialecten in het Gelders Rivierengebied, West-Brabants
- 4. Noord-Belgisch
  - 4.1 CENTRAAL BRABANTS
  - 4.2 Periferisch Brabants
    - 4.2.1 Zeeuws
    - 4.2.2 Brabants
  - 4.3 Periferisch Vlaams
  - 4.4 CENTRAAL VLAAMS
- 5 Limburgs

Vervolgens wordt de verdeling binnen de Nedersaksische dialecten (hoofdstuk 5), de Friese variëteiten (hoofdstuk 6), de Hollandse, Utrechtse en Veluwse dialecten (hoofdstuk 7), de Noord-Brabants dialecten (hoofdstuk 8), de Zeeuwse dialecten (hoofdstuk 9), de Belgisch-Brabantse dialecten (hoofdstuk 10), de Vlaamse dialecten (hoofdstuk 11) en de Limburgse dialecten (hoofdstuk 12) tot in detail besproken. Vele bronnen uit de traditionele dialectologie worden hierbij betrokken. Heel mooi is te zien hoe de FFM daar, waar verschillende bronnen elkaar tegenspreken, uitsluitsel geeft. Wat de indeling in hoofdstukken betreft volgden H & H niet consequent de indeling volgens het dendrogram wanneer het gaat om de Noord-Belgische en Zeeuwse dialecten.

H & H maken onderscheid tussen kerndialecten en randdialecten. In het overzicht hierboven zijn de kerndialecten in kleine hoofdletters weergegeven. In het Nedersaksische gebied zijn volgens H & H twee kernen aanwezig, namelijk één in Groningen en het aangrenzende deel van Drenthe, en één in Twente. In het dendrogram waarin de Nedersaksische dialecten nader worden bekeken (bladzijde 65), zien we twee hoofdgroepen, namelijk één met de kerndialecten, en één met de randdialecten. Dit dendrogram suggereert dat de verwantschap tussen de beide kernen (Groningen en het aan-

grenzende deel van Drenthe enerzijds, en Twente anderzijds) groter is dan de verwantschap tussen een kern en de aangrenzende randdialecten. H & H wijzen erop dat beide gebieden gesitueerd zijn aan de oostelijke landrand (blz. 64), en dat de tussenliggende strook "is gaan fungeren als een wig die beide kerngebieden uit elkaar heeft gedreven" als gevolg van "taalinvloeden vanuit het Hollands-Brabantse centrum." (blz. 68/69).

Opvallend in hun hoofdindeling is ook dat een groep van Veluwse overgangsdialecten geclusterd is met de Friese kern- en randdialecten. De verwantschap tussen deze Veluwse overgangsdialecten en het Fries is zelfs groter dan de verwantschap tussen de Nedersaksische kern- en randdialecten. Het groepje Veluwse overgangsdialecten wordt door H & H besproken in het hoofdstuk over de Hollandse, Utrechtse en Veluwse dialecten (blz. 132), en niet in het hoofdstuk over de Friese dialecten. Als verklaring voor de verwantschap met het Fries verwijzen H & H naar Van Loey (1970) die erop wijst dat de Veluwe in een ver verleden deel uitmaakte van het Friese rijk.

Uit de resultaten die H & H presenteren in sectie 7.4 blijkt dat het Algemeen Nederlands (AN) het sterkst verwant is met het dialect van Haarlem, dat valt onder het Noord-Hollands. Beschouwen we nu het dendrogram dat de hoofdindeling weergeeft, dan blijkt het Limburgs het verst van het Noord-Hollands verwijderd te zijn, gevolgd door het Nedersaksisch. H & H schrijven op bladzijde 201: "Uit ons onderzoek blijkt die afstand, waar het gaat om de klank- en vormverhoudingen, voor de Nedersaksische en de Limburgse dialecten nog groter te zijn dan voor de Friese varianten". Voor wie grote waarde hecht een aan bijzondere status van het Fries is dit een teleurstellende uitkomst.

## 5 *Polyphon versus de kloon*

Zoals gemeld gebruiken H & H het door hen ontwikkelde programma 'Polyphon' om de featurefrequenties te bepalen. In deze sectie willen we de resultaten van onze kloon vergelijken met de resultaten van 'Polyphon'.

### 5.1 *Afstanden*

Zoals gemeld in sectie 2.3 hebben we ten behoeve van onze kloon zeven variëteiten gekozen uit de lijst van 156 variëteiten van H & H. Elk van deze

zeven variëteiten is gegeven als uitgangspunt ten opzichte waarvan de rangorde van de overige variëteiten is gegeven. De zeven respectievelijke rangordelijsten bevatten alle informatie om een gelijkheidsmatrix af te leiden die de onderlinge correlatiecoëfficiënten tussen de zeven dialecten bevat. Daarmee kregen we de volgende matrix:

	Bierum	Losser	Grouw	Rosmalen	Borssele	Poeke	Kerkrade
Bierum	1.00000	0.99491	0.98750	0.97800	0.98493	0.98896	0.95959
Losser	0.99491	1.00000	0.99043	0.98323	0.98816	0.98862	0.96553
Grouw	0.98750	0.99043	1.00000	0.99410	0.99517	0.99046	0.97337
Rosmalen	0.97800	0.98323	0.99410	1.00000	0.99341	0.98419	0.98508
Borssele	0.98493	0.98816	0.99517	0.99341	1.00000	0.99475	0.98089
Poeke	0.98896	0.98862	0.99046	0.98419	0.99475	1.00000	0.97226
Kerkrade	0.95959	0.96553	0.97337	0.98508	0.98089	0.97226	1.00000

Vervolgens bepaalden we onderlinge correlatie met behulp van onze kloon. Dit gaf de volgende matrix:

	Bierum	Losser	Grouw	Rosmalen	Borssele	Poeke	Kerkrade
Bierum	1.00000	0.99419	0.98450	0.97405	0.98327	0.98801	0.95287
Losser	0.99419	1.00000	0.99114	0.98729	0.99132	0.99078	0.96821
Grouw	0.98450	0.99114	1.00000	0.99509	0.99634	0.99018	0.97264
Rosmalen	0.97405	0.98729	0.99509	1.00000	0.99586	0.98463	0.98385
Borssele	0.98327	0.99132	0.99634	0.99586	1.00000	0.99449	0.98204
Poeke	0.98801	0.99078	0.99018	0.98463	0.99449	1.00000	0.97060
Kerkrade	0.95287	0.96821	0.97264	0.98385	0.98204	0.97060	1.00000

We berekenden ook de correlatie tussen de matrix van H & H en de matrix die gegenereerd werd door onze kloon. Daarbij gebruikten we de halve matrices en lieten bovendien de middendiagonalen buiten beschouwing. De correlatie bleek gelijk te zijn aan 0.97833 met  $p=0.000$  (tweezijdig). Er is dus sprake van een significante correlatie (met  $\alpha=0.05$ ).

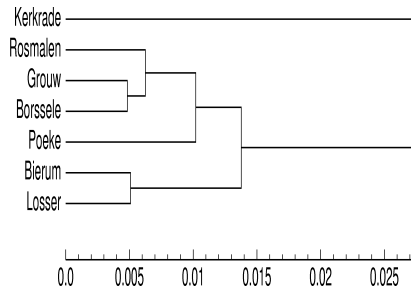
We toetsten ook of de correlatiecoëfficiënten van de matrix van H & H significant verschillend zijn ten opzichte van de correlatiecoëfficiënten van de matrix die gegenereerd werden door onze kloon. We gebruikten de t-test voor gekoppelde paren, eveneens op basis halve matrices zonder middendiagonalen. Hiermee kregen we  $t(20)=0.189$  en  $p=0.852$  (tweezijdig). We vonden dus geen bewijs voor de hypothese dat de reeksen significant verschillend zijn (met  $\alpha=0.05$ ). Ook vonden we geen bewijs voor de hypothese dat de correlatiecoëfficiënten van H & H significant hoger zijn dan die van onze kloon ( $t(20)=0.189$  en  $p=0.426$  (rechtseenzijdig)) of lager ( $t(20)=-0.189$  en  $p=0.426$  (linkseenzijdig)).

## 5.2 Classificatie

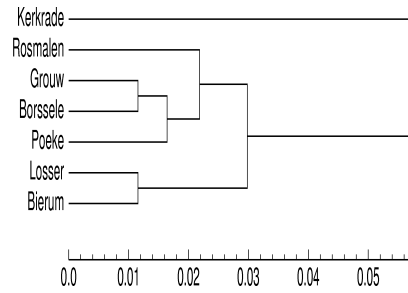
Op basis van de matrix passen H & H cluster-analyse toe. Uit sectie 1.2 van hun boek krijgen we de indruk dat H & H gebruik hebben gemaakt van SPSS (een gangbaar statistisch computerprogramma). Daarbij kozen zij voor ‘Average Linkage (Between Groups)’ (zie blz. 12 en 56). Nu is het zo dat SPSS de getallen in de matrix niet *direct* als afstanden interpreteert, maar deze zelf berekent op basis van de vectoren (althans versie 8.0). Daarbij kan de gebruiker opgeven of de rijen dan wel de kolommen als vectoren beschouwd moeten worden. Standaard berekent SPSS het verschil tussen twee vectoren als de Euclidische afstand (de wortel uit de som van de gekwadrateerde verschillen). Op deze manier maakt het niet uit of een afstandsmatrix dan wel een gelijkheidsmatrix gebruikt wordt.

Echter bij gebruik van vectoren als basis voor cluster-analyse wordt de afstand tussen de dialecten van elk te vormen paar bepaald op basis van *alle* afstanden tot andere plaatsen, en niet alleen op basis van de werkelijke afstand tot elkaar. Wij menen daarom dat het beter is de gegevens op zo direct mogelijke wijze te gebruiken met een minimum aan tussentijdse manipulaties. We hebben daarom de correlatiecoëfficiënten  $r$  omgezet naar afstanden door berekening van  $1-r$ . Vervolgens hebben we cluster-analyse toegepast waarbij de afstanden *direct* gebruikt worden.

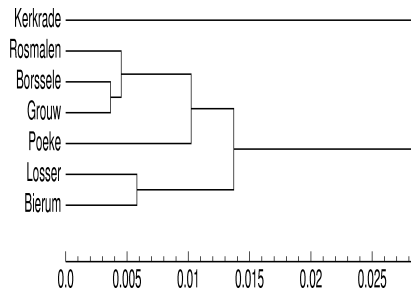
De Figuren 2a en 2b tonen dendrogrammen als resultaat van toepassing van cluster-analyse op basis van de matrix met de gegevens van H & H. In Figuur 2a zijn de afstanden direct gebruikt, en in Figuur 2b zijn de afstanden net als in SPSS herberekend. De Figuren 3a en 3b tonen dendrogrammen als resultaat van toepassing van cluster-analyse op basis van de matrix met gegevens die door onze kloon werden gegenereerd. In Figuur 3a zijn de afstanden direct gebruikt, en in Figuur 3b zijn de afstanden net als in SPSS herberekend.



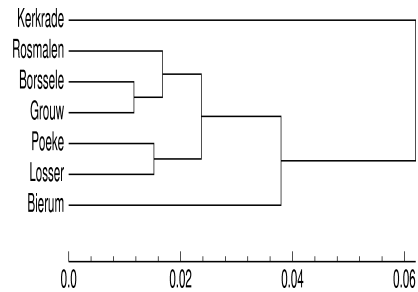
*Figuur 2a*



*Figuur 2b*



*Figuur 3a*



*Figuur 3b*

In alle gevallen hebben we, evenals H & H, ‘average linkage (between groups)’, ook bekend als ‘unweighted average’, gebruikt. Uit de figuren blijkt dat bij direct gebruik van de afstanden de matrix die door onze kloon werd gegenereerd dezelfde boomstructuur geeft als de matrix van H & H. Er is alleen verschil in de lengte van de takken (dus verschil in afstand tussen dialecten of dialectgroepen). Verder blijkt uit de figuren dat het gebruik van directe afstanden tot iets andere resultaten leidt dan het gebruik van herbere-

kende afstanden, met name als we letten op de positie van het dialect van Poeke.

### 5.3 Rangorden

In de hoofdstukken 5 t/m 12 wordt meerdere keren een rangorde gegeven van 155 variëteiten ten opzichte van een uitgangsvariëteit. In zeven gevallen is de rangorde ook in de vorm van een kaart gegeven. De corresponderende zeven uitgangsvariëteiten zijn de variëteiten die we in ons experiment gebruikten. Binnen een rangorde onderscheiden H & H *ringen*. De uitgangsvariëteit wordt in de lijst of op de kaart met een (0) gemarkeerd. In de eerste ring bevinden zich de dialecten met een correlatie van 0.99999 t/m 0.99850 ten opzichte van de uitgangsvariëteit. Deze zijn met een (1) gemarkeerd. De intervalgrootte van deze ring is dus gelijk aan 0.00150. In de tweede ring bevinden zich de dialecten met een correlatie van 0.99849 t/m 0.99750 ten opzichte van de uitgangsvariëteit. Deze zijn met een (2) gemarkeerd. De intervalgrootte van deze ring is dus gelijk aan 0.00100. H & H spreken hier van "het tweede duizental". Dit is echter niet juist omdat de intervalgrootte van deze tweede ring kleiner is dan van de eerste, en bovendien is geen sprake van een 'duizental', maar van een 'duizendstental'. Behalve de tweede ring hebben ook alle volgende ringen een intervalgrootte van 0.00100.

De keuze voor de grootte van de ringen lijkt vrij willekeurig. De ringen worden gebruikt op de Kaarten 3 t/m 10. Zouden op deze kaarten de rangordenummers 1 t/m 156 gebruikt worden in plaats van de ringnummers, dan zou het beeld aanzienlijk minder overzichtelijk zijn. De ringnummers worden ook veel gebruikt in de tekst om te verwijzen naar de positie van (een groep) dialecten in de rangordetabel.

In hoofdstuk 5 (de Nedersaksische dialecten) geven H & H in Lijst 3 en op Kaart 3 de rangorde ten opzichte van Bierum. In de tabel hieronder laten we de rangorde voor onze zeven dialecten zien. De linker kolom geeft de rangorde volgens Lijst 3 van H & H, terwijl in de rechter kolom de rangorde gegeven is volgens de resultaten van onze kloon. We zien dat kleine verschillen in de correlatiecoëfficiënten niet geleid hebben tot verschillende rangorden. Wel leiden verschillen in duizendstentallen tot andere ringen.

1. Bierum	1.00000	1.	Bierum	1.00000
2. Losser	0.99491	2.	Losser	0.99419
3. Poeke	0.98896	3.	Poeke	0.98801
4. Grouw	0.98750	4.	Grouw	0.98450
5. Borssele	0.98493	5.	Borssele	0.98327
6. Rosmalen	0.97800	6.	Rosmalen	0.97405
7. Kerkrade	0.95959	7.	Kerkrade	0.95287

In Lijst 4 en op Kaart 4 geven H & H de rangorde ten opzichte van Losser. In de tabel hieronder vermeldt de linker kolom de ordening volgens H & H en de rechter kolom de ordening volgens onze kloon. We zien dat in de rangorde volgens onze kloon de dialecten van Poeke en Borssele van plaats zijn verwisseld. Het verschil in correlatiecoëfficiënten is 0.00046, kleiner dan de intervalgrootte van één ring (0.00100).

1. Losser	1.00000	1.	Losser	1.00000
2. Bierum	0.99491	2.	Bierum	0.99419
3. Grouw	0.99043	3.	Grouw	0.99114
4. Poeke	0.98862	4.	Borssele	0.99132
5. Borssele	0.98816	5.	Poeke	0.99078
6. Rosmalen	0.98323	6.	Rosmalen	0.98729
7. Kerkrade	0.96553	7.	Kerkrade	0.96821

In hoofdstuk 6 (de taalsituatie in Friesland) wordt door H & H in Lijst 5 en op Kaart 5 de rangorde ten opzichte van Grouw gegeven. In de tabel hieronder geven we deze rangorde voor onze zeven dialecten in de linker kolom, terwijl in de rechter kolom de rangorde gegeven is volgens de resultaten van onze kloon. Opnieuw zijn in de ordening volgens de kloon twee dialecten van plaats verwisseld, namelijk Poeke en Losser. In de rangordelijst van 156 variëteiten behoren beide dialecten tot dezelfde ring en staan ze achter elkaar met als rangordenummers respectievelijk 124 en 125,

1. Grouw	1.00000	1.	Grouw	1.00000
2. Borssele	0.99517	2.	Borssele	0.99634
3. Rosmalen	0.99410	3.	Rosmalen	0.99509
4. Poeke	0.99046	4.	Losser	0.99114
5. Losser	0.99043	5.	Poeke	0.99018
6. Bierum	0.98750	6.	Bierum	0.98450
7. Kerkrade	0.97337	7.	Kerkrade	0.97264

In hoofdstuk 7 (Hollands, Utrechts en Veluws) wordt in Lijst 6 en op Kaart 6 de rangorde ten opzichte van het Algemeen Nederlands gegeven. Het is mooi om te zien dat het dialect van Haarlem het dichtste bij het Algemeen Nederlands staat, geheel in overeenstemming met de gangbare opvatting (blz. 126). Omdat we ons beperkten tot de variëteiten waarvoor we in de RND een transcriptie konden vinden, kon onze kloon geen rangorde ten opzichte van het algemeen Nederlands genereren. We geven daarom geen verdere bespreking van deze rangorde.

In hoofdstuk 8 (Noord-Brabants) van het boek van H & H is in Lijst 7 en op kaart 7 de rangorde ten opzichte van het dialect van Rosmalen afgedrukt. De tabel hieronder toont in de linker kolom de rangorde voor onze zeven dialecten volgens Lijst 7 van H & H, terwijl in de rechter kolom de rangorde getoond wordt volgens de resultaten van onze kloon. De verschillen tussen H & H en onze kloon zijn hier beduidend groter dan in de voorgaande rangordelijsten. Grouw en Borssele zijn van plaats verwisseld. Het verschil in correlatiecoëfficiënten volgens H & H (0.00069) ligt ruim binnen de grootte van één ring (0.00100). Verder zijn Losser en Kerkrade van plaats verwisseld. Het verschil in correlatiecoëfficiënten volgens H & H (0.00186) ligt nu niet binnen de grootte van één ring. Wel is in beide gevallen het dialect van Bierum het verst verwijderde dialect.

1. Rosmalen	1.00000	1.	Rosmalen	1.00000
2. Grouw	0.99410	2.	Borssele	0.99586
3. Borssele	0.99341	3.	Grouw	0.99509
4. Kerkrade	0.98508	4.	Losser	0.98729
5. Poeke	0.98419	5.	Poeke	0.98463
6. Losser	0.98323	6.	Kerkrade	0.98385
7. Bierum	0.97800	7.	Bierum	0.97405

In hoofdstuk 9 (Zeeuws) wordt in Lijst 8 en op kaart 8 de rangorde ten opzichte van het dialect van Borssele vermeld. In de tabel hieronder geven we deze rangorde voor onze zeven dialecten in de linker kolom, terwijl in de rechter kolom de rangorde gegeven is volgens de resultaten van onze kloon. We zien dat de dialecten van Rosmalen en Poeke verwisseld zijn in de ordening volgens de kloon. Het verschil in correlatiecoëfficiënten (0.00134) valt niet binnen de grootte van één ring (0.00100).

1. Borssele	1.00000	1.	Borssele	1.00000
2. Grouw	0.99517	2.	Grouw	0.99634
3. Poeke	0.99475	3.	Rosmalen	0.99586
4. Rosmalen	0.99341	4.	Poeke	0.99449
5. Losser	0.98816	5.	Losser	0.99132
6. Bierum	0.98493	6.	Bierum	0.98327
7. Kerkrade	0.98089	7.	Kerkrade	0.98204

In hoofdstuk 10 (Belgisch-Brabants) van het boek van H & H wordt in Lijst 8b de rangorde ten opzichte van het dialect van Aarschot gegeven. Deze rangorde wordt niet gevisualiseerd in de vorm van een kaart. We hebben dit dialect niet betrokken in ons experiment en geven daarom geen verdere bespreking van de rangorde.

In hoofdstuk 11 (Vlaams) wordt in Lijst 9 en op kaart 9 de rangorde ten opzichte van het dialect van Poeke gegeven. In de tabel hieronder vermeldt de linker kolom de ordening volgens H & H en de rechter kolom de ordening volgens onze kloon. In de rangorde volgens de kloon valt de verschuiving van het dialect van Losser op.

1. Poeke	1.00000	1.	Poeke	1.00000
2. Borssele	0.99475	2.	Borssele	0.99449
3. Grouw	0.99046	3.	Losser	0.99078
4. Bierum	0.98896	4.	Grouw	0.99018
5. Losser	0.98862	5.	Bierum	0.98801
6. Rosmalen	0.98419	6.	Rosmalen	0.98463
7. Kerkrade	0.97226	7.	Kerkrade	0.97060

In hoofdstuk 12 (Limburgs) wordt door H & H in Lijst 10 en op Kaart 10 de rangorde ten opzichte van het dialect van Kerkrade gegeven. In de tabel hieronder vermeldt de linker kolom de ordening volgens H & H en de rechter kolom de ordening volgens onze kloon. We zien dat kleine verschillen in de correlatiecoëfficiënten niet geleid hebben tot verschillende rangorden.

1. Kerkrade	1.00000	1.	Kerkrade	1.00000
2. Rosmalen	0.98508	2.	Rosmalen	0.98385
3. Borssele	0.98089	3.	Borssele	0.98204
4. Grouw	0.97337	4.	Grouw	0.97264
5. Poeke	0.97226	5.	Poeke	0.97060

6. Losser	0.96553	6.	Losser	0.96821
7. Bierum	0.95959	7.	Bierum	0.95285

We zien dus dat de kloon in het algemeen de resultaten voor H & H aardig reproduceert. Verschillen vallen voor een deel binnen een ring. Daar waar ze niet binnen een ring vallen, valt moeilijk te zeggen welke van de twee rangorden de betere is. Om meer zicht te krijgen op de oorzaak van de verschillen kijken we in sectie 5.4 naar de afzonderlijke feature-percentages.

#### 5.4 Kenmerken

In de hoofdstukken 5 t/m 12 geven H & H meerdere keren een lijst waarbij voor één bepaald feature de percentages vermeld worden voor alle 156 dialecten. Daarbij worden de dialecten zodanig geordend dat het dialect met het hoogste percentage bovenaan, en die met het laagste percentage onderaan de lijst is afgedrukt. Het bestuderen van featurewaarden maakt duidelijk wat kenmerkend is voor een groep dialecten. Juist de features verklaren waarom de FFM sommige dialecten als verwant, en andere juist als heel verschillend herkent. Zo laten H & H op de bladzijden 73 t/m 75 zien dat Nedersaksische dialecten zich kenmerken door een hoger percentage syllabische klanken (bijvoorbeeld *spinnen* wordt in Nedersaksische dialecten gewoonlijk uitgesproken als [spɪŋ]), waarbij binnen de Nedersaksische dialecten de Groningse dialecten het hoogste percentage syllabische klanken hebben.

In de verschillende hoofdstukken bespreken H & H uiteindelijk 14 features, namelijk: vocalisch, voor, laag, polair, lang, nasaal, anterieur, posterieur, laryngaal, sonorant, stemhebbend, hoog, continuant, syllabisch. Eigenlijk is het jammer dat niet ook voor de resterende zeven features rangordes gegeven zijn, eventueel in een aparte bijlage.

Uit de rangordelijsten van de verschillende features kunnen we voor de zeven variëteiten de volgende matrix afleiden:

	Bie- rum	Los- ser	Grouw	Ros- malen	Bors- sele	Poe- ke	Kerk- rade
vocalisch	3.8	5.6	12.1	13.7	12.0	11.4	14.0
voor	5.8	19.1	15.2	15.4	17.8	16.7	16.0
laag	3.9	14.8	12.3	12.6	13.6	13.4	13.3
polair	10.3	8.9	10.3	8.6	13.5	13.7	11.1

lang	11.6	10.2	6.8	5.1	8.4	14.8	9.2
nasaal	14.6	14.5	13.2	11.4	12.2	13.8	9.6
anterieur	51.5	50.5	48.0	46.2	47.9	51.2	37.6
posterieur	8.9	9.5	7.6	9.2	8.7	9.1	14.1
laryngaal	1.5	2.0	2.6	2.2	0.0	0.0	2.4
sonorant	26.8	24.8	24.4	23.8	24.4	26.4	23.6
stemhebbend	42.9	37.6	37.5	39.5	39.3	42.2	37.6
hoog	4.1	3.0	5.2	4.2	4.8	2.7	7.0
continuant	25.1	24.6	22.8	25.5	25.7	26.3	27.5
syllabisch	5.9	4.1	1.9	0.0	0.2	0.9	0.0

Met behulp van onze kloon hebben we daarnaast de volgende matrix verkregen:

	Bie- rum	Los- ser	Grouw	Ros- malen	Bors- sele	Poe- ke	Kerk- rade
vocalisch	4.0	6.9	12.7	14.2	12.8	12.3	15.0
voor	16.8	18.2	15.6	15.3	17.3	16.8	15.8
laag	14.5	13.7	11.7	12.3	12.8	12.9	12.4
polair	10.5	7.2	10.5	8.7	12.2	12.9	10.7
lang	12.2	10.5	7.0	5.1	8.7	15.2	8.9
nasaal	16.0	16.0	14.1	12.6	13.4	14.9	10.0
anterieur	52.8	52.1	47.5	47.3	48.6	51.7	37.5
posterieur	9.4	9.5	7.1	8.9	9.0	8.9	13.9
laryngaal	1.3	1.9	2.5	1.9	0.0	0.0	2.2
sonorant	27.4	26.7	24.9	24.3	25.1	27.5	24.0
stemhebbend	41.2	38.8	36.6	38.6	38.9	41.4	36.2
hoog	3.0	3.0	4.8	3.7	4.0	2.3	6.8
continuant	25.9	26.3	23.1	25.7	26.5	27.0	27.9
syllabisch	5.9	2.5	1.9	0.0	1.0	2.5	0.0

We hebben vervolgens een verschilmatrix bepaald, waarbij we de waarden van H & H aftrokken van de waarden volgens onze kloon. Deze matrix ziet er als volgt uit:

	Bie- rum	Los- ser	Grouw	Ros- malen	Bors- sele	Poe- ke	Kerk- rade	Totaal
vocalisch	0.2	1.3	0.6	0.5	0.8	0.9	1.0	5.3
voor	1.0	-0.9	0.4	-0.1	-0.5	0.1	-0.2	3.2
laag	0.6	-1.1	-0.6	-0.3	-0.8	-0.5	-0.9	4.8
polair	0.2	-1.7	0.2	0.1	-1.3	-0.8	-0.4	4.7
lang	0.6	0.3	0.2	0.0	0.3	0.4	-0.3	2.1
nasaal	1.4	1.5	0.9	1.2	1.2	1.1	0.4	7.7
anterieur	1.3	1.6	-0.5	1.1	0.7	0.5	-0.1	5.8
posterieur	0.5	0.0	-0.5	-0.3	0.3	-0.2	-0.2	2.0

laryngaal	-0.2	-0.1	-0.1	-0.3	0.0	0.0	-0.2	0.9
sonorant	0.6	1.9	0.5	0.5	0.7	1.1	0.4	5.7
stemhebbend	-1.7	1.2	-0.9	-0.9	-0.4	-0.8	-1.4	7.3
hoog	-1.1	0.0	-0.4	-0.5	-0.8	-0.4	-0.2	3.4
continuuant	0.8	1.7	0.3	0.2	0.8	0.7	0.4	4.9
syllabisch	0.0	-1.6	0.0	0.0	0.8	1.6	0.0	4.0
-----								
<i>totaal</i>	<i>10.2</i>	<i>14.9</i>	<i>6.1</i>	<i>6.0</i>	<i>9.4</i>	<i>9.1</i>	<i>6.1</i>	<i>61.8</i>

We zien dat de verschillen over het algemeen klein zijn, per feature per dialect altijd beneden de 2.0%. In de tabel zijn de rij- en kolomtotalen cursief weergegeven. Deze zijn bepaald door sommatie over de absolute verschillen. Voor wat betreft de features zien we de grootste verschillen bij de features ‘nasaal’ en ‘stemhebbend’. Zoals hierboven gemeld in sectie 3.4 passen H & H de transcripties ten aanzien van deze features aan, terwijl onze kloon werkt op basis van teksten waarin de door H & H voorgestelde aanpassingen niet zijn doorgevoerd. Voor wat betreft de dialecten vinden we het grootste verschil bij Losser, gevolgd door Bierum. Hiermee is ook verklaard waarom Losser in de rangordelijsten volgens de kloon een aantal keren een andere positie heeft dan in de rangordelijsten volgens H & H (zie hierboven sectie 5.3). Ten slotte kunnen andere (grotere) verschillen verklaard worden uit het feit dat H & H de complete teksten gebruikten, terwijl wij ons beperkten tot alleen de ‘echte’ zinnen.

## 6. Conclusies

De FFM werd door H & H ontwikkeld om te komen tot een objectieve dialectclassificatie. Per dialect worden frequenties van kenmerken bepaald. Een dialect wordt op die manier gekarakteriseerd door een reeks kenmerkfrequenties. Een histogram dat deze frequenties weergeeft vormt als het ware een plaatje van het dialect. De mate van overeenkomst tussen twee dialecten wordt berekend als de correlatie tussen de respectievelijke reeksen van kenmerkfrequenties. Wanneer tussen alle dialecten onderling op die manier de correlatie-coëfficiënten bekend zijn, kan men op basis daarvan door middel van cluster-analyse komen tot een indeling in groepen.

H & H pasten hun methode toe op teksten uit de RND. Belangrijkste nadelen van het RND-materiaal zijn verschil in tijd van opname, het feit dat de vragenlijsten die corresponderen met de verschillende delen niet altijd helemaal gelijk zijn, en dat verschillende transcribenten verschillende notaties

gebruiken. Het eerste nadeel leidt vooral tot vertekening tussen geografisch ver uit elkaar gelegen dialecten, het verschil in opnametijd tussen dialecten uit delen van aangrenzende gebieden is nooit groter dan 30 jaar. Het tweede nadeel zal bij de FFM niet tot (grote) vertekening leiden, zolang de transcripties per dialect maar voldoende omvang hebben (zie H & H sectie 2.1.3, blz. 23). Het derde nadeel weegt minder zwaar zolang het net van plaatsen niet te dicht is (zie H & H sectie 2.2, blz. 24). Dialectverschillen zullen dan waarschijnlijk een grotere invloed hebben op het eindresultaat dan notatieverschillen (zie H & H sectie 2.2, blz. 24).

In het kader van dit artikel gebruikten wij een kloon en pasten die toe op een kleine selectie van zeven variëteiten. Bij de digitalisatie van de transcripties van deze variëteiten digitaliseerden we alleen de complete zinnen.

De hoofdingeling zoals die door H & H met de FFM gegenereerd werd, heeft als drie belangrijkste groepen het Limburgs, het Nedersaksisch en de rest. Die rest heeft als drie subgroepen het Fries, alle overige Nederlandse dialecten, en het Noord-Belgisch. De term ‘Frankisch’ kan in deze context niet anders gezien worden dan als een verzamelnaam voor alle dialecten die noch Nedersaksisch, noch Fries zijn.

Met behulp van onze kloon berekenden we de correlatiecoëfficiënten tussen de zeven dialecten en vergeleken die met de correlatiecoëfficiënten die door H & H gegeven zijn in verschillende lijsten in hun boek. De correlatiecoëfficiënten werden door ons omgezet naar afstanden door iedere coëfficiënt te vervangen door  $1 - r$ . Op basis van deze afstanden konden we rechtstreeks cluster-analyse toepassen. In beide gevallen kregen we een classificatie met dezelfde structuur.

In het boek van H & H worden rangorden gegeven ten opzichte van een uitgangsdialect. We hebben hetzelfde gedaan, waarbij we elk van de zeven dialecten als uitgangspunt namen. Twee lijsten hadden dezelfde ordening als bij H & H, bij de overige lijsten waren meer of minder grote verschillen. Lijsten waren nooit compleet verschillend. Bijzonder interessant is de rangordelijst met het Algemeen Nederlands als uitgangspunt. Heel mooi laten H & H met de FFM zien dat van alle dialecten het dialect van Haarlem het meest op het Algemeen Nederlands lijkt, precies volgens het stereotype.

H & H geven ook rangordelijsten op basis van de percentages van een feature. In totaal worden voor 14 features rangordelijsten gegeven, waarbij het dialect met het grootste percentage het eerste element, en dat met het

laagste percentage het laatste element is. We hebben de percentages volgens onze kloon vergeleken met de percentages volgens H & H op basis van de zeven dialecten. Verschillen waren nooit groter dan 2%. De grootste totale afwijking betrof het feature 'nasaal', gevolgd door het feature 'stemhebbend'. Dit is te verklaren uit het feit dat H & H ten aanzien van deze features de transcriptie aanpassen. Voor onze kloon hebben we dat niet gedaan. Van de dialecten was de totale afwijking over de 14 features het grootst voor Lossler. In deze rangordelijsten per feature kunnen we soms heel mooi datgene terugvinden dat we als karakteristiek ervaren voor een dialect. Het feit bijvoorbeeld dat in het Gronings veel 'geknauwd' wordt vonden we terug in het hoge percentage syllabische klanken.

Het feit dat onze transcripties ten aanzien van nasaliteit en stemhebbendheid niet zijn aangepast, en het feit dat in onze transcripties losse begrippen niet zijn opgenomen, verklaren de verschillen tussen de resultaten van H & H en de resultaten van onze kloon. We kunnen concluderen dat H & H een heel complete beschrijving van hun methodiek geven, voldoende compleet om een kloon te bouwen die (bijna) dezelfde resultaten genereert.

Ten slotte kunnen we concluderen dat de FFM een bruikbaar instrument is waarmee taalverhoudingen op objectieve wijze bepaald en stereotypes betreffende taalverwantschap getoetst kunnen worden. De FFM laat ook de waarde zien van featuresystemen, wat van groot belang is voor de fonologie. Het boek van H & H is een waardevol boek dat een schat aan informatie bevat uit de traditionele dialectologie, en een complete beschrijving van een belangrijke eerste stap om te komen tot een betrouwbare computationele vergelijking en classificatie van dialecten in het Nederlandse taalgebied. We willen benadrukken dat H & H de eersten waren die een classificatie van het complete Nederlandse taalgebied genereerden met behulp van een computationele methode.

Op deze eerste stap zijn sinds de eerste introductie van de FFM in Hoppenbrouwers & Hoppenbrouwers (1988) verdere stappen gevolgd. In Nerbonne en Heeringa (1998) wordt een aanpak gepresenteerd waarbij gebruik gemaakt wordt van de Levenshtein-afstand, een maat die de afstand tussen twee woorden bepaalt. Bij deze aanpak wordt meer informatie uit hetzelfde materiaal gehaald omdat woorden als taalkundige eenheden worden onderscheiden, en bovendien de volgorde van klanken in een woord in rekening wordt gebracht. Het resultaat van deze aanpak is een indeling waarbij Fries,

Saksisch en Frankisch de drie belangrijkste groepen vormen, precies zoals in de traditionele dialectologie beschreven is. Anders dan bij de resultaten van H & H vormt het Frankisch hier daadwerkelijk een dialectologische groep, waarbij ook het Limburgs keurig onder het Frankisch valt. Deze latere aanpak, die ons een stap verder brengt, wordt vreemd genoeg in het boek van H & H niet genoemd, hoewel ze hem zouden hebben moeten kennen.

#### *Verwijzingen*

- Blancquaert, E. & W. Pée, *Reeks Nederlandse Dialectatlassen (1925-1982)*, De Sikkel, Antwerpen.
- Chomsky, N. & M. Halle (1968), *The Sound Pattern of English*, Harper & Row, New York.
- Heeringa, W. (2001), 'De selectie en digitalisatie van dialecten en woorden uit de Reeks Nederlandse Dialectatlassen'. In: *TABU, Bulletin voor Taalwetenschap*, Rijksuniversiteit Groningen, Faculteit der Letteren, Groningen, 30, (1/2), pp. 61-103, 2001.
- Hoppenbrouwers, C. & G. (1988), 'De Feature frequentie methode en de classificatie van Nederlandse dialecten'. In: *TABU, Bulletin voort taalwetenschap*, Rijksuniversiteit Groningen, Faculteit der Letteren, Groningen, 18, pp. 51-92.
- Hoppenbrouwers, C. (1990), *Het regiolect: van dialect tot algemeen Nederlands*, Coutinho, Muiderberg.
- Jain, A. K. & R. C. Dubes (1988), *Algorithms for clustering data*, Prentice Hall, Englewood Cliffs, N. J..
- Jacobson, R. O., C. G. M. Fant & M. Halle (1952), *Preliminaries to speech analysis, the distinctive features and their correlates*, M. I. T. Press, Cambridge Mass.
- Ladefoged, P. (1975), *A Course in Phonetics*, Harcourt Brace Jovanovich, New York.
- Loey, A. van (1970), *Schönfelds Historische Grammatica van het Nederlands*, 8e druk, Thieme & Cie., Zutphen.
- Nerbonne, J. & W. Heeringa (1998), 'Computationale vergelijking en classificatie van dialecten'. In: *Taal en Tongval* 50(2), pp. 164-193.
- The Principles of the Phonetic Association* (1949), International Phonetic Association, London.
- Vries, J. W. de, R. E. C. Willemys & P. Burger (1993), *Het verhaal van een taal: negen eeuwen Nederlands*, Prometheus, Amsterdam.