



KATHOLIEKE UNIVERSITEIT LEUVEN  
FACULTEIT LETTEREN  
SUBFACULTEIT TAALKUNDE

# Tussen spreek- en standaardtaal

Een corpusgebaseerd onderzoek naar de situationele, regionale en sociale verspreiding van enkele morfo-syntactische verschijnselen uit het gesproken Belgisch-Nederlands

Koen Plevoets

Promotoren:  
Prof. dr. Dirk Geeraerts  
Prof. dr. Dirk Spielman

Proefschrift ingediend tot het  
behalen van de graad van  
Doctor in de Taal- en  
Letterkunde: Germaanse Talen

Leuven 2008



# Dankwoord

Het is een afgesloten cliché dat een doctoraat nooit zonder hulp van anderen tot stand komt. Het is alleen maar een afgesloten cliché omdat het ook zo onomstotelijk waar is. Daarom had ik graag enkele mensen bedankt.

In de eerste plaats denk ik hierbij aan mijn beide promotoren, professor Dirk Geeraerts en professor Dirk Speelman. Professor Dirk Geeraerts dank ik omdat hij de voorbije vier jaren het meest leerrijke klankbord voor mijn ideeën is geweest dat ik me maar kon wensen: meermaals heeft hij mij ervoor behoed om redeneringsgewijs ietwat te kort door de bocht te gaan; op andere tijdstippen heeft hij me net aangespoord om de door mij ingeslagen weg vooral verder te bewandelen. Dat ik professor Dirk Speelman bedank voor de meer praktische kant van het doctoraat zal wellicht weinig verbazing wekken; het verdient evenwel enige nadruk dat één en ander soms in de spreekwoordelijke details kan zitten: “av.py” zijn maar vier letters met een punt ertussen, maar zonder dat Pythonscript was dit proefschrift er eenvoudigweg nooit gekomen.

In het verlengde hiervan wil ik graag mijn dank betuigen aan de ‘Nijmeegse partners’ van dit doctoraatsproject (dossiernummer: G.0490.04): professor Harald Baayen, professor Roeland van Hout, en Karen Keune. Hun momenten van feedback betekenden steevast een bron van verlichting.

Hulp kwam af en toe ook uit niet-linguïstische hoek. Ik dank professor Koen Matthijs en professor Koen Van Eijck van de faculteit Sociale Wetenschappen ervoor dat ze me het werk van Heck Becker over de verschillende generaties van de twintigste eeuw onder de aandacht gebracht hebben. De mensen van het KBC-Infodoc dank ik voor het welwillend ter beschikking stellen van hun archief. Mijn oude statistiekpromotor, professor Anne-Marie De Meyer, dank ik voor haar bibliografische aanwijzingen omtrent Correspondentie-analyse.

Natuurlijk kan ik in dit dankwoord niet heen om wat vier jaar lang mijn academische ‘thuishaven’ is geweest: QLVL. Stef, Gert, Kris, José, Stefania, Ann, Sofie, Caroline, Dylan, Dagmar, Yves, Leen, Eline, Natalia en Frieda, dank je voor het creëren van een aangename en productieve werksfeer, waarin zich meer dan eens een boeiende discussie over tussen-, spreek- en/of standaardtaal ontspoon. Door jullie heb ik mijn ideeën alleen maar aangescherpt. Sofie, Dylan en Leen verdienen misschien wel een speciale vermelding: het was een genoegen om met jullie het kantoor te delen.

Toch kon de boog niet altijd gespannen staan, en gelukkig was er dan steeds een vangnet om op/in terug te vallen. Raf en Kristel, Yves en Kathleen, bedankt voor de concerten nu en dan: ze waren altijd een welkome afwisseling in soms hectische tijden. In het bijzonder wil ik An Bruyninckx nog danken voor het dak boven mijn hoofd dat ze me gegeven heeft tijdens die eerste weken van mijn doctoraat in januari 2004.

Nele en Toon zijn in de loop van dit project onmisbaar gebleken: niet alleen hebben ze beiden wekelijks tijd voor me vrij proberen te maken; bovendien heeft Nele in de allerlaatste fase de hele lay-out van dit proefschrift verzorgd, zodat ze zich met recht de 'co-auteur' ervan mag noemen. Bedankt, allebei!

Daarmee ben ik ten slotte aanbeland bij de twee mensen die wel het meeste voor me betekend hebben en dat nog steeds doen: pa en ma, ik kan geen zin bedenken die onder woorden brengt wat ik jullie verschuldigd ben, vandaar dat ik maar zal houden op een simpel maar oprecht: bedankt... voor alles!

# Inhoudsopgave

<b>Hoofdstuk 1   Inleiding</b> .....	7
1.1 Nogmaals de taaltoestand in Vlaanderen – <i>status quaestionis</i> .....	7
1.2 Naar een beschrijving van het Soapvlaams – <i>hypothesen</i> .....	17
<b>Hoofdstuk 2   Corpus Gesproken Nederlands</b> .....	19
2.1 Het design van het CGN .....	19
2.2 De sprekersinformatie in het CGN .....	21
2.2.1 <i>sex</i> .....	21
2.2.2 <i>birthYear</i> .....	21
2.2.3 <i>birthRegion</i> .....	21
2.2.4 <i>eduLevel</i> .....	22
2.2.5 <i>occLevel</i> .....	22
2.3 De mark-up van het CGN .....	25
2.4 Het dialect in het CGN .....	26
<b>Hoofdstuk 3   Taalvariabelen</b> .....	29
3.1 De verkenning van het CGN .....	29
3.2 De operationalisering van de taalvariabelen .....	33
3.2.1 <i>De adnominale taalvariabelen</i> .....	34
3.2.2 <i>De diminutivische taalvariabele</i> .....	52
3.2.3 <i>De pronominale taalvariabelen</i> .....	52
<b>Hoofdstuk 4   Methodologie</b> .....	65
4.1 De bivariate techniek .....	66
4.2 De multivariate techniek .....	70
<b>Hoofdstuk 5   Bivariate analyses</b> .....	73
5.1 Register .....	73
5.2 Regio .....	86
5.3 Decennium .....	96
5.4 Generatie .....	101

5.5	Sekse .....	106
5.6	Opleidingsniveau.....	110
5.7	Beroepsniveau.....	114
<b>Hoofdstuk 6</b>	<b>  Multivariate analyse</b> .....	<b>123</b>
6.1	Varianten .....	124
6.2	Variëteiten – hoofdeffecten.....	129
6.3	Variëteiten – interacties .....	131
6.3.1	<i>Register &amp; Regio</i> .....	131
6.3.2	<i>Register &amp; Generatie</i> .....	135
6.3.3	<i>Register &amp; Sekse</i> .....	139
6.3.4	<i>Register &amp; Educatieniveau</i> .....	141
6.3.5	<i>Register &amp; Beroepsniveau</i> .....	143
6.3.6	<i>Regio &amp; Generatie</i> .....	149
6.3.7	<i>Regio &amp; Sekse</i> .....	152
6.3.8	<i>Regio &amp; Educatieniveau</i> .....	154
6.3.9	<i>Regio &amp; Beroepsniveau</i> .....	156
6.3.10	<i>Generatie &amp; Sekse</i> .....	159
6.3.11	<i>Generatie &amp; Educatieniveau</i> .....	161
6.3.12	<i>Generatie &amp; Beroepsniveau</i> .....	163
6.3.13	<i>Sekse &amp; Educatieniveau</i> .....	167
6.3.14	<i>Sekse &amp; Beroepsniveau</i> .....	169
6.3.15	<i>Educatieniveau &amp; Beroepsniveau</i> .....	171
<b>Hoofdstuk 7</b>	<b>  Besluit</b> .....	<b>175</b>
7.1	Samenvatting van de resultaten .....	175
7.2	Beoordeling van de resultaten.....	179
<b>Bibliografie</b>	.....	<b>183</b>

# Hoofdstuk 1 | Inleiding

Op het einde van het vorige millennium is er nogal wat commotie ontstaan rond een nieuwe ‘taal’ die in Vlaanderen steeds vastere grond onder de voeten aan het krijgen zou zijn. Het zou gaan om een halfbakken en minderwaardige taal, een taal die geboren is uit een gebrek aan durf dan wel aan wil om goed Nederlands te praten, een taal die kortom een smet werpt op het aanschijn van het Vlaamse taallandschap. Desondanks gingen meer en meer Vlamingen zich er in hun dagdagelijkse taalgedrag op verlaten, zodat die taal zich een stevig verankerde positie in de maatschappelijke omgang wist te verwerven. Van lieverlee kwam daarbij echter de status van het officiële Standaardnederlands, zoals die van oudsher in formele taalsituaties de norm uitmaakte, in toenemende mate in het gedrang. Eén en ander zou uiteindelijk zelfs de VRT – in Vlaanderen de vertegenwoordiger bij uitstek van een aantrekkelijk, duidelijk en correct Nederlands – ertoe bewegen een campagne op te zetten om de ‘taalverloedering’ tegen te gaan en het gebruik van de standaardtaal te promoten (Hendrickx 2001. *Laat ons ne keer te goei naar onszelf luisteren. Taalsignaal voor 6 miljoen Vlamingen*).

Wat was er nu – en is er nog steeds – aan de hand? In een eerste paragraaf zullen we een overzicht geven van de literatuur die er gedurende de laatste jaren over het nieuwe talige fenomeen verschenen is. Onze belangstelling gaat daarbij vooral uit naar het descriptieve aspect: in dit proefschrift zijn we geïnteresseerd in het feitelijke taalgebruik in Vlaanderen van vandaag de dag, en wat de verscheidene auteurs daarover beweren. In een tweede paragraaf zullen we één en ander dan operationaliseren tot enkele toetsbare hypothesen.

## 1.1 Nogmaals de taaltoestand in Vlaanderen – *status quaestionis*

De discussie werd op gang getrokken door journalist Geert Van Istendael, die in zijn boek *Het Belgisch Labyrint* (eerste druk: 1989) fulmineerde tegen wat hij het ‘Verkavelingsvlaams’ noemt:

“Verkavelings-Vlaams, het is de taal van een nieuwsoortig, door en door vals Vlaams zelfvertrouwen, het is de taal die uit minachting voor de spraak van gewone mensen en uit angst voor Nederlands geboren is, een wangedrocht is

het, die taal van het nieuwe Vlaanderen, dat blaakt van intellectuele luiheid. Het allerergste is dat het waardeloze Verkavelings-Vlaams steeds vaker door gewone mensen wordt overgenomen. Hoorde ik laatst de slager niet zeggen: *ik heb e goe contact naar mijne zoon toe?*” (Van Istendael 2005: 124)

Daarmee is meteen een eerste karakterisering van dat Verkavelingsvlaams gegeven. Het gaat meer bepaald om tal van verschijnselen op vrijwel elk domein van de grammatica (uitspraak, morfologie, lexicon, syntaxis), maar de meest prominente ervan zijn toch wel de flectionele (oftewel, morfosyntactische): de (masculiene) genusflexie (“mijne”), de diminuering op *-ke*, en verder nog een hele reeks van dialectische pronomina (*ekik, degij*) – zie De Vriendt (1989), Haeseryn (1996), en Goossens (2000) voor uitgebreide inventarisaties van deze en soortgelijke kenmerken; een uitvoerige bespreking ervan stellen we voorlopig uit tot hoofdstuk 3.

Toch was het Verkavelingsvlaams geenszins iets nieuws – wat Van Istendael zelf trouwens ook toegeeft, als hij de geciteerde passage laat volgen door de woorden: “Er wordt inderdaad gewag gemaakt, ook door filologen en linguïsten, over een Belgische variant van het Nederlands” (Van Istendael 2005: 124). Vanuit de strekking van zijn betoog lijkt hij daarmee echter niet zozeer te doelen op de autonome standaardtaalvorming zoals die voor het eerst geanalyseerd werd in Jaspaert (1986), als wel op wat Willemyns in navolging van Goossens het “Belgisch Beschaafd” genoemd heeft (zie Willemyns 1987; nadere details omtrent het Vlaamse standaardiseringsproces vindt men overigens in Geeraerts, Grondelaers & Speelman 1999, Van der Wal & van Bree 1992, Van de Velde 1996a en Willemyns & Daniëls 2003). Al in 1970 stelde die laatste namelijk vast dat het toenemende gebruik van het Standaardnederlands in Vlaanderen vooral aanleiding gaf tot de vorming van een endogeen Vlaamse bovenregionale standaard die met name in de informele registers steeds meer tot norm begon te worden (zie Goossens 1970, een studie waarop hij in latere publicaties steeds verder geborduurd heeft en met name in Goossens 1975 tot een alomvattend kader uitgewerkt; vergelijk verder Goossens 1971, 1973, 1978, 1981a, 1981b, 1983a, 1983b, 1987 en 1990 – uiteindelijk zou Goossens evenwel voor de term ‘Schoon Vlaams’ opteren: zie Goossens 2000).

Zoals de titel van Goossens’ oorspronkelijke artikel al aangeeft, nam die autonome en endogene substandaardisering daarbij de vorm aan van een verbreiding van voornamelijk Brabantse elementen, een proces dat sindsdien bekend staat als de “Brabantse expansie”. Die hypothese werd in de jaren tachtig uitvoerig getest door de kring van onderzoekers rondom Geerts, die er een overtuigende mate van evidentie voor verzamelden (zie Geerts 1983, en verder het syntheseartikel Geerts 1985). Meer recentelijk nog heeft Willemyns dan weer het verband gelegd met het proces van dialectverlies dat de Brabantse provincie het sterkst getroffen zou hebben (zie Willemyns 1997, zie ook Hinskens, Hoppenbrouwers & Taeldeman 1993).



In het licht van dat alles besluit Craps in 2002 dan ook:

“Wat is precies Schoon Vlaams / Verkavelingsvlaams? Als dat wordt onderzocht, lijkt de conclusie onvermijdelijk dat het een recent ontdekt (niet ontstaan) gesproken Nederlands is, waarover veel negatiever wordt gedaan dan nodig is, omdat het toevallig in de aandacht is. Er is zeker een continuïteit tussen de taalvormen die Van Coetsem (1957) als typisch Belgisch beschreef, de vormen die Jan Goossens (1966) in zijn artikel “Hij zei dattem van niks en wist” citeert uit de mond van licentiestudenten Germaanse Filologie, de exogene, niet-Nederlandse woorden in het onderzoek van Geerts, Creten, Deprez, Van den Broeck, Knops en anderen uit het midden van de jaren zeventig (bijv. Deprez & Geerts 1978), de vormen met t-deletie “da” en “nie” die de houders van de legitieme competentie gebruikten volgens Jaspaert (1986) en het Verkavelingsvlaams dat plots het onderwerp van onderzoek wordt: een grote stroom gesproken substandaard-Vlaams komt al minstens sinds 1957 op ons af.” (Craps 2002: 82)

Het nieuwe aan het Verkavelingsvlaams/Schoon Vlaams/Belgisch Beschaafd aan het einde van de jaren tachtig en het begin van de jaren negentig is dan ook veeleer het feit dat het steeds meer in situaties gebruikt wordt, waar vroeger alleen Standaardnederlands te horen viel – een proces dat met Haugen (1972) bekend staat als ‘elaboration of function’. Dat wordt retorisch gezien het treffendst verwoord in Taeldeman (1992), waarbij de opkomende substandaardvariëteit eveneens de naam toebedeeld krijgt waarmee er vanaf dan in vrijwel elke wetenschappelijke bespreking erover naar verwezen zal worden – ‘tussentaal’:

“Dat steeds meer Vlamingen in steeds meer omstandigheden tussentaal (...) gebruiken, constitueert m.i. de tweede pijler van Vlaanderens ware taalverdriet [naast taalonzekerheid, K.P.]. Uiteraard gaat het er daarbij niet om dat wij het bestaansrecht van tussentaal in het continuüm tussen dialect en A.N. in minder formele situaties ontkennen, maar wel om het feit dat het in het ‘nieuwe’ Vlaanderen de nieuwe maatstaf van een nieuwe ‘taalvaardigheid’ dreigt te worden.” (Taeldeman 1992: 36-37)

Daar onmiddellijk op volgend geeft Taeldeman dan aan waar volgens hem het probleem ligt – de verspreiding van tussentaal is te wijten aan het feit dat haar sprekers tot de maatschappelijk hogere klassen behoren:

“Dat laatste is niet de schuld van de vele Vlamingen voor wie (door historische omstandigheden) die tussentaal de uiterste grens van hun *kunnen* markeert, maar van de stilaan even talrijke gouvgenoten die (via hun opleiding) wel beter kunnen maar *niet willen*. Het is een bekend fenomeen in de geschiedenis: goedkoop arrivisme/opportunisme op de rug van maatschappelijke kansloosheid. In ons geval betekent dat concreet dat een groot deel van de Vlaamse ‘elite’ zich bewust aan (taal)normverlaging en –vervalsing bezondigt (...). De vervalsing ligt juist hierin dat een registerbepaalde (vrij informele) variëteit een omni-situationele bruikbaarheid

te torsen krijgt, wat erop neerkomt dat hamburgertaal voor de modale Vlaming bevorderd wordt tot een driesterrig taalfestijn.” (Taeldeman 1992: 37)

Daarmee zit Taeldeman overigens op volstrekt dezelfde lijn als Van Istendael, die het begrip ‘Verkavelingsvlaams’ zelfs *definieert* in termen van de maatschappelijke status van haar sprekers:

“Verkavelings-Vlaams, dat is de taal die gesproken wordt in de betere villa’s op de verkavelde grond van onze verminkte dorpen. Het is de taal van de jongens en de meisjes die naar een deftige school gaan en andere kinderen uitlachen omdat die zo onbeschaafd praten.” (Van Istendael 2005: 123)

De substandaardisering in Vlaanderen wordt volgens aantal auteurs met andere woorden gedragen door een spraakmakende gemeente, een economische en intellectuele elite die genoeg prestige geniet opdat hun taalgedrag nagevolgd zou worden – Willemyns hypothetiseert bovendien nog dat ook vrouwen daarin hun uit de sociolinguïstiek bekende voortrekkersrol zouden spelen (zie Willemyns 2002: 16, en Willemyns 2005: 32). Tussentaal hangt in hun ogen onlosmakelijk samen met het proces van emancipatie zoals Vlaanderen dat na de Tweede Wereldoorlog doorlopen heeft. In dat verband sommen zowel Van de Velde als Janssens vijf factoren op – materiële evenals culturele – die tot de genese van de tussentaal hebben bijgedragen, in dalende volgorde van belangrijkheid (zie Van de Velde 1996b: 266-268 en Janssens 2004: 174-175):

1. De negatieve attitude tegenover het dialect
2. De gebrekkige beheersing van het Standaardnederlands en daaraan gekoppeld de negatieve attitude ertegenover, die beide het resultaat zijn van de formalistische (‘zeg niet, zeg wel’) en schoolmeesterachtige manier waarop het taalonderwijs in Vlaanderen georganiseerd werd
3. De toegenomen mobiliteit, zowel geografisch als sociaal, van de Vlamingen
4. Een “conglomeraat” van economische en politieke ontwikkelingen die hebben geleid tot zelfbewustwording, en daarmee gepaard gaande een bepaalde mate van zelfgenoegzaamheid en regionalisering
5. De veranderingen in het medialandschap, in casu de opkomst van de commerciële zenders zoals VTM

Voor de eerste twee – culturele – factoren hebben in de literatuur daarbij de meeste aandacht gekregen; over de meer materiële derde, vierde en vijfde factor wordt er merkbaar minder uitgeweid. Zo haalt De Schutter bijvoorbeeld eveneens de dubbelslachtige afwijzing aan – enerzijds van het dialect maar anderzijds ook van het Standaardnederlands – zoals die al door Van Istendael onder woorden was gebracht (eerste citaat):

“Wat de meeste taalbeschouwers en alle taalbewakers tegen dit soort “tussentaal” hebben, is dat die zowel de rijkdom van de standaardtaal als de spontaneïteit van het dialect mist: het is een taal die uit angst geboren wordt: angst om dialect te spreken, omdat dat niet “netjes” is, en angst om de standaardtaal te spreken, omdat men denkt die niet aan te kunnen, of omdat men wel zeker weet dat men die niet aankan. De tussentaal lijkt in die omstandigheden uitkomst te bieden: dialect is het in geen geval, daarvoor staat de uitspraak garant (en ook het afwijzen van gestigmatiseerde dialectvormen zoals *kinnekes* of *ulielulder*), maar het aantal aperte afwijkingen van de “echte” standaardtaal is wel zo groot dat iedereen meteen door moet hebben dat het om een bewuste keuze gaat. En op die manier, zo wordt verwacht of althans gehoopt, vallen ook echte “fouten”, te wijten aan onkunde, in de rechte plooi.” (De Schutter 1998: 243-244)

Daarmee zoekt De Schutter resoluut aansluiting bij de bekende these over de ‘taalonzekerheid’ van de Vlamingen (voor een overzicht omtrent dit onderwerp, zie Deprez 1981 maar tevens het vermelde artikel van Taeldeman 1992: 34-36). Eén en ander leidt De Schutter er zelfs toe het bestaan van een eventueel endogeniseringsproces eenvoudigweg te ontkennen:

“Wat er dus echt aan de hand is, laat zich als volgt samenvatten: in hun angst voor het “dialectische” gaan veel Belgen, ook Brabanders, ook die uit het culturele centrum van dat gebied, zover dat ze ook elementen die objectief gezien aan hun dialect en de standaardtaal gemeenschappelijk zijn, proberen te vermijden. Het feit dat ze (al even onzekere) sprekers uit andere gebieden andere woorden en uitdrukkingen horen gebruiken, is dan aanleiding genoeg om naar dat “exogene” gebruik over te stappen; cf. De Schutter (1973) en, heel recent, Lebbe (199[8]).” (De Schutter 1998: 246)

Debrabandere besluit zijn analyse van “het echec van de ABN-actie in Vlaanderen” dan weer als volgt:

“Het is ook mijn overtuiging dat Nederlands te vaak als alternatief tegenover het dialect werd voorgesteld. Het dialect was als het ware iets minderwaardigs. Toen de *Vereniging voor Beschaafde Omgangstaal* (VBO) deze vergissing inzag, wijzigde ze in 1979 haar naam in *Vereniging Algemeen Nederlands* (VAN). Niet het dialect is de vijand van het Nederlands. In 1971 (...) wees ik er al op dat niet het dialect het Standaardnederlands bedreigt, maar het Zuid-Nederlands of de ‘kromme schooltaal’. Vandaag noemen we die kromtaal tussentaal of ‘Belgisch-Nederlands’. De ABN-acties hebben de Vlamingen van hun moedertaal, het authentieke dialect, afgewend, in de hoop dat ze zich in plaats daarvan van correct Nederlands zouden bedienen. We hebben ABN aangeprezen, zonder te beseffen dat de Vlamingen geen ABN konden spreken of schrijven, omdat ze die taal onvoldoende kenden. Heel veel jonge Vlaamse ouders denken er goed aan te doen niet langer dialect te spreken met hun kinderen, zonder te weten dat het ‘Schoon Vlaams’ dat ze hun kinderen leren, eigenlijk geen Nederlands is en ook helemaal niet mooi. M.a.w., we hebben het dialect gestigmatiseerd, Vlaanderen zijn dialect afgeleerd en het naar de tussentaal gedreven, al was dat net niet de

bedoeling. Intussen kennen onze jongelui hun mooie dialect niet meer, en kennen ze nog altijd geen Nederlands. Jammer.” (Debrabandere 2005: 31)

De theoretisch meest uitgesponnen interpretatie vinden we ten slotte bij Geeraerts, die in verscheidene artikels (zie Geeraerts 1990, 1993, 2002, 2003) de opkomst van de tussentaal heeft proberen te duiden aan de hand van enerzijds het postmoderne tijdsklimaat met haar onthiërarchisering van de klassieke opdeling tussen hoge en lage cultuur en de flexibilisering van omgangsvormen die daarmee gepaard gaat, en anderzijds de verstoorde verhouding tot het emancipatie-ideaal die het zelfbeeld van de modale Vlaming zou tarten: door de eeuwenlange overheersing van Vlaanderen door vreemde mogendheden zou er in de Vlaamse ‘volksaard’ een zekere distantie geslopen zijn ten opzichte van openbaarheid en gezag. Nu de Vlaming met de socio-economische verbeteringen van na de Tweede Wereldoorlog echter zelf steeds meer tot een autonome burger wordt, belet die attitude hem er paradoxalerwijs in om zich ook volmondig te identificeren met zijn (nieuwe) rol als maatschappelijk geëmancipeerde actor, en meer bepaald de status als beleids- en spraakmaker die daar nu eenmaal per definitie mee samenhangt. De herinnering aan het verleden doet hem er met andere woorden wantrouwig voor terugdeinzen om zich als normbepaler op te werpen. Linguïstisch gezien komt dat tot uiting in een aarzeling om zich van de taalvariëteit te bedienen die bij zijn maatschappelijke positie hoort: het Standaardnederlands. In plaats daarvan trekt de Vlaming zich terug in de aloud vertrouwde geborgenheid van een dialectisch getinte tussentaal, en behoudt ten overstaande van de standaardtaal daarentegen alleen maar een zogenaamde *zondagse-pakmentaliteit* (Geeraerts 2001a, zie verder ook De Schutter 1973: 122, Taeldeman 1992: 50 en Lebbe 1997: 125): “zoals een zondags pak is de hoogste taalnorm iets waarvan de noodzaak buiten kijf staat, maar je er echt goed in voelen doe je niet” (Geeraerts 2001a: 343) – Geeraerts verklaart de tussentaal kortom in termen van een (doorwerking van een) historische *aliënatie* [weliswaar eerder in Hegeliaanse dan Feuerbachiaanse zin], een gedachte waarop ook in Goossens (1991) gealludeerd wordt.

Naarmate de klemtoon zich daarentegen van de culturele naar de materiële factoren verplaatst, groeit er ook enige appreciatie voor het verschijnsel tussentaal. Zo wordt Geeraerts’ pleinvreestheorie uitdrukkelijk door De Caluwe in zijn ‘vierde stelling’ gekapitteld:

“De alomtegenwoordigheid van tussentaal bij de zogenaamde elite, is tekenend voor het toegenomen zelfbewustzijn van de Vlamingen. Lieten die zich vroeger nog gemakkelijk culpabiliseren met betrekking tot hun gebrekkige beheersing van de standaardtaal, dan gebruiken ze nu onbeschroomd tussentaal waar en wanneer zij dat prettig vinden.” (De Caluwe 2002: 62)

De Caluwe vervolgt deze passage dan met een excursie naar Goossens' aanmerkingen over de "inkapseling" van Vlaanderen (zie Goossens 2000: 6-7).

Hij wordt in zijn visie bijgetreden door Cajot, die in enkele artikels (Cajot 1998, 1999a, 1999b, 2000) de tussentaal heeft willen verklaren vanuit een combinatie van de derde factor – de toegenomen mobiliteit – en de vijfde factor – de rol van de media:

"Voor de lagere dialectvervangende registers is er een autonoom standaardiseringsproces op gang gekomen dat verre van beëindigd is. De nieuwe televisiezenders vormen de voedingsbodem waarop zulke standaardtaalvorming kan gedijen: de Vlaamse productiehuisen moeten vooral aan de reclamezenders dagelijks verscheidene uren amusement leveren. De toegenomen mobiliteit en de verdichting van de communicatiestromen versnellen de behoefte aan en de ontwikkeling van zulk bovenregionaal uitdrukingsmiddel. In eerste instantie wordt deze evolutie gekenmerkt door individuele pogingen tot distantiëring van het dialect in de richting van de schrijftaal. Aanvankelijk treden daarin enorme fluctuaties op, maar naarmate de communicatie toeneemt, groeit er om redenen van verstaanbaarheid en prestige een nivelleringstendens, zet een geleidelijke stabilisering in, en wordt de ruimte voor variaties kleiner." (Cajot 1998: 1003-1004)

Daaruit trekt Cajot dan een inmiddels veel geciteerde conclusie: "Deze omgangstaal is integendeel een zelfstandige talige grootheid (aan het worden), d.w.z. moedertaal van veel dialectlozen en doeltaal van veel dialectsprekenden: het tastbaar maar wankel resultaat van een autonome informele taalstandaardisering." (Cajot 1998: 1005)

Los van de precieze factoren waaraan elke auteur belang wil hechten lijkt iedereen het er verder evenwel over eens te zijn dat de tussentaal in Vlaanderen zonder meer een *andere* taal dan het Standaardnederlands is: een taal die met andere woorden simpelweg náást de officiële standaardtaal zou bestaan – wat natuurlijk al blijkt uit de hele serie namen die ervoor in omloop zijn. De Vlaamse taalsituatie wordt daarmee primair opgevat als één van *diglossie*. De meest uitgesproken formulering in die trant is afkomstig van Goossens:

"In ieder geval kan men aannemen dat Vlaanderen onder de genoemde voorwaarden op weg is naar een nieuwe tweetaligheid. Dit moet letterlijk opgevat worden: het gaat om twee talen met een eigen, linguïstisch beschrijfbaar structuur." (Goossens 2000: 8)

Ook Beheydt beweert: "In het Zuiden is op dit moment in de spreektaal een dubbele standaard aan het groeien" (Beheydt 2003: 160), een proces dat hij omschrijft als 'destandaardisering'. Het minst aan de verbeelding overlatend is misschien nog wel

het feit dat de taalraadsman van de VRT en als dusdanig de officieuze Vlaamse taalnorbepaler bij uitstek, Ruud Hendrickx, in zijn Taalcharter onomwonden stelt: “Tussentaal is geen verworven register van de Nederlandse standaardtaal, zoals de Umgangssprache in Duitsland en de informele omgangstaal in Nederland dat wel zijn” (Hendrickx 1998a). Elders heeft Hendrickx zijn taalbeleid in nog een aantal artikels verder uiteengezet (zie Hendrickx 1998b, 1999, 2003).

De diglossie van het Belgisch-Nederlands – en dan meer bepaald Goossens’ karakterisering ervan in termen van ‘beschrijfbaarheid’ – raakt aan wat voor de meeste auteurs uiteindelijk de steen des aanstoots in de hele tussentaaldiscussie vormt: het gebrek aan beschrijving ervan. Jaspers stelt één en ander scherp als hij het wijt aan de dominante standaardtaalideologie die enkel het Standaardnederlands als legitieme variëteit beschouwt:

“Eigen aan een ideologie is dat feiten die niet passen binnen het ideologische kader onopgemerkt blijven of weggeredeneerd worden. Gal & Irvine spreken dan van ‘deletie’ (1995: 974). Dat verklaart wellicht de virtuele afwezigheid van tussentaalbeschrijving, en het onvermeld blijven van de overeenkomsten tussen tussentaal en de standaardtaal. Wat eveneens onvermeld blijft, is dat veel tussentaalsprekers waarschijnlijk regelmatig en efficiënt gebruik maken van de standaardvariëteit (vooral in geschreven, maar ook in gesproken vorm). Ook blijft de in Vlaanderen bestaande meertaligheid of diglossie en het massale code-wisselen onopgemerkt. Verder is er sprake van deletie doordat verzet tegen de standaardideologie en standaardvariëteit nauwelijks vernoemd wordt. Het is opvallend dat er geen taalkundige beschrijving bestaat van waar tussentaal wordt gesproken, op welke momenten, door wie en waarom.” (Jaspers 2001: 145-146)

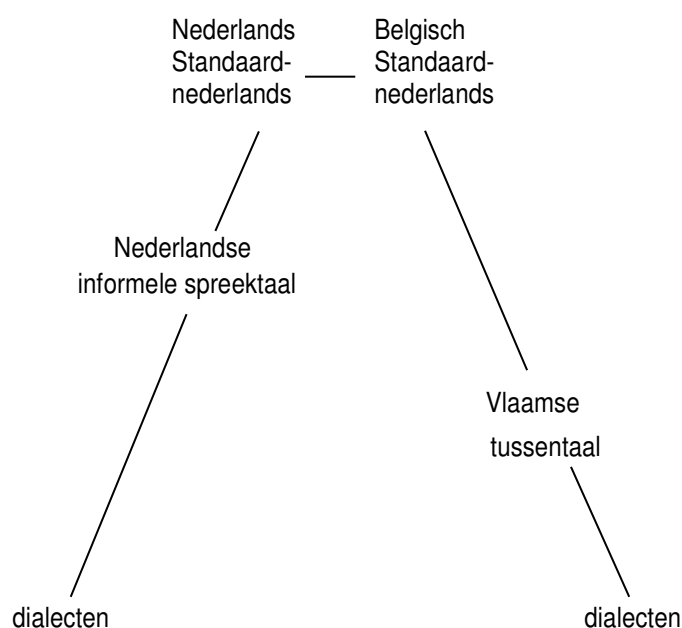
Later zou Jaspers zijn pijlen zelfs expliciet op het Taalcharter richten – zie Jaspers (2002), daarbij evenwel vinnig van repliek gediend in Hendrickx (2002); zie verder nog Jaspers (2005).

Toch behoeft het opgehangen beeld omtrent de descriptieve leemte enige bijstelling. Zo biedt Ryckeboer (2004) bijvoorbeeld (voor de periode van 1993 tot en met 2003) een uitgebreide catalogus van licentiaatsscripties en –verhandelingen waarin verscheidene facetten van de Belgisch-Nederlandse variatiestructuur uitgelicht worden – zie in dat verband ook nog de studies gebundeld in Belemans & Vandekerckhove (1999).

Van belang is in dat opzicht het onderzoek dat door Geeraerts en zijn onderzoeksgroep QLVL uitgevoerd is, omdat hier op een kwantitatieve en corpusgebaseerde manier getracht wordt om het hele veld van de Belgisch-Nederlandse grammatica te bestrijken: lexicale variatie in Geeraerts, Grondelaers & Spielman (1999), Grondelaers (2000) en Van Gijssel (2007), morfologische variatie in

Tummers (2005) en syntactische variatie in De Sutter (2005). Relevant voor de specifiek morfosyntactische kenmerken van de tussentaal zijn vooral Geeraerts, Penne & Vanswegenoven (2000) over de verschillen tussen de soaps *Thuis* en *Familie* en Van Gijssel, Geeraerts & Speelman (2004) over taalgebruik in reclamespots – de eerste studie wordt daarbij gerecapituleerd in Geeraerts (2001b), de tweede dient in menig opzicht als inspiratiebron voor het Gentse onderzoek naar ondertitelingsstrategieën zoals gerapporteerd in Slembrouck & Van Herreweghe (2004). Beide artikels zijn overigens nadere uitwerkingen van licentiaatsverhandelingen: zie Penne & Vanswegenoven 1999 en Van Gijssel 2001 respectievelijk; daarvan zijn er verder wel meer: zie Binnemans 2003 en Zenner 2007 bijvoorbeeld, Vandenhende 2001 en Impe 2006 focussen dan weer meer op attitudes, terwijl De Bisschop 2003 een analyse van de taalpolitieke discussie biedt – ook in Leuven is er met andere woorden door studenten nogal wat onderzoek uitgevoerd.

De voornaamste vaststelling van het Leuvense onderzoek is daarbij dat de heterogeniteit tussen het formele en informele register in België beduidend *groter* is dan de heterogeniteit tussen analoge registers in Nederland (zie Geeraerts en De Sutter 2003 voor een overzichtsartikel, en verder nog Geeraerts 1999a), een gegeven dat als volgt gevisualiseerd kan worden (Figuur 1.1):



Figuur 1.1: De stilistische stratificatie van het Nederlands

Geeraerts bestempelt deze eigenschap van het Belgisch-Nederlands met name als een *taalkloof* (Geeraerts 1999b) – wat trouwens niet hetzelfde is als de bovenvermelde notie van diglossie: diglossie behelst namelijk de aanwezigheid

binnen een taalgemeenschap van twee vormelijk, en dus kwalitatief, verschillende talen. Een taalkloof duidt veeleer op een kwantitatief verschil: het feit dat er tussen twee of meer spreek situaties grote verschillen in gebruiksfrequentie bestaan van bepaalde taalvarianten (waarbij die laatste voor alle situaties kwalitatief dezelfde set is) – een taalkloof is daarmee in zeker zin het tegenovergestelde van het begrip ‘uniformiteit’, zoals geoperationaliseerd in Geeraerts, Grondelaers & Spielman (1999); zie ook hoofdstuk 4.

Daarnaast is ook nog Vandekerckhove (2004) interessant omdat het een uniek vergelijkingspunt voor onze studie biedt: Vandekerckhove bestudeert namelijk enkele van de meest prototypische kenmerken van het Belgisch-Nederlands – te weten: de pronominale aanspreekvormen – aan de hand van hetzelfde corpus dat wij zullen gebruiken, met name: het CGN (zie hoofdstuk 2). Haar constatering is daarbij dat in het informele register het G-systeem voor heel Vlaanderen de default blijft en bij de jongere generaties zelfs nog in gebruiksfrequentie toeneemt, behalve voor de West-Vlaamse sprekers die zich eerder van de J-vormen bedienen, “waardoor – ironisch genoeg – het niet-dialectische informele taalgebruik van de meest dialectvaste regio van Nederlandstalig België (...) voor deze variabele het dichtst de standaardtaalnorm benadert” (Vandekerckhove 2004: 990); zie in dat verband ook nog Vandekerckhove (2006), Vandekerckhove & Van Rooy (2005), en in het bijzonder voor het West-Vlaams: Lebbe (1997).

Een punt waar we tot slot nog op willen terugkomen is de problematische invulling van het begrip standaardtaal zoals dat uit de voorgaande bespreking wel gebleken moge zijn. De verankering van de tussentaal in het Vlaamse taallandschap maakt het namelijk niet meer zo eenduidig te bepalen welk taalgebruik nu precies in Vlaanderen de standaard uitmaakt in de ‘technische’ zin van het woord: de door de taalgemeenschap algemeen aanvaarde norm – of zoals de officiële grammatica van het Nederlands het zelf uitdrukt: “de taal waarin geen elementen of structuren voorkomen die duidelijk opvallen als niet-algemeen” (ANS: 16). Het zou met name best wel eens kunnen dat de acceptatie van heel wat verschijnselen die onder de noemer ‘tussentaal’ ressorteren wijder verbreid is dan doorgaans wordt vermoed (of toegegeven), een hypothese die enigszins bevestigd wordt in de recent verschenen licentiaatsverhandelingen van Vandenhende (2001) en Impe (2006): het Standaardnederlands wordt door de informanten in beide attitudestudies weliswaar ondubbelzinnig beladen met ‘overt prestige’, maar de tussentaal al even onomwonden met ‘covert prestige’. Het doet de vraag rijzen in hoeverre tussentaalkenmerken niet gewoon de Vlaamse versie zijn van spreektaalverschijnselen zoals De Vries (2001) ze voor het Nederlandse Nederlands opsomt. In dat geval zouden ze echter ‘impliciet’ gestandaardiseerd zijn, en zijn benamingen als ‘tussentaal’ en/of ‘standaardtaal’ zonder meer misleidend. Ze verwijzen mogelijkerwijs alleen maar naar een zuiver virtuele standaard, waarvan het normatieve statuut enkel in naam bestaat. Dat alles hangt samen met een oud en



bekend zeer in de Vlaamse taalpolitiek (zie Jaspaert 1989, 1990), met name dat taalnormen in Vlaanderen steevast opgevat worden als Platoonse absoluta in plaats van sociaal geëmergeerde, arbitraire conventies. Het probleem wordt acuut gesteld door Haeseryn, die daar eveneens het Taalcharter toe aangrijpt:

“In het *Taalcharter* staat: “In Vlaanderen wordt de standaardtaal (...) niet of nauwelijks gedragen door een “spraakmakende gemeente”. In onze buurlanden wordt die mede gevormd door politici, bedrijfsleiders en academici, maar in Vlaanderen kan hun taal bezwaarlijk een voorbeeld worden genoemd”. Hier worden naar mijn gevoel twee zaken door elkaar gehaald. Dat de standaardtaal niet of nauwelijks door een spraakmakende gemeente gedragen zou worden, lijkt mij een wat voorbarige conclusie. Ten eerste is dat min of meer in tegenspraak met wat eerder gezegd is over het ‘bewust verzorgde taalgebruik van de taalgevoelige Vlamingen’ en ten tweede moet eerst nog maar eens vastgesteld worden of de spraakmakende gemeente in Vlaanderen *principieel* verschillend is van die in andere taalgemeenschappen. Waarom zouden politici, bedrijfsleiders en academici bij voorbaat van de spraakmakende gemeente uitgesloten moeten worden? Wat men bedoelt te zeggen is dat men vindt dat een deel van die politici enz. – we kunnen ook niet iedereen over een kam scheren – niet voldoet aan een bepaalde norm, die kennelijk als enig juiste wordt beschouwd. Maar hiermee keren we de zaken om en dreigen we in een cirkelredenering terecht te komen. Ook op het specifieke terrein van taalnormvorming is dus dringend nader (onder meer attitudineel) onderzoek gewenst. Als uit zulk onderzoek naar voren zou komen – ik zeg ‘als’, want misschien valt het allemaal reuze mee – dat de norm van de spraakmakende gemeente (gedeeltelijk) anders is dan men gedacht had of wenselijk had, dan is dat zo, of men het nu leuk vindt of niet (...).” (Haeseryn 2004: 485)

In het licht van deze begripsverwarring zullen we dan ook één en ander terminologisch (her-)ijken. Termen zoals ‘standaardtaal’ en/of ‘tussentaal’ zullen we in de rest van dit proefschrift niet langer *an sich* gebruiken. Voor verschijnselen die tot de in Vlaanderen officieel gesanctioneerde ‘taal’ behoren zullen we daarentegen het steeds gangbaarder wordende *Journaalnederlands* hanteren. Kenmerken van de ‘nieuwe’ Vlaamse ‘taal’ zullen we met Geeraerts dan weer betitelen als *Soapvlaams* (zie Geeraerts 1998).

## 1.2 Naar een beschrijving van het Soapvlaams – *hypothesen*

Op basis van wat er in het bovenstaande over het Soapvlaams geponeerd is, verwachten we (gemakshalve specificeren we de hypothesen al meteen in termen van de taalexterne parameters uit het CGN – zie hoofdstuk 2):

- Met betrekking tot de *taalstructuur*: dat de diglossie van het Belgisch-Nederlands tot uiting komt in twee onderscheiden clusters Soapvlaams en Journaalnederlands

- Met betrekking tot *register*: dat het Soapvlaams sterk met de informele spreesituaties geassocieerd is en Journaalnederlands met de formele, waarbij er bovendien een aanzienlijke afstand – oftewel, een ‘taalkloof’ – tussen beide registers bestaat
- Met betrekking tot *regio*: dat het Soapvlaams vooral met de Brabantse regio geassocieerd zal zijn, waarbij de overige Vlaamse provincies krachtens de werking van de Brabantse expansie er een zekere invloed van te kennen zullen geven, behalve misschien voor West-Vlaanderen, waar de dialectische achtergrond nog sterk kan meespelen (Vandekerckhove)
- Met betrekking tot *leeftijd*: dat het Soapvlaams vooral door de jongere generaties gesproken zal worden
- Met betrekking tot *sekse*: dat het Soapvlaams enigszins een associatie met vrouwelijke sprekers laat zien (Willemysn)
- Met betrekking tot sociale *klasse*: dat het Soapvlaams sterk met de maatschappelijk hogere klassen geassocieerd zal zijn – in zowel economisch opzicht, bijvoorbeeld: beroepsniveau, als intellectueel: educatieniveau (zie hoofdstuk 2)

Daarnaast willen we er ons in dit proefschrift uitdrukkelijk op toeleggen om de inventaris aan vormkenmerken enigszins uit te breiden. Op die manier hopen we om een representatiever beeld van de Vlaamse spreektaal op te hangen dan tot nog toe gangbaar was. Sinds kort kan dat namelijk moeiteloos door de release van het Corpus Gesproken Nederlands.

In het volgende hoofdstuk willen we dat corpus dan ook kort voorstellen, om het daarna in het derde hoofdstuk aan een grondige verkenning te onderwerpen. Dat zal dan leiden tot de operationalisering van een set van relevante taalvariabelen. Het vierde hoofdstuk schetst vervolgens de methodologie waarmee we onze variatiestudie zullen uitvoeren. Het vijfde en zesde hoofdstuk bevatten de analyses: het vijfde behandelt elke taalexterne factor apart; het zesde neemt ze allemaal samen op in één globale analyse. Dit proefschrift beëindigen we in een zevende hoofdstuk met enkele concluderende beschouwingen.

## Hoofdstuk 2 | Corpus Gesproken Nederlands

Het corpus dat we voor onze doeleinden zullen gebruiken is het Corpus Gesproken Nederlands, afgekort als het CGN. Het CGN is met name geschikt omdat het uitdrukkelijk een “plausibele” steekproef van het hedendaagse gesproken Nederlands wil zijn (Oostdijk 2000a, 2000b, 2000c). Zo is er bijvoorbeeld naar gestreefd om sprekers uit alle lagen van de maatschappij te vertegenwoordigen. Daarnaast is het CGN vooral ook een geannoteerd corpus: voor tal van linguïstische kenmerken zijn de opnames getagd (zie Schuurman et al. 2003). De hierna volgende bespreking zal daarom uit vier punten bestaan. Eerst wordt het design en daarmee samenhangend de interne stratificatie van het CGN besproken. Daarna wordt er ingegaan op de sociale informatie die het CGN over haar sprekers verschaft. Omdat we onze gegevens door middel van reguliere expressies uit het CGN zullen halen, wordt vervolgens de mark-up van de eigenlijke CGN-bestanden aan de hand van een voorbeeld toegelicht. In het licht van onze onderzoeksvraag naar het Soapvlaams wordt tot slot nog uit de doeken gedaan hoe het CGN dialectische en/of substandaardtalige elementen markeert.

### 2.1 Het design van het CGN

Het CGN bestaat uit opnames van gesproken Nederlands zoals dat rond de millenniumwissel te horen valt (het project duurde van 1998 tot 2003). De totale grootte bedraagt zo'n 1000-tal uren spraak, wat neerkomt op ongeveer 10 miljoen woorden. Daarvan is 2/3 afkomstig uit Nederland, en 1/3 uit Vlaanderen, conform de populatieverhouding tussen beide landen. Omdat ons onderzoek primair de Vlaamse spreektaal betreft, gebruiken we enkel dat Vlaamse deel.

Bij de samenstelling van het CGN is er voorts rekening mee gehouden dat taal – en a fortiori spreektaal – kan verschillen naargelang de spreesituatie. Bijgevolg is het CGN gestratificeerd in een aantal subcorpora die gekarakteriseerd kunnen worden in termen van parameters zoals dialogisch-monologisch, spontaan-voorbereid, privé-publiek, en dergelijke (zie Oostdijk 2002). In het CGN-jargon worden die subcorpora ‘componenten’ genoemd. Een overzicht van alle componenten (samen met hun respectieve corpusgroottes) geeft tabel 2.1.

Label	Component	Aantal woorden	VL	NL
a	Spontane conversaties ('face-to-face')	2.626.172	878.383	1.747.789
b	Interviews met leraren Nederlands	565.433	315.554	249.879
c	Telefoondialogen opgenomen m.b.v. platform	1.208.633	465.096	743.537
d	Telefoondialogen opgenomen m.b.v. minidisc recorder	853.371	343.167	510.204
e	Zakelijke onderhandelingen	136.461	0	136.461
f	Interviews en discussie uitgezonden op radio en televisie	790.269	250.708	539.561
g	Discussie, debatten, vergaderingen (m.n. politieke)	360.328	138.819	221.509
h	Lessen	405.409	105.436	299.973
i	Spontane commentaren (o.a. sport) uitgezonden op radio en televisie	208.399	78.022	130.377
j	Actualiteitenrubrieken en reportages uitgezonden op radio en televisie	186.072	95.206	90.866
k	Nieuwsbulletins uitgezonden op radio en televisie	368.153	82.855	285.298
l	Beschouwingen en commentaren uitgezonden op radio en televisie	145.553	65.386	80.167
m	Missen, lezingen, plechtige toespraken	18.075	12.510	5.565
n	Colleges, voordrachten, lezingen	140.901	79.067	61.834
o	Voorgelezen teksten	903.043	351.419	551.624
<b>Totaal</b>		<b>8.916.272</b>	<b>3.261.628</b>	<b>5.654.644</b>

*Tabel 2.1: De componenten van het CGN*

Component e ('zakelijke onderhandelingen') bevat enkel taal materiaal uit Nederland, zodat er voor Vlaanderen 14 componenten overblijven. Die zullen de basis vormen voor onze bestudering van de registervariatie in het gesproken Belgisch-Nederlands.

## 2.2 De sprekersinformatie in het CGN

Het Vlaamse deel van het CGN telt 1351 sprekers. Elk van hen werd, in zoverre dat haalbaar bleek, voor zo'n 23 kenmerken geclassificeerd (zie Oostdijk 2000d – ontbrekende informatie voor een bepaald kenmerk wordt weergegeven door middel van een 'X'; als de identiteit van een spreker echter in het geheel niet te achterhalen valt, dan wordt er 23 keer het label 'unknown' toegekend). De volledige lijst van sprekerskenmerken is te vinden op

[http://lands.let.kun.nl/cgn/doc\\_Dutch/topics/version\\_1.0/metadata/speakers.htm](http://lands.let.kun.nl/cgn/doc_Dutch/topics/version_1.0/metadata/speakers.htm).

Voor ons onderzoek zullen we er daar evenwel maar vijf van gebruiken, te weten: sex, birthYear, birthRegion, eduLevel, en occLevel.

### 2.2.1 sex

Verreweg de meest voor zichzelf sprekende factor is natuurlijk sex, het biologische geslacht van de spreker. Mannelijke sprekers krijgen het label 'sex1', vrouwelijke sprekers 'sex2'. Voor onze analyses zullen we de numerieke codering evenwel vervangen door hoofdletter 'M', respectievelijk 'F'.

### 2.2.2 birthYear

De factor birthYear geeft het geboortjaar van de spreker. De labels zijn dan ook telkens een getal van vier cijfers, waarvan de eerste twee '19' (alle sprekers zijn geboren gedurende de twintigste eeuw). In de analyse zullen we de geboortejaren evenwel groeperen in enkele leeftijdsklassen. Meer bepaald zullen we twee schema's hanteren. Enerzijds vergelijken we verschillende *decennia* van sprekers met elkaar (bijvoorbeeld, de twintigers, de dertigers, de veertigers, etc.). Anderzijds gaan we ook te rade bij het sociologische werk van Henk Becker (1992) over de *generaties* van de twintigste eeuw. De precieze afbakening van de cohorten volgt in 5.4.

### 2.2.3 birthRegion

Regionale informatie over de spreker vat het CGN door middel van drie factoren: birthRegion (regio waar de spreker geboren is), resRegion (regio waar de spreker woonachtig is), en eduRegion (regio waar de spreker gestudeerd heeft). Voor ons onderzoek viel onze keuze daarbij op birthRegion, omdat die factor wellicht het meest betrouwbaar de dialectische achtergrond van de spreker weerspiegelt – hoewel we de verschillen met resRegion eveneens kort zullen aanstippen. Er wordt een hele reeks regio's onderscheiden (voor een exhaustieve lijst, zie opnieuw

[http://lands.let.kun.nl/cgn/doc\\_Dutch/topics/version\\_1.0/metadata/speakers.htm](http://lands.let.kun.nl/cgn/doc_Dutch/topics/version_1.0/metadata/speakers.htm)).

Vrijwel de belangrijkste daarvan zijn volgende vier:

- regV1: Antwerpen en Vlaams-Brabant
- regV2: Oost-Vlaanderen
- regV3: West-Vlaanderen
- regV4: Limburg

In de analyses zullen we evenwel respectievelijk de labels 'bra', 'ovl', 'wvl', en 'lim' gebruiken.

#### 2.2.4 *eduLevel*

De factor *eduLevel* catalogeert het opleidingsniveau van de spreker. Het is een ternaire factor:

- edu1: hoog
- edu2: midden
- edu3: laag

De labels die we voor de analyses zullen gebruiken zijn respectievelijk 'hie', 'mid', en 'low'.

#### 2.2.5 *occLevel*

Het beroepsniveau van de spreker wordt ten slotte aangeduid door *occLevel*. In tegenstelling tot Nederland onderscheidt het CGN voor Vlaanderen 10 niveaus (voor de precieze details, zie nogmaals

[http://lands.let.kun.nl/cgn/doc\\_Dutch/topics/version\\_1.0/metadata/speakers.htm](http://lands.let.kun.nl/cgn/doc_Dutch/topics/version_1.0/metadata/speakers.htm)).

Deze zijn:

- occA: werkzaam in het hoger management of bestuur
- occB: beroep dat een hogere opleiding veronderstelt
- occC: werkzaam als onderzoeker of docent aan universiteit of instelling voor hoger onderwijs
- occD: werkzaam in administratieve functie of een functie in de zorgsector
- occE: ongeschoold beroep
- occF: werkzaam als zelfstandige
- occG: politicus
- occH: werkzaam in de media (journalist, reporter) of als artiest

- occl: student, stagiair
- occJ: niet beroepsmatig actief

Aan een handvol sprekers kent het CGN een combinatie van categorieën toe. Zo slaat 'occC+G' bijvoorbeeld op een hoogleraar die tevens politicus is. Zulke combinatiecategorieën zullen we evenwel (terug) tot één van de tien hoofdcategorieën reduceren, en in Tabel 2.2 wordt aangegeven hoe.

<b>Combinatie</b>	<b>Voorbeeld</b>	<b>Hoofdcategorie</b>
occA+B	voorzitter sportorganisatie & arts	occB
occA+B+C	directeur & architect & docent	occA
occA+B+F	voorzitter sportorganisatie & arts & sportman	occB
occA+C	directeur & professor, directeur-generaal & professor, onderdirectrice school & leerkracht, directeur school & leerkracht	occA
occA+G	bedrijfsleider & parlamentslid	occA
occB+C	advocaat & professor, architect & docent, arts & professor	occC
occB+C+G	advocaat & professor & senator	occC
occB+C+H	architect & docent & criticus	occB
occB+D	priester & maatschappelijk werker	occD
occB+G	advocaat & senator, arts & parlamentslid, advocaat & gemeenteraadslid, advocaat & parlamentslid (6), arts & senator, advocaat & politicus (4)	occB
occB+H	priester & publicist	occH
occC+D	leerkracht & therapeute, docent & verpleegster, leerkracht & werkmeester (technisch adviseur)	occD
occC+G	leerkracht & burgemeester, docent & politicus, professor & minister (2), leerkracht & parlamentslid	occC
occC+H	leerkracht & journalist, universitaire staf & correspondent, onderwijzer & sportverslaggever, leerkracht & auteur, professor & producer, leerkracht & recensent, lesgever & kunstenaars, opleider & auteur, docent hogeschool & journalist	occH
occD+F	vormingswerker & lasser, diëtiste & landbouwster	occF
occD+G	maatschappelijk assistent & burgemeester, verpleegkundige & gemeenteraadslid	occD
occF+H	kok & artistiek medewerker	occF
occG+H	gemeenteraadslid & presentatrice (& auteur), politicus & journalist, parlamentslid & journalist	occH

Tabel 2.2: De combinatiecategorieën bij beroepsniveau



### 2.3 De mark-up van het CGN

Naast sprekerskenmerken codeert het CGN ook voor tal van linguïstische features. Zo zijn alle opnames orthografisch getranscribeerd (voor het protocol dat daarbij gehanteerd werd, zie Goedertier & Goddijn 2000). Vervolgens zijn deze transcripties verrijkt met woordsoortinformatie in de vorm van ‘Part-Of-Speech’- oftewel POS-tags, en verder zijn ze ook gelemmatiseerd (Van Eynde 2004). Dit is het zogenaamde ‘kerncorpus’, dat voorts vergezeld wordt van een intern lexicon waarbij allerlei scheidbare woorden lexicologisch gekoppeld zijn (Piepenbrock 2004). Daarnaast heeft een deel van zo’n 1 miljoen woorden zowel een brede fonetische transcriptie gekregen (Gillis 2001) als een syntactische annotatie (Hoekstra et al. 2003). De fonetische transcriptie is daarbij gekoppeld aan de orthografische transcriptie – een proces dat ‘oplijning’ heet (Binnenpoorte 2002/4). Tot slot is er nog een deel van 250.000 woorden prosodisch geannoteerd (Martens 2002).

Het corpus dat we in deze studie zullen consulteren is het kerncorpus. Voor Vlaanderen is dat verspreid over 3911 tekstbestanden in ASCII-formaat (alleen is de extensie .pos in plaats van .txt). Ieder bestand is opgedeeld in zogenaamde ‘annotation units’, die elk ruwweg samenvallen met een zin. Zo’n annotation unit ziet er als volgt uit:

```
<au id="5" s="V40175" tb="5.106">
ik    VNW(pers,pron,nomin,vol,1,ev)    ik
heb   WW(pv,tgw,ev)                   hebben
nu    BW()                               nu
minder VNW(onbep,grad,stan,prenom,zonder,agr,comp)  weinig
tijd  N(soort,ev,basis,zijd,stan)      tijd
om    VZ(init)                          om
te    VZ(init)                          te
lezen WW(inf,vrij,zonder)              lezen
.     LET()                               .
```

Een annotation unit begint met andere woorden met een regel tussen vishaken. Op die regel vinden we 4 gegevens, van elkaar gescheiden door een spatie: ‘au’ ter aanduiding van het begin van een nieuwe annotation unit, een volgnummer (‘id’), het id van de spreker van de uiting (‘s’), en een indicatie van het aantal seconden dat sinds het begin van het gesprek verstreken is (‘tb’). Dan volgt de eigenlijke uiting waarbij elk woord op een nieuwe regel staat. Die regels zijn allemaal van dezelfde vorm: token + tab + POS-tag + tab + lemma (+ ‘newline’). Een annotation unit eindigt tot slot met de sequentie leesteken + tab + ‘LET()’ + tab + nogmaals het leesteken. Het CGN onderscheidt daarbij drie mogelijke leestekens: punt (.), beletseltoken (...), en vraagteken (?). Daarna volgt de nieuwe annotation unit natuurlijk op een nieuwe regel.

De annotation units volgen elkaar zo op tot het einde van het bestand. Dat is meteen ook het einde van een gesprek, aangezien elk bestand in het kerncorpus de (geannoteerde) weerslag is van één volledig gesprek. Het Vlaamse kerncorpus bevat met andere woorden 3911 verschillende gesprekken.

## 2.4 Het dialect in het CGN

De primaire focus van het CGN ligt op het bovengewestelijke, en dus veelal standaardtalige, Nederlands (zie Oostdijk 2000a: 281). Substandaardtalige elementen – niet ongewoon in gesproken taal, zeker wat Vlaanderen betreft – worden daarbij tweevoudig gemarkeerd, en wel op verschillende annotatieniveaus. Bij de orthografische transcriptie wordt aan woorden die niet tot het CGN-lexicon behoren een extra tag toegevoegd. Daarnaast krijgen deze woorden aparte labels tijdens de POS-tagging. Opmerkelijk is overigens wel dat dit niet eenvormig gebeurt is: hoewel de eerste annotatielaag in principe de input vormt voor de tweede, zijn er in werkelijkheid heel wat woorden waarbij de ene annotatie de andere niet ‘dekt’. Louter manuele verificatie kon hier soelaas bieden.

Tabel 2.3 geeft de orthografische codes voor de acht manieren waarop een bepaald woord niet tot het CGN-lexicon kan behoren. Bij de transcriptie wordt zo’n code gewoon aan het einde van een woord aangehecht:

Tag	Categorie
*v	Woorden uit een <b>vreemde taal</b> (bijv.: tomorrow*v)
*d	<b>Dialectwoorden</b> (bijv.: kortewagen*d; keuje*d)
*z	<b>Woorden</b> uit de standaardtaal die <b>zwaar dialectisch zijn uitgesproken</b>
*n	<b>Nieuwe woorden</b> (bijv.: zerotolerantie*n)
*t	<b>Nieuwe tussenwerpsels</b> (bijv.: amaa!*t; hoppa*t)
*a	<b>Afgebroken woorden</b> (bijv.: uitges*a; verpr*a)
*u	Weergeven van de <b>uitspraak</b> en <b>versprekingen</b> (bijv.: boink*u; gespreken*u)
*x	Woorden die <b>niet goed verstaanbaar</b> zijn

Tabel 2.3: De acht orthografische tags

Voor onze doeleinden is categorie ‘\*d’ natuurlijk het meest toepasselijk (hoewel we de andere zeven ook hebben gecontroleerd; vooral categorie ‘\*z’ is daarbij – begrijpelijkerwijs – soms erg relevant gebleken).

Verder hanteert het CGN nog een hele reeks conventies voor meer specifieke gevallen. Zo worden de typisch Zuid-Nederlandse aanspreekvormen *ge* en *gij*, alsook de verkleinwoorden op *-ke* beschouwd als standaardtalig, en worden bijgevolg niet gemarkeerd (zie regel D2b en D2c respectievelijk in Goedertier & Goddijn 2000: 8-9). Anderzijds worden de stapelvormen *ekik* en *degij* getranscribeerd met “k\*d” en “de\*d” op één regel en “ik” en “gij” op de volgende (D4b en D4c in Goedertier & Goddijn 2000: 10). Over *zezij* vermeldt het CGN-protocol evenwel niets. Ten slotte wordt het gereduceerd pronomen *me* (1.mv.) genoteerd als “me\*d” (D4c – Goedertier & Goddijn 2000: 10), maar over het enclitische ‘*m*’ (3.enk.masc.) wordt opnieuw niets gezegd (wat dan ook tot uiting komt in de totale chaos waarmee dat pronomen geannoteerd is – zie hoofdstuk 3).

Naast orthografisch worden dialectische woorden dus ook nog op POS-niveau gemarkeerd. Meer bepaald gebeurt dat door middel van een aparte set tags, die erdoor gekenmerkt worden dat ze eindigen op de waarde ‘dial’. Voorts is de informatie die ze aanbieden minder gedetailleerd dan die voor de standaardtalige woorden. De motivatie die daarbij aangevoerd wordt, is precies dat de morfologie van de Nederlandse dialecten zo sterkt afwijkt van die van het Standaardnederlands (zie Van Eynde 2004: 7). Een voorbeeld is:

elken\*d          VNW(onbep,det,dial) elken

De volledige lijst van POS-tags is te vinden in Van Eynde (2004: 62-63).



## Hoofdstuk 3 | Taalvariabelen

Het opstellen van de taalvariabelen zal bewust rekening houden met de eigenschappen van het CGN, en dan niet alleen om de voor de hand liggende reden dat de automatische zoekinstructies (de zogenaamde ‘queries’) gebaseerd moeten zijn op de mark-up van de CGN-bestanden. Nog voor we queries kunnen definiëren, willen we vooral ook weten wat er aan substandaardtalig taalgebruik zoal voorhanden is in het gesproken Belgisch-Nederlands. Daarom zullen we in een eerste deel even verkennend nagaan wat het CGN allemaal als dialectisch markeert. Drie groepen van verschijnselen blijken daarbij dan als relevant uit de bus te komen. De opgedane inzichten zullen we vervolgens aanwenden bij de operationalisering van relevante taalvariabelen. Dat zijn er 62 in totaal, en die vormen dus de basis voor de variationele analyses.

### 3.1 De verkenning van het CGN

Voor we taalvariabelen zullen opstellen, willen we vooral de verschijnselen onder de loep nemen die het CGN als substandaardtalig markeert. Dat zijn dus diegene ofwel met een orthografische ‘\*d’- en (in mindere mate) ‘\*z’-tag, ofwel met een ‘dial’-tag op POS-niveau<sup>1</sup>. Zo’n aanpak heeft dan het voordeel dat we een getrouwer beeld krijgen van wat de Vlaamse spreektaal nu precies inhoudt dan tot voor kort gangbaar was. Tot nog toe is de aan het Soapvlaams gewijde literatuur, als ze al een overzicht van kenmerken gaf, namelijk steevast beperkt geweest tot een of andere anekdotische registratie van de meest frappante daarvan. Het CGN is daarentegen een representatieve steekproef van het gesproken Belgisch-Nederlands. Van de dialectisch getagde vormen mag dan ook verwacht worden dat ze elk hun eigen stek hebben weten te veroveren in het hedendaagse Vlaamse taallandschap.

Het meest in het oog springende is daarbij natuurlijk de bonte verscheidenheid, en dat heeft alles te maken met de dialectische achtergrond. Vooral de uitspraakvarianten enerzijds en de lexicale afwijkingen anderzijds zijn veelvuldig vertegenwoordigd. Omdat onze interesse evenwel primair uitgaat naar de flexie

---

<sup>1</sup> Alle relevante AV-workshops kunnen teruggevonden worden op de bijgevoegde CD-ROM.

zullen we deze twee vormen van substandaardtaligheid louter ter illustratie overlopen.

Uitspraakvarianten treffen we vooral aan bij de grammaticale woordsoorten: de voorzetsels, de voegwoorden, en de tussenwerpsels. Voorbeelden zijn substandaard realisaties van *aan*, *allee*, *als*, *bij*, *dat*, *door*, *ja*, *juist*, *maar*, *met*, *naar*, *nee*, *nu*, *ook*, en *voor*. Zoals blijkt gaat het hier meestal om de bekende woordfinale deleties. Lexicale afwijkingen komen dan weer voor bij de inhoudswoorden zoals de substantieven, adjectieven, werkwoorden, bijwoorden en verder nogmaals tussenwerpsels: *ambetant*, *azo*, *bleiten*, *bomma*, *chance* ('geluk'), *dierf* (=durfde, verl. tijd van *durven*), *ele(n)triek*, *gans*, *goesting*, *hof*, *jawadde*, *kost* (=kon, verl. tijd van *kunnen*), *lijk* ('zoals'), *machien*, *morgend*, *peis* (van *peinzen* 'denken'), *salukes*, *sebiet*, *sloef*, *zakes*, en *zenne*. Bij de adjectieven vinden we evenwel ook enkele representanten van de eerste flectionele groep van varianten: de attributieve vormen uitgaande op *-en*, zoals bijvoorbeeld: *anderen*, *gansen*, *goeien*, *groten*, *halven*, *helen*, *kleinen*, *laatsten*, *nieuwen*, en *ouwen*.

Die uitgang *-en* is historisch gezien een accusatiefuitgang bij masculiene nomina (zie Van Loey 1964: 120). Mettertijd drong hij evenwel ook door in subjectpositie en werd daarbij geherinterpreteerd als een genusuitgang, een ontwikkeling die kaderde in het algemene proces van deflexie (Hermkens 1985, Nijen Twilhaar 1992). Naast adjectieven komt hij dan ook voor bij tal van andere adnominale elementen. In het CGN zijn dat het bepaald lidwoord *den*, de onbepaalde lidwoorden *ne*, *nen*, *ene*, en *enen* (waarbij de eerste twee fonetische reducties zijn van de laatste twee), en dan nog een hele rits van determinatorische voornaamwoorden, waarvan de meeste trouwens ook als zelfstandig gebruikt kunnen voorkomen: het proximaal aanwijzende *dezen*, de distaal demonstrativa *dieje*, *diejen*, *dien*, *diene*, en *dienen*, de onbepaalde voornaamwoorden *elken*, *ene*, *enen*, *gene*, *genen*, en *ieder*, de possessiva *hare*, *haren*, *hunne*, *hunnen*, *mijne* en *mijnen* samen met hun gereduceerde tegenhangers *m'ne* en *m'nen*, voorts *onzen*, *uwe*, *uwen*, *zijne* en *zijnen* ook weer met hun reducties *z'ne* en *z'nen*, en ten slotte het vragende *welken*. Het optreden van de *-n* wordt daarbij overigens gereguleerd door de fonetische context (Taeldeman 1980, 1983): enkel als de eerstvolgende klank een vocaal is, een h, d, t, in mindere mate een b, en voor sommige dialecten ook een r (er is met andere woorden nogal wat variatie – zie De Schutter 1966, 1967 voor een bespreking ervan).

Veel sporadischer zijn de typisch dialectische vormen die aangetroffen worden bij de bezittelijke voornaamwoorden. De Brabantse varianten zijn *(h)ulle*, en *(h)ullen*; de (West-)Vlaamse zijn *(h)ulder*, *uldere*, *ulderen*, *(h)under*, *junder*, en *(h)underen*. De schaarsheid van de frequenties heeft daarbij te maken met het uitgangspunt van het CGN om in principe vooral het bovengewestelijke taalgebruik te samplen. Een ander

gevolg daarvan is dat vormen zoals *huldere*, *hulderen*, *hundere*, *undere*, en *underen* in het geheel ontbreken.

Een totaal andere vorm van adnominale variatie betreft dan weer *dees* en *ons* in constructies als *ons vader* en dergelijke. Dit is precies het omgekeerde wat we hierboven besproken hebben: waar het zonet ging om de aanwezigheid van een flexie-uitgang op plaatsen waar er standaardtalig gezien eigenlijk geen hoort te zijn, behelst dit juist de afwezigheid van flexie op posities waar het wel zou moeten. Als één van de typische verschillen tussen Noord- en Zuid-Nederlands is het een bekend verschijnsel. Een grondige studie ervan is Tummers (2005).

De tweede grote groep van verschijnselen vormen de persoonlijke voornaamwoorden. De bekendste zijn ongetwijfeld de aanspreekvormen. Daarvan markeert het CGN enkel de zogenaamde D-vormen, te weten: *-de*, of de stapelvormen *-degij* en zelfs *-dege*. Het zijn relicten van het Middelnederlandse enkelvoudspronomen *du* (de ongetagde G-vormen *ge* en *gij* gaan terug op het meervoudspronomen *ghi*; zie Van Loey 1964: 136-141). Met de opkomende standaardisatie is *du* verdwenen, om te worden vervangen door de Hollandse vormen *je* en *jij* (zie Berteloot 2003; de meest recente studie over de geschiedenis van de aanspreekvormen is Vermaas 2002).

De vormen *-dege* en *-degij* hebben daarbij als extra eigenaardigheid dat het stapelvormen zijn (in het CGN worden de *ge* en *gij* op de volgende regel geschreven). Hetzelfde vinden we bij *-ekik* (1.enk) en *-zezij* (3.enk & mv), hoewel het CGN alleen de eerste markeert. Die is volgens Van Loey (1964: 139) ontstaan naar analogie van *-degij*; over de tweede wordt niets vermeld.

Al deze vormen zijn enclitisch: ze treden enkel op in inversie, en hechten zich daarbij aan het finiete werkwoord (De Schutter 1989). Dat is ook het geval bij de bekende Zuid-Nederlandse subject-*'m* (3.enk.masc). Een uitzondering vormt *me* (1.mv). Het is weliswaar ontstaan in geïnverteerde positie – Van Loey (1964: 136) verklaart het vanuit een assimilatie van *we* met de slot-n, die werkwoorden in het meervoud nu eenmaal hebben – van waaruit het vervolgens echter ook in de niet-geïnverteerde positie doorgedrongen is.

Nadere beschouwing van de tot dusver behandelde vormen onthult nog een ander verschijnsel: reduplicatieve constructies zoals *ik gaan-ekik* of *ge zijt gij*, waar een variant van het niet-geïnverteerde pronomen na het finiete werkwoord nog eens herhaald wordt (Vandeweghe 2000). De vorm vóór de persoonsvorm kan fonetisch vol (*ik*; *gij*) dan wel gereduceerd (*'k*; *ge*) zijn, die erna is evenwel steevast een volle vorm – eventueel een stapelvorm; zulke 'retriplicatieve' constructies komen in het

CGN echter maar een handvol keren voor, dus nemen we ze niet apart op. Nuyts (1995) schrijft reduplicaties daarbij twee mogelijke functies toe: ofwel zijn ze van het 'informatie-structurele type', waarbij het dan gaat om het focaliseren van de referent van het pronomen; ofwel zijn ze van het 'empathische type', waarbij de spreker empathie tegenover de hoorder uitdrukt. Het CGN voorziet geen markering voor dit verschijnsel, aangezien het eerder syntactisch van aard is. Het is echter geenszins afwezig bij ongetagde pronomina zoals *hij*, *we* en *wij*, en *ze* en *zij*.

Dat 'm überhaupt gemarkeerd voorkomt, mag overigens een wonder heten, gezien de totale verwaarlozing van het pronomen in de protocollaire richtlijnen van het CGN. Iets analogoos is er met *ikke* aan de hand: ondanks het feit dat het CGN het officieel als standaardtalig beschouwt, komt het in werkelijkheid meermaals getagd voor. Over de zogenaamd 'emfatische' status van het pronomen zijn de meningen trouwens verdeeld, en dat vanwege de mogelijkheid van *ikke* om wel degelijk persoonsvormen te regeren (zie, bijvoorbeeld, de discussie tussen Hoeksema 2000 en Zwart 2000). Ook in het CGN komt *ikke* inderdaad dikwijls voor bij een vervoegd werkwoord, meer bepaald in niet-geïnverteerde positie, en dus voegen we het pronomen aan onze lijst van taalvariabelen toe.

Aan de andere kant is het misschien niet verrassend dat pronomina zoals 'm en *ikke* toch getagd kunnen worden aangetroffen. De vormenrijkdom bij de pronomina is ronduit indrukwekkend, getuige voorts de hele reeks dialectvormen. Het volledige overzicht is: *geille*, *gelle*, *gijle*, *gijlen*, *gijlie*, *gijlle*, *gijllie*, *gulder*, *gulle*, *gullie*, *gunder*, *hulder*, *hulle*, *ich*, *julle*, *junder*, *mich*, *ulder*, *ulle*, *ullie*, *under* (niet: *undere*), *wieder*, *wijle*, *wijlen*, *wijlie*, *wulder*, *wulle*, *zieder*, *ziender*, *zijle*, *zijlie*, *zijllie*, *zul*, *zulder*, *zulle*, *zun*, en *zunder*. Tot op zekere hoogte zijn dit louter spellingvarianten natuurlijk, maar verder weerspiegelt deze lijst vooral de nog immer vigerende dialectverscheidenheid van Vlaanderen.

De derde en laatste groep is er ten slotte één die eigenlijk niet gemarkeerd wordt. Zoals vermeld beschouwt het CGN, naast de aanspreekvormen *ge* en *gij*, namelijk ook de diminutief op *-ke* als een gewoon standaardtalig verschijnsel. Die diminutief op *-ke* was het oorspronkelijke diminueringsprocedé in het Middelnederlands (zie Van Loey 1964: 225-232 voor een gedetailleerde beschrijving). Door de standaardisering werd het evenwel geleidelijk aan verdrongen door de Hollandse diminutief op *-je* (een proces waarbij overigens allerhande subtiele pragmatische factoren zijn komen kijken – zie Bakema 1997). Bijgevolg willen we de diminutief op *-ke* dan ook in onze analyses opnemen.



### 3.2 De operationalisering van de taalvariabelen

Als taalvariabele beschouwen we zoals vanouds een set van verschillende vormelijke strategieën die (nagenoeg) dezelfde functie vervullen. Daarmee sluiten we aan bij het zogenaamde ‘profielgebaseerde’ model van taalvariatie zoals beschreven in Geeraerts *et al.* (1994). Variatie wordt er onderverdeeld in drie typen: semasiologische variatie, conceptueel onomasiologische variatie, en formeel onomasiologische variatie, wat respectievelijk neerkomt op variatie tussen de verschillende functies van een welbepaalde vorm, variatie tussen de functies onderling, en variatie tussen de vormelijke realisaties van een bepaalde talige functie. Bij dat laatste type, waaronder onze studie natuurlijk ressorteert, definieert elke aparte reeks vormvarianten een zogenaamd profiel.

Rest ons dan aan te geven op basis waarvan we de profielen zullen afbakenen: welke vormen zijn alternatieven van elkaar? Vanuit onze in hoofdstuk 1 geschetste vraagstelling naar de standaardisatie van Vlaanderen, gaan we daarbij primair te rade bij de taalpolitieke literatuur. Een taalvariabele vatten we met name op als een alternantie tussen (één of meer) standaard en substandaard varianten voor een welbepaalde niche in het taalsysteem.

Voor de adnominale en de diminutivische variabelen is dit triviaal; bij de pronominale variabelen is het daarentegen nodig om enkele verdere onderscheidingen te maken. Zo treden er bij de pronomina tal namelijk tal van spreektaalige verschijnselen op, die voor een onderzoek naar het gesproken Belgisch-Nederlands natuurlijk van belang zijn. Het gaat dan allereerst om fonetische reducties: ‘k voor *ik*, *je* en *ge* voor respectievelijk *jij* en *gij*, *–ie* voor *hij*, *ze* voor *zij*, *we* voor *wij*, *me* voor *mij*, en ‘m voor *hem* (voor een studie omtrent vocaalreducties, zie Kloots 2005). Het enclitische *–ie* is daarbij van bijzonder belang omdat het de meest directe variant vormt voor de subjectvorm ‘m (van de diachronie van *–ie* is evenwel niets zeker; zie Van Loey 1964: 141-142; de draad wordt weer opgenomen in Nieuwenhuijsen 2002 en Van Bree 2004). Een ander sjobbolet van het Belgisch-Nederlands is het gebruik van *ze* in objectpositie, en dat zowel bij 3.enk.fem als 3.mv. In het eerste geval is *ze* een variant voor *haar* (of beter *d’r* – zie de Rooij 1994); in het tweede voor *hen* en *hun*. Overigens vertonen die onderling nog interessante variatie (de Rooij 1990).

Voorts bergen de bezittelijke voornaamwoorden nog een typisch Zuid-Nederlands kenmerk in zich. Het betreft hier meer bepaald reflexivumconstructies van een (onverbogen) possessivum + ‘eigen’: *mijn eigen*, *uw eigen*, *zijn eigen*, *haar eigen*, *ons eigen*, *hun eigen*, en zelfs *je eigen*. Omdat het hier weer om een meer syntactisch verschijnsel gaat, zijn deze vormen niet substandaardtalig gemarkeerd. Toch nemen we ze op, zelfs in weerwil van hun lage frequenties.

Als laatste variabele nemen we nog de reciprookpronomina *elkaar* en *mekaar* op – maar niet *elkander*, dat te weinig voorkomt. Weliswaar is geen van beide varianten op zich substandaard, maar natuurlijk zijn het in de Nederlandse grammatica oude bekenden als het op stijlverschillen aankomt, en net daardoor relevant voor ons onderzoek (de meest recente studie is Hüning 2006).

Naast extra varianten zijn er echter ook vormen die niet in de analyses opgenomen zullen worden. Zo is bij *ene* en *enen* de ambiguïteit tussen lidwoord, telwoord, of onbepaald voornaamwoord te groot om nog relevant te zijn. De beslissing bestaat er bijgevolg in om deze twee vormen in het geheel buiten beschouwing te laten. Het voornaamste criterium om vormen uit te sluiten is evenwel van meer praktische aard: de schaarsheid van hun frequenties. Dat was al zo bij *elkander*; het is met name het geval bij de dialectische varianten, zowel de determinatorische, als de pronominale. Ook het interrogativum *welken* vervalt om die reden. Alle andere varianten kunnen geoperationaliseerd worden tot taalvariabelen, en de volgende drie paragrafen zullen uit de doeken doen hoe dat in concreto aangepakt is. Achtereenvolgens komen aan bod: de adnominale taalvariabelen, de diminutivische taalvariabele, en de pronominale taalvariabelen.

### 3.2.1 De adnominale taalvariabelen

Omdat de adnominale flexie volgens de literatuur afhangt van het nominaal genus, is het nodig om het nomen bij elke determinator te identificeren. Daarbij hebben we voor een pragmatische oplossing gekozen. We zijn namelijk gewoon op zoek gegaan naar het eerstvolgende nomen na de determinator. Dat hebben gedaan met behulp van een Pythonscript<sup>2</sup>.

Het algoritme verliep daarbij in de volgende drie stappen:

- Zoek binnen de ‘annotation unit’ het eerstvolgende token met als pos-tag N(...).
- Check of die pos-tag van de vorm N(soort,ev,basis,(zijd|genus),stan) is.
- Indien ja, neem het lemma bij het token op; indien neen, ga naar de volgende observatie.

Bij de tweede stap maken we gebruik van de woordsoortinformatie die het CGN meegeeft om na te gaan of het nomen potentieel relevant is. Nomina worden in het CGN namelijk getagd voor vijf features (zie Van Eynde 2004: 13-18): type (soort- of

---

<sup>2</sup> Alle Pythoncode is verzameld in het script “tsscript.py” dat staat op de meegeleverde CD-ROM.

eigenaam), getal (enkel- of meervoud), graad (basis of gediminueerd), genus (zijdig, of onzijdig), en naamval (standaard, genitief, of datief). Het zal duidelijk zijn dat enkel die nomina in aanmerking komen die voor elk van de vijf features de eerste waarde hebben. Enkele nomina worden in het CGN evenwel ondergespecificeerd met betrekking tot het feature genus, en krijgen dan de generische waarde 'genus'. Het betreft hier nomina zoals *riool*, *filter*, *soort* en dergelijke. Voor de volledigheid hebben we ze eveneens opgenomen.

Bij de standaardtalig verbogen determinatoren levert deze zoekopdracht een lijst van 14270 lemmata op<sup>3</sup>. Het genus van elk hebben we vervolgens opgezocht in de elektronische versie van de 14-de editie van Van Dale Groot Woordenboek van de Nederlandse Taal (vanaf nu afgekort als GvD14). Dat geeft de volgende cijfers:

Genus	types	tokens
zn.	3	4
de (m.)	3637	39967
de (m.&v.)	70	737
de (v.)	3019	38048
het	48	212
het & de (v.)	7	46
het & de	21	121
de	2003	30616
de & het	27	210
de (m.) & het	42	142
de (v.) & het	13	63
het & de (m.)	24	163
de (m./v.)	61	237
mv.	2	3

Nr nomina = AMBIGU:

types:	638
tokens:	21232

Nr nomina NT in GvD14:

types:	4655
tokens:	9147

De GvD14 hanteert met andere woorden 14 genuscategorieën. Vier daarvan zijn duidelijk hoofdcategorieën: *het*, *de (m.)*, *de (v.)*, en *de* zonder meer; de overige tien zijn meng- dan wel restcategorieën. Verder zijn er dan nog een aantal nomina waarvan het genus afhankelijk van de betekenis en dus ambigu is. Ten slotte zijn er ook enkele nomina die niet in de GvD14 voorkomen; het gaat hier veelal om samenstellingen en afleidingen.

---

<sup>3</sup> De genuslijst waarvan sprake is onder de naam "generata.txt" op de CD-ROM te vinden is.

De veertiendelige genusclassificatie uit de GvD14 moge een accuraat beeld schetsen van de genusvariatie in het Nederlandse taalgebied, voor een variationele studie als de onze is ze evenwel enigszins te granulair. We willen deze classificatie dan ook iets meer op onze doelstellingen afstemmen. Zo laten we de restcategorieën *zn.* en *mv.* bijvoorbeeld vallen. Daarnaast hebben we besloten om de mengcategorieën te reduceren tot een of andere hoofdcategorie. Daartoe hebben we van elk nomen uit zo een mengcategorie de treffers manueel gecontroleerd. De gemeenkunnige categorie *de (m./v.)* hebben we daarbij in zijn geheel tot de generische *de* gerekend. Voorts hebben we van de ambigue nomina het genus gekozen dat intuïtief het meest prototypische is. De samenstellingen en afleidingen hebben het genus van het rechterlid, respectievelijk het suffix toegewezen gekregen.

Het volgende overzicht geeft de cijfers van de determinatoren met een substandaardtalige flexie-uitgang:

Genus	types	tokens
<i>zn.</i>	0	0
<i>de (m.)</i>	1123	7623
<i>de (m.&amp;v.)</i>	19	99
<i>de (v.)</i>	119	235
<i>het</i>	21	69
<i>het &amp; de (v.)</i>	0	0
<i>het &amp; de</i>	2	2
<i>de</i>	197	688
<i>de &amp; het</i>	2	4
<i>de (m.) &amp; het</i>	16	35
<i>de (v.) &amp; het</i>	3	3
<i>het &amp; de (m.)</i>	6	11
<i>de (m./v.)</i>	11	22
<i>mv.</i>	2	2

Nr nomina = AMBIGU:

types:	228
tokens:	2965

Nr nomina NT in GvD14:

types:	582
tokens:	966

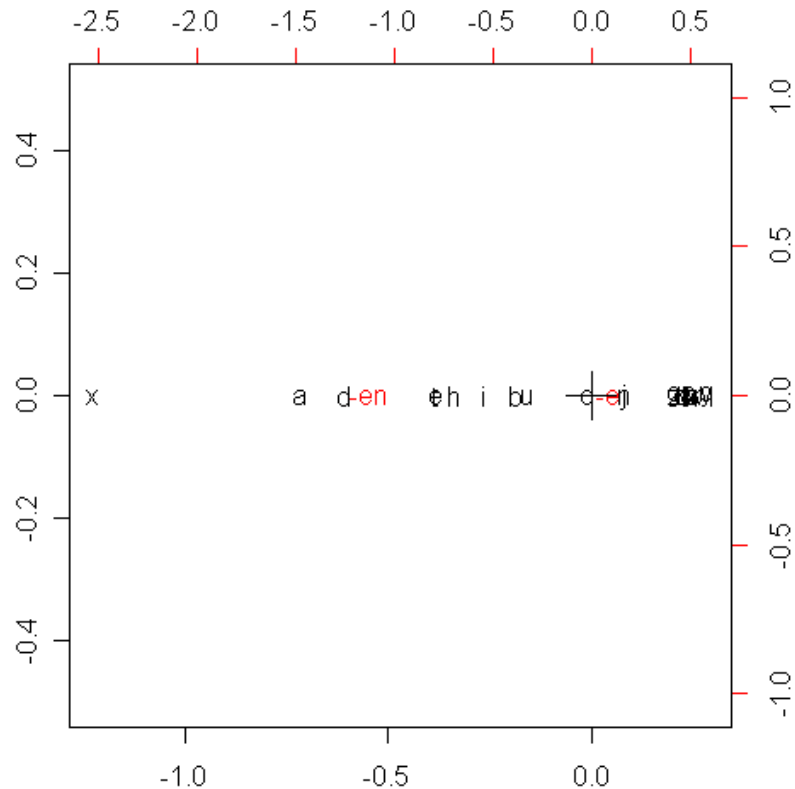
Opvallend zijn daarbij de frequenties bij de niet-masculiene nomina. Voor categorie *de* is dat nog enigszins begrijpelijk, gezien de bekende neiging tot masculinisering, die nu eenmaal ook in Vlaanderen werkzaam is (zie Geerts 1988). De scores voor *de (v.)* en *het* bewijzen echter dat die masculiniseringstendens inmiddels verder gevorderd is dan doorgaans wordt verondersteld. In het licht hiervan hebben we dan ook besloten om elke adnominale taalvariabele uit te splitsen per genuscategorie – en dat zijn er na de simplificatie waarvan hierboven sprake dus vier. Het totaal aantal adnominale taalvariabelen wordt zo met andere woorden verviervoudigd.

Rest er nog de kwestie van de verbindings-n, die volgens de literatuur dus enkel optreedt voor een vocaal, h, d, t, minder vaak voor een b, en nog minder voor een r. Om die stelling op zijn waarde te toetsen, hebben we het aantal voorkomens van de eerstvolgende klank na de determinator geteld. Dat is gebeurd voor de determinatoren op sjwa enerzijds (dat zijn dus *de, ne, gene, elke, iedere, deze, dieje, diene, mijne, m'ne, uwe, zijne, z'ne, hare, onze, hunne*, en de adjectieven op *-e*), en die op sjwa plus n anderzijds (te weten: *den, nen, genen, elken, iederen, dezen, diejen, dien, dienen, mijnen, m'nen, uwen, zijnen, z'nen, haren, onzen, hunnen*, en de adjectieven op *-en*). De volgende tabel geeft de frequenties:

	<b>-e</b>	<b>-en</b>
a	3610	1299
b	6256	886
c	2462	45
d	4498	1393
e	2958	634
f	2367	33
g	4246	88
h	3570	703
i	1630	277
j	860	48
k	8128	87
l	3423	41
m	6919	93
n	2163	125
o	3477	291
p	4474	51
q	20	0
r	5072	28
s	6506	59
t	4965	1073
u	1131	151
v	6955	99
w	4568	37
x	6	4
y	13	0
z	2583	56

Tabel 3.1: Frequenties van opeenvolgende klanken

De associaties tussen rijen en kolommen (berekend door middel van correspondentieanalyse, zie hoofdstuk 4) zijn dan als volgt:



Figuur 3.1: Associaties van opeenvolgende klanken

De positie van de *x* is niet relevant: het heeft te maken met het feit dat het CGN onverstaanbare uitingen transcribeert als *xxx*. Verder is het duidelijk dat de stelling klopt: de *-n* treedt merkbaar frequenter op bij de vocalen, *b*, *d*, *t*, en *h* (maar niet *r*), en onderling is er heel wat variatie. Dat leidt ons ertoe om met de fonetische context eveneens rekening te houden. Daarbij houden we echter wel onze primaire onderzoeksvraag naar de standaardisering van het Soapvlaams voor ogen: varianten zijn maar relevant voor zover ze ook (kunnen) concurreren met Journaalnederlandse vormen. Concreet betekent dat dan dat de fonetische context maar van belang is voor zes determinatoren, met name: *de*, *deze*, *elke*, *iedere*, *onze*, en de adjectieven. Bij alle andere is zowel de uitgang op *-e* als op *-en* daarentegen Soapvlaams (tegenover een Journaalnederlandse onverbogen variant), en doet het er dus eigenlijk niet zoveel meer toe welke fonotactische restricties er precies rusten op het optreden van een eufonische verbindingsklank die de *-n* toch is. De verbogen varianten bij deze variabelen zullen we dan ook cumuleren tot één waarde.

We eindigen deze paragraaf met een overzicht van de operationalisering van alle adnominale taalvariabelen. Daarbij zullen we de in Geeraerts *et al.* (1999) gehanteerde conventie volgen om de namen van de variabelen in KLEINKAPITAAL weer te geven, en de namen van de varianten in *cursief*. Omdat grammaticale variabelen zoals de onze evenwel niet meteen een voor de hand liggende naam hebben, kiezen we veelal de meest prototypische, standaardtalige variant (gemakshalve introduceren we ook een numerieke codering).

## 1. DE (MASCULIEN)

1.1 *de.m*                      Voorbeeld : *de hond*

Regex : `^de\tLID\(\bep,stan,rest\)`  
Genus : de (m.)  
Filter : REGEX + `\t.+n[aeiouhbdt]`

1.2 *den.m*                      Voorbeeld : *den hond*

Regex : `^den\*d\tLID\(\bep,dial\)`  
Genus : de (m.)

## 2. DE (COMMUUN)

2.1 *de.c*                      Voorbeeld : *de helft*

Regex : `^de\tLID\(\bep,stan,rest\)`  
Genus : de  
Filter : REGEX + `\t.+n[aeiouhbdt]`

2.2 *den.c*                      Voorbeeld : *den helft*

Regex : `^den\*d\tLID\(\bep,dial\)`  
Genus : de

## 3. DE (FEMINIEN)

3.1 *de.f*                      Voorbeeld : *de tv*

Regex : `^de\tLID\(\bep,stan,rest\)`  
Genus : de (v.)  
Filter : REGEX + `\t.+n[aeiouhbdt]`

3.2 *den.f*                      Voorbeeld : *den tv*

Regex : `^den\*d\tLID\(\bep,dial\)`  
Genus : de (v.)

## 4. EEN (MASCULIEN)

4.1 *een.m*                      Voorbeeld :    *een man, een hond*

Regex : `^een\tLID\(\onbep,stan,agr\)`  
Genus : de (m.)

4.2 *ne.n.m*                      Voorbeeld :    *ne man, nen hond*

Regex : `^nen?*\d\tLID\(\onbep,dial\)`  
Genus : de (m.)  
Filter : NOT ( REGEX + `\tne\nkeer\t` )

## 5. EEN (COMMUUN)

5.1 *een.c*                        Voorbeeld :    *een foto, een helft*

Regex : `^een\tLID\(\)`  
Genus : de

5.2 *ne.n.c*                        Voorbeeld :    *ne foto, nen helft*

Regex : `^nen?*\d\tLID\(\onbep,dial\)`  
Genus : de  
Filter : NOT ( REGEX + `\tne\nkeer\t` )

## 6. EEN (FEMINIEN)

6.1 *een.f*                        Voorbeeld :    *een kopie, een tv*

Regex : `^een\tLID\(\onbep,stan,agr\)`  
Genus : de (v.)

6.2 *ne.n.f*                        Voorbeeld :    *ne kopie, nen tv*

Regex : `^nen?*\d\tLID\(\onbep,dial\)`  
Genus : de (v.)  
Filter : NOT ( REGEX + `\tne\nkeer\t` )

## 7. GEEN (MASCULIEN)

7.1 *geen.m*                      Voorbeeld :    *geen man, geen hond*

Regex : `^geen\tVNW\(\onbep,det,stan,prenom,zonder,agr\)`  
Genus : de (m.)

7.2 *gene.n.m*                      Voorbeeld :    *gene man, genen hond*

Regex : `^genen?*\d\tVNW\(\onbep,det,dial\)`



Genus : de (m.)

## 8. GEEN (COMMUUN)

8.1 *geen.c*                      Voorbeeld : geen *foto*, geen *helft*

Regex : `^geen\tVNW\(\onbep,det,stan,preom,zonder,agr\)`

Genus : de

8.2 *gene.n.c*                      Voorbeeld : gene *foto*, genen *helft*

Regex : `^genen?\*d\tVNW\(\onbep,det,dial\)`

Genus : de

## 9. GEEN (FEMINIEN)

9.1 *geen.f*                      Voorbeeld : geen *kopie*, geen *tv*

Regex : `^geen\tVNW\(\onbep,det,stan,preom,zonder,agr\)`

Genus : de (v.)

9.2 *gene.n.f*                      Voorbeeld : gene *kopie*, genen *tv*

Regex : `^genen?\*d\tVNW\(\onbep,det,dial\)`

Genus : de (v.)

## 10. ELKE (MASCULIEN)

10.1 *elke.m*                      Voorbeeld : elke *hond*

Regex : `^elke\tVNW\(\onbep,det,stan,preom,met-e,evz\)`

Genus : de (m.)

Filter : REGEX + `\t.+ \n[aeiouhbd]`

10.2 *elken.m*                      Voorbeeld : elken *hond*

Regex : `^elken\*d\tVNW\(\onbep,det,dial\)`

Genus : de (m.)

## 11. IEDERE (MASCULIEN)

11.1 *iedere.m*                      Voorbeeld : iedere *hond*

Regex : `^iedere\tVNW\(\onbep,det,stan,preom,met-e,evz\)`

Genus : de (m.)

Filter : REGEX + `\t.+ \n[aeiouhbd]`

11.2 *iedereen.m*                      Voorbeeld : *iedereen hond*

Regex : `^iedereen*d\tVNW\(\onbep,det,dial\)`  
Genus : de (m.)

## 12. IEDERE (COMMUUN)

12.1 *iedere.c*                              Voorbeeld : *iedere helpt*

Regex : `^iedere\tVNW\(\onbep,det,stan,prenom,met-e,evz\)`  
Genus : de  
Filter : REGEX + `\t.+n[aeiouhbd]`

12.2 *iedereen.c*                              Voorbeeld : *iedereen helpt*

Regex : `^iedereen*d\tVNW\(\onbep,det,dial\)`  
Genus : de

## 13. DEZE (MASCULIEN)

13.1 *deze.m*                                      Voorbeeld : *deze hond*

Regex : `^deze\tVNW\(\aanw,det,stan,prenom,met-e,rest\)`  
Genus : de (m.)  
Filter : REGEX + `\t.+n[aeiouhbd]`

13.2 *dezen.m*                                      Voorbeeld : *dezen hond*

Regex : `^dezen*d\tVNW\(\aanw,det,dial\)`  
Genus : de (m.)

## 14. DEZE (COMMUUN)

14.1 *deze.c*                                      Voorbeeld : *deze helpt*

Regex : `^deze\tVNW\(\aanw,det,stan,prenom,met-e,rest\)`  
Genus : de  
Filter : REGEX + `\t.+n[aeiouhbd]`

14.2 *dezen.c*                                      Voorbeeld : *dezen helpt*

Regex : `^dezen*d\tVNW\(\aanw,det,dial\)`  
Genus : de

## 15. DEZE (FEMINIEN)

15.1 *deze.f* Voorbeeld : *deze tv*Regex : `^deze\tVNW\(\aanw,det,stan,prenom,met-e,rest\)`Genus : de (v.)Filter : REGEX + `\t.\n[aeiouhbd]`15.2 *dezen.f* Voorbeeld : *dezen tv*Regex : `^dezen\*d\tVNW\(\aanw,det,dial\)`Genus : de (v.)

## 16. DIE (MASCULIEN)

16.1 *die.m* Voorbeeld : *die man, die hond*Regex : `^die\tVNW\(\aanw,det,stan,prenom,zonder,rest\)`Genus<sup>4</sup> : de (m.)16.2 *dieje.n.m* Voorbeeld : *dieje man, diejen hond*Regex : `^diejen?\*d\tVNW\(\aanw,det,dial\)`Genus : de (m.)16.3 *diene.n.m* Voorbeeld : *diene man, dienen hond*Regex : `^diene?n?\*d\tVNW\(\aanw,det,dial\)`Genus : de (m.)

## 17. DIE (COMMUUN)

17.1 *die.c* Voorbeeld : *die foto, die helpt*Regex : `^die\tVNW\(\aanw,det,stan,prenom,zonder,rest\)`Genus : de17.2 *dieje.n.c* Voorbeeld : *dieje foto, diejen helpt*Regex : `^diejen?\*d\tVNW\(\aanw,det,dial\)`Genus : de17.3 *diene.n.c* Voorbeeld : *diene foto, dienen helpt*Regex : `^diene?n?\*d\tVNW\(\aanw,det,dial\)`Genus : de

---

<sup>4</sup> In één workshop (die-a-2) is er nog een extra filter, omdat er aanvankelijk iets misgelopen was – zie tsscript.py

## 18. DIE (FEMINIEN)

18.1 *die.f* Voorbeeld : die *kopie*, die *tv*

Regex : `^die\tVNW\(\aanw,det,stan,prenom,zonder,rest\)`  
Genus : de (v.)

18.2 *dieje.n.f* Voorbeeld : dieje *kopie*, diejen *tv*

Regex : `^diejen?*\d\tVNW\(\aanw,det,dial\)`  
Genus : de (v.)

18.3 *diene.n.f* Voorbeeld : diene *kopie*, dienen *tv*

Regex : `^diene?n?*\d\tVNW\(\aanw,det,dial\)`  
Genus : de (v.)

## 19. MIJN (MASCULIEN)

19.1 *mijn.m* Voorbeeld : mijn *man*, mijn *hond*

Regex : `^mijn\tVNW\(\bez,det,stan,vol,1,ev,prenom,zonder,agr\)`  
Filter : NOT ( REGEX + `\t.+\\neigen\t` )  
Genus : de (m.)

19.2 *mijne.n.m* Voorbeeld : mijne *man*, mijnen *hond*

Regex : `^mijnen?*\d\tVNW\(\bez,det,dial\)`  
Genus : de (m.)

19.3 *m.n.m* Voorbeeld : m'n *man*, m'n *hond*

Regex : `^m'n\tVNW\(\bez,det,stan,red,1,ev,prenom,zonder,agr\)`  
Filter : NOT ( REGEX + `\t.+\\neigen\t` )  
Genus : de (m.)

19.4 *m.ne.n.m* Voorbeeld : m'ne *man*, m'nen *hond*

Regex : `^m'nen?*\d\tVNW\(\bez,det,dial\)`  
Genus : de (m.)

## 20. MIJN (COMMUUN)

20.1 *mijn.c* Voorbeeld : mijn *foto*, mijn *helft*

Regex : `^mijn\tVNW\(\bez,det,stan,vol,1,ev,prenom,zonder,agr\)`  
Filter : NOT ( REGEX + `\t.+\\neigen\t` )  
Genus : de

20.2 *mijne.n.c* Voorbeeld : mijne *foto*, mijnen *helft*

Regex : `^mijnen?*d\|t\|V\|N\|W\|(bez, det, dial\)`  
Genus : de

20.3 *m.n.c*                      Voorbeeld :    *m'n foto, m'n helft*

Regex : `^m'n\|t\|V\|N\|W\|(bez, det, stan, red, 1, ev, prenom, zonder, agr\)`  
Filter : NOT ( REGEX + `\|t.\|neigen\|t` )  
Genus : de

20.4 *m.ne.n.c*                      Voorbeeld :    *m'ne foto, m'nen helft*

Regex : `^m'nen?*d\|t\|V\|N\|W\|(bez, det, dial\)`  
Genus : de

## 21. MIJN (FEMINIEN)

21.1 *mijn.f*                      Voorbeeld :    *mijn kopie, mijn tv*

Regex : `^mijn\|t\|V\|N\|W\|(bez, det, stan, vol, 1, ev, prenom, zonder, agr\)`  
Filter : NOT ( REGEX + `\|t.\|neigen\|t` )  
Genus : de (v.)

21.2 *mijne.n.f*                      Voorbeeld :    *mijne kopie, mijnen tv*

Regex : `^mijnen?*d\|t\|V\|N\|W\|(bez, det, dial\)`  
Genus : de (v.)

21.3 *m.n.f*                      Voorbeeld :    *m'n kopie, m'n tv*

Regex : `^m'n\|t\|V\|N\|W\|(bez, det, stan, red, 1, ev, prenom, zonder, agr\)`  
Filter : NOT ( REGEX + `\|t.\|neigen\|t` )  
Genus : de (v.)

21.4 *m.ne.n.f*                      Voorbeeld :    *m'ne kopie, m'nen tv*

Regex : `^m'nen?*d\|t\|V\|N\|W\|(bez, det, dial\)`  
Genus : de (v.)

## 22. JOUW (MASCULIEN)

22.1 *je.m*                      Voorbeeld :    *je man, je hond*

Regex : `^je\|t\|V\|N\|W\|(bez, det, stan, red, 2v, ev, prenom, zonder, agr\)`  
Filter : NOT ( REGEX + `\|t.\|neigen\|t` )  
Genus : de (m.)

22.2 *jouw.m*                      Voorbeeld :    *jouw man, jouw hond*

Regex : `^jouw\|t\|V\|N\|W\|(bez, det, stan, vol, 2v, ev, prenom, zonder, agr\)`  
Genus : de (m.)

- 22.3 *jullie.m*                      Voorbeeld :    *jullie man, jullie hond*
- Regex :    `^jullie\tVNW\(\bez, det, stan, nadr, 2v, mv, prenom, zonder, agr\)`  
Genus :    de (m.)
- 22.4 *uw.m*                              Voorbeeld :    *uw man, uw hond*
- Regex :    `^uw\tVNW\(\bez, det, stan, vol, 2, getal, prenom, zonder, agr\)`  
Filter :    NOT ( REGEX + `\t.+ \neigen\t` )  
Genus :    de (m.)
- 22.5 *uwe.n.m*                          Voorbeeld :    *uwe man, uwen hond*
- Regex :    `^uwen? \*d\tVNW\(\bez, det, dial\)`  
Genus :    de (m.)

## 23. JOUW (COMMUUN)

- 23.1 *je.c*                                  Voorbeeld :    *je foto, je helpt*
- Regex :    `^je\tVNW\(\bez, det, stan, red, 2v, ev, prenom, zonder, agr\)`  
Filter :    NOT ( REGEX + `\t.+ \neigen\t` )  
Genus :    de
- 23.2 *jouw.c*                              Voorbeeld :    *jouw foto, jouw helpt*
- Regex :    `^jouw\tVNW\(\bez, det, stan, vol, 2v, ev, prenom, zonder, agr\)`  
Genus :    de
- 23.3 *jullie.c*                              Voorbeeld :    *jullie foto, jullie helpt*
- Regex :    `^jullie\tVNW\(\bez, det, stan, nadr, 2v, mv, prenom, zonder, agr\)`  
Genus :    de
- 23.4 *uw.c*                                  Voorbeeld :    *uw foto, uw helpt*
- Regex :    `^uw\tVNW\(\bez, det, stan, vol, 2, getal, prenom, zonder, agr\)`  
Filter :    NOT ( REGEX + `\t.+ \neigen\t` )  
Genus :    de
- 23.5 *uwe.n.c*                          Voorbeeld :    *uwe foto, uwen helpt*
- Regex :    `^uwen? \*d\tVNW\(\bez, det, dial\)`  
Genus :    de

## 24. JOUW (FEMINIEN)

- 24.1 *je.f*                                  Voorbeeld :    *je kopie, je tv*
- Regex :    `^je\tVNW\(\bez, det, stan, red, 2v, ev, prenom, zonder, agr\)`  
Filter :    NOT ( REGEX + `\t.+ \neigen\t` )

Genus : de (v.)

24.2 *jouw.f*                      Voorbeeld :    *jouw kopie, jouw tv*

Regex : `^jouw\tVNW\(\bez, det, stan, vol, 2v, ev, prenom, zonder, agr\)`

Genus : de (v.)

24.3 *jullie.f*                      Voorbeeld :    *jullie kopie, jullie tv*

Regex : `^jullie\tVNW\(\bez, det, stan, nadr, 2v, mv, prenom, zonder, agr\)`

Genus : de (v.)

24.4 *uw.f*                            Voorbeeld :    *uw kopie, uw tv*

Regex : `^uw\tVNW\(\bez, det, stan, vol, 2, getal, prenom, zonder, agr\)`

Filter : NOT ( REGEX + `\t.+ \neigen\t` )

Genus : de (v.)

24.5 *uwe.n.f*                      Voorbeeld :    *uwe kopie, uwen tv*

Regex : `^uwen? \*d\tVNW\(\bez, det, dial\)`

Genus : de (v.)

## 25. ZIJN (MASCULIEN)

25.1 *zijn.m*                        Voorbeeld :    *zijn man, zijn hond*

Regex : `^zijn\tVNW\(\bez, det, stan, vol, 3, ev, prenom, zonder, agr\)`

Filter : NOT ( REGEX + `\t.+ \neigen\t` )

Genus : de (m.)

25.2 *zijne.n.m*                      Voorbeeld :    *zijne man, zijnen hond*

Regex : `^zijnen? \*d\tVNW\(\bez, det, dial\)`

Genus : de (m.)

25.3 *z.n.m*                         Voorbeeld :    *z'n man, z'n hond*

Regex : `^z'n\tVNW\(\bez, det, stan, red, 3, ev, prenom, zonder, agr\)`

Filter : NOT ( REGEX + `\t.+ \neigen\t` )

Genus : de (m.)

25.4 *z.ne.n.m*                      Voorbeeld :    *z'ne man, z'nen hond*

Regex : `^z'nen? \*d\tVNW\(\bez, det, dial\)`

Genus : de (m.)

## 26. ZIJN (COMMUUN)

26.1 *zijn.c* Voorbeeld : *zijn foto, zijn helft*

Regex : `^zijn\tVNW\(\bez, det, stan, vol, 3, ev, prenom, zonder, agr\)`  
Filter : NOT ( REGEX + `\t.+ \neigen\t` )  
Genus : de

26.2 *zijne.n.c* Voorbeeld : *zijne foto, zijnen helft*

Regex : `^zijnen? \*d\tVNW\(\bez, det, dial\)`  
Genus : de

26.3 *z.n.c* Voorbeeld : *z'n foto, z'n helft*

Regex : `^z'n\tVNW\(\bez, det, stan, red, 3, ev, prenom, zonder, agr\)`  
Filter : NOT ( REGEX + `\t.+ \neigen\t` )  
Genus : de

26.4 *z.ne.n.c* Voorbeeld : *z'ne foto, z'nen helft*

Regex : `^z'nen? \*d\tVNW\(\bez, det, dial\)`  
Genus : de

## 27. ZIJN (FEMINIEN)

27.1 *zijn.f* Voorbeeld : *zijn kopie, zijn tv*

Regex : `^zijn\tVNW\(\bez, det, stan, vol, 3, ev, prenom, zonder, agr\)`  
Filter : NOT ( REGEX + `\t.+ \neigen\t` )  
Genus : de (v.)

27.2 *zijne.n.f* Voorbeeld : *zijne kopie, zijnen tv*

Regex : `^zijnen? \*d\tVNW\(\bez, det, dial\)`  
Genus : de (v.)

27.3 *z.n.f* Voorbeeld : *z'n kopie, z'n tv*

Regex : `^z'n\tVNW\(\bez, det, stan, red, 3, ev, prenom, zonder, agr\)`  
Filter : NOT ( REGEX + `\t.+ \neigen\t` )  
Genus : de (v.)

## 28. HAAR (MASCULIEN)

28.1 *haar.m* Voorbeeld : *haar man, haar hond*

Regex : `^haar\tVNW\(\bez, det, stan, vol, 3, ev, prenom, zonder, agr\)`  
Filter : NOT ( REGEX + `\t.+ \neigen\t` )  
Genus : de (m.)



28.2 *hare.n.m* Voorbeeld : *hare man, haren hond*

Regex : `^haren?*d\TVNW\(\bez,det,dial\)`  
Genus : de (m.)

## 29. HAAR (COMMUN)

29.1 *haar.c* Voorbeeld : *haar foto, haar helpt*

Regex : `^haar\TVNW\(\bez,det,stan,vol,3,ev,prenom,zonder,agr\)`  
Filter : NOT ( REGEX + `\t.+neigen\t` )  
Genus : de

29.2 *hare.n.c* Voorbeeld : *hare foto, haren helpt*

Regex : `^haren?*d\TVNW\(\bez,det,dial\)`  
Genus : de

## 30. HAAR (FEMINIEN)

30.1 *haar.f* Voorbeeld : *haar kopie, haar tv*

Regex : `^haar\TVNW\(\bez,det,stan,vol,3,ev,prenom,zonder,agr\)`  
Filter : NOT ( REGEX + `\t.+neigen\t` )  
Genus : de (v.)

30.2 *hare.n.f* Voorbeeld : *hare kopie, haren tv*

Regex : `^haren?*d\TVNW\(\bez,det,dial\)`  
Genus : de (v.)

## 31. ONZE (MASCULIEN)

31.1 *onze.m* Voorbeeld : *onze hond*

Regex : `^onze\TVNW\(\bez,det,stan,vol,1,mv,prenom,met-e,rest\)`  
Genus : de (m.)  
Filter : REGEX + `\t.+n[aeiouhbd]`

31.2 *onzen.m* Voorbeeld : *onzen hond*

Regex : `^onzen*d\TVNW\(\bez,det,dial\)`  
Genus : de (m.)



Regex : ^hunnen?\*d\|t\|VNW\|(bez, det, dial\|)  
Genus : de

## 36. HUN (FEMINIEN)

36.1 *hun.f* Voorbeeld : *hun kopie, hun tv*

Regex : ^hun\|t\|VNW\|(bez, det, stan, vol, 3, mv, prenom, zonder, agr\|)  
Filter : NOT ( REGEX + \|t.\|neigen\|t )  
Genus : de (v.)

36.2 *hunne.n.f* Voorbeeld : *hunne kopie, hunnen tv*

Regex : ^hunnen?\*d\|t\|VNW\|(bez, det, dial\|)  
Genus : de (v.)

## 37. ADJECTIEF (MASCULIEN)

37.1 *adj.e.m* Voorbeeld : *een mooie hond*

Regex : ^\|S\|e\|t\|ADJ\|(prenom, (basis\|comp\|sup), met-e, stan\|)  
Genus : de (m.)  
Filter : REGEX + \|t.\|n\|[aeiouhbd]

37.2 *adj.n.m* Voorbeeld : *ne mooien hond*

Regex : ^\|S\|en\|^\*d\|t\|ADJ\|(dial\|)  
Genus : de (m.)

## 38. ADJECTIEF (COMMUUN)

38.1 *adj.e.c* Voorbeeld : *een slechte helft*

Regex : ^\|S\|e\|t\|ADJ\|(prenom, (basis\|comp\|sup), met-e, stan\|)  
Genus : de  
Filter : REGEX + \|t.\|n\|[aeiouhbd]

38.2 *adj.n.c* Voorbeeld : *ne slechten helft*

Regex : ^\|S\|en\|^\*d\|t\|ADJ\|(dial\|)  
Genus : de

## 39. ADJECTIEF (FEMINIEN)

39.1 *adj.e.f* Voorbeeld : *een grote tv*

Regex : ^\|S\|e\|t\|ADJ\|(prenom, (basis\|comp\|sup), met-e, stan\|)

Genus : de (v.)  
Filter : REGEX + \t.+n[aeiouhbd]

39.2 *adj.n.f*                      Voorbeeld : *ne* groten *tv*

Regex : ^\S+en\\*d\tADJ\dial\  
Genus : de (v.)

Van *elken* komt er geen enkele treffer voor met een feminien of 'commuun' substantief, en dus vervallen de corresponderende variabelen. Hetzelfde geldt voor zowel *ieder* als *z'ne(n)* bij feminiene nomina.

### 3.2.2 De diminutivische taalvariabele

De diminutiefvariatie coderen we als een eenvoudige dichotome variabele: soortnamen eindigend op *-je(s)* tegenover soortnamen eindigend op *-ke(s)*. De variabele ziet er dus als volgt uit:

#### 40. DIMINUTIEF

40.1 *dim.j*                      Voorbeeld : *het* stoeltje, stoeltjes

Regex : ^\S+jes?\tN\soort,[em]v,dim[^\]\*\

40.2 *dim.k*                      Voorbeeld : *het* stoeleke, stoelekes

Regex : ^\S+kes?\tN\soort,[em]v,dim[^\]\*\

### 3.2.3 De pronominale taalvariabelen

Bij de pronomina is het voornaamste onderscheid dat we moeten maken dat tussen geïnverteerde en niet-geïnverteerde subjectvormen enerzijds, en anderzijds dat tussen reflexieve en niet-reflexieve objectvormen. Het eerste doen we automatisch; het tweede manueel. De reguliere expressie voor de niet-geïnverteerde pronomina is:

*variant* + \t.+n\S+\tWW\pv,

Die voor de geïnverteerde is:

\t(?:WW\pv,|VG\onder\)|VNW\((?:betr|vrag|vb),).+n + *variant*

Daarbij maken we twee uitzonderingen. De stapelvormen zijn altijd geïnverteerd; de reduplicaties nooit geïnverteerd. Die reduplicaties beantwoorden aan de volgende regex:

`^ + variant1 + \t.+ \n + variant2`

Verder hebben we ook een aantal vormen manueel gedesambigueerd. Dat is met name gebeurd voor *ze* en *jullie*, die voor het naamvalsfeature enkel de generische waarde 'standaard' krijgen.

Het grootste probleempunt vormde evenwel 'm. Door de veronachtzaming ervan in de CGN-protocollen komen de substandaardtalige subjectvormen zelfs voor bij het volle *hem*. De oplossing bestond er dan ook in om alle treffers van zowel *hem* als 'm manueel te controleren.

Zoals vermeld is manuele desambiguering eerder regel dan uitzondering bij de objectvormen. Voor *je*, *me*, *mij*, *ons* en *u* maakt het CGN namelijk geen onderscheid tussen het personale en het reflexivum. Bij *zich* wordt er dan weer niet gedifferentieerd tussen 3-de en 2-de persoon.

Daarmee zijn we beland bij de reflexiva van de vorm possessivum + 'eigen'. Die werden automatisch uit de treffers van de bezittelijke vormen gefilterd, en wel door middel van deze reguliere expressie:

`variant + \t.+ \neigen \t`

Voor de vormen van de 3-de persoon *zijn eigen*, *haar eigen*, en *hun eigen* hebben we weer het criterium toegepast dat een taalvariabele in ons onderzoek eerst en vooral het verschil in distributie tussen het Journaalnederlands en het Soapvlaams moet vatten. De drie varianten in kwestie zijn elk een substandaardtalig alternatief voor het standaardtalige *zich*, maar onderling wordt hun variatie bepaald door minder relevante factoren zoals sekse en getal van de referent(en). Bijgevolg hebben we deze vormen eveneens gecumuleerd, en wel tot het generische *eigen.3*.

Dit zijn tot slot de operationalisering van de pronominale variabelen:

#### 41. IK (NIET-GEÏNVERTEERD)

##### 41.1 *ik.not*

Voorbeeld : *ik kom mee*

**Regex** : `^ik\tVNW\(\(pers,pron,nomin,vol,1,ev\)`  
**Filter1** : NOT ( `^k\*d\tVNW\(\(pers,pron,dial\)\t'k\n + REGEX )`  
**Filter2** : NOT ( `\t(?:WW\(\(pv,|VG\(\(onder\)|VNW\(\(\(?:betr|vrag|vb)\),\)+\n + REGEX )`

Filter3 : NOT ( REGEX + \tik\n.+tWW\ (pv, .+ \n (? : ik | 'k (? : \\*d) ?) ) \t )

41.2 *k.not*                      Voorbeeld : 'k kom mee

Regex : ^k\tVNW\ (pers, pron, nomin, red, 1, ev\

Filter1 : NOT ( REGEX + \tik\nik\t )

Filter2 : NOT ( \t (? : WW\ (pv, | VG\ (onder\)) | VNW\ ((? : betr | vrag | vb), .) .+ \n + REGEX )

Filter3 : NOT ( REGEX + \tik\n.+tWW\ (pv, .+ \n (? : ik | 'k (? : \\*d) ?) ) \t )

41.3 *ikke*                      Voorbeeld : ikke kom mee

Regex : ^ikke\tVNW\ (pers, pron, nomin, nadr, 1, ev\

41.4 *ik.dbl*                      Voorbeeld : ik kom ik mee, ik kom (e)kik mee

Regex : ^ik\tVNW\ (pers, pron, nomin, vol, 1, ev\

Filter1 : NOT ( ^k\\*d\tVNW\ (pers, pron, dial\)\t'k'n + REGEX )

Filter2 : NOT ( \t (? : WW\ (pv, | VG\ (onder\)) | VNW\ ((? : betr | vrag | vb), .) .+ \n + REGEX )

Filter3 : REGEX + \tik\n.+tWW\ (pv, .+ \n (? : ik | 'k (? : \\*d) ?) ) \t

41.5 *k.dbl*                      Voorbeeld : 'k kom ik mee, 'k kom (e)kik mee

Regex : ^k\tVNW\ (pers, pron, nomin, red, 1, ev\

Filter1 : NOT ( REGEX + \tik\nik\t )

Filter2 : NOT ( \t (? : WW\ (pv, | VG\ (onder\)) | VNW\ ((? : betr | vrag | vb), .) .+ \n + REGEX )

Filter3 : REGEX + \tik\n.+tWW\ (pv, .+ \n (? : ik | 'k (? : \\*d) ?) ) \t

## 42. IK (GEÏNVERTEERD)

42.1 *ik.inv*                      Voorbeeld : kom ik mee?

Regex : ^ik\tVNW\ (pers, pron, nomin, vol, 1, ev\

Filter1 : NOT ( ^k\\*d\tVNW\ (pers, pron, dial\)\t'k'n + REGEX )

Filter2 : \t (? : WW\ (pv, | VG\ (onder\)) | VNW\ ((? : betr | vrag | vb), .) .+ \n + REGEX

Filter3 : NOT ( ^ (? : ik | 'k (? : \\*d) ?) \t .+ \n .+ WW\ (pv, .+ \n + REGEX )

42.2 *k.inv*                      Voorbeeld : kom 'k mee?

Regex : ^k\tVNW\ (pers, pron, nomin, red, 1, ev\

Filter1 : NOT ( ^k\\*d\tVNW\ (pers, pron, dial\)\t'k'n + REGEX )

Filter2 : \t (? : WW\ (pv, | VG\ (onder\)) | VNW\ ((? : betr | vrag | vb), .) .+ \n + REGEX

Filter3 : NOT ( ^ (? : ik | 'k (? : \\*d) ?) \t .+ \n .+ WW\ (pv, .+ \n + REGEX )

42.3 *k.ik*                      Voorbeeld : kom-(e)kik mee?

Regex : ^ik\tVNW\ (pers, pron, nomin, vol, 1, ev\

Filter1 : ^k\\*d\tVNW\ (pers, pron, dial\)\t'k'n + REGEX

Filter2 : NOT ( ^ (? : ik | 'k (? : \\*d) ?) \t .+ \n .+ WW\ (pv, .+ \n + REGEX )

## 43. JIJ (NIET-GEÏNVERTEERD)

43.1 *je.not* Voorbeeld : *je komt mee*

Regex : ^je\tVNW\(\pers,pron,nomin,red,2v,ev\  
Filter : NOT ( \t(?:WW\(\pv,|VG\(\onder)|VNW\((?:betr|vrag|vb),).+\n + REGEX )

43.2 *jij.not* Voorbeeld : *jij komt mee*

Regex : ^jij\tVNW\(\pers,pron,nomin,vol,2v,ev\  
Filter : NOT ( \t(?:WW\(\pv,|VG\(\onder)|VNW\((?:betr|vrag|vb),).+\n + REGEX )

43.3 *jullie.not* Voorbeeld : *jullie komen mee*

Regex : ^jullie\tVNW\(\pers,pron,stan,nadr,2v,mv\  
Manueel : subjectvorm – objectvorm  
Filter : NOT ( \t(?:WW\(\pv,|VG\(\onder)|VNW\((?:betr|vrag|vb),).+\n + REGEX )

43.4 *ge.not* Voorbeeld : *ge komt mee*

Regex : ^ge\tVNW\(\pers,pron,nomin,red,2,getal\  
Filter1 : NOT ( ^de\\*d\tVNW\(\pers,pron,dial)\tde\n + REGEX )  
Filter2 : NOT ( \t(?:WW\(\pv,|VG\(\onder)|VNW\((?:betr|vrag|vb),).+\n + REGEX )  
Filter3 : NOT ( REGEX + \tge\n.+tWW\(\pv,.\n(?:ge|gij|de\\*d)\t

43.5 *gij.not* Voorbeeld : *gij komt mee*

Regex : ^gij\tVNW\(\pers,pron,nomin,vol,2,getal\  
Filter1 : NOT ( ^de\\*d\tVNW\(\pers,pron,dial)\tde\n + REGEX )  
Filter2 : NOT ( \t(?:WW\(\pv,|VG\(\onder)|VNW\((?:betr|vrag|vb),).+\n + REGEX )  
Filter3 : NOT ( REGEX + \tgij\n.+tWW\(\pv,.\n(?:ge|gij|de\\*d)\t

43.6 *u.not* Voorbeeld : *u komt mee*

Regex : ^u\tVNW\(\pers,pron,nomin,vol,2b,getal\  
Filter : NOT ( \t(?:WW\(\pv,|VG\(\onder)|VNW\((?:betr|vrag|vb),).+\n + REGEX )

43.7 *ge.dbl* Voorbeeld : *ge komt gij mee, ge komt-degij mee*

Regex : ^ge\tVNW\(\pers,pron,nomin,red,2,getal\  
Filter1 : NOT ( ^de\\*d\tVNW\(\pers,pron,dial)\tde\n + REGEX )  
Filter2 : NOT ( \t(?:WW\(\pv,|VG\(\onder)|VNW\((?:betr|vrag|vb),).+\n + REGEX )  
Filter3 : REGEX + \tge\n.+tWW\(\pv,.\n(?:ge|gij|de\\*d)\t

43.8 *gij.dbl* Voorbeeld : *gij komt gij mee, gij komt-degij mee*

Regex : ^gij\tVNW\(\pers,pron,nomin,vol,2,getal\  
Filter1 : NOT ( ^de\\*d\tVNW\(\pers,pron,dial)\tde\n + REGEX )  
Filter2 : NOT ( \t(?:WW\(\pv,|VG\(\onder)|VNW\((?:betr|vrag|vb),).+\n + REGEX )  
Filter3 : REGEX + \tgij\n.+tWW\(\pv,.\n(?:ge|gij|de\\*d)\t

## 44. JIJ (GEÏNVERTEERD)

44.1 *je.inv* Voorbeeld : *kom je mee?*

Regex : `^je\tVNW\(\pers,pron,nomin,red,2v,ev\)`  
Filter : `\t(?:WW\(\pv,|VG\(\onder\)|VNW\((?:betr|vrag|vb),)+\n + REGEX`

44.2 *jij.inv* Voorbeeld : *kom jij mee?*

Regex : `^jij\tVNW\(\pers,pron,nomin,vol,2v,ev\)`  
Filter : `\t(?:WW\(\pv,|VG\(\onder\)|VNW\((?:betr|vrag|vb),)+\n + REGEX`

44.3 *jullie.inv* Voorbeeld : *komen jullie mee?*

Regex : `^jullie\tVNW\(\pers,pron,stan,nadr,2v,mv\)`  
Manueel : subjectvorm – objectvorm  
Filter : `\t(?:WW\(\pv,|VG\(\onder\)|VNW\((?:betr|vrag|vb),)+\n + REGEX`

44.4 *ge.inv* Voorbeeld : *komt ge mee?*

Regex : `^ge\tVNW\(\pers,pron,nomin,red,2,getal\)`  
Filter1 : `NOT ( ^de\*d\tVNW\(\pers,pron,dial\)\tde\n + REGEX )`  
Filter2 : `\t(?:WW\(\pv,|VG\(\onder\)|VNW\((?:betr|vrag|vb),)+\n + REGEX`  
Filter3 : `NOT ( ^(?:ge|gij)\t.+ \n.+ WW\(\pv,.\n + REGEX )`

44.5 *gij.inv* Voorbeeld : *komt gij mee?*

Regex : `^gij\tVNW\(\pers,pron,nomin,vol,2,getal\)`  
Filter1 : `NOT ( ^de\*d\tVNW\(\pers,pron,dial\)\tde\n + REGEX )`  
Filter2 : `\t(?:WW\(\pv,|VG\(\onder\)|VNW\((?:betr|vrag|vb),)+\n + REGEX`  
Filter3 : `NOT ( ^(?:ge|gij)\t.+ \n.+ WW\(\pv,.\n + REGEX )`

44.6 *u.inv* Voorbeeld : *komt u mee?*

Regex : `^u\tVNW\(\pers,pron,nomin,vol,2b,getal\)`  
Filter : `\t(?:WW\(\pv,|VG\(\onder\)|VNW\((?:betr|vrag|vb),)+\n + REGEX`

44.7 *de.inv* Voorbeeld : *komt-de mee?*

Regex : `^de\*d\tVNW\(\pers,pron,dial\)`  
Filter1 : `NOT ( \tde\ngij\t )`  
Filter2 : `NOT ( ^(?:ge|gij)\t.+ \n.+ WW\(\pv,.\n + REGEX )`

44.8 *de.gij* Voorbeeld : *komt-degij mee?*

Regex : `^de\*d\tVNW\(\pers,pron,dial\)`  
Filter1 : `REGEX + \tde\ngij\tVNW\(\pers,pron,nomin,vol,2,getal\)`  
Filter2 : `NOT ( ^(?:ge|gij)\t.+ \n.+ WW\(\pv,.\n + REGEX )`



## 45. HIJ (NIET-GEÏNVERTEERD)

45.1 *hij.not* Voorbeeld : *hij komt mee*

Regex : ^hij\tVNW\(\pers,pron,nomin,vol,3,ev,masc\  
Filter1 : NOT ( \t(?:WW\(\pv,|VG\(\onder)|VNW\((?:betr|vrag|vb),)+\n + REGEX )  
Filter2 : NOT ( REGEX + \thij\n.+tWW\(\pv,+\nhij\t )

45.2 *hij.dbl* Voorbeeld : *hij komt hij mee*

Regex : ^hij\tVNW\(\pers,pron,nomin,vol,3,ev,masc\  
Filter1 : NOT ( \t(?:WW\(\pv,|VG\(\onder)|VNW\((?:betr|vrag|vb),)+\n + REGEX )  
Filter2 : REGEX + \thij\n.+tWW\(\pv,+\nhij\t

## 46. HIJ (GEÏNVERTEERD)

46.1 *hij.inv* Voorbeeld : *komt hij mee?*

Regex : ^hij\tVNW\(\pers,pron,nomin,vol,3,ev,masc\  
Filter1 : \t(?:WW\(\pv,|VG\(\onder)|VNW\((?:betr|vrag|vb),)+\n + REGEX  
Filter2 : NOT ( ^hij\t.+n.+WW\(\pv,+\n + REGEX )

46.2 *ie* Voorbeeld : *komt-ie mee?*

Regex : ^ie\tVNW\(\pers,pron,nomin,red,3,ev,masc\

46.3 *m.inv* Voorbeeld : *komt-'m mee?*

Regex : ^m\\*d\tVNW\(\pers,pron,dial\  
Manueel : subjectvorm – objectvorm

## 47. ZIJ – FEM (NIET-GEÏNVERTEERD)

47.1 *ze.not.vr* Voorbeeld : *ze komt mee*

Regex : ^ze\tVNW\(\pers,pron,stan,red,3,ev,fem\  
Manueel : subjectvorm – objectvorm  
Filter1 : NOT ( REGEX + \tze\nzij\t )  
Filter2 : NOT ( \t(?:WW\(\pv,|VG\(\onder)|VNW\((?:betr|vrag|vb),)+\n + REGEX )  
Filter3 : NOT ( REGEX + \tze\n.+tWW\(\pv,+\n(?:ze|zij)\t )

47.2 *zij.not.vr* Voorbeeld : *zij komt mee*

Regex : ^zij\tVNW\(\pers,pron,nomin,vol,3v,ev,fem\  
Filter1 : NOT ( ^ze\t.+n + REGEX )  
Filter2 : NOT ( \t(?:WW\(\pv,|VG\(\onder)|VNW\((?:betr|vrag|vb),)+\n + REGEX )  
Filter3 : NOT ( REGEX + \tzij\n.+tWW\(\pv,+\n(?:ze|zij)\t )

47.3 *ze.dbl.vr* Voorbeeld : *ze komt zij mee*

Regex : ^ze\tVNW\(\pers,pron,stan,red,3,ev,fem\



Regex : `^we\tVNW\(\pers,pron,nomin,red,1,mv\)`  
Filter1 : NOT ( `\t(?:WW\(\pv,|VG\(\onder\)|VNW\((?:betr|vrag|vb),)+\n` + REGEX )  
Filter2 : REGEX + `\twe\n.+tWW\(\pv,+\n(?:we|wij|me\*d)\t`

49.5 *wij.dbl* Voorbeeld : *wij komen wij mee*

Regex : `^wij\tVNW\(\pers,pron,nomin,vol,1,mv\)`  
Filter1 : NOT ( `\t(?:WW\(\pv,|VG\(\onder\)|VNW\((?:betr|vrag|vb),)+\n` + REGEX )  
Filter2 : REGEX + `\twij\n.+tWW\(\pv,+\n(?:we|wij|me\*d)\t`

## 50. WIJ (GEÏNVERTEERD)

50.1 *we.inv* Voorbeeld : *komen we mee?*

Regex : `^we\tVNW\(\pers,pron,nomin,red,1,mv\)`  
Filter1 : `\t(?:WW\(\pv,|VG\(\onder\)|VNW\((?:betr|vrag|vb),)+\n` + REGEX  
Filter2 : NOT ( `^(?:we|wij|me\*d)\t.+ \n.+WW\(\pv,+\n` + REGEX )

50.2 *wij.inv* Voorbeeld : *komen wij mee?*

Regex : `^wij\tVNW\(\pers,pron,nomin,vol,1,mv\)`  
Filter1 : `\t(?:WW\(\pv,|VG\(\onder\)|VNW\((?:betr|vrag|vb),)+\n` + REGEX  
Filter2 : NOT ( `^(?:we|wij|me\*d)\t.+ \n.+WW\(\pv,+\n` + REGEX )

50.3 *me.inv* Voorbeeld : *komen me mee?*

Regex : `^me\*d\tVNW\(\pers,pron,dial\)`  
Filter1 : `\t(?:WW\(\pv,|VG\(\onder\)|VNW\((?:betr|vrag|vb),)+\n` + REGEX  
Filter2 : NOT ( `^(?:we|wij|me\*d)\t.+ \n.+WW\(\pv,+\n` + REGEX )

## 51. ZIJ – MV (NIET-GEÏNVERTEERD)

51.1 *ze.not.mv* Voorbeeld : *ze komen mee*

Regex : `^ze\tVNW\(\pers,pron,stan,red,3,mv\)`  
Manueel : subjectvorm – objectvorm  
Filter1 : NOT ( REGEX + `\tze\nzij\t` )  
Filter2 : NOT ( `\t(?:WW\(\pv,|VG\(\onder\)|VNW\((?:betr|vrag|vb),)+\n` + REGEX )  
Filter3 : NOT ( REGEX + `\tze\n.+tWW\(\pv,+\n(?:ze|zij)\t` )

51.2 *zij.not.mv* Voorbeeld : *zij komen mee*

Regex : `^zij\tVNW\(\pers,pron,nomin,vol,3p,mv\)`  
Filter1 : NOT ( `^ze\t.+ \n` + REGEX )  
Filter2 : NOT ( `\t(?:WW\(\pv,|VG\(\onder\)|VNW\((?:betr|vrag|vb),)+\n` + REGEX )  
Filter3 : NOT ( REGEX + `\tzij\n.+tWW\(\pv,+\n(?:ze|zij)\t` )

51.3 *ze.dbl.mv* Voorbeeld : *ze komen zij mee*

Regex : `^ze\tVNW\(\pers,pron,stan,red,3,mv\)`  
Manueel : subjectvorm – objectvorm

Filter1 : NOT ( REGEX + \tze\nzij\t )  
Filter2 : NOT ( \t(?:WW\ (pv,|VG\ (onder\)|VNW\ ((?:betr|vrag|vb),).+\n + REGEX )  
Filter3 : REGEX + \tze\n.+ \tWW\ (pv,.\n(?:ze|zij))\t

51.4 *zij.dbl.mv* Voorbeeld : *zij komen zij mee*

Regex : ^zij\tVNW\ (pers,pron,nomin,vol,3p,mv\  
Filter1 : NOT ( ^ze\t.+ \n + REGEX )  
Filter2 : NOT ( \t(?:WW\ (pv,|VG\ (onder\)|VNW\ ((?:betr|vrag|vb),).+\n + REGEX )  
Filter3 : REGEX + \tzij\n.+ \tWW\ (pv,.\n(?:ze|zij))\t

## 52. ZIJ – MV (GEÏNVERTEERD)

52.1 *ze.inv.mv* Voorbeeld : *komen ze mee?*

Regex : ^ze\tVNW\ (pers,pron,stan,red,3,mv\  
Manueel : subjectvorm – objectvorm  
Filter1 : NOT ( REGEX + \tze\nzij\t )  
Filter2 : \t(?:WW\ (pv,|VG\ (onder\)|VNW\ ((?:betr|vrag|vb),).+\n + REGEX  
Filter3 : NOT ( ^(?:ze|zij)\t.+ \n.+ WW\ (pv,.\n + REGEX )

52.2 *zij.inv.mv* Voorbeeld : *komen zij mee?*

Regex : ^zij\tVNW\ (pers,pron,nomin,vol,3p,mv\  
Filter1 : NOT ( ^ze\t.+ \n + REGEX )  
Filter2 : \t(?:WW\ (pv,|VG\ (onder\)|VNW\ ((?:betr|vrag|vb),).+\n + REGEX  
Filter3 : NOT ( ^(?:ze|zij)\t.+ \n.+ WW\ (pv,.\n + REGEX )

52.3 *ze.zij.mv* Voorbeeld : *komen zezij mee?*

Regex : ^ze\tVNW\ (pers,pron,stan,red,3,mv\  
Manueel : subjectvorm – objectvorm  
Filter1 : REGEX + \tze\nzij\t  
Filter2 : NOT ( ^(?:ze|zij)\t.+ \n.+ WW\ (pv,.\n + REGEX )

## 53. MIJ (OBJECTVORM)

53.1 *me.obj* Voorbeeld : *hij ziet me*

Regex : ^me\tVNW\ (pr,pron,obl,red,1,ev\  
Manueel : objectvorm – reflexief

53.2 *mij.obj* Voorbeeld : *hij ziet mij*

Regex : ^mij\tVNW\ (pr,pron,obl,vol,1,ev\  
Manueel : objectvorm – reflexief

## 54. JOU

54.1 *je.obj* Voorbeeld : *ik zie je*

Regex : `^je\tVNW\(\pr,pron,obl,red,2v,getal\)`

Manueel : objectvorm – reflexief

54.2 *jou* Voorbeeld : *ik zie jou*

Regex : `^jou\tVNW\(\pers,pron,obl,vol,2v,ev\)`

54.3 *jullie.obj* Voorbeeld : *ik zie jullie*

Regex : `^jullie\tVNW\(\pers,pron,stan,nadr,2v,mv\)`

Manueel : subjectvorm – objectvorm

54.4 *u.obj* Voorbeeld : *ik zie u*

Regex : `^u\tVNW\(\pr,pron,obl,vol,2,getal\)`

Manueel : objectvorm – reflexief

## 55. HEM

55.1 *hem* Voorbeeld : *je ziet hem*

Regex : `^hem\tVNW\(\pers,pron,obl,vol,3,ev,masc\)`

Manueel : subjectvorm – objectvorm

55.2 *m.obj* Voorbeeld : *je ziet 'm*

Regex : `^m\tVNW\(\pers,pron,obl,red,3,ev,masc\)`

Manueel : subjectvorm – objectvorm

## 56. HAAR (OBJECTVORM)

56.1 *haar.obj* Voorbeeld : *je ziet haar*

Regex : `^haar\tVNW\(\pers,pron,obl,vol,3,getal,fem\)`

56.2 *ze.obj.vr* Voorbeeld : *je ziet ze*

Regex : `^ze\tVNW\(\pers,pron,stan,red,3,ev,fem\)`

Manueel : subjectvorm – objectvorm

## 57. HEN

57.1 *hen* Voorbeeld : *zij ziet hen*

Regex : `^hen\tVNW\(\pers,pron,obl,vol,3p,mv\)`

57.2 *hun.obj* Voorbeeld : *zij ziet hun*  
Regex : `^hun\tVNW\(\(pers,pron,obl,vol,3p,mv\)`

57.3 *ze.obj.mv* Voorbeeld : *zij ziet ze*  
Regex : `^ze\tVNW\(\(pers,pron,stan,red,3,mv\)`  
Manueel : subjectvorm – objectvorm

## 58. MIJ (REFLEXIEF)

58.1 *me.ref* Voorbeeld : *ik was me*  
Regex : `^me\tVNW\(\(pr,pron,obl,red,1,ev\)`  
Manueel : objectvorm – reflexief

58.2 *mij.ref* Voorbeeld : *ik was mij*  
Regex : `^mij\tVNW\(\(pr,pron,obl,vol,1,ev\)`  
Manueel : objectvorm – reflexief

58.3 *mijn.eigen* Voorbeeld : *ik was mijn eigen*  
Regex1 : `^mijn\tVNW\(\(bez,det,stan,vol,1,ev,prenom,zonder,agr\)`  
Regex2 : `^m'n\tVNW\(\(bez,det,stan,red,1,ev,prenom,zonder,agr\)`  
Filter : REGEX + `\t.+neigen\t`

## 59. JE (REFLEXIEF)

59.1 *je.ref* Voorbeeld : *jij wast/jullie wassen je*  
Regex : `^je\tVNW\(\(pr,pron,obl,red,2v,getal\)`  
Manueel : objectvorm – reflexief

59.2 *u.ref* Voorbeeld : *gij wast u*  
Regex : `^u\tVNW\(\(pr,pron,obl,vol,2,getal\)`  
Manueel : objectvorm – reflexief

59.3 *je.eigen* Voorbeeld : *jij wast je eigen*  
Regex : `^je\tVNW\(\(bez,det,stan,red,2v,ev,prenom,zonder,agr\)`  
Filter : REGEX + `\t.+neigen\t`

59.4 *uw.eigen* Voorbeeld : *gij wast uw eigen*  
Regex : `^uw\tVNW\(\(bez,det,stan,vol,2,getal,prenom,zonder,agr\)`  
Filter : REGEX + `\t.+neigen\t`

59.5 *zich.2* Voorbeeld : *u wast zich*

Regex : `^zich\tVNW\(\refl,pron,obl,red,3,getal\)`

Manueel : 2-de persoon – 3-de persoon

## 60. ONS (REFLEXIEF)

60.1 *ons.ref*                      Voorbeeld : *wij wassen ons*

Regex : `^ons\tVNW\(\pr,pron,obl,vol,1,mv\)`

Manueel : objectvorm – reflexief

60.2 *ons.eigen*                      Voorbeeld : *wij wassen ons eigen*

Regex : `^ons\tVNW\(\bez,det,stan,vol,1,mv,prenom,zonder,evon\)`

Filter : REGEX + `\t.+neigen\t`

## 61. ZICH

61.1 *zich.3*                      Voorbeeld : *hij/zij wast/wassen zich*

Regex : `^zich\tVNW\(\refl,pron,obl,red,3,getal\)`

Manueel : 2-de persoon – 3-de persoon

61.2 *eigen.3*                      Voorbeeld : *hij/zij wast/wassen zijn/haar/hun eigen*

Regex1 : `^zijn\tVNW\(\bez,det,stan,vol,3,ev,prenom,zonder,agr\)`

Regex2 : `^z'n\tVNW\(\bez,det,stan,red,3,ev,prenom,zonder,agr\)`

Regex3 : `^haar\tVNW\(\bez,det,stan,vol,3,ev,prenom,zonder,agr\)`

Regex4 : `^hun\tVNW\(\bez,det,stan,vol,3,mv,prenom,zonder,agr\)`

Filter : REGEX + `\t.+neigen\t`

## 62. ELKAAR

62.1 *elkaar*                      Voorbeeld : *wij wassen elkaar*

Regex : `^elkaar\tVNW\(\recip,pron,obl,vol,persoon,mv\)`

62.2 *mekaar*                      Voorbeeld : *wij wassen mekaar*

Regex : `^mekaar\tVNW\(\recip,pron,obl,vol,persoon,mv\)`

Van reduplicatief *zij* (3.enk – *zij* komt *zij* mee) zijn er in het CGN geen treffers.





## Hoofdstuk 4 | Methodologie

Voor de analyses zullen we de taalvariabelen kruistabuleren met de taalexterne factoren uit het CGN (daarmee bedoelen we dan enerzijds de interne stratificatie en anderzijds de sprekerskenmerken – zie 2.1 en 2.2). Elke analyse zal met andere woorden vertrekken van een contingentietabel: een tabel met het frequentieaantal voor elke combinatie van een taalvariant en een taalvariëteit – de verschillende variëteiten, die we zullen vergelijken, zijn natuurlijk gespecificeerd in de categorieën van de externe factoren. Aan de ene kant kan dat gebeuren voor elke factor apart. De analyse is dan bivariaat, en de techniek daarvoor zal in 4.1 worden besproken. Aan de andere kant kunnen we ook alle factoren samen opnemen in een globale multivariate analyse. De daarbij aangewende techniek staat beschreven in paragraaf 4.2.

Hoe dan ook haken we resoluut in op de uniformiteitsmetingen uit Geeraerts *et al.* (1999): wat ons primair interesseert zijn *afstanden tussen taalvariëteiten*. De rijen en kolommen van een frequentietabel kunnen namelijk telkens geïnterpreteerd worden als dimensies van een meetkundige ruimte. Daarbinnen vormt dan de ‘complementaire’ factor – kolommen in het geval van rijen; rijen in het geval van kolommen – een *datawolk van punten*, waartussen dus afstanden berekend kunnen worden. In onze studie zullen we daarbij de conventie hanteren om de taalvariabelen en –varianten als rijen te construeren, en de variëteiten als kolommen:

		Variëteit X	Variëteit Y	...	Variëteit Z
Variabele 1	Variant 1.1				
	...				
	Variant 1.I <sup>(1)</sup>				
...	...				
Variabele K	Variant K.1				
	...				
	Variant K.I <sup>(K)</sup>				
TOTAAL		$n_{+X}$	$n_{+Y}$	...	$n_{+Z}$

Tabel 4.1: Een taalvariabele×taalvariëteit-tabel

De uniformiteit tussen twee taalvariëteiten is zo de meetkundige afstand tussen de corresponderende punten in de ruimte omspannen door de taalvarianten (waarbij er voorts een specifieke metriek gedefinieerd wordt – zie 4.1 verderop).

Daarnaast zullen we eveneens de complementaire afstanden berekenen: die tussen de varianten in termen van de variëteiten. Het voordeel daarvan is dat we deze afstandsmeting en die tussen de variëteiten kunnen samenbrengen in één en dezelfde grafische plot: de ‘biplot’ – waarvan de constructie overigens de dimensie-reductietechniek veronderstelt, die verderop in dit hoofdstuk zal worden toegelicht. Daaruit blijkt dan de grote meerwaarde van een biplot: een biplot visualiseert zo namelijk niet alleen de afstanden tussen rijen onderling en kolommen onderling, maar onthult ook de *associaties* van rijen mét kolommen (en vice versa). Als dusdanig is een biplot een uiterst compacte en summiere weergave van *alle* informatie die een contingentietabel bevat.

#### 4.1 De bivariate techniek

Uniformiteit, zoals gespecificeerd in Geeraerts *et al.* (1999), wordt gemeten aan de hand van de volgende formule (toegepast op variëteiten X en Y in tabel 4.1):

$$U(X, Y) = \frac{1}{K} \sum_{k=1}^K \sum_{i=1}^{I^{(k)}} \min(r_{iX}^{(k)}, r_{iY}^{(k)})$$

waarbij:

$$r_{ij}^{(k)} = \frac{n_{ij}^{(k)}}{n_{+j}^{(k)}}$$

Daarmee is de uniformiteit vooral een similariteitsmaat. De onderliggende afstandsmetriek waarop de uniformiteit gebaseerd is, is het gemiddelde absolute verschil van de relatieve frequenties:

$$D(X, Y) = \frac{1}{K} \sum_{k=1}^K \frac{1}{2} \sum_{i=1}^{I^{(k)}} |r_{iX}^{(k)} - r_{iY}^{(k)}|$$

Bewezen kan namelijk worden dat  $U(X, Y) = 1 - D(X, Y)$ . Voor verdere details omtrent de uniformiteitsmaat evenals een vergelijking ervan met andere afstands- en associatiematen, zie Speelman *et al.* (2003).

De kleine wijziging die we daarbij zullen aanbrengen, is dat we de afstandsmetingen zullen verrichten op *gestandaardiseerde* variabelen (in de statistische zin van het woord; dit wil zeggen: gecontroleerd voor variantie en covariantie). Voor ons onderzoek willen we namelijk de zogenaamde ‘structurele verschillen’ in kaart brengen (voor de notie *structural difference*, zie Agresti 2002: 80-86). Metingen op ongestandaardiseerde variabelen (zoals de uniformiteitsmaat die doet) zijn daarvoor evenwel te onbetrouwbaar.

Inachtneming van standaardisering van de variabelen leidt tot de volgende metriek:

$$D(X, Y) = \frac{1}{\frac{1}{n_{+X}} - \frac{1}{n_{+Y}}} (\vec{x} - \vec{y})^T \Sigma^{-1} (\vec{x} - \vec{y})$$

Hierbij stellen de vectoren  $\vec{x}$  en  $\vec{y}$  de hele kolommen van variëteit X, respectievelijk Y voor gedeeld door hun kolomtotalen  $n_{+X}$  en  $n_{+Y}$ , en is  $\Sigma$  de covariantiematrix van de taalvariabelen:

$$\Sigma = \begin{bmatrix} r_{1+}(1 - r_{1+}) & \cdots & -r_{1+}r_{I+} \\ \vdots & \ddots & \vdots \\ -r_{I+}r_{1+} & \cdots & r_{I+}(1 - r_{I+}) \end{bmatrix}$$

De bovenstaande metriek is de zogenaamde *chi-kwadraatmetriek* (Anderson 2003: 80). Het is de geijkte afstandsmaat voor het verschil tussen twee multivariate variabelen (deze maat is namelijk een Mahalanobisafstand), waar de variabelen met name een multinomiale verdeling hebben – zoals bijvoorbeeld de factoren in een contingentietabel.

De chi-kwadraatmetriek ontleent haar naam aan het feit dat ze aantoonbaar equivalent is aan de Pearson- $X^2$ -statistiek van de kruistabel die gevormd wordt door de twee multinomiaalvariabelen samen te binden tot een dichotome factorvariabele:

		Variëteit X	Variëteit Y
Variabele 1	Variant 1.1		
	...		
	Variant 1.I <sup>(1)</sup>		
...	...		
Variabele K	Variant K.1		
	...		
	Variant K.I <sup>(K)</sup>		
TOTAAL		n <sub>+X</sub>	n <sub>+Y</sub>

Tabel 4.2: Variëteiten X en Y samengebonden tot een frequentietabel

De formule voor de Pearson- $X^2$  is natuurlijk:

$$X^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{\left(n_{ij} - \frac{n_{i+}n_{+j}}{n_{++}}\right)^2}{\frac{n_{i+}n_{+j}}{n_{++}}}$$

Bijgevolg kunnen we eveneens de  $X^2$ -statistiek gebruiken om de afstanden tussen rijen en kolommen in een contingentietabel te meten, en dat is dan ook precies de aanpak die we hier zullen volgen. Onze techniek is daarmee nagenoeg analoog aan die van Correspondentieanalyse (Benzécri 1992, Greenacre 1984). We zullen namelijk de frequentie in elke cel van de tabel vervangen door zijn  $X^2$ -deviatie:

$$d_{ij} = \frac{n_{ij} - \frac{n_{i+}n_{+j}}{n_{++}}}{\sqrt{\frac{n_{i+}n_{+j}}{n_{++}}}}$$

om uit de aldus bekomen tabel vervolgens een overzichtelijke representatie van de variatie te extraheren. Daartoe maken we gebruik van ‘Singuliere-Waardeontbinding’.

Singuliere-Waardeontbinding – afgekort: SVD – is een matrixalgebraïsche techniek om een hoogdimensionale matrix (zoals een tabel met een groot aantal rijen en/of kolommen, bijvoorbeeld) te approximeren in enkele ruimtelijke dimensies. Het eigene aan die approximatie is daarbij dat ze geoptimaliseerd is: de bedoeling is om zoveel mogelijk informatie uit de oorspronkelijke matrix te behouden in zo weinig mogelijk dimensies (voor de mathematische details achter SVD, zie Strang 2003). Omdat wij daarbij enkel geïnteresseerd zijn in visualisering, zullen we ons beperken tot twee dimensies. Dat zijn dan de plots waarvan boven sprake.

De hier geschetste benadering is overigens volkomen identiek aan de werkwijze om voor elk paar van multinomiaalvariabelen de afstand te meten, en de matrix van die afstanden vervolgens te ontbinden met behulp van Multidimensionele Schaalanalyse (MDS – Cox & Cox 2001, Borg & Groenen 2005). Vanuit onze initiële vraagstelling naar afstandsmeting moge zulks op het eerste gezicht enigszins meer voor de hand liggen. Beide procedures geven evenwel net dezelfde resultaten – voor het bewijs daarvan, zie Greenacre (1984: 81-82).

Ten slotte dienen we er evenwel nog rekening mee te houden dat de rijen van onze tabellen gepartitioneerd zijn in taalvariabelen. We willen namelijk uitdrukkelijk enkel de variatie *binnen* de taalvariabelen meten – zoals vermeld in hoofdstuk 3 zijn we principieel uit op de zogenaamde ‘formeel onomasiologische variatie’. Elke onderlinge variatie *tussen* de taalvariabelen dient met andere woorden geëlimineerd te worden. Het verband tussen totale, binnen- en tussenvariatie wordt daarbij uitgedrukt in de Stelling van Huyghens (Greenacre 1984: 203-204):

$$X_{TOTAAL}^2 = X_{BINNEN}^2 + X_{TUSSEN}^2$$

met:

$$X_{TOTAAL}^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{\left(n_{ij} - \frac{n_{i+}n_{+j}}{n_{++}}\right)^2}{\frac{n_{i+}n_{+j}}{n_{++}}} = \sum_{k=1}^K \sum_{i=1}^{I^{(k)}} \sum_{j=1}^J \frac{\left(n_{ij} - \frac{n_{i+}n_{+j}}{n_{++}}\right)^2}{\frac{n_{i+}n_{+j}}{n_{++}}}$$

$$X_{BINNEN}^2 = \sum_{k=1}^K \sum_{i=1}^{I^{(k)}} \sum_{j=1}^J \frac{\left(n_{ij} - \frac{n_{i+}n_{+j}^{(k)}}{n_{++}^{(k)}}\right)^2}{\frac{n_{i+}n_{+j}}{n_{++}}}$$

$$X_{TUSSEN}^2 = \sum_{k=1}^K \sum_{j=1}^J \frac{\left(n_{+j}^{(k)} - \frac{n_{+j}n_{++}^{(k)}}{n_{++}}\right)^2}{\frac{n_{+j}n_{++}^{(k)}}{n_{++}}}$$

De tabellen, die we zullen ontbinden, stellen we dan ook op aan de hand van de  $X_{BINNEN}^2$ .

## 4.2 De multivariate techniek

Voor de multivariate analyse hebben we een techniek nodig die de verschillende taalexterne factoren samen bestudeert. Een mogelijk vertrekpunt – en één waarvan we in deze studie dan ook zullen uitgaan – is om tussen alle externe factoren de interactievariabele te berekenen, en die vervolgens te kruistabuleren met de taalvariabelen. De contingentietabel ziet er dan als volgt uit:

Factor M		Var X'				...	Var Z'						
...		...				...	...						
Factor 1		Var X	...	Var Z	...	Var X	...	Var Z	...	Var X	...	Var Z	...
Variabele 1	Variant 1.1												
	...												
	Variant 1.I <sup>(1)</sup>												
...	...												
Variabele K	Variant K.1												
	...												
	Variant K.I <sup>(K)</sup>												
TOTAAL		$n_{+X...X'}$	...	$n_{+Z...X'}$	...	$n_{+X...X'}$	...	$n_{+Z...X'}$	...	$n_{+X...Z'}$	...	$n_{+Z...Z'}$	...

Tabel 4.3: De tabel van de taalvariabelen  $\times$  de interactie tussen alle taalexterne factoren samen

Vervolgens moeten we bij de externe factoren nog een manier vinden om de effecten op de meer algemenere niveaus te reconstrueren. De interactie tussen de externe factoren is op zich natuurlijk niet interessant maar alleen voor zover we er generalisaties uit kunnen halen – dat is in zekere zin zelfs de hele bedoeling van ons onderzoek. Zulke generalisaties zijn echter triviaal: elk hoger effect kan namelijk afgezonderd worden door per rij de frequenties in de relevante kolommen – dat zijn dus die kolommen die onder het effect in kwestie ressorteren – gewoon te sommeren (dat volgt overigens uit de Wet van de Totale Waarschijnlijkheid). Alle externe effecten kunnen zo bij wijze van spreken hun eigen kolom krijgen.

De  $X^2$ -deviaties voor deze gegeneraliseerde effecten, nodig om de heterogeniteit te meten, zijn verder geheel volgens formule. De celdeviatie voor rij  $i$  en gesommeerd over de  $J^{(l)}$  kolommen van effect  $l$  (waarbij  $l$  dus een specifieke subset van de kolommen representeert:  $l=1 \dots L$ ) is met andere woorden:

$$d_{i+}^{(l)} = \frac{n_{i+}^{(l)} - \frac{n_{i+} n_{++}^{(k)(l)}}{n_{++}^{(k)}}}{\sqrt{\frac{n_{i+} n_{++}^{(l)}}{n_{++}}}}$$

Het zijn de kolommen van celdeviaties zoals deze, die dan ook geplot zullen worden.

Daartoe maken we gebruik van de intrinsieke eigenschap van SVD die de kolomruimte definieert in termen van de rijruimte (en vice versa). We herhalen nog eens even de hoofdlijnen van SVD. De SVD van om het even welke matrix  $X$  is het product van matrix  $U$ ,  $\Sigma$ , en  $V^T$ :

$$X = U\Sigma V^T$$

Daarbij geeft  $U$  voor elke rij in  $X$  de coördinaten in de approximatieve ruimte, doet  $V^T$  hetzelfde voor de kolommen van  $X$ , en is  $\Sigma$  een matrix die aangeeft hoeveel variatie elke dimensie van die ruimte verklaart – de dimensies in  $U$  en  $V^T$  zijn namelijk genormaliseerd (dat wil zeggen dat ze herschaald zijn opdat hun lengte gelijk is aan 1). Om de ‘werkelijke’ rij- en kolomcoördinaten te bekomen, moeten we matrices  $U$  en  $V^T$  dan ook vermenigvuldigen met  $\Sigma$ . De coördinaten voor de rijen en de kolommen van  $X$  staan met andere woorden in  $U\Sigma$ , respectievelijk  $\Sigma V^T$ .

Nu kan de bovenstaande gelijkheid herschreven worden als:

$$U^T X = \Sigma V^T$$

Dat komt erop neer dat de kolomcoördinaten evengoed verkregen kunnen worden door matrix  $X$  te vermenigvuldigen met  $U^T$ .

Voor elke ‘extra’ kolom – die we bijvoorbeeld noteren als  $\mathbf{x}_*$  – geldt bij uitbreiding hetzelfde: vermenigvuldigen met  $U^T$  levert de coördinaten ervan in de approximatieve ruimte op – genoteerd als  $\mathbf{v}_*^T$ :

$$U^T \vec{x}_* = \Sigma \vec{v}_*^T$$

Dat zal dan ook de manier zijn waarop we de posities van de taalexterne effecten bepalen.

Samengevat: eerst ontbinden we de celdeviaties in de globale tabel. De taalvarianten – de rijen – plotten we daarbij meteen: de verhouding tussen de varianten krijgt namelijk haar meest nauwkeurige representatie in de coördinaten berekend op basis van alle mogelijke combinaties van de externe factoren. Voor die externe factoren – de kolommen – selecteren we echter op voorhand de gewenste effecten, waarvan we de coördinaten vervolgens via de hier geschetste ‘omweg’ aan de biplot zullen toevoegen.

Hoewel we onze aanpak tot slot met Correspondentieanalyse in verband gebracht hebben, verschilt onze multivariate techniek evenwel aanzienlijk van een geijkte Multipele Correspondentieanalyse (MCA). MCA construeert namelijk een tabel van correlaties tussen *alle* factoren (zowel linguïstische als externe), die vervolgens ontbonden wordt. Daarentegen ontbindt onze techniek zoals aangegeven een kruistabel van enerzijds taalvariabelen met anderzijds (de interactie tussen) de externe factoren. Het verschil zit er met andere woorden in dat de afstanden zoals die bij MCA berekend worden óók de associaties tussen de externe factoren *onderling* incorporeren (voor een verdere bespreking van MCA evenals aanverwante technieken, zie Greenacre & Blasius 2006). In onze techniek weerspiegelen ze louter en alleen verschillen in taalgebruik. Zulks sluit natuurlijk nauwer aan bij onze onderzoeksvraag naar formeel onomasiologische variatie.



## Hoofdstuk 5 | Bivariate analyses

Zoals aangegeven vatten we de analyses aan door voorlopig nog voor elke externe factor apart de taalvariatie te meten. Achtereenvolgens komen aan bod: register (5.1), regio (5.2), leeftijd (5.3 en 5.4), sekse (5.5), opleidingsniveau (5.6), en beroepsniveau (5.7). Bij leeftijd zullen we de sprekers op twee verschillende manieren in cohorten onderverdelen: eerst op basis van decennia (5.3), en daarnaast op basis van generaties (5.4) – zie de paragrafen in kwestie voor meer uitleg. Verder maken we voor de regiofactor gebruik van de geboorteregio van de sprekers in het CGN, hoewel we eveneens kort een vergelijking zullen maken met de woonregio (dit is de regio waar de spreker ten tijde van de spraakopname woonachtig was). Register is ten slotte gebaseerd op de veertienvoudige stratificatie van het CGN in ‘componenten’.

### 5.1 Register

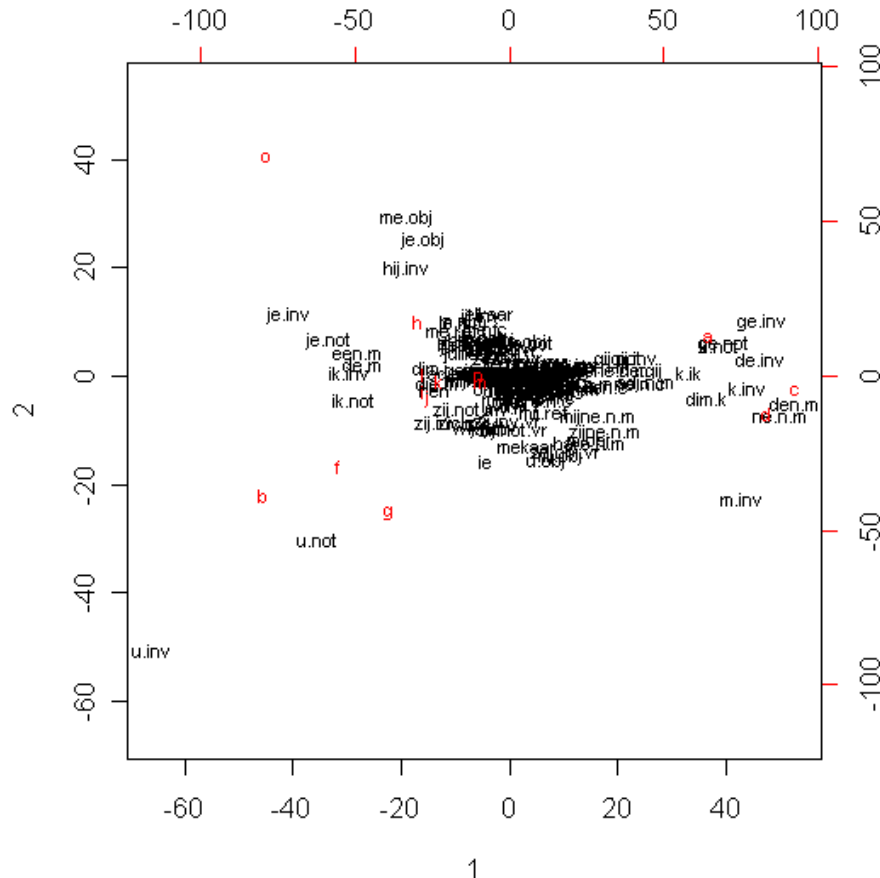
Die componentindeling staat voor alle duidelijkheid nog eens in Tabel 5.1 (vergelijk 2.1):

<b>Label</b>	<b>Component</b>	<b>Aantal woorden (Vlaanderen)</b>
a	Spontane conversaties ('face-to-face')	878.383
b	Interviews met leraren Nederlands	315.554
c	Telefoondialogen opgenomen m.b.v. platform	465.096
d	Telefoondialogen opgenomen m.b.v. minidisc recorder	343.167
e	Zakelijke onderhandelingen	0
f	Interviews en discussie uitgezonden op radio en televisie	250.708
g	Discussie, debatten, vergaderingen (m.n. politieke)	138.819
h	Lessen	105.436
i	Spontane commentaren (o.a. sport) uitgezonden op radio en televisie	78.022
j	Actualiteitenrubrieken en reportages uitgezonden op radio en televisie	95.206
k	Nieuwsbulletins uitgezonden op radio en televisie	82.855
l	Beschouwingen en commentaren uitgezonden op radio en televisie	65.386
m	Missen, lezingen, plechtige toespraken	12.510
n	Colleges, voordrachten, lezingen	79.067
o	Voorgelezen teksten	351.419
<b>Totaal</b>		<b>3.261.628</b>

*Tabel 5.1: De componenten van het CGN*

Figuur 5.1 geeft dan de resultaten. Met betrekking tot de veertien componenten kunnen we een zestal clusters onderscheiden. Links bovenaan vormt component o (voorgelezen tekst) een singleton cluster. Onderaan groepen component b (interviews met leraren Nederlands), f (interviews en discussies uitgezonden op radio en televisie) en g (discussies, debatten, en vergaderingen) bij elkaar. Helemaal rechts treffen we de face-to-face conversaties (component a) aan samen met de telefoondialogen (c en d). Verder zien we in het midden van de plot nog drie clustertjes. De lessen opgenomen in het klaslokaal (component h) staan daarbij enigszins apart. Veruit de interessantste cluster is echter die van componenten i (spontane commentaren uitgezonden op radio en televisie), j (actualiteitenrubrieken en reportages uitgezonden op radio en televisie), k (niewsbulletins uitgezonden op radio en televisie), en l (beschouwingen en commentaren uitgezonden op radio en televisie). Al deze componenten hebben namelijk met elkaar gemeen dat ze

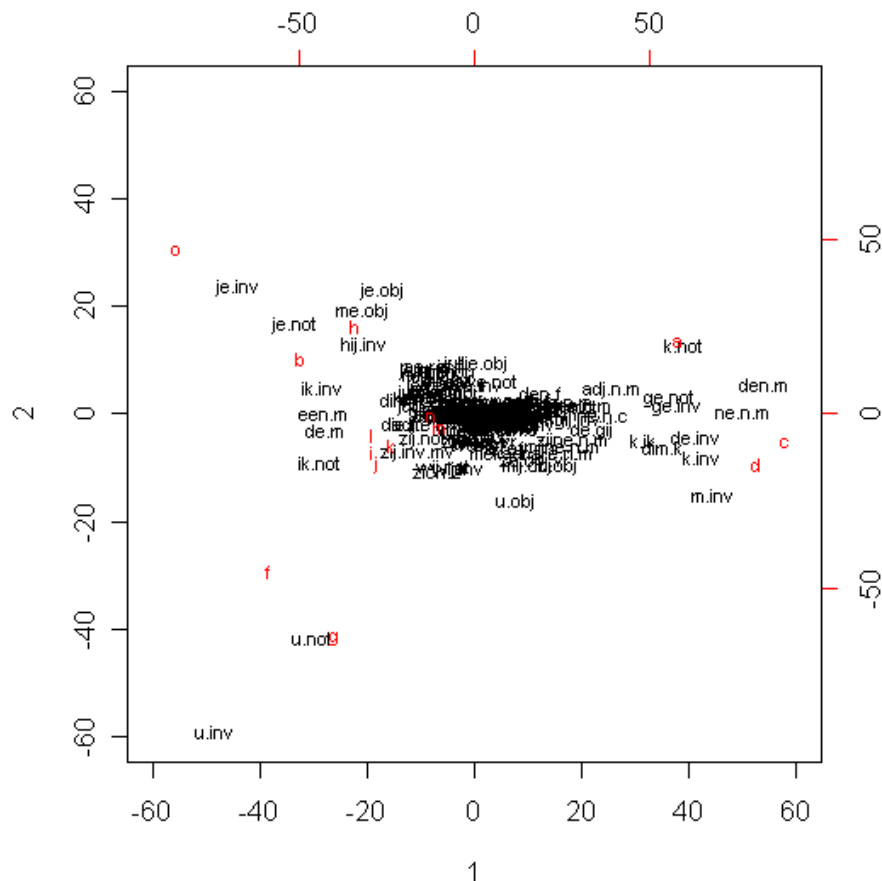
'uitgezonden op radio en televisie' zijn. Deze cluster representeert dan ook de positie van een voor ons onderzoek belangrijke taalvariëteit in het Vlaamse taallandschap: het Journaalnederlands. Tot slot clustert component m (missen, lezingen, plechtige toespraken) nog samen met n (colleges, voordrachten, lezingen) – ook dat is niet toevallig: in de prereleases van het CGN vormden deze twee componenten nog één component.



Figuur 5.1: Biplot – register

Wat al meteen opvalt aan deze resultaten is toch de positie van component b: hoe komt het dat interviews met leraren Nederlands vrijwel hetzelfde taalgebruik vertonen als dat van zulke formele spreesituaties als interviews, discussies en debatten – waarvan het merendeel van de participanten overigens politici zijn? Het antwoord op die vraag ligt bij een eigenaardigheid in de opmaak van component b: de rol van de interviewer wordt namelijk steeds door één en dezelfde persoon ingevuld (spreker-id: V40100). Zulks leidt natuurlijk tot een vertekening: hoewel component b in se als dialogisch geboekstaafd staat, weerspiegelt het taalgebruik in deze component asymmetrisch de talige idiosyncrasieën van één enkele spreker. Daarmee kan dan ook de correlatie van component b met f en g verklaard worden. Bij manuele verificatie is namelijk gebleken dat de interviewer in kwestie de leraren consequent toespreekt in een vrij formele stijl – wat zich bijvoorbeeld uit in het gebruik van *u* als

aanspreekvorm. Een dialogeerstijl zoals deze is verder nogal typisch voor de manier waarop politieke debatten gevoerd worden. Bijgevolg clustert component b samen met f en g, een cluster die echter volkomen artificieel is. De oplossing bestaat er daarom in om de uitingen van de interviewer geheel uit de dataset te verwijderen. Component b bevat voortaan met andere woorden louter het taalgebruik van de interviewee's – de leerkrachten.



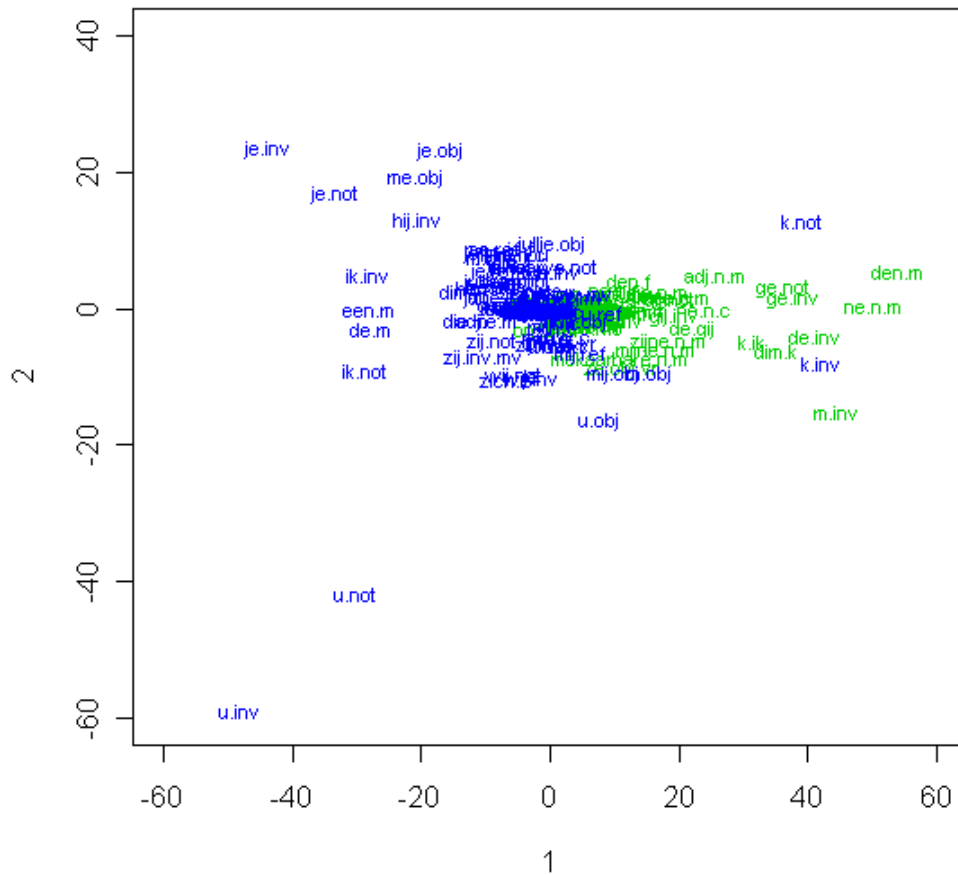
*Figuur 5.2: Biplot – register: zonder spreker V40100*

De resultaten voor deze ‘gekuiste’ steekproef worden dan gevisualiseerd in Figuur 5.2. We zien dat component b zich nu ophoudt bij component h – wat toch een meer ‘natuurlijke’ positie voor b genoemd mag worden: beide componenten bevatten namelijk het taalgebruik van dezelfde maatschappelijke groep sprekers: de leerkrachten. Globaal bemerken we nu vijf clusters:

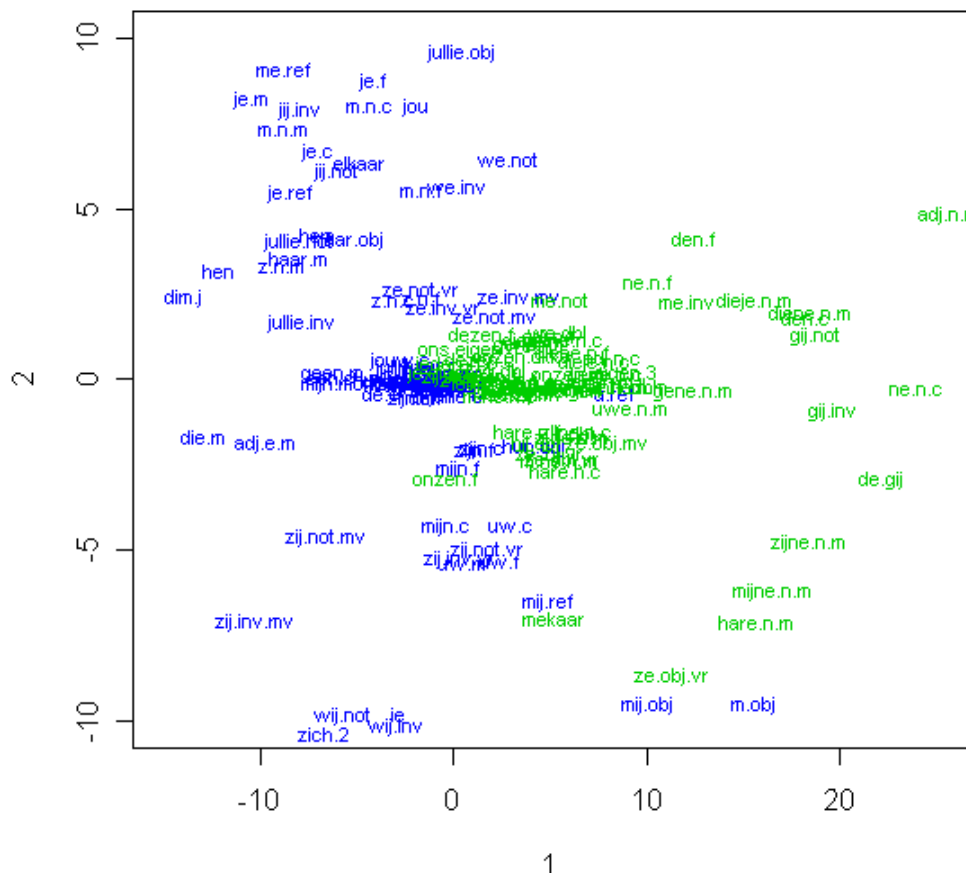
- Een cluster van componenten o, b, en h links bovenaan.
- Een cluster van f en g linksonder.
- Daar vrijwel midden tussenin een cluster van componenten i, j, k, en l.
- Ietwat rechts daarvan een cluster van m en n.
- Uiterst rechts een cluster van componenten a, c, en d.

Daarmee kan ook een interpretatie aan de beide assen van de plot gegeven worden. De horizontale as constitueert duidelijk een continuüm tussen spontaan taalgebruik (rechts) en tot wat we in tegenstelling daarmee ‘gemonitord’ taalgebruik zouden kunnen noemen (links). De verticale as differentieert dan weer tussen het ‘debattaire’ taalgebruik beneden en wat we in het licht van de maatschappelijke status van de sprekers in kwestie boven maar als ‘schools’ taalgebruik zullen bestempelen – een distictie die we op basis van de respectieve spreekstijlen die we in deze situaties mogen verwachten (en ook daadwerkelijk aantreffen: zie verderop) ook nog zouden willen interpreteren als het onderscheid tussen vertrouwelijk en beleefd.

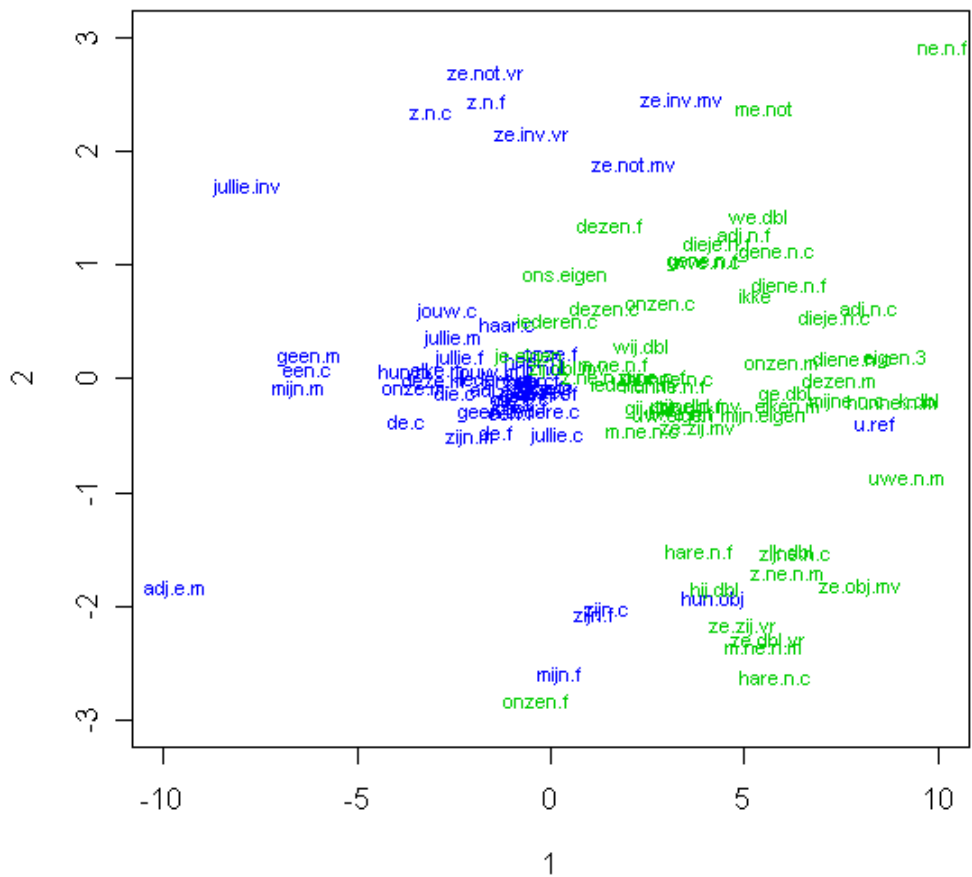
Wat is vervolgens de associatie van de registers met de taalvarianten – de zwarte punten in de plot? Om die vraag afdoende te kunnen beantwoorden, geven we van de varianten aparte plots, telkens op een kleiner interval inzoomend (Figuren 5.3, 5.4, 5.5, en 5.6). Vanuit onze onderzoeksvraag naar de verhouding tussen het Journaalnederlands en het Soapvlaams hebben we de Journaalnederlandse vormen in het blauw gezet, en de Soapvlaamse in het groen. Daarbij drongen zich overigens enkele keuzes op: de objectvorm *u* evenals het possessivum *uw* hebben we als Journaalnederlands geassocieerd – het geflecteerde *uwe(n)* is dan Soapvlaams. Verder beschouwen we de objectvorm *ze* (zowel 3.enk.fem als 3.mv) als Soapvlaams; de corresponderende vormen *haar*, respectievelijk *hen* en *hun* zijn dus Journaalnederlands. Tot slot hebben we bij de reciproke voornaamwoorden – geheel arbitrair – *elkaar* bij het Journaalnederlands ondergebracht, terwijl *mekaar* bij het Soapvlaams.



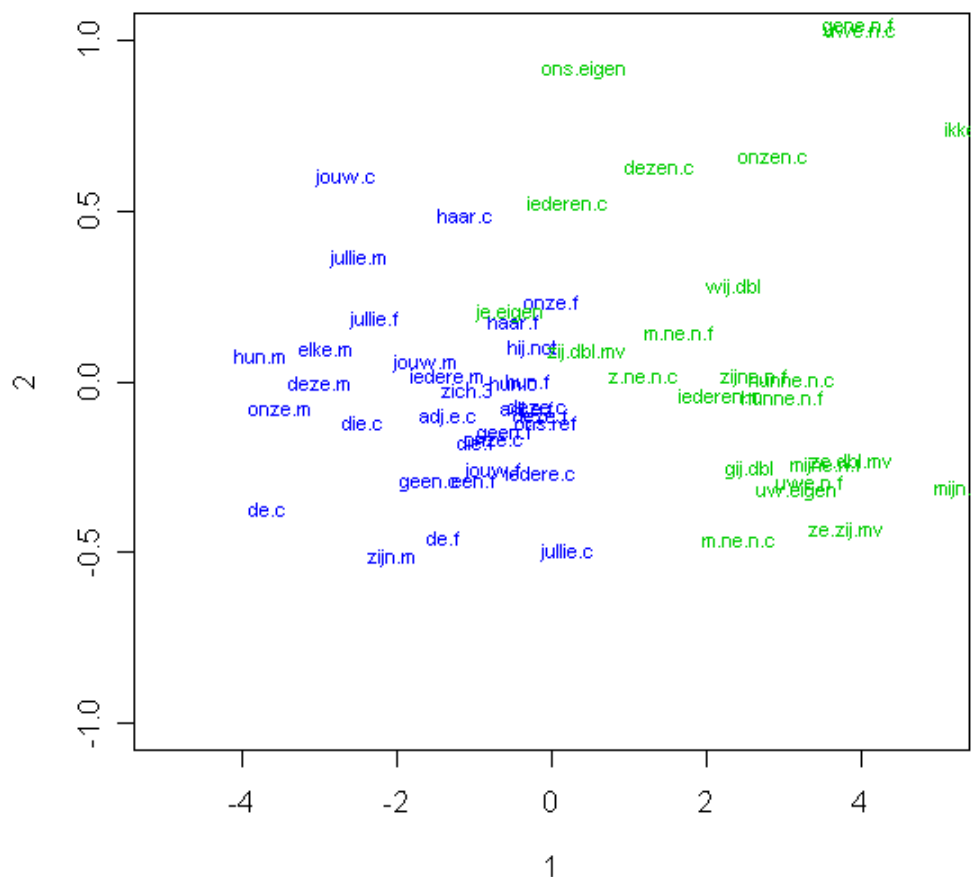
Figuur 5.3: Register – varianten: detail [-60;60] horizontaal en [-60;40] verticaal



Figuur 5.4: Register – varianten: detail [-15;20] horizontaal en [-10;10] verticaal



Figuur 5.5: Register – varianten: detail [-10;10] horizontaal en [-3;3] verticaal



Figuur 5.6: Register – varianten: detail [-5;5] horizontaal en [-1;1] verticaal

We zien dan dat het onderscheid tussen Journaalnederlands en Soapvlaams nagenoeg samenvalt met de negatieve en de positieve helft van de horizontale as (enkel in het dichtbevolkte centrum zijn er een handvol varianten die voor wat overlapping zorgen): de Soapvlaamse vormen zitten in het positieve rechterdeel, geassocieerd met de componenten a-c-d; terwijl de Journaalnederlandse vormen in het negatieve linkerdeel, geassocieerd met de drie clusters o-b-h, i-j-k-l, en f-g (m-n heeft als cluster in het centrum geen uitgesproken voorkeur voor één van beide variëteiten). De enige sterke afwijking vormt het fonetisch gereduceerde 'k (zowel niet- als wel geïnverteerd), dat zich eerder Soapvlaams gedraagt – wat omgekeerd natuurlijk een reden kan zijn om deze variant(-en) als dusdanig te categoriseren.

De verticale as van de plot staat daarmee vooral in voor de variatie binnen de officiële standaardtaal in Vlaanderen, het Journaalnederlands: er is zoals aangegeven een onderscheid tussen het schoolse Nederlands van de leerkrachten en het taalgebruik in de politieke debatten. Nu springen daarbij in het bijzonder de specifieke afstanden in het oog (voor de coördinaten van de rode punten lette men op de schaalverdeling aan de boven- en rechterkant van de biplot, aangegeven door de rode streepjes): de cluster o-b-h 'laadt' namelijk gemiddeld gezien op het punt +40 van de verticale as, terwijl de cluster f-g op -60, wat een geraamd verschil van 100 eenheden uitmaakt. Als we daar evenwel de afstand bij betrekken tussen cluster i-j-k-l – zoals gezegd de prototypische representant van het Journaalnederlands in ons corpus – en wat uit onze analyse als de cluster van het Soapvlaams naar voren komt, a-c-d, dan merken we dat beide clusters zich ongeveer op de respectieve posities -25 en +75 (van de horizontale as) bevinden, zodat ook hier het verschil 100 eenheden bedraagt. De variatie *binnen* het Journaalnederlands is met andere woorden even groot als de variatie *tussen* het Journaalnederlands en het Soapvlaams. Dat lijkt ons toch geen onbelangrijke vaststelling: met name, het begrip 'taalkloof' wordt er enigszins mee *uitgehouden*. Geeraerts *et al.*, van wie het begrip afkomstig is (zie hoofdstuk 1), hebben het gelanceerd ter aanduiding van de onmiskenbaar grotere afstand tussen het informele en formele taalgebruik in België ten opzichte van de afstand tussen beide registers in Nederland. Daarbij hebben zij evenwel maar twee registers in ogenschouw genomen (meer bepaald, etalagemateriaal voor het informele taalgebruik en tijdschriftenmateriaal voor het formele – zie Geeraerts, Grondelaers & Speelman 1999: 7-8). Hun registerspectrum is daarmee echter te beperkt om ook eventuele variatie binnen het formeel hoogste register op te merken, waardoor zij dat formele taalgebruik misschien wat te snel gehypostaseerd hebben tot een ongedifferentieerd, monolithisch geheel. Dat er wel degelijk enige variatie voorhanden is, blijkt nu uit onze analyse (voor een soortgelijke studie omtrent deze problematiek, zie Biber 1995). Dat die bovendien van dezelfde grootte is als het verschil tussen het formele en het spontane taalgebruik, wijst er volgens ons op dat we hier eerder te maken hebben met stijldistincties binnen één, enkelvoudige taal (genaamd: het 'Algemeen Belgisch-Nederlands'). Voor het 'officiële' Journaalnederlands is zulks natuurlijk geheel onproblematisch: het



onderscheid tussen vertrouwelijk en beleefd taalgebruik (weerspiegeld in de verticale dimensie van de plot) behoort van oudsher gewoon tot het grammaticale systeem van het Belgisch-Nederlands. In verband daarmee is het wel vermeldenswaard dat de interne differentiatie van het Belgisch-Nederlands en dan voornamelijk het gebrek aan aandacht ervoor in de traditionele taalpolitiek net één van de hetere hangijzers in het hele tussentaaldebat vormt, zoals het volgende citaat van Taeldeman stellig illustreert:

“De normdiscussie in Vlaanderen baadt in een sfeer van variatievijandigheid. Er wordt m.a.w. (te veel) voorbijgegaan aan een fundamenteel kenmerk en een belangrijke functie van (elke levende) taal, nl. dat ze als middel tot sociale + situationele profilering fataal variatie vertoont/moet vertonen. De hele normdiscussie cirkelt in Vlaanderen dus te veel rond een bij voorbaat geamputeerd standaardtaalbeeld.” (Taeldeman 1992: 33)

De gelijke afstand van de gemonitorde, Journaalnederlandse registers tot de spontane registers (a-c-d) verbiedt het ons eerlijkheidshalve echter om voor het Soapvlaams hierop een uitzondering te maken: het Soapvlaams is dan ook een (spontane) spreekstijl die binnen de variatiegrenzen van een algemeen Belgisch-Nederlandse omgangstaal valt, precies zoals het vertrouwelijke en het beleefde jargon dat doen. In het licht daarvan dunkt het ons ten slotte niet langer opportuun om de Vlaamse taalsituatie te beschrijven in termen van een zogezegde *taalkloof*: bij wijze van spreken benoemt het begrip ‘taalkloof’ alleen maar wat het niet betekent.

In die optiek is daarom misschien nog wel het belangrijkste aspect aan de verhouding tussen de Journaalnederlandse (blauwe) en de Soapvlaamse (groene) varianten de naadloze overgang tussen beide groepen: Journaalnederlands en Soapvlaams vormen geen twee aparte clusters, maar ‘raken’ elkaar daarentegen. Ons inziens wordt hiermee die andere opvatting over het Belgisch-Nederlands weerlegd, met name als zou het Vlaamse taallandschap *diglossisch* zijn. Vanuit een diglossiesituatie zouden we namelijk, zoals gezegd, twee afzonderlijke, onderscheiden clusters verwachten (hoewel we terdege beseffen dat een theoretisch begrip zoals diglossie zich wellicht niet zomaar laat reduceren tot simpele frequentieverschillen van welbepaalde taalvarianten; daar staat evenwel tegenover dat zulke meer operationele criteria naar onze mening nu ook weer niet helemaal genegeerd kunnen worden). Het beeld van de Vlaamse taalsituatie dat uit deze distributie van Soapvlaams en Journaalnederlands dan ook oprijst is er kortom inderdaad één van een stilistisch continuüm binnen één en dezelfde taal – die we dus het (Algemeen) Belgisch-Nederlands noemen – dat daarmee de facto als geheel in Vlaanderen de omgangstalige standaard uitmaakt.

De sjibbolets voor het Soapvlaams zijn (zie Figuur 5.3): *ge* (wel en niet-geïnverteerd), *den* en *ne(n)* bij masculiene nomina, de invertete pronomina *–de*, *(e)kik*,

en 'm, en ten slotte de diminutief op *-ke*. Voor het Journaalnederlands zijn dat de corresponderende alternatieven: *je* (niet-inverte subjectvorm, inverte subjectvorm, en objectvorm), *de* en *een* (nogmaals bij masculina), het geïnvverteerde *hij*, en voorts contrasteert het volle *ik* (wel en niet-invert) natuurlijk met de gereduceerde variant(en) van daarnet. De verticale as ondervangt ten slotte het afwijkende gedrag van de – beleefde – subject-*u* (weerom wel en niet-invert), die daarmee de gebruikelijke aanspreekstrategie in de debattaire componenten f en g vormt. Bemerkt daarbij de positie van de object-*u*: dat bevindt zich precies tussen de *ge-* en de *u-*subjecten in, geheel conform de polysemie van dat pronomen (*u* is objectvorm bij het paradigma van zowel de eerste als de tweede). Op de object-*me* komen we zo meteen terug.

Vanaf *adj.n.m* (de adjectieven op *-en* bij masculiene nomina) begint er bij de Soapvlaamse varianten een soort 'hoefijzer' van vormen waarvan het effect enerzijds minder uitgesproken is, maar die anderzijds nog voldoende van de overige varianten verschillen om als representanten van het Soapvlaams te kunnen gelden – getuige bijvoorbeeld de afstand tot de cluster in het centrum (Figuur 5.4). Het gaat meer bepaald over: het subject-*me* (wel en niet-geïnvverteerd), *den* en *ne(n)* maar nu bij commune evenals feminiene nomina (zoals nog zal blijken is de heterogeniteit bij het masculiene genus in haar geheel groter dan bij de overige twee genera; opvallend is alleen dat dit effect sterker is voor de lidwoorden dan voor alle andere determinatoren), voorts zowel *dieje(n)* als *dienne(n)* bij masculina (dat beide flexiealternatieven van het demonstrativum *die* dezelfde distributie vertonen – iets wat bij de twee andere genera eveneens opduikt; zie verderop – is toch ook geen onbelangrijke vaststelling, aangezien elk een andere dialectische achtergrond heeft: de vorm op *-je(n)* is Brabants, die op *-ne(n)* Vlaams – vergelijk hoofdstuk 3), en verder zijn er nog de aanspreekvormen *gij* (wel en niet-invert) en *-degij*, de possessiva *zijne(n)*, *mijne(n)*, en *hare(n)* bij masculina, het object-*ze* voor 3.enk.fem., en ten slotte *mekaar*. Dat de *mij*-vormen (zowel de objectvorm als het reflexivum) eveneens in deze buurt voorkomen heeft er bij nader inzien mee te maken dat het onderscheid tussen vertrouwelijk en beleefd taalgebruik langs de verticale dimensie zonder meer (ook) de vorm aanneemt van een differentiatie van fonetische volle en gereduceerde vormen; een feit waarop de object-'*m* dan weer een uitzondering vormt – zie verderop. Zo is reflexieve *me* bijvoorbeeld net aan de bovenkant van Figuur 5.4 te vinden (over *mijn eigen* aanstonds meer). Naast *me.ref* treffen we nog het gereduceerde possessivum *m'n* aan (en nog wel voor de drie genera), terwijl de volle vormen onderaan de plot staan (links boven *mekaar*; *mijn.m* bevindt zich weliswaar iets meer naar boven, in de buurt van *geen.m* en *een.c*, maar toch nog steeds onder *m.n.m* – zie Figuur 5.5). Verder zien we – rechts van *m'n* – ook de beide varianten van *we* (wel en niet-invert); de volle *wij*-vormen vinden we helemaal onderaan Figuur 5.4. Hetzelfde geldt voor de vier varianten van het subject-*ze*: *ze.not.vr*, *ze.inv.vr*, *ze.not.mv*, en *ze.inv.mv* (net onder *we*): weer staan de corresponderende *zij*-vormen onderaan Figuur 5.4 (in de linkerhoek). Ter hoogte van de *ze*-vormen zien we ten slotte nog de drie varianten van *z'n* (*z.n.m* staat daarbij iets meer naar links, bij

*haar.m*): de volle varianten zitten in het kleine clustertje van zowel Journaalnederlandse als Soapvlaamse vormen bij *onzen.f* (zie Figuur 5.6 – ook hier is het masculiene *zijn* iets meer naar boven te vinden, deze maal tussen *de.c* en *de.f*). Dat er zich tussen al deze gereduceerde vormen in Figuur 5.4 verder ook nog de hele reeks volle J-vormen bevinden – te weten: *jij* (wel en niet-invert), *jou*, subject-*jullie* (ook weer wel en niet-invert), en object-*jullie* – is dan ook alleen maar een schijnbare tegenspraak: de gereduceerde alternatieven op *je* staan zoals aangegeven namelijk helemaal bovenaan de verticale as (het reflexieve *je* is daarbij van een andere orde aangezien het concurreert met vormen zoals *u*, *uw eigen*, *je eigen*, en *zich* – zie verderop). Blijkbaar laadt het onderscheid vol-gereduceerd met andere woorden voor de J-vormen (in hun geheel) hoger op die verticale dimensie dan voor de andere pronomina. Dat geldt a fortiori voor de J-vormen *onderling*: voor de drie genera zitten de volle possessiva (*jouw* en *jullie*) namelijk in de cluster in het centrum (zie Figuur 5.6), en dus onder de gereduceerde *je*-alternatieven – die zich bij de volle vormen van de pronomina ophouden. Dat *elkaar* voorts in de buurt van de J-pronomina verschijnt, relateren we, in het licht van de positie van *mekaar*, dan weer met het verschil in formaliteit dat er tussen beide varianten bestaat (omgekeerd bewijst dit dan dat onze beslissing om *elkaar* en *mekaar* als respectievelijk Journaalnederlands en Soapvlaams te labelen terecht was). Daarmee samenhangend is ook de positie van het inverté *ie* (ter hoogte van *wij.not* en *wij.inv*) geen uitzondering – het volle *hij.inv* zagen we bijvoorbeeld net helemaal bovenaan bij de *je*-cluster – het moet namelijk concurreren met *m.inv* (en is daarbij op de koop toe sowieso een nogal ‘onvlaamse’ reductievorm, wat enig afwijkend gedrag eens te meer begrijpelijk maakt).

Van de kleine cluster onder de centrumcluster leert Figuur 5.5 ons dat het hier voornamelijk om de bezittelijke voornaamwoorden van de derde persoon enkelvoud gaat. Dat is trouwens een algemeen gegeven: de possessiva van alle andere personen – en alle andere determinatoren tout court – bevinden zich in de centrumcluster; de enige uitzondering is *zijne.n.f* dat eveneens in die centrumcluster zit (zie Figuur 5.6). Daarnaast zien we in dit groepje ook nog de twee reductievormen die we bij de bezittelijke voornaamwoorden voor het masculiene genus gedefinieerd hebben: *m.ne.n.m* en *z.ne.n.m*. De corresponderende reductievormen voor het commune en het feminiene genus zitten daarbij allemaal in de centrumcluster (vergelijk Figuur 5.6 – we herhalen dat er in ons corpus geen enkele attestatie is van *z'ne(n)* met een feminien nomen). Zulks reflecteert dan een tweede algemeen gegeven, met name dat de verticale dimensie van de plot differentieert tussen fonetisch gereduceerde vormen (boven) en volle vormen (onder): het volle *mijne.n.m* en *zijne.n.m* zijn we namelijk nog onderaan in de hoefijzercluster tegengekomen. Dat het trouwens de masculiene vormen zijn die buiten de centrumcluster vallen illustreert tot slot het zo-even (bij de lidwoorden) vermelde principe – dat daarmee terecht voor een derde algemeen gegeven kan doorgaan – dat het effect in spreiding

het sterkste is bij het masculiene genus. Zo zien we hier bijvoorbeeld *hare.n.c* en *hare.n.f* opduiken, waar *hare.n.m* nog in de hoefijzercluster stond.

Verder zien we nog enkele varianten van het pronomen *ze*. Het gaat meer bepaald over de stapelvorm *ze* voor vrouwelijk enkelvoud (*ze.zij.vr*), het gereduplicateerde *ze* voor vrouwelijk enkelvoud (*ze.dbl.vr*), en het object-*ze* voor meervoud (*ze.obj.mv*). De enkelvoudig-vrouwelijke pedant van die laatste hebben we beneden nog aangetroffen in de hoefijzercluster (zie Figuur 5.4). Voeg daar nu bij dat alle andere Soapvlaamse varianten van het meervoud-*ze* – te weten: de stapelvorm *ze.zij.mv*, het reduplicatieve *ze.dbl.mv*, en verder de volle variant *zij.dbl.mv* – in de centrumcluster zitten (zie Figuur 5.6), dan moeten we toch vaststellen dat de variatie merkbaar groter is voor 3.enk.fem dan voor 3.mv (van het volle *zij* voor 3.enk.fem wijzen we er nogmaals op dat er in het corpus geen treffers van voorkomen). Daar valt voorts nog bij op hoe anders beide object-*ze*'s zich gedragen in vergelijking met de subject-*ze*'s, en dit in weerwil van de gebruikelijke labeling van deze vormen (door standaardgrammatica's zoals de ANS bijvoorbeeld – zie ANS: 243) als weliswaar typisch Zuid-Nederlands maar daarmee nog niet substandaard: de object-*ze*'s gedragen zich onmiskenbaar als Soapvlaamse varianten (ook dat corroboreert onze eerdere beslissing om *ze* bij die variëteit onder te brengen; verder lijkt dit bij uitbreiding eventueel ook een verklaring te bieden voor het gedrag van *m.obj* – vergelijk supra). Ten slotte constateren we dat *ze.obj.mv* een rechtstreekse concurrent dicht bij zich in de buurt krijgt: het object-*hun*. Het verschil met *hen* (ter hoogte van *dim.j* – Figuur 5.4) is duidelijk; de conclusie kan dan ook alleen maar luiden dat *hen* inderdaad de meest formele variant van de drie is (zoals de standaardgrammatica's specificeren – zie ANS: 248), maar dat *hun* en *ze* in Vlaanderen nagenoeg synoniem zijn.

Bij *uwe.n.m* (dit is: *uwe(n)* bij masculiene nomina), en *u.ref* (het reflexieve en dus intrinsiek substandaardtalige *u*) begint wat we de 'centrumcluster' hebben genoemd. Die bestaat voor het overige uit determinatoren, reflexiva, en – tenminste wat de Soapvlaamse varianten ervan betreft – gereduplicateerde vormen (zie Figuur 5.6). Daarbij vinden we alle aangehaalde tendensen erin terug. Zo is de spreiding bij de determinatoren weer het grootste voor het masculiene genus; vergelijk bijvoorbeeld *hunne.n.m* (net boven *u.ref*) en *uwe.n.m* (net eronder) met hun corresponderende alternatieven in het midden van de cluster – bij *dezen* en *onzen* lijkt er meer aan de hand te zijn, maar dat effect is twijfelachtig: de varianten *dezen.c*, *dezen.f*, en *onzen.f* tellen elk maar één luttel observatie in de dataset (net zoals *iederene.c* trouwens) wat hun precieze positie natuurlijk erg onbetrouwbaar maakt. Het is ook hier dat we de *die*-varianten voor commune en feminiene nomina aantreffen: *dieje.n.c*, *diene.n.c*, *dieje.n.f*, en *diene.n.f* – waarbij het nogmaals aangestipt zij dat er geen noemenswaardig verschil is tussen de Brabantse *-je(n)*-vormen en de Vlaamse *-ne(n)*-vormen. Daarnaast zien we – bij de reduplicaties – opnieuw de tweedeling tussen fonetisch gereduceerde en volle varianten: *we.dbl* staat hoger dan *wij.dbl*,

*k.dbl* (rechts boven *u.ref*) staat hoger dan *ik.dbl* (in de cluster van possessiva onder de centrumcluster; net achter *zijne.n.c* – zie voor deze vier varianten Figuur 5.5). Voorts staat *ge.dbl* nog enigszins hoger dan *gij.dbl*, maar *ze.dbl.mv* staat echter ontegensprekelijk lager dan *zij.dbl.mv* (zie voor deze vier varianten Figuur 5.6). Voor de laatste vier vormen lijkt het onderscheid gereduceerd-vol zich eerder langs de horizontale as te situeren. Dat zou dan verder ook stroken met het contrast tussen *ik* en 'k, dat zich eveneens langs deze dimensie beweegt (zie supra). Misschien moet daarom de hypothese geopperd worden dat het proces van fonetische reductie niet 'lineair' is maar voor verschillende pronomina andere determinanten kent.

Zoals gezegd treffen we in de centrumcluster de geflecteerde en dus Soapvlaamse vormen van het bezittelijke *uw* aan – *uwe.n.c* helemaal bovenaan Figuur 5.6 bij *gene.n.f*; *uwe.n.f* links onderaan tussen *gij.dbl* en *uw.eigen*. De ongeflecteerde, Journaalnederlandse alternatieven bevinden zich heel wat meer naar beneden – in Figuur 5.4 links boven *mij.ref*, ter hoogte van *zij.not.vr* en *zij.inv.vr* staan ze alle drie. Zulks is dan een duidelijk verschil met alle andere determinatoren, waar de Journaalnederlandse varianten meer naar links in plaats van naar beneden te vinden zouden zijn. Dat *uw* toch deze distributie vertoont, heeft natuurlijk te maken met de polysemie die ook de positie bepaalde van *u.obj* tussen de beleefde *u*-subjecten en de vertrouwelijke *ge/gij*-subjecten in. Toch moet het opgemerkt worden dat dit effect voor het possessivum klaarblijkelijk minder sterk is, getuige daarvan de centrale en dus weinig uitgesproken positie van deze vormen. Dat de Soapvlaamse flexievormen zelfs in de dichte centrumcluster voorkomen – en dus niet eens in de hoefijzercluster, bijvoorbeeld – is daarvoor alleen maar des te kenmerkender. Soortgelijke bedenkingen kunnen trouwens eveneens gemaakt worden ten aanzien van *ikke*: hoewel het vaak wordt aangehaald als het voorbeeld bij uitstek van het gebrek aan registergevoeligheid bij de modale Vlaamse taalgebruiker, bewijst zijn weinig geprononceerde positie in de centrumcluster (net onder *diene.n.f* – Figuur 5.5) dat die reputatie veruit overroepen is.

Een andere groep van varianten waarvan het effect zo zwak is dat ze in de centrumcluster zitten zijn ten slotte de reflexiva – vergelijk bijvoorbeeld de verhouding van *mijn.eigen* met die van *me.ref* en *mij.ref*. Het verschil tussen *u.ref* enerzijds en *je.eigen* en *uw.eigen* anderzijds (die laatste bevindt zich links onderaan, net boven *ze.zij.mv*) is groter dan dat tussen die laatste twee, en opmerkelijk analoog aan dat tussen *eigen.3* (helemaal links in Figuur 5.5) en *zich.3* (bijna in het midden van de centrumcluster – Figuur 5.6). Tekenend is tot slot de positie van *je.eigen* dat als enige Soapvlaamse variant het nulpunt overschrijdt en in de negatieve, Journaalnederlandse helft van de horizontale as gelokaliseerd is. Natuurlijk is zulks evenwel begrijpelijk vanuit de morfologische geleiding van de vorm: aan de ene kant is het een Soapvlaams reflexivum op *eigen*, maar aan de andere kant is het als constructie opgebouwd uit een Journaalnederlandse J-vorm.

Hiermee zijn we dan in staat om enkele conclusies te formuleren met betrekking tot de registervariatie in het gesproken Belgisch-Nederlands. We hebben gezien dat het onderscheid tussen Journaalnederlands en Soapvlaams langs de horizontale as van de plot loopt, de dimensie die de spontane componenten a-c-d afzet tegen de meer gemonitorde componentclusters o-b-h, i-j-k-l, en f-g (zoals gezegd zijn m en n niet specifiek met één van beide variëteiten geassocieerd). Opmerkelijk daarbij is dat dit onderscheid de vorm aanneemt van een continuüm, wat de bekende diglossiethese zonder meer weerlegt. Er zijn louter enkele uitschieters, die dan ook terecht geboekstaafd staan als sjibboletvormen (voor het Soapvlaams, met name: de G-aanspreekvormen, het subject-‘m, en de diminutief op –ke). Verder blijkt nog bij de determinatoren de variatie voor het masculiene genus groter te zijn dan voor het commune en het feminiene genus (die zich op hun beurt overigens analoog gedragen). Op de verticale as wordt er vervolgens onderscheiden tussen het ‘schoolse’ taalgebruik uit componenten o-b-h enerzijds en de ‘debattaire’ spreekstijl in f-g anderzijds – met de representant bij uitstek van het Journaalnederlands, cluster i-j-k-l, daar precies tussenin. Van belang daarbij is dat de gemiddelde afstand tussen o-b-h en f-g nagenoeg dezelfde is als die tussen i-j-k-l en a-c-d op de horizontale dimensie, zodat er van een eventuele ‘taalkloof’ in het Belgisch-Nederlands geen sprake kan zijn: de variatie binnen het Journaalnederlands is namelijk even groot als de variatie tussen het Journaalnederlands en het Soapvlaams. Wat de taalvarianten betreft onderscheidt de verticale as verder nog twee distincties: ten eerste die tussen de J- respectievelijk de U-vormen, en ten tweede die tussen de fonetisch gereduceerde tegenover de volle realisaties. Met dat alles kan deze dimensie niet anders geïnterpreteerd worden dan als de grammaticale parameter vertrouwelijk-beleefd (tenminste, als we die in ‘structurele’ zin opvatten). Zulks betekent evenwel dat *niet* die parameter het gebruik van het Soapvlaams bepaalt: de dimensie die dat wel doet staat – letterlijk en figuurlijk – ‘loodrecht’ op het onderscheid tussen vertrouwelijk en beleefd.

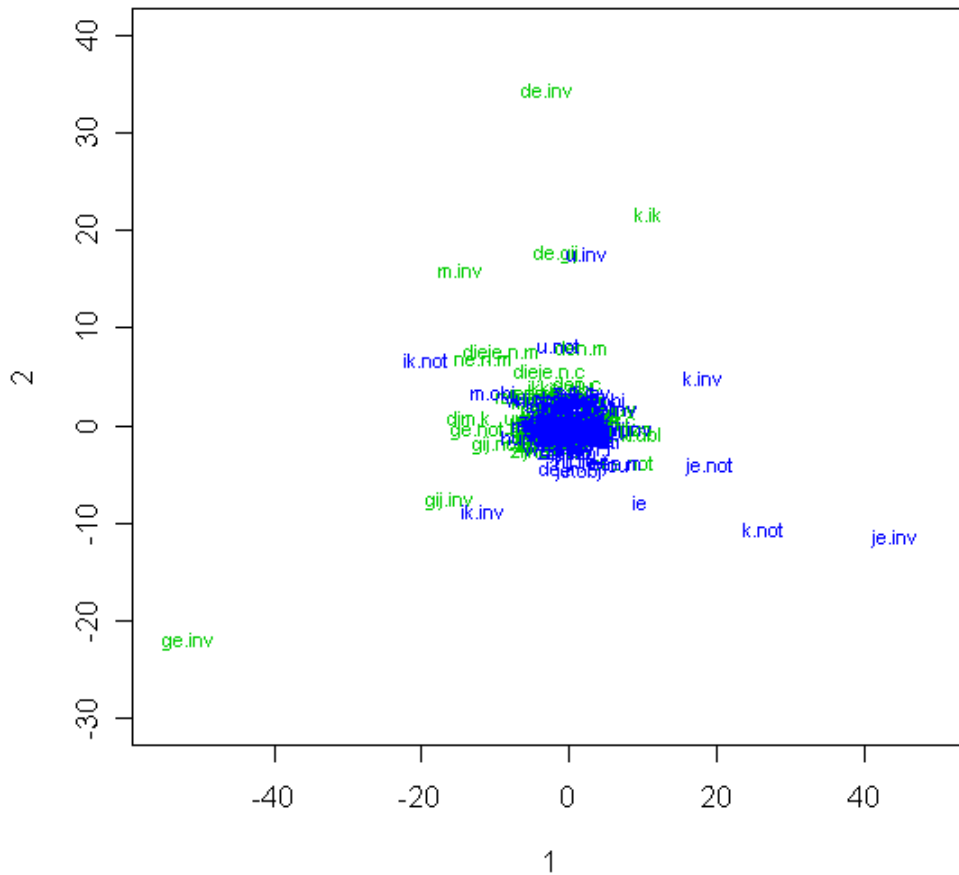
## 5.2 Regio

Voor de bestudering van de regiovariatie baseren we ons zoals gezegd op het geboortjaar van de sprekers. Het CGN definieert daarbij vier categorieën (zie 2.2.3):

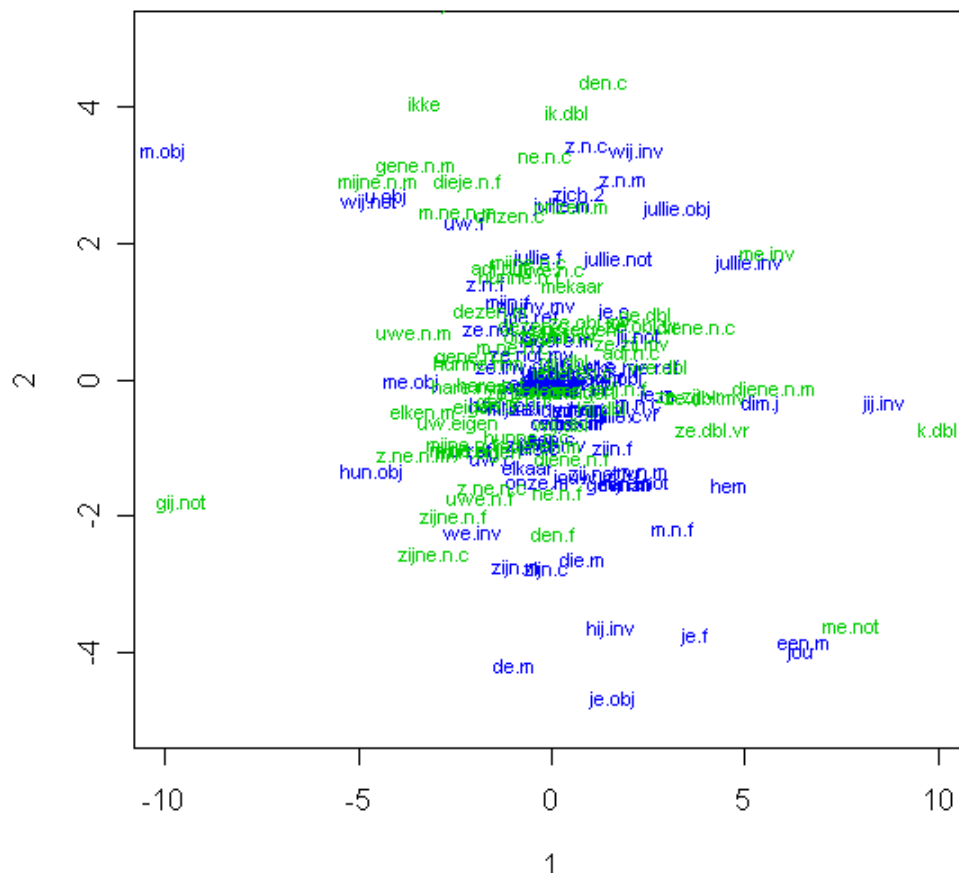
Label	Regio
bra	Vlaams-Brabant & Antwerpen
ovl	Oost-Vlaanderen
wvl	West-Vlaanderen
lim	Limburg

Tabel 5.2: De vier regio's in het CGN



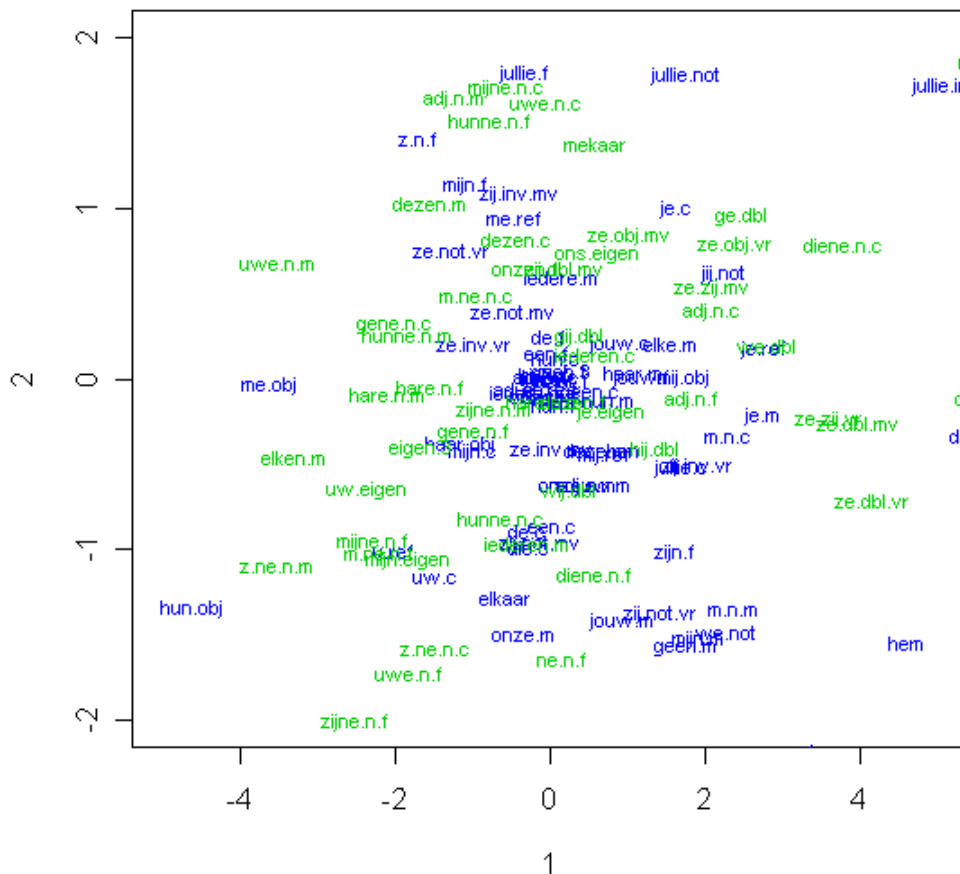


Figuur 5.8: Regio – varianten: detail [-50;50] horizontaal en [-30;40] verticaal

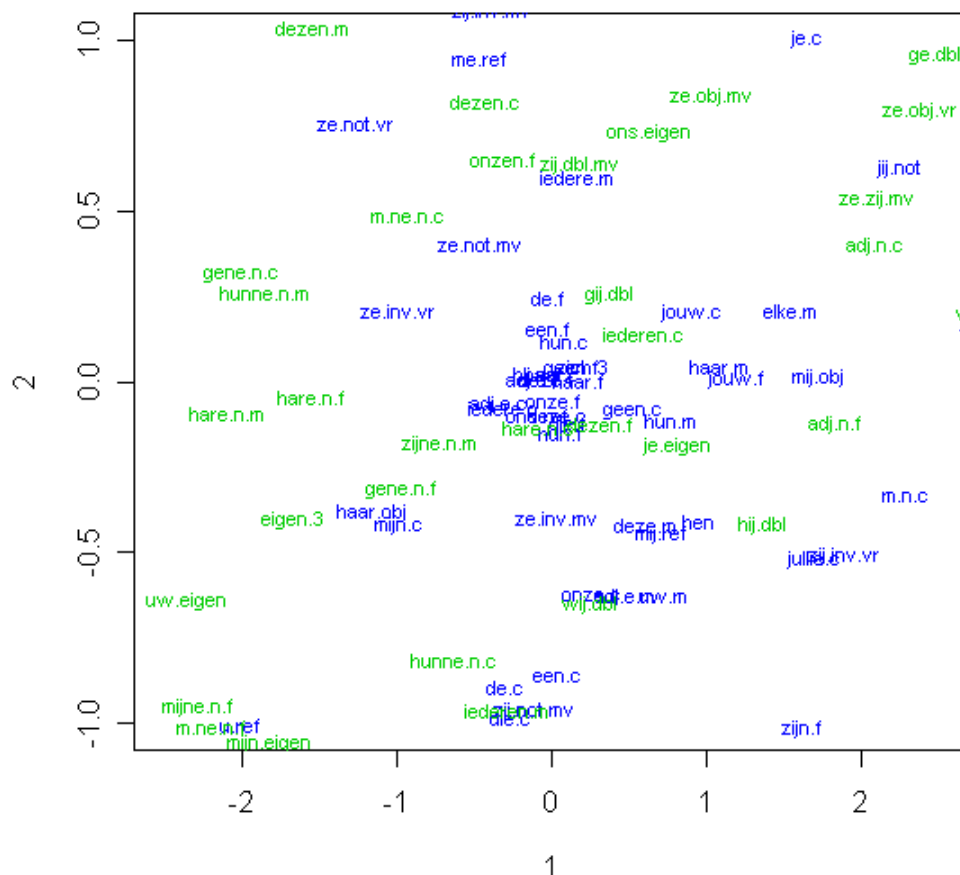


Figuur 5.9: Regio – varianten: detail [-10;10] horizontaal en [-5;5] verticaal





Figuur 5.10: Regio – varianten: detail [-5;5] horizontaal en [-2;2] verticaal



Figuur 5.11: Regio – varianten: detail [-3;3] horizontaal en [-1;1] verticaal

Figuren 5.8, 5.9, 5.10, en 5.11 geven meer details omtrent de varianten – opnieuw zijn de Soapvlaamse vormen groen, en de Journaalnederlandse blauw. Het blijkt dan dat de regio's vooral van elkaar verschillen in hun gebruik van de inverte aanspreekvormen: het enclitische *-de* is Brabants, *ge* is Limburgs, en *je* is (West-)Vlaams. Die laatste corroborereert dan de stelling van Vandekerckhove (zie hoofdstuk 1) over de paradox die West-Vlaanderen met betrekking tot het standaardisatieproces belichaamt: sommige West-Vlaamse dialecten hebben namelijk historisch gezien al een – fonetisch gereduceerde – J-vorm in aanspreekfunctie; de associatie van West-Vlaanderen met een J-vorm wijst er dan ook *niet* op dat deze regio meer gestandaardiseerd zou zijn dan de overige drie regio's, wel integendeel. Verdere evidentie voor Vandekerckhoves these kan daarbij nog gevonden worden in het feit dat West-Vlaanderen bijvoorbeeld in het geheel niet geassocieerd is met de volle J-vormen *jij* en *jullie*: die bevinden zich namelijk meer aan de bovenkant van de cluster (zie Figuren 5.9 en 5.10).

Voorts moet er nog op gewezen worden dat sterke associatie van Limburg met *ge* – en in mindere mate *gij* – vooral een 'negatief' effect is: het betekent niet dat *ge* en *gij* niet gebruikt zouden worden in Brabant of de Vlaamse provincies – de attestaties uit het corpus leren namelijk dat zulks wel degelijk het geval is – maar wel dat er in Limburg *uitsluitend* van *ge* en *gij* gebruik gemaakt wordt (de Brabantse *-de* en de Vlaamse *je* zijn zoals gezegd dialectvarianten typisch voor hun regio). Dat uit zich bijvoorbeeld in het feit dat de associatie van Limburg met de G-vormen zich enkel tot inversie beperkt: dat is net de positie waar het moet concurreren met de Brabantse *-de* en *-degij*. De corresponderende vormen voor de niet-geïnverteerde positie – te weten: *ge.not* en *gij.not*, waarvoor Brabant met andere woorden geen extra varianten heeft – staan dan ook heel wat hoger in de plot (de Vlaamse *je*-vormen zijn nog anders omdat die wel kunnen voorkomen in niet-inverte positie, wat dan ook af te lezen is aan de coördinaten van *je.not*).

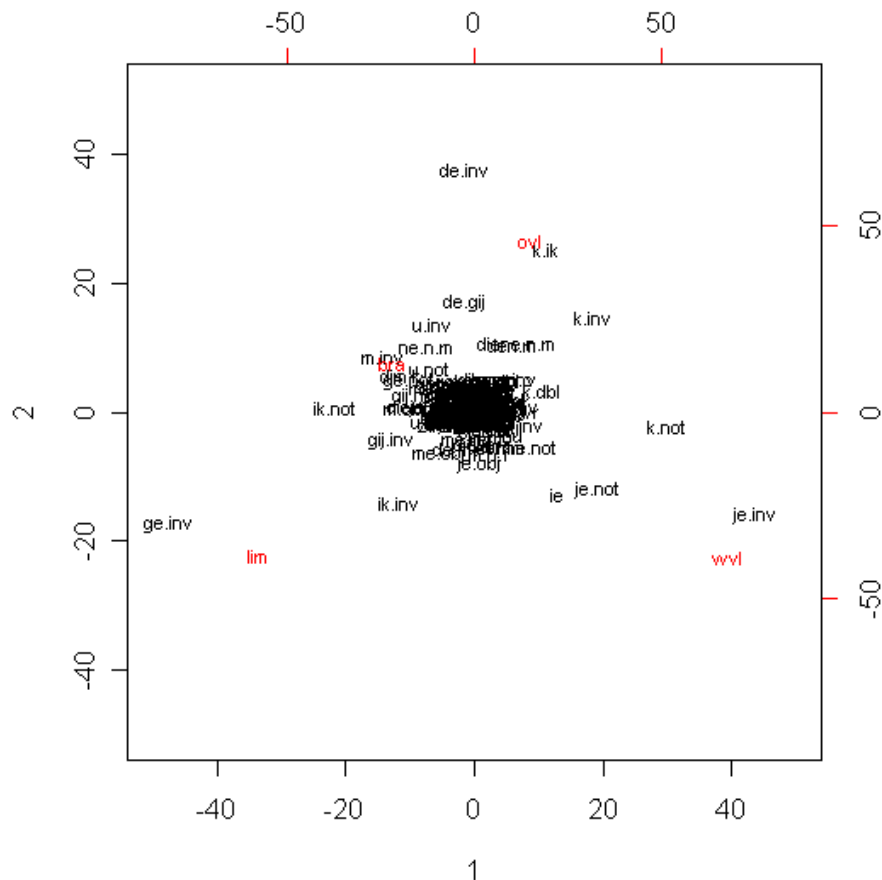
De aanspreekvormen zijn echter niet de enige markers van een regio. Zo stellen we vast dat het gereduceerde 'k – dat zich in het registerspectrum nog zo afwijkend gedroeg; zie 5.1 hierboven – vooral een Vlaams verschijnsel is, en weer zowel in wel als niet-inverte positie. Daarmee lijkt deze variant vooral de sjwaloze [k] te coderen, die inderdaad typisch is voor sommige Vlaamse dialecten (zie: de Schutter, van den Berg, Goeman & de Jong 2005). De beide volle *ik*-vormen zijn dan 'niet-Vlaams' (in de zin van: Brabants en Limburgs). De stapelvormen *-degij* en *(e)kik* zijn dan weer duidelijk Brabants. Tot slot weerspiegelt de plot ook de alternantie tussen de Brabantse 'm (linksboven) en de (West-)Vlaamse *ie* (rechtsonder).

In de centrumcluster kunnen we verder nog enkele tendenzen ontwaren. Zo valt op dat het leeuwendeel van de groene punten aan de 'niet-Vlaamse' linkerzijde van de plot ligt. Vrijwel de enige Soapvlaamse varianten aan de rechterkant van de plot zijn

die van de demonstrativa *diene(n)*, de subjecten op *me* (1.mv), en de reduplicatievormen (enkel *ik.dbl* lijkt om onverklaarbare redenen een uitzondering). Voor Brabant en Limburg blijven er met andere woorden de diminutief op *-ke*, het pronomen *ikke*, de reflexiva op *eigen*, en de hele reeks geflecteerde determinatoren over. Zulks is natuurlijk allemaal conform de dialectische achtergrond van deze vormen. Vooral in de verhouding tussen de *die*-varianten komt dat goed tot uiting, aangezien de Brabantse *dieje(n)*-vormen inderdaad meer naar boven te vinden zijn (zie Figuur 5.8 en 5.9).

Hoe dan ook, het feit dat de overgrote meerderheid van alle varianten – zowel Journaalnederlandse als Soapvlaamse – in het centrum van de plot samenklitten wijst zoals gezegd op een erg gehomogeniseerde distributie. In het licht van de hele problematiek rond de standaardisatie van Vlaanderen is dat natuurlijk een belangrijk gegeven. Het beduidt namelijk dat niet alleen de standaardisering in Vlaanderen al behoorlijk ver gevorderd is, maar dat bovendien ook veel Soapvlaamse elementen – ondanks hun normatief gestigmatiseerde status – net deel uitmaken van dat proces. In dat verband dienen we daarom nog enige aandacht te wijden aan de rol van de Brabantse regio als taalcentrum van de verbreiding van de Soapvlaamse vormen (de theorie van de ‘Brabantse expansie’ – zie hoofdstuk 1). Eén van de implicaties van die theorie is namelijk dat de linguïstische invloed van Brabant het sterkst voelbaar zal zijn in de aangrenzende provincies. De opdeling van de plot in een ‘Vlaamse’ rechter- en een ‘niet-Vlaamse’ linkerkant lijkt nu één en ander te bevestigen: Limburg zit met Brabant in de linkerhelft van de plot, en Oost-Vlaanderen hadden we al aangeduid als een overgangszone tussen Brabant en West-Vlaanderen; daarmee ondergaat de enige niet aan Brabant grenzende provincie in Vlaanderen inderdaad het minste Brabantse invloed.

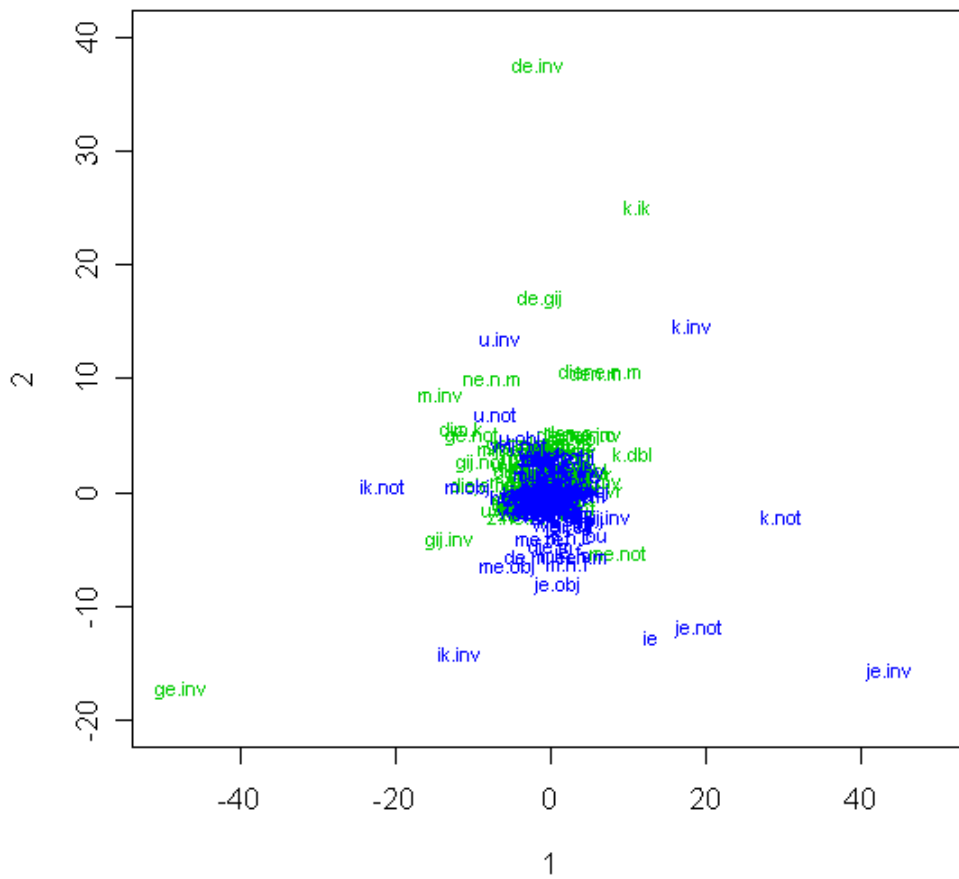
Als afsluiter willen we nog even wijzen op een verschil bij woonregio, de regio waar de spreker op het tijdstip van de opname van het gesprek woonachtig is (gecodeerd als de variabele *resRegion*). Zoals Figuur 5.12 aantoon, leveren de woonregio’s vrijwel hetzelfde beeld op als de geboorteregio’s, behalve voor het Brabantse gebied:



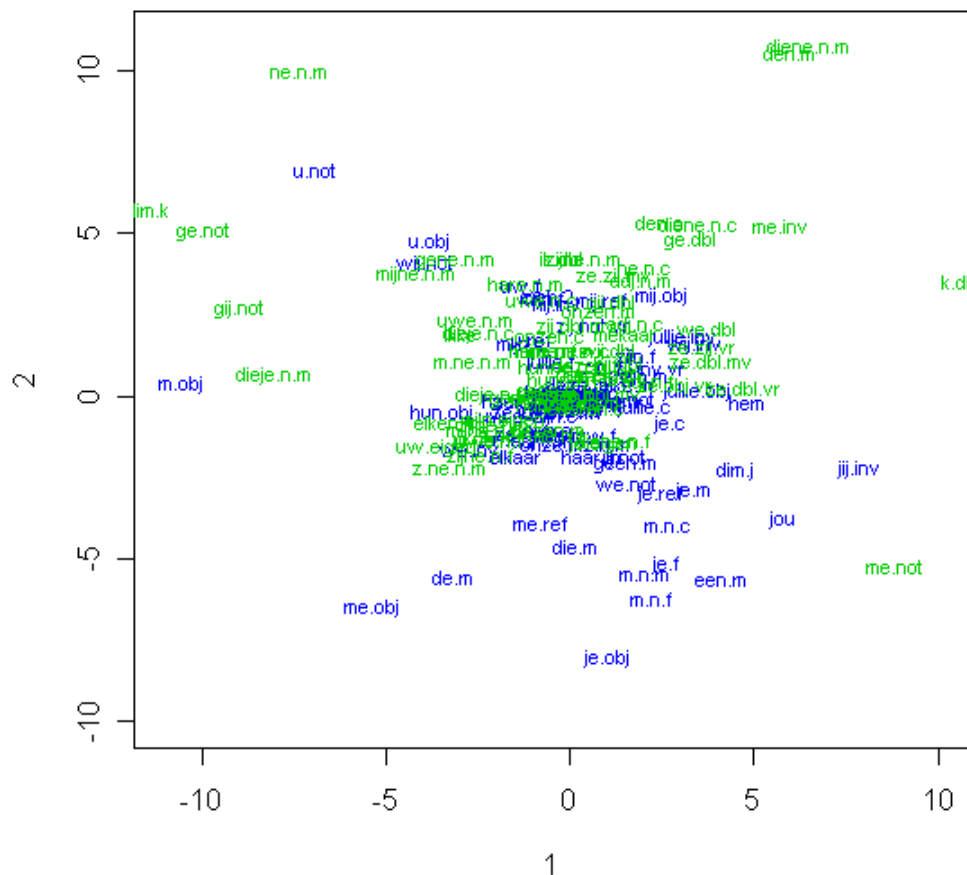
Figuur 5.12: Biplot – woonregio

Brabant is verschoven naar het centrum van de plot (waardoor omgekeerd de status van Oost-Vlaanderen als overgangsgebied des sterker naar voren komt). Dat lijkt ons met name geïnterpreteerd te kunnen worden in het licht van de migratiestromen zoals Vlaanderen die na de Tweede Wereldoorlog gekend heeft. De Brabantse regio is namelijk niet alleen het geografische zwaartepunt van Vlaanderen, maar ook het economische en culturele (zie Goossens 1970 en Geerts 1983 en de referenties aldaar), en geniet dan ook een zeker prestige. Nu manifesteert zich zulks ondermeer in een inwijking van Vlamingen uit de meer perifere regio's. Brabant wordt op linguïstisch vlak daardoor een potpourri: de taal die in het Brabantse gebied te horen valt is steeds minder zuiver Brabants, maar meer en meer een kakofonie van alle Belgisch-Nederlandse dialecten (we veronderstellen even dat de 'nieuwelingen' hun oorspronkelijke tongval grotendeels behouden – daarmee gaan we echter gelijk uit van een nogal belangrijke hypothese: met name, dat het prestige van de Brabantse regio blijkbaar toch niet zo groot is dat er massale overname plaatsvindt). Als plek om te wonen etaleert Brabant met andere woorden de algehele dialectverscheidenheid van Vlaanderen: "Brabant is een talige miniatuurversie van Vlaanderen". In de plot uit zich dat dan in een centrale(-re) positie.

Bij de taalvarianten valt er verder niets op dat grondig verschilt van wat we bij de geboorteregio vastgesteld hebben. Figuren 5.13, 5.14, 5.15, en 5.16 geven we daarom alleen maar voor de volledigheid.



Figuur 5.13: Woonregio – varianten: detail [-50;50] horizontaal en [-20;40] verticaal



Figuur 5.14: Woonregio – varianten: detail [-10;10] horizontaal en [-10;10] verticaal



### 5.3 Decennium

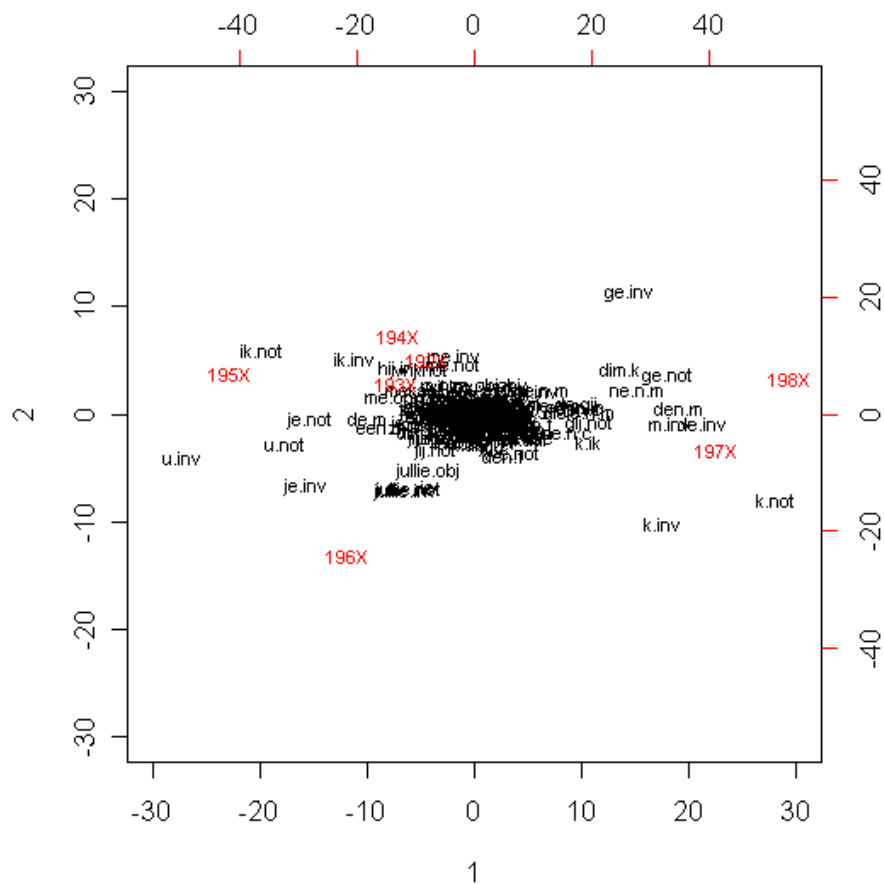
De derde variabele die we willen bestuderen is leeftijd. Het CGN geeft hiervoor het geboortejaar van de spreker, en dus zullen we cohorten afbakenen (zie 2.2.2). Daartoe zullen we meer bepaald twee verschillende schema's hanteren: één in termen van het decennium waarin de sprekers geboren zijn, en een ander in termen van hun sociologische generatie – de details omtrent dat laatste komen in de volgende paragraaf (5.4) aan bod; hier behandelen we het eerste schema.

In het CGN is de oudste spreker geboren in 1917 en de jongste in 1987. Omdat de spreker uit 1917 evenwel de enige is uit het decennium 1910-1919, groeperen we hem bij het volgende decennium. Op die manier kunnen er zeven decennia onderscheiden worden:

<b>Label</b>	<b>Decennium</b>
192X	1920-1929
193X	1930-1939
194X	1940-1949
195X	1950-1959
196X	1960-1969
197X	1970-1979
198X	1980-1989

*Tabel 5.3: De zeven decennia op basis van het geboortejaar*

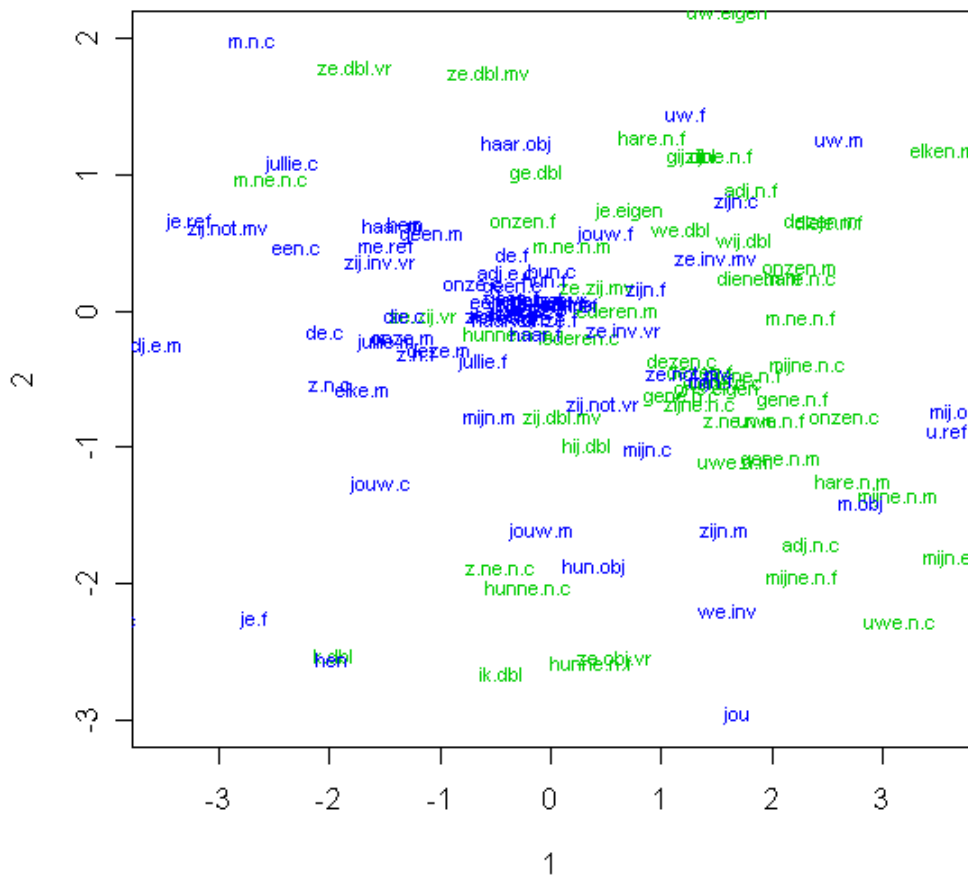




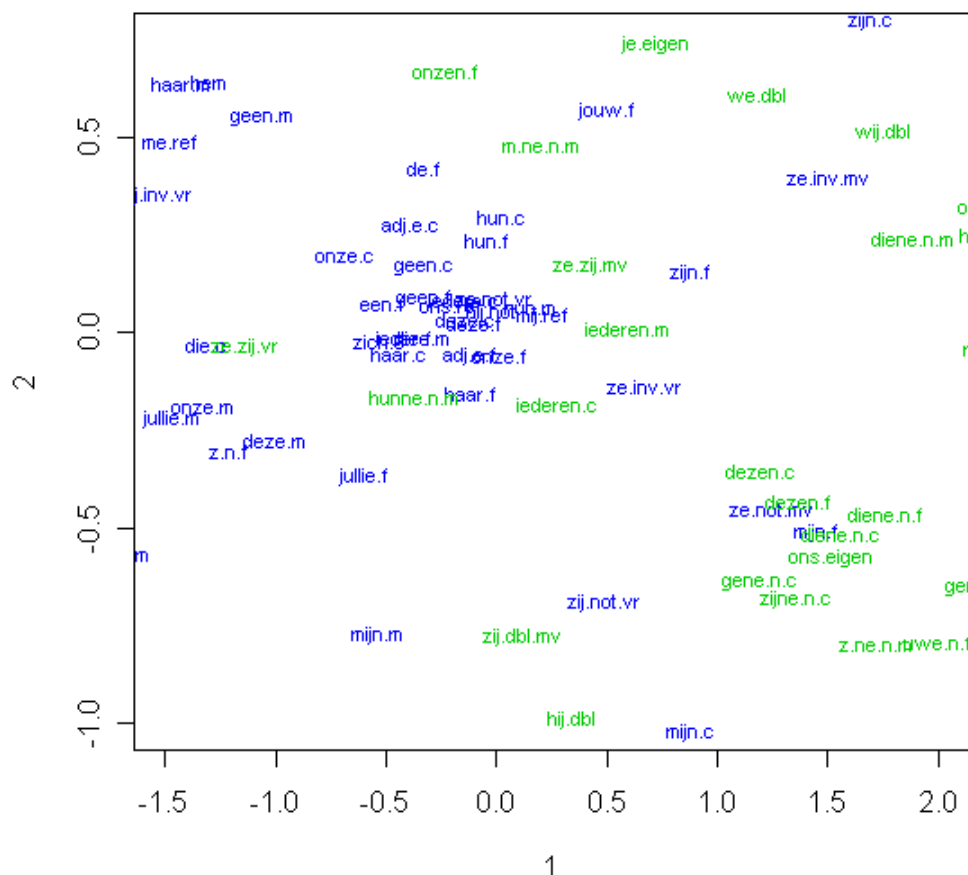
Figuur 5.17: Biplot – decennium

Figuur 5.17 geeft de biplot van de decennia. We bemerken daarbij een nagenoeg lineaire trend, zowel wat de taalvarianten als wat de decennia zelf betreft (alleen 6X wijkt om een onverklaarbare reden wat naar beneden af; zie evenwel 5.4 verderop). Wat die laatste betreft vinden we de oudste drie cohorten (dit is dus: 2X, 3X, en 4X) in het centrum van de plot, vanaf de jaren 50 verschuift het dan naar rechts, om in de jaren 70 helemaal naar de linkerkant om te slaan.





Figuur 5.20: Decennium – varianten: detail [-4;4] horizontaal en [-3;2] verticaal



Figuur 5.21: Decennium – varianten: detail [-1,5;2] horizontaal en [-1;1] verticaal

De plots van de taalvarianten (Figuren 5.18, 5.19, 5.20, en 5.21) onthullen wat er daarbij linguïstisch gezien aan de hand is. We stellen dan vast dat – los van het centrum van de plot, waar er nogal wat overlapping is – de Journaalnederlandse blauwe varianten allemaal links van de plot te vinden zijn, en de Soapvlaamse groene varianten rechts. Net zoals bij de componenten (5.1) vormen de gereduceerde 'k-subjecten ook hier de uitzondering op de regel, door zich als Soapvlaamse elementen te gedragen. Verder zien we weer een geleidelijke overgang tussen Journaalnederlands en Soapvlaams (in plaats van een strikte scheiding), wat zoals gezegd een indicatie voor standaardisering is. Opnieuw is er daarbij louter sprake van 'uitschietende' sijbboletvormen (en dat is eveneens dezelfde set – voor het Soapvlaams: wel en niet-invert *ge* en zelfs *–de*, de diminutief op *–ke*, *den* bij masculina evenals *ne(n)*, en tot slot de subject-'*m*'). De lidwoorden zijn evenwel de enige varianten waar de spreiding bij het masculiene genus groter is dan bij de andere twee genera: blijkbaar is dit effect voor de leeftijdsvariabele niet van kracht (hetzelfde geldt voor het onderscheid tussen fonetisch gereduceerde en volle vormen, wat er natuurlijk mee te maken heeft dat er hier geen verticale dimensie is die vertrouwelijk van beleefd distingeert).

Sprekers uit de jaren 50 en 60 maken kortom veelvuldig gebruik van het Journaalnederlands; die uit de jaren 70 en 80 eerder van het Soapvlaams – alle oudere sprekers gedragen zich ten opzichte van beide variëteiten verder neutraal. Op grond van de socio-economische condities die de maatschappelijke status van deze cohorten gedurende de tweede helft van de twintigste eeuw bepaald hebben interpreteren we één en ander als een confirmatie van de in de literatuur (zie hoofdstuk 1) geschetste verklaring aan de hand van emancipatie. De oudste sprekers (geboren in de jaren 20, 30 en 40) zijn voornamelijk diglossische sprekers: als zij standaardtaal proberen te praten, dan is hun taal eerder een mengeling van Journaalnederlands en hun oorspronkelijke dialect – 'tussentaal' in de eigenlijke zin van het woord. Dat komt omdat zij nog niet voldoende taalonderwijs hebben om competente sprekers van de 'officiële' standaardtaal te zijn. Vanaf de jaren 50 treden er sprekers naar voren die hebben kunnen genieten van de verbeteringen die de maatschappelijke veranderingen van na de Tweede Wereldoorlog met zich meegebracht hebben: zo zijn zij – zowel geografisch als sociaal – mobieler geworden door enerzijds de stijging van de welvaart en daaraan gekoppeld de uitbouw van de materiële infrastructuur, maar anderzijds ook door de opkomst van de massamedia en de democratisering van het onderwijs (zie: Boone, Gaus & Scholliers 1982, Gaus 2001, en Vandenbroeke 1982 voor basisreferenties omtrent de sociale geschiedenis van België en Vlaanderen). Vooral die laatste twee factoren maken dat hun taal een toenemende graad van standaardisering reflecteert – waarbij dat vooralsnog opgevat dient te worden als in de eerste plaats gericht op het Journaalnederlands. In de jaren 70 staat er dan een generatie op voor wie de maatschappelijke verworvenheden een vanzelfsprekendheid geworden zijn: de emancipatie van het Vlaamse openbare leven is inmiddels al zo ver gevorderd dat deze sprekers – quasi-probleemloos –

kunnen opklimmen tot de hoogste sociale echelons en daarmee per definitie in Vlaanderen de maatschappelijke elite uitmaken. Linguïstisch gezien gaan zij daardoor willens nillens als spraakmaker fungeren: hun prestigieuze positie maakt dat hun eigen taalgebruik steeds meer tot leidraad gaat dienen. Voor de ontwikkeling van de standaardtaal in Vlaanderen heeft zulks natuurlijk zo zijn consequenties: daar er namelijk altijd idiolectische verschillen tussen sprekers onderling bestaan, en meer bepaald niet elke spreker het Journaalnederlands tot in de puntjes beheerst, zal er voor het sociale veld als geheel onvermijdelijk een toename in gebruik komen van endogeen Belgisch-Nederlandse elementen. In die zin lijkt het ons dan ook weinig plausibel om één en ander als louter een ‘age grading’-fenomeen te bestempelen: het CGN is namelijk samengesteld rond de millenniumwissel (zie hoofdstuk 2), zodat een gemiddelde spreker uit het decennium 1970-1979 op dat moment 25 jaar oud is; daarmee is hij al lang in de maatschappelijke orde opgenomen.

Op die manier is de verbreiding van het Soapvlaams in de jaren 80 toe te schrijven aan het proces dat oorspronkelijk net het tegenovergestelde ervan beoogde: de ontvoogding van Vlaanderen beduidde zoals bekend namelijk dat Vlamingen zich de Journaalnederlandse standaardtaal aan zouden leren. Hoewel geen van de auteurs uit hoofdstuk 1 het er expliciet over heeft, is een ‘dialectische omwenteling’ zoals deze evenwel niet zo ongewoon voor verschijnselen van maatschappelijke aard. Met name de socioloog Abram de Swaan heeft de informalisering, zoals die in het algemeen vanaf het midden van de jaren 60 op begon te komen, proberen te verhelderen aan de hand van een verschuiving van, wat hij noemt, een ‘bevelshuishouding’ naar een ‘onderhandelingshuishouding’ (zie de Swaan 1983 voor een bundel opstellen hieromtrent): naarmate voorheen bevoogde bevolkingsgroepen krachtens hun emancipatie zowel in economisch als intellectueel opzicht slagvaardiger worden, zal de gevestigde orde ertoe gedwongen worden om steeds meer met deze ‘nieuwelingen’ rekening te gaan houden – de oude garde moet in plaats van bevelen overgaan tot onderhandelen. Ons dunkt de Vlaamse standaardisering hiervan met name een treffende illustratie.

## 5.4 Generatie

Omdat het effect bij de decennia best wel eens op een aan de gang zijnde taalverandering zou kunnen wijzen (en dus net niet op een vertraagde of zelfs stilgevallen standaardisering, zoals soms nogal eens geponeerd wordt – zie bijvoorbeeld recent nog Van Marle 2002), willen we de rol van leeftijd ook op een sociologisch ietwat meer gemotiveerde manier bestuderen. Daartoe baseren we ons op het standaardwerk van socioloog Henk Becker over de verschillende generaties van de twintigste eeuw: Becker, Henk (1992) *Generaties en hun kansen*. Becker onderscheidt vijf generaties:

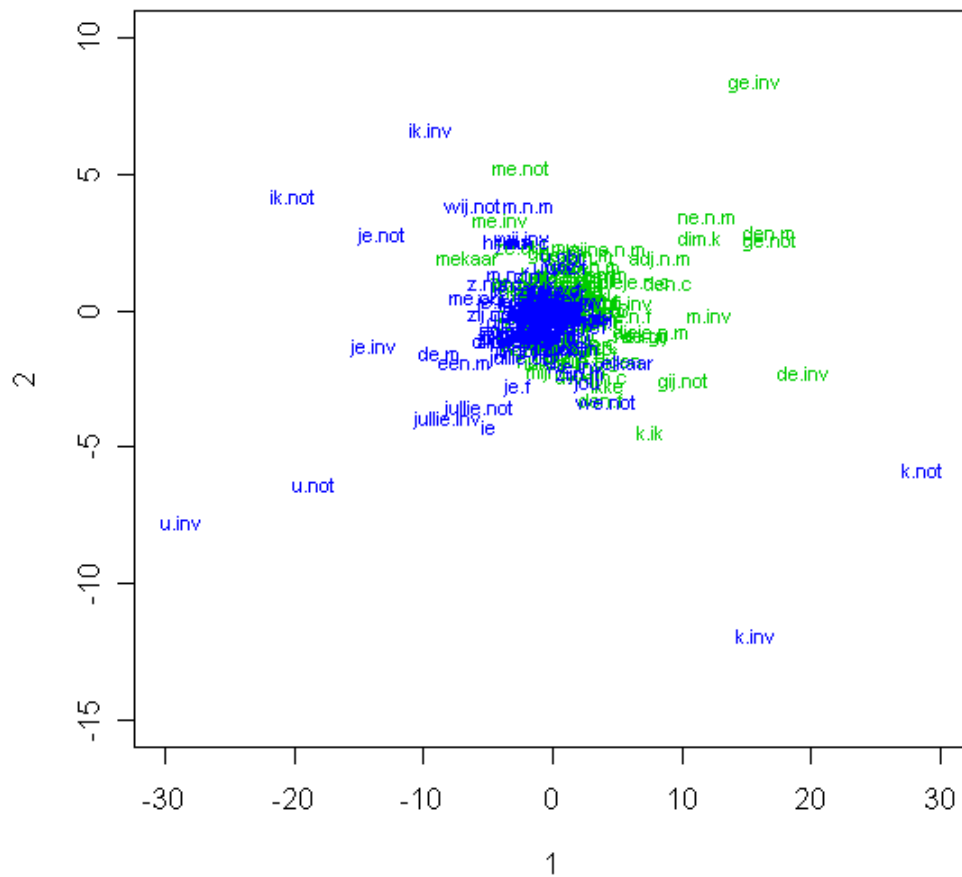
<b>Label</b>	<b>Generatie</b>	<b>Periode</b>
pre	vooorlogse generatie	1910-1929
sil	stille generatie	1930-1939
pro	protestgeneratie	1940-1954
los	verloren generatie	1955-1970
pra	pragmatische generatie	1971-nu

*Tabel 5.4: De vijf generaties uit Becker, Henk (1992)*

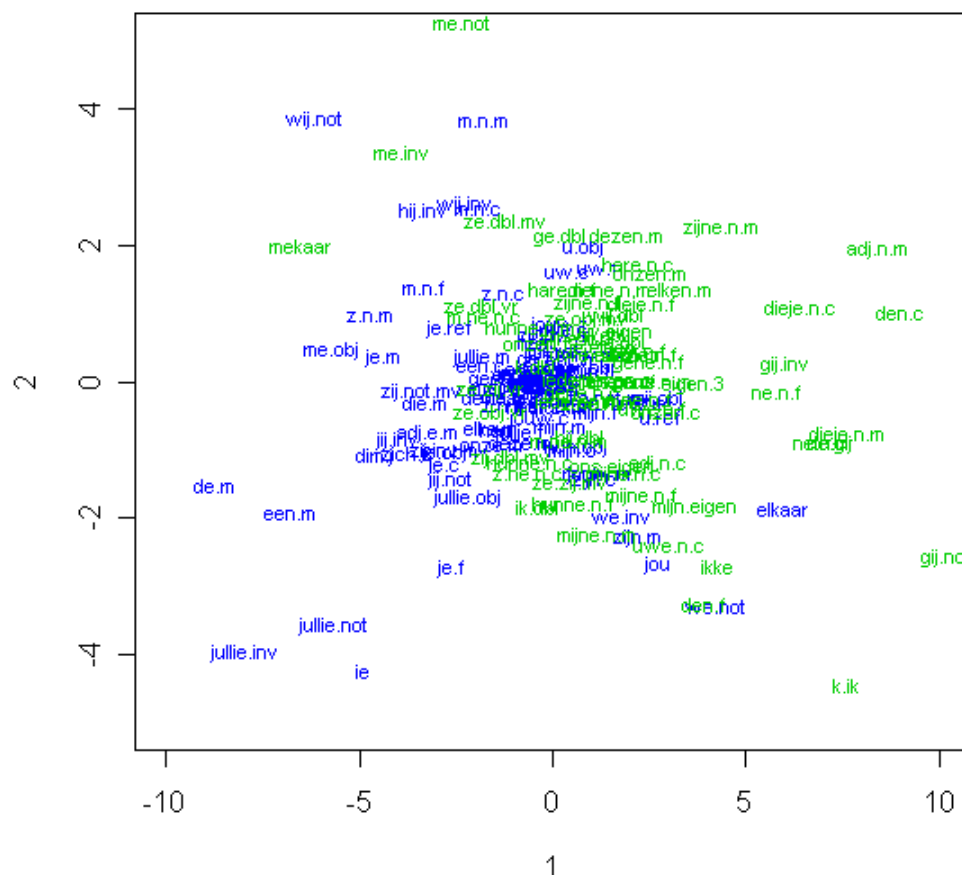
Voor verdere details omtrent deze indeling verwijzen we naar het boek van Becker zelf.

In de biplot (Figuur 5.22) zien we dan het effect van de decennia inderdaad terugkeren: de oudste generaties (de ‘vooorlogse’ en de ‘stille’) bevinden zich in het centrum van de plot, waarna er tot en met de ‘protest-’ en de ‘verloren’ generatie een steeds verdere verschuiving naar rechts is, om ten slotte met de ‘pragmatische’ generatie ten slotte helemaal de andere richting uit te gaan. De trend is ook hier weer lineair (enkel de verloren generatie vertoont, net zoals de zestigers dat deden in de analyse van de decennia, een duik naar onder. Aangezien zich daar de  $u$ -subjecten bevinden, zou dit eventueel een interactie met register kunnen beduiden: misschien is het bijvoorbeeld zo dat de politici in het CGN toevallig allemaal uit de verloren generatie – dan wel de jaren zestig – stammen. Of zulks inderdaad het geval is zal de multivariate analyse in hoofdstuk 6 uitwijzen).



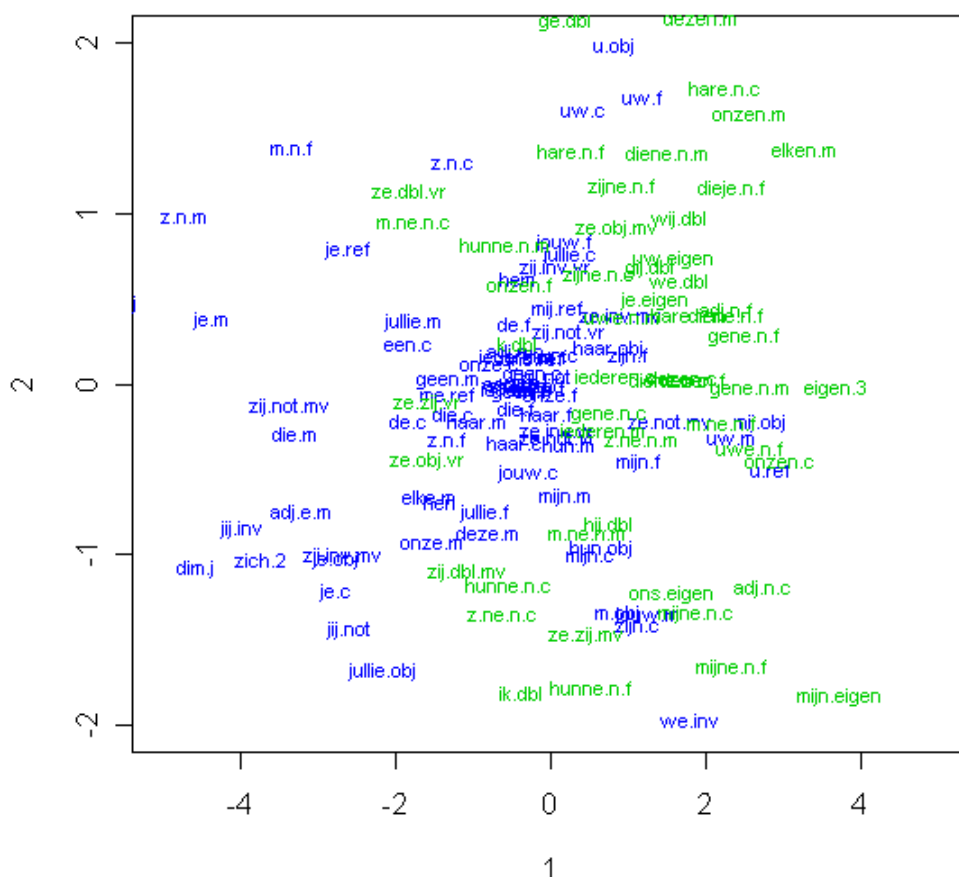


Figuur 5.23: Generatie – varianten: detail [-30;30] horizontaal en [-15;10] verticaal

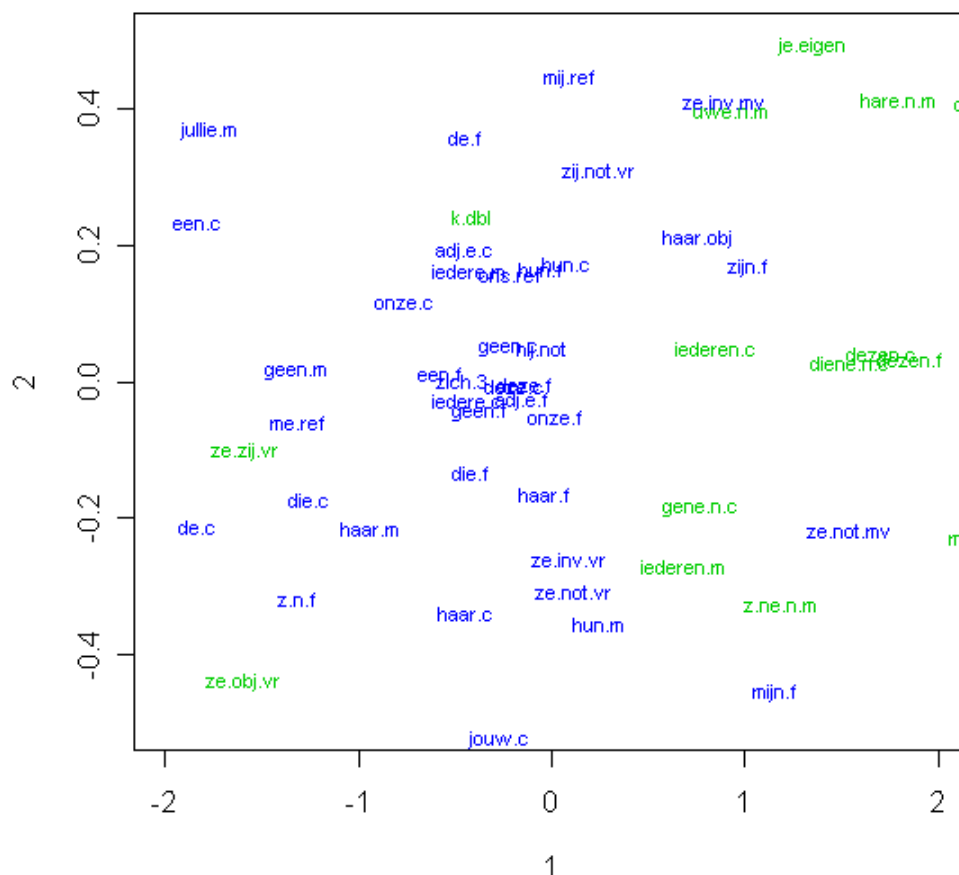


Figuur 5.24: Generatie – varianten: detail [-10;10] horizontaal en [-5;5] verticaal





Figuur 5.25: Generatie – varianten: detail [-5;5] horizontaal en [-2;2] verticaal

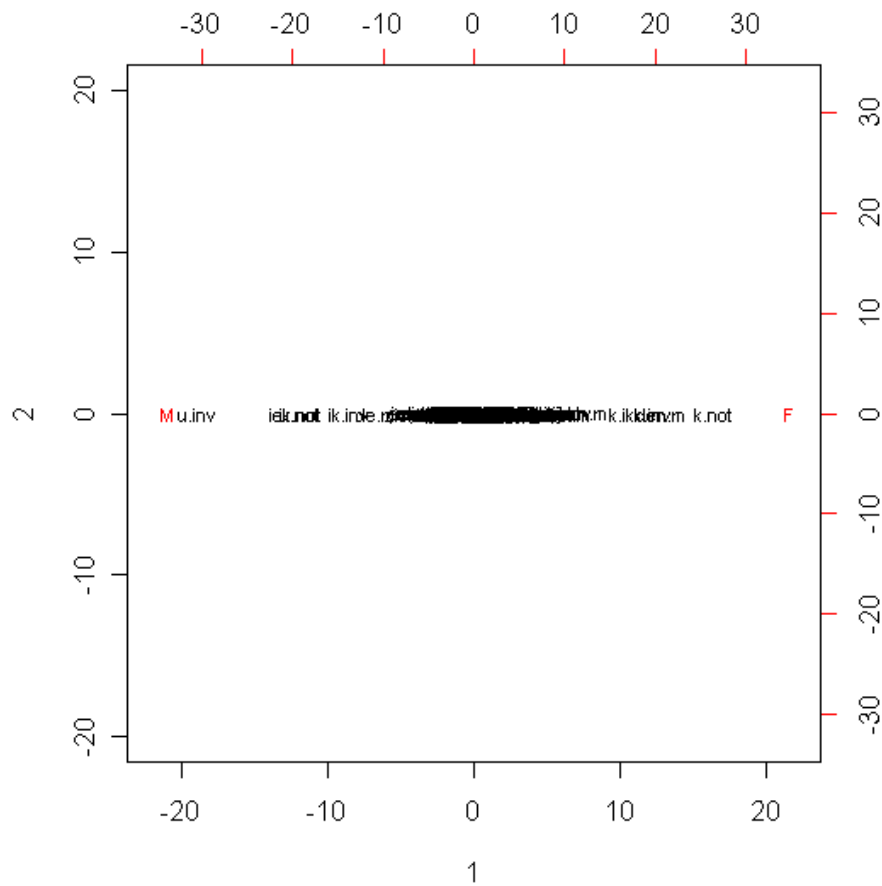


Figuur 5.26: Generatie – varianten: detail [-2;2] horizontaal en [-0,5;0,5] verticaal

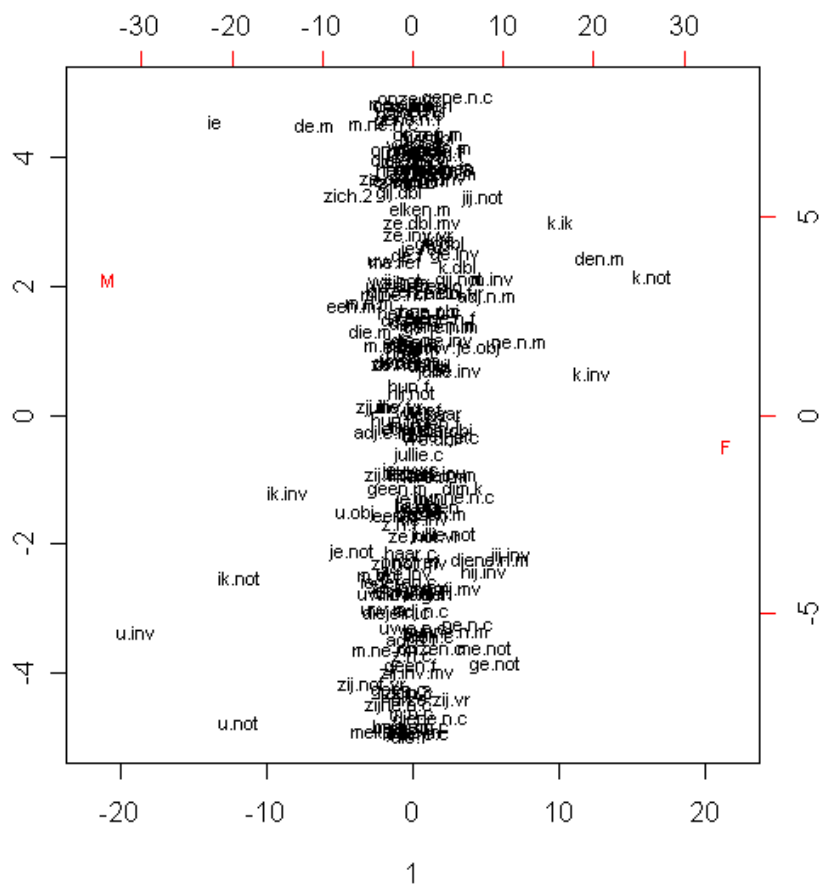
De eindconclusie met betrekking tot de leeftijdsvariatie luidt dan ook dat er zich vanaf ongeveer 1970 een nieuwe episode heeft aangediend in het standaardisatieproces van Vlaanderen: waar standaardisering voorheen nog de vorm aannam van een verbreiding van het Journaalnederlands, geeft de standaardtaalontwikkeling van vandaag een verregaande mate van endogenisering te kennen. Bij de uitbouw van zijn standaardtaal is de Vlaming niet langer bereid om binnen de 'strakke' lijntjes van de geijkte norm te blijven, maar put daarentegen steeds meer uit zijn inheemse taaleigen. De opkomst van het Soapvlaams behelst zo kortom een niet te verwaarlozen uitbreiding van de autonomie van het Belgisch-Nederlands.

## 5.5 Sekse

Voor de factor sekse moeten we op voorhand aanstippen dat de approximatieve ruimte unidimensioneel zal zijn, oftewel een rechte. De reden daarvoor is van zuiver wiskundige aard: een approximatieve ruimte zoals wij die berekenen heeft nu eenmaal altijd hoogstens  $\min(\text{het aantal rijen} - 1, \text{het aantal kolommen} - 1)$  dimensies (vergelijk het aantal vrijheidsgraden in een chikwadrattest:  $(\text{het aantal rijen} - 1) \times (\text{het aantal kolommen} - 1)$ ); we gebruiken bijvoorbeeld ook de chikwadratmetriek om afstanden te meten – zie hoofdstuk 4). De biplot ziet er met andere woorden uit zoals Figuur 5.27. Zulks is natuurlijk nogal onhandig om te interpreteren. Ter vergroting van de leesbaarheid van de plot, voegen we aan de verticale dimensie dan ook een extra portie witte ruis toe. Dat zal de datapunten namelijk uit elkaar trekken, zodat er uit de plot toch nog enige informatie te puren valt. Het zij echter beklemtoond dat we dit enkel uit praktische overwegingen doen: de variatie die we hiermee op de verticale dimensie introduceren is geheel arbitrair – ze betekent niets. Anders uitgedrukt komt het erop neer dat de verticale dimensie 'eigenlijk niet bestaat' (wat we symbolisch dan ook zullen aangeven door het label bij de verticale as weg te laten). De biplot krijgt daarmee een vorm zoals Figuur 5.28.

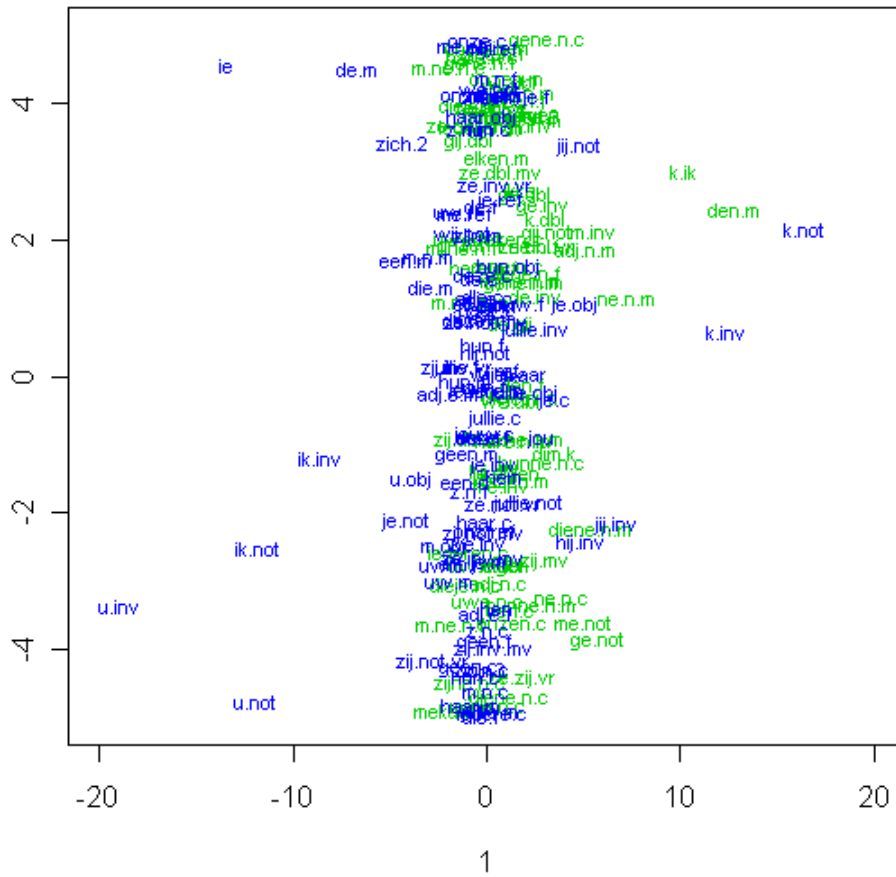


Figuur 5.27: Biplot – sekse

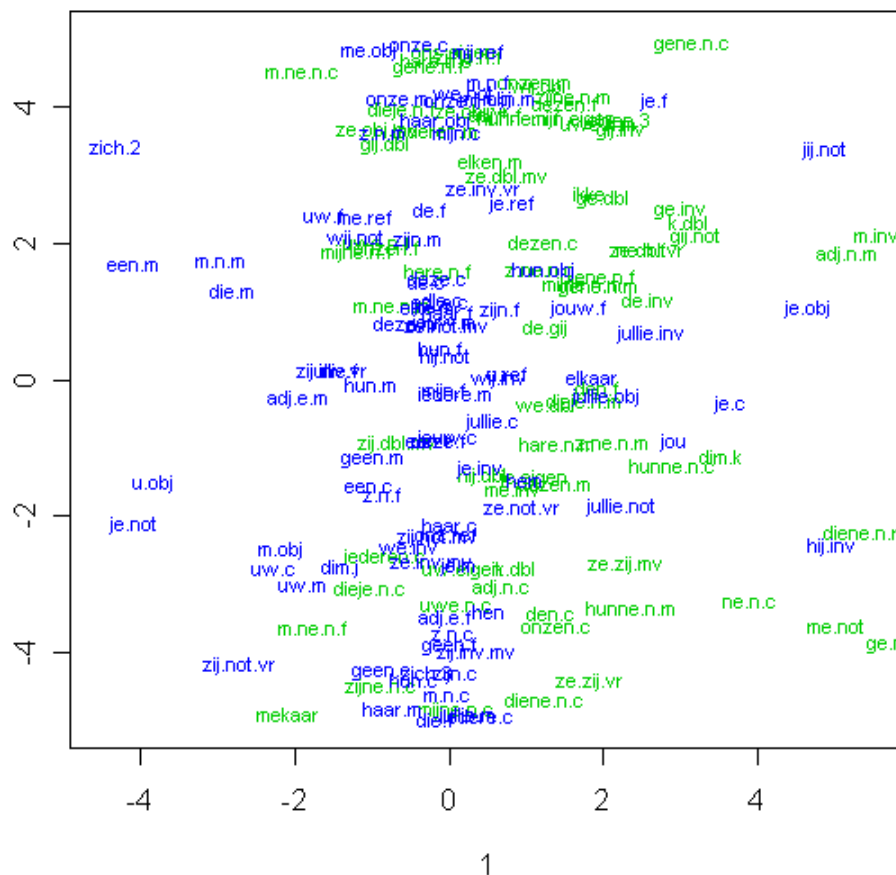


Figuur 5.28: Biplot – sekse: met artificiële verticale dimensie

Figuur 5.29 geeft verder de distributie van de taalvarianten, en Figuur 5.29 zoomt daarbij iets dieper in op het interval tussen -5 en 5. We zien dan geen echt markante verschillen tussen de sekse, wat de in hoofdstuk 1 gehypothetiseerde rol van de vrouwen ontkracht: de Journaalnederlandse varianten en Soapvlaamse varianten lopen gewoon door elkaar. Veel valt er dan verder ook niet op te maken uit het feit dat de mannen in de linkerhelft van de plot staan, terwijl de vrouwen in de rechterhelft. Eventueel kan men opperen dat de mannen vooralsnog meer naar een formele spreekstijl neigen dan vrouwen, gegeven de associatie met de *u*-subjecten aan de linkerkant. De verklaring hiervoor zou echter wel eens veeleer op een sociaal stereotype kunnen berusten dan op een verschil in louter taalgebruik. Uit de analyse van de registers (5.1) weten we namelijk dat het gebruik van de beleefdheidsvorm een kenmerk is van componenten f en g, waar politici met elkaar debatteren. Als dan blijkt dat de mannelijk sprekers meer de beleefdheidsvorm gebruiken, dan zou dat misschien eerder kunnen duiden op het feit dat het – in weerwil van de emancipatie – toch nog altijd in hoofdzaak mannen zijn die de politieke arena bevolken. In dat geval zou dit effect (net zoals we zulks voor leeftijd gehypothetiseerd hebben – zie 5.3 en 5.4) een interactie zijn tussen register en sekse. De multivariate analyse zal daarover uitsluitsel bieden; zie hoofdstuk 6.



Figuur 5.29: Sekse – varianten: detail [-20;20] horizontaal en [-5;5] verticaal



Figuur 5.30: Sekse – varianten: detail [-5;5] horizontaal en [-5;5] verticaal

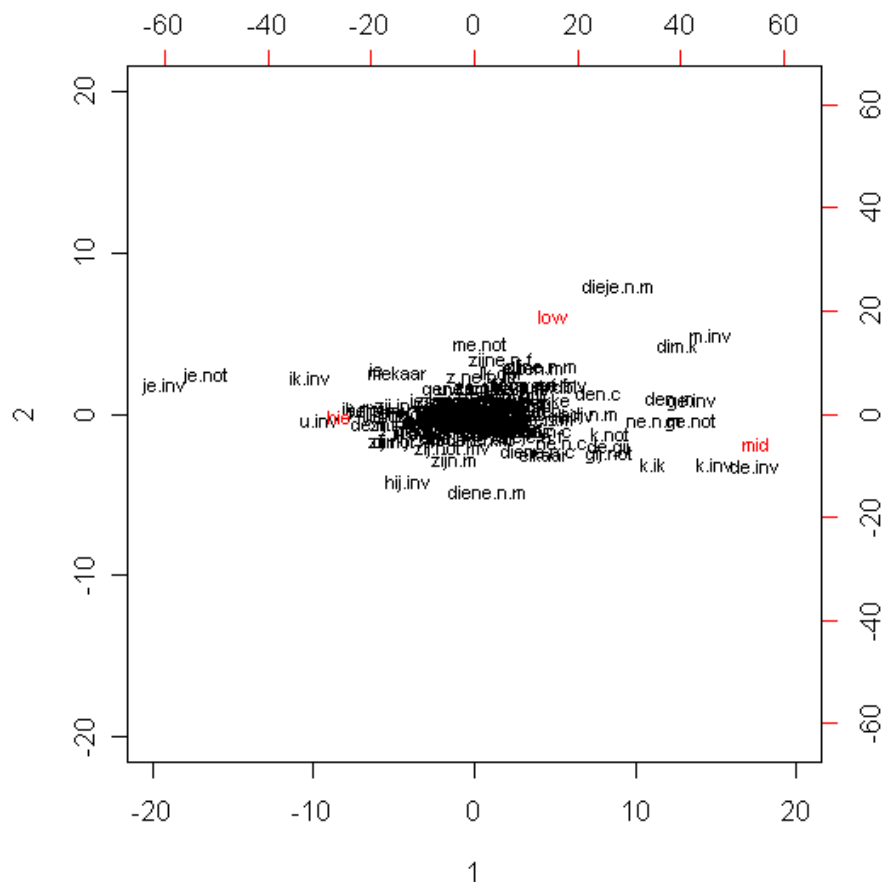
## 5.6 Opleidingsniveau

Voor het opleidingsniveau van de sprekers herhalen we dat er drie categorieën onderscheiden worden (zie 2.2.4):

<b>Label</b>	<b>Educatieniveau</b>
hie	hoog
mid	midden
low	laag

*Tabel 5.5: De drie opleidingsniveaus in het CGN*

De biplot (Figuur 5.31) geeft opnieuw de lineaire trend weer die we eveneens bij de leeftijdsvariabele tegengekomen zijn. De hoog opgeleide sprekers staan in de linkerhelft van de plot, de laag opgeleiden in het centrum, en de middencategorie bevindt zich helemaal rechts. Daarbij zij het evenwel opgemerkt dat de positie van die laatste geprononceerder is dan die van de eerste: respectievelijk (ongeveer) +20 en -10 op de horizontale as.



Figuur 5.31: Biplot – educatieniveau

Ook voor de varianten keert hetzelfde patroon terug (zie Figuren 5.32, 5.33, 5.34, en 5.35): een continuüm – let bovendien op de grote mate van overlapping – tussen de Journaalnederlandse vormen links en de Soapvlaamse vormen rechts, de bekende set van uitschieters (het 'm-subject, de diminutief op *-ke*, *den* en *ne(n)* bij masculina, wel en niet-invert *ge*, het enclitische *-de*, en zelfs het reduplicatieve (*e*)*kik*), en de subject-'*k* nog maar eens als uitzondering.







Eén en ander heeft zo zijn consequenties: sprekers met een hogere opleiding zijn namelijk in staat om gebruik te maken van de Journaalnederlandse variëteit, de laagst opgeleiden alterneren veelal tussen Journaalnederlands en Soapvlaams, terwijl de middencategorie tot slot van het Soapvlaams haar idioom heeft gemaakt. Daarmee is de hypothese (zie hoofdstuk 1) als zou het Soapvlaams uitgedragen worden door de hoger opgeleiden zonder meer gefalsificeerd. Belangrijk in dat verband is evenwel het feit dat het linguïstische continuüm Journaalnederlands-Soapvlaams *niet* parallel loopt aan de sociale ladder hoog-midden-laag: midden en laag hebben daarvoor van plaats verwisseld. Hoe moeten we dat nu interpreteren? Ons lijkt het vooral te maken te hebben met de successieve episodes waarin Vlaamse taalgebruikers zich de verschillende spreekstijlen van het Belgisch-Nederlands eigen maken: in eerste instantie verwerft men bijvoorbeeld de basisgrammatica van de taal. Voor Vlaanderen is dat typisch een mengeling van het Journaalnederlands en het dialect (nogmaals: ‘tussentaal’ in de eigenlijke zin van het woord – zie 5.3), en bijgevolg staan de laagst opgeleide sprekers in het centrum van de plot. In de daarop volgende fase (die we dus correleren aan de educatieve middencategorie) ontwikkelen sprekers hoofdzakelijk de omgangstalige standaardvariëteit van hun taal, en in het Belgisch-Nederlands is dat natuurlijk het Soapvlaams. Het is pas vanaf het hoogste maatschappelijke niveau dat sprekers in situaties verzeild geraken waar de officiële standaardtaal vereist is, vandaar dat de hoogst opgeleiden in de plot meer naar het Journaalnederlands neigen. Toch mogen we dat nu ook weer niet overdrijven. Het boven gesignaleerde verschil in prominentie (op de horizontale as) met de educatieve middencategorie geeft namelijk aan dat zij één en ander niet met volle overgave doen. Voor hen geldt dan ook alsnog hetzelfde als wat we voor de overige twee opleidingsgraden geponeerd hebben: de standaard waar de modale Vlaming zijn taalgedrag naar richt ligt in de eerste plaats aan de Soapvlaamse kant van het spectrum, en dan pas aan de Journaalnederlandse.

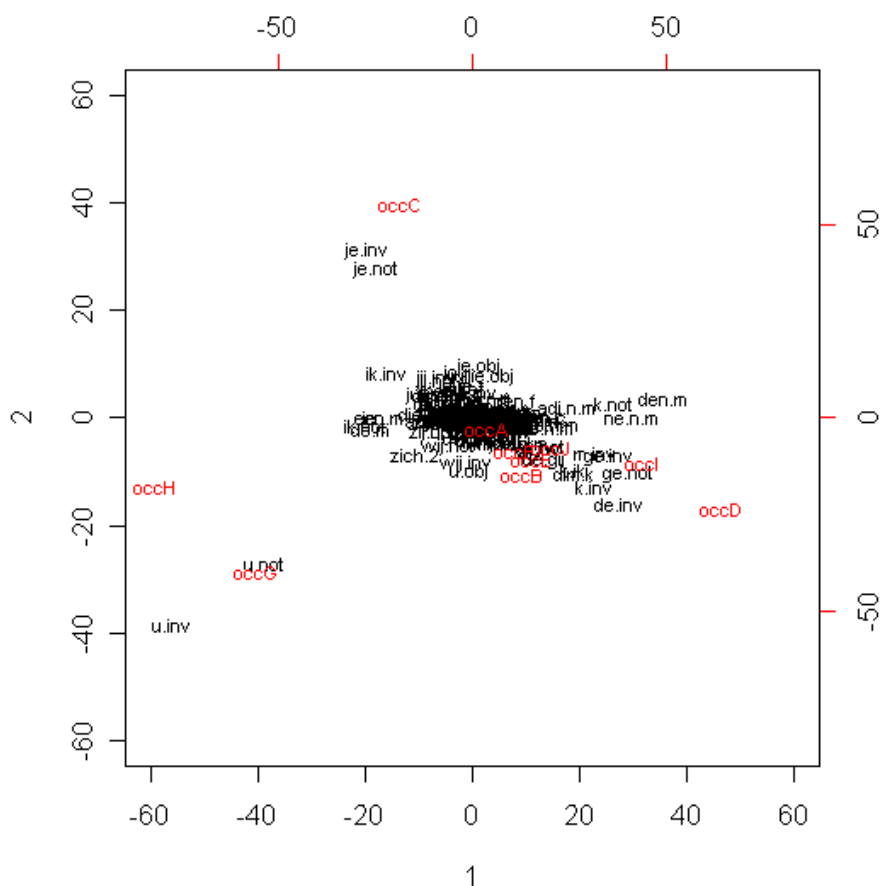
## 5.7 Beroepsniveau

Tot slot hanteert de factor beroepsniveau zoals gezegd een tienvoudige indeling (zie 2.2.5):

<b>Label</b>	<b>Beroepsniveau</b>
occA	werkzaam in het hoger management of bestuur
occB	beroep dat een hogere opleiding veronderstelt
occC	werkzaam als onderzoeker of docent aan universiteit of instelling voor hoger onderwijs
occD	werkzaam in administratieve functie of een functie in de zorgsector
occE	ongeschoold beroep
occF	werkzaam als zelfstandige
occG	politicus
occH	werkzaam in de media (journalist, reporter) of als artiest
occl	student, stagiair
occJ	niet beroepsmatig actief

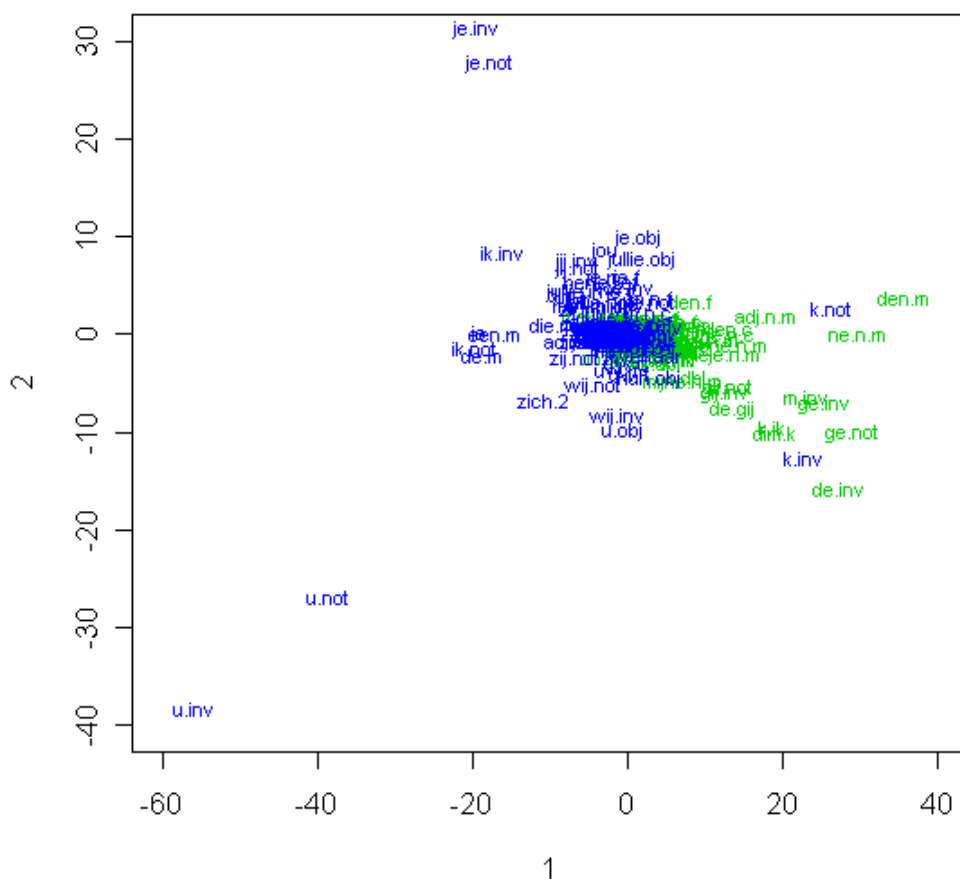
*Tabel 5.6: De tien beroepsniveaus in het CGN*

Enigszins simplificerend kunnen we stellen dat categorie occA de managers representeert, occB de hoger opgeleiden, occC de academici, occD de ambtenaren (en beambten), occE de ongeschoolden, occF de zelfstandigen, occG de politici, occH de mediafiguren allerhande, occl de studenten (en stageairs), en occJ de werklozen.

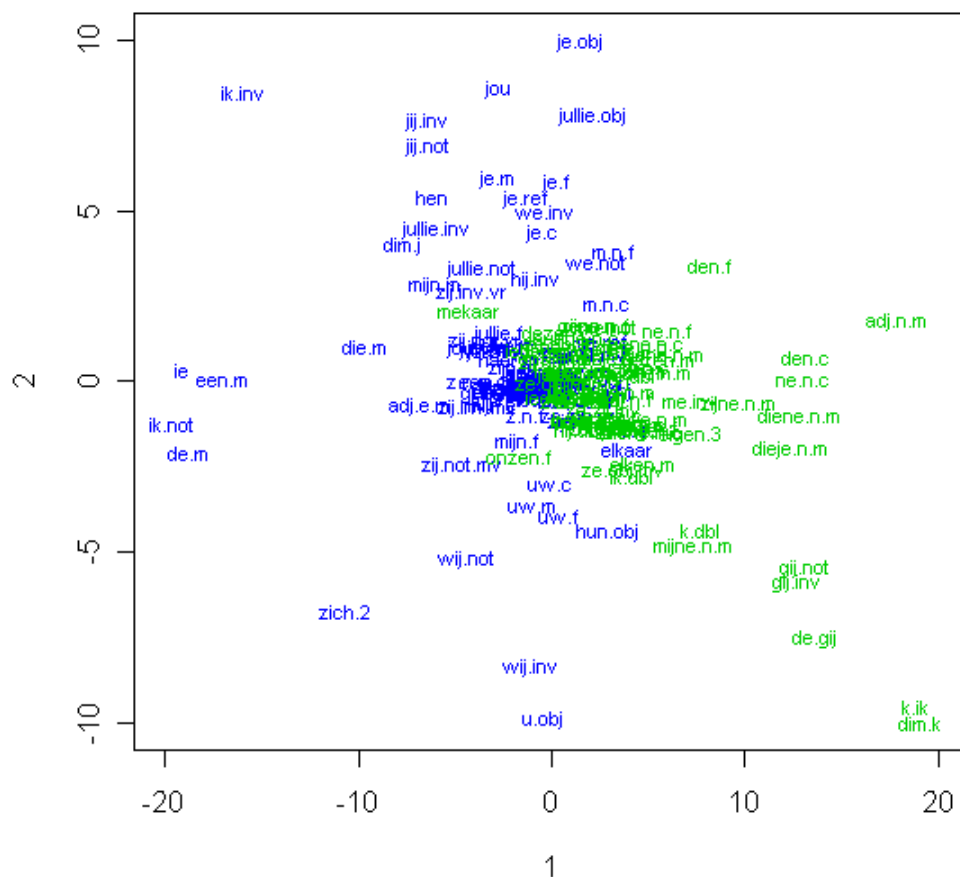


Figuur 5.36: Biplot – beroepsniveau

De biplot voor de tien beroepscategorieën is dan Figuur 5.36. We treffen de politici (occG) en de mediafiguren (occH) linksonder aan, de academici (occC) linksboven, de studenten (occI) en ambtenaren (occD) helemaal rechts van de plot, en verder clusteren alle overige categorieën (occA, occB, occE, occF, en occJ) samen in het centrum.



Figuur 5.37: Beroepsniveau – varianten: detail [-60;40] horizontaal en [-40;30] verticaal



Figuur 5.38: Beroepsniveau – varianten: detail [-20;20] horizontaal en [-10;10] verticaal



Nadere details omtrent de precieze verhouding van de taalvarianten geven Figuren 5.37, 5.38, 5.39 en 5.40. We bemerken weer een geleidelijke overgang van de Journaalnederlandse vormen links naar de Soapvlaamse rechts (nogmaals met uitzondering van de gereduceerde 'k-subjecten), maar deze keer is de plot terug tweedimensionaal: de *u*-subjecten links onderaan, de J-vormen bovenaan. De gereduceerde *je*-subjecten zijn daarbij duidelijk opnieuw de uitschieters; alle andere J-vormen bevinden zich gewoon aan de bovenkant van de centrumcluster met de possessiva zelfs in het midden daarvan – zie Figuur 5.38 en 5.39, respectievelijk. Naast de aanspreekvormen (zoals verwacht staan de G-vormen rechts) komen we verder de bekende reeks van sjibbolets tegen (*den* en *ne(n)* bij masculina, de subject-'*m*, de diminutief op *-ke*, en (*e*)*kik*). Tot slot valt nog op dat *hen* zich ter hoogte van de J-vormen ophoudt, terwijl *hun.obj* en *ze.obj.mv* meer aan de rechterbenedenkant te vinden zijn (onder *elkaar*, meer bepaald).

Het meest opmerkelijke aan deze resultaten is evenwel de treffende gelijkheid met die van de registers (zie 5.1). Dat is niet toevallig: heel wat componenten zijn namelijk net gedefinieerd in termen van een beroepsklasse. Zo bevatten componenten b en h bijvoorbeeld het taalgebruik van leraren (in een privé-interview dan wel voor de klas), terwijl de debattaire componenten f en g dat van politici. Met betrekking tot de beroepsgroepen zien we de J-vormen dan ook opnieuw geassocieerd met een onderwijsgevende klasse zoals de academici (de J-vormen treden kortom nogmaals tevoorschijn als 'schoolse' varianten), en de beleefde U-vormen met de politici – dat de mediafiguren soortgelijk taalgedrag vertonen heeft dan weer te maken met het feit dat het hier veelal om interviewende of modererende reporters en journalisten gaat, waarmee hun taalgebruik 'als vanzelf' formeler wordt (vergelijk in dat verband bijvoorbeeld de interviewster uit component b – zie de bespreking in 5.1).

Eén en ander toont echter aan hoe problematisch een begrip zoals 'homogeniteit' of 'uniformiteit' wel is. Van een standaardtaal wordt in de regel namelijk verondersteld dat ze 'homogeen'/'uniform' is ten opzichte van sociale klassen (zie bijvoorbeeld Geeraerts 2003: 28, maar ook de ANS: 16). We stellen hier evenwel vast dat één van de sterkste contrasten uit het hele Belgisch-Nederlandse taalspectrum net een verschil binnen het officieel standaardtalige Journaalnederlands behelst tussen twee beroepsgroepen die toch wel prototypische exponenten ervan mogen heten. Natuurlijk gaat het daarbij om een grammaticale distinctie (met name: vertrouwelijk-beleefd, inderdaad gerepresenteerd in de verticale dimensie van de plot – zie 5.1) die *als zodanig* homogeen/uniform over de verschillende bevolkingsgroepen verspreid is; toch kan men er niet blind voor blijven dat zulks in de feitelijke taalpraktijk tot uiting komt als heterogeniteit, wat de notie van uniformiteit – ook hier weer – enigszins uitholt.

Aan de Soapvlaamse (rechter-)kant van de plot kunnen we voorts in de positie van studenten en ambtenaren een bevestiging zien van al eerder geconstateerde effecten. De studenten corroboreren namelijk wat we bij leeftijd opgemerkt hebben: met name, dat de verbreiding van het Soapvlaams gedragen wordt door de jongere sprekers. De ambtenaren kunnen we dan weer in verband brengen met de middencategorie van het educatieniveau: sprekers uit de middenklasse van de Vlaamse maatschappij schromen niet om hun toevlucht te nemen tot het Soapvlaams.

Alle overige beroepsniveaus clusteren samen in het centrum van de plot, waarbij de lichte afbuiging naar de Soapvlaamse rechterhelft er geen twijfel over laat bestaan welke stijl in Vlaanderen uiteindelijk de default vormt. Verrassend is evenwel de sociale heterogeniteit van deze cluster: het gaat hier namelijk enerzijds om managers en hoger opgeleiden, en anderzijds om ongeschoolden, zelfstandigen en werklozen. Op grond van eerdere bevindingen is de positie van de laatste drie daarbij voorspelbaar: het betreft hier klassen uit de lagere regionen van de maatschappij, waarvan we al weten zij – al dan niet uit een gebrekkige beheersing – een dialectisch gekleurd Nederlands hanteren. Voor de eerste twee geldt dat daarentegen niet: deze bevolkingsklassen behoren juist tot de maatschappelijke toplaag en zouden we daarom ook meer naar de Journaalnederlandse linkerkant verwachten. Een tabel van het aantal sprekers per educatieniveau (Tabel 5.7) werpt enig licht op de zaak:

	<b>occA</b>	<b>occB</b>	<b>occC</b>	<b>occD</b>	<b>occE</b>	<b>occF</b>	<b>occG</b>	<b>occH</b>	<b>occl</b>	<b>occJ</b>
<b>hie</b>	62	90	331	186	0	15	98	241	34	6
<b>mid</b>	6	0	1	66	5	19	4	20	85	6
<b>low</b>	0	0	1	3	4	3	1	3	3	4

*Tabel 5.7: Sprekersaantallen – Educatieniveau X Beroepsniveau*

We zien dan dat de groep van de managers (occA) uit die twee educatieniveaus bestaat die in hun taalgebruik net elkaars uitersten zijn: de Journaalnederlands sprekende hoog opgeleiden en de Soapvlaams pratende middencategorie (vergelijk 5.6). In het licht hiervan kunnen we de positie van de managers dan ook als een ‘compromis’ van beide opleidingsniveaus beschouwen: ‘occA’ weerspiegelt zowel het taalgedrag van ‘hie’ als dat van ‘mid’. Ronduit paradoxaal zijn echter de occupationeel hoger opgeleiden (occB), die qua educatiegraad inderdaad alleen hoger opgeleiden tellen en die we dan ook met name in het linkerdeel verwachten aan te treffen. De sprekersverdeling van de hoogste opleidingsgraad in Tabel 5.7 leert evenwel dat de meest prominente vertegenwoordigers ervan enerzijds de academici (occC) en anderzijds de mediafiguren (occH) zijn, oftewel de twee



exponenten bij uitstek van het Journaalnederlands. Daarmee kan deze educatiegraad eveneens opgevat worden als een 'compromis': ze staat namelijk in tussen aan de ene kant de bevolkingsgroepen die in Vlaanderen *intellectuele* elite uitmaken (academici en mediafiguren) en aan de andere kant de beroepsklasse die dat met betrekking tot de *economische* elite (occB) doet. Met die opdeling kunnen we ten slotte de (in hoofdstuk 1 gehypothetiseerde) vraag beantwoorden in hoeverre het Soapvlaams geschraagd wordt door een maatschappelijke elite. Het antwoord daarop kan niet anders dan genuanceerd zijn: het Journaalnederlands representeert nog altijd de standaard voor wie op culturele gronden prestige geniet; degenen die hun status daarentegen aan meer materiële factoren danken neigen veel meer naar het Soapvlaams.

Met bespiegelingen van deze aard hebben we evenwel al een begin gemaakt met het vergelijken van effecten van verschillende factoren met elkaar. Zulks kan evenwel gemakkelijk tot overhaaste conclusies leiden als we niet over een multivariate techniek beschikken waarmee we één en ander op een meer onderbouwde manier kunnen valideren. Die multivariate analyse komt nu in hoofdstuk 6 aan bod.



## Hoofdstuk 6 | Multivariate analyse

Nu de externe factoren afzonderlijk bestudeerd zijn, willen we ze allemaal samen opnemen in één globale multivariate analyse. Dat zal ons enerzijds namelijk in staat stellen om effecten van verschillende factoren met elkaar te vergelijken. Zo weten we dat het gebruik van het Soapvlaams bijvoorbeeld aan de ene kant te maken heeft met informele spreesituaties (registereffect – zie 5.1), en aan de andere kant met de pragmatische generatie (leeftijdseffect – zie 5.4); het is evenwel niet duidelijk welke van deze effecten nu het sterkst doorweegt: als Vlamingen het Soapvlaams hanteren, komt dat dan vooral omdat ze zich in een informele spreesituatie bevinden, of eerder doordat het gaat om sprekers die tot de pragmatische generatie behoren? Anderzijds kan een multivariate analyse ons eveneens wijzen op eventuele interacties tussen de factoren. Dat de mogelijkheid daartoe allesbehalve louter hypothetisch blijft, is bij de bivariate analyses al op enige plaatsen gebleken – zie 5.5 en 5.6. In beide gevallen vormen de hoofdeffecten van de factoren het uitgangspunt voor de hele analyse; dit hoofdstuk vertrekt dan ook van Figuur 6.1.







De plots tonen weer de naadloze overgang van het Journaalnederlands (links) naar het Soapvlaams (rechts), opnieuw met de bekende uitzonderingen evenals uitschieters – waarover zo meteen meer. We stellen met andere woorden vast dat de continuüm die we voor elk van de factoren apart vastgesteld hebben volkomen bewaard blijft in een analyse die alle factoren tegelijkertijd bestudeert. In het licht van wat in 5.1 hierover gezegd hebben, wijst het er alleen maar op hoe monoglossisch het taalgebruik in Vlaanderen vandaag de dag wel is.

Ook de reeks van uitschieters bedraagt hetzelfde ‘selecte clubje’: *den* en *ne(n)* bij masculiene nomina, de aanspreekvormen *–de* en *ge* (die laatste wel en niet-geïnverteerd), de diminutief op *–ke*, en de subject-‘*m*’ (en in mindere mate het reduplicatieve *(e)kik*); de Journaalnederlandse sjibboletvormen zijn dan weer de corresponderende alternatieven: *een* en *de* bij masculina, de aanspreekvormen *je* en *u* (beide zowel wel als niet-geïnverteerd), en tot slot het volle *ik* (ook hier wel en niet-invert) – daarentegen gedragen de diminutief op *–je*, evenals het invert *hij* en *ie* zich voor het Journaalnederlands blijkbaar minder prominent.

De posities van de lidwoorden bevestigen bovendien het spreidingseffect onder de genera: bij een masculien nomen vertonen de determinatoren meer variatie dan bij commune of feminiene nomina. Naast de lidwoorden komt dat vooral goed tot uiting bij de *die*-vormen – waar verder de Brabantse varianten op *–je(n)* trouwens allemaal onderaan de centrumcluster te vinden zijn, terwijl de Vlaamse *–ne(n)*-varianten bovenaan; zie Figuur 6.3 en 6.4.

Het sjibbolet van het Journaalnederlands blijft echter toch het onderscheid tussen de vertrouwelijke J-vormen en de beleefde U-vormen langs de verticale as – die daarmee gelijk weer de dimensie vertrouwelijk-beleefd representeert (in tegenstelling tot de horizontale as, die dus differentieert tussen Journaalnederlands en Soapvlaams). Dit lijkt er op het eerste gezicht ook een verklaring voor te bieden waarom de gereduceerde *je*-varianten hoger in de plot staan dan de volle *jij*, *jullie*, en *jou* die zich links bovenaan de centrumcluster ophouden – waarbij we overigens opnieuw moeten constateren dat de pronomina van de tweede persoon in hun geheel hoger op de verticale as laden dan de determinatorische vormen: de *je*-possessiva bevinden zich dan wel op gelijke hoogte als de vernoemde volle pronomina, de volle possessiva *jouw* en *jullie* zitten echter in het midden van de centrumcluster. Bij de analyse van de registers (5.1) hebben we namelijk opgemerkt dat de verticale as eveneens het onderscheid tussen fonetisch gereduceerd en vol constitueert. Bij nader inzien is zulks hier evenwel niet het geval. De enige variabelen waarvan de gereduceerde varianten eveneens hoger staan dan de volle vormen zijn die van de pronomina van de eerste persoon meervoud: *me.not* en *me.inv* net boven de centrumcluster (Figuur 6.3), daaronder *we.not*, *we.dbl*, en *we.inv*, *wij.dbl* helemaal in het midden van de centrumcluster (net onder *mekaar* – zie Figuur 6.5), en *wij.not*

en *wij.inv* ten slotte onder de centrumcluster (Figuur 6.3). Bij de possessiva van de eerste en derde persoon enkelvoud daarentegen zitten de gereduceerde varianten *m'n-e(n)* en *z'n-e(n)* – dus zowel de ongeflecteerd Journaalnederlandse als de geflecteerd Soapvlaamse – lager dan de volle varianten *mijn-e(n)* en *zijn-e(n)*, en dat voor de drie genera; zie Figuur 6.4 en 6.5. Bovendien lijkt er voor *ze* (zowel vrouwelijk enkelvoud als meervoud) eerder een tegenstelling tussen links en rechts te bestaan: het gereduceerde *ze* is zelfs één van de weinige Journaalnederlandse varianten die het nulpunt van de horizontale as overschrijdt en zich in de Soapvlaamse rechterhelft bevindt (Figuur 6.3 en 6.4), net zoals de *'k* (wel en niet-invert).

Dat leidt ons tot de zogenaamde ‘uitzonderingen’ (die voor de overlapping tussen Journaalnederlands en Soapvlaams zorgen). Naast *'k* en *ze* zien we ook nog *mij* (zowel objectvorm als reflexivum), de object-*hun* evenals *-'m*, en ten slotte *u* (zowel objectvorm als reflexivum). Die laatste heeft/hebben natuurlijk te maken met polysemie: enerzijds behorend tot het vertrouwelijke G-systeem, en anderzijds tot het beleefde U-systeem. De positie van de object-*'m* corroboreert dan weer het vastgestelde principe dat ook de horizontale as distingeert tussen fonetisch gereduceerde en volle vormen: *hem* staat namelijk meer naar links (ter hoogte van *jouw.m* – zie Figuur 6.3 – dus eveneens meer naar boven). Hetzelfde geldt verder voor *haar*: in Figuur 6.4 is *haar.obj* linksonder te vinden, terwijl *ze.obj.vr* rechtsboven.

In het licht hiervan lijkt het toch enigszins paradoxaal dat het volle *mij* in de Soapvlaamse rechterhelft van de plot voorkomt: volgens het fonetische principe zou het gereduceerde *me* hier beter passen. Beide reductievormen staan echter in de Journaalnederlandse linkerkant: *me.obj* helemaal links van Figuur 6.3 (net boven *die.m*); *me.ref* helemaal links van Figuur 6.4 (schuin boven *hen*). De positie van *mij* lijkt ons evenwel geduid te kunnen worden in associatie met het reflexivum *mijn eigen*: doordat die vorm als constructie van possessivum + *eigen* intrinsiek een fonetische ‘zware’ vorm is, zal het de volle vorm *mij* zijn in plaats van het gereduceerde *me*, die als informeel opgevat wordt. Het object-*mij* wordt dan door analogie eveneens informeel beschouwd. Dat deze hypothese niet volslagen uit de lucht gegrepen is, moge blijken uit het samen clusteren van *mij.ref* en *mijn.eigen*: beide varianten hebben met andere woorden nagenoeg dezelfde distributie. Verder kan hiermee eveneens het gedrag van de object-*hun* verklaard worden: *hun.obj* bevindt zich ten opzichte van varianten *hen* (helemaal links van Figuur 6.4) en *ze.obj.mv* (schuin onder *mij.obj*; Figuur 6.4) in de rechterkant van de plot omdat het geassocieerd is met het Soapvlaamse reflexivum *hun eigen*.

Daarmee zijn de uitzonderingen aan de Soapvlaamse rechterkant uitgeput. De enige uitzondering aan de linkerkant wordt gevormd door het reflexivum *je eigen*, dat als Soapvlaamse constructie met een Journaalnederlandse J-vorm haar ambigue status



in het Belgisch-Nederlands weerspiegeld ziet (de vormen *onzen.f* en *iedereen.c* laten we verder buiten beschouwing, omdat het hier telkens varianten betreft met maar één luttel treffer in het hele corpus).

## 6.2 Variëteiten – hoofdeffecten

Met de verhouding tussen de varianten afgehandeld kunnen we overgaan tot een bespreking van de externe factoren (zie Figuur 6.1). We zien daarbij dat vrijwel alle hoofdeffecten uit de bivariate analyses terugkeren. Voor register bemerken we de componenten b, en h weer linksboven sterk geassocieerd met de J-vormen (maar niet component o, die namelijk meer naar het midden verschoven is – zie verderop), f en g linksonder bij de U-vormen, in het centrum de cluster van ‘Journaalnederlandse’ componenten i, j, k, en l, met daarnaast nog m en n (hoewel er tussen de twee laatste nu blijkbaar wat meer afstand is), en ten slotte a, c, d helemaal aan de rechterkant geassocieerd met de Soapvlaamse vormen. Bij leeftijd – waar we ons voor de multivariate analyse enkel toeleggen op de generaties – constateren we weer een geleidelijke verschuiving naar de linkerkant van de plot van de vooroorlogse generatie tot en met de verloren generatie (waarbij er hier duidelijk geen verschil is met de protestgeneratie; de stille generatie houdt zich overigens bij het punt van component m op), om met de pragmatische generatie naar de rechterkant over te slaan. Bij educatieniveau is de verdeling opnieuw: hoog opgeleid links, middencategorie rechts en laag opgeleid in het midden van de plot (‘low’ staat daarbij precies tussen occE en occF), waarbij ook hier de positie van de tweede prominenter is dan die van de eerste. Tot slot keert ook voor beroepsniveau dezelfde verhouding terug: occC (academici) linksboven bij de J-vormen, occG (politici) en occH (mediafiguren) linksonder bij de U-vormen, occD (ambtenaren) en occI (studenten) rechts bij de Soapvlaamse varianten, en verder scholen occA (managers), occB (hoger opgeleiden), occE (ongeschoolden), occF (zelfstandigen), en occJ (werklozen) weer allemaal samen in het neutraal Belgisch-Nederlandse centrum van de plot, nogmaals lichtelijk neigend naar de Soapvlaamse kant.

Toch moeten we nog terugkomen op de verschuiving van component o, waarmee deze component op één lijn komt te staan met de clusters i-j-k-l en a-c-d langs de horizontale as van de plot, die zoals het gezegd het Soapvlaams van het Journaalnederlands differentieert. Zulks bevestigt namelijk onze interpretatie van deze as als een *spontaniteitsdimensie*: het Soapvlaams is een spontane spreekstijl; het Journaalnederlands is daarentegen meer gemonitord – het is met name een mediataal, en eigenlijk eerder nog een schrijftaal. Op de verticale as valt het linguïstische onderscheid tussen de vertrouwelijke J-vormen en de beleefde U-vormen enerzijds dan weer samen met het extralinguïstische onderscheid tussen de schoolse variëteiten b, en h (en occC) en de debataire variëteiten f, en g (en occG en occH) anderzijds, waarmee deze as in alle opzichten een (*in-*)*formaliteitsdimensie*

genoemd kan worden – als naam klinkt ‘formaliteitsdimensie’ misschien beter, ware het niet dat de formele categorieën in de negatieve helft staan en de informele in de positieve, zodat de verticale dimensie van formeel naar informeel loopt, wat we dan ook aangeven door het voorvoegsel *in-* tussen haken te plaatsen. Eén en ander betekent evenwel dat de grammaticale distinctie vertrouwelijk-beleefd niet de parameter is die het al dan niet optreden van het Soapvlaams beregelt: de dimensie die dat wel doet staat, zowel letterlijk als figuurlijk, ‘loodrecht’ daarop.

Voor regio is het beeld vervolgens bepaald anders dan we bij de bivariate analyse vastgesteld hebben, hoewel nu ook weer niet geheel oninterpreteerbaar. Oost- en West-Vlaanderen zitten boven de centrumcluster – met West-Vlaanderen zelfs uiterst linksboven – geassocieerd met de *'k*-subjecten, de *me*-subjecten (eerste persoon meervoud), en de *diene(n)*-varianten; Brabant en Limburg zitten net onder de centrumcluster en krijgen dus de *dieje(n)*-varianten bij zich. De plot is met andere woorden verdeeld in een Vlaamse bovenkant en een ‘niet-Vlaamse’ onderkant; daarbij komt de dialectische achtergrond van West-Vlaanderen in de afwijkende positie van die provincie sterk tot uiting (vergelijk de stelling van Vandekerckhove – 5.2).

Het enige echte verschil met de bivariate analyses treffen we aan bij sekse. Waar we in 5.5 nog geen uitgesproken contrast tussen mannelijke en vrouwelijke sprekers konden vaststellen, zien we hier de mannen linksonder de centrumcluster en de vrouwen rechtsboven. De conclusie moet dan ook luiden dat er in Vlaanderen wel degelijk een sekseverschil is qua taalgebruik: mannen spreken door de band genomen meer Journaalnederlands dan vrouwen, die veeleer binnen de vertrouwelijke spreekstijl blijven en daarbij vooral alterneren tussen de J- en de G-vormen.

Daarnaast is de multivariate analyse zoals gezegd eveneens uiterst geschikt om effecten van verschillende factoren met elkaar te vergelijken. Daarbij stippen we enkel de belangrijkste tendenzen aan; de rest spreekt namelijk voor zich. We hebben bijvoorbeeld gezien dat het taalgebruik van component o (voorgelezen tekst), het hoog opgeleide educatieniveau en de protest- en verloren generatie alle vier variëteiten van het Journaalnederlands zijn; Figuur 6.1 toont evenwel duidelijk aan dat het effect sterker is voor component o dan voor de andere drie variëteiten: als er in Vlaanderen Journaalnederlands gesproken wordt, dan heeft dat zoals aangegeven eerder te maken met het feit dat het taalgebruik een voorgelezen tekst (en dus schrijftaal) is, dan met het feit dat de spreker hoog opgeleid is dan wel uit de protest- of de verloren generatie stamt. Omgekeerd manifesteert de schoolse status van de J-vormen zich treffend in de associatie van register h met beroepsgroep occC (academici): de lesgevers spreken dus doorgaans zoals in de lessen. Bij de U-vormen geldt hetzelfde voor component f en de mediafiguren (occh) enerzijds, en

component g en de politici (occG) anderzijds – waarbij het verschil tussen deze twee clusters onderling dan weer steunt op het feit dat de debatten in f uitgezonden zijn terwijl in g niet; daarmee zijn er in g, in tegenstelling tot f, namelijk geen mediafiguren. In die optiek krijgen ook onze speculaties uit 5.7 in verband met de opsplitsing tussen een economische en een intellectuele elite overigens wat meer fundamenteel: we zien occA bijvoorbeeld inderdaad ongeveer in het midden van de opleidingsgraden ‘hie’ en ‘mid’ staan, terwijl ‘hie’ zich dan weer vrijwel in het brandpunt van de driehoek occC-occH-occB bevindt. Aan de Soapvlaamse rechterkant constateren we ten slotte nog evenmin een verschil is tussen het taalgebruik in component a (face-to-face conversaties) en dat van de pragmatische generatie – en bij uitbreiding het taalgebruik van de educatieve middencategorie.

### 6.3 Variëteiten – interacties

Naast een vergelijking van de hoofdeffecten zijn we ook geïnteresseerd in de manier waarop de verschillende externe factoren met elkaar interageren. We geven daarom de interactie voor elk paar van factoren, de zogenaamde tweewegse interacties<sup>5</sup>. Voor zes factoren – register, regio, generatie, sekse, educatieniveau, en beroepsniveau – zijn er dat vijftien.

#### 6.3.1 Register & Regio

Als we de registerverdeling voor de vier regio's met elkaar vergelijken<sup>6</sup> (Figuur 6.6, 6.7, 6.8, en 6.9), dan merken we dat de regio's grotendeels de verhouding van de registers volgen: de gesprekssituatie heeft met andere woorden een grotere impact op het taalgebruik dan de dialectische achtergrond van de spreker. Enkel de drie informele componenten a, c, en d kennen ten opzichte van hun respectieve hoofdeffecten veel interne variatie.

Voor Limburg zijn c en d bijvoorbeeld meer naar het centrum opgeschoven, en a naar beneden. Het gedrag van c en d is daarbij typerend voor Limburg: Limburg vertoont in het geheel minder heterogeniteit dan de andere drie regio's – vergelijk de plot van Brabant, dat het taalcentrum heet van de verbreiding van het Soapvlaams. Zulks wijten we aan het feit dat de typisch Limburgse dialectkenmerken (de bekende *ich*, *mich*, en *dich* – zie 3.1) niet in het corpus vertegenwoordigd zijn. Daarmee beschikken we voor de Limburgers enkel over het bovengewestelijk neutrale Belgisch-Nederlands, want dat is de spreekstijl die zij in algemenere situaties

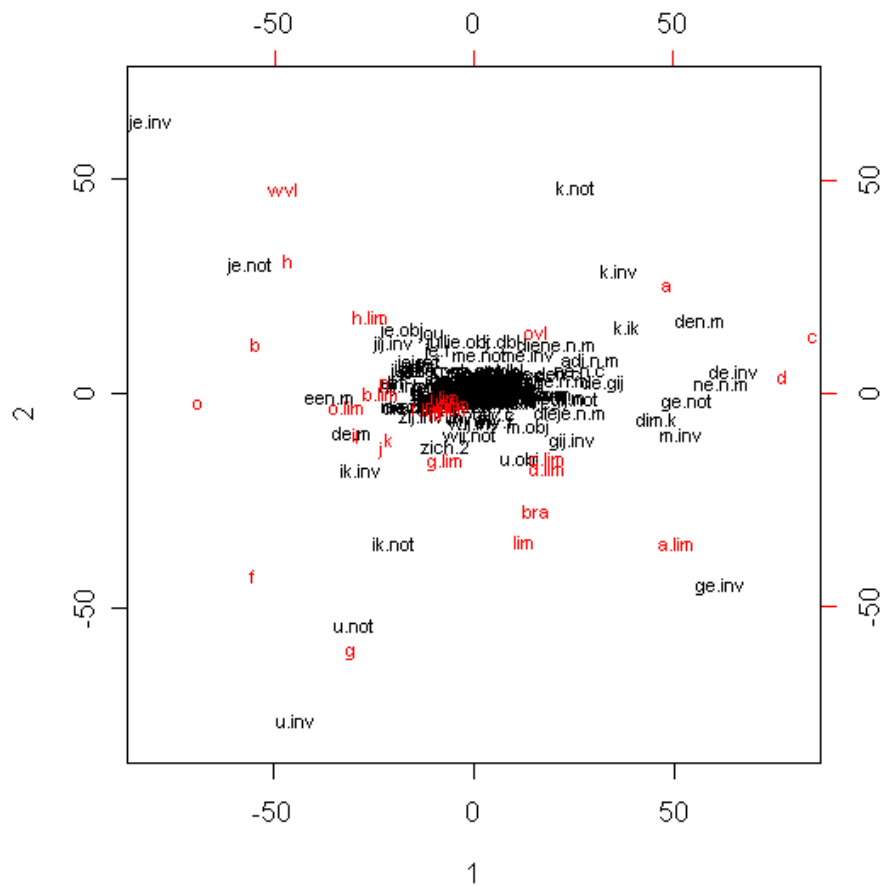
<sup>5</sup> De drie-, vier-, en vijfwegse interacties laten we hier buiten beschouwing, omdat het ons te ver zou voeren; geïnteresseerden kunnen ze evenwel vinden op de CD-ROM.

<sup>6</sup> De precieze presentatie van de figuren wordt vooral door praktische overwegingen ingegeven; de volledige serie van interactieplots staat op de CD-ROM.

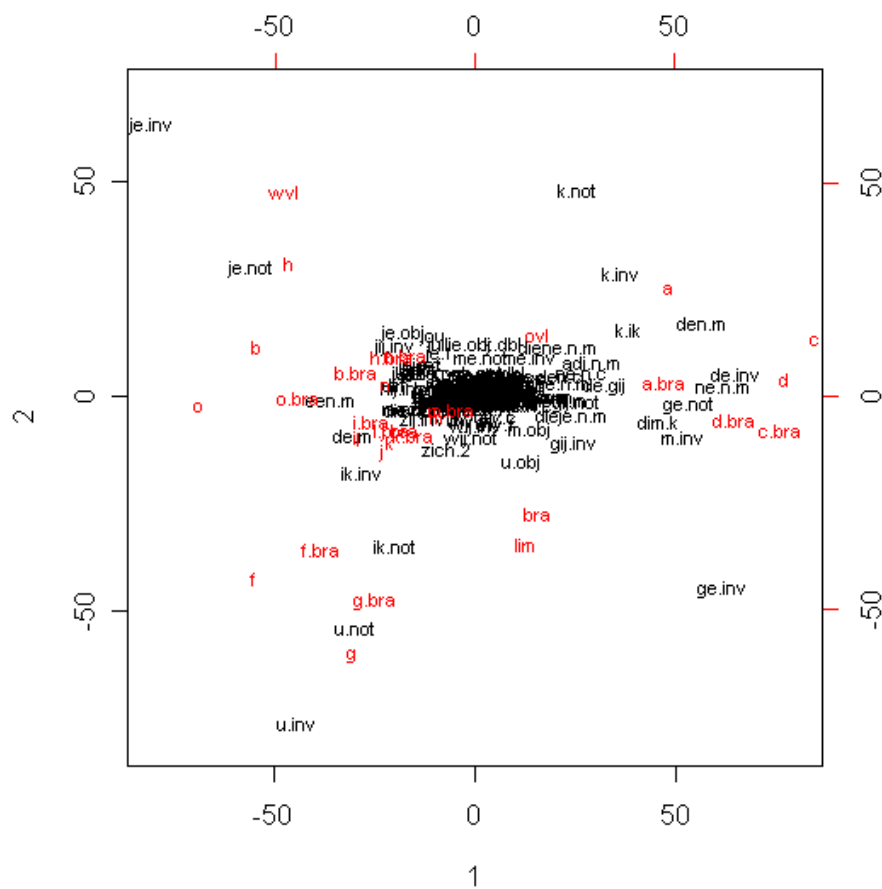
hanteren (wat dan meteen ook het verschil tussen c en d enerzijds en a anderzijds verklaren zou: telefoongesprekken zijn namelijk al sneller interregionaal dan conversaties met een vriend). Voeg daar nog bij dat de Brabantse kenmerken – die zo sterk de kleur van het Soapvlaams bepalen – in Limburg weinig of niet ingeburgerd zijn, dan wordt het begrijpelijk waarom Limburg zich ‘terugtrekt’ in het centrum van de plot. Ook de lage positie van component a kan in dit licht begrepen worden: a is met name geassocieerd met de *ge.inv* rechtsonder. Zoals we bij de bivariate analyse al opgemerkt hebben (zie 5.2), is *ge* in Limburg de gebruikelijke informele aanspreekvorm, en komen *–de* (Brabants) of *je* (Vlaams en Journaalnederlands) er nauwelijks voor.

Verder bevestigen de plots het statuut van Oost-Vlaanderen als dialectische overgangszone: aan de ene kant is het Oost-Vlaamse registerpatroon vrijwel een kopie van het Brabantse, maar aan de andere kant is het als geheel toch meer naar de ‘Vlaamse’ bovenhelft van de plot verschoven – vergelijk vooral weer de posities van de componenten a,c, en d. West-Vlaanderen stemt daarbij ten slotte overeen met Oost-Vlaanderen, alleen valt er hier dan weer een verschuiving naar linksboven waar te nemen: de plaats van de J-vormen, die zoals we inmiddels weten (zie 5.2) homoniem zijn met de endogeen West-Vlaamse dialectvarianten.

Onze conclusie in verband met de interne variatie van het Belgisch-Nederlands luidt dan ook dat op de informele taalsituaties na het taalgebruik in Vlaanderen regionaal gezien sterk gehomogeniseerd is. Enkel in de registers die gelden als de representanten van het Soapvlaams bij uitstek – a, c, en d – komt de dialectische achtergrond van de regio’s nog sterk tot uiting (waarbij we in het achterhoofd dienen te houden dat het corpus voor Limburg niet het volledige variatiespectrum bevat).



Figuur 6.6: Interactie Register – Limburg



Figuur 6.7: Interactie Register – Brabant



### 6.3.2 Register & Generatie

In de plots van register en generatie (Figuur 6.10, 6.11, 6.12, 6.13 en 6.14) zien we voor het eerst duidelijk de voortschrijding van de standaardisering. Dat neemt met name de vorm aan van een steeds toenemende *functionele differentiatie*. Waar de eerste twee generaties registermatig nog behoorlijk homogeen zijn – de interacties voor de vooroorlogse en de stille generatie worden daarbij vooral bepaald door de hoofdeffecten bij generatie – vertonen de volgende twee generaties meer variatie: de interacties voor de protst- en de verloren generatie volgen eerder de hoofdeffecten van de registerverdeling. Opvallend is met andere woorden dat het gebruik van Soapvlaamse elementen in meer informele situaties niet geschuwd wordt, getuige de posities van componenten a, c en d. Met de pragmatische generatie wordt ten slotte de registerdifferentiatie enerzijds wel bewaard, maar ze vertoont anderzijds een algehele opschuiving richting Soapvlaams.

Met de vermaarde socioloog Norbert Elias interpreteren we deze voortgang in termen van een toegenomen standaardtaalbeheersing. We weten namelijk dat er vanaf de Tweede Wereldoorlog in Vlaanderen een generatie opstaat die beter taalonderwijs genoten heeft – zie 5.3. In zijn magnum opus *Über den Prozess der Zivilisation* (in 1982 in het Nederlands vertaald) merkt Elias met betrekking tot versoepeling van omgangsvormen dan ook op dat die optreedt “in een samenleving waarin en hoge mate van terughoudendheid vanzelfsprekend is geworden en waarin zowel vrouwen als mannen er absoluut zeker van kunnen zijn dat ieder individu in toom wordt gehouden door een sterke zelfbeheersing en strikt geregelde omgangsvormen. Het is een losser worden, een versoepeling die volkomen binnen het kader van een bepaald ‘gecivileerd’ standaardgedrag blijft, dat wil zeggen binnen het kader van de binding en omvorming van de affecten, die in zeer hoge mate als gewoonte aangekweekt en automatisch geworden zijn” (Elias 1982: 253-254). Hoe beter een spreker met andere woorden zijn normen kent, hoe sneller hij er ook afwijkingen van zal introduceren, en wel precies omdát hij die normen beter kent: daarmee is hij namelijk meer in staat om te taxeren waar en wanneer zulks geoorloofd is.

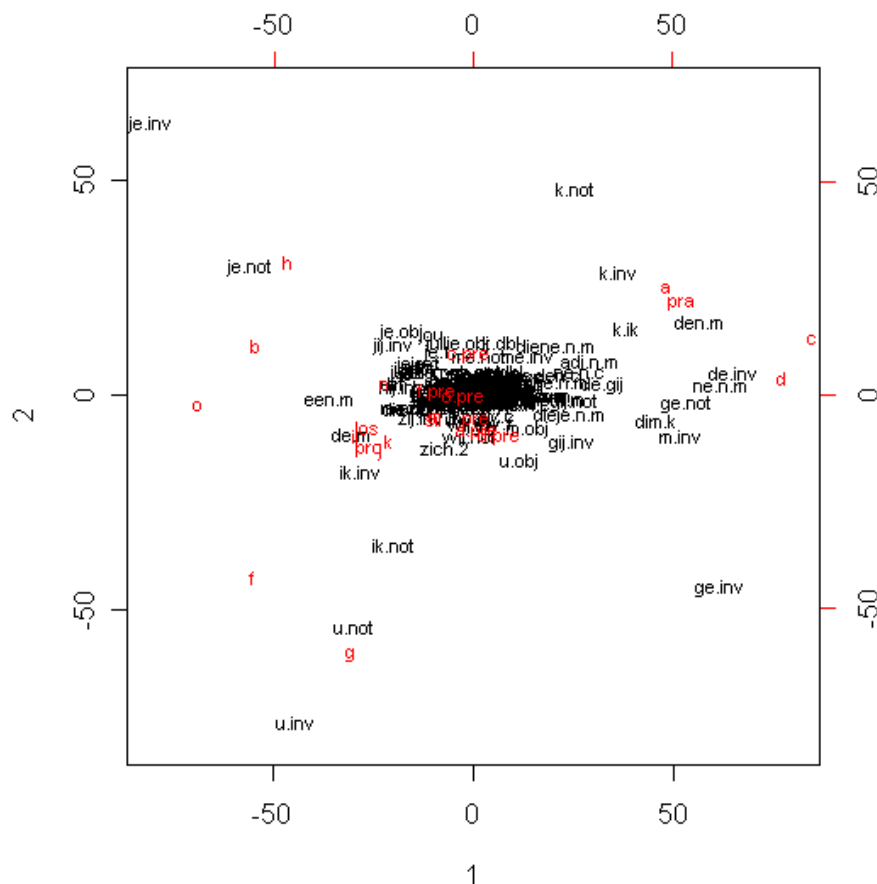
Ook de ‘verstrakking’ in de daarop volgende fase, zoals hier de pragmatische generatie die laat zien, kan in het licht van Elias’ civilisatietheorie geduid worden. Met name Elias’ Amsterdamse volgeling Cas Wouters (naast Joop Goudsblom en de in 5.3 vermelde Abram de Swaan) schrijft over de jaren 80:

“De informalisering van de omgangsvormen stagneerde, maar in de eropvolgende fase van formalisering werden niet alle informalisering uit de voorafgaande fase tenietgedaan. Veel van de gedrags- en gevoelsalternatieven die in het laatste informaliseringstijdperk ingang vonden, bleven intact; ze werden in de fase van formalisering in de standaard van

gevoel en gedrag overgenomen en in die zin geformaliseerd.” (Wouters 1990: 110)

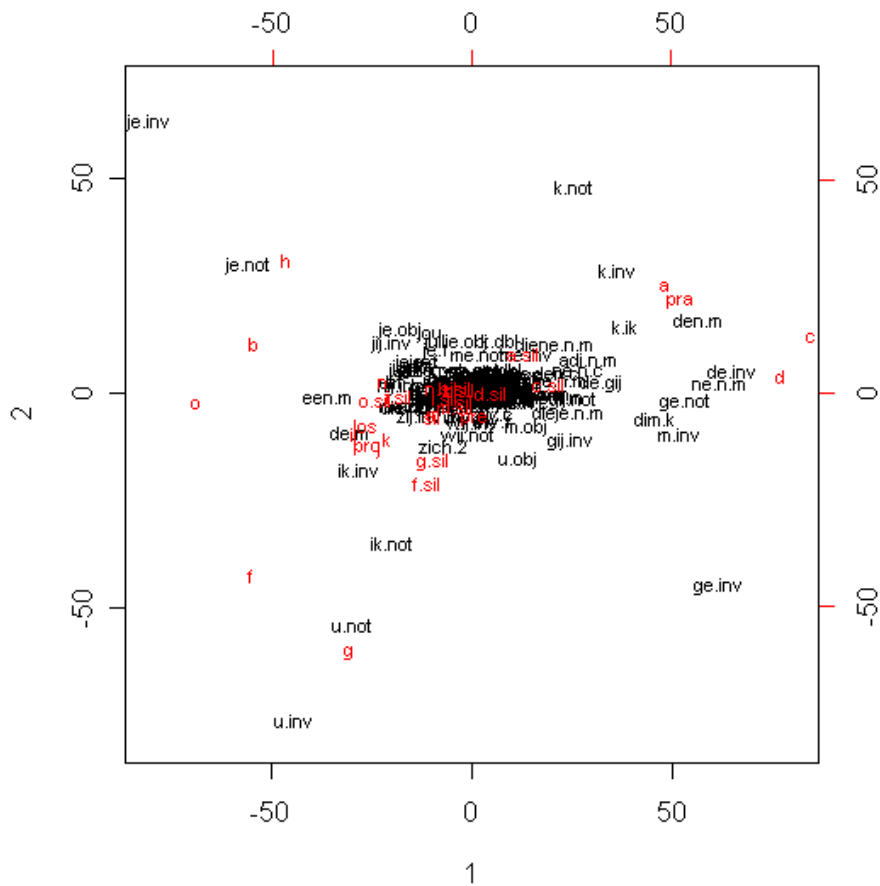
Dat sluit dan aan bij wat we in 5.3 en 5.4 over endogenisering en autonomisering opgemerkt hebben: naarmate een stijgend aantal Vlamingen tot de maatschappelijke elite en daarmee ook de spraakmakende gemeente gaan behoren, zal de voorhanden differentiatie steeds meer opgevat worden als een inherent onderdeel van de talige conventie zoals elke taalgebruiker die moet internaliseren. De variatie wordt kortom ingeschreven in de grammatica.

Toch willen we ten slotte nog de aandacht vragen voor het feit dat de geschetste differentiërende ontwikkeling haar startpunt vindt in het centrum van de plot – de positie van de twee oudste generaties, maar ook de plaats waar de Soapvlaamse varianten en de Journaalnederlandse varianten elkaar ‘raken’. Dat lijkt er namelijk op te wijzen dat het continuüm tussen beide spreekstijlen (vergelijk 5.1) in zekere zin ‘opgeheven’ wordt: het standaardiseringsproces in Vlaanderen behelst net een *uitkristallisatie* ervan in verscheidene variëteiten – in die optiek kunnen we hier trouwens ook een bevestiging in zien van Craps’ vermoeden omtrent de continuïteit van het substandaard taalgebruik (zie hoofdstuk 1): Soapvlaams heeft altijd al bestaan; het wordt nu alleen scherper gedistingeerd.

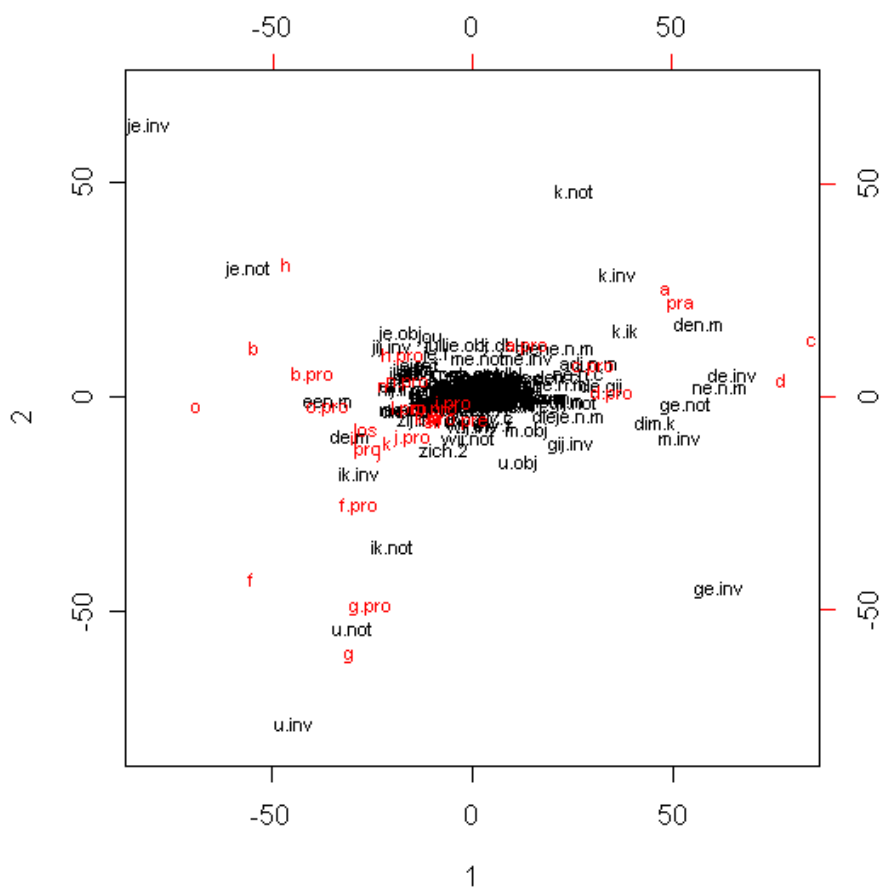


Figuur 6.10: Interactie Register – Vooroorlogse generatie

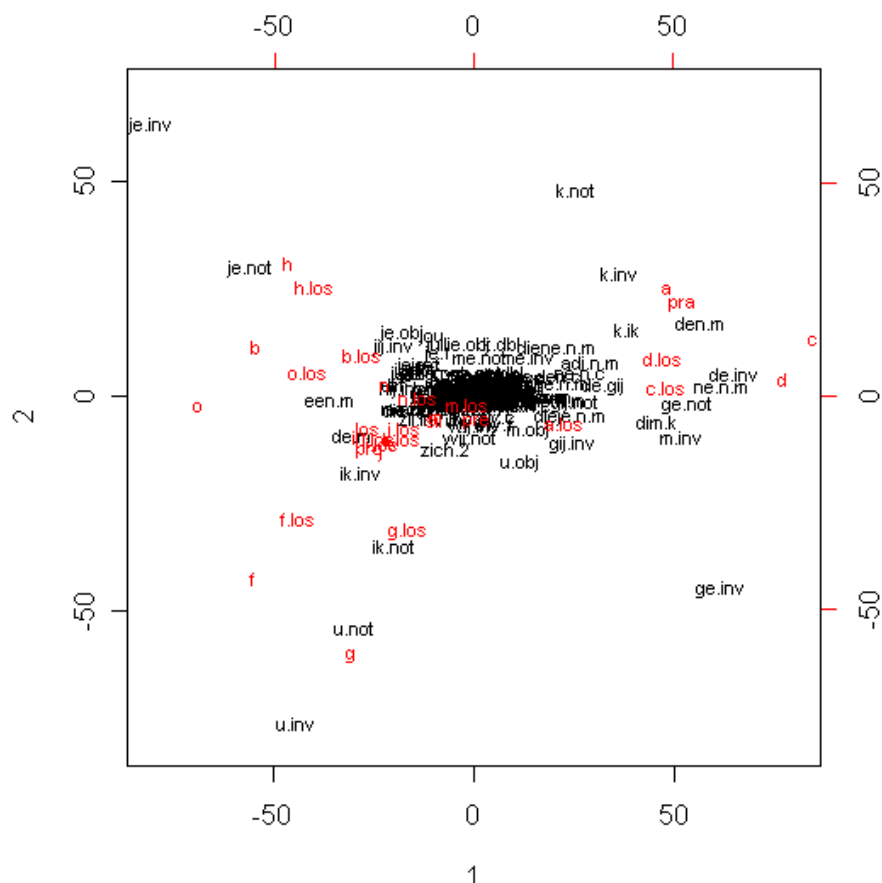




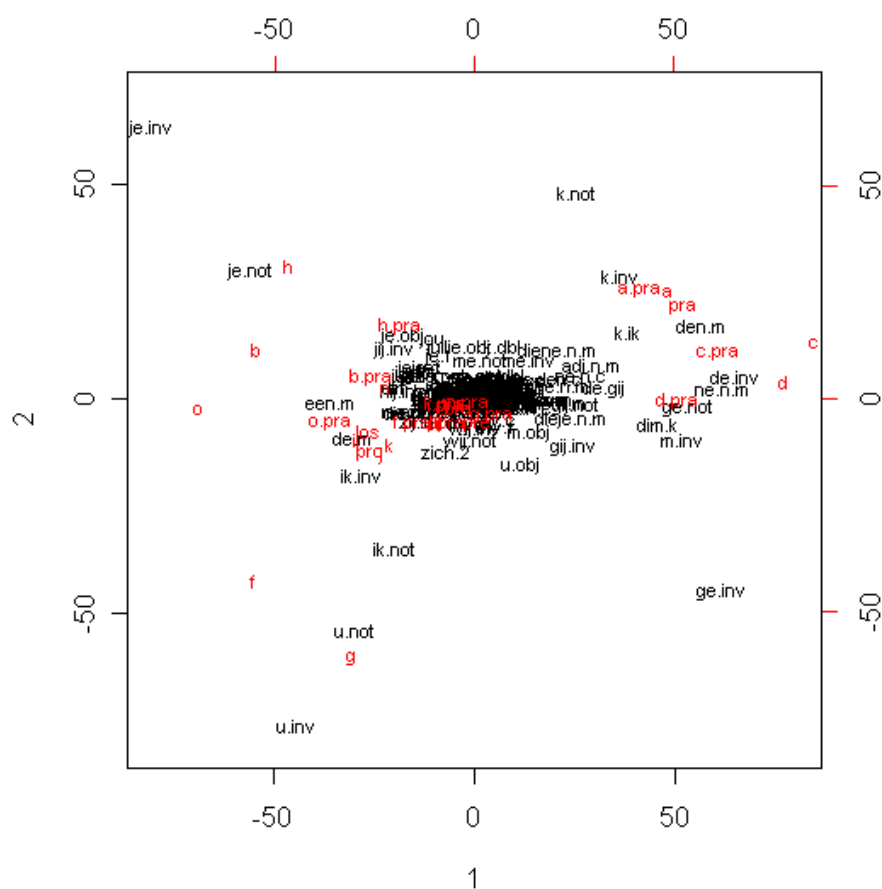
Figuur 6.11: Interactie Register – Stille generatie



Figuur 6.12: Interactie Register – Protestgeneratie



Figuur 6.13: Interactie Register – Verloren generatie



Figuur 6.14: Interactie Register – Pragmatische generatie

### 6.3.3 Register & Sekse

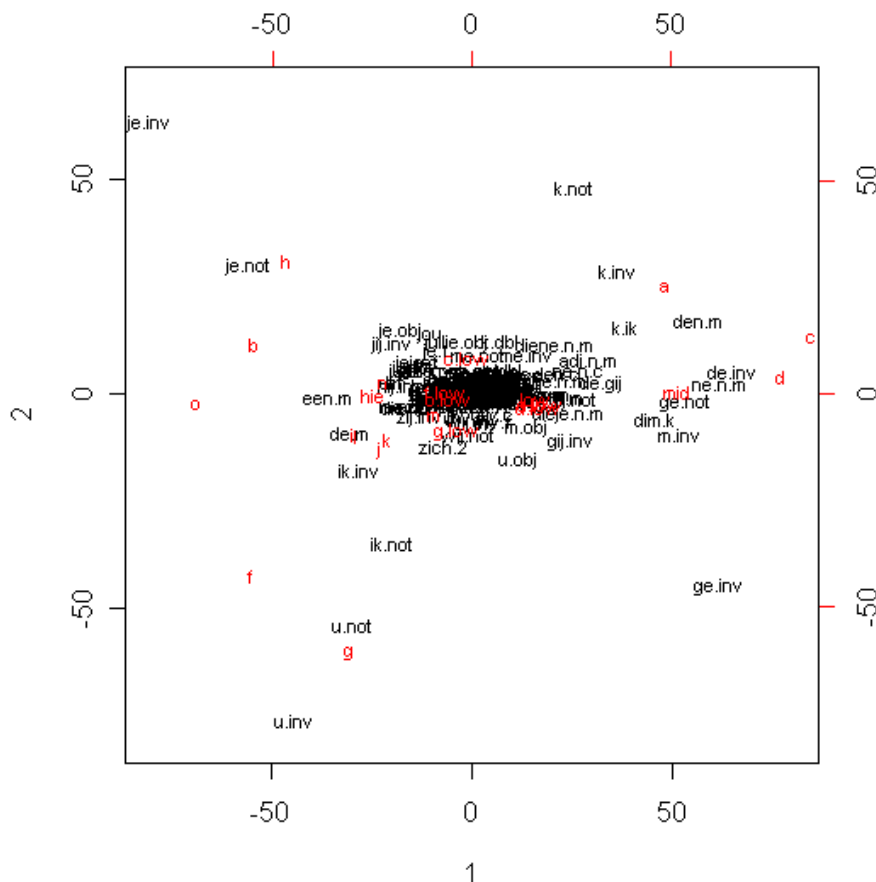
Voor register en sekse stellen we enerzijds vast dat de interacties bij beide sekse nogmaals hoofdzakelijk volgens het registerpatroon verdeeld zijn; anderzijds respecteren ze eveneens de diagonaal tussen vrouwelijk rechtsboven en mannelijk linksonder (Figuur 6.15 en 6.16). Opnieuw zijn het daarbij de informelere registers, in casu de telefoondialogen, die de meeste interne variatie vertonen: in component c verschuiven de mannen naar rechtsboven in plaats van naar linksonder zoals in component d (in b staan de mannen dan weer meer naar linksboven dan de vrouwen, maar het verschil zou wel eens te gering om veelbetekenend te kunnen zijn).

De interacties tussen register en sekse zijn bijgevolg niets anders dan de gecombineerde effecten van beide hoofdfactoren (anders uitgedrukt, 'ze bestaan eigenlijk niet'). De enige uitzondering vormt component h (lessen), waar de interactie met de vrouwen sterker is dan het hoofdeffect: blijkbaar doen vrouwelijke leerkrachten voor de klas extra hun best om JournaalNederlands te spreken – let in dat verband overigens op het toch wel markante verschil met hun mannelijke collega's (h.M) in het neutraal Belgisch-Nederlandse centrum van de plot.



### 6.3.4 Register & Educatieniveau

Bij de interactie tussen register en educatieniveau zien we hetzelfde effect opduiken als bij register en generatie (Figuur 6.17, 6.18, en 6.19): naarmate de kwaliteit van het onderwijs stijgt, treedt er functionele differentiatie op. Zo clusteren de interacties van het laagste educatieniveau allemaal samen rond het hoofdeffect in het centrum van de plot (dit wil zeggen: de interacties die ook daadwerkelijk beschikbaar zijn, want voor deze opleidingscategorie ontbreken er nogal wat componenten: met name b, h, i, j, k, l, m, en n) – ook hier vormt het centrum kortom het startpunt voor de taalontwikkeling. De distributie bij het hoogste educatieniveau wordt daarnaast vrijwel geheel bepaald door de hoofdeffecten van de registers. Ten slotte zit de educatieve middencategorie daar overal ergens tussenin (hoewel er ook hier enkele componenten ontbreken: b, h, en m – daardoor vallen de interacties met de hoogste opleidingsgraad voor deze drie registers volledig samen met hun respectieve hoofdeffecten). De verklaring is daarbij eveneens dezelfde: beter onderwijs leidt tot een betere vertrouwdheid met de gangbare stijlverschillen. Ook voor het educatieniveau impliceert zulks geenszins een afwijzing van de Soapvlaamse varianten in de meer informele taalsituaties.



Figuur 6.17: Interactie Register – Laag opgeleid



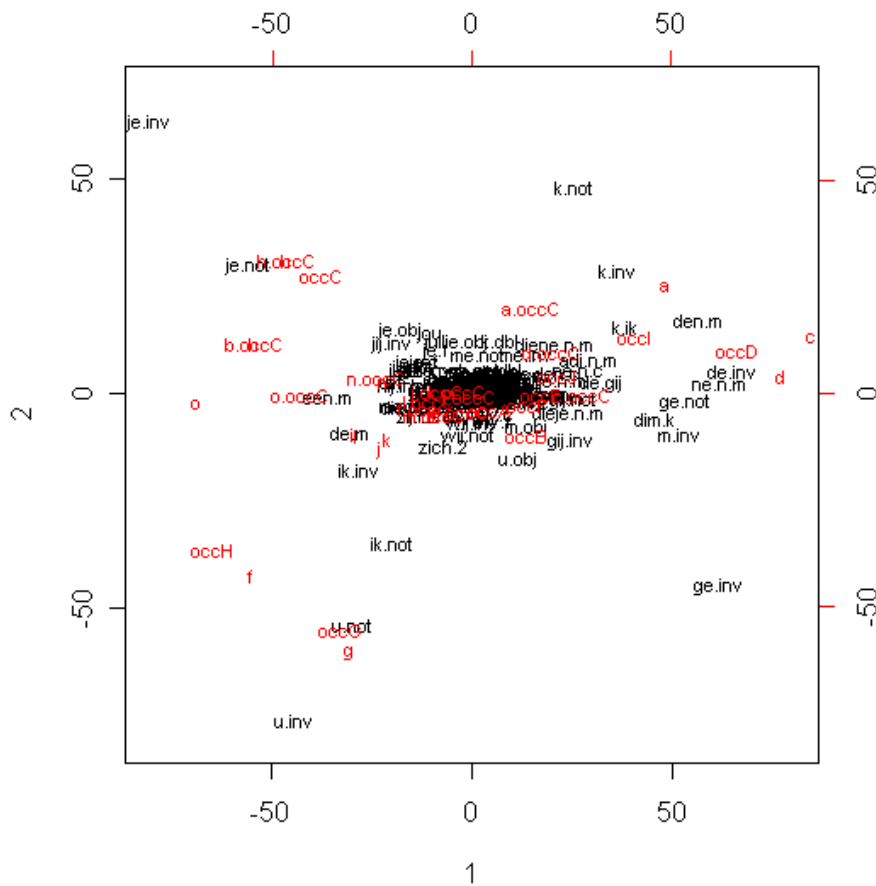
### 6.3.5 Register & Beroepsniveau

Voor de interactie tussen register en beroepsniveau moeten we op voorhand opmerken dat er voor heel wat combinaties geen sprekers en dus geen treffers in het corpus zijn (waarvan we wegens technische redenen geen overzicht kunnen geven). Bij de interpretatie is dus enige behoedzaamheid geboden. Van de interacties die er wel zijn geven Figuren 6.20 tot en met 6.29 de verdeling van de registers voor elk beroepsniveau. We constateren een opvallend grote mate van homogeniteit, zowel binnen de beroepscategorieën (de componenten clusteren voor elke categorie samen in het centrum van de plot) als ertussen (de plots voor de verschillende categorieën lijken allemaal op elkaar). Enkel occC (academici) is enerzijds ietwat gevoeliger voor de posities van componenten b-h-o, terwijl occD (ambtenaren) en occl (studenten) anderzijds dan weer meer voor die van a-c-d – met occD toch nog iets prominenter dan occl. Deze effecten verschillen evenwel niet zo heel erg veel dan van wat louter op basis van de hoofdeffecten te voorspellen is. Ook voor register en beroepsniveau lijken de interacties met andere woorden neer te komen op de gecombineerde effecten van beide hoofdfactoren.

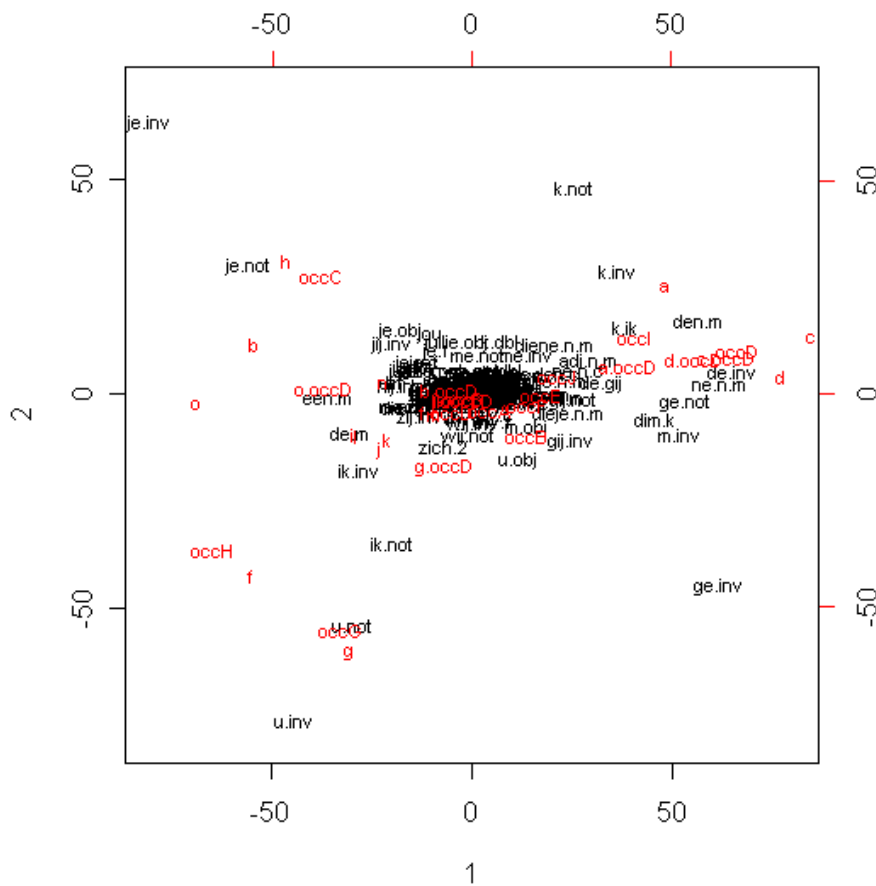
Het enige beroepsniveau dat verder nog een vermeldenswaardig effect laat zien is occH (mediafiguren). De interacties voor occH vallen namelijk nagenoeg samen met de hoofdeffecten bij de registers voor die componenten die ‘uitgezonden op radio en televisie’ zijn, te weten: i, j, k, l, en f. Dat is dan volgens verwachting: de sprekers die uit hoofde van hun beroep de belichaming van het JournaalNederlands (zouden moeten) zijn, zijn dat bij de uitoefening ervan – als ze op radio of televisie verschijnen – ook daadwerkelijk. Vergelijk component f vervolgens echter met g, het register dat eveneens interviews en debatten bevat, alleen met dat verschil dat deze *niet* uitgezonden op radio en televisie zijn. De interactie die voor component g met het hoofdeffect correspondeert is niet die van occH (maar die van occG: politici); de interactie g.occH bevindt zich gewoon in de centrumcluster (tussen de hoofdeffecten van component m en occA, meer bepaald). De conclusie die we hieraan verbinden is dat de mediafiguren ‘tijdens de officiële werkuren’ dan wel het toonbeeld van het JournaalNederlands mogen zijn; daarbuiten zijn ze het beslist niet: dan hanteren ze meer het neutrale Belgisch-Nederlands.



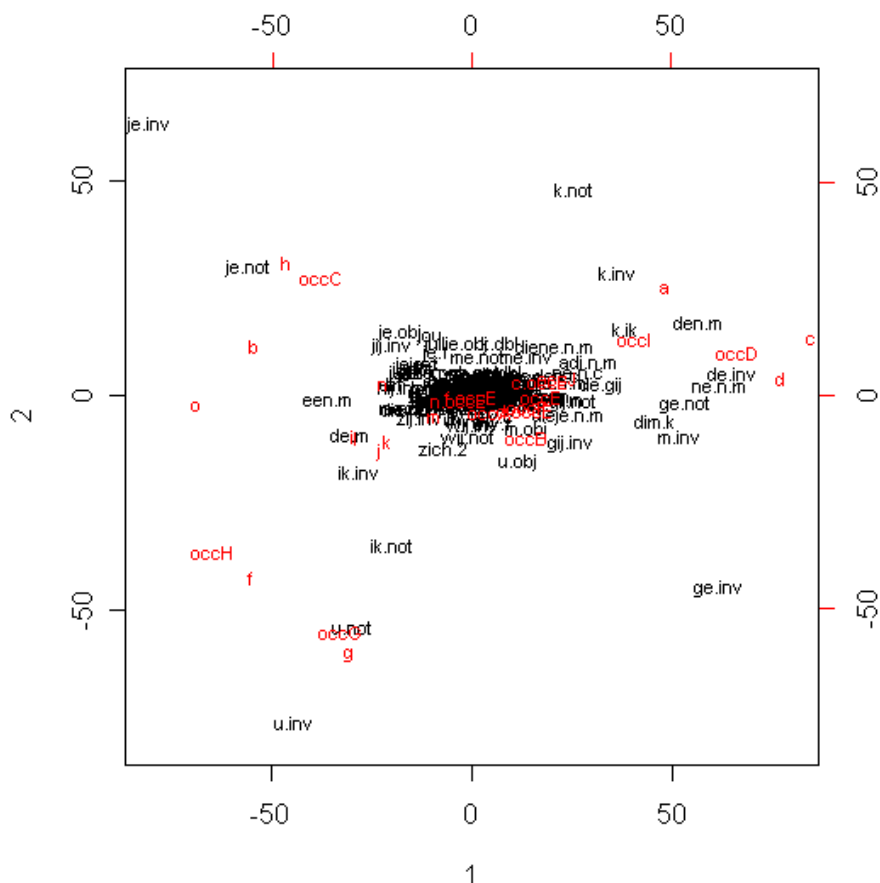




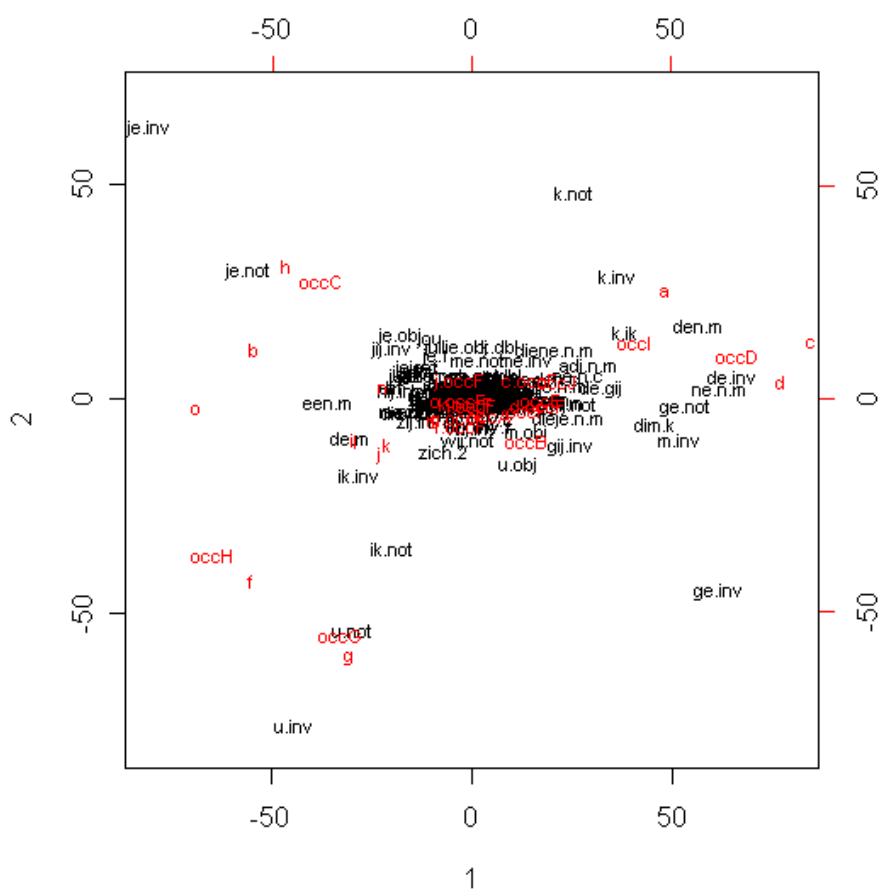
Figuur 6.22: Interactie Register – Academics (occC)



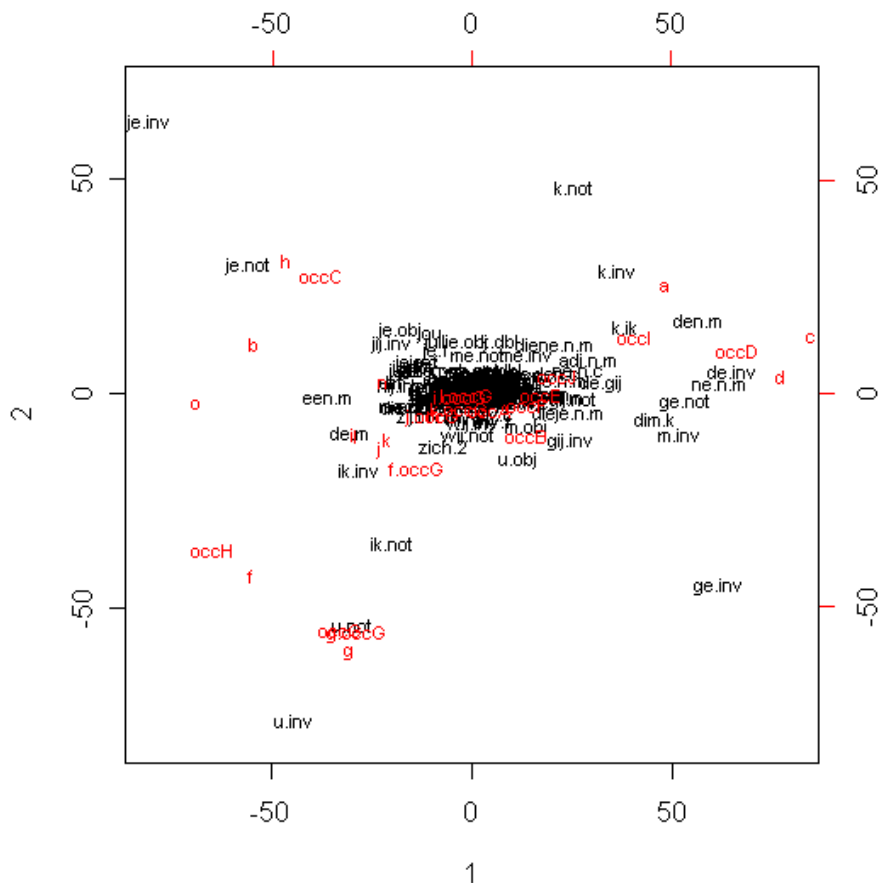
Figuur 6.23: Interactie Register – Ambtenaren (occD)



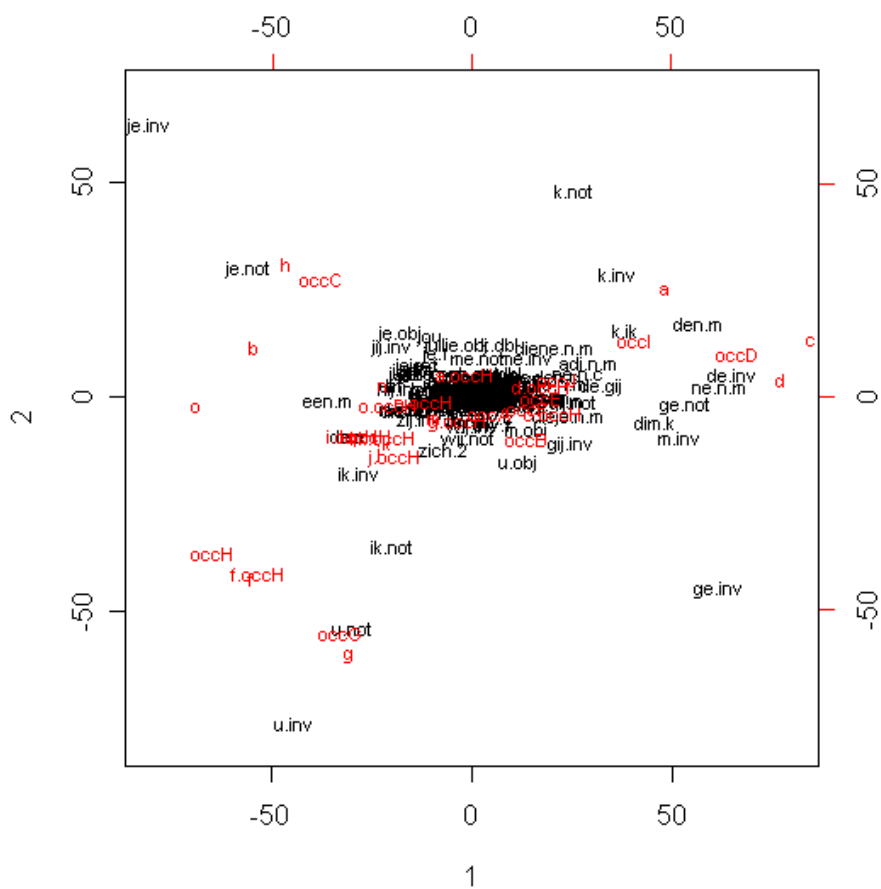
Figuur 6.24: Interactie Register – Ongeschoolden (occE)



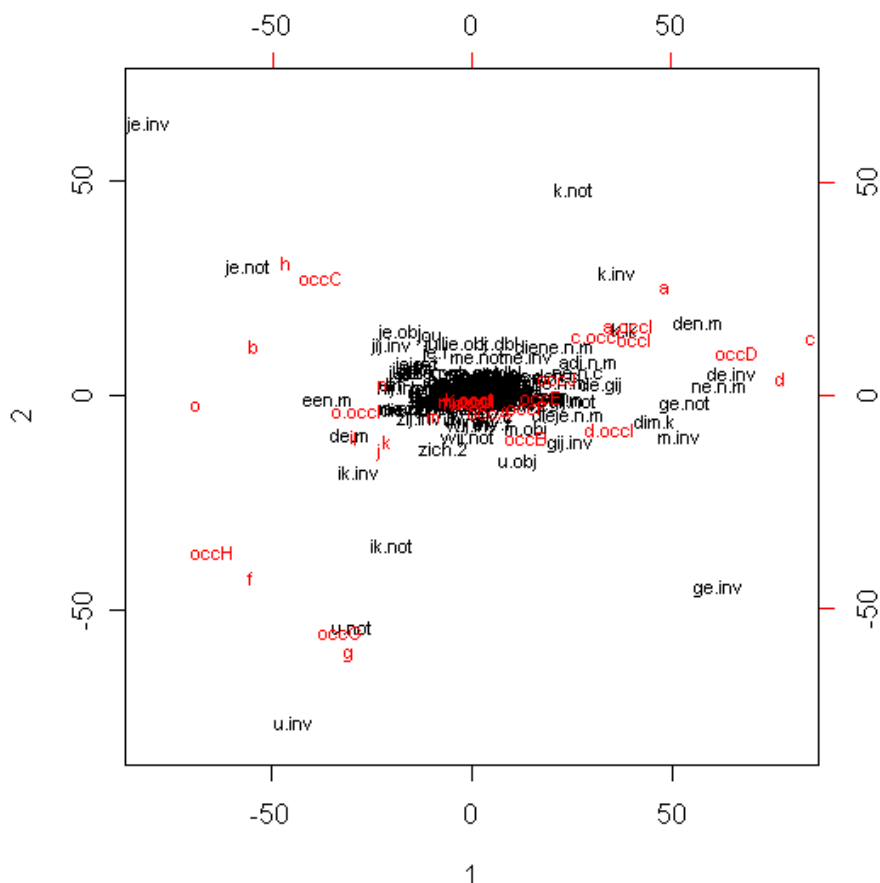
Figuur 6.25: Interactie Register – Zelfstandigen (occF)



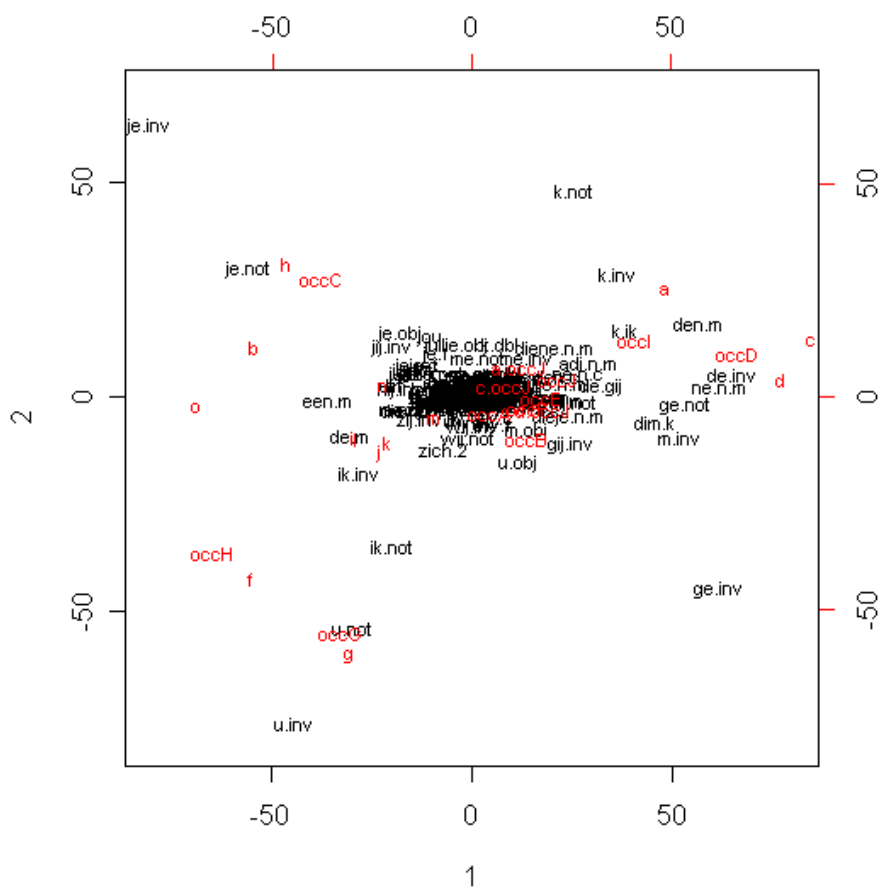
Figuur 6.26: Interactie Register – Politici (occG)



Figuur 6.27: Interactie Register – Mediafiguren (occh)



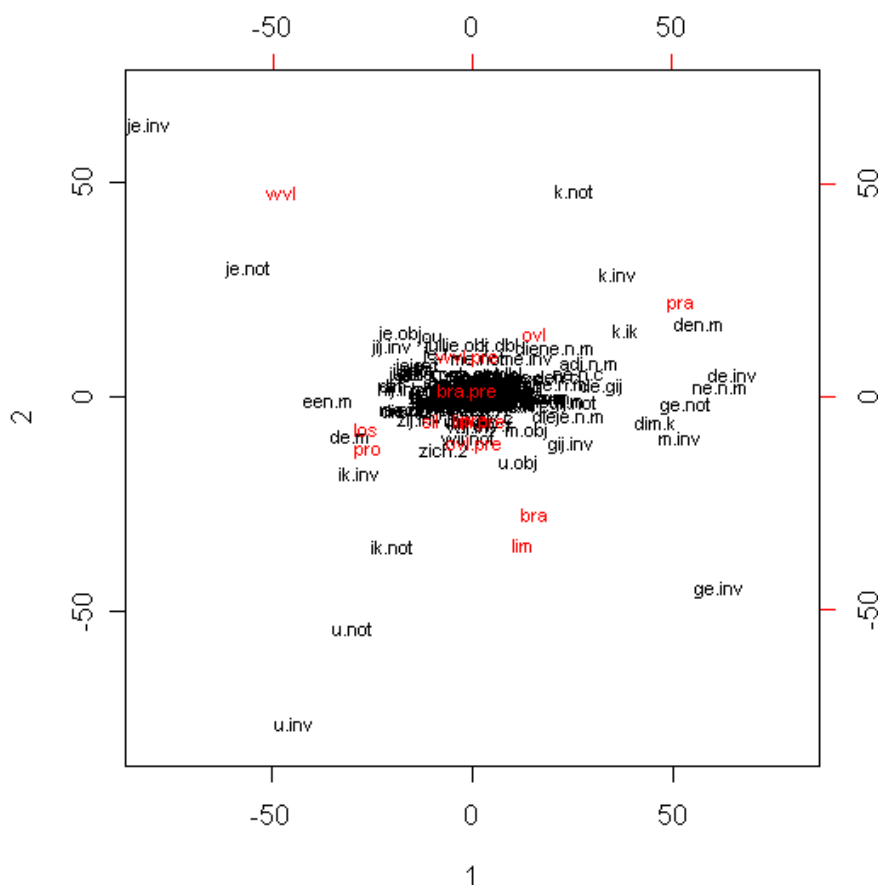
Figuur 6.28: Interactie Register – Studenten (occl)



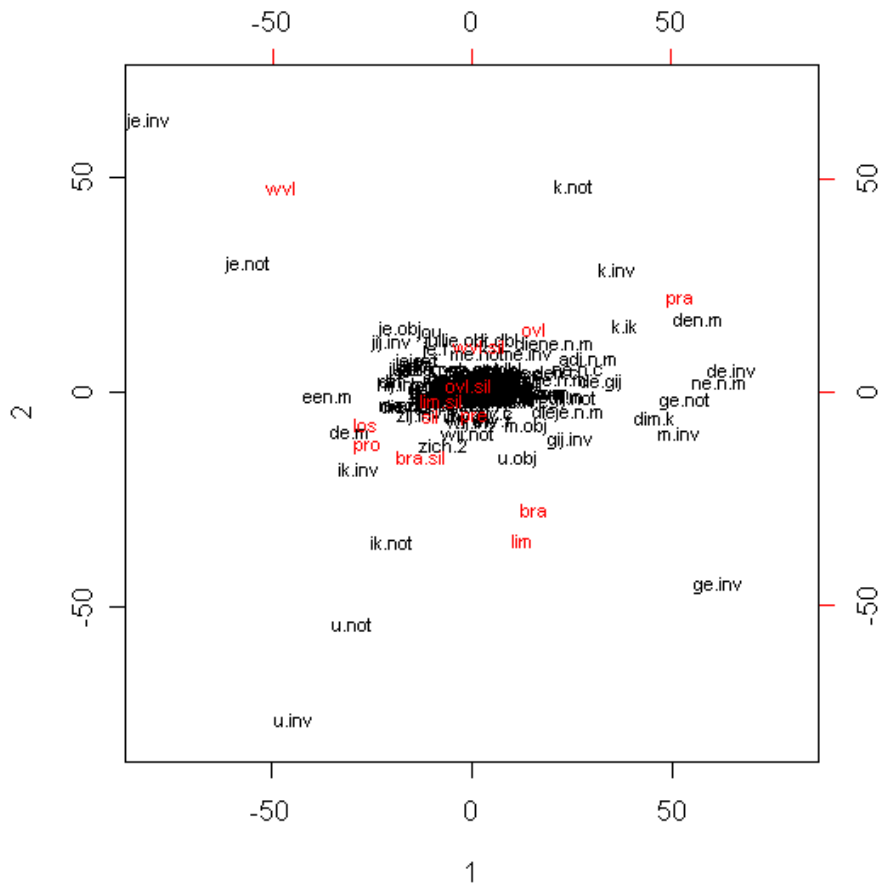
Figuur 6.29: Interactie Register – Werklozen (occJ)

### 6.3.6 Regio & Generatie

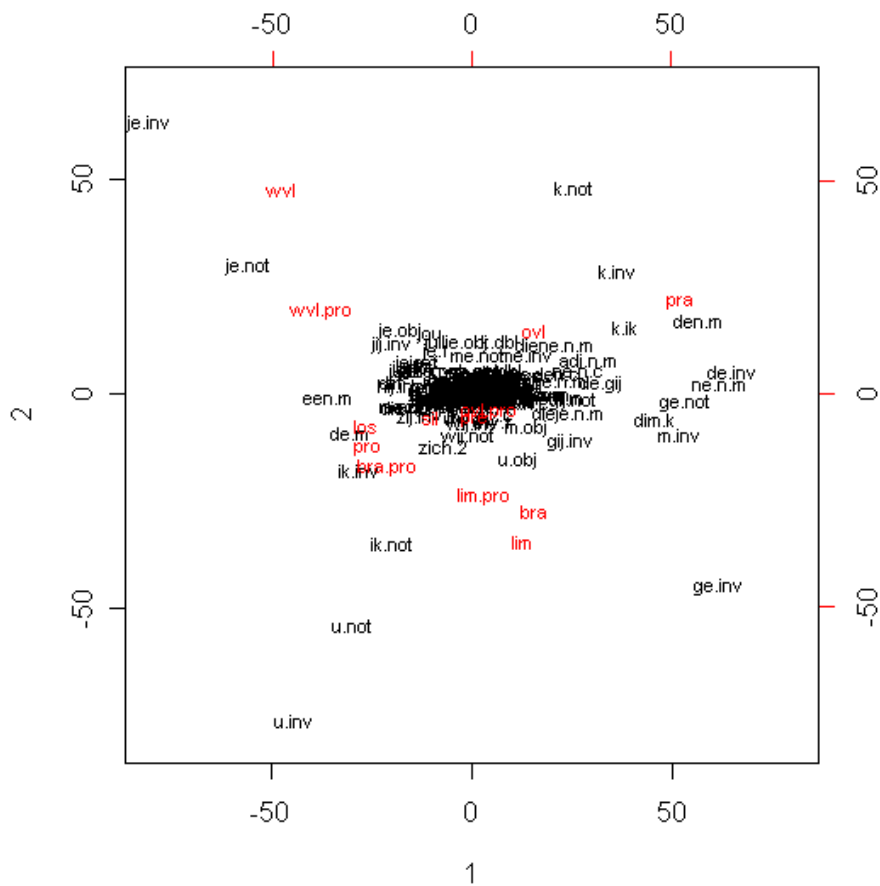
De plots voor regio en generatie (Figuur 6.30, 6.31, 6.32, 6.33, en 6.34) onderstrepen nogmaals de revolutie die de pragmatische generatie in de aanblik van het Vlaamse taallandschap heeft ingeluid. Van de vooroorlogse tot en met de verloren generatie is er een lage graad van regionale heterogeniteit (met *wvl.pro* en *lim.pro* als enige uitzonderingen: de dialectische variatie is bij de generatie die zoals we weten als eerste de standaardisatiebeweging inzet – de protestgeneratie – duidelijk nog niet helemaal weggewerkt; daar staat tegenover dat de verloren generatie dan weer een algehele verschuiving naar de Journaalnederlandse linkerkant van de plot laat zien). Met de pragmatische generatie wordt het taalgebruik evenwel opeens sterk regiolectisch gekleurd. Vooral de provincies West-Vlaanderen en Limburg springen daarbij opnieuw in het oog: hun interacties vallen vrijwel samen met de hoofdeffecten. De kroon wordt echter gespannen door Brabant, dat in de positie van *bra.pra* uiterst rechts – het is de meest extreme uitschieter van de hele analyse – geen twijfel laat bestaan omtrent het voortouw dat deze regio neemt in de verbreiding van het Soapvlaams.



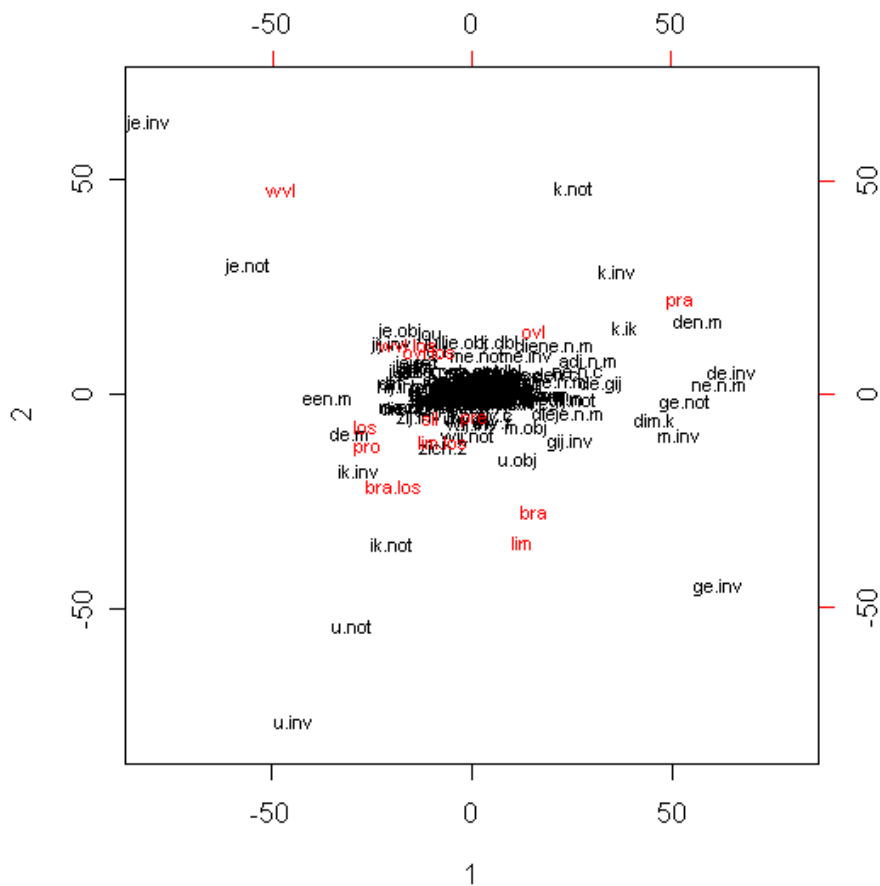
Figuur 6.30: Interactie Regio – Vooroorlogse generatie



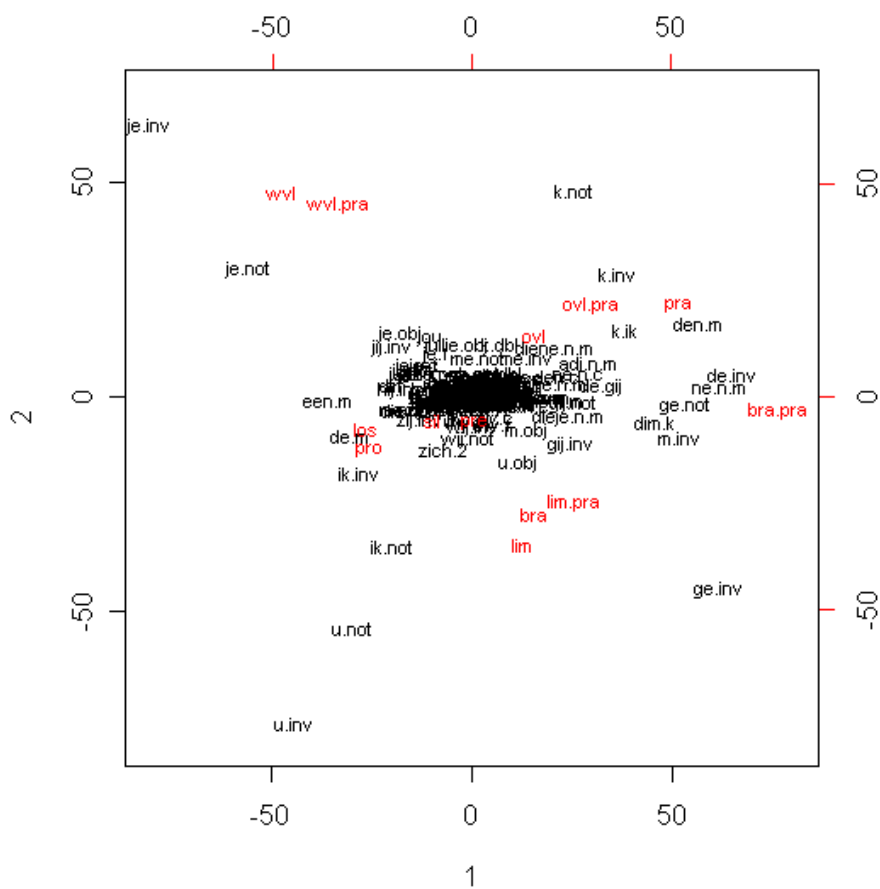
*Figuur 6.31: Interactie Regio – Stille generatie*



*Figuur 6.32: Interactie Regio – Protestgeneratie*



Figuur 6.33: Interactie Regio – Verloren generatie

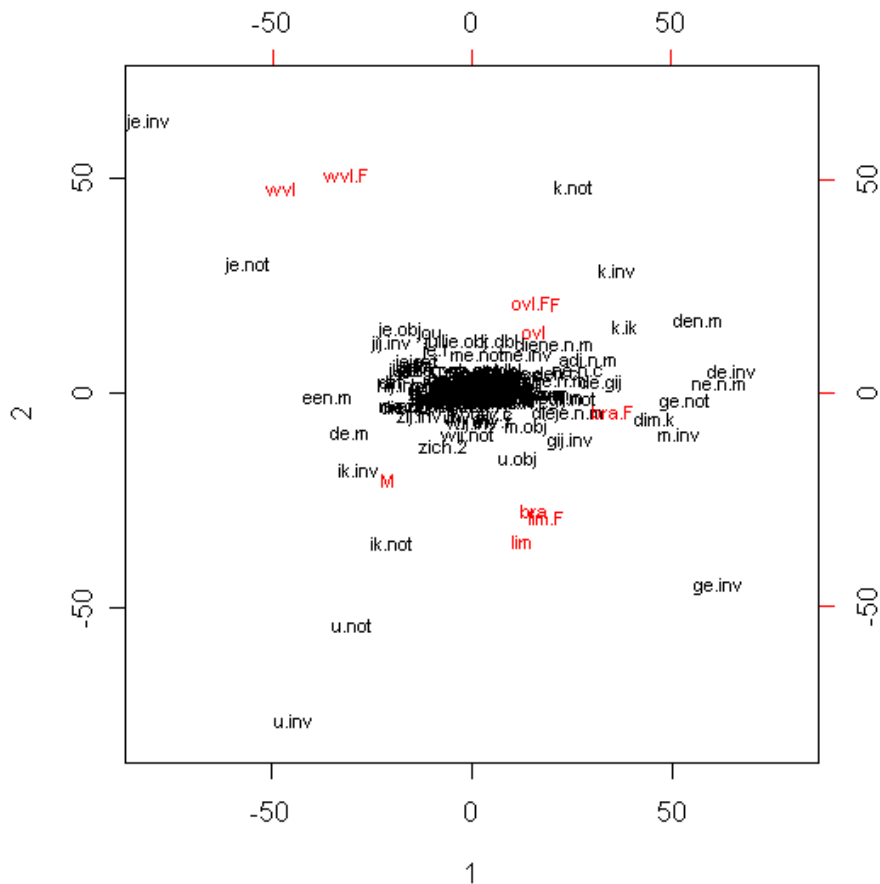


Figuur 6.34: Interactie Regio – Pragmatische generatie

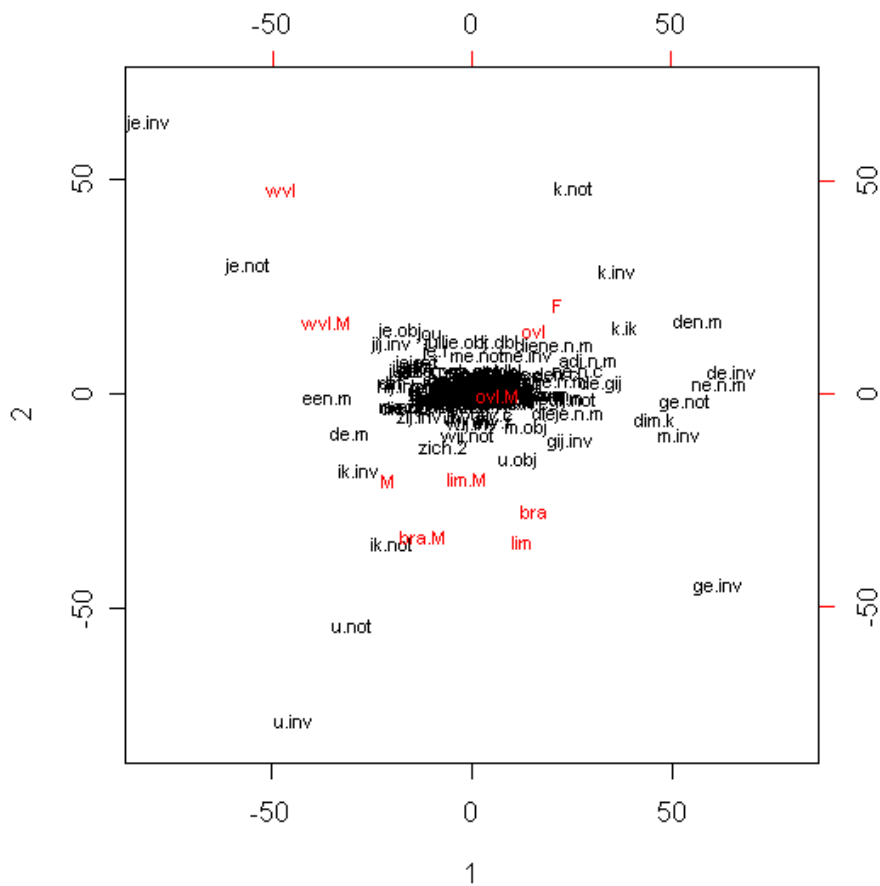
### 6.3.7 Regio & Sekse

Bij regio en sekse merken we weer dat de interacties de diagonaal van rechtsboven naar linksonder respecteren (Figuur 6.35 en 6.36). Opvallend is daarbij dat de interacties voor de mannen stevast ergens in het midden staan tussen enerzijds het hoofdeffect van de mannelijke sprekers en anderzijds dat van de regio, terwijl de interacties voor de vrouwen – op één uitzondering na – vlak bij de regio-effecten zelf te vinden zijn. De interpretatie die we daaraan hechten is dan dat vrouwen toch meer aan hun dialectische achtergrond vasthouden dan mannen (hoewel die hun tongval nu ook weer niet geheel opgeven, getuige de nog steeds grote spreiding van de regio's). De uitzondering waarvan sprake zijn de vrouwelijke sprekers uit Brabant, die net stellig de associatie met het Soapvlaams opzoeken (een eventuele tweede zijn de mannelijke Limburgers, die zich in vergelijking met de vrouwen naar linksboven bewegen; een andere verklaring dan de in 6.3.1 vermelde afwezigheid van kenmerkend Limburgse dialectvormen, waardoor het taalgebruik in Limburg als vanzelf homogener lijkt, hebben we daarvoor evenwel niet). Daarmee is het Soapvlaams het Poldernederlands van Vlaanderen: ook het Soapvlaams wordt blijkbaar gedragen door (jonge en hoog opgeleide) vrouwen uit het taalcentrum – zie Stroop 1998.





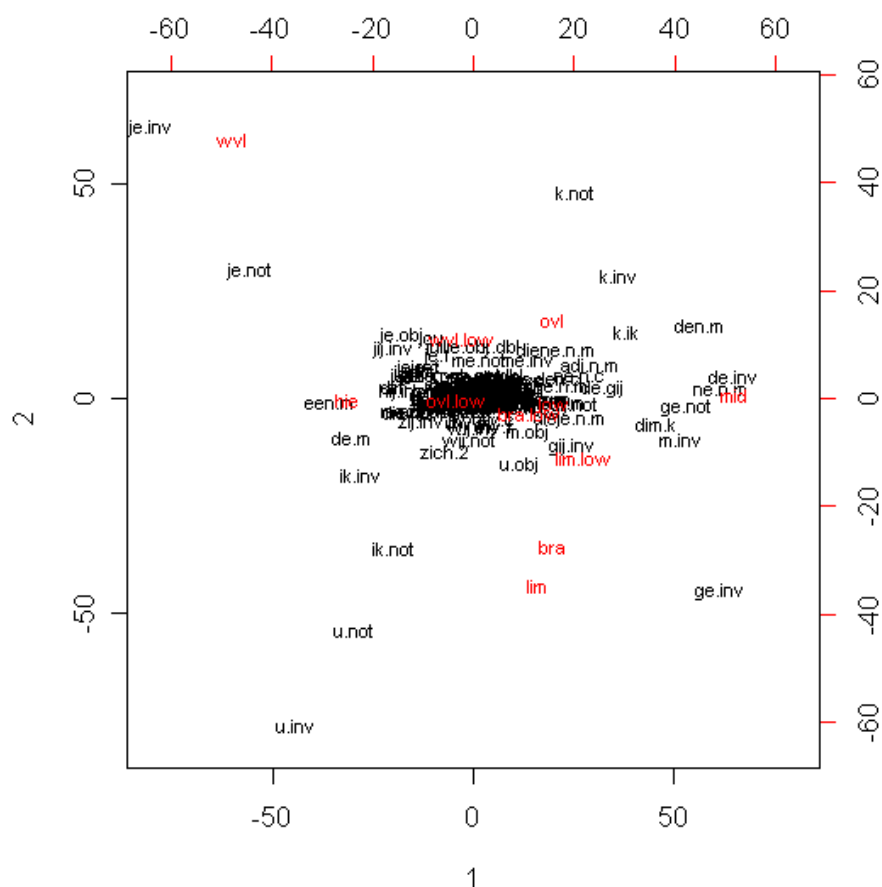
*Figuur 6.35: Interactie Regio – Vrouwen*



*Figuur 6.36: Interactie Regio – Mannen*

### 6.3.8 Regio & Educatieniveau

Met regio en educatieniveau bemerken we dan de paradox dat toenemende standaardisatie niet noodzakelijk leidt tot supraregionale homogeniteit (Figuur 6.37, 6.38, en 6.39). Voor het laagste opleidingsniveau clusteren de vier regio's bijvoorbeeld nog samen rondom het hoofdeffect in het centrum van de plot; die homogeniteit hebben we in 6.3.4 al in verband gebracht met een gebrekkig ontwikkeld taalgevoel. De educatieve middencategorie vertoont dan enigszins meer variatie, waarbij vooral opvalt dat de provincies Limburg en West-Vlaanderen niet delen in de opmars naar het Soapvlaams, die dit opleidingsniveau in het algemeen karakteriseert. Dat blijft enkel voorbehouden voor Brabant (de regio die de bakermat vormt voor het Soapvlaams als bovengewestelijke informele omgangsvariëteit) en Oost-Vlaanderen (het gebied dat zodanig door die variëteit geïnfecteerd is dat het het epitheton 'overgangszone' draagt). De Brabantse inslag van het Soapvlaams komt dus nog maar eens uit de verf. Ten slotte zien we bij het hoogste opleidingsniveau weliswaar een verschuiving naar het Journaalnederlands maar tegelijkertijd ook een toename van de spreiding tussen de regio's – de verdeling correspondeert bijvoorbeeld bijna met die van de hoofdeffecten. Dat koppelen we eveneens aan de in 6.3.2 geschetste notie van standaardtaalbeheersing: door hun scholing kennen hoger opgeleiden namelijk de standaardtaal dan wel beter, maar net daardoor hebben ze tevens een veel verfijnder gevoel voor stijlverschillen.



Figuur 6.37: Interactie Regio – Laag opgeleid

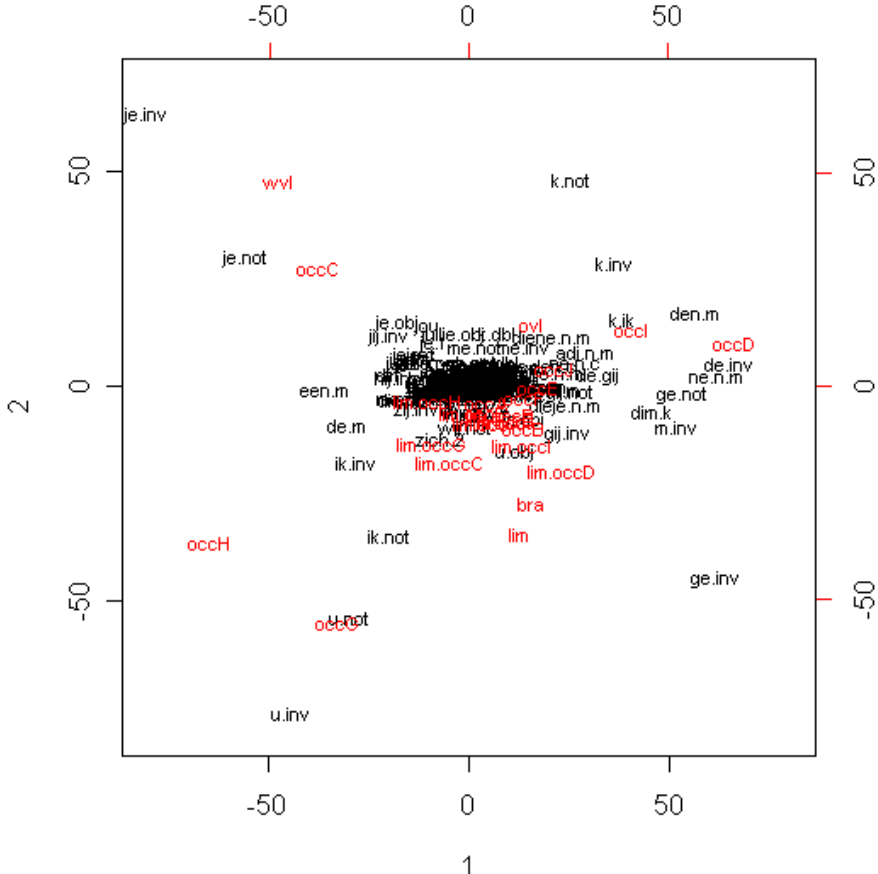


### 6.3.9 Regio & Beroepsniveau

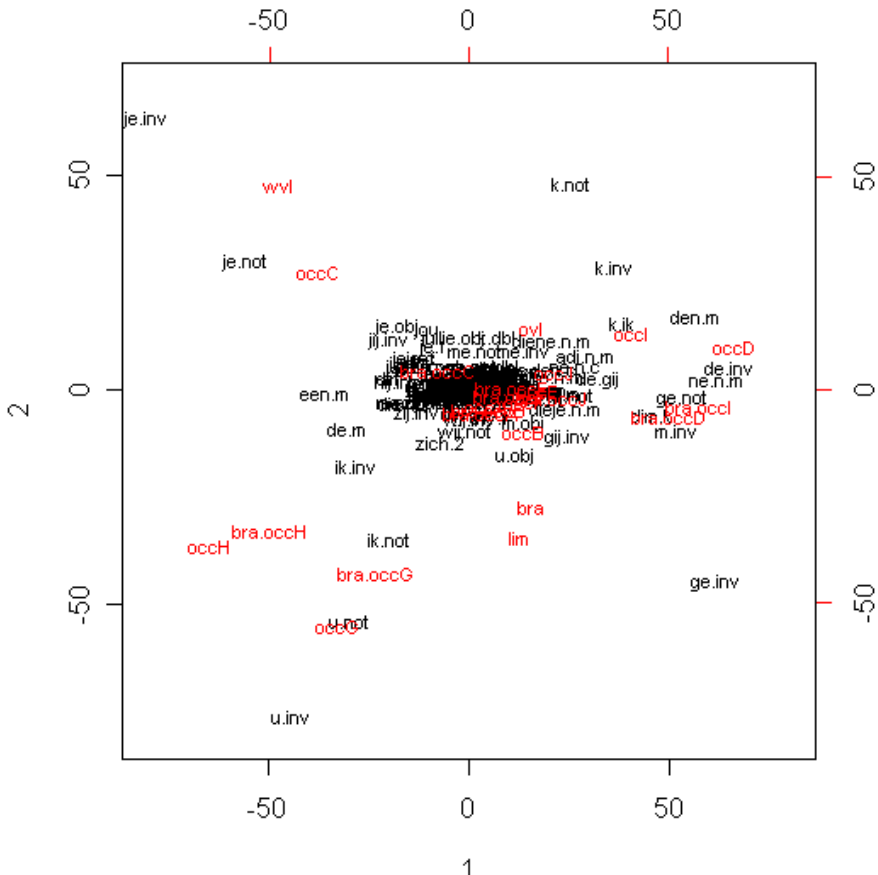
Bij regio en beroepsniveau merken we opnieuw dat de beroepscategorieën die zich neutraal in het centrum van de plot ophouden (occA, occB, occE, occF, en occJ) ook intern homogeen zijn, terwijl de met het Soapvlaams geassocieerde categorieën veel variatie vertonen (Figuur 6.40, 6.41, 6.42, en 6.43). Zo zitten de studenten (occl) en ambtenaren (occD) voor West-Vlaanderen ver boven de centrumcluster, voor Oost-Vlaanderen er net rechtsboven, voor Brabant helemaal aan de rechterkant, en voor Limburg juist onder de centrumcluster. De betekenis daarvan dunkt ons de volgende: Soapvlaams is (ook hier) vooral een Brabants fenomeen, Limburgers spreken meer neutraal Belgisch-Nederlands, terwijl de West-Vlamingen zich dan weer sterk door hun dialect laten leiden, en Oost-Vlaanderen is andermaal de overgangszone tussen Brabant en West-Vlaanderen. Zulks geldt in principe voor beide beroepsniveaus, hoewel er blijkbaar toch ook verschillen zijn. In Oost-Vlaanderen neigen de studenten bijvoorbeeld meer naar het centrum van de plot – *ovl.occl* is tussen *adj.n.m* en *de.gij* te vinden. De Oost-Vlaamse studenten gaan met andere woorden niet mee in het gebruik van het Soapvlaams door hun streekgenoten uit de administratieve sector, maar spreken een neutraler Belgisch-Nederlands.

Moge de bevindingen tot dusver enigszins voorspelbaar zijn op basis van vorige resultaten, ronduit verwonderlijk is echter het feit dat we hetzelfde variatiepatroon tegenkomen bij de academici, de beroepsklasse die doorgaans toch geassocieerd wordt met de J-vormen. Ook voor occC vinden we de Limburgers namelijk net onder de centrumcluster, Brabant er middenin, Oost-Vlaanderen er lichtelijk boven, en West-Vlaanderen nog verder naar boven. Het gebruik van de J-vormen door deze beroepsklasse is dan ook vooral weer een zaak van dialectische achtergrond; Brabantse en Limburgse academici spreken veeleer het meer neutrale Belgisch-Nederlands. Dat betekent evenwel dat de bevolkingsgroep die bij wijze van spreken uit hoofde van hun status ‘het goede voorbeeld zou moeten geven’ dat in de praktijk duidelijk niet overal doet.

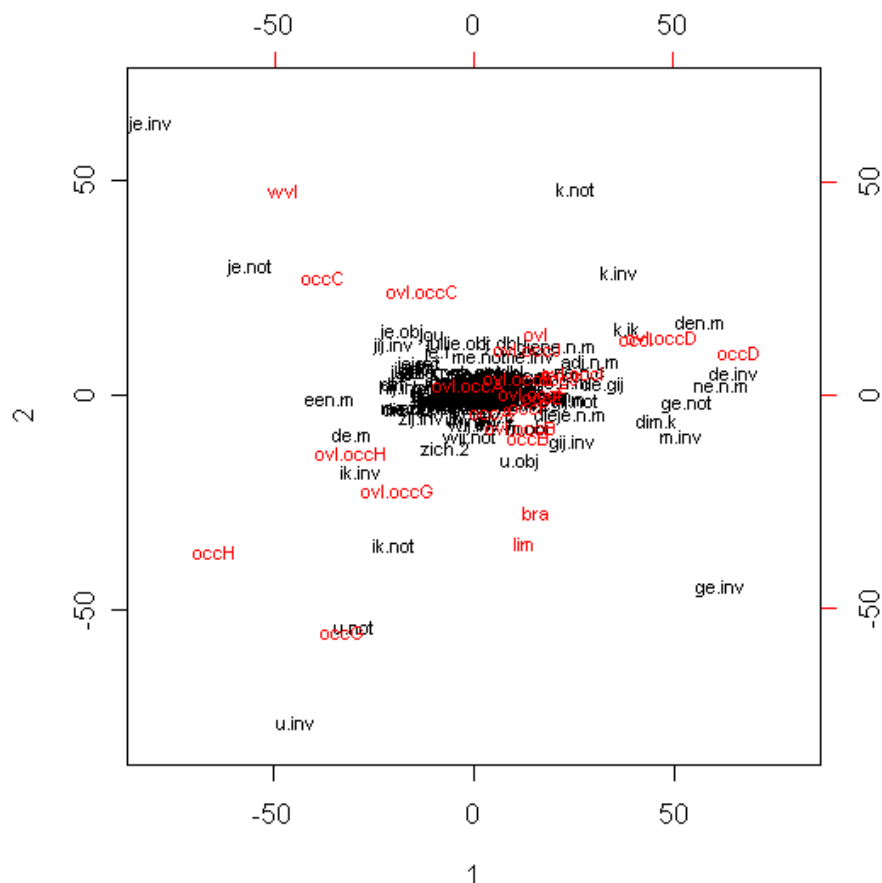
Tot slot vermelden we nog (en daar gaan we het verder dan ook bij laten) dat ook de politici (occG) en mediafiguren (occhH) regionale variatie kennen. Voor beide beroepsniveaus vertonen de regio’s een cline van neutraal naar beleefd taalgebruik (de U-vormen), die loopt van Limburg over de beide Vlaamse provincies naar het Brabantse gebied.



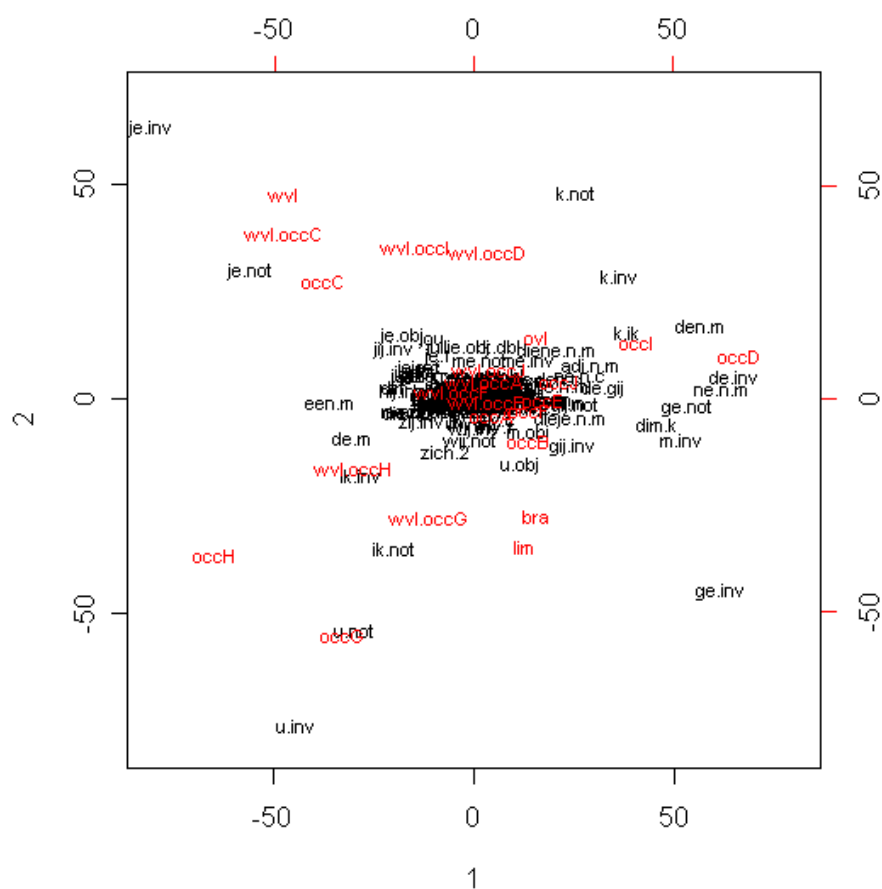
Figuur 6.40: Interactie Limburg – Beroepsniveau



Figuur 6.41: Interactie Brabant – Beroepsniveau



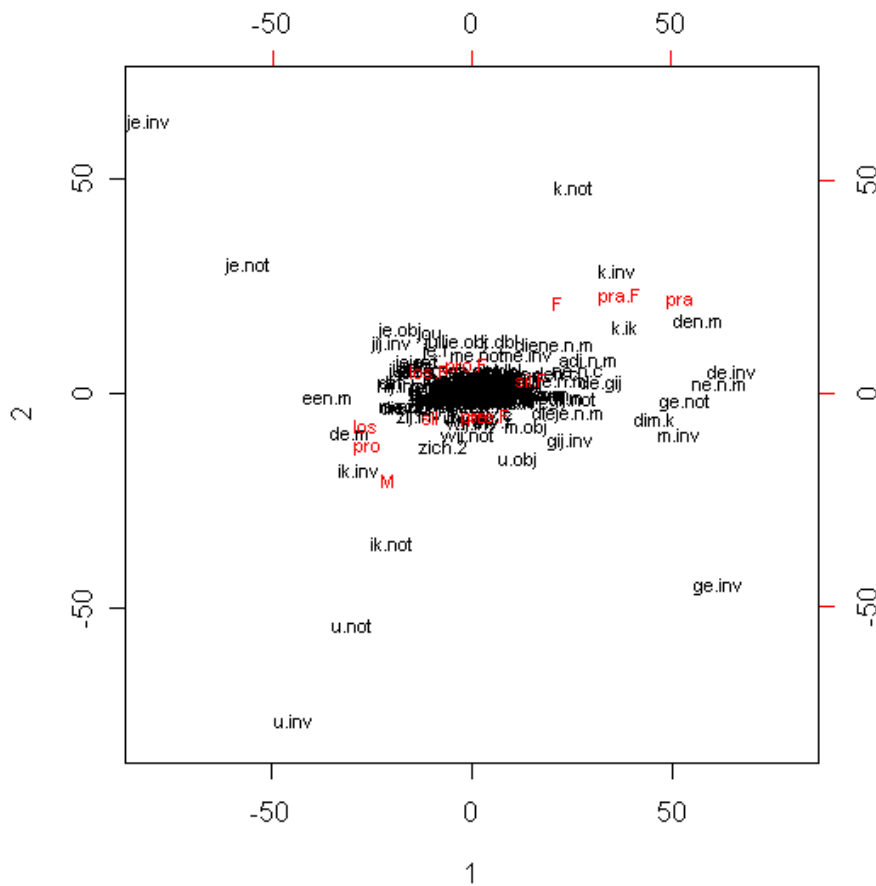
Figuur 6.42: Interactie Oost-Vlaanderen – Beroepsniveau



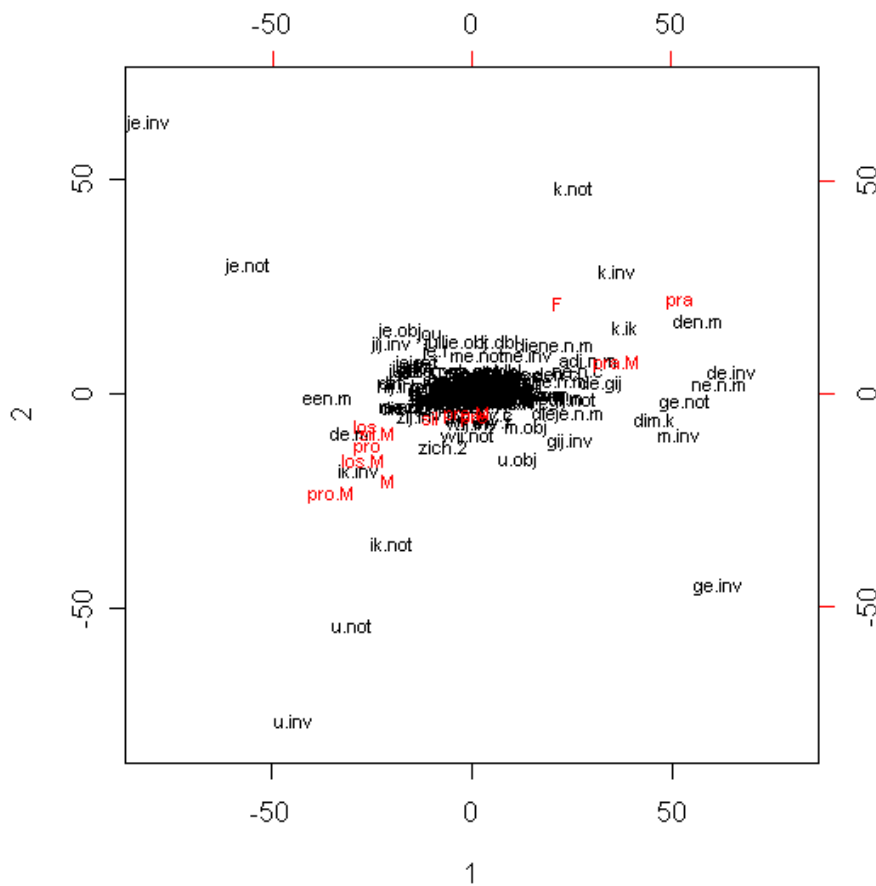
Figuur 6.43: Interactie West-Vlaanderen – Beroepsniveau

### 6.3.10 Generatie & Sekse

De plots voor generatie en sekse (Figuur 6.44 en 6.45) onthullen dat de ‘gendering’ van het Belgisch-Nederlands – het uitkristalliseren van het verschil tussen de sekses – begint met de stille generatie. Voor de rest zijn de interacties evenwel weinig meer dan de gecombineerde effecten van de hoofdfactoren. Zo is er tot de pragmatische generatie weer een algehele opschuiving naar het Journaalnederlands – meer bepaald, de beleefde spreekstijl ervan (U-vormen); vooral de mannen uit de protestgeneratie lijken daarin de spits te willen afbijten. Daarna volgt er voor beide sekses de bekende kentering naar het Soapvlaams. In alle generaties zijn de vrouwelijke sprekers daarbij opnieuw rechtsboven van de mannelijke sprekers te vinden (en in dit geval mag dit letterlijk genomen worden: de vrouwen zitten namelijk allemaal in de bovenhelft van de plot, terwijl de mannen – op die van de pragmatische generatie na – in de benedenhelft).



Figuur 6.44: Interactie Generatie – Vrouwen

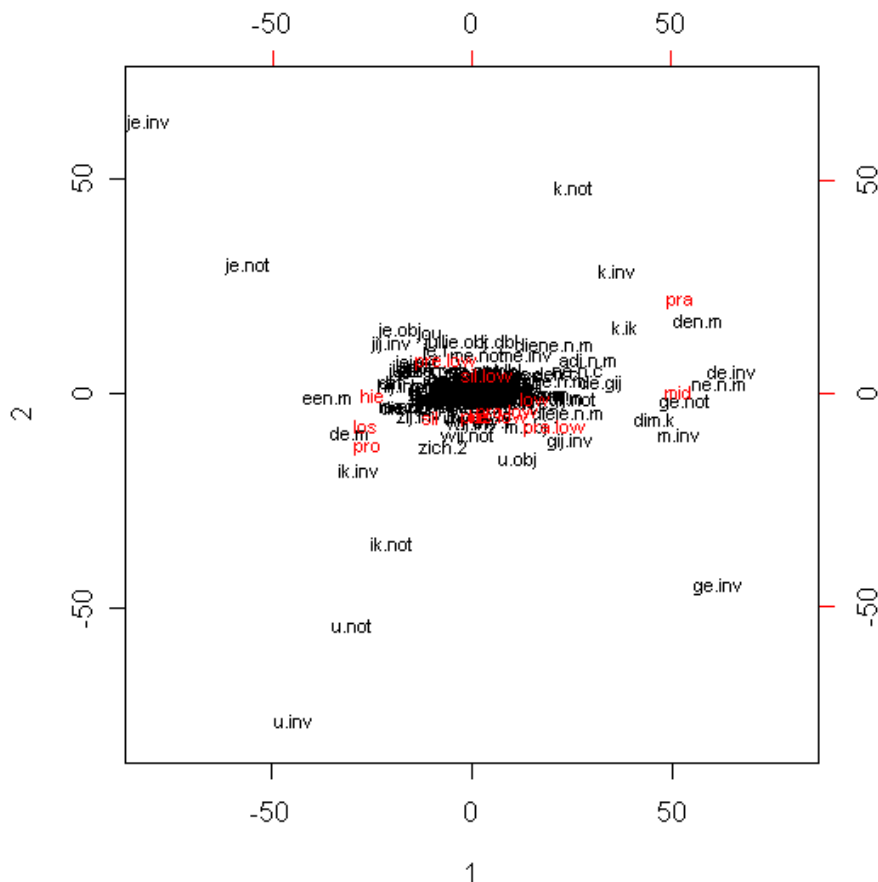


Figuur 6.45: Interactie Generatie – Mannen



### 6.3.11 Generatie & Educatieniveau

Omdat we de verschillen tussen de generaties gecorreleerd hebben aan verschillen in de kwaliteit van het onderwijs, zou de indruk kunnen bestaan dat de interactie tussen generatie en educatieniveau niet veel meer behelst dat wat we tot nog toe al te weten zijn gekomen (Figuur 6.46, 6.47, en 6.48). Zo is er bij de vooroorlogse en de stille generatie weinig variatie tussen de opleidingsniveaus. Alle sprekers clusteren – opnieuw – samen in het centrum van de plot, en voor de laagst opgeleiden zal daar verder nooit enige verandering in komen: zij blijven neutraal Belgisch-Nederlands spreken. Vanaf de protest- en de verloren generatie zet de standaardisering zich in. Die komt er hoofdzakelijk op neer dat de hoogst opgeleiden in deze cohorten het Journaalnederlands beginnen te hanteren. Met de pragmatische generatie gaan de hoogopgeleiden dan uiteindelijk overstag voor het Soapvlaams. Frappant is echter dat de educatieve middencategorie uit de protest- en de verloren generatie daarbij in het neutrale centrum van de plot blijft staan en dus niet de overeenkomstige beweging naar het Soapvlaams kent, die zo eigen is aan deze opleidingsgraad – pas vanaf de pragmatische generatie zal die er komen. De verklaring voor deze ‘inertie’ zou wel eens kunnen liggen in de positie van *los.mid*, dat ten opzichte van *pro.mid* naar rechtsboven verschuift. Dat zou er namelijk op kunnen wijzen dat de kiem van het Soapvlaams gezocht moet worden bij de middenklasse van de verloren generatie.



Figuur 6.46: Interactie Generatie – Laag opgeleid

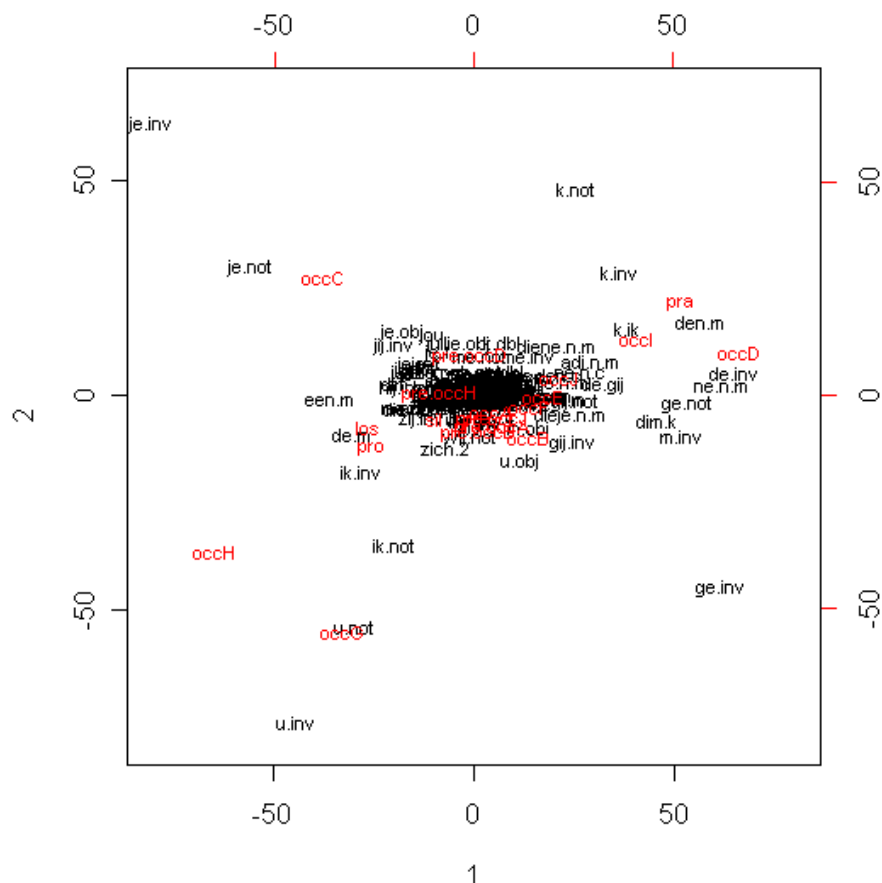


### 6.3.12 Generatie & Beroepsniveau

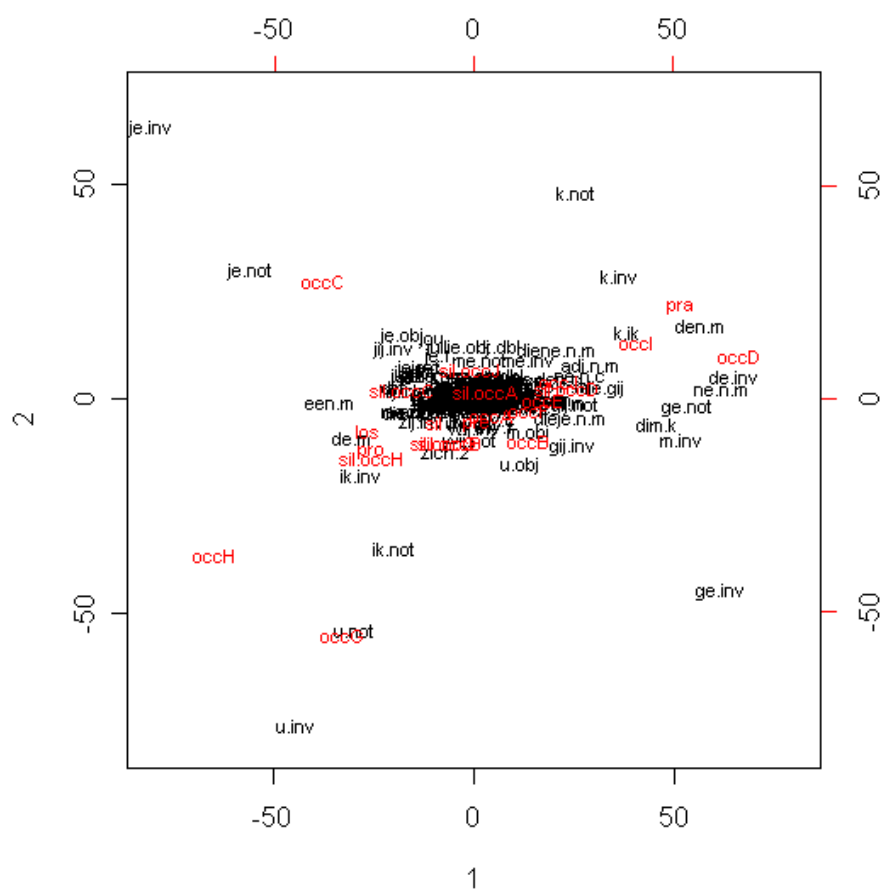
De interactie van generatie met beroepsniveau biedt in elk geval evidentie voor de hypothese dat het Soapvlaams al in de middenklasse bij de verloren generatie embryonaal aanwezig is (Figuur 6.49, 6.50, 6.51, 6.52, en 6.53). Dat zien we met name aan het gedrag van één van de uitgesproken vertegenwoordigers van het Soapvlaams: *occD*, de ambtenaren (over de andere, *occl* oftewel de studenten, valt niks zinnigs te zeggen omdat er in deze bevolkingsgroep enkel sprekers uit de pragmatische generatie zijn; voor *occl* is er met andere woorden geen generationele variatie). Als met de protestgeneratie de talige differentiatie namelijk intreedt, dan blijft de interactie *pro.occD* nog in de centrumcluster staan (ter hoogte van de hoofdeffecten van *occE* en *occF*). Bij de verloren generatie neigt *los.occD* echter al meer naar rechts, in de richting van het Soapvlaams – om in de pragmatische generatie ten slotte nog verder op te schuiven.

Daarmee is de verloren generatie zonder meer een schizofrene generatie. Ook voor de drie beroepscategorieën die namelijk met het Journaalnederlands geassocieerd zijn (*occC*, *occG*, en *occH*), bevestigt de verloren generatie de in de lijn van de verwachting liggende ontwikkelingen. Bij de *academici* gaat ze bijvoorbeeld door op de aan de gang zijnde verschuiving naar de J-vormen linksboven – dat de pragmatische generatie vervolgens naar het Soapvlaams afbuigt spreekt in dat verband overigens boekdelen. Bij de politici en mediafiguren respecteert ze dan weer de door de mediafiguren uit de stille generatie geïnstigeerde evolutie (zie Figuur 6.50) naar het gebruik van de U-vormen.

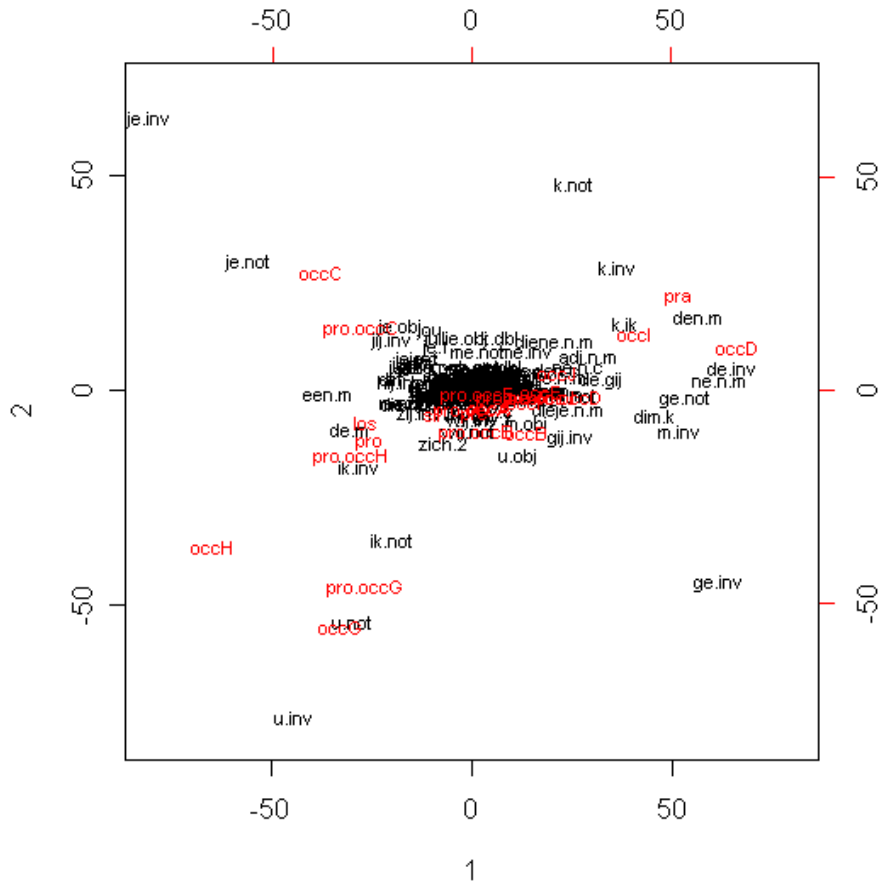
Voor de rest corroboreren de interacties tussen generatie en beroepsniveau alleen maar eerder vastgestelde effecten. Zo kennen de variëteiten waarvan het hoofdeffect al in het centrum van de plot staat ook in de tijd weinig tot geen verandering: sprekers uit categorie *occA*, *occB*, *occE*, *occF*, en *occJ* blijven over de verschillende generaties heen een neutraal Belgisch-Nederlands spreken. Tekenend is wel dat ze vanaf de pragmatische generatie daarbij het gezelschap krijgen van mediafiguren en politici.



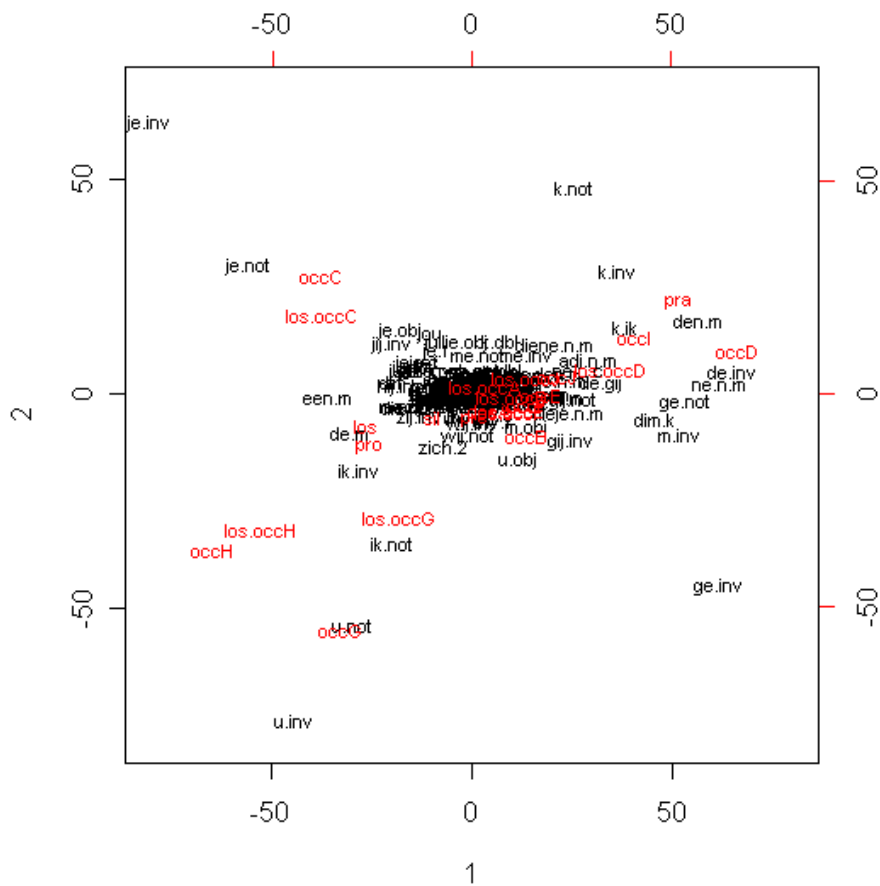
Figuur 6.49: Interactie Vooroorlogse generatie – Beroepsniveau



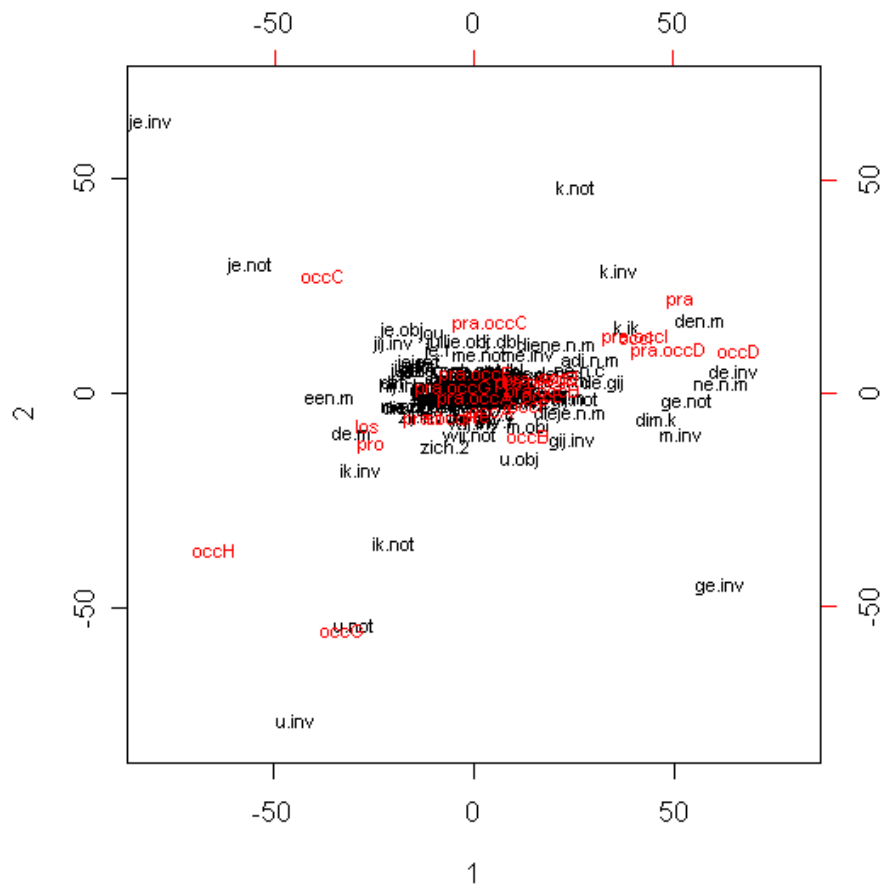
Figuur 6.50: Interactie Stille generatie – Beroepsniveau



Figuur 6.51: Interactie Protestgeneratie – Beroepsniveau



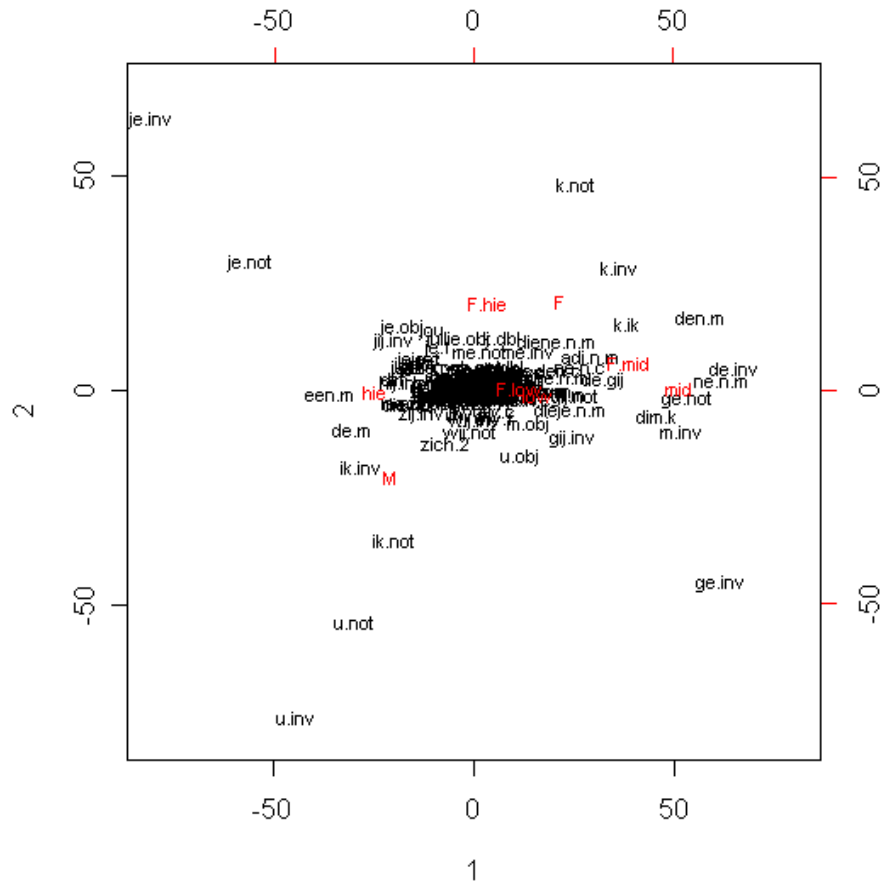
Figuur 6.52: Interactie Verloren generatie – Beroepsniveau



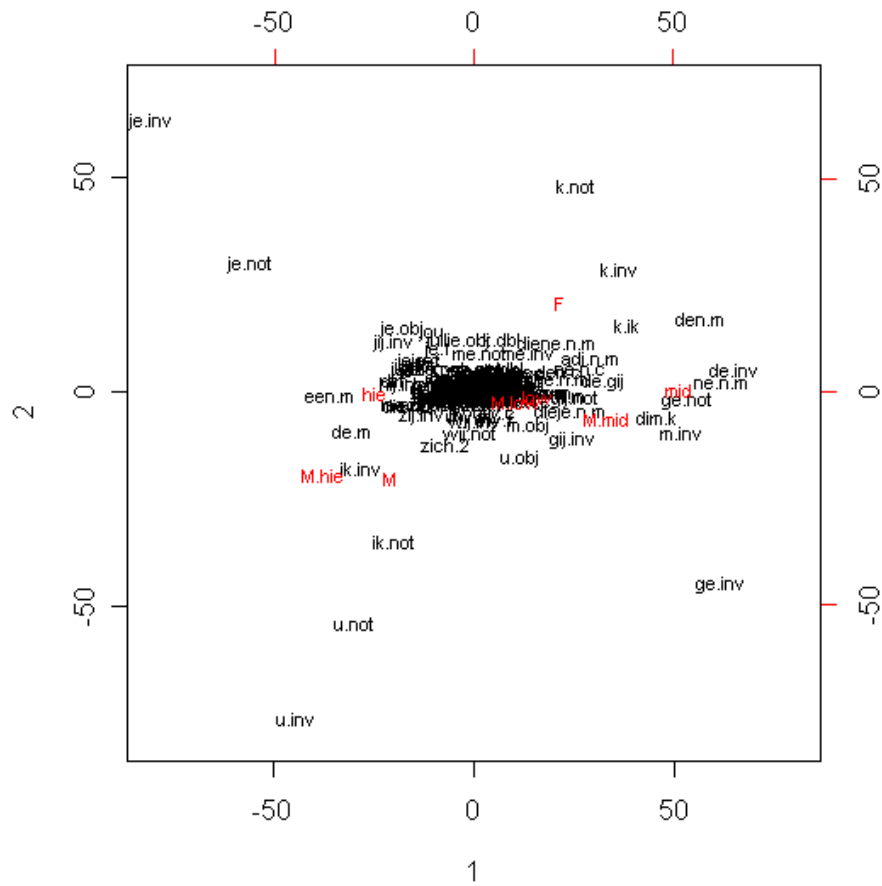
Figuur 6.53: Interactie Pragmatische generatie – Beroepsniveau

### 6.3.13 Sekse & Educatieniveau

Bij de interactie tussen sekse en educatieniveau vallen vooral de spreidingseffecten op (Figuur 6.54 en 6.55). Voor de laagst opgeleiden, die neutraal Belgisch-Nederlands spreken, is er geen noemenswaardig verschil tussen de sekses. Bij de middencategorie treedt er dan met de verschuiving naar het Soapvlaams ook enige variatie tussen mannen en vrouwen op. De sekseverschillen zijn evenwel het scherpst afgelijnd bij de hoog opgeleide sprekers, wat nog maar eens aantoont hoezeer betere scholing kan leiden tot meer differentiatie. De exacte posities van de interacties zijn daarbij in het bijzonder opmerkelijk. De hoog opgeleide vrouwen staan namelijk ongeveer in het midden van de J- en de G-vormen, en op hun beurt bevinden de mannen zich vrijwel precies tussen de J- en de U-vormen in. Het hoofdeffect bij de sekses – vrouwen hanteren typisch een vertrouwelijke spreekstijl, waarbinnen ze dan alterneren tussen Journaalnederlandse en Soapvlaamse varianten; mannen switchen daarentegen tussen vertrouwelijke en beleefd taalgebruik, maar houden zich daarbij dan weer meer aan het Journaalnederlands – komt met andere woorden bij de hoogst opgeleiden treffend tot uiting.



Figuur 6.54: Interactie Vrouwen – Educatieniveau



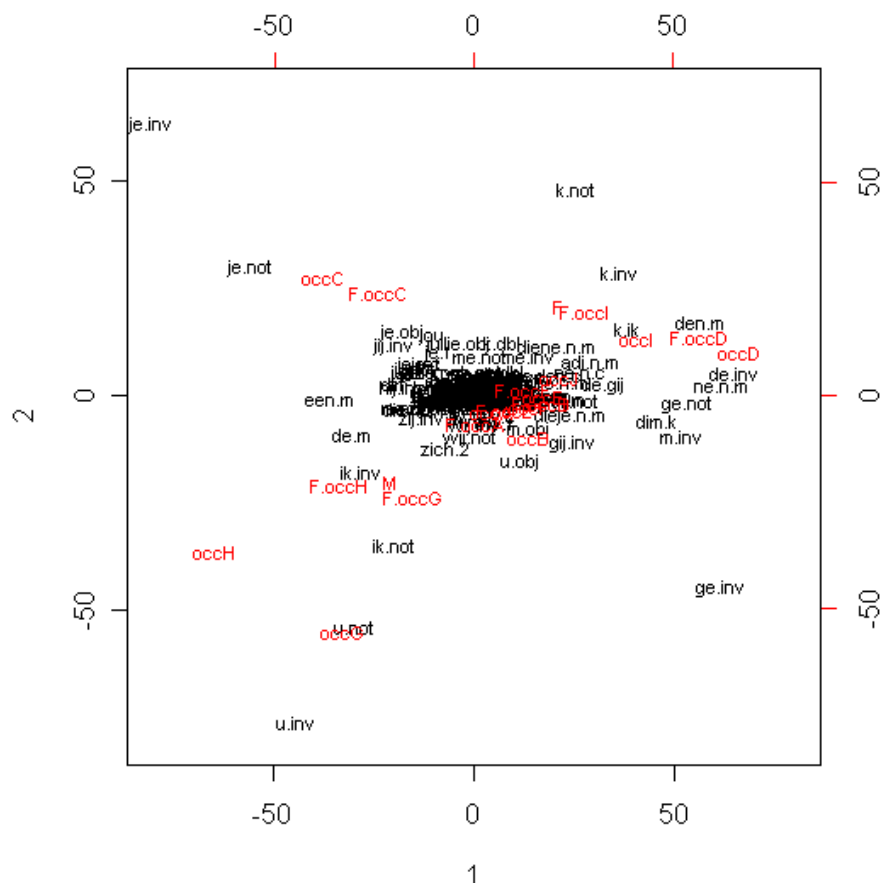
Figuur 6.55: Interactie Mannen – Educatieniveau



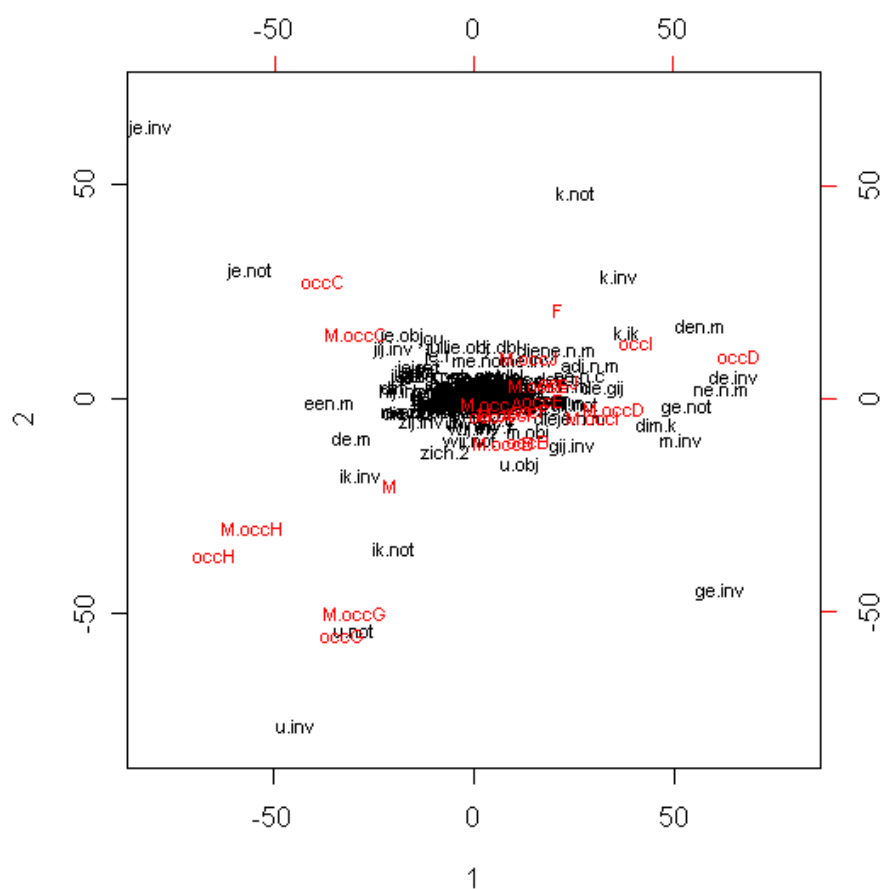
### 6.3.14 Sekse & Beroepsniveau

Voor de interactie tussen sekse en beroepsniveau zijn het nogmaals de categorieën in de Soapvlaamse helft van het Belgisch-Nederlands die onze aandacht trekken (Figuur 6.56 en 6.57). Zo vertonen de mannen uit beroepsniveaus *occl* en *occD* niet alleen een behoorlijke duik naar beneden in vergelijking met de vrouwen, bij de vrouwen is er tussen beide beroepsklassen bovendien ook onderling een verschil in welke factor het sterkst het gedrag van de interactie bepaalt: voor *F.occl* is dat de sekse (F), terwijl voor *F.occD* het beroepsniveau (*occD*). Anders gezegd, zowel de studentes als de ambtenaressen zijn notoire representanten van het Soapvlaams, maar de studentes spreken meer zoals vrouwen in het algemeen, terwijl de ambtenaressen eerder zoals een modale ambtenaar. De interactie *F.occl* bevestigt overigens opnieuw de voortrekkersrol van jonge, hoog opgeleide (want: student) vrouwen – vergelijk: 6.3.7.

Verder incorporeren de interacties opnieuw de gecombineerde effecten van de hoofdfactoren: de globale spreiding is die van het beroepsniveau, waarbij op de koop toe telkens de diagonaal van vrouwelijk rechtsboven naar mannelijk linksonder gerespecteerd wordt. Zo zijn de academici (*occC*) voor beide sekse weer met de J-vormen geassocieerd, de politici (*occG*) en mediafiguren (*occH*) met de U-vormen, en tot slot de hele reeks van managers (*occA*), hoger opgeleiden (*occB*), ongeschoolden (*occE*), zelfstandigen (*occF*), en werklozen (*occJ*) met het neutrale Belgisch-Nederlands – hoewel bij *occA*, *occE*, en *occJ* de verhouding tussen de sekse omgekeerd is, en de mannen dus boven de vrouwen staan; een verklaring hebben we daar voorlopig niet voor, maar gezien de geringe grootte van de verschillen is dat misschien ook niet nodig.



Figuur 6.56: Interactie Vrouwen – Beroepsniveau



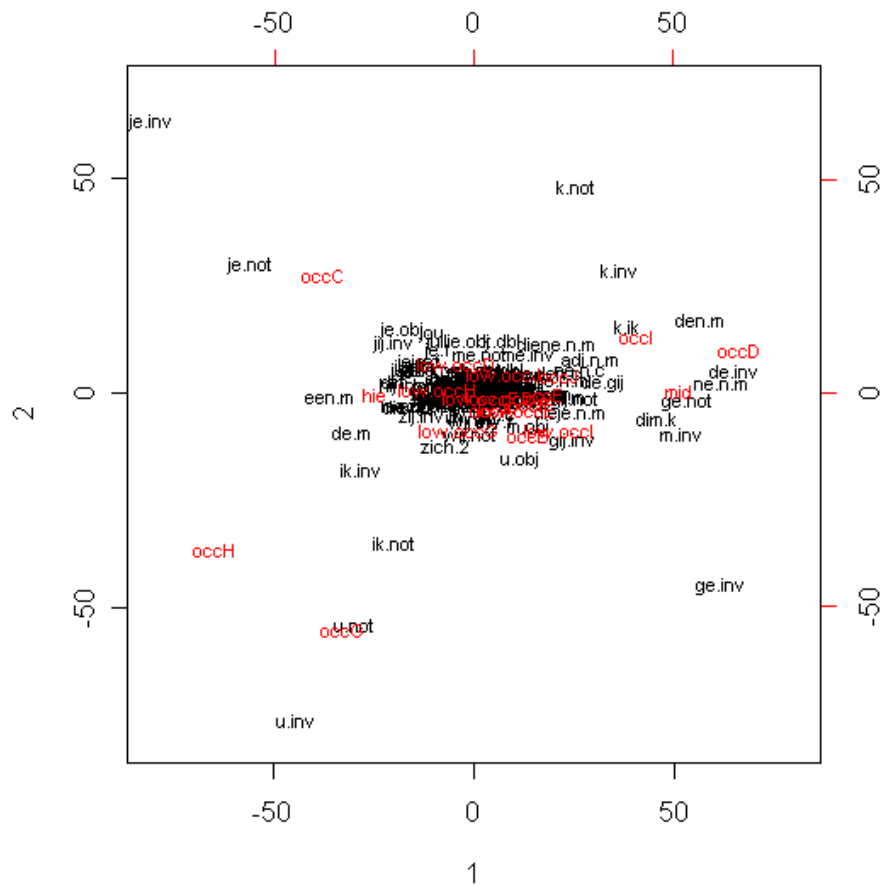
Figuur 6.57: Interactie Mannen – Beroepsniveau

### 6.3.15 Educatieniveau & Beroepsniveau

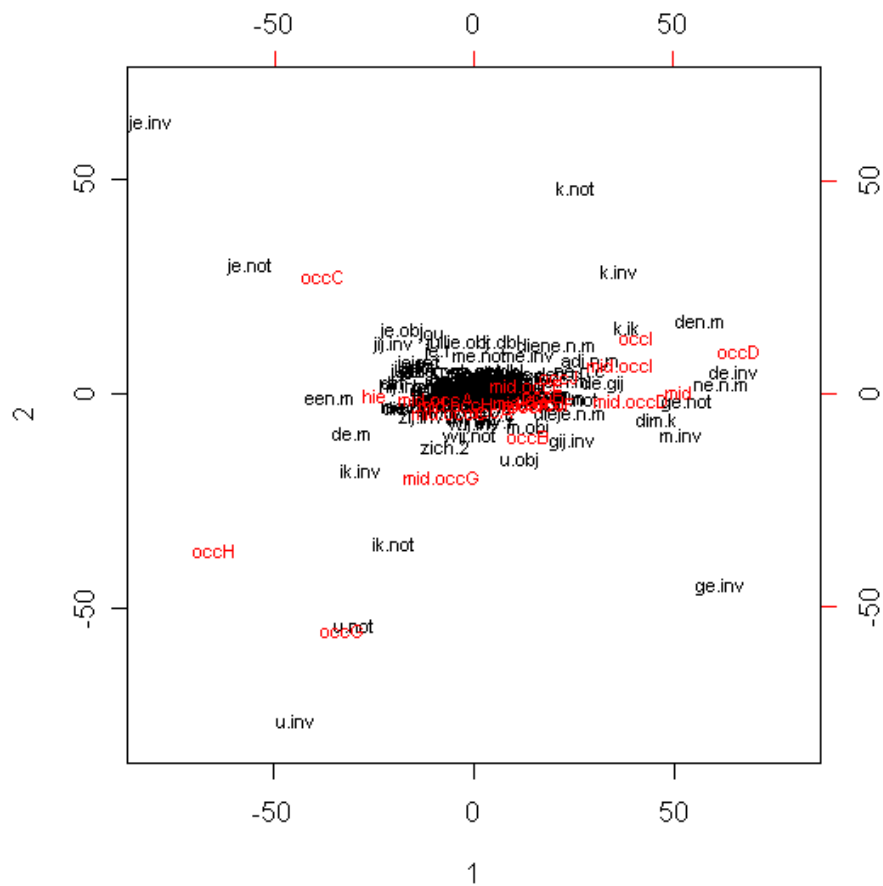
Bij de interactie tussen educatie- en beroepsniveau valt tot slot opnieuw het spreidingseffect op: naarmate de educatiegraad stijgt, neemt de variatiebreedte tussen de beroepsklassen eveneens toe – en opnieuw fungeert het centrum van de plot daarbij als startpunt voor de ontwikkeling (Figuur 6.58, 6.59, en 6.60). Vooral het verschil van de hoogst opgeleiden met de middencategorie springt daarbij in het oog: waar het verschil met de laagst opgeleiden er voor de educatieve middencategorie hoogstens in bestaat dat de studenten en ambtenaren hun bekende ruk naar het Soapvlaams aanvatten en de politici naar de U-vormen, zien we bij de hoogst opgeleiden de interacties de extreme posities naast hun respectieve hoofdeffecten innemen. Enkel de studenten vormen hierop een uitzondering, doordat zij naar de J-vormen afbuigen. Daarmee duikt er ook hier een verschil op tussen de twee beroepsklassen die het Soapvlaams representeren: terwijl de ambtenaren een gestage toevlucht tot het Soapvlaams nemen (over de verschillende educatieniveaus), zoeken de studenten vanaf de hoogste opleidingsgraad toch meer het Journaalnederlands op.

Het kloof tussen de hoog opgeleiden en de educatieve middencategorie is het grootst bij de academici en de mediafiguren: voor beide beroepsklassen staan de interacties van het middenniveau zelfs nog samen met die van de laag opgeleiden in het neutrale centrum van de plot (in tegenstelling tot de politici, bijvoorbeeld, die veeleer een geleidelijke overgang van laag opgeleid centrum naar hoog opgeleid hoofdeffect te kennen geven). Ons lijkt dit gedrag vooral een drang tot distinctie te beduiden: in de vernoemde beroepsklassen maken hoger opgeleiden namelijk in Vlaanderen de culturele en talige elite uit – mediafiguren en onderwijsgevenden belichamen net de Journaalnederlandse norm (vandaar dat de interacties dan ook nagenoeg met de hoofdeffecten samenvallen). Die status legitimeren zij door zich scherp te onderscheiden van de neutraal Belgisch-Nederlands sprekende lagere regionen in de maatschappij. In het contrast met het taalgebruik uit de lagere klassen beklemtoont de linguïstische elite het maatschappelijke verschil dat er tussen haar en de niet-elite bestaat, en verstevigt daarmee haar cultureel hiërarchische positie.

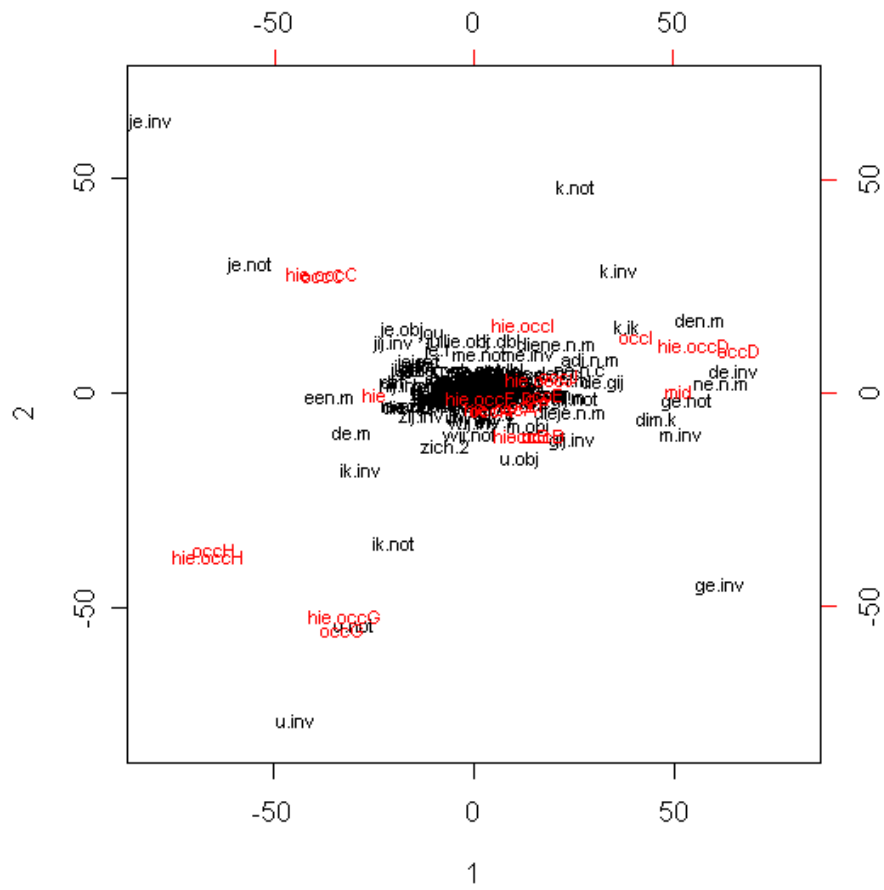
De klassen die ten slotte geen last van enig distinctiestreven hebben zijn de vijf beroepsniveaus die het neutrale Belgisch-Nederlands hanteren: occA, occB, occE, occF, en occJ. Zoals gewoonlijk voor categorieën die zich in het centrum van de plot ophouden, vertonen ze geen enkele interne variatie met betrekking tot de educatieniveaus – waarbij we echter wel moeten herhalen (vergelijk: 5.7) dat beroepsniveau occA van de managers geen lager opgeleiden bevat, occB van de hoger opgeleiden (logischerwijs) enkel uit hoog opgeleiden bestaat, en occE van de ongeschoolden omgekeerd geen hoger opgeleiden telt.



Figuur 6.58: Interactie Laag opgeleid – Beroepsniveau



Figuur 6.59: Interactie Mid-opgeleid – Beroepsniveau



*Figuur 6.60: Interactie Hoog opgeleid – Beroepsniveau*



## Hoofdstuk 7 | Besluit

In dit proefschrift hebben we de status en positie van de Vlaamse tussentaal, het zogenaamde Soapvlaams, nader willen bepalen. Op basis van de probleemstelling en de hypothesen die daaruit volgden zoals beschreven in hoofdstuk 1 hebben we daartoe voor een corpusgebaseerde aanpak gekozen; het corpus waarvan we gebruik gemaakt hebben is het Corpus Gesproken Nederlands (CGN). Na een korte bespreking ervan in hoofdstuk 2 hebben we het CGN in hoofdstuk 3 aan een grondige verkenning onderworpen, om er zo een totaal van 62 relevante taalvariabelen uit te distilleren. Vervolgens hebben we in hoofdstuk 4 de kwantitatieve methodologie uit de doeken gedaan waarmee we onze variatiestudie zouden uitvoeren. In hoofdstuk 5 hebben we daarbij elke taalexterne factor uit het CGN apart behandeld in de zogenaamde bivariate analyses. In hoofdstuk 6 hebben we ten slotte alle factoren samen opgenomen in één globale multivariate analyse.

Daarmee kunnen we nu overgaan tot enkele concluderende beschouwingen. In 7.1 worden alle resultaten nog eens samengevat; in 7.2 volgt dan een beoordeling ervan.

### 7.1 Samenvatting van de resultaten

Het belangrijkste resultaat (constructiefouten zoals het feit dat de rol van interviewer in component b maar door één en dezelfde persoon ingevuld wordt laten we natuurlijk buiten beschouwing) is misschien nog wel een *negatief* effect (zie 5.1): het Journaalnederlands en het Soapvlaams vormen niet twee aparte clusters, maar vertonen daarentegen een naadloze overgang van en naar elkaar. Hoogstens is er sprake van enkele uitschieterende sjibboletvormen, en voor het Soapvlaams zijn dat het gereduceerde 'k (dat daarmee meteen ook een 'uitzondering' is: een Journaalnederlandse variant die zich als een Soapvlaamse gedraagt – daarvan zijn er nog wel meer, maar 'k is toch de meest prominente), het enclitische (*e*)kik, *den* en *ne(n)* bij masculiene nomina, de aanspreekvormen *-de* en *ge*, de diminutief op *-ke*, en tot slot de subject-'*m*. Beide lidwoorden illustreren daarbij een ander variationeel principe: met name, dat de spreiding het grootst is onder de determinatoren bij masculina. Ten slotte hebben we ook nog een aantal fonetische reductieverschijnselen geconstateerd, maar het effect ervan was uitgesproken noch transparant. Hoe dan ook, het continuüm tussen Soapvlaams en Journaalnederlands

vatten we zonder meer op als een weerlegging van de these als zou de Vlaamse taalsituatie diglossisch zijn – enige reserves ten aanzien van onze nogal operationalistische invulling van dat begrip niet te na gesproken. Omgekeerd betekent één en ander evenwel dat het geen zin heeft om het Journaalnederlands en het Soapvlaams nog langer als twee verschillende talen tegenover elkaar te stellen; in Vlaanderen wordt één, enkelvoudige taal gesproken, (Algemeen) Belgisch-Nederlands, *waarbinnen* Journaalnederlandse varianten alterneren met Soapvlaamse varianten.

Voorts valt het onderscheid Soapvlaams-Journaalnederlands niet samen met de grammaticale distinctie vertrouwelijk-beleefd: de tegenstelling tussen de J- en de U-vormen verloopt namelijk volgens de verticale as van de biplot, waar de differentiatie tussen Soapvlaams en Journaalnederlands nog langs de horizontale as loopt. Daarmee wordt het al dan niet optreden van het Soapvlaams niet beregeld door de parameter vertrouwelijk versus beleefd: de dimensie die dat wel doet staat in meerdere betekenissen van het woord 'loodrecht' daarop.

De J-varianten zijn daarbij in elk opzicht schoolse vormen: ze zijn geassocieerd met de beroepsklasse van de academici (occC), en worden door leraren gehanteerd zowel voor de klas (component h) als in een privé-interview (component b). De U-vormen zijn daarentegen meer debataire varianten: ze worden voornamelijk in de discussiecomponenten f en g gebruikt door politici (occG) en mediafiguren (occH) – als die laatsten tenminste ook daadwerkelijk op radio en televisie komen; is dat niet zo, dan hanteren deze sprekers veeleer neutraal Belgisch-Nederlands (zie 6.3.5). Het Journaalnederlands (bij uitstek vertegenwoordigd in componenten i, j, k en l) en de voorgelezen tekst (component o) zitten precies tussen beide cluster in, waarbij de laatste verder van het Soapvlaams af staat dan de eerste: in de media duikt er met andere woorden net iets meer Soapvlaams op dan in geschreven materiaal. Dat betekent dat het Journaalnederlands als spreekstijl in hoofdzaak een mediataal, en misschien eerder nog een schrijftaal is.

Soapvlaams treffen we dan weer aan in meer informele gespreksituaties zoals face-to-face conversaties (component a) en telefoondialogen (component c en d), en bij studenten (occl) en ambtenaren (occD). Typisch voor deze variëteiten is dat het hier telkens een laag betreft die zich juist onder het hoogste niveau bevindt: het register is enerzijds te hoog voor zuiver dialect maar anderzijds laag genoeg voor het gebruik van omgangstalige varianten, terwijl de sprekers ofwel het openbare leven nog niet betreden hebben ofwel tot een maatschappelijke klasse behoren die net niet tot de elite gerekend kan worden. Soapvlaams is dan ook verder geassocieerd met sprekers uit de educatieve middencategorie evenals de jongste (dit wil zeggen: de pragmatische) generatie – hoewel er wat die laatste betreft meer aan de hand is; zie verderop. Stilistisch is het onderscheid tussen Journaalnederlands en Soapvlaams



daarmee een continuüm op een *spontaniteitsdimensie* (horizontale as); verticale as van de J- naar de U-vormen constitueert dan weer een *(in-)formaliteitsdimensie*.

Alle andere variëteiten clusteren samen in het centrum van de plot. De sprekers in deze categorieën spreken neutraal Belgisch-Nederlands – ‘tussentaal’ in de eigenlijke zin van het woord – hoewel de meesten onder hen een lichte voorkeur voor het Soapvlaams vertonen. Het gaat dan ook om een bonte keur uit alle lagen van de samenleving. Meer bepaald treffen we hier managers (occA) en hoger opgeleiden (occB) aan, maar ook ongeschoolden (occE), zelfstandigen (occF), en werklozen (occJ), waarbij de positie van de eerste twee beroepsgroepen aanleiding heeft gegeven tot een opdeling van de Vlaamse maatschappelijke elite in een *economische* tegenover een *intellectuele* (zie 5.7): de eerste neigt daarbij dus eerder naar het Soapvlaams; de tweede houdt (nog) aan het Journaalnederlands vast. Ten slotte vinden we hier ook nog de (educatief) laagst opgeleiden evenals de twee oudste generaties – de vooroorlogse en de stille generatie.

Met die drie laatste cohorten representeert het centrum van de plot het natuurlijke startpunt voor het standaardisatietraject zoals Vlaanderen dat gedurende de twintigste eeuw heeft doorlopen: naarmate de generaties elkaar opvolgden is namelijk ook het onderwijspeil – alsmede het niveau van welvaart – van de gemiddelde Vlaming gestegen (zie 5.3). Daarbij hebben we gezien dat de standaardisering in de volgende twee generaties – de protest- en de verloren generatie – vooralsnog de vorm aannam van een beweging naar het Journaalnederlands. Vanaf de pragmatische generatie komt er dan echter een ‘dialectische omslag’ naar het Soapvlaams, een verschuiving die zich niet zomaar tot een ‘age grading’-fenomeen laat reduceren (zie 5.3): hetzelfde effect kunnen we namelijk eveneens voor het decennium 1970-1979 vaststellen, waarvan de gemiddelde spreker ten tijde van de compilatie van het CGN (dit is: rond de millenniumwissel) toch 25 jaar oud is en daarmee al lang in de maatschappelijke orde opgenomen. Dat leidt ons er dan ook toe om de waargenomen taalontwikkeling zonder meer te interpreteren als een – aan de gang zijnde – taalverandering.

Als taalcentrum van die verandering fungeert zonder twijfel de Brabantse regio (het samenstel van de provincies Antwerpen en Vlaams-Brabant). Dat leiden we niet alleen uit het feit dat de Soapvlaamse varianten dialectologisch gezien doorgaans een Brabantse achtergrond hebben, ook enkele interacties wijzen in die richting: bij leeftijd zijn bijvoorbeeld de Brabanders van de pragmatische generatie (die dus de generatie van het Soapvlaams genoemd mag worden) de meest extreme uitschieter in de hele analyse (6.3.6); bij sekse is dan weer de interactie tussen Brabant en de vrouwelijke sprekers sterker dan het gecombineerde effect van beide hoofdfactoren (6.3.7). Dat laatste sluit trouwens aan op een in de sociolinguïstiek bekend verschijnsel: met name, dat vrouwen over het algemeen meer het vertrouwelijke

jargon hanteren dan mannen maar daarbij naargelang de situatie gevoeliger zijn voor prestigevormen, waarmee zij vaak de instigator van taalveranderingen in het informele register zijn (de vergelijking met het Poldernederlands, zoals we die in 6.3.7 gemaakt hebben, dient zich daarbij wel aan). Als het Vlaamse taallandschap dan ook inderdaad een gendereffect laat zien waarbij mannen vooral Journaalnederlands spreken en daarbij alterneren tussen de vertrouwelijke en de beleefde varianten, terwijl vrouwen eerder binnen een informele spreekstijl blijven en daarin switchen tussen de Journaalnederlandse en de Soapvlaamse vormen, dan lijkt Vlaanderen daarin alle andere taalgemeenschappen te weerspiegelen.

Dat Brabant het taalcentrum van de versoapvlaamsing is betekent hoegenaamd nog niet dat de overige Vlaamse provincies nu allemaal in gelijke mate voor het Soapvlaams overstag gaan. Niet alleen merken we dat aan het feit dat sprekers uit de perifere gebieden van Vlaanderen die in Brabant komen wonen daarbij nog sterk hun oorspronkelijke tongval behouden, met name de West- en Oost-Vlaamse sprekers verdelen de plot in een Vlaamse bovenkant en een 'niet-Vlaamse' onderkant. Voorts komt de dialectische achtergrond van de verschillende regio's sterk tot uiting – vooral voor West-Vlaanderen dienen we er rekening mee te houden dat de gereduceerde *je*-vormen historisch al tot sommige West-Vlaamse dialecten behoren (vergelijk de stelling van Vandekerckhove in 5.2); voor Limburg heeft het CGN dan weer weinig dialectkenmerken geregistreerd zodat deze provincie consistent neutraal Belgisch-Nederlands lijkt te spreken. Dat alles leidt soms tot enkele opmerkelijke interacties. Zo valt er niet alleen in de lagere en informele regionen van het spreektaalspectrum (componenten a,c en d, en verder de studenten en de ambtenaren) zoals verwacht nog heel wat variatie te bespeuren. Ook onder de academici, een beroepsklasse die toch wel bovenaan de sociale ladder staat, is er behoorlijk veel regionale heterogeniteit (zie 6.3.9).

Daarmee zijn we tot slot beland bij wat wel het meest pregnante effect uit de hele analyse mag heten: met name, het feit dat de bevolkingsgroepen die op grond van hun maatschappelijke positie de standaardisering (kunnen) inzetten daarbij niet zozeer opschuiven naar een welbepaalde spreekstijl (in casu: het Journaalnederlands), maar veeleer hun taalgebruik *differentiëren* in verschillende variëteiten. Dat geldt zowel voor de leeftijdscategorieën als voor het educatieniveau – wat nogmaals niet zo verwonderlijk is, aangezien het peil van ontwikkeling in Vlaanderen na de Tweede Wereldoorlog ook in het algemeen doorheen de tijd gestegen is. Bij leeftijd geven de sprekers geboren in de jaren 40, 50 en 60 – de protest- en de verloren generatie – bijvoorbeeld een steeds toenemende diversificatie te kennen in hun interacties met register, regio, sekse, en beroepsklasse (waarna er met de pragmatische generatie zoals gezegd een kentering naar het Soapvlaams volgt). Bij educatiegraad vallen interacties voor de hoogst opgeleiden met betrekking tot deze vier factoren zelfs nagenoeg samen met de respectieve hoofdeffecten – wat voor bepaalde bevolkingsklassen dan ook

aanleiding geeft tot een interpretatie in termen van distinctiedrift (zie 6.3.15). In beide gevallen ligt het vertrekpunt van de ontwikkeling daarbij in het centrum van de plot, de plaats waar de Journaalnederlandse en de Soapvlaamse varianten met andere woorden elkaar 'raken'. Het continuüm tussen Journaalnederlands en Soapvlaams komt daarmee in zekere zin evenwel ten einde: het wordt namelijk *uitgekristalliseerd* in een veelheid aan variëteiten.

Onze bevindingen laten zich daarom ook als volgt formuleren: terwijl het *taalgebruik* (zoals vertegenwoordigd in de varianten – de zwarte punten in de plot) als door een *centripetale* kracht samen wordt gehouden, kennen de *taalgebruikers* (de variëteiten, oftewel de rode punten) daarentegen een *centrifugale* beweging. Een paradox zoals deze lijkt ons een treffende samenvatting van onze resultaten.

## 7.2 Beoordeling van de resultaten

Met Norbert Elias hebben we de differentiatie bij generatie en educatiegraad gekoppeld aan een verhoogde standaardtaalbeheersing en registergevoeligheid (zie 6.3.2): als sprekers beter weten wat de officiële norm is, zullen ze in onofficiële situaties informele elementen des te geprononceerder tot uiting laten komen. Daarmee is de uitkristallisatie van de Vlaamse spreektaal aan net hetzelfde proces van emancipatie toe te schrijven dat tot de dialectische omslag van de pragmatische generatie heeft geleid: aan de ene kant betekende dat namelijk het aanleren aan de Vlamingen van een algemeen inzetbare standaardtaal (in casu: het Journaalnederlands), maar aan de andere kant steeg daardoor ook de participatie aan het openbare leven, zodat er als vanzelf een verdere endogenisering van het taalgebruik optrad (men denke hierbij aan de opmerkingen van Abram de Swaan over de verschuiving van een bevelshuishouding naar een onderhandelingshuishouding – zie 5.3): er zijn namelijk onvermijdelijk idiolectische verschillen tussen sprekers onderling ('elke spreker spreekt zoals hij gebekt is'), en in het bijzonder geldt voor Vlaanderen dat niet elke taalgebruiker het Journaalnederlands even machtig is. Voor de differentiatie brengt zulks dan met zich mee dat ze eveneens opgenomen zal worden in het geheel van talige conventies, zoals elke spreker die moet internaliseren (vandaar dat we één en ander juist als 'uitkristallisatie' bestempeld hebben; vergelijk in dat verband verder de notie '(her)formalisering' van Cas Wouters – zie 6.3.2): de variatiestructuur van het Belgisch-Nederlands wordt niet langer beschouwd als een gebrekkige realisatie van een ideële standaard, maar integendeel als ingeschreven in de grammatica.

Dat schept enkele interessante perspectieven voor attitudeonderzoek – waarmee we aan het einde van onze descriptieve studie gekomen zijn en dan ook enige suggesties voor toekomstig onderzoek willen doen. De hier geschetste interpretatie maakt namelijk aannemelijk dat de jongste generaties Vlaamse taalgebruikers de

variatiestructuur van het Belgisch-Nederlands in haar geheel positief zullen evalueren, dat wil zeggen: niet alleen het gebruik van Journaalnederlandse varianten op formeel niveau zal worden gewaardeerd, maar dus ook het gebruik van Soapvlaamse varianten op informeel niveau. Dat zou overigens geen geringe constatering zijn: in tegenstelling tot wat Taeldeman (1992) en De Schutter (1998) bijvoorbeeld beweren, houdt het namelijk in dat de jongste sprekers van het Belgisch-Nederlands geen last (meer) zullen hebben van de beruchte ‘taalonzekerheid’ – zie hoofdstuk 1. Daarmee zou dan een heuse barrière in het Vlaamse standaardiseringsproces genomen zijn. Vooralsnog bestaat er echter geen systematische uitgewerkte studie omtrent attitudes tegenover hedendaags taalgedrag, maar de resultaten in Vandenhende (2001) en Impe (2006) lijken de hier geopperde hypothese alvast te ondersteunen.

In die optiek is één van de belangrijkste conclusies uit ons onderzoek nog wel de problematische inhoud van het begrip ‘uniforme’ of ‘homogene’ taal, en de daaraan verbonden notie van een ‘taalkloof’. Niet alleen met betrekking tot de interacties van generatie en educatieniveau gaat toenemende standaardisering gepaard met een uitkristallisatie van variëteiten, ook bij de bivariate analyses hebben we voor zowel een taalinterne (register – zie 5.1) als een taalexterne factor (beroepsniveau – zie 5.7) gezien dat de variatie *binnen* het officieel standaardtalige Journaalnederlands vrijwel even groot is als de variatie *tussen* dat Journaalnederlands en het ‘substandaardtalige’ Soapvlaams. Het doet de vraag rijzen naar de precieze verhouding tussen standaardtaligheid aan de ene kant en uniformiteit/homogeniteit aan de andere kant. Uit onze analyse lijken zulke begrippen namelijk naar voren te komen als de negatie van iedere vorm van standaardisatie eerder dan als de consequentie ervan.

In het licht van dat alles willen we daarom afsluiten met een persoonlijke bedenking achteraf. Met de conventionalisering van het Soapvlaams dunkt er ons namelijk een nieuwe vorm van sociale ongelijkheid in de maak: één die draait rond het beheersen van een zo rijkelijk mogelijk geschakeerde taalvariëteit. Wie dat niet doet – of weigert het spel mee te spelen – zal zichzelf in stijgende mate stigmatiseren als onvoldoende competent en zich zo enigszins buiten de maatschappij plaatsen. Het gebruik van het Soapvlaams is met andere woorden *gecommodificeerd*, en het spreken van Belgisch-Nederlands dreigt daarmee steeds meer neer te komen op een *display of riches*. Die is evenwel ‘*embarrassing*’ aangezien die sprekers net degenen zijn die zich dat op grond van hun maatschappelijke positie en alle sociale voordelen die daarbij horen – zoals beter onderwijs bijvoorbeeld – ook kunnen veroorloven: uitgerkend zij verkeren in de mogelijkheid om het ‘linguïstische kapitaal’ de facto te verwerven. Daarmee zijn de kiemen gelegd voor een symbolische machtsstrijd: in een (postmoderne) samenleving zoals de onze waar de informalisering alle lagen van het maatschappelijk verkeer doordrongen heeft is het bezit van een vlotte spreekstijl namelijk een krachtig wapen. Het heeft er daarom alle schijn van dat het Belgisch-

Nederlands in de toekomst een groeiende maatschappelijke tweespalt te kennen zal geven tussen linguïstische *haves* versus *have-nots*.



## Bibliografie

Agresti, A. 2002. *Categorical data analysis*. Hoboken: Wiley.

Anderson, T. 2003. *An introduction to multivariate statistical analysis*. Hoboken: Wiley.

ANS, zie: Haeseryn *et al.* 1997.

Bakema, P. 1997. "Het onvoltooid verleden verkleinwoord". *Tijdschrift voor Nederlandse taal- en letterkunde* 113 (3), 201-221.

Becker, H. 1992. *Generaties en hun kansen*. Amsterdam: Meulenhoff.

Beheydt, L. 2003. "De moeizame weg van een standaardtaal. De externe en interne druk op het Nedderlands". In: J. Stroop (red.) *Waar gaat het Nederlands naartoe? Panorama van een taal*. Amsterdam: Bert Bakker, 152-163.

Belemans, R. & R. Vandekerckhove 1999. *Variation in (sub)standard language*. Belgian Journal of Linguistics 13. Amsterdam: John Benjamins.

Benzécri, J.-P. 1992. *Correspondence analysis handbook*. New York: Dekker.

Berteloot, A. 2003. "Van 'du' naar 'ghi'. Waarom het pronomen 'du' uit het Nederlands verdween". *Tijdschrift voor Nederlandse taal en letterkunde* 119 (3), 204-217.

Biber, D. 1995. *Dimensions of register variation. A cross-linguistic comparison*. Cambridge: Cambridge University Press.

Binnemans, T. 2003. *Taal en identiteit: een functionele analyse van de taalvariatie in het televisieprogramma '2x Enkel'*. Leuven: Licentiaatsverhandeling.

- Binnenpoorte, D. 2002/4. *Protocol voor manuele verificatie van automatisch gegenereerde woordsegmentaties*.  
[http://lands.let.kun.nl/cgn/doc\\_Dutch/topics/version\\_1.0/annot/word\\_align/wrd\\_prot\\_nl.pdf](http://lands.let.kun.nl/cgn/doc_Dutch/topics/version_1.0/annot/word_align/wrd_prot_nl.pdf).
- Boone, M., H. Gaus & P. Scholliers 1982. *Dagelijks leven. Sociaal-culturele omstandigheden vroeger en nu*. Deurne: Baart.
- Borg, I. & P. Groenen 2005. *Modern multidimensional scaling. Theory and applications*. New York: Springer.
- Cajot, J. 1998. "Een omgangstaal voor alledag. Vlaanderens eigen weg". *Streven* 65 (11), 999-1008.
- Cajot, J. 1999a. "Nieuwsnederlands en Verkavelingsvlaams: de zorgen in een binationaal taalgebied". In: B. De Soomer, E. Moonen & L. Renders (red.) *Kanwilverstaan. Festschrift – Feestbundel Jos Wilmots*. Diepenbeek: Limburg Universitair Centrum, 41-62.
- Cajot, J. 1999b. "Vlaanderens eeuwige weg naar een standaardtaal". In: J. Cajot (red.) *Bedrijfsbeheer & Taalbedrijf. Jubileumboek 30 jaar VLEKHO. 46 bijdragen over bedrijfsbeleid, recht, (ver)taal- en letterkunde*. Brussel: Hogeschool voor Wetenschap en Kunst. Vlaamse Economische Hogeschool, 367-379.
- Cajot, J. 2000. "De omgangstaal van Vlaanderen". *Over Taal* 39 (1), 3-6.
- Cox, T. & M. Cox 2001. *Multidimensional scaling*. Boca Raton: Chapman and Hall/CRC.
- Craps, G. 2002. "De humanisering van het taalbeleid in Vlaanderen: een analyse en een vooruitblik". In: J. De Caluwe, D. Geeraerts, S. Kroon, V. Mamadouh, R. Soetaert, L. Top & T. Vallen (red.) *Taalvariatie en Taalbeleid. Bijdragen aan het taalbeleid in Nederland en Vlaanderen*. Antwerpen & Apeldoorn: Garant, 69-86.
- De Bisschop, K. 2003. *De recente taalpolitieke discussie in Vlaanderen. Een analyse*. Leuven: Licentiaatsverhandeling.



- Debrabandere, F. 2005. "Het echech van de ABN-actie in Vlaanderen". *Nederlands van nu* 53 (1), 27-31.
- De Caluwe, J. 2002. "Tien stellingen over functie en status van tussentaal in Vlaanderen". In: J. De Caluwe, D. Geeraerts, S. Kroon, V. Mamadouh, R. Soetaert, L. Top & T. Vallen (red.) *Taalvariatie en Taalbeleid. Bijdragen aan het taalbeleid in Nederland en Vlaanderen*. Antwerpen & Apeldoorn: Garant, 57-68.
- De Caluwe, J. D. Geeraerts, S. Kroon, V. Mamadouh, R. Soetaert, L. Top & T. Vallen 2002. *Taalvariatie en Taalbeleid. Bijdragen aan het taalbeleid in Nederland en Vlaanderen*. Antwerpen & Apeldoorn: Garant.
- Den Boon, T. & D. Geeraerts 2005. *Van Dale Groot Woordenboek van de Nederlandse taal*. Antwerpen: Van Dale Lexicografie.
- Deprez, K. 1981. *Naar een eigen identiteit. Resultaten en evaluatie van tien jaar taalsociologisch en sociolinguïstisch onderzoek betreffende de standaardtaal in Vlaanderen: perspectieven voor verder onderzoek*. Kessel-Lo.
- De Rooij, J. 1990. "Over 'hun' en 'hen', en 'hun': vorm en functie van de niet-gereduceerde voorwerpsvormen van het persoonlijk voornaamwoord in de derde persoon meervoud, in standaardtaal, ouder Nederlands en dialect". *Taal en Tongval* 42 (2), 107-147.
- De Rooij, J. 1994. "'D'r' en 'ze': over de gereduceerde niet-onderwerpsvormen van het persoonlijk voornaamwoord in de derde persoon vrouwelijk enkelvoud". *Acta Universitatis Wratislaviensis. Neerlandica Wratislaviensia* 7, 161-181.
- De Schutter, G. 1966. "De verbuiging van het onbepaald lidwoord en van een aantal bijvoeglijke woorden in de Zuidnederlandse dialecten". *Taal en Tongval* 18 (3), 71-99.
- De Schutter, G. 1967. "De verbuiging van het onbepaald lidwoord en van een aantal bijvoeglijke woorden in de Zuidnederlandse dialecten". Gent: Gentse bijdragen 1967, 48-76.
- De Schutter, G. 1973. "Eksogeen taalgebruik in Zuid-Nederland". In: *Album Willem Pee*. Tongeren, 117-123.

- De Schutter, G. 1989. *Pronominale clitica in de Nederlandse dialecten*. Antwerp Papers in Linguistics 58. Antwerpen: Universiteit Antwerpen, Universitaire Instelling Antwerpen, Departement Germaanse, Afdeling Linguïstiek.
- De Schutter, G. 1998. "Talen, taalgemeenschappen en taalnormen in Vlaams-België". *Verslagen en mededelingen van de Koninklijke Academie voor Nederlandse Taal- en Letterkunde* 108 (2-3), 227-251.
- De Schutter, G., B. van den Berg, T. Goeman & T. de Jong 2005. *Morfologische atlas van de Nederlandse dialecten*. Amsterdam: Amsterdam University Press.
- De Sutter G. 2005. *Rood, groen, corpus! Een taalgebruiksgebaseerde analyse van woordvolgordevariatie in tweeledige werkwoordelijke eindgroepen*. Leuven: Doctorale dissertatie.
- De Swaan, A. 1983. *De mens is de mens een zorg. Opstellen 1971-1981*. Amsterdam: Meulenhoff.
- De Vriendt, S. 1989. "Nederlands in Nederland en Vlaanderen". *Acta Universitatis Wratislaviensis. Neerlandica Wratislaviensia* 4, 91-114.
- De Vries, J. 2001. *Onze Nederlandse spreektaal*. Den Haag: Sdu.
- Elias, N. 1982. *Het civilisatieproces. Sociogenetische en psychogenetische onderzoeken*. Utrecht: Het Spectrum.
- Friedl, J. 2002. *Mastering regular expressions*. Sebastopol: O'Reilly.
- Gaus, H. 2001. *Politieke en sociale evolutie van België*. Leuven: Garant.
- Geeraerts, D. 1990. "Het dialect en de dialectiek, de Verlichting en het Vlaams". *Dietsche Warande en Belfort* 135 (4), 432-441.
- Geeraerts, D. 1993. "Postmoderne taalattitudes". *Streven* 60 (4), 346-353.
- Geeraerts, D. 1998. "VRT-Nederlands en Soap-Vlaams". *Nederlands van nu* 46 (4), 75-77.

- Geeraerts, D. 1999a. "Taalselectie en woordkeuze: VRT-Nederlands en soap-Vlaams". In: R. Welters (red.) *Grenzeloze selectie. 75 jaar Katholieke Universiteit Nijmegen*. Nijmegen: Uitgeverij KU Nijmegen, 19-28.
- Geeraerts, D. 1999b. "De Vlaamse taalkloof". *Over Taal* 38 (2), 30-34.
- Geeraerts, D. 2001a. "Een zondagspak? Het Nederlands in Vlaanderen: gedrag, beleid, attitudes". *Ons Erfdeel* 44 (3), 337-343.
- Geeraerts, D. 2001b. "Everyday language in the media: the case of Belgian Dutch soap series". In: M. Kammerer, K.-P. Konerding, A. Lehr, A. Storrer, C. Thimm & W. Wolski (red.) *Sprache im Alltag. Beiträge zu neuen Perspektiven in der Linguistik Herbert Ernst Wiegand zum 65. Geburtstag gewidmet*. Berlin & New York: Mouton de Gruyter, 281-291.
- Geeraerts, D. 2001c. "On measuring lexical convergence". In: A. Soares da Silva (red.) *Linguagem e Cognição. A Perspectiva da Linguística Cognitiva*. Braga: Associação Portuguesa de Linguística, 51-61.
- Geeraerts, D. 2002. "Rationalisme en nationalisme in de Vlaamse taalpolitiek". In: J. De Caluwe, D. Geeraerts, S. Kroon, V. Mamadouh, R. Soetaert, L. Top & T. Vallen (red.) *Taalvariatie en Taalbeleid. Bijdragen aan het taalbeleid in Nederland en Vlaanderen*. Antwerpen & Apeldoorn: Garant, 87-104.
- Geeraerts, D. 2003. "Cultural models of linguistic standardization". In: R. Dirven, R. Frank & M. Pütz (red.) *Cognitive models in language and thought. Ideology, metaphors and meanings*. Berlin: Mouton de Gruyter, 25-68.
- Geeraerts, D. & G. De Sutter 2003. "'Ma wa zegdegij nu? Da kanekik nie verstaan zelle!' Taalgedrag, taalbeleid en taalattitudes in Vlaanderen". In: J. Stroop (red.) *Waar gaat het Nederlands naartoe? Panorama van een taal*. Amsterdam: Bert Bakker, 54-64.
- Geeraerts, D., S. Grondelaers & P. Bakema 1994. *The structure of lexical variation*. Berlin & New York: Mouton de Gruyter.
- Geeraerts, D., S. Grondelaers & D. Speelman 1999. *Convergentie en divergentie in de Nederlandse woordenschat. Een onderzoek naar kleding- en voetbaltermen*. Amsterdam: Meertens Instituut.

- Geeraerts, D., A. Penne & V. Vanswegenoven 2000. "'Thuis'-taal en 'Familie'-taal. Taalgebruik in Vlaamse soaps". In: S. Gillis, J. Nuyts & J. Taeldeman (red.) *Met taal om de tuin geleid. Een bundel opstellen voor Georges de Schutter ter gelegenheid van zijn pre-emeritaat*. s.n., 161-170.
- Geerts, G. 1983. "Brabant als centrum van de standaardtaalontwikkeling in Vlaanderen". *Forum der Letteren* 24 (1), 55-63.
- Geerts, G. 1985. "Taalvariatie en taalnormen in Vlaanderen". *Verslagen en mededelingen van de Koninklijke Academie voor Nederlandse Taal- en Letterkunde* 1, 85-112.
- Geerts, G. 1988. "Genusfouten: hollanditis in Vlaanderen?". *Verslagen en mededelingen van de Koninklijke Academie voor Nederlandse Taal- en Letterkunde* 1, 68-78.
- Gillis, S. 2001. *Protocol voor brede fonetische transcriptie*.  
[http://lands.let.kun.nl/cgn/doc\\_Dutch/topics/version\\_1.0/annot/phonetics/fon\\_prot.pdf](http://lands.let.kun.nl/cgn/doc_Dutch/topics/version_1.0/annot/phonetics/fon_prot.pdf).
- Goedertier, W. & S. Goddijn 2000. *Protocol voor Orthografische Transcriptie*.  
[http://lands.let.kun.nl/cgn/doc\\_Dutch/topics/version\\_1.0/annot/orthography/ort\\_prot.pdf](http://lands.let.kun.nl/cgn/doc_Dutch/topics/version_1.0/annot/orthography/ort_prot.pdf).
- Goossens, J. 1970. "'Belgisch beschaafd Nederland' en Brabantse expansie". *De Nieuwe Taalgids* 63, 54-70.
- Goossens, J. 1971. "Algemeen Vlaamse woordenschat". *Wetenschappelijke Tijdingen* 30 (1), 33-36.
- Goossens, J. 1973. "De Belgische uitspraak van het Nederlands". *De Nieuwe Taalgids* 66 (3), 230-240.
- Goossens, J. 1975. "De ontwikkeling van het gesproken Nederlands in Vlaanderen". *Nu Nog* 23, 51-62.
- Goossens, J. 1978. "Nederlandse taalintegratie of Vlaamse taalfrustratie". In: G. Geerts (red.) *Taal of taaltje? Een bloemlezing taalpolitieke beschouwingen over het Nederlands*. Leuven: Acco, 235-248.

- Goossens, J. 1981a. "Willen en half durven in Vlaanderen" *Nederlands van nu* 29 (5), 126-133.
- Goossens, J. 1981b. "De gevolgen van 150 jaar België voor het taalgebruik in Vlaanderen". *Verslagen en mededelingen van de Koninklijke Academie voor Nederlandse Taal- en Letterkunde* 2, 196-213.
- Goossens, J. 1983a. "Heeft het Nederlands variatiegrenzen in het Vlaams?" *Nederlands van nu* 31 (3), 69-78.
- Goossens, J. 1983b. "De gevolgen van het Belgisch staatsverband voor het taalgebruik in Vlaanderen". *Acta Universitatis Wratislaviensis. Neerlandica Wratislaviensia* 1, 161-167.
- Goossens, J. 1987. "Het gebruik van dialect en Algemeen Nederlands en de evolutie ervan". *Verslagen en mededelingen van de Koninklijke Academie voor Nederlandse Taal- en Letterkunde* 3, 289-307.
- Goossens, J. 1990. "Vlaamse pronominale tussenstabilisering". In: J. B. den Besten (red.) *Vragende wijs: vragen over tekst, taal en taalgeschiedenis. Bundel aangeboden aan Leopold Peeters bij zijn afscheid als Hoogleraar Historische Taalkunde van het Nederlands aan de Universiteit van Amsterdam*. Amsterdam: Rodopi, 89-94.
- Goossens, J. 1991. "Die Standardisierung des Niederländischen Belgien als Verfremdung". In: Eijirō Iwasaki (red.) *Begegnung mit dem 'Fremden': Grenzen, Traditionen, Vergleiche. Akten des VIII. Internationalen Germanisten-Kongresses, Tokyo 1990*. München: Ludicum Verlag, 406-412.
- Goossens, J. 2000. "De toekomst van het Nederlands in Vlaanderen". *Ons Erfdeel* 43 (1), 3-13.
- Goudsblom, J. 1988. *Taal en sociale werkelijkheid. Sociologische stukken*. Amsterdam: Meulenhoff.
- Greenacre, M. 1984. *Theory and applications of correspondence analysis*. Londen: Academic Press.

- Greenacre, M. 2007. *Correspondence analysis in practice*. Boca Raton: Chapman and Hall/CRC.
- Greenacre, M. & J. Blasius 2006. *Multiple correspondence analysis and related methods*. Boca Raton: Chapman and Hall/CRC.
- Grondelaers, S. 2000. *De distributie van niet-anaforsch er buiten de eerste zinsplaats. Sociolexicologische, functionele en psycholinguïstische aspecten van er's status als presentatief signaal*. Leuven: Doctorale dissertatie.
- Grondelaers, S., H. Van Aken, D. Speelman & D. Geeraerts 2001. "Inhoudswoorden en preposities als standaardiseringsindicatoren. De diachrone en synchrone status van het Belgische Nederlands". *Nederlandse Taalkunde* 6 (3), 179-202.
- Grondelaers, S., D. Geeraerts, D. Speelman & J. Tummers 2001. "Lexical standardisation in internet conversations. Comparing Belgium and The Netherlands". In: J. Fontana, L. McNally, M. Teresa Turell & E. Vallduví (red.) *Proceedings of the first International Conference on Language Variation in Europe*. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada, Unitat de Investigació Lingüística, 90-100.
- GvD of GvD14, zie: Den Boon & Geeraerts 2005.
- Haeseryn, W. 1996. "Grammaticale verschillen tussen het Nederlands in België en het Nederland in Nederland: een poging tot inventarisatie". In: R. van Hout & J. Kruijssen (red.) *Taalvariaties. Toonzettingen en modulaties op een thema*. Dordrecht: Foris Publications, 109-129.
- Haeseryn, W. 2004. "Belgisch Nederlands voor Vlamingen!". In: J. De Caluwe, G. De Schutter, M; Devos & J. Van Keymeulen (red.) *Taeldeman, man van de taal, schatbewaarder van de taal*. Gent: Academia Press, 481-489.
- Haeseryn, W., K. Romijn, G. Geerts, J. de Rooij & M. C. Van den Toorn 1997. *Algemene Nederlandse Spraakkunst*. Deurne: Wolters Plantyn.
- Haugen, E. 1972. *The ecology of language*. Stanford: Stanford University Press.
- Hendrickx, R. 1998a. *Het Taalcharter*.  
[http://vrtaal.net/taaldatabanken\\_master/taalbeleid/taalcharter.shtml](http://vrtaal.net/taaldatabanken_master/taalbeleid/taalcharter.shtml).

- Hendrickx, R. 1998b. "Algemeen Belgisch Nederlands, de vlag en de lading". *Neerlandia* 102 (5), 28-30.
- Hendrickx, R. 1999. "De taal van de VRT". *Nederlands van nu* 47 (2), 39-41.
- Hendrickx 2001. *Laat ons ne keer te goei naar onszelf luisteren. Taalsignaal voor 6 miljoen Vlamingen.*  
[http://vrtaal.net/taaldatabanken\\_master/taalbeleid/tussentaalcampagne.shtml](http://vrtaal.net/taaldatabanken_master/taalbeleid/tussentaalcampagne.shtml).
- Hendrickx, R. 2002. "Een bedenkelijke analyse". *Over Taal* 41 (3), 58-59.
- Hendrickx, R. 2003. "Wat is Standaardtaal?". *Script* 38, 5-7.
- Hermkens, H. 1985. "Van casus-n naar genus-n". *De Nieuwe Taalgids* 78 (6), 537-538.
- Hinskens, F., C. Hoppenbrouwers & J. Taeldeman 1993. *Dialectverlies en Regiolectvorming*. Themanummer: Taal en Tongval 6. Amsterdam: Meertens Instituut.
- Hoeksema, J. 2000. "Ikke als default-nominatief". *Tabu* 30 (1-2), 27-46.
- Hoekstra, H., M. Moortgat, B. Renmans, M. Schoupe, I. Schuurman & T. Van der Wouden 2003. *CGN syntactische annotatie.*  
[http://lands.let.kun.nl/cgn/doc\\_Dutch/topics/version\\_1.0/annot/syntax/syn\\_prot.pdf](http://lands.let.kun.nl/cgn/doc_Dutch/topics/version_1.0/annot/syntax/syn_prot.pdf).
- Hoppenbrouwers, C. 1990. *Het regiolect. Van dialect tot Algemeen Nederlands.* Muiderberg: Dick Coutinho.
- Hüning, M. 2006. "Reciprociteit in het Nederlands. De geschiedenis van 'elkaar' en 'mekaar'". *Nederlandse Taalkunde* 11 (3), 185-217.
- Impe, L. 2006. *Een attitudeel mixed-guise onderzoek naar tussentaal in Vlaanderen.* Leuven: Licentiaatsverhandeling.

- Janssens, G. 2004. "Ontstaan, ontwikkeling en positie van de 'Vlaamse tussentaal'". In: S. Kiedroń & A. Kowalska-Szubert (red.) *Thesaurus polyglottus et flores quadrilingues. Festschrift für Stanisław Prędoła zum 60. Geburtstag*. Wrocław: Oficyna Wydawnicza ATUT, 173-178.
- Jaspaert, K. 1986. *Statuut en structuur van standaardtalig Vlaanderen*. Leuven: Universitaire Pers.
- Jaspaert, K. 1989. "Van oude normen, de dingen die voorbij gaan: over de Nederlands-Vlaamse taalverhouding". *Kultuurleven* 56 (8), 16-23.
- Jaspaert, K. 1990. "Het Nederlands in Vlaanderen". In: *Jouw taal: sociale invloed op taalgebruik*. Utrecht: Stichting voor Publieksvoorlichting over Wetenschap en Techniek, 55-61.
- Jaspers, J. 2001. "Het Vlaamse stigma. Over tussentaal en normativiteit". *Taal en Tongval* 53 (2), 129-153.
- Jaspers, J. 2002. "De VRT-taalpolitiek. Een bedenkelijk signaal". *Over Taal* 41 (3), 51-57.
- Jaspers, J. 2005. "Algemeen Nederlands en gelijke kansen". *Over Taal* 44 (1), 6-11.
- Kloots, H. 2005. *Vocaalreductie in het Standaardnederlands in Vlaanderen en Nederland*. Antwerpen: Doctorale dissertatie.
- Lebbe, D. 1997. "Schoon Vlaams. Een onderzoek naar syntactische en morfosyntactische aspecten van tussentaal door leperlingen gesproken". *Taal en Tongval* 49 (2), 158-175.
- Lebbe, D. 1998. "Exogeen taalgebruik in Vlaanderen: 25 jaar later". In: R. van Bezooijen, J. Stroop & J. Taeldeman (red.) *Standaardisering in Noord en Zuid*. Themanummer: *Taal en Tongval* 10, 113-125.
- Martens, J.-P. 2002. *Protocol voor de prosodische annotatie*.  
[http://lands.let.kun.nl/cgn/doc\\_Dutch/topics/version\\_1.0/annot/prosody/pr\\_prot.pdf](http://lands.let.kun.nl/cgn/doc_Dutch/topics/version_1.0/annot/prosody/pr_prot.pdf).



- Mertz, D. 2003. *Text processing in Python*. Boston: Addison Wesley.
- Nieuwenhuijsen, P. 2002. "Een onbekend, maar hoogfrequent woord". In: S. Kiedroń & A. Kowalska-Szubert (red.) *170 jaar neerlandistiek in Silezië: V Colloquium Neerlandicum Universitatis Wratislaviensis, Wrocław 12-15 oktober 2000*. Wrocław: Oficyna Wydawnicza ATUT-Wrocławskie Wydawnictwo Oświatowe, 85-92.
- Nijen Twilhaar, J. 1992. "Deflexie en genusverlies". *De Nieuwe Taalgids* 85 (2), 147-160.
- Nuyts, J. 1995. "Subjectspronomina en dubbele pronominale constructies in het Antwerps". *Taal en Tongval* 47 (1), 43-58.
- Oostdijk, N. 2000a. "Het Corpus Gesproken Nederlands". *Nederlandse Taalkunde* 5 (3), 280-284.
- Oostdijk, N. 2000b. "Building a Corpus of Spoken Dutch". In: P. Monachesi (red.) *Computational Linguistics in The Netherlands 1999. Selected papers from the tenth CLIN meeting*. Utrecht Institute of Linguistics OTS, 147-158.
- Oostdijk, N. 2000c. "The Spoken Dutch Corpus. Overview and first evaluation". In: M. Gravididou, G. Carayannis, S. Markantantou, S. Piperidis & G. Stainhaouer (red.) *LREC-2000 (Second International Conference on Language Resources and Evaluation) Proceedings*. Volume 2, 887-894.
- Oostdijk, N. 2000d. "Meta-Data in the Spoken Dutch Corpus Project". In: D. Broeder, H. Cunningham, N. Ide, D. Roy, H. Thompson & P. Wittenburg (red.) *LREC-2000 Workshop Proceedings of the workshop on meta-descriptions and annotations schemes for multi-modal/multi-media language resources*, 21-25.
- Oostdijk, N. 2002. "The design of the Spoken Dutch Corpus". In: P. Peters, P; Collins & A. Smith (red.) *New frontiers of corpus research*. Amsterdam: Rodopi, 105-112.
- Penne, A. & V. Vanswegenoven 1999. "*Oenoemdegij?*": een onderzoek naar het Verkavelingsvlaams in de soapseries 'Thuis' en 'Familie'. Leuven: Licentiaatsverhandeling.

- Piepenbrock, R. 2004. *Taalkundig protocol voor de lexicologische koppeling*.  
[http://lands.let.kun.nl/cgn/doc\\_Dutch/topics/version\\_1.0/annot/lex\\_linkup/lxk\\_p  
rot.pdf](http://lands.let.kun.nl/cgn/doc_Dutch/topics/version_1.0/annot/lex_linkup/lxk_p<br/>rot.pdf).
- Plevoets, K., D. Speelman & D. Geeraerts 2007. "A corpus-based study of modern colloquial 'Flemish'". In: Stephan Elspass, Nils Langer, Joachim Scharloth & Wim Vandebussche (eds.) *Germanic Language Histories 'from Below' (1700-2000)*. Berlijn & New York: Mouton de Gruyter, 179-188.
- Plevoets, K., D. Speelman & D. Geeraerts (te verschijnen). "The distribution of T/V pronouns in Netherlandic and Belgian Dutch". In: Klaus P. Schneider & Anne Barron (eds.) *Variational Pragmatics*. Amsterdam & Philadelphia: John Benjamins.
- Rietveld, T. & R. van Hout 2005. *Statistics in language research: analysis of variance*. Berlijn: Mouton de Gruyter.
- Ryckeboer, H. 2004. "Proefschriften, scripties en verhandelingen in verband met de variatielinguïstiek van het Nederlands". *Taal en Tongval* 56 (1), 81-112.
- Schuurman, I., M. Schoupe, H. Hoekstra & T. Van der Wouden 2003. "CGN, an annotated corpus of spoken Dutch". In: *Proceedings of the 4th International workshop on linguistically interpreted corpora (LINC-03). 14 April 2003. Budapest, Hungary*, 101-108.
- Slembrouck, S. & M. Van Herreweghe 2004. "Teletekstondertiteling en tussentaal: de pragmatiek van het alledaagse". In: J. De Caluwe, G. De Schutter, M; Devos & J. Van Keymeulen (red.) *Taaldeman, man van de taal, schatbewaarder van de taal*. Gent: Academia Press, 853-876.
- Speelman, D., S. Grondelaers & D. Geeraerts 2003. "Profile-based linguistic uniformity as a generic method for comparing language varieties". *Computers and the Humanities* 37, 317-337.
- Speelman, D., S. Grondelaers & D. Geeraerts 2006. "A profile-based calculation of region and register variation: the synchronic and diachronic status of the two main national varieties of Dutch". In: A. Wilson, D. Archer & P. Rayson (red.) *Corpus linguistics around the world*. Amsterdam: Rodopi, 195-202.

- Strang, G. 2003. *Introduction to linear algebra*. Wellesley: Wellesley-Cambridge Press.
- Stroop, J. 1998. *Poldernederlands. Waardoor het ABN verdwijnt*. Amsterdam: Bert Bakker.
- Stroop, J. 2003. *Waar gaat het Nederlands naartoe? Panorama van een taal*. Amsterdam: Bert Bakker.
- Taeldeman, J. 1980. "Inflectional aspects of adjectives in the dialects of Dutch-speaking Belgium". In: W. Zonneveld, F. van Coetsem & O. W. Robinson (red.) *Studies in Dutch phonology*. Den Haag: Nijhoff, 223-254.
- Taeldeman, J. 1983. "Adjektieffleksie in de Zuidnederlands dialecten". In: J. Stroop (red.) *Nederlands dialectonderzoek. Artikelen uit de periode 1927-1982*. Amsterdam: Huis aan de drie grachten, 211-229.
- Taeldeman, J. 1992. "Welk Nederlands voor Vlamingen?" *Nederlands van nu* 40 (2), 33-52.
- Tummers, J. 2005. *Het naakt(e) adjectief. Kwantitatief-empirisch onderzoek naar de adjectivische buigingsalternantie bij neutra*. Leuven: Doctorale dissertatie.
- Van Bezooijen, R. J. Stroop & J. Taeldeman 1997. *Standaardisering in Noord en Zuid*. Themanummer: Taal en Tongval 10. Amsterdam: Meertens Instituut.
- Van Bree, C. 1987. *Historische grammatica van het Nederlands*. Dordrecht: Foris Publications.
- Van Bree, C. 2004. "De Zuid-Hollandse trochee op '-ie' in de enclise". In: S. Daalder, T. Janssen & J. van Leuvensteijn (red.) *Taal in verandering. Artikelen aangeboden aan Arjan Leuvensteijn bij zijn afscheid van de opleiding Nederlandse Taal en Cultuur aan de Vrije Universiteit Amsterdam*. Amsterdam: Stichting Neerlandistiek VU, 121-136.

- Vandekerckhove, R. 2004. "Waar zijn je, jij en jou(w) gebleven? Pronominale aanspreekvormen in het gesproken Nederlands van Vlamingen". In: J. De Caluwe, G. De Schutter, M; Devos & J. Van Keymeulen (red.) *Taeldeman, man van de taal, schatbewaarder van de taal*. Gent: Academia Press, 981-993.
- Vandekerckhove, R. 2006. "Chattaal, tienertaal en taalverandering: (sub)standaardiseringsprocessen in Vlaanderen" *Handelingen van de Koninklijke Zuidnederlandse Maatschappij voor Taal- en Letterkunde en Geschiedenis* 59, 139-158.
- Vandekerckhove, R. & A. Van Rooy 2005. "De chattaal van Vlaamse jongeren". *Over Taal* 44 (2), 30-30.
- Vandenbroeke, C. 1982. *Sociale geschiedenis van het Vlaamse volk*. Leuven: Kritak.
- Vandenhende, A. 2001. *Taalattitude bij studenten: een onderzoek naar de houding van Vlaamse studenten tegenover het Standaardnederlands en haar dialecten*. Leuven: Licentiaatsverhandeling.
- Van der Wal, M & C. van Bree 1992. *Geschiedenis van het Nederlands*. Utrecht: Het Spectrum.
- Van de Velde, H. 1996a. *Variatie en verandering in gesproken Standaard-Nederlands (1935-1993)*. Nijmegen: Doctorale dissertatie.
- Van de Velde, H. 1996b. "Verkavelingsvlaams. Wat is mij dat nu?". In: R. van Hout & J. Kruijssen (red.) *Taalvariaties. Toonzettingen en modulaties op een thema*. Dordrecht: Foris Publications, 261-271.
- Vandeweghe, W. 2000. "Pronominale reduplicatie: sporen in het AN". In: V. De Tier, Magda Devos & J. Van Keymeulen (red.) *Nochtans was scherp van zin: huldealbum Hugo Ryckeboer. Een bundel artikelen aangeboden aan Hugo Ryckeboer voor zijn 65ste verjaardag*. Gent: Universiteit Gent, Vakgroep Nederlandse Taalkunde, 439-443.

- Van Eynde, F. 2004. *Part Of Speech tagging en lemmatisering van het Corpus Gesproken Nederlands*.  
[http://lands.let.kun.nl/cgn/doc\\_Dutch/topics/version\\_1.0/annot/pos\\_tagging/tg\\_prot.pdf](http://lands.let.kun.nl/cgn/doc_Dutch/topics/version_1.0/annot/pos_tagging/tg_prot.pdf).
- Van Gijsel, S. 2001. *Taalgebruik in reclame: een variationele en functionele analyse*. Leuven: Licentiaatsverhandeling.
- Van Gijsel, S. 2007. *Sociovariation in Lexical Richness. A quantitative, corpus linguistic analysis*. Leuven: Doctorale dissertatie.
- Van Gijsel, S., D. Geeraerts & D. Speelman 2004. "A functional analysis of the linguistic variation in Flemish spoken commercials". In: G. Purnelle, C. Fairon & A. Dister (red.) *Le poids des mots. Proceedings of the 7th International Conference on the statistical analyses of textual data*. Louvain-la-neuve : Presses Universitaires de Louvain, 1136-1144.
- Van Istendael, G. 2005. *Het Belgisch labyrint. Wakker worden in een ander land*. Amsterdam: Arbeiderspers.
- Van Loey, A. 1964. *Schönfeld's Historische grammatica van het Nederlands: klankleer, vormleer en woordvorming*. Zutphen: Thieme.
- Van Marle, J. 2002 "Standard Dutch in Flanders: the case of a delayed standardization?". In: P. Wiesinger, H. Derkits, H. Van Uffelen & B. Algot Sørensen (red.) *Akten des X. Internationalen Germanistenkongresses Wien 2000: "Zeitenwende – die Germanistik auf dem Weg vom 20. ins 21. Jahrhundert"*. Bern: Lang, 35-40.
- Vermaas, J. 2002. *Veranderingen in de Nederlandse aanspreekvormen: van de dertiende t/m de twintigste eeuw*. Utrecht: LOT.
- Willemyns, R. 1987. "Norm en variatie in Vlaanderen". In: J. de Rooij (red.) *Variatie en norm in de standaardtaal*. Amsterdam: Meertens Instituut, 143-164.
- Willemyns, R. 1997. "Dialektverlust im niederländischen Sprachraum". *Zeitschrift für Dialektologie und Linguistik* 64 (2), 129-154.

- Willemyns, R. 2002. "De toekomst van het Nederlands en het Nederlands van de toekomst". *Deus ex Machina* 26 (100), 12-17.
- Willemyns, R. 2005. "Verkavelingsbrabants. Werkt het integratiemodel ook voor tussentalen?". *Neerlandica extra muros* 43 (3), 27-40.
- Willemyns, R. & W. Daniëls 2003. *Het verhaal van het Vlaams. De geschiedenis van het Nederlands in de Zuidelijke Nederlanden*. Antwerpen: Standaard Uitgeverij.
- Wouters, C. 1990. *Van minnen en sterven. Informalisering van omgangsvormen rond seks en dood*. Amsterdam: Bert Bakker.
- Zenner, E. 2007. *Expeditie Tussentaal. Een corpusgebaseerd onderzoek naar inter- en intrasprekervariatie in het gebruik van de Vlaamse tussentaal in "Expeditie Robinson"*. Leuven: Licentiaatsverhandeling.
- Zwart, J.-W. 2000. "'Ikke' en de default-naamval". *Tabu* 30 (3-4), 175-182.