



# Variatie in het Nederlandse genus

Een multimethodologisch perspectief

**Evelien Van den Branden**

Masterproef aangeboden binnen de opleiding  
master in de Taalkunde

Promotor prof. dr. Stefania Marzo  
Co-promotor prof. dr. Freek Van de Velde

Academiejaar 2015-2016

197.399 tekens





# Variatie in het Nederlandse genus

Een multimethodologisch perspectief

**Evelien Van den Branden**

Masterproef aangeboden binnen de opleiding  
master in de Taalkunde

Promotor prof. dr. Stefania Marzo  
Co-promotor prof. dr. Freek Van de Velde

Academiejaar 2015-2016

197.399 tekens



## Samenvatting

Het Nederlands beschikt over een grammaticaal genussysteem, waarin pronomina en lidwoorden congrueren met het grammaticale geslacht van hun referent. Onderzoek heeft echter aangetoond dat steeds meer sprekers de tendens vertonen om voornaamwoorden te selecteren op basis van de semantische eigenschappen van de referent. Zo toonde Audring (2006) aan dat het pronomina-systeem van het Noord-Nederlands gebaseerd is op de *Individuation Hierarchy*: hoe hoger een item zich op de hiërarchie bevindt, hoe groter de kans dat er voor het pronomina *hij* of *zij* wordt gekozen; hoe minder geïndividualiseerd een item is, hoe waarschijnlijker er voor het pronomina *het* wordt gekozen. Variatie in de keuze voor bepaalde lidwoorden (of aanwijzende voornaamwoorden) treedt vervolgens ook op, maar wordt vaak in verband gebracht met het taalgebruik van sprekers van allochtone afkomst. Volgens taalkundigen zouden zij het lidwoord *de* overmatig gebruiken, omdat ze het Nederlands niet volledig onder de knie hebben of omdat ze aan de hand van die overgeneralisering hun groepslidmaatschap willen uitdrukken.

Ondanks de vele studies naar variatie in het Nederlandse lidwoordensysteem blijven een aantal vragen onbeantwoord. In deze masterproef vragen we ons ten eerste af in welke mate etnolecten van allochtone sprekers op het vlak van lidwoordtoekenning afwijken van het taalgebruik van moedertaalsprekers en of er inderdaad een overgeneralisering van het lidwoord *de* optreedt. Ten tweede willen we nagaan of er een hersemantiseringsproces in het lidwoordensysteem van etnolecten aan de gang is, zoals Audring (2006) aantoonde in het pronomina-systeem van het Noord-Nederlands. Ten derde trachten we te onderzoeken of aspecten van taalverwerving een rol spelen bij lidwoordtoekenning of dat andere, sociale factoren een invloed uitoefenen. Om een antwoord te kunnen formuleren op deze drie onderzoeksvragen hebben we nood aan een multimethodologische onderzoeksopzet. Enerzijds maken we gebruik van een corpusonderzoek, anderzijds voeren we een formele taalttest uit bij drie leeftijdsgroepen van allochtone afkomst.

De belangrijkste bevindingen van dit onderzoek waren dat de bestudeerde etnolecten (*Marokkaans Nederlands* en *Citétaal*) op het vlak van lidwoordtoekenning afwijken van het taalgebruik van moedertaalsprekers en dat er een tendens tot overgeneralisering van het lidwoord *de* optreedt. De resultaten uit de formele taalttest wezen wel uit dat de respondenten bij sommige testwoorden opvallend vaak opteerden voor het grammaticale lidwoord, wat deed vermoeden dat ze tijdens de taalttest (bewust) rekening hielden met de overgeneralisering van *de*. Vervolgens werd ook de hersemantiseringstheorie in zekere mate bevestigd. We stelden namelijk een tendens vast tot overgeneralisering van het lidwoord *de* in de hoogst-geïndividualiseerde klassen, terwijl we in de laagst-geïndividualiseerde klassen een lichte tendens tot overgeneralisering van *het* konden waarnemen. Wel moeten we daarbij opmerken dat die hersemantisering voornamelijk uitgelokt leek te worden door woorden die lidwoordalternantie vertonen in regionale variëteiten. Verder bleken ook aspecten van taalverwerving een rol te spelen. De kinderen van allochtone afkomst scoorden immers significant lager op de taalttest dan de bevroegde adolescenten en jongvolwassenen. De uiteenlopende scores van de jongeren in het corpusonderzoek en de taalttest deden ten slotte vermoeden dat er ook sociale factoren, zoals groepslidmaatschap, meespelen.



## Dankwoord

Na een periode van intensief werk leg ik met het schrijven van dit dankwoord de laatste hand aan mijn masterproef. Omdat een soortgelijk project nooit het werk is van één persoon, maak ik graag van de gelegenheid gebruik om een aantal mensen te bedanken.

In het bijzonder wil ik mijn promotor en co-promotor, prof. dr. Marzo en prof. dr. Van de Velde, bedanken. Zonder hun advies, opbouwende feedback en persoonlijke begeleiding had ik deze masterproef nooit kunnen voltooien. Daarnaast wil ik ook Karlien Franco bedanken voor de tijd die ze vrijmaakte om samen te overleggen en voor de papers die ze mailde. Dankzij haar bevindingen kon ik bepaalde aspecten van mijn scriptieonderwerp nauwkeuriger onderzoeken. Ook wil ik dr. Eline Zenner graag bedanken voor haar hulp bij de statistische analyse van het veldonderzoek.

Speciale dank gaat ook uit naar de leerkrachten en directie van de scholen die hebben deelgenomen aan het veldonderzoek. Bedankt voor jullie medewerking en het warme onthaal!

Verder heb ik ook veel te danken aan mijn familie. Ik wil graag mijn moeder bedanken voor haar steun en het nalezen van mijn masterproef. Ook mijn vader wil ik bedanken voor zijn aanmoedigingen en voor al zijn hulp bij het verwerken van de onderzoeksresultaten. Mijn zussen, Hannelore en Lieselot, voor de goede raad en ideeën.

Ten slotte wil ik ook al mijn vrienden en vriendinnen bedanken: Silke voor de steun en het luisterende oor; Kristina voor de wederzijdse aanmoediging en de uren die we samen in de bibliotheek doorbrachten; al mijn andere medestudenten voor de goede raad, de mooie momenten en de ontspanning als het nodig was.





# Inhoud

Samenvatting.....	I
Dankwoord .....	III
Lijst van tabellen.....	VII
Lijst van figuren .....	IX
Inleiding.....	1
<b>Hoofdstuk 1: Theoretisch kader .....</b>	<b>3</b>
1.1. Taaltheoretisch luik.....	3
1.1.1. Het Nederlandse genussysteem.....	3
1.1.2. Hersemantisering en Hierachy of individuation.....	6
1.1.3. Verandering in het Zuid-Nederlandse genussysteem .....	8
1.1.4. De tendens tot hersemantisering.....	9
1.2. Sociolinguïstisch luik.....	10
1.2.1. Etnisch taalgebruik als uiting van groepslidmaatschap?.....	11
1.2.2. Etnolect: een proces van etnische identificering en stereotypering?.....	12
1.2.3. Percepties en attitudes over ‘Citétaal’ in Limburg.....	14
1.3. Taalverwervingsonderzoek .....	16
1.3.1. Eerstetaalverwervingsonderzoek.....	16
1.3.2. Tweedetaalverwervingsonderzoek .....	17
<b>Hoofdstuk 2: Onderzoeksvragen en hypotheses .....</b>	<b>21</b>
2.1. Onderzoeksvragen.....	21
2.2. Hypotheses.....	21
2.2.1. Onderzoeksvraag 1.....	22
2.2.2. Onderzoeksvraag 2.....	22
2.2.3. Onderzoeksvraag 3.....	24
2.3. Omschrijving van het onderzoek.....	24
<b>Hoofdstuk 3: Corpusonderzoek.....</b>	<b>27</b>
3.1. Methode.....	27
3.1.1. Selectie van de corpora .....	27
3.1.2. Selectie van de onderzochte woorden.....	29
3.1.3. Doorzoeken van de corpora .....	33
3.2. Analyse en resultaten.....	33
3.2.1. Overzicht van de variabelen.....	33
3.2.2. Analyse en resultaten exclusief zelfstandige naamwoorden met een alternerend lidwoordgebruik .....	35
3.2.3. Analyse en resultaten exclusief ‘raam’.....	38
3.2.4. Logistische regressieanalyse .....	43
3.2.5. ‘Random forest’ en ‘Conditional inference tree’.....	43
3.3. Conclusie .....	47
<b>Hoofdstuk 4: Veldonderzoek.....</b>	<b>49</b>
4.1. Methode.....	49
4.1.1. Beschrijving van de dataverzameling.....	49
4.1.1.1. Proefpersonen en afnameplaats.....	50
4.1.1.2. Beschrijving van de taak.....	51

4.1.2. Pilootstudie .....	54
4.2. Analyse en resultaten .....	<b>55</b>
4.2.1. Exploratie van de verkregen data.....	55
4.2.1.1. Verkenning van de totaalscores ‘grammaticaal genus’ .....	55
4.2.1.2. Verkenning van de variabele ‘leeftijdscategorie’ .....	58
4.2.1.3. Vergelijking tussen de leeftijdsgroepen ‘kinderen’ en ‘jongeren’ .....	60
4.2.1.4. Vergelijking scores op ‘de-woorden’ en ‘het-woorden’ .....	62
4.2.1.5. Vergelijking scores op individuele woorden .....	63
4.2.2. ‘Random forest’ en ‘Conditional inference tree’ .....	66
4.2.2.1. Overzicht van de variabelen .....	66
4.2.2.2. Bespreking van de ‘random forest’ en ‘conditional inference tree’ .....	68
4.3. Conclusie .....	70
<b>Hoofdstuk 5: Conclusie en discussie .....</b>	<b>73</b>
5.1. Eerste onderzoeksvraag .....	73
5.2. Tweede onderzoeksvraag.....	75
5.3. Derde onderzoeksvraag.....	78
Referenties .....	81
Bijlagen .....	i
Bijlage 1: Geattesteerde woorden per corpus .....	i
Bijlage 2: Reguliere expressies .....	v
Bijlage 3: Verkennende plots corpusonderzoek.....	ix
Bijlage 4: Draaitabellen corpusonderzoek.....	xi
Bijlage 4A: Moroccorp .....	xi
Bijlage 4B: Citétaalcorpus.....	xiii
Bijlage 4C: Condiv .....	xiv
Bijlage 5: Leesteksten veldonderzoek .....	xix
Bijlage 5A: Leestekst kinderen (7-8 jaar).....	xix
Bijlage 5B: Leestekst adolescenten (14-15 jaar) en jongvolwassenen (18-21 jaar) .....	xxii

## Lijst van tabellen

<b>Tabel 1</b>	Grammaticaal geslacht in het Oost-Vlaams	4
<b>Tabel 2</b>	Pronominaal woordgeslacht in hedendaags Standaardnederlands: enkelvoudige paradigma's	5
<b>Tabel 3</b>	Geselecteerde corpora voor het corpusonderzoek	26
<b>Tabel 4</b>	Geselecteerde woorden corpusonderzoek per woordgeslacht, per semantische klasse	29
<b>Tabel 5</b>	Geselecteerde woorden met een afwijkend woordgeslacht in supra-regionale variëteiten	30
<b>Tabel 6</b>	Overzicht van de variabelen, niveaus en eventuele codes (corpusonderzoek)	32
<b>Tabel 7</b>	Percentage standaardrealisatie van de geselecteerde woorden (corpusonderzoek)	38
<b>Tabel 8</b>	Overzicht output regressiemodel (corpusonderzoek)	41
<b>Tabel 9</b>	Geselecteerde woorden veldonderzoek	51
<b>Tabel 10</b>	Percentages 'grammaticaal genus' per testwoord uit het <i>Moroccorp</i> , <i>Citétaalcorpus</i> en de taaltest	55
<b>Tabel 11</b>	Resultaten Anova-test en Bonferroni-correctie (veldonderzoek)	57
<b>Tabel 12</b>	Percentages 'grammaticaal genus' per testwoord, per leeftijdscategorie (veldonderzoek)	62
<b>Tabel 13</b>	Overzicht van de variabelen en hun bijhorende niveaus (veldonderzoek)	64
<b>Tabel 14</b>	Percentages 'grammaticaal genus' per testwoord uit het <i>Moroccorp</i> , <i>Citétaalcorpus</i> en de taaltest (kinderen en jongeren)	73



## Lijst van figuren

<b>Figuur 1</b>	Standaard- en niet-standaardrealisatie in de verschillende semantische klassen (corpusonderzoek)	34
<b>Figuur 2</b>	Standaard- en niet-standaardrealisatie van grammaticale <i>de-</i> en <i>het-woorden</i> in de verschillende semantische klassen over alle taalvariëteiten heen (corpusonderzoek)	35
<b>Figuur 3</b>	Standaard- en niet-standaardrealisatie van grammaticale <i>de-</i> en <i>het-woorden</i> in de verschillende semantische klassen in <i>Marokkaans Nederlands</i> en <i>Citétaal</i> (corpusonderzoek)	36
<b>Figuur 4</b>	Standaard- en niet-standaardrealisatie van grammaticale <i>de-</i> en <i>het-woorden</i> in etnolect en autochtoon taalgebruik (corpusonderzoek)	37
<b>Figuur 5</b>	Standaard- en niet-standaardrealisatie in etnolect en autochtoon taalgebruik per variëteit (corpusonderzoek)	37
<b>Figuur 6</b>	Standaard- en niet-standaardrealisatie van grammaticale <i>de-</i> en <i>het-woorden</i> in de verschillende semantische klassen in etnolecten (corpusonderzoek)	39
<b>Figuur 7</b>	Standaard- en niet-standaardrealisatie van grammaticale <i>de-</i> en <i>het-woorden</i> in de verschillende semantische klassen in het <i>Morocccorp</i> (corpusonderzoek)	40
<b>Figuur 8</b>	Standaard- en niet-standaardrealisatie van grammaticale <i>de-</i> en <i>het-woorden</i> in de verschillende semantische klassen in het <i>Citétaalcorpus</i> (corpusonderzoek)	40
<b>Figuur 9</b>	Conditional permutation variable importance voor de random forest met de predictoren <i>De.het</i> , <i>Typedeterminator</i> , <i>Etnolect</i> en <i>Klasse</i> (corpusonderzoek)	42
<b>Figuur 10</b>	Conditional inference tree (corpusonderzoek)	44
<b>Figuur 11</b>	Percentages ‘grammaticaal genus’ per testwoord (veldonderzoek)	54
<b>Figuur 12</b>	Totaalscores ‘grammaticaal genus’ per semantische klasse (veldonderzoek)	55
<b>Figuur 13</b>	Percentage ‘grammaticaal genus’ per semantische klasse, per woordgeslacht (veldonderzoek)	56
<b>Figuur 14</b>	Scores ‘grammaticaal genus’ per semantische klasse, per leeftijdscategorie (veldonderzoek)	58
<b>Figuur 15</b>	Scores ‘grammaticaal genus’ per semantische klasse, per woordgeslacht van de leeftijdscategorie ‘kinderen’ (veldonderzoek)	59

<b>Figuur 16</b>	Scores 'grammaticaal genus' per semantische klasse, per woordgeslacht van de leeftijdscategorie 'jongeren' (veldonderzoek)	60
<b>Figuur 17</b>	Totaalscore 'grammaticaal genus' per woordgeslacht, per leeftijdscategorie (veldonderzoek)	61
<b>Figuur 18</b>	Conditional permutation variable importance voor de random forest met alle predictoren (veldonderzoek)	66
<b>Figuur 19</b>	Conditional inference tree (veldonderzoek)	67

## Inleiding

Variatie in het Nederlandse genussysteem is de laatste jaren reeds intensief onderzocht. Zo bogen verschillende taalkundigen zich over de vraag waarom er in de voorbeeldzin “*Het meisje staat op het perron. Ze wacht op de trein*” met een vrouwelijk persoonlijk voornaamwoord (*ze*) wordt verwezen naar een onzijdig zelfstandig naamwoord (*meisje*). Eerder dan het gebruik van het voornaamwoord *ze* in de voorbeeldzin als ‘foutief’ te bestempelen, wees Audring (2006) er in haar studie op dat het genussysteem van het Noord-Nederlands onderhevig is aan een hersemantiseringsproces. Het oorspronkelijk grammaticale genussysteem, dat in de voorbeeldzin het voornaamwoord *het* vereist, wordt geleidelijk aan vervangen door een semantisch systeem, waarin voornaamwoorden congrueren met het biologische geslacht van hun referent. Het hersemantiseringsproces zou gebaseerd zijn op de graad van individualisering volgens de *Individuation Hierarchy*: hoe sterker de referent als een zelfstandig individu wordt beschouwd, hoe groter de kans wordt dat er voor een mannelijk of vrouwelijk voornaamwoord wordt gekozen; hoe abstracter en minder geïndividualiseerd een item daarentegen is, hoe vaker er met een onzijdig voornaamwoord naar verwezen wordt. Studies naar het genussysteem van het Zuid-Nederlands (De Vos 2009; Franco 2012; De Vogelaer 2006) hebben bovendien aangetoond dat er bij Zuid-Nederlandse pronominale verwijzing ook in zekere mate rekening wordt gehouden met de semantische kenmerken van de referent, hoewel er in heel wat Zuid-Nederlandse dialecten nog relatief sterk wordt vastgehouden aan het conservatieve grammaticale systeem.

In tegenstelling tot de eerste voorbeeldzin, doet de zin “*De meisje wacht op de trein*” bij moedertaalsprekers veel meer stof oplaaien. Lidwoordvariatie – en dan voornamelijk het overmatig gebruik van *de* – wordt immers vaak in verband gebracht met het taalgebruik van allochtone sprekers, die het Nederlandse genussysteem niet onder de knie zouden hebben. Taalverwervingsonderzoek (Cornips & De Vogelaer 2009; Cornips 2008) benadrukt in dat verband hoe moeilijk het grammaticale genussysteem voor verwervers van het Nederlands is en dat tweedetaalverwervers zelfs op volwassen leeftijd nog afwijken van de standaardvorm. Sociolinguïstisch onderzoek (Cornips 2008; Nortier & Dorleijn 2008; Marzo & Ceuleers 2011) wijst er bovendien op dat afwijkingen van de grammaticale standaardvorm van lidwoorden erg aanwezig zijn in de perceptie van moedertaalsprekers en dat dat linguïstisch kenmerk voldoende blijkt om taalgebruikers als ‘allochtoon’ of ‘sprekers van een etnolect’ te bestempelen. De etnolectsprekers zelf zouden volgens sommige taalkundigen mogelijk (bewust) afwijken van de standaardtaal met als doel erkend te worden als lid van een bepaalde groep (Nortier & Dorleijn 2008).

Ondanks de vele studies naar variatie in het genussysteem van autochtone en allochtone variëteiten, zijn een aantal belangrijke zaken nog niet onderzocht. Zo focussen studies naar het taalgebruik van allochtone sprekers van het Nederlands zich eerst en vooral hoofdzakelijk op het overmatig gebruik van het lidwoord *de*, dat als een typisch kenmerk van etnolecten wordt beschouwd. Zelden is er echter onderzoek uitgevoerd naar het volledige lidwoordensysteem van etnolecten en heeft men zich afgevraagd of het lidwoord *het* bij bepaalde woorden ook overmatig gebruikt wordt. Bijgevolg is er ook nog nooit onderzocht of etnolectsprekers bij lidwoordtoekenning mogelijk gebruik maken van een alternatief, semantisch systeem, zoals Audring (2006) dat aantoonde bij pronominale verwijzing in autochtone variëteiten. Anders dan *de meisje* op te vatten als een afwijking van de standaardvorm, zou het lidwoord *de* volgens een semantisch systeem correct gebruikt worden om te

verwijzen naar een vrouwelijke referent (*meisje*). Wanneer er in etnolecten effectief sprake zou zijn van zo'n alternatief semantisch systeem, rijst bovendien de vraag of etnolecten dan wel zozeer verschillen van autochtone variëteiten, waar in het pronominale domein immers ook een proces van hersemantisering aan de gang is.

Ten tweede dient nog beter onderzocht te worden in welke mate aspecten van taalwerving een rol spelen bij lidwoordtoekenning in etnolecten. We vragen ons met andere woorden af of het nieuwe, semantische systeem al zichtbaar is bij kinderen en later merkbaar blijft in het taalgebruik van jongeren en volwassenen van allochtone afkomst. Of spelen er bij jongeren nog andere, sociale factoren mee?

In deze masterproef zullen we een antwoord trachten te formuleren op al deze vragen. Om alle taalkundige visies samen te brengen en om een omvattend inzicht te kunnen krijgen in de kwestie hebben we nood aan een combinatie van meerdere onderzoeksmethodes. Enerzijds maken we gebruik van een grootschalig corpusonderzoek, anderzijds voeren we een taalverwervingsexperiment uit. Aan de hand van een combinatie tussen die twee onderzoeksmethodes willen we eerst en vooral nagaan in welke mate het genussysteem van etnolecten afwijkt van het genussysteem van autochtone variëteiten (onderzoeksvraag 1). We onderzoeken daarbij zowel Noord-Nederlandse als Zuid-Nederlandse variëteiten en we focussen ons enkel op het lidwoordensysteem. Daarnaast willen we ook te weten komen in hoeverre het hersemantiseringsproces reeds van toepassing is op het lidwoordensysteem van beide variëteiten (onderzoeksvraag 2) en in welke mate afwijkingen van het grammaticale genussysteem in etnolecten verklaard kunnen worden aan de hand van sociale factoren of aspecten van taalverwerving (onderzoeksvraag 3).

De paper is opgebouwd uit vijf hoofdstukken. Zo bevat het eerste hoofdstuk een overzicht van alle relevante literatuur, waarin variatie in het genussysteem van autochtone variëteiten en Nederlandse etnolecten wordt beschreven. Omdat verschillende taalkundige takken zich reeds over dat verschijnsel hebben gebogen, zullen we variatie in het Nederlandse genussysteem vanuit meerdere invalshoeken bespreken. In het tweede hoofdstuk zullen we vervolgens de onderzoeksvragen doornemen en daarbij ook enkele hypotheses formuleren. We geven daarbij bovendien een inleiding op het multimethodologisch onderzoek, dat in de twee daaropvolgende hoofdstukken aan bod komt. Het derde hoofdstuk omsluit enerzijds de onderzoeksmethode en onderzoeksresultaten van het corpusonderzoek; het vierde hoofdstuk bevat anderzijds de bespreking van het veldonderzoek. De belangrijkste bevindingen, alsook een kritische evaluatie en een aansporing tot verder onderzoek zullen ten slotte in het vijfde hoofdstuk worden besproken.



# Hoofdstuk 1: Theoretisch kader

In dit eerste hoofdstuk zullen we alle literatuur aanhalen die als theoretische basis diende voor het latere empirisch onderzoek. Variatie in het Nederlandse genussysteem is doorheen de jaren niet alleen een veelvuldig besproken onderwerp geworden, maar is bovenal vanuit meerdere taalkundige perspectieven bestudeerd. Om die reden valt dit eerste hoofdstuk uiteen in drie delen. In het eerste deel (zie 1.1.) focussen we ons op enkele taaltheoretische visies, waarin verandering en variatie in het gebruik van het Nederlandse genus onder meer worden toegeschreven aan een zich ontwikkelend hersemantiseringsproces. In het tweede deel (zie 1.2.) bekijken we de variatie vervolgens vanuit sociolinguïstisch standpunt en leggen we de nadruk op jongerentalen, etnolecten en taalgebruik als uiting van groepslidmaatschap. Het derde deel (zie 1.3.) omvat ten slotte enkele bevindingen uit het taalverwervingsonderzoek, waarin de moeilijkheden bij de verwerving van het Nederlandse genussysteem veelvuldig worden beschreven.

## 1.1. Taaltheoretisch luik

Dat het Nederlandse genussysteem een complex gegeven is, daar zijn heel wat taalkundigen het over eens. Reductie van het genussysteem, het verschil tussen het adnominale en het pronominale domein en het contrast tussen de standaardtaal en regionale variëteiten zouden in taaltheoretische opvattingen aan de basis liggen van een groeiende variatie in het gebruik van genera. Om die standpunten beter te begrijpen, zullen we eerst en vooral een overzicht geven van de werking van het Nederlandse genussysteem (zie 1.1.1.). Daarna bekijken we hoe de variatie vanuit een theorie van hersemantisering verklaard kan worden (zie 1.1.2.) en of die variatie in het Noord-Nederlands en in het Zuid-Nederlands op dezelfde manier optreedt (zie 1.1.3.). Omdat hersemantisering van het Nederlandse genus voornamelijk onderzocht is op het pronominale domein, gaan we in het laatste deel (zie 1.1.4.) na in welke mate definitieve lidwoorden ook aan verandering onderhevig zijn.

### 1.1.1. Het Nederlandse genussysteem

Reductie en verlies in genussystemen is bij Indo-Europese talen een gekend fenomeen. Wanneer een duidelijke gendermarkering van het zelfstandig naamwoord afwezig is of wanneer de paradigma's van adjectieven en lidwoorden gereduceerd worden, verliezen talen de mogelijkheid om bepaalde woordgeslachten in nominale constituenten duidelijk te onderscheiden (Audring 2006: 85). Dat is ook het geval bij het Nederlands, waar het originele drie-generasysteem in de loop van de 17<sup>de</sup> eeuw vervangen werd door een twee-generasysteem (Franco 2013: 3).

Historisch gezien werd er in het Nederlands immers een onderscheid gemaakt tussen drie geslachten: mannelijk, vrouwelijk en onzijdig. Adnominale elementen, zoals lidwoorden en adjectieven, werden volgens dat drie-generasysteem verbogen en congrueerden in geslacht met hun bijhorende zelfstandige naamwoord of nominale constituent. Ook persoonlijke en bezittelijke voornaamwoorden namen het geslacht over van het antecedent waaraan ze refereerden. Wat de motivatie voor het woordgeslacht betreft, zien we dat het geslacht van enkele naamwoorden correspondeerde met het natuurlijke geslacht van hun referenten, maar dat het woordgeslacht van

het overgrote deel van de naamwoorden grammaticaal bepaald was zonder enige semantische motivering. Om die reden spreken taalkundigen van een *grammaticaal genussysteem* (De Vogelaer 2010: 167).

In de loop der jaren werd het drie-generasysteem echter vervangen door een twee-generasysteem, waardoor we in het hedendaagse Standaardnederlands enkel nog een onderscheid maken tussen *de-woorden* en *het-woorden* (Haeseryn et al. 1984: 39). Het historisch mannelijke en vrouwelijke geslacht vallen nu samen onder het commuun geslacht *de*, dat zich onderscheidt van het onzijdige geslacht *het* (Cornips & De Vogelaer 2009: 1). Cornips & Hulk (2008: 269) halen daarbij aan dat volgens een woordenboekgebaseerde schatting (token-analysis) ongeveer 75% van de Nederlandse zelfstandige naamwoorden tot het commuun geslacht behoort en slechts 25% tot het onzijdige. Bovendien zou in een doorlopende tekst de relatieve distributie van *de-* en *het-woorden* ruw geschat 2:1 zijn (Van Berkum 1996), wat de hoge frequentie van *de-woorden* nog eens benadrukt.

Op het adnominale domein zien we dat het huidige twee-generasysteem bepaalt “welk lidwoord gebruikt wordt [...], hoe adjectieven verbogen worden, en welke aanwijzende, vragende en betrekkelijke voornaamwoorden er gebruikt worden” (De Vogelaer 2010: 168). Toch moeten we opmerken dat vele gesproken variëteiten in Vlaanderen (en enkele in Nederland) op het adnominale domein nog steeds gebruik maken van het oudere drie-generasysteem. Dat geldt voornamelijk voor Vlaamse dialecten en regionaal gekleurde variëteiten (Cornips & De Vogelaer 2009: 1). Zo vertonen de Oost-Vlaamse dialecten het volgende systeem:

	mannelijk	vrouwelijk	onzijdig
bepaald lidwoord	<i>de(*n)</i> (h)ond	<i>de</i> fles	't huis
onbepaald lidwoord	<i>ne(*n)</i> (h)ond	<i>een</i> fles	<i>ee(n)</i> huis
voornaamwoord	o.a. <i>mijne(*n)</i>	o.a. mijn	o.a. mijn
adjectief	<i>ne zieke(*n)</i> (h)ond	<i>een lege</i> fles	<i>ons groot</i> huis

Tabel 1: Grammaticaal geslacht in het Oost-Vlaams (Cornips & De Vogelaer 2009: 1; De Vos 2009: 85)

Bovenstaande tabel toont duidelijk aan dat genussystemen van gesproken variëteiten en dialecten van het Zuid-Nederlands veel uitgebreider zijn dan het systeem van het hedendaagse Standaardnederlands. Doordat sprekers van zulke dialecten ook op het adnominale domein voortdurend genusonderscheidingen waarnemen, is het vanzelfsprekend dat ze gemakkelijker onthouden welk grammaticaal genus een bepaald woord heeft. Ze ondervinden dan ook veel minder moeilijkheden bij pronominale verwijzing in de Standaardtaal (De Vos 2009: 86).

Volgens Cornips & De Vogelaer (2009: 3) is het op het pronominale domein dat de impact van de overgang naar een nieuw genussysteem het duidelijkst te merken is. Het onderscheid tussen het mannelijke en vrouwelijke geslacht, dat in het huidige twee-generasysteem volledig verdwenen is, blijft namelijk wel bestaan bij de persoonlijke en bezittelijke voornaamwoorden<sup>1</sup>. Het paradigma van de persoonlijke voornaamwoorden is daarbij het rijkste, omdat er daar nog steeds een duidelijk

<sup>1</sup> In adnominaal gebruik treedt er bij bezittelijke voornaamwoorden congruentie op met het antecedent en niet met het hoofd:

De jongen [M] **zijn** [M] pen [V]

\*De jongen [M] **haar** [V] pen [V]

onderscheid wordt gemaakt tussen drie geslachten (Audring 2006: 85-86), zoals wordt aangetoond in onderstaande tabel.

Pronoun	Case		Masc.	Fem.	Neuter
personal	NOM		hij/ie	zij/ze	het/(e)t
	OBL		hem/m	haar/(d)r	het/(e)t
possessive	NOM/OBL		zijn/z'n	haar/(d)r	zijn/z'n
			<b>Common</b>	<b>Common</b>	<b>Neuter</b>
relative	NOM/OBL		die	die	dat
demonstrative	NOM/OBL	proximal	deze	deze	dit
		distal	die	die	dat

Tabel 2: Pronominaal woordgeslacht in hedendaags Standaardnederlands: enkelvoudige paradigma's (Audring 2006: 86)

Door het onderscheid tussen het adnominale en pronominale domein ontstaat er volgens Audring (2006: 87-88) een zekere discrepantie in het Noord-Nederlandse genussysteem. Ze haalt daarbij aan dat wanneer men met een persoonlijk voornaamwoord naar een zelfstandig naamwoord wil verwijzen, er bepaalde congruentieregels in werking treden en dat de toepassing van die regels erg bemoeilijkt wordt door het feit dat het antecedent minder genuswaarden heeft dan het persoonlijke voornaamwoord. Strikt gesproken kunnen we volgens haar zelfs niet beweren dat het commune geslacht exact hetzelfde is als het mannelijke en vrouwelijke geslacht samen of dat het onzijdige geslacht in het twee-generasysteem dezelfde status heeft als dat van het drie-generasysteem. In de praktijk zou zich dat bij sprekers uiten in een verloren kennis van het mannelijke en vrouwelijke geslacht en zouden ze vaak niet meer in staat zijn om met het juiste grammaticale voornaamwoord naar een zelfstandig naamwoord te verwijzen. Pronominale referentie in het hedendaagse Standaardnederlands bevindt zich daardoor volgens Audring in een chaotische staat: "Zelfstandige naamwoorden van het commune geslacht nemen mannelijke, vrouwelijke en onzijdige voornaamwoorden, en onzijdige zelfstandige naamwoorden kunnen combineren met commune, mannelijke of vrouwelijke voornaamwoorden" (Audring 2006: 88).

Kraaikamp (2012: 204) meent echter dat het verschil in het aantal genera tussen lidwoorden en persoonlijke voornaamwoorden in de praktijk niet zozeer problemen oplevert. In gesproken taalgebruik zou het vrouwelijke geslacht volgens haar enkel gebruikt worden om te verwijzen naar vrouwelijke identiteiten en daardoor geen grote rol meer spelen in het grammaticale genussysteem van het Standaardnederlands. Ook Klom & De Vogelaer (2016) trekken de redenering van Audring (2006) in twijfel en menen dat deflexie en verlies van duidelijke adnominale genuswaarden aan de basis liggen van het probleem.

Welke theorie ook het meest waarschijnlijk lijkt, taalkundigen, zoals Devos (2009), De Vogelaer (2010), Audring (2006) en Kraaikamp (2012), zijn het erover eens dat er in het Noord-Nederlandse genussysteem op het pronominale domein een hersemantiseringsproces aan de gang is.

### 1.1.2. Hersemantisering en Hierachy of individuation

In het Nederlands kunnen er, wat de overeenkomst van het pronominale geslacht met zijn antecedent betreft, in het algemeen twee tendensen worden waargenomen (Klom & De Vogelaer 2016: 3). Eerst en vooral treft men, voornamelijk in het Noord-Nederlands, een zekere *masculinisering* aan, waarbij er naar etymologisch vrouwelijke woorden (zoals *koe*) verwezen wordt met een mannelijk voornaamwoord (Geeraerts 1992; Geerts 1966). Dat zou leiden tot een systeem waarin het voornaamwoord *hij* kan verwijzen naar alle *de-woorden* en het voornaamwoord *het* naar alle *het-woorden*. Ten tweede lijkt het grammaticale genussysteem in zekere mate vervangen te worden door een nieuw semantisch systeem. Opnieuw zijn het de Noord-Nederlandse variëteiten waar die vernieuwing het sterkst wordt waargenomen (De Vogelaer 2010). Dat proces van *hersemantisering* wordt duidelijk beschreven door De Vos (2009: 83):

“De ontwikkeling die men *hersemantisering* noemt, is een wijdverbreid verschijnsel dat in heel wat Indo-Europese talen geattesteerd is. Het verschijnsel bestaat erin dat het oorspronkelijke grammaticale systeem van pronominaal genus wordt vervangen door een nieuw systeem dat gebruik maakt van semantisch gemotiveerde pronomina. [...] In het semantische systeem dat het oude vervangt, stemt het genus van het pronomina niet langer overeen met het genus van het antecedent, maar wordt de keuze van het pronomina gebaseerd op de conceptuele eigenschappen van de referent.”

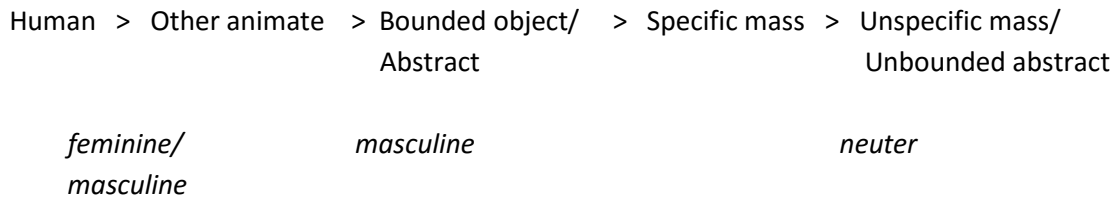
*Hersemantisering* zou onder andere verklaren waarom er in onderstaand voorbeeld uit Kraaikamp (2012: 196) een vrouwelijk voornaamwoord wordt gebruikt om te refereren aan een onzijdig antecedent:

- (1) *Het meisje ging snel naar school,*  
DET.NEUT girl went quickly to school,  
*want ze was te laat.*  
because she was too late.  
'The girl left for school quickly, because she was late.'

In het Noord-Nederlands lijkt het hersemantiseringsproces grotendeels betrekking te hebben op de parameters [mannelijk/vrouwelijk] en [telbaar/massa] (Audring 2006: 101). Nog specifieker zou de keuze voor een mannelijk, vrouwelijk of onzijdig voornaamwoord bepaald worden door de positie van de referent op de *Individuation Hierachy* (Audring 2006; Siemund 2008). Die hiërarchie kan gezien worden als een uitbreiding van de *Animacy Hierachy* van Silverstein (1976), waarvan de sectie *inanimate* in specifiekere klassen werd onderverdeeld (Kraaikamp 2012: 200). Elementen die links op de *Individuation Hierarchy* staan, vertonen een hoge graad van individualisering en hebben bijgevolg duidelijke karaktertrekken en een afgebakende vorm. Aan de rechterkant van de hiërarchie bevinden zich daarentegen minder begrensde entiteiten en moeilijk te specificeren concepten (Kraaikamp 2012: 200). Op basis van die individualiseringsgraad denken taalkundigen de keuze voor het pronomina te kunnen verklaren: “Hoe meer geïndividueerd een item is, hoe groter de kans dat er een niet-neutraal pronomina wordt gekozen. Omgekeerd geldt dat hoe abstracter en minder geïndividueerd een item is, hoe waarschijnlijker er voor een neutraal pronomina wordt gekozen” (De Vos 2009: 83). De niet-neutrale pronomina worden verder onderverdeeld in mannelijke en

vrouwelijke voornaamwoorden op basis van het natuurlijke geslacht van de referent (Audring 2006: 103). Op die manier komt men tot de volgende hiërarchie:

(2) Gender and Individuation Hierarchy in Northern Dutch (Audring 2006: 103)



Helemaal links op de hiërarchie bevinden zich de categorieën *human* en *other animate*, waaraan de hoogste graad van individualisering wordt toegeschreven. Op basis van eerdere bevindingen van Köpcke & Zubin (1996)<sup>2</sup> splitsen Klom & De Vogelaer (2016: 15-16) de klasse *other animate* nog verder op in *high animate* en *low animate*. Tijdens hun onderzoek kwamen ze immers tot de conclusie dat pronominale verwijzing met een onzijdig voornaamwoord vaker voorkomt in de *low animate*-groep, terwijl er in de *high animate*-groep vaker verwezen werd met een niet-neutraal voornaamwoord. In het midden van de hiërarchie bevindt zich vervolgens de klasse *bounded object/abstract*. Die klasse bevat entiteiten die fysisch of conceptueel een duidelijk gebonden vorm vertonen, zoals *boek* of *probleem* (Audring 2006: 102). Kraaikamp (2012) maakt nog een verdere opsplitsing tussen die twee klassen en plaatst de *bounded objects* meer naar links op de hiërarchie dan de *abstracts*. Vervolgens bevat de klasse *specific mass* specifieke instanties of massa's, zoals *mijn moeders soep* of *zijn thee* (Kraaikamp 2012: 201). Het laagst op de hiërarchie staat ten slotte de klasse *unspecific mass/unbounded abstract*, die woorden bevat als *sneeuw*, *zand* of *liefde*.

In onderstaande voorbeelden wordt de semantische overeenkomst tussen pronomen en referent aan de hand van de *Individuation Hierarchy* aangetoond. De voorbeelden zijn afkomstig uit Audring (2006) en werden ook overgenomen in Kraaikamp (2012). In voorbeeld (3) zien we hoe een mannelijk persoonlijk voornaamwoord wordt gebruikt om te verwijzen naar een onzijdig *bounded object*:

- (3) Moet je nog wat informatie over dat **boek** [N] hebben?  
Dan moet 'k 'm [M] ook nog niet gaan inleveren.  
'Do you need some more information about that book? Then I shouldn't return it (lit: him) yet.' (Audring 2006: 95)

In voorbeeld (4) wordt er met een onzijdig pronomen verwezen naar een grammaticaal vrouwelijke referent, omdat die referent zich helemaal onderaan op de *Individuation Hierarchy* bevindt:

- (4) 't zit toch ook bij **olijfolie** [C] een beetje in hoe  
't [N] geconserveerd wordt.  
'Even with olive oil, it matters how it is preserved.' (Audring 2006: 96)

---

<sup>2</sup> In hun studie toonden Köpcke & Zubin (1996) aan dat de klassen *high animate* en *low animate* zich anders gedragen met betrekking tot hun genusovereenkomst in het Duits. Typische voorbeelden waren *Elephant* (M, 'olifant') als *high animal* en *Hummel* (F, 'hommel') als *low animal* (Klom & De Vogelaer 2016: 9).

Voorbeeld (5) toont ten slotte aan dat ook de keuze voor aanwijzende voornaamwoorden semantisch gemotiveerd kan zijn. Een commuun pronomen wordt gebruikt om te verwijzen naar een onzijdig *bounded object*:

- (5) heb jij een **fototoestel** [N]? –  
nee ik kan **die** [C] wel van m'n broer lenen.  
'Have you got a camera? – No, I can borrow my brother's.' (Audring 2006: 98)

### 1.1.3. Verandering in het Zuid-Nederlandse genussysteem

Het hersemantiseringsproces volgens de *Individuation Hierarchy* van Audring (2006), dat volgens sommige taalkundigen in het Noord-Nederlands al volledig zou hebben plaatsgevonden, lijkt op het eerste gezicht in schril contrast te staan met het genussysteem van het zuiden. Zoals we eerder zagen (zie 1.1), behouden heel wat Zuid-Nederlandse dialecten op het adnominale domein immers het conservatieve drie-generasysteem dat gebaseerd is op syntactische congruentie, waardoor sprekers relatief weinig problemen ondervinden bij pronominale verwijzing (De Vos 2009: 85). De vraag rijst daarom hoe die geografische verschillen verklaard kunnen worden. Enerzijds zouden we te maken kunnen hebben met een divergerende tendens tussen het Noord-Nederlands en het Zuid-Nederlands, anderzijds kan het gaan om een algemene ontwikkeling die zich in het noorden sneller doorzet dan in het zuiden (De Vogelaer 2006: 92).

Onderzoek naar pronominale verwijzing in het Zuid-Nederlandse genussysteem (De Vogelaer 2006; De Vos 2009) heeft aangetoond dat die laatste optie het meest waarschijnlijk is. Vlaamse taalverwervende kinderen (6 tot 7 jaar) zouden bijvoorbeeld een aantal tendensen vertonen die grotendeels overeenstemmen met het Noord-Nederlandse semantische genussysteem. Zo gebruiken ze zelden een vrouwelijk voornaamwoord om te verwijzen naar een niet-menselijke referent, verwijzen ze opvallend vaak naar telbare entiteiten met een mannelijk voornaamwoord en verkiezen ze het neutrale *het* om te refereren aan niet-telbare entiteiten (De Vogelaer 2006: 100). De semantische regels die de Vlaamse kinderen gebruiken, zouden daardoor in scherp contrast staan met het (eerder) grammaticale systeem van hun ouders.

Ook het onderzoek van De Vos (2009) naar pronominale verwijzing bij (dialect)sprekers uit het Oost-Vlaamse Moerzeke toont aan dat het Zuid-Nederlandse genussysteem nog grotendeels gebaseerd is op grammaticale regels, maar dat het toch al enkele semantische tendensen vertoont. Die worden voornamelijk voortgestuwd door de jongste leeftijdsgroepen. De Vos (2009: 86-87) haalt bovendien aan dat de dialecten uit het zuiden stilaan verdwijnen en worden verdrongen door tussentalige variëteiten. Door die afnemende dialectkennis (en bijgevolg het verlies van genusgevoel) lijkt het volgens haar zeer waarschijnlijk dat het drie-generasysteem van zuidelijke dialecten zal evolueren naar een twee-generasysteem en dat pronominale verwijzing niet langer gebaseerd zal zijn op syntactische regels. Eerder toonde Hoppenbrouwers (1983) ook al aan dat er een correlatie bestaat tussen de dialectbekwaamheid van Zuid-Nederlandse sprekers en hun kennis van het grammaticale drie-generasysteem.

Het onderzoek van Franco (2013) toont ten slotte ook aan dat Vlaamse sprekers de neiging vertonen om voornaamwoorden te laten congrueren met het natuurlijke geslacht van hun referent. Enerzijds

bevestigen die onderzoeksresultaten dus het belang van individualisering bij pronominale verwijzing in het Zuid-Nederlands, maar anderzijds blijken er toch een aantal kleine verschillen te bestaan tussen het hersemantiseringsproces in Nederland en Vlaanderen. Uit het onderzoek van Franco (2012: 51) kwam namelijk naar boven dat bij Zuid-Nederlandse pronominale verwijzing de invloed van het semantische geslacht kleiner wordt bij niet-humane bezielde referenten en dat de parameter [+/- telbaar] minder van belang is. Franco (2012) stelde daarom een *Individuation Hierarchy* op voor pronominale verwijzing in het Zuid-Nederlands, die als een verkorte versie van de *Individuation Hierarchy* van Audring (2006) beschouwd kan worden:

(6) The Individuation Hierarchy in Southern Dutch (Franco 2012: 44)

Human [male/female]	>	Other animate [male/female]	>	Inanimate
<i>masculine/feminine/ common</i>				<i>neuter</i>

Ondanks de relatief grote gelijkenis tussen het Noord-Nederlands en het Zuid-Nederlands voor wat het belang van individualiseringsgraad bij pronominale verwijzing betreft, stelde Franco (2013: 29) vast dat het genussysteem van definiete lidwoorden in gesproken variëteiten van het Zuid-Nederlands geenszins dezelfde tendensen vertoont als het pronominale genussysteem. Om die bevinding beter te begrijpen, bekijken we in de volgende paragraaf eerst enkele andere resultaten van eerdere onderzoeken naar hersemantisering in het Nederlandse lidwoordensysteem.

#### 1.1.4. De tendens tot hersemantisering

Een belangrijk onderzoek naar verandering in genussystemen en de invloed van woordklassen is de studie van Corbett (1979), waarin genuscongruentie gerepresenteerd wordt op een universele *Agreement Hierarchy*. Bij elementen die zich links op de hiërarchie bevinden, zouden sprekers meer geneigd zijn om ze syntactisch te laten congrueren met hun referent, terwijl elementen aan de rechterkant van de hiërarchie eerder semantische overeenkomst zouden vertonen. Corbett (1979: 216) meende immers dat hoe groter de afstand tussen een woord en zijn referent in een zin was, hoe groter de kans op semantische verwijzing werd. Aangezien de afstand tussen persoonlijke voornaamwoorden en hun referenten het grootst is, zou die woordklasse bijgevolg het eerst onderhevig zijn aan het hersemantiseringsproces. Pas later zouden veranderingen ten voordele van het semantische systeem zich verspreiden naar andere woordklassen, zoals de definiete lidwoorden (De Vos 2009:87).

(7) Agreement Hierarchy (Corbett 1979: 204, met specificering van Kraaikamp 2012: 199)

attributive - predicate - relative pronoun - personal pronoun	
lexical agreement	semantic agreement

Onderzoek in andere talen toonde aan dat hersemantisering van andere woordklassen zich al voltrokken zou hebben in enkele Zuidelijk-Centrale variëteiten van het Italiaans (Fernández-Ordóñez 2009), terwijl het hersemantiseringsproces zich in het Nederlands voorlopig zou beperken tot het pronominale domein. Toch menen taalkundigen, zoals Klom & De Vogelaer (2016: 5), dat het bestaan van een aantal Nederlandse zelfstandige naamwoorden met een dubbel genus erop kan wijzen dat ook het Nederlandse adnominale domein aan een zekere semantisering toe is. Een woord als *diamant* neemt bijvoorbeeld het gemeen lidwoord *de* wanneer het geïnterpreteerd wordt als een telbaar object, maar het onzijdig lidwoord *het* wanneer het als een massa wordt beschouwd (Semplicini 2012: 152).

Een recent onderzoek van Roodenburg & Hulk (2009) toonde bovendien aan dat Noord-Nederlandse kinderen tijdens de verwerving van het Nederlandse lidwoord in beperkte mate rekening houden met semantische kenmerken. Zo zou hun keuze tussen het bepaald lidwoord *de* of *het* beïnvloed worden door het grammaticale woordgeslacht enerzijds, en de semantische parameter [+/-telbaar] anderzijds. Eentalige kinderen tussen 4 en 7 jaar zouden relatief systematisch opteren voor het bepaald lidwoord *het* wanneer het antecedent het semantische kenmerk [-telbaar] bezit.

Ondanks die tendensen tot semantisering in het Noord-Nederlands kwam Franco (2013) tijdens haar onderzoek tot de conclusie dat de keuze voor het gemeen lidwoord *de* op plaatsen waar het onzijdig lidwoord *het* is vereist, in het Zuid-Nederlands onmogelijk kan worden toegeschreven aan de individualiseringsgraad van de referent. Eerder zou variatie veroorzaakt worden door conflicten tussen het genussysteem van de Zuid-Nederlandse standaardtaal en dat van regionale variëteiten. Zoals Geeraerts (2001) en Plevoets (2008) aanhalen, zou er in Vlaanderen immers een betrekkelijk groot verschil bestaan tussen de standaardtaal en de supra-regionale *tussentaal*, die voornamelijk beïnvloed wordt door Brabantse dialectkenmerken. Wanneer we aannemen dat het genussysteem van beide variëteiten lichtelijk verschillend is, zou de keuze voor het gemeen geslacht opgevat kunnen worden als supra-regionaal equivalent voor het onzijdige geslacht in de standaardtaal (Franco 2013: 31-32)<sup>3</sup>. Het veelvuldige gebruik van het lidwoord *de* op plaatsen waar de standaardtaal *het* vereist, zou volgens Plevoets (2008) voornamelijk voorkomen bij sprekers met een lage opleidingsgraad, aangezien zij in het onderwijs minder contact met de standaardtaal hebben ervaren.

## 1.2. Sociolinguïstisch luik

In tegenstelling tot bovenstaande taaltheoretische visies, waar variatie in het Nederlandse genussysteem aan de hand van een hersemantiseringsproces intern linguïstisch verklaard wordt, heeft men variatie bij Nederlandse definitieve lidwoorden vanuit sociolinguïstisch standpunt eerder proberen toe te lichten aan de hand van de functie van taal binnen een sociale groep en de maatschappij. De overgeneralisering van het gemeen lidwoord *de* (het gebruik van *de* op plaatsen waar *het* is vereist) zou zo voornamelijk gelinkt worden aan het taalgebruik van immigranten en hun

---

<sup>3</sup> Uit het werk van Pauwels (1983) en het *Woordenboek der Nederlandsche Taal* (De Vries & Te Winkel 1864-1998) blijkt bijvoorbeeld dat het woord *boek* mannelijk is in enkele Zuid-Nederlandse regionale variëteiten. De constructie *de(n) boek* kan op die manier opgevat worden als (supra-)regionaal equivalent voor de standaardtaalige constructie *het boek*.



nakomelingen (Cornips 2008: 105). In dit tweede deel van het eerste hoofdstuk zullen we daarom nagaan waarom zulke ‘etnische’ variëteiten interessant zijn om taalverandering te onderzoeken en in welke mate groepslidmaatschap een rol speelt in het taalgebruik van allochtone jongeren (zie 1.2.1.). Binnen die context is het belangrijk om ook even stil te staan bij het proces van ‘etnische’ identificatie en stereotypering (zie 1.2.2.), om ten slotte in het laatste deel de nadruk te kunnen leggen op één specifieke variëteit, namelijk *Citétaal* (zie 1.2.3.).

### 1.2.1. Etnisch taalgebruik als uiting van groepslidmaatschap?

Over het algemeen wordt aangenomen dat de rol van taalcontact bij taalverandering nauwelijks overschat kan worden. In bepaalde omstandigheden zou taalcontact immers leiden tot het ontstaan van nieuwe variëteiten van een taal, die *contactvariëteiten* worden genoemd (Van der Sijs 2014: 117-118). Voor het Noord-Nederlands is bijvoorbeeld bekend dat zulke nieuwe variëteiten, die op het vlak van uitspraak, woordenschat, woordvorming en grammatica afwijkingen vertonen van het Standaardnederlands, in overzeese gebieden ontstonden ten gevolge van de 17<sup>de</sup>-eeuwse Nederlandse kolonisatie. Recenter – en meer van belang voor dit onderzoek – zijn enkele nieuwe Nederlandse *etnolecten* die voornamelijk vanuit Turkse en Marokkaanse immigratiecontext in Vlaanderen en Nederland ontstonden (Van der Sijs 2014: 118). Kenmerkend voor *etnolecten* is dat zulke taalvariëteiten nooit uitgroeien tot aparte talen<sup>4</sup>, aangezien de sprekers ervan onder voortdurende invloed van het Standaardnederlands staan. Bovendien rijst de vraag of *etnolecten* wel een duidelijke eenheid vormen, aangezien niet alle taalgebruikers in dezelfde mate afwijken van het Standaardnederlands (Van der Sijs 2014: 120-121).

Uit het onderzoek van Cornips (2008: 105-107) kwam naar voren dat latere generaties van Turkse en Marokkaanse immigranten, die in Nederland geboren werden en het Nederlands bijgevolg als één van hun moedertalen verwierven, nog steeds een overgeneralisering van het gemeenschappelijke woord *de* vertonen. Interessant genoeg stelde Cornips (2008: 107) bovendien vast dat het verdwijnen van het grammaticale genussysteem in Nederlandse contactvariëteiten geen onbekend verschijnsel is. Eerder zouden ook het *Negerhollands*, *Zuid-Afrikaans* en *Surinaams-Nederlands* dezelfde tendens vertoond hebben (Muysken 2001; Donaldson 1993; Cornips 2005). Daarnaast zou de overgeneralisering van één bepaald woordgeslacht ook zijn vastgesteld bij immigrantengemeenschappen in de voorsteden van Zweden (Kotsinas 2001) en Denemarken (Quist 2008).

De hamvraag blijft echter tot in welke mate we in de hedendaagse immigratiecontext kunnen blijven spreken in termen van *etnolecten* en *contactvariëteiten*. Zoals Nortier en Dorleijn (2008: 128) opmerkten, ontstaan contactvariëteiten (in de strikte zin van het woord) immers steeds vanuit een zekere noodzaak om met de ander te communiceren, terwijl bij enkele Noord-Nederlandse taalvariëteiten, zoals *Moroccan flavored Dutch*, werd vastgesteld dat ze bij sprekers eerder opdoken uit het verlangen om zich van bepaalde mensen en groepen te onderscheiden. Jongeren zouden met andere woorden (min of meer) bewust afwijken van de standaardtaal met als doel “erkend te worden als iemand die rondhangt met z’n vriendengroep” (Nortier & Dorleijn 2008: 132).

---

<sup>4</sup> Wanneer dat wel gebeurt, wordt de taalvariëteit niet meer beschouwd als *etnolect*, maar als *creooltaal* (Van der Sijs 2014: 120-121)

Overgeneralisering van het lidwoord *de* zou zo opgevat kunnen worden als een uiting van groepslidmaatschap, zoals ook blijkt uit onderstaand citaat (*J* en *M* zijn de interviewers):

- (8) S: *Dat is het slechte Nederlands*  
J: *En heeft dat ook een naam?*  
S: *Ja, niet echt, maar 't is in principe dan eh lidwoorden die gebruik je dan expres verkeerd*  
M: *Ja ja die gebruik je dan exprès verkeerd, net als-*  
S: *Ja dus*  
J: *Die meisje*  
S: *Die huis zeg ik dan. Terwijl ik weet ik bedoel ik weet heus wel dat het dat huis is maar 't staat zo dom als ik dat op straat zeg, als ik zeg*  
M: *Ja*  
S: *Als ik zeg dat huis*  
M: *Ja ja*  
S: *'t Is gewoon die huis. Maar als ik met jullie spreek dan wordt 't gewoon dat huis.*

(Nortier & Dorleijn 2008: 132)

Vanuit die visie zou de 'onvolledige' verwerving van het Nederlandse genussysteem bij jongeren van allochtone afkomst gezien kunnen worden als betekenisvol, in die zin dat ze lijken te voldoen aan een set van taalregels die bepaald werd door de sprekers van de sociale groep waartoe ze willen behoren (Cornips 2008: 120).

Bovendien werd vastgesteld dat zulk 'afwijkend' taalgebruik niet alleen voorkomt bij leden van etnische minderheidsgroepen, maar dat het ook optreedt bij jongeren van autochtone afkomst. Het *Murks* zou bijvoorbeeld rond de millenniumwisseling gesproken zijn geweest door Nederlandstalige jongeren in zeer informele *in-group* situaties. Zonder een hecht contact te hebben met leden van etnische minderheidsgroepen, baseerden de jongeren hun taalgebruik op stereotypen, voornamelijk omdat ze het sterk en cool vonden klinken (jongens) of om er de spot mee te drijven (meisjes) (Nortier & Dorleijn 2008: 127).

Ook *Citétaal*, de taalvariëteit die later besproken zal worden (zie 1.2.3.), wordt vandaag de dag door Limburgse autochtone jongeren gesproken, mogelijk ook als uiting van groepslidmaatschap. Vooraleer we daarop terugkomen, bespreken we eerst of we mogen en kunnen spreken van iemands 'etnische identiteit' en of de term 'etnisch taalgebruik' al dan niet te stereotyperend is.

## 1.2.2. Etnolect: een proces van etnische identificering en stereotypering?

Een belangrijke opmerking die Cornips (2008: 114) in dat verband maakte, is dat hoewel we tot nu toe voortdurend gesproken hebben over 'Marokkaanse' of 'Turkse' sprekers, we er vrijwel zeker van kunnen zijn dat geen enkel individu dat tot die categorie wordt gerekend, zijn ervaringen als immigrant, zijn opvoeding, zijn geloof of culturele achtergrond op dezelfde manier ervaart. Er is met andere woorden volgens Cornips (2008) geen enkele reden waarom de categorisatie als 'Marokkaan' door buitenstaanders ook groepslidmaatschap zou betekenen vanuit *insiders'* standpunt. Net door sociaal-etnische groepen op voorhand te categoriseren, lopen we de kans om uit het oog te verliezen

dat zo bepaalde leden van etnische groepen een etnische identiteit niet altijd op dezelfde linguïstische manier construeren (Fought 2006: 38). Dat blijkt ook uit volgend citaat:

“There are different types of contexts for the construction of ethnicity through language, including at the very last: (a) the situational context which can vary from one conversation to the next or even within a single conversation, (b) the broader context of an individual’s life history, and (c) the social context of how the community, or a segment of the community, views ethnicity at a particular point in time. Essential in the concept of ethnolect is the important role of both self-identification and the perceptions and attitudes of others in the construction of ethnic identity” (Cornips 2008: 114-115)

Anderzijds worden de termen *etniciteit* en *etnolect* wel betekenisvol in termen van de hedendaagse immigratiecontext. Turkse en Marokkaanse gastarbeiders die in de jaren '60 in de Vlaamse en Nederlandse industriesector tewerkgesteld werden en met hun daaropvolgende generaties in arme stadswijken woonden, werden op basis van hun sociaaleconomische en religieuze achtergrond door de autochtone bevolking grotendeels als één entiteit beschouwd, wat leidde tot de ideologische oppositie tussen 'wij' en 'zij', of meer specifiek tussen een 'autochtone' en een 'allochtone' identiteit (Cornips 2008: 115). Hoewel de overgeneralisering van het lidwoord *de*, zoals we later in de paragraaf over taalverwervingsonderzoek zullen bespreken (zie 1.3.), een linguïstisch probleem blijkt voor elke taalverwerper, wordt dat proces volgens De Rooij (2005) pas betekenisvol in een context van etnische identificatie, waarin de 'allochtone' groep zich door afwijkend taalgebruik onderscheidt van de dominante 'autochtone' groep. In tegenstelling tot kinderen uit Engelse immigrantenfamilies, zijn Marokkaanse en Turkse jongeren nu eenmaal zeer zichtbaar aanwezig in de maatschappij: ze leven niet alleen in etnische minderheidsgroepen, maar ze worden bovendien vaak bekritiseerd door omstanders en politici (Cornips 2008: 115-116). Het is in die context dat de overgeneralisering van het commuun lidwoord *de* de status krijgt van een stereotype (Labov 1994) en dat één bepaald linguïstisch kenmerk voldoende blijkt om taalgebruikers als leden van de 'allochtone groep' te bestempelen (Tabouret-Keller 1997: 317). In onderstaand fragment is bijvoorbeeld te lezen hoe drie jongeren van Surinaamse afkomst het Antilliaans in Rotterdam stereotyperen aan de hand van afwijkend gebruik van het aanwijzend voornaamwoord (*die* in plaats van *dat*):

- (9) *Interviewer:* *kun jij ook Antilliaans.*  
*Ronald:* *dumpen+...*  
*Interviewer:* *hah?*  
*Gerard:* *gewoon straattaal Antilliaans xx.*  
*Interviewer:* *xx een voorbeeld?*  
*Gerard:* *ja dat, ik schop **die ding** terug.*  
*meta:* *they are imitating young Antillian, he laughs*  
*Ronald:* *ik heb gezien dat je xx.*  
*meta:* *they are imitating young Antillians*  
*Vincent:* *ik rende voor die bus, maar die bus rende weg.*  
*meta:* *they are imitating young Antillians*  
*Ronald:* *ik heb gezien dat je die jongen hebt geschiet.*  
*meta:* *they are imitating young Antillians*  
*Vincent:* *ik schiet je met **die mes***

*meta:*                *they are imitating young Antillians*  
*Ronald:*             *jongen met die broek met die korte mouwen.*  
*meta:*                *they are imitating young Antillians*  
*Interviewer:*       *maar dat is geen Antilliaans?*  
*meta:*                *she laughs*  
*Vincent:*            *zo praten ze.*  
*Interviewer:*       *zo praten ze.*

(Cornips & de Rooij 2004, geciteerd door Cornips 2008: 116-117)

Aan de hand van dit voorbeeld kunnen we concluderend stellen dat variatie in het Nederlandse genussysteem vanuit sociolinguïstisch perspectief een cruciale rol blijkt te spelen bij het stereotyperen van ‘de ander’. Niet alleen vatten buitenstaanders de overgeneralisering van het algemeen lidwoord *de* op als een linguïstisch kenmerk van jongeren met een allochtone achtergrond, maar ook de jongeren zelf zouden het mogelijk (bewust) gebruiken om hun groepslidmaatschap te benadrukken (zie 1.2.2.) (Cornips 2008: 120).

### 1.2.3. Percepties en attitudes over ‘Citétaal’ in Limburg

Het is binnen die context van etnische identificatie en stereotypering dat we *Citétaal* in dit onderzoek bespreken. Oorspronkelijk gebruikt om te verwijzen naar de taal van Italiaanse migratiearbeiders, die kort na WOII in geconcentreerde gemeenschappen in de mijnstreken van de provincie Limburg woonden (de zogenaamde *cités*), duidde de term *Citétaal* later op de contactvariëteit die na de aankomst van arbeiders uit Turkije en Marokko in de mijngetto’s ontstond. Door die smeltkroes van verschillende talen en culturen en door het Nederlandstalige onderwijs dat de kinderen van de tweede en derde generatie ontvingen, werd de lokaal gesproken variëteit van het Nederlands steeds belangrijker, zelfs met een zekere afname in het gebruik van de moedertaal tot gevolg (Marzo & Ceuleers 2011: 454). Heden ten dage zouden de termen *Citétaal*, *Algemeen Cités* of *Cités* meer algemeen gebruikt worden om te verwijzen naar het taalgebruik van jongeren in stadswijken van de provincie Limburg, zoals Genk (Marzo 2015: 4).

Hoewel de term *Citétaal* strikt gesproken geen duidelijk afgebakende taalvariëteit omvat, blijkt hij wel aanwezig te zijn in de perceptie van mensen die de variëteit beweren te spreken, evenals in de perceptie van diegenen die de variëteit menen te herkennen. Wanneer er aan zulke mensen wordt gevraagd wat *Citétaal* onderscheidt van andere taalvariëteiten, beantwoorden ze die vraag meestal met de opsomming van een aantal fonologische, grammaticale en lexicale kenmerken. Heel vaak komen zo een reeks morfosyntactische overgeneralisaties ter sprake, die ook vaak met andere Europese etnolecten, zoals *Straattaal* of *Türkendeutsch*, in verband worden gebracht. De overgeneralisering van het algemeen lidwoord *de* ten koste van het onzijdig lidwoord *het* (*de meisje*) of het gegeneraliseerde gebruik van het aanwijzende voornaamwoord *die* (*die meisje*) zijn daarvan de meest opgesomde voorbeelden. Andere vaak aangehaalde kenmerken van *Citétaal* zijn het weglaten van lidwoorden en preposities (*gaat ge cinema?*), het gebruik van leenwoorden en leenvertalingen (*parleren*, ‘spreken’ en *bordel<bordello*, ‘chaos’), de stijgende en dalende intonatiepatronen en de palato-alveolarisatie van [s] in [ʃ] in woorden zoals *stijl* (Marzo & Ceuleers 2011: 451-452).

Zoals hierboven reeds vermeld, is uit onderzoek (Marzo & Ceuleers 2011; Marzo 2015) gebleken dat ook Limburgse jongeren van autochtone afkomst *Citétaal* onder hun vrienden gebruiken. Ze zouden *Citétaal* jong, grappig en cool vinden klinken en de taalvariëteit beschouwen als een symbool van de lokale Vlaamse jeugd en de buurt waar ze zijn opgegroeid, eerder dan dat ze zichzelf affiliëren met een bepaalde etnische groep (Marzo 2015: 4, 17). Dat taalgebruik zou grotendeels in de hand gewerkt worden door lokale televisiezenders, reclameslogans en de stedelijke hiphopcultuur (Marzo en Ceuleers 2011: 455).

Ondanks dat groeiend idee van een lokale variëteit, blijft *Citétaal* in de ogen van vele mensen echter nog steeds een associatie met multi-ethniciteit en migratieklassen oproepen. Voornamelijk zouden sprekers die de variëteit niet beweren te gebruiken, *Citétaal* op die manier stigmatiseren en bekritisieren, waardoor stereotiepe ideeën ontstaan, zoals de onwil van de allochtone jongeren om correcte Standaardtaal te spreken (Marzo 2015: 4).

*Citétaal*, als sociale constructie in de hoofden en conversaties van sprekers, staat bovendien nooit op zichzelf. Zoals Marzo (2015: 9) opmerkt, bekijken sprekers een taalvariëteit als *Citétaal* steeds in relatie tot andere variëteiten, met in het bijzonder in relatie tot de algemene standaardtaal. Door de grote variëteit aan beschrijvingen van *Citétaal*, die verwijzen naar plaats (Genk, Limburg), mensen, leeftijd, sociale klasse en migratie, ontstaat er een duidelijk dialectisch indexicaal veld met een positieve en een negatieve pool: door sommigen wordt *Citétaal* opgevat als een “lokale variëteit om trots op te zijn”, terwijl anderen ze zien als “onbeschofte, incorrecte *slang* die men dient te vermijden” (Marzo 2015: 9-10).

Een treffend voorbeeld daarvan kan gevonden worden in een politieke campagne voor de gemeenteraadsverkiezingen van Genk in 2011-2012. De lokale socialistische partij en de lokale nationalistische partij hadden eerder al openlijk met elkaar gediscussieerd over het gebruik van *Citétaal* in hun campagne, toen de socialistische partij de slogan “*Genk, de stad van de shtijl*” bekend maakte. Door de duidelijke palatalisatie van [s] in [ʃ], zou de slogan al snel met *Citétaal* in verband worden gebracht, wat de constructie van een nieuwe, lokale identiteit van de stad Genk in de hand moest werken. De nationalistische partij, die *Citétaal* sterk associeerde met de oorspronkelijke historische en sociale waarden van migratie, bekritiseerde de slogan echter meteen en liet in haar campagnebrochure weten hoe belangrijk ‘correct’ taalgebruik was voor een goede integratie in de maatschappij (Marzo & Svendsen 2015: 25).

Aan de hand van bovenstaand voorbeeld wordt duidelijk hoe er tegenwoordig een aanzienlijk verschil bestaat tussen de status van *Citétaal* en haar aantrekkelijkheid. Hoewel *Citétaal* in het algemeen wordt opgevat als alledaags en vriendelijk, schrijft men duidelijk een lagere sociale status aan die taalvariëteit toe. Die lage status is hoofdzakelijk te wijten aan de voortdurende associatie van *Citétaal* met vreemde afkomst en etnie en werkt op die manier een zekere stigmatisering in de hand. Het *Algemeen Nederlands* zou daarentegen zeer hoog scoren voor zowel status als aantrekkelijkheid (Marzo 2015: 14-15).

### 1.3. Taalverwervingsonderzoek

Naast de hierboven besproken sociolinguïstische verklaringen voor de overgeneralisering van het commuun lidwoord *de*, benadrukken taalkundigen vanuit taalverwervingsperspectief de enorme moeilijkheden die met de verwerving van het Nederlandse genussysteem gepaard gaan. In dit derde en laatste deel van het eerste hoofdstuk bekijken we daarom de belangrijkste bevindingen uit eerstetaalverwervingsonderzoek (zie 1.3.1.) en tweedetaalverwervingsonderzoek (zie 1.3.2.)

#### 1.3.1. Eerstetaalverwervingsonderzoek

Een belangrijke studie over de verwerving van generasystemen bij eentalige kinderen is die van Mills (1986), waarin een vergelijking wordt gemaakt tussen het Engels en het Duits met betrekking tot de aard van pronominale congruentieregels en de snelheid waarmee het genussysteem wordt verworven. Mills (1986) meent dat hoe prominenter een regel aanwezig is in een taal, hoe vroeger hij wordt verworven (De Paepe & De Vogelaer 2008: 4). In het Duits, een taal met een duidelijk grammaticaal genussysteem, slagen jonge kinderen er al op zeer vroege leeftijd in om het geslacht van het subjectpronomen te laten overeenstemmen met het geslacht van zijn antecedent. Rond 5-6-jarige leeftijd zouden kinderen in hun pronomingebruik nog nauwelijks afwijken van het grammaticale genus en het regelsysteem bijgevolg consequent toepassen. In het Engels daarentegen, waar de keuze voor het subjectpronomen semantisch gemotiveerd is, ervaren jonge kinderen meer problemen bij de verwerving van het pronominale gebruik. Op een leeftijd van 5-6 jaar zouden Engelse kinderen bijvoorbeeld nog steeds de neiging vertonen om het pronomen *he* te veralgemenen (De Paepe & De Vogelaer 2008: 4).

Een verklaring voor dat verschil in verwervingssnelheid is volgens Mills (1986) niet zozeer te vinden in de classificatie van de taal als grammaticaal of semantisch genussysteem, maar eerder in de zichtbaarheid en de consistentie van de regels. Het is volgens haar net omdat een semantisch systeem meer variatie toelaat in het pronominale gebruik, dat het trager verworven zal worden. Prominentie en grote zichtbaarheid van de congruentieregels leiden daarentegen tot snellere verwerving, zoals ook blijkt uit de verwerving van het grammaticale genussysteem van het Duits. Zeer prominent en zichtbaar in het Duits is bijvoorbeeld de regel dat woorden die op een sjwa (-[ə]) eindigen, vrouwelijk zijn. Uit het onderzoek van Mills (1986) bleek dat Duitse kinderen die regel dan ook al op 3-jarige leeftijd onder de knie te hebben. Regels met een kleiner lexicaal bereik of met meer uitzonderingen zijn daarentegen veel moeilijker te verwerven. Zo zouden de regels die bepalen dat woorden die eindigen op *-ft/* en *-cht/* vrouwelijk zijn, of dat woorden die eindigen op *-et/* onzijdig zijn, nog afwezig zijn bij 8-jarige Duitse kinderen (Mills 1986: 70-80).

Onderzoek uit het Frans (Maillart 2003, Van der Velde 2003) heeft bovendien aangetoond dat ook de verwerving van het Franse grammaticale genus bij pronominale verwijzing op zeer jonge leeftijd plaatsvindt. In tegenstelling tot het Duits, heeft het Frans echter geen genusaanduidende naamvalsuitgangen, maar vertoont de taal wel adjectivale verbuiging en wordt er een onderscheid gemaakt in de lidwoordvorm van mannelijke en vrouwelijke woorden. Pronominale genusfouten zouden bij Franse taalverwervende kinderen zeldzaam zijn en bijna volledig verdwijnen rond een leeftijd van 6 jaar. De relatieve zichtbaarheid en de stabiliteit van de regels zouden hier aan de basis liggen van de snelle verwerving.

In tegenstelling tot het Duits, Engels en Frans, is Nederlandse genusverwerving hoofzakelijk bestudeerd vanuit het adnominale domein (De Vogelaer 2010: 4). Daarbij valt onmiddellijk op dat genustoekenning in het Nederlands noch in het pronominale drie-generasysteem, noch bij het onderscheid tussen *de*- en *het*-woorden, onderhevig is aan duidelijke formele regels (De Paepe & De Vogelaer 2008: 5). Hoewel de ANS wel enkele regelmatigheden vermeldt (bv. woorden die eindigen op *-heid* zijn vrouwelijk, woorden die eindigen op *-ment* zijn onzijdig), betreffen ze slechts een beperkt deel van de Nederlandse woordenschat. Dat zou kunnen verklaren waarom Nederlandstalige kinderen meer problemen ondervinden bij de verwerving van het definiëte lidwoord dan Franse en Duitse kinderen (De Vogelaer 2010: 4). Een vaak vastgestelde tendens bij jonge Nederlandstalige kinderen is bijvoorbeeld de overgeneralisering van het algemeen lidwoord *de*, ten koste van het onzijdige lidwoord *het* (Van der Velde 2003: 124-129). Volgens Van der Velde (2003) zouden die fouten verdwijnen rond de leeftijd van 6 jaar. De Paepe & De Vogelaer (2008: 4-5) concluderen daaruit dat afwijkingen van het grammaticale woordgeslacht na een leeftijd van 6 jaar zouden duiden op een afgenomen belang van het grammaticale genussysteem, zeker wanneer die afwijkingen optreden bij woorden die kinderen al op zeer vroege leeftijd verwerven.

### 1.3.2. Tweedetaalverwervingsonderzoek

Zoals uit voorgaande paragraaf over eerstetaalverwervingsonderzoek duidelijk is gebleken, onderscheidt het Nederlands zich vanuit taalverwervingsperspectief van andere talen zoals het Duits en het Frans, door een overgeneralisering van het algemeen lidwoord *de* ten koste van het onzijdig lidwoord *het* en door de betrekkelijk late leeftijd waarop eentalige kinderen het genussysteem verwerven (Cornips & Hulk 2008: 270). De vraag rest nu of tweetalige kinderen diezelfde tendensen vertonen tijdens hun verwerving van het Nederlandse genussysteem en of ze eventueel nog grotere problemen ondervinden. Vanuit dat opzicht zijn etnolecten zeer interessant, omdat ze heel wat niet-standaardvormen bevatten, die voortkomen uit de moeilijkheden van tweedetaalleerders bij de verwerving van het Nederlands.

Vele tweedetaalverwervingsstudies benadrukken in dat verband hoe weinig evident het kenmerk grammaticaal geslacht (*de* versus *het*) voor verwervers van het Nederlands is. Zelfs op volwassen leeftijd zouden tweedetaalverwervers het Nederlandse genussysteem nog niet altijd volledig onder de knie hebben (Cornips & De Vogelaer 2009: 5). Cornips (2008) toont in haar onderzoek bijvoorbeeld aan dat jongvolwassenen van Turkse en Marokkaanse afkomst in het Nederlands nog steeds de tendens vertonen om het algemeen lidwoord *de* te gebruiken op plaatsen waar het onzijdig lidwoord *het* is vereist:

- |         |                     |                  |                           |
|---------|---------------------|------------------|---------------------------|
| (10) a. | <i>de</i>           | geloof           | [Guray, Turkish descent]  |
|         | <i>the</i> -COMMON  | faith-NEUTER     |                           |
| b.      | <i>de</i>           | laatste jaar     | [Hassan, Turkish descent] |
|         | <i>the</i> -COMMON  | last year-NEUTER |                           |
| c.      | <i>de</i>           | filmpje          | [Badir, Moroccan descent] |
|         | <i>the</i> -COMMON  | movie-NEUTER+DIM |                           |
| d.      | <i>die</i>          | nummer           | [Hasan, Turkish descent]  |
|         | <i>this</i> -COMMON | number-NEUTER    |                           |

Voorbeelden uit Cornips (2008: 107-108)

Opmerkelijk genoeg werd vastgesteld dat het taalgebruik van de Turkse en Marokkaanse jongvolwassen enkel op het NP-domein<sup>5</sup> afwijkt van dat van moedertaalsprekers. Op andere domeinen, zoals bij ontkenning, expletieve constructies of woordvolgorde, zouden deze tweedetaalverwervers de regels van de Nederlandse standaardtaal zonder al te veel uitzondering respecteren. Sommigen zouden zelfs enkele kenmerken van lokale substandaardvariëteiten vertonen, wat een relatief grote bekwaamheid in het Nederlands doet vermoeden (Cornips 2002: 292).

Anders dan in het eerder besproken sociolinguïstisch onderzoek (zie 1.2.), waar de overgeneralisering van het lidwoord *de* werd gezien als een uiting van groepslidmaatschap, wordt de moeilijke verwerving van het Nederlandse genussysteem vanuit taalverwervingsperspectief opgevat als een taalverwervingsprobleem dat in het merendeel van de gevallen in een staat van fossilisatie terecht komt. In tegenstelling tot eentalige kinderen, die na veel moeilijkheden het Nederlandse lidwoordensysteem rond 6-jarige leeftijd verwerven, zouden tweetalige kinderen van Turkse en Marokkaanse afkomst al snel stuiten op de grenzen van hun verwervingscapaciteiten ten koste van het onzijdige lidwoord *het* (Cornips, Van der Hoek & Verwer 2006: 5; Cornips 2008: 109).

Dat fossilisatieproces blijkt niet alleen op te treden bij tweetalige kinderen uit etnische minderheids-groepen, maar ook bij tweetalige (Engels-Nederlandse) kinderen uit families van de middenklasse (Cornips 2008: 113). Unsworth (2006: 451) stelde tijdens haar onderzoek immers vast dat haar Engels-Nederlandse proefpersonen (met een gemiddelde leeftijd van 10,5 jaar) bij onzijdige substantieven het lidwoord *de* in 74,6% van de gevallen overgeneraliseerden. Toch betwijfelt ze of we de fossilisatie van die langdurige overgeneralisering mogen beschouwen als het ultieme kenmerk van NT2-verwervers of dat andere factoren, zoals ondermaatse input, ook een rol kunnen spelen (Unsworth 2006: 457).

Cornips en Hulk (2008) bespraken in dat verband vier factoren die bij tweedetaalverwervers een invloed zouden kunnen uitoefenen op het al dan niet slagen in het aanleren van genus-onderscheiding bij definiëte lidwoorden: leeftijd van *onset*<sup>6</sup>, duur van de blootstelling, kwaliteit van input en invloed van de andere taal. Uit meerdere onderzoeken (Cornips & Hulk 2008; Unsworth et al. 2014) kwamen leeftijd van *onset* en (duur van) input als twee belangrijke factoren naar voren. Brouwer et al. (2008) onderzochten bovendien of er een correlatie bestaat tussen de frequentie van (onzijdige) woorden en de snelheid waarmee het grammaticale lidwoord wordt verworven. Zoals eerder aangehaald (zie 1.1.1.), komt het onzijdige lidwoord *het* immers relatief weinig voor in vergelijking met het algemeen lidwoord *de* (token-frequentie), waardoor het nog minder frequent aanwezig zou zijn in de input van tweetalige kinderen (Cornips & Hulk 2008: 279). Brouwer et al. (2008: 91-93) stelden dan ook vast dat tweetalige kinderen tussen 11 en 13 jaar in staat waren om op een consequente manier de grammaticale lidwoorden van frequente naamwoorden te selecteren, terwijl ze een veel inconsequenter patroon vertoonden bij minder frequente woorden. Dat

---

<sup>5</sup> Zelfs binnen het NP-domein zijn er zekere verschillen. Sommige inflectionele regels worden zonder problemen verworven, terwijl andere veel moeilijker blijken. Zie Pijpops & Van de Velde. (To appear, 2016). *Etnolect speakers and Dutch partitive adjectival inflection: a corpus analysis. Taal & Tongval*, 67(2), 343-371.

<sup>6</sup> Met 'leeftijd van *onset*' wordt de leeftijd bedoeld waarop de taalleerder voor het eerst aan de vreemde taal wordt blootgesteld.



frequentie-effect zou aantonen dat tweetalige verwerving van het Nederlandse genussysteem wel degelijk gevoelig is aan inputfactoren.

Of de tweede taal ten slotte een invloed uitoefent op de verwerving van het Nederlandse grammaticale lidwoord bij tweetalige kinderen, daar zijn taalkundigen het niet over eens. In principe kan verondersteld worden dat de aanwezigheid van een genussysteem in de tweede taal een positieve invloed heeft op de verwerving van het Nederlandse genussysteem, zeker wanneer er een structurele of morfologische overlap aanwezig is (Cornips, Van der Hoek & Verwer 2006: 8). Brugman (2008) toonde zo aan dat tweetalige kinderen van Turkse en Marokkaanse afkomst meer problemen vertonen bij de verwerving van het Nederlandse lidwoord dan de groep Spaans-Nederlandse kinderen die onderzocht werd door Reijers (2007). Cornips, Van der Hoek & Verwer (2006: 8-9) bekwamen daarentegen geen significante resultaten voor eventuele crosslinguïstische effecten. Hoewel de Marokkaanse (Arabische/Berber) kinderen in hun moedertaal wel over een genussysteem voor definiete lidwoorden beschikken en de Turkse kinderen niet, bleek dat geen significante invloed te hebben op de snelheid of de mate waarin ze het Nederlandse lidwoordensysteem verwierven. Volgens Cornips (2008: 112) lijkt het er dus niet op dat de overgeneralisering van het lidwoord *de* zich beperkt tot één bepaalde (Turkse of Marokkaanse) groep met een specifieke culturele, nationale of taalkundige achtergrond.



# Hoofdstuk 2: Onderzoeksvragen en hypotheses

De drie taalkundige luiken die in het voorgaande hoofdstuk uitvoerig werden besproken en waarin onderzoekers vanuit verschillende perspectieven variatie in het Nederlandse genussysteem hebben proberen te verklaren, zullen we in dit tweede hoofdstuk aan de hand van de onderzoeksvragen en hypotheses met elkaar trachten te verbinden. Niet alleen willen we zo een nieuwe kijk bieden op wat aangehaalde studies reeds over variatie in het Nederlandse genussysteem hebben beschreven, maar we hopen ook dat de bevindingen uit de drie theoretische luiken elkaar in zekere mate kunnen aanvullen. Omdat taalcontact een belangrijke rol speelt bij taalverandering (zie 1.2.1.), spitst dit huidige onderzoek zich hoofdzakelijk toe op variatie in genustoekenning in etnolectische variëteiten van het Nederlands. Zo zullen we het *Marokkaans Nederlands* en *Citétaal* nauwkeuriger bestuderen. Nadat we een beschrijving hebben gegeven van de onderzoeksvragen (zie 2.1.) en de hypotheses (zie 2.2.), zullen we overgaan tot een korte omschrijving van het onderzoek (zie 2.3.).

## 2.1. Onderzoeksvragen

De drie onderzoeksvragen waarin we de drie taalkundige luiken met elkaar proberen te verbinden en waarop we doorheen ons onderzoek een antwoord zullen trachten te formuleren, luiden als volgt:

### Onderzoeksvraag 1:

Wijken etnolecten (*Marokkaans Nederlands* en *Citétaal*) af van het taalgebruik van reguliere moedertaalsprekers wat genustoekenning betreft? Zo ja, is er in het *Marokkaans Nederlands* en *Citétaal* dan sprake van een overgeneralisering van het algemeen lidwoord *de*?

### Onderzoeksvraag 2:

Wat is de rol van het hersemantiseringsproces in de genustoekenning van adnominale elementen in etnolecten en autochtone variëteiten?

### Onderzoeksvraag 3

Wat is de rol van taalverwerving? Wordt verandering in genustoekenning in etnolecten voortgestuwd door tweedetaalverwervers die het grammaticale genussysteem van het Nederlands moeilijk onder de knie krijgen? Of spelen andere (sociale) factoren eerder een rol?

## 2.2. Hypotheses

Op basis van de literatuur die in het eerste hoofdstuk werd aangehaald, hebben we een aantal hypotheses geformuleerd, die hieronder per onderzoeksvraag staan opgesomd. In tegenstelling tot de eerste onderzoeksvraag, waarin we ons afvragen of er taalintern verschillen bestaan tussen beide variëteiten, focussen we ons in de twee laatste onderzoeksvragen hoofdzakelijk op verklaringen voor die mogelijke verschillen. Aan de hand van onderzoeksvraag 2 gaan we na of we die verklaring op taaltheoretisch (en taalintern) vlak kunnen vinden, terwijl we in onderzoeksvraag 3 eerder inzetten

op verklaringen uit taalwervingsonderzoek en de sociolinguïstiek. Logischerwijs houdt die werkwijze in dat de eerste onderzoeksvraag positief moet worden beantwoord alvorens er sprake kan zijn van de twee laatste onderzoeksvragen.

### 2.2.1. Onderzoeksvraag 1

*Wijken etnolecten (Marokkaans Nederlands en Citétaal) af van het taalgebruik van reguliere moedertaalsprekers wat genustoekenning betreft? Zo ja, is er in het Marokkaans Nederlands en Citétaal dan sprake van een overgeneralisering van het algemeen lidwoord *de*?*

❖ **Hypothese 1:** Etnolecten wijken - wat genustoekenning betreft - af van het taalgebruik van moedertaalsprekers. We treffen er een overgeneralisering van het algemeen lidwoord *de* aan.

Meerdere studies die in het sociolinguïstisch luik werden besproken (Cornips 2008; Nortier & Dorleijn 2008; Marzo & Ceuleers 2011; Marzo 2015), hebben gewezen op variatie in genustoekenning in etnolectisch taalgebruik. Cornips (2008: 105-107) toonde zo aan dat latere generaties van Turkse en Marokkaanse immigranten, die in Nederland geboren werden en het Nederlands bijgevolg als één van hun moedertalen verwierven, nog steeds een overgeneralisering van het algemeen lidwoord *de* vertonen. Ook Marzo & Ceuleers (2011: 451-452) geven aan dat hoewel het etnolect *Citétaal* strikt gesproken geen duidelijk afgebakende taalvariëteit omvat, mensen de variëteit wel menen te herkennen en de overgeneralisering van het algemeen lidwoord *de* zelfs als één van de belangrijkste kenmerken aanhalen.

Vele taalverwervingsonderzoeken benadrukken bovendien hoe weinig evident het kenmerk grammaticaal geslacht (*de* versus *het*) voor verwervers van het Nederlands is. In tegenstelling tot eentalige kinderen, die na veel moeilijkheden het Nederlandse lidwoordensysteem rond 6-jarige leeftijd verwerven (Van der Velde 2003: 124-129), zouden tweetalige kinderen van Turkse en Marokkaanse afkomst al snel stuiten op de grenzen van hun verwervingscapaciteiten ten koste van het onzijdige lidwoord *het* (Cornips, Van der Hoek & Verwer 2006: 5; Cornips 2008: 109). Zelfs op volwassen leeftijd zouden tweedetaalverwervers het Nederlandse genussysteem nog niet volledig onder de knie hebben (Cornips & De Vogelaer 2009: 5) en bijgevolg in een staat van fossilisatie terecht komen. Aangezien etnolecten kenmerken vertonen die ook in variëteiten van tweedetaalverwervers voorkomen (Van der Sijs 2014), stellen we in deze hypothese dat etnolecten op het vlak van genustoekenning afwijken van het taalgebruik van moedertaalsprekers en dat er een overgeneralisering van het algemeen lidwoord *de* optreedt (Cornips & De Vogelaer 2009; Cornips 2008).

### 2.2.2. Onderzoeksvraag 2

*Wat is de rol van het hersemantiseringsproces in de genustoekenning van adnominale elementen in etnolecten en autochtone variëteiten?*

❖ **Hypothese 1:** Zowel etnolecten als autochtone variëteiten zijn onderhevig aan een hersemantiseringsproces in het adnominale domein, hoewel de hersemantisering in etnolectisch taalgebruik al duidelijker zichtbaar is.

De universele *Agreement Hierarchy* van Corbett (1979), waarin de invloed van woordklassen op genuscongruentie aan de basis ligt van verandering in genussystemen, doet vermoeden dat het hersemantiseringsproces zich ook in het Nederlands vanuit het pronominale domein verder zal uitbreiden naar andere woordklassen, zoals de definiëte lidwoorden. Recent onderzoek (Fernández-Ordóñez 2009) toont immers aan dat die expansie zich in andere talen, zoals in enkele Zuidelijk-Centrale variëteiten van het Italiaans, reeds heeft voltrokken. Omdat contactvariëteiten en etnolecten een grote rol spelen bij taalverandering (Van der Sijs 2014), is het mogelijk dat het lidwoordensysteem van het *Marokkaans Nederlands* en *Citétaal* al veel sterker onderhevig is aan hersemantisering in het adnominale domein dan het Nederlands dat door moedertaalsprekers wordt gesproken. De *Individuation Hierarchy* van Audring (2006), die werd opgesteld voor pronominale verwijzing, zou volgens die redenering ook in zekere mate van toepassing zijn op het lidwoordensysteem van beide etnolecten. De overgeneralisering van het algemeen lidwoord *de* zou dan voornamelijk optreden bij referenten die het hoogst op de hiërarchie staan, terwijl we bij laaggeïndividualiseerde referenten eerder een overgeneralisering van het onzijdig lidwoord *het* zouden aantreffen.

Onderzoek naar hersemantisering in autochtone variëteiten (Audring 2006; Kraaikamp 2012; Klom & De Vogelaer 2016; Franco 2013) toont vervolgens aan dat genustoekenning in het pronominale domein zowel in het Noord-Nederlands als het Zuid-Nederlands steeds vaker gebeurt op basis van semantische overeenkomsten. Hoewel het bestaan van een aantal Nederlandse zelfstandige naamwoorden met een dubbel genus erop kan wijzen dat ook het adnominale domein aan een zekere semantisering toe is (Klom & De Vogelaer 2016; Semplicini 2012) en het onderzoek van Roodenburg & Hulk (2009) aantoont dat Noord-Nederlandse kinderen tijdens de verwerving van het Nederlandse lidwoord reeds in beperkte mate rekening houden met semantische kenmerken, bewees de studie van Franco (2013) dat variatie in het Zuid-Nederlandse adnominale domein onmogelijk aan hersemantisering kan worden toegeschreven. Om die reden menen we dat variatie in het adnominale domein van autochtone variëteiten in mindere mate wordt voortgestuwd door een hersemantiseringsproces dan in etnolectisch taalgebruik.

❖ **Hypothese 2:** De overgeneralisering van het algemeen lidwoord *de* is een inherent kenmerk van *Citétaal* en *Marokkaans Nederlands*. Er is weinig of geen sprake van een hersemantisering in het lidwoordensysteem van beide etnolecten.

Aangezien taalverwervingsonderzoek (Van der Velde 2003; Cornips & Hulk 2008; Cornips & De Vogelaer 2009) de overgeneralisering van het algemeen lidwoord *de* bij eerste- en tweedetaalverwervers toeschrijft aan de grote moeilijkheden die met de verwerving van het Nederlandse lidwoordensysteem gepaard gaan en tweedetaalverwervingsonderzoek (Cornips, Van der Hoek & Verwer 2006) zelfs spreekt van een fossilisatie-effect, zouden we de overgeneralisering van het algemeen lidwoord *de* kunnen beschouwen als een inherent kenmerk van beide etnolecten. In deze hypothese komt er geen hersemantisering aan te pas, maar verloopt de keuze voor de lidwoorden eerder willekeurig, met een voorkeur voor het lidwoord *de* wanneer het grammaticale geslacht van de referent niet is gekend. Ook sociolinguïstisch onderzoek (Cornips 2008; Nortier & Dorleijn 2008) waarin het overgeneralisering van het lidwoord *de* wordt gezien als een uitdrukking van groepslidmaatschap, pleit niet per se in het voordeel van een hersemantiseringsproces.

### 2.2.3. Onderzoeksvraag 3

*Wat is de rol van taalverwerving? Wordt verandering in genustoekenning in etnolecten voortgestuwd door tweedetaalverwervers die het grammaticale genussysteem van het Nederlands moeilijk onder de knie krijgen? Of spelen andere (sociale) factoren eerder een rol?*

❖ **Hypothese 1:** Zowel een taalverwervingsprobleem als sociale factoren liggen aan de basis van de overgeneralisering van het commuun lidwoord *de* in etnolecten.

De moeilijkheden die jonge tweedetaalverwervers ervaren bij de verwerving van het Nederlandse lidwoordensysteem, zouden volgens tweedetaalverwervingsonderzoek (Van der Velde 2003; Cornips & Hulk 2008; Cornips & De Vogelaer 2009) aan de basis liggen van de overgeneralisering van het commuun lidwoord *de*. Sociolinguïstisch onderzoek (Cornips 2008) heeft bovendien aangetoond dat die overgeneralisering een cruciale rol blijkt te spelen bij het stereotyperen van 'de ander'. Volgens deze hypothese zouden jongeren die als kind een taalverwervingsprobleem hebben ervaren en het grammaticale lidwoordensysteem van het Nederlands nog steeds niet volledig onder de knie hebben, zich steeds bewuster worden van het feit dat buitenstaanders hen op basis van die overgeneralisering als 'allochtoon' bestempelen. Om die reden zouden ze het commuun lidwoord *de* in bepaalde contexten (bewust) overgeneraliseren om op die manier hun groepslidmaatschap uit te drukken.

### 2.3. Omschrijving van het onderzoek

Om een antwoord te kunnen formuleren op de drie onderzoeksvragen en om na te gaan welke van de hierboven opgesomde hypothesen mogelijk correct blijken, werd een onderzoek uitgevoerd dat een corpusonderzoek en een veldonderzoek combineert. Het corpusonderzoek biedt ons eerst en vooral de mogelijkheid om aan de hand van data uit vier geselecteerde corpora (twee allochtone en twee autochtone) te bepalen of er een verschil bestaat in het genussysteem van etnolecten en autochtone variëteiten en of er mogelijk een hersemantiseringsproces aan de gang is. Aan de hand van een veldonderzoek gaan we daarnaast na of de resultaten uit het corpusonderzoek bevestigd worden en of leeftijd een invloed uitoefent op het gebruik van de standaardvorm van het lidwoord.

Die multimethodologische onderzoeksopzet maakt het mogelijk om inzichten te verwerven die we wellicht aan de hand van één onderzoeksmethode niet verkregen zouden hebben en biedt ons bijgevolg ook de gelegenheid om de twee uitgekozen etnolectische taalvariëteiten in verschillende onderzoekssettings te bestuderen. Zo zullen de taaldata uit het corpusonderzoek afkomstig zijn uit gesprekken die door allochtone jongeren in informele omstandigheden (op chatkanalen of in spontane interacties onder vrienden) werden gevoerd, terwijl we de data uit het veldonderzoek aan de hand van een formele taaltest verzamelden. Het werken met data van verschillende leeftijdsgroepen maakt het bovendien mogelijk om een iets genuanceerder beeld te schetsen van de huidige situatie.

Hoewel het corpusonderzoek grotendeels betrekking heeft op de eerste twee onderzoeksvragen en het veldonderzoek hoofdzakelijk focust op de laatste onderzoeksvraag, zullen beide onderzoeken de andere onderzoeksonderwerpen ook in beperkte mate behandelen en elkaar op die manier

aanvullen. In de twee volgende hoofdstukken zullen we de onderzoeksmethode, de analyse en resultaten van het corpusonderzoek (zie hoofdstuk 3) en het veldonderzoek (zie hoofdstuk 4) gedetailleerder bespreken.





# Hoofdstuk 3: Corpusonderzoek

Om enerzijds na te gaan of het lidwoordensysteem van etnolectische taalvariëteiten (*Marokkaans Nederlands* en *Citétaal*) door een overgeneralisering van het gemeen lidwoord *de* afwijkt van het lidwoordensysteem van reguliere moedertaalsprekers (onderzoeksvraag 1) en om anderzijds te onderzoeken of beide variëteiten onderhevig zijn aan een hersemantiseringsproces in het adnominale domein (onderzoeksvraag 2), werd een corpusonderzoek uitgevoerd waarin zowel taaldata van autochtone als allochtone sprekers worden geanalyseerd. In dit derde hoofdstuk zullen we eerst een beschrijving geven van de onderzoeksmethode (zie 3.1.) en zullen we daarna overgaan tot de bespreking van de data-analyse en onderzoeksresultaten (zie 3.2.). De belangrijkste bevindingen zullen ten slotte in een concluderende paragraaf worden samengevat (zie 3.3).

## 3.1. Methode

### 3.1.1. Selectie van de corpora

Voor het corpusonderzoek werd gebruik gemaakt van drie corpora: het *Moroccorp*, het *Citétaalcorpus* en het *Condiv*. Na een korte beschrijving van de geselecteerde corpora bespreken we enkele methodologische afwegingen die met de selectie van de data gepaard gingen.

Zoals de naam al doet vermoeden, bevat het *Moroccorp* schriftelijk communicatiemateriaal van Marokkaans-Nederlandse taalgebruikers. Meer bepaald gaat het om een Nederlands chattaalcorpus van ruim 10.000.000 woorden dat in de zomer van 2012 werd samengesteld op basis van logs van twee chatkanalen: *#maroc* (*irc.marocchat.net*) en *#maroc.nl* (*irc.scarynet.org*). Hoewel het Nederlands de voertaal is, richten de chatkanalen zich duidelijk op Nederlanders met een Marokkaanse achtergrond, van wie het taalgebruik in de literatuur beschouwd wordt als de etnolectische variëteit *Marokkaans Nederlands* of *Moroccan flavored Dutch* (Ruetten & Van de Velde 2013: 456-457).

Het *Citétaalcorpus* bevat daarentegen transcripties van mondelinge, spontane interacties tussen 60 Limburgse jongeren van allochtone en autochtone afkomst, opgenomen in de periode van 2013-2014. Om de validiteit van de data te verhogen werd aan de deelnemende jongeren gevraagd om in drie verschillende interactiecontexten (thuis/school/vrije tijd) zelf opnames te maken. De jongeren waren op het moment van deelname allemaal tussen 17 en 22 jaar oud en waren van Vlaamse, Marokkaanse, Italiaanse of Turkse origine. Alle jongeren waren bovendien in België geboren en hadden het Nederlands als moedertaal (naast eventuele andere talen). Aangezien we veronderstellen dat het hierboven vermelde *Moroccorp* hoofdzakelijk etnolectisch taalgebruik van Marokkaanse Nederlanders bevat, hebben we uit het *Citétaalcorpus* enkel de interacties van jongeren van allochtone afkomst geselecteerd.

Ten slotte werd het *Condiv* gebruikt als controlecorpus. Dat corpus bevat uitsluitend geschreven taalmateriaal en is in totaal 47.397.581 woorden groot. Naast een opsplitsing tussen Nederlands en Vlaams materiaal, wordt er in het *Condiv* ook een onderscheid gemaakt tussen *Krantentaal* en *Internettaal*. Het is van het subcorpus met *Internettaal* dat we het *IRC*-deelcorpus in ons onderzoek

betrekken. Net zoals het *Moroccorp*, werd het *Internet Relay Chat*-subcorpus samengesteld op basis van logs van diverse chatkanalen. In totaal bevat het *IRC* 6.965.291 woorden afkomstig van Nederlandse chatkanalen en 8.207.007 woorden afkomstig van Vlaamse chatkanalen. Het chatmateriaal uit het subcorpus met *Internettaal* werd verzameld rond 1998 (Grondelaers et al. 2000).

In de tabel hieronder wordt een overzicht gegeven van de geselecteerde corpora. Alle corpora bevatten taalgebruik dat in informele, minder gecontroleerde contexten voorkomt. Binnen dat informele register maken we nog een verder onderscheid tussen CMC (*computer mediated communication*) en gesproken, mondelinge interacties.

	Vlaanderen		Nederland	
Informeel register	Autochtoon	Allochtoon	Autochtoon	Allochtoon
CMC	<i>Condiv (IRC)</i>		<i>Condiv (IRC)</i>	<i>Moroccorp</i>
Gesproken interacties		<i>Citétaalcorpus</i>		

Tabel 3: Geselecteerde corpora voor het corpusonderzoek

Een eerste belangrijke methodologische afweging die we bij de selectie van de corpora moeten maken, betreft de vraag of we het gesproken taalmateriaal uit het *Citétaalcorpus* mogen vergelijken met het geschreven taalmateriaal uit het Vlaamse deelcorpus van het *IRC*. Aangezien het *IRC* chattaal bevat, menen we dat dat geen groot probleem vormt. Zoals Grondelaers et al. (2000: 358) terecht opmerken, kan de taalvariant die het *IRC* voortbrengt, immers het best beschreven worden als “geschreven gesproken Nederlands”. Chatters zouden tijdens hun schriftelijke conversaties zo goed mogelijk proberen om een normaal spreektempo te benaderen en bijgevolg de noodzaak voor een standaardnorm en correcte spelling achterwege laten. Bovendien blijkt het Vlaamse *IRC*-deelcorpus vol te staan met dialectismen en spreektaalfenomenen, zoals enclisie (*kunde, hebde*), proclisie (*kzijn*) en T-deletie (*wa, da, moe*) (Grondelaers et al. 2000: 358). Om die reden oordelen we dat een vergelijking tussen de twee corpora verantwoord is.

Ten tweede moeten we opmerken dat de chatkanalen die gelogd werden om het *Moroccorp* samen te stellen, in principe ook toegankelijk zijn voor niet-Marokkaanse Nederlanders. Hoewel er kwalitatieve aanwijzingen zijn (zoals gebruikersnamen, referenties naar Marokko en lexicale of grammaticale kenmerken van het *Marokkaans Nederlands*) dat het merendeel van de chatters van Marokkaanse origine is, moeten we er toch rekening mee houden dat er mogelijk ook taaldata van niet-Marokkaanse Nederlanders in het corpus aanwezig zijn. Dat beschouwen we niet als problematisch, omdat een zekere hoeveelheid ruis in een corpus onvermijdelijk is (Ruetten & Van de Velde 2013: 259).

Iets problematischer is de omvang van het *Citétaalcorpus*. In tegenstelling tot het *Moroccorp* en het Nederlandse en Vlaamse deel van het *IRC*, die met respectievelijk 10.000.000 woorden, 6.965.291 woorden en 8.207.007 woorden ongeveer even groot zijn, bevat het allochtone deel van het *Citétaalcorpus* slechts ca. 150.000 tokens. Tendensen inzake de overgeneralisering van het lidwoord *de* of een hersemantisering van het Nederlandse lidwoordensysteem die we in het *Citétaalcorpus* zouden vaststellen, kunnen daarom moeilijk veralgemeend worden. Wel kunnen we aan de hand van

de geselecteerde taaldata een kleinschalig inzicht geven in de huidige taalsituatie en moeten we voornamelijk de noodzaak voor een verdere uitbreiding van het *Citétaalcorpus* benadrukken.

Ondanks de beperktere omvang van het *Citétaalcorpus*, moeten we wel opmerken dat we dankzij de keuze voor gesproken taal materiaal minder te maken krijgen met ruis op de taaldata. Net omdat de deelnemende jongeren zelf opnames maakten, zijn we er zeker van dat ze allen Vlamingen van allochtone afkomst zijn.

Een laatste afweging die we ten slotte moeten vermelden, betreft de ouderdom van de corpora. In tegenstelling tot het *Moroccorp* en het *Citétaalcorpus*, die allebei recent werden samengesteld, dateert het taal materiaal uit het *Condiv* reeds van 1998. Die ouderdom kan een mogelijke beperking vormen voor ons onderzoek, zeker wanneer zou blijken dat de hersemantisering van het Nederlandse genussysteem de laatste jaren een sterke opmars heeft gekend. Aangezien er momenteel geen recenter corpus met Nederlands chatmateriaal van dezelfde omvang als het *Condiv* voorhanden is, hebben we voornamelijk uit praktische overwegingen dit controlecorpus geselecteerd.

Ondanks al deze methodologische afwegingen, menen we dat een vergelijking tussen het *Moroccorp* en het Nederlandse *IRC*-deelcorpus enerzijds en een vergelijking tussen het *Citétaalcorpus* en het Vlaamse *IRC*-deelcorpus anderzijds een goed inzicht kan bieden in de manier waarop sprekers van allochtone afkomst omgaan met het Nederlandse lidwoordensysteem.

### **3.1.2. Selectie van de onderzochte woorden**

Omdat we in de tweede onderzoeksvraag willen nagaan of het lidwoordensysteem van etnolecten en autochtone variëteiten onderhevig is aan een hersemantiseringsproces, gebeurde de selectie van de woorden waarop de drie corpora onderzocht werden, hoofdzakelijk op basis van de semantische klasse waartoe ze behoren. Daarbij werd voornamelijk gewerkt met de semantische klassen uit *Individuation Hierarchy* van Audring (2006), zoals die in het eerste hoofdstuk reeds vermeld werd (zie 1.1.2.) en hieronder nog eens afgebeeld staat (zie voorbeeld 11).

Wanneer andere studies (Klom & De Vogelaer 2016; De Vos 2009) echter relevante suggesties deden voor kleine aanpassingen van die hiërarchie, werden die in het merendeel van de gevallen doorgevoerd. Zo werd er ten eerste voor gekozen om binnen de categorie *other animate* een onderscheid te maken tussen de categorieën *high animate* en *low animate* (Klom & De Vogelaer 2016) en werd ten tweede de klasse *collective nouns* toegevoegd, die het mogelijk maakte om ook menselijke collectiva (zoals *politie* of *leger*) in het onderzoek te betrekken (De Vos 2009). De klasse *specific mass* uit de *Individuation Hierarchy* van Audring (2006) werd daarentegen buiten beschouwing gelaten, aangezien de corpora te weinig woorden van die semantische klasse bleken te bevatten die bovendien met een lidwoord gecombineerd werden. Op die manier kwamen we tot 6 semantische klassen, die in voorbeeld 12 staan opgesomd.

(11) Gender and Individuation Hierarchy in Northern Dutch (Audring 2006: 103)

Human > Other animate > Bounded object/  
Abstract Unbounded abstract  
> Specific mass > Unspecific mass/

(12) Hiërarchie Corpusonderzoek (naar Audring 2006, De Vos 2009, Klom & De Vogelaer 2016)

Human > High animate > Low animate > Collective > Bounded object/  
nouns Abstract Unspecific mass  
> Unbounded abstract/

Binnen die 6 semantische klassen zit bovendien de parameter *telbaar* [+/-] vervat. *Telbaarheid* en meer bepaald de tegenstelling *count/mass* blijken immers een zeer belangrijke rol te spelen binnen het hersemantiseringsproces (Audring 2006; Klom & De Vogelaer 2016). Alle woorden die tot de semantische klassen *human*, *high animate*, *low animate*, *bounded object/abstract* en *collective nouns* behoren zijn telbaar<sup>7</sup>; alle woorden die tot de klasse *unbounded abstract/unspecific mass* behoren, zijn dat niet.

Wanneer de resultaten van ons corpusonderzoek eerder zouden pleiten in het voordeel van de *Individuation Hierarchy in Southern Dutch* van Franco (2012), waarin er enkel een onderscheid wordt gemaakt tussen de klassen *human*, *other animate* en *inanimate* (zie 1.1.3.), kunnen we stellen dat onze opsplitsing in semantische klassen volgens de *Individuation Hierarchy* van Audring (2006) nauwkeuriger was dan eigenlijk noodzakelijk bleek. In dat geval vallen de klassen *human* en *collective nouns* samen in de klasse *human*, vormen de klassen *high animate* en *low animate* samen de klasse *other animate* en behoren de klassen *bounded object/abstract* en *unbounded abstract/unspecific mass* beide tot de klasse *inanimate*.

Per semantische klasse werden vervolgens *de-woorden* en *het-woorden* geselecteerd. Van de 101 woorden die in totaal voor dit corpusonderzoek uitgezocht werden, behoren er 75 tot het algemeen geslacht *de* en 26 tot het onzijdige geslacht *het*. Met die aantallen proberen we tegemoet te komen aan de verdeling die door Cornips & Hulk (2008: 269) werd aangehaald en die op basis van een woordenboekgebaseerde schatting (token-analyse) stelt dat ongeveer 75% van de Nederlandse zelfstandige naamwoorden tot het algemeen geslacht behoort en ongeveer 25% tot het onzijdige geslacht. Binnen de categorie *de-woorden* werd er bovendien een opsplitsing gemaakt tussen mannelijke en vrouwelijke woorden met als doel de verdeling tussen die twee woordgeslachten constant te houden. Zo werden er uiteindelijk 38 mannelijke en 37 vrouwelijke woorden geselecteerd. Het grammaticale geslacht van de geselecteerde woorden werd bepaald aan de hand van *Van Dale Groot Woordenboek van de Nederlandse taal* (Den Boon, Geeraerts & Van der Sijs, editie 2005, 14<sup>de</sup> herziene uitgave).

De selectie van de woorden zelf gebeurde grotendeels op basis van eerder onderzoek naar hersemantisering in het Nederlandse genussysteem. Zo werd er gebruik gemaakt van enkele

---

<sup>7</sup>Om de 'telbaarheid' van de geselecteerde woorden te bepalen, werd een morfosyntactisch criterium gehanteerd: wanneer een enkelvoudig, onbepaald lidwoord of een telwoord het geselecteerde woord vooraf kan gaan, dan wordt het woord als 'telbaar' beschouwd.

geselecteerde woorden uit het veldonderzoek van Klom & De Vogelaer (2016) en dat van De Vos (2009) en werden enkele opvallende bevindingen uit het corpusonderzoek van Audring (2006) in de woordenlijst betrokken<sup>8</sup>. Daarnaast werden ook de frequentielijsten van de drie corpora geraadpleegd om er zeker van te zijn dat de geselecteerde woorden in de drie corpora voorkwamen. Door de beperkte omvang van het *Citétaalcorpus* bleek dat echter moeilijker dan gedacht. Onderstaande woordenlijst bevat daarom de 101 woorden die initieel voor het onderzoek geselecteerd werden. Gedetailleerde lijsten met de woorden waarvan er effectief attestaties in de verschillende corpora werden gevonden, zijn achteraan opgenomen in bijlage (zie Bijlage 1).

	<i>Mannelijk</i>	<i>Vrouwelijk</i>	<i>Onzijdig</i>
Human	bakker, broer, dokter, vader	juffrouw, moeder, tante, zangeres	broertje, kind, meisje, zusje
High animate	aap, beer, ezel, haan, hond, leeuw, olifant, tijger, vis	duif, geit, kat, kip, koe, muis, mus, slang	konijn, paard, schaap, varken
Low animate	kakkerlak, kever, vlinder, worm	mier, mug, slak, spin, wesp	insect <sup>9</sup>
Collective nouns	raad	brandweer, politie	korps , leger
Bounded object/Abstract	appel, auto, bal, fiets, lepel, rok, sleutel, stoel  vorm, winter	bank, broek, deur, kast, tafel, vaas, vork  kleur, lente, ramp, vondst	boek, dak, glas, hemd, huis, raam  geheim, probleem
Unbounded abstract/ Unspecific mass	haat, pijn, adem, ouderdom, wind  chocolade, koffie, pasta, rijst, suiker	liefde, vriendschap, koorts  boter, melk, saus, soep	geluk, verdriet  brood, goud, hout, water, zand

Tabel 4: Geselecteerde woorden corpusonderzoek per woordgeslacht, per semantische klasse

<sup>8</sup> Op basis van Audring (2006) werden de woorden *suiker*, *broertje* en *zusje* geselecteerd.

- Uit het corpusonderzoek van Audring (2006:99) kwam naar voor dat het algemeen woord *suiker* met een onzijdig aanwijzend voornaamwoord gecombineerd wordt:

Ook voor **suiker** [C] geldt de stelregel: zolang u **dit** [N] met mate binnenkrijgt

‘Also for sugar the rule holds: as long as you eat that in moderation’

(periodical *Consumentengids*, January 2005 p. 70)

- Audring (2006:105) maakt een onderscheid tussen *broertje/zusje* en *meisje/mannetje* op basis van verwantschap (*kinship*). Het voornaamwoord van *zusje* en *broertje* zou sneller corresponderen met het biologische geslacht van de referent dan dat van *meisje* en *mannetje*, omdat de spreker de referenten persoonlijk kent.

<sup>9</sup>Aangezien er in het Nederlands nauwelijks onzijdige woorden bestaan die tot de semantische klasse *low animate* behoren, werd het meer algemene woord *insect* geselecteerd. De interpretatie van *insect* als behorend tot de klasse *low animate* (telbaar) wordt gegarandeerd door de context.

Om bovendien tegemoet te komen aan de conclusie van Franco (2013), waaruit bleek dat variatie in het Zuid-Nederlandse lidwoordensysteem grotendeels verklaard kan worden door een afwijkend woordgeslacht in supra-regionale variëteiten, werden diverse dialectwoordenboeken geraadpleegd. Aan de hand van het werk van Pauwels (1938) en het *Woordenboek der Nederlandsche Taal* van De Vries & Te Winkel (1864-1998) werd zo nagaan welke van de 101 geselecteerde woorden een afwijkend geslacht vertoonden in regionale (dialect)variëteiten. Onderstaande tabel bevat een opsomming van die woorden, vergezeld van de woordgeslachten die ze afhankelijk van hun dialectregio aannemen. Enkel de vetgedrukte woorden vertonen ook alternantie in het gebruik van definiëte lidwoorden en aanwijzende voornaamwoorden:

	Mannelijk	Vrouwelijk	Onzijdig
<b>boek</b>	X		X
broek	X	X	
chocolade	X	X	
dokter	X	(X) <sup>10</sup>	
<b>ezel</b>	X		X
fiets	X	X	
haat	X	X	
kat	(X) <sup>11</sup>	X	
<b>kleur</b>		X	X
koffie	X	X	
<b>leger</b>	X		X
lente	X	X	
<b>melk</b>		X	X
<b>ouderdom</b>	X	X	
<b>raam</b>	X	X	X
<b>rijst</b>	X	X	X
<b>suiker</b>	X	X	X
tafel	X	X	
vaas	X	X	
vork	X	X	
vorm	X	X	
<b>water</b>		X	X
<b>winter</b>	X		X
<b>zand</b>	X		X

Tabel 5: Geselecteerde woorden met een afwijkend woordgeslacht in supra-regionale variëteiten

Of het afwijkend woordgeslacht in supra-regionale variëteiten effectief een invloed uitoefent op het lidwoordensysteem in de onderzochte etnolecten, zullen de resultaten van het onderzoek moeten uitwijzen.

<sup>10</sup> Enkel in de betekenis van *dokteres* (De Vries & Te Winkel 1864-1998)

<sup>11</sup> Enkel in de betekenis van *kater* (De Vries & Te Winkel 1864-1998)

### 3.1.3. Doorzoeken van de corpora

Het doorzoeken van de vier geselecteerde corpora gebeurde aan de hand van het programma *AntConc*. De reguliere expressies<sup>12</sup> waren zo opgesteld dat ze telkens een geselecteerd zelfstandig naamwoord bevatten, voorafgegaan door een aanwijzend voornaamwoord of een bepaald lidwoord. Tussen het zelfstandige naamwoord en de determinator konden bovendien tot maximum drie woorden voorkomen, wat het mogelijk maakte om ook constructies met een bijwoord en/of een adjectief tussen de determinator en het zelfstandige naamwoord in de analyse te betrekken. Het maximumaantal van drie woorden werd gekozen om een teveel aan overtollige observaties te vermijden. Ten slotte werd ook rekening gehouden met mogelijke variatie in de schrijfwijze van de geselecteerde zelfstandige naamwoorden, bepaalde lidwoorden en aanwijzende voornaamwoorden. Voor de schrijfwijze van de bepaalde lidwoorden werden de zes volgende variaties in de analyse betrokken: *de, het, et, t, 't*.

De totale dataset (zonder overtollige observaties) bestond uit 12965 attestaties, waarvan zo'n 55% afkomstig was uit het *Morocccorp*. Het Vlaamse en het Nederlandse deel van het *Condiv* leverden beide ongeveer 22% van de attestaties op, terwijl het *Citétaalcorpus* met 103 attestaties sterk ondervertegenwoordigd was (0,79%).

## 3.2. Analyse en resultaten

Voor de statistische analyse werd gebruik gemaakt van het softwarepakket R en de R packages *vcd* (Meyer et al. 2015), *lme4* (Bates et al. 2015), *dplyr* (Wickham & Francois 2015) en *stringr* (Wickham 2015). De resultaten zijn statistisch geanalyseerd aan de hand van Chi-kwadraattests en Fisher Exact Tests voor de significantie tussen twee categorische variabelen en Cramér's V tests voor de effectgrootte. Naast die bivariate testen is er ook een logistische regressieanalyse uitgevoerd. In die techniek wordt een lineaire relatie voorgesteld tussen de onafhankelijke variabelen en de *logit* (*log*-transformatie van de *odds*) van de binaire responsvariabele. In een derde deel van de analyse hanteren we ten slotte de techniek van *random forests* en *conditional inference trees*.

Hieronder zullen we eerst een overzicht geven van alle variabelen die in de analyse betrokken werden (zie 3.2.1.), om daarna over te gaan tot de bespreking van de meest opmerkelijke resultaten uit de statistische analyse. Aangezien de integratie van zelfstandige naamwoorden met een alternerend lidwoordgebruik een ander resultaat opleverde, zullen we de bevindingen exclusief (zie 3.2.2.) en inclusief (zie 3.2.3.) die woorden apart bespreken.

### 3.2.1. Overzicht van de variabelen

Tabel 6 bevat een overzicht van alle variabelen en hun bijhorende niveaus die in de statistische analyse werden opgenomen. Eventuele codes voor niveaus of variabelen die doorheen de analyse in tabellen of grafieken gebruikt zullen worden, staan in de derde kolom opgesomd.

---

<sup>12</sup> Voor de reguliere expressies: zie Bijlage 2

Variabelen	Niveaus	Codes
<b>Responsvariabele</b>		
- Realisatie	<i>standaard, niet_standaard</i>	
<b>Verklarende variabelen</b>		
- Genus	<i>de, het</i>	Voor variabele: De.Het
- Klasse	<i>human, high animate, low animate, collective nouns, bounded object/abstract, unbounded abstract/unspecific mass</i>	Voor niveaus: <i>Hum, HiAn, LoAn, Coll, B.Obj.Abst, U.Abst.mass</i>
- Etnolect	<i>autochtoon Ndl, etnolect</i>	
- Variëteit	<i>Nederland, Vlaanderen</i>	
- TypeDeterminator	<i>aanwijzend voornaamwoord, lidwoord</i>	

Tabel 6: Overzicht van de variabelen, niveaus en eventuele codes (corpusonderzoek)

De responsvariabele *Realisatie* bevat het verschil tussen standaardgebruik of niet-standaardgebruik van het grammaticale lidwoord of aanwijzend voornaamwoord. Die variatie proberen we te verklaren aan de hand van verschillende verklarende variabelen.

*Genus* drukt ten eerste het grammaticale woordgeslacht van de geselecteerde woorden uit: commuun (*de*) of onzijdig (*het*). Op die manier kunnen we nagaan of niet-standaardrealisatie even vaak optreedt bij beide woordgeslachten of dat we slechts bij één van de twee genera een overgeneralisering kunnen waarnemen.

De variabele *Klasse* omvat vervolgens zes niveaus die de verschillende semantische klassen uitdrukken volgens (een kleine aanpassing van) de *Individuation Hierarchy* van Audring (2006). In Tabel 4 was al eerder te zien tot welke semantische klasse de 101 geselecteerde woorden behoren. Wanneer we te maken zouden hebben met een hersemantiseringsproces in het adnominale domein, zou onder andere de variabele *Klasse* een significante invloed moeten uitoefenen op de responsvariabele.

De variabelen *Etnolect* en *Variëteit* werden vervolgens opgesteld op basis van de eigenschappen van de corpora. *Etnolect* bevat zo twee niveaus (*autochtoon Ndl* en *etnolect*), die respectievelijk attestaties uit het Nederlandse en Vlaamse deel van het *Condiv* en attestaties uit het *Morocccorp* en *Citétaalcorpus* omvatten. Op basis van de variabele *Etnolect* kunnen we onderzoeken of niet-standaardrealisatie van de determinator vaker voorkomt in etnolecten of in het taalgebruik van moedertaalsprekers. Binnen de variabele *Variëteit* maken we bovendien een onderscheid tussen *Nederland* (Nederlandse deel van het *Condiv* en *Morocccorp*) en *Vlaanderen* (Vlaamse deel van het *Condiv* en *Citétaalcorpus*) om na te gaan of niet-standaardrealisatie zich in één van beide landen vaker voordoet.

*TypeDeterminator* bevat ten slotte twee verschillende niveaus, die toelaten om een onderscheid te maken tussen de woorden die met een aanwijzend voornaamwoord in de corpora geattesteerd werden en diegene die met een bepaald lidwoord voorkwamen. Mogelijk oefent het type van determinator een invloed uit op standaardrealisatie.



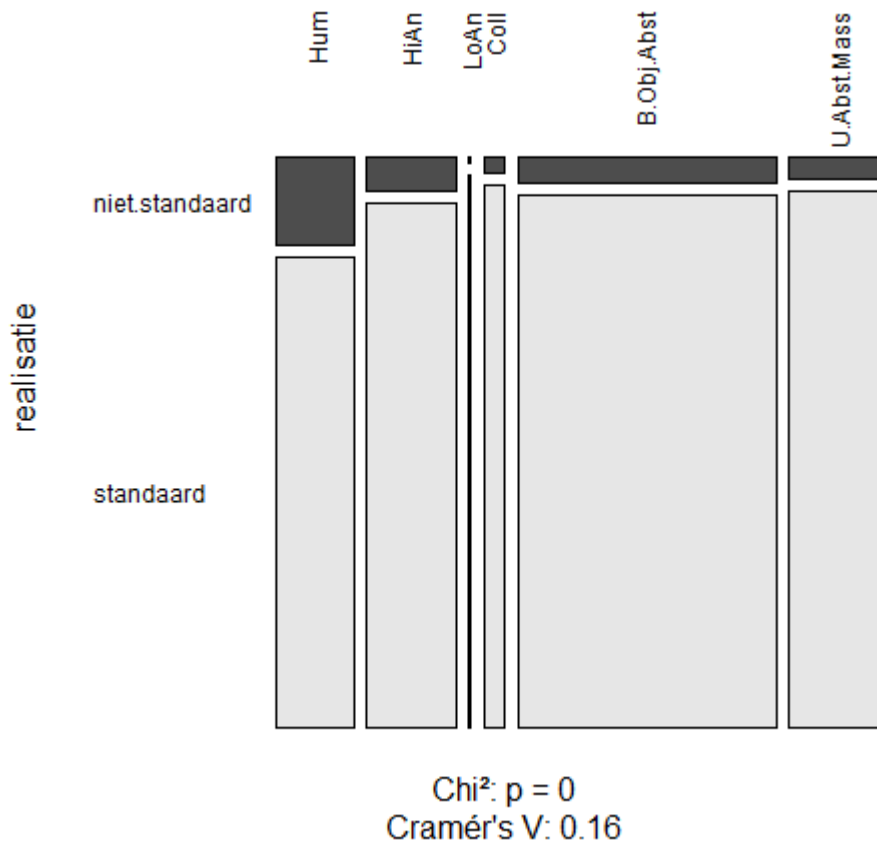
### 3.2.2. Analyse en resultaten exclusief zelfstandige naamwoorden met een alternerend lidwoordgebruik

Om mogelijke dialectinterferentie bij de keuze voor een definitief lidwoord of aanwijzend voornaamwoord onder controle te houden, werd in eerste instantie beslist om de twaalf zelfstandige naamwoorden die volgens het werk van Pauwels (1938) en het *WNT* (De Vries & Te Winkel 1864-1998) in bepaalde dialectregio's lidwoordalternantie vertonen, niet in de dataset op te nemen. Op basis van die gefilterde dataset werden er vervolgens een aantal verkennende plots<sup>13</sup> opgesteld, die grotendeels enkele bevindingen uit de literatuurstudie bevestigden. Ten eerste werd al gauw duidelijk dat de etnolectische corpora significant meer niet-standaardvormen van het lidwoord bevatten dan de corpora met taaldata van moedertaalsprekers ( $\chi^2$ :  $p=0$ , Cramér's V: 0.2). In de Vlaamse corpora kwamen niet-standaardvormen bovendien net iets vaker voor dan in de Nederlandse. Ten tweede bleek ook het grammaticale geslacht van het lidwoord een significante invloed uit te oefenen op standaardrealisatie ( $\chi^2$ :  $p=0$ , Cramér's V: 0.4). Bij grammaticale *het-woorden* werd immers vaker de niet-standaardvorm aangetroffen dan bij grammaticale *de-woorden*. Enerzijds bevestigen die twee resultaten dus de overgeneralisering van het algemeen lidwoord *de* in etnolecten, anderzijds zijn ze te algemeen om met zekerheid iets te zeggen over een mogelijk hersemantiseringsproces.

Figuur 1, die werd opgesteld op basis van taaldata uit de vier corpora, toont aan dat de semantische klasse waartoe een woord behoort, een significante invloed uitoefent op standaardrealisatie ( $\chi^2$ :  $p=0$ , Cramér's V: 0.16). Zo zien we dat woorden uit de semantische klassen *human* en *high animate* significant vaker voorkomen met een niet-standaardvorm van de bijhorende determinator dan woorden uit de vier andere semantische klassen. Ondanks die significante invloed van *Klasse* op de responsvariabele *Realisatie*, kunnen we echter nog steeds niet met zekerheid spreken van een hersemantiseringsproces in het adnominale domein. Onderstaande plot biedt ons immers enkel informatie over de standaardrealisatie van de determinatoren, maar vertelt ons niet welk genus niet-standaardrealisatie in de semantische klassen uitlokt. Om dat te weten te komen, zouden we de interactie tussen *Genus* en *Klasse* nauwkeuriger moeten bestuderen.

---

<sup>13</sup> Zie Bijlage 3



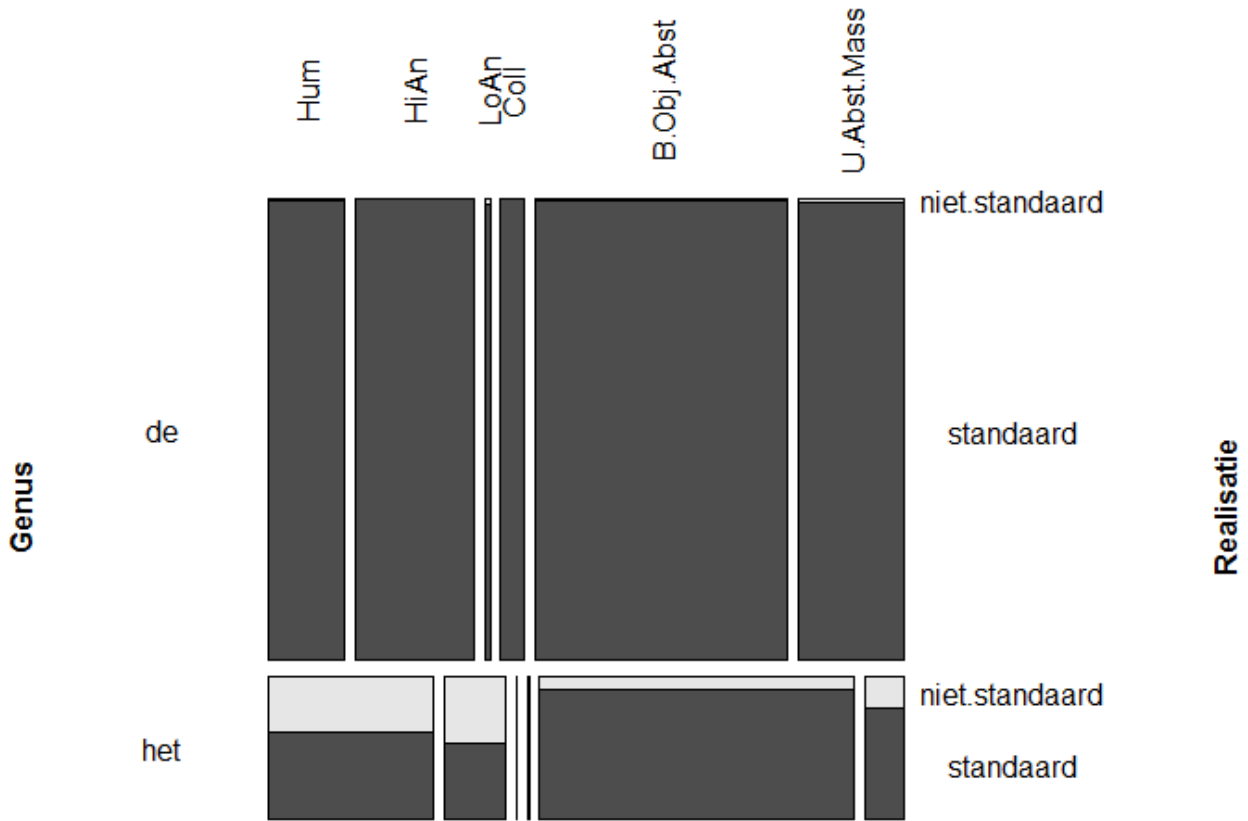
Figuur 1: Standaard- en niet-standaardrealisatie in de verschillende semantische klassen (corpusonderzoek)

De *mosaicplots* op de volgende bladzijden werden opgesteld om na te gaan of er een significante interactie bestond tussen de variabelen *Genus* en *Klasse*, wat een duidelijker inzicht zou moeten bieden in de eventuele aanwezigheid van hersemantisering. Wanneer die hersemantisering zich inderdaad voordoet, zou niet-standaardrealisatie van hoog-geïndividualiseerde referenten voornamelijk moeten voorkomen bij grammaticale *het-woorden* en niet-standaardrealisatie van laag-geïndividualiseerde woorden bij grammaticale *de-woorden*. Zowel de *mosaicplot* van alle taalvariëteiten als dat van de etnolecten (zie Figuur 2 en Figuur 3) tonen echter aan dat het effect van niet-standaardrealisatie voornamelijk aanwezig is bij *het*, wat enkel de overgeneralisering van *de* bevestigt en dus niet wijst op een hersemantisering van *het*. Het kleine effect van de niet-standaardvorm binnen *de-woorden* van de laagst-geïndividualiseerde semantische klassen, dat uit beide *mosaicplots* nog viel af te leiden, bleek bovendien niet significant, zelfs niet wanneer we enkel naar het marginale effect keken (Fisher Exact test:  $p = 0.1035$ )<sup>14</sup>.

<sup>14</sup> De Fisher Exact Test werd uitgevoerd, omdat een Chi-kwadraattest onbetrouwbaar was door te dunbevolkte cellen voor de verwachte frequenties.

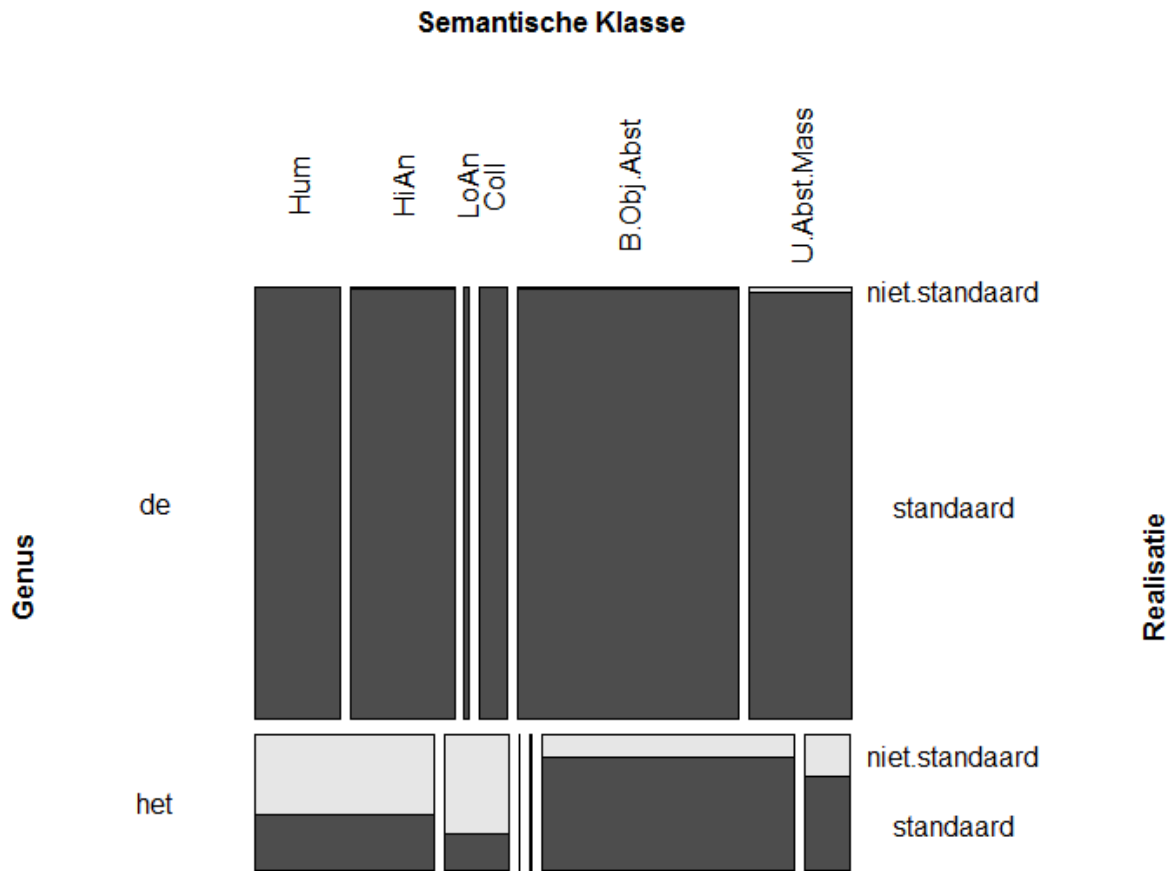
# Alle variëteiten

## Semantische Klasse



Figuur 2: Standaard- en niet-standaardrealisatie van grammaticale *de-* en *het-woorden* in de verschillende semantische klassen over alle taalvariëteiten heen (corpusonderzoek)

# Etnolect

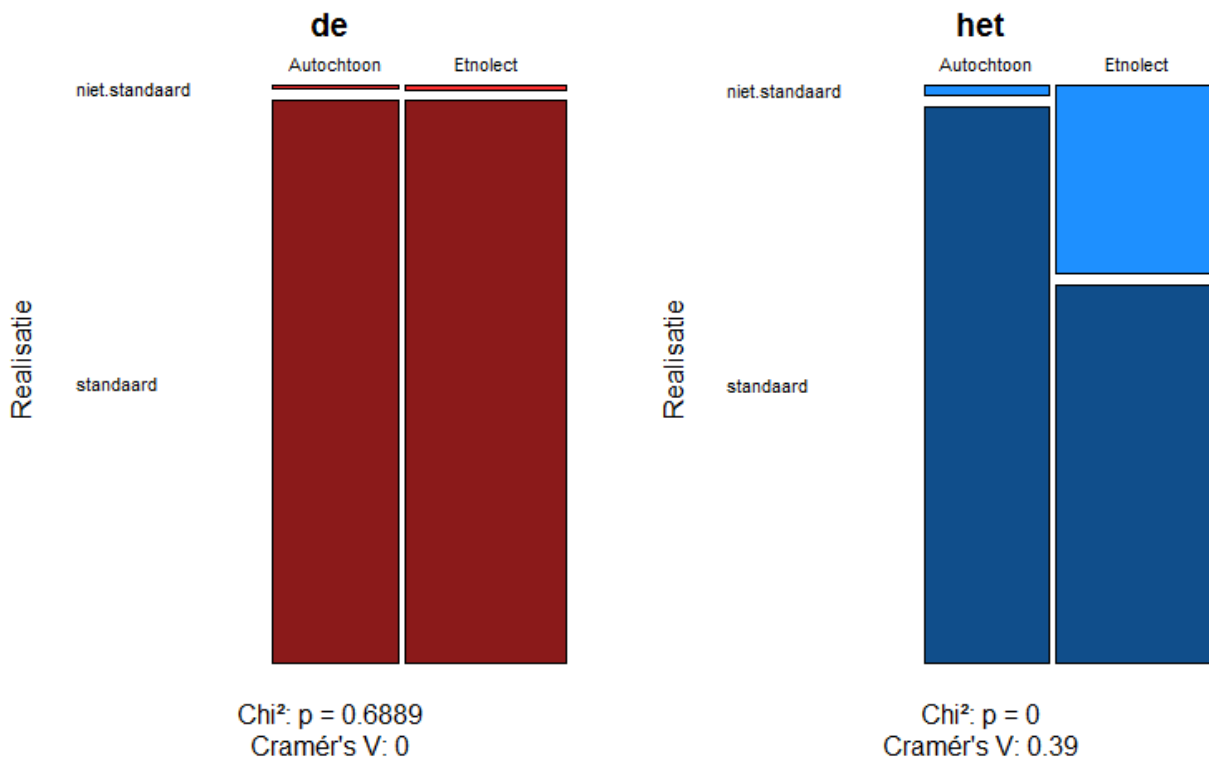


Figuur 3: Standaard- en niet-standaardrealisatie van grammaticale *de*- en *het*-woorden in de verschillende semantische klassen in *Marokkaans Nederlands* en *Citétaal* (corpusonderzoek)

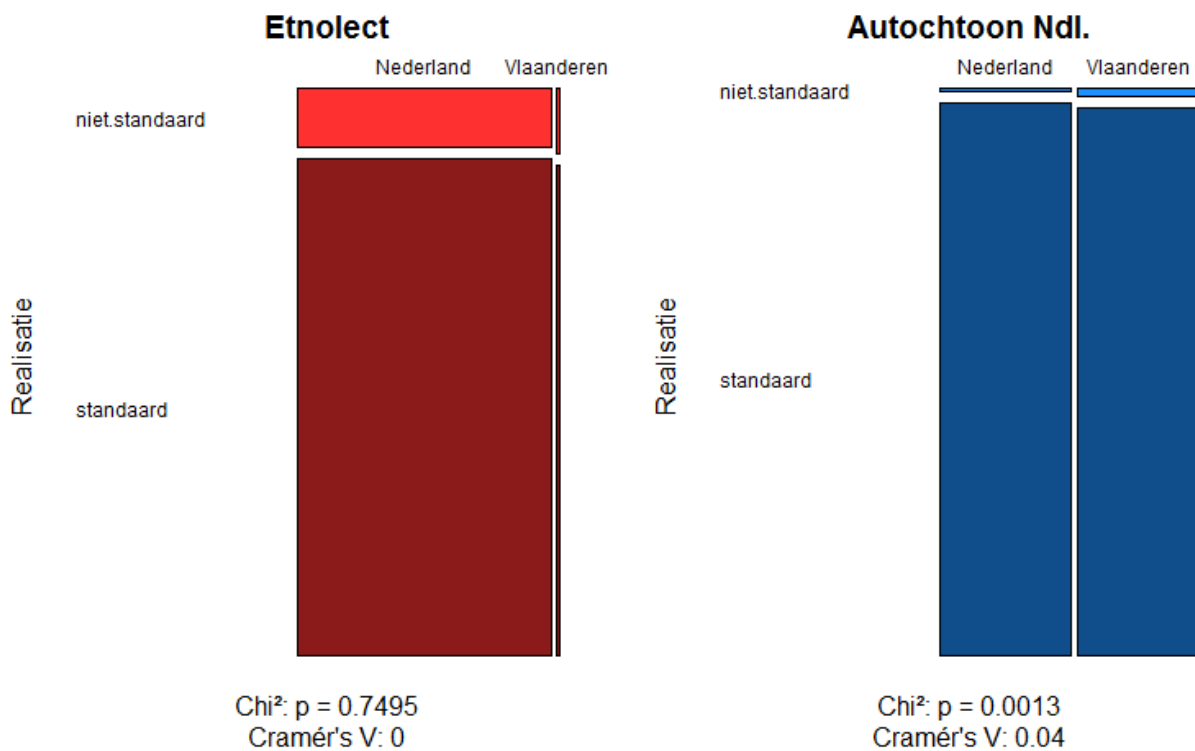
### 3.2.3. Analyse en resultaten exclusief 'raam'

De vraag rijst echter of de integratie van de twaalf woorden met een alternerend lidwoordgebruik andere resultaten zou opleveren. Wanneer we in meer detail nagingen welke zelfstandige naamwoorden met een alternerend lidwoordgebruik in de autochtone taalvariëteiten van de corpora voorkwamen met de niet-standaardvorm van hun bijbehorende lidwoord, bleek dat slechts om een vijftal woorden te gaan: *boek*, *kleur*, *raam*, *winter* en *zand*. Bovendien bleek enkel het woord *raam* in het Nederlandse deel van het *Condiv* met een niet-standaardvorm gecombineerd te worden. Aangezien we er door de grootte van het *Moroccorp* vanuit kunnen gaan dat het grootste deel van de geobserveerde effecten in etnolecten zijn toe te schrijven aan het *Marokkaans Nederlands* (Nederlanders), werd er besloten om van nu af aan enkel het woord *raam* uit de dataset te halen. De kans dat afwijkingen op de woorden uit de lijst van Pauwels en het WNT aan dialectinterferentie te wijten zijn, is immers miniem, wanneer die woorden ook in het Nederlandse deel van het *Condiv* niet met een niet-standaardtalige determinator voorkomen. Op basis van die gefilterde dataset werden nieuwe plots opgesteld, waarvan er vier hieronder staan afgebeeld. Opnieuw bevestigden die resultaten de overgeneralisering van het algemeen lidwoord *de* bij etnolecten ( $\chi^2: p=0$ , Cramér's V:

0.39), maar bleek wel dat de niet-standaardvorm van het lidwoord *de* niet significant vaker optreedt bij etnolecten dan bij de autochtone variëteiten ( $\chi^2: p=0.6889$ , Cramér's V: 0).



Figuur 4: Standaard- en niet-standaardrealisatie van grammaticale *de*- en *het*-woorden in etnolect en autochtoon taalgebruik (corpusonderzoek)



Figuur 5: Standaard- en niet-standaardrealisatie in etnolect en autochtoon taalgebruik per variëteit (corpusonderzoek)

Onderstaande tabel, waarin het percentage standaardrealisatie van alle geselecteerde woorden staat afgebeeld, laat ook uitschijnen dat de woorden waarbij het minst vaak een standaardvorm van het bijbehorende lidwoord wordt gebruikt (20%-60%), allemaal onzijdige *het-woorden* zijn. Bovendien behoren die woorden ofwel tot de twee hoogst-geïndividualiseerde klassen *human* of *high animate* (*schaap, konijn, meisje, zusje*) ofwel tot de laagst-geïndividualiseerde klasse *unbounded abstract/unspecific mass* (*brood*).

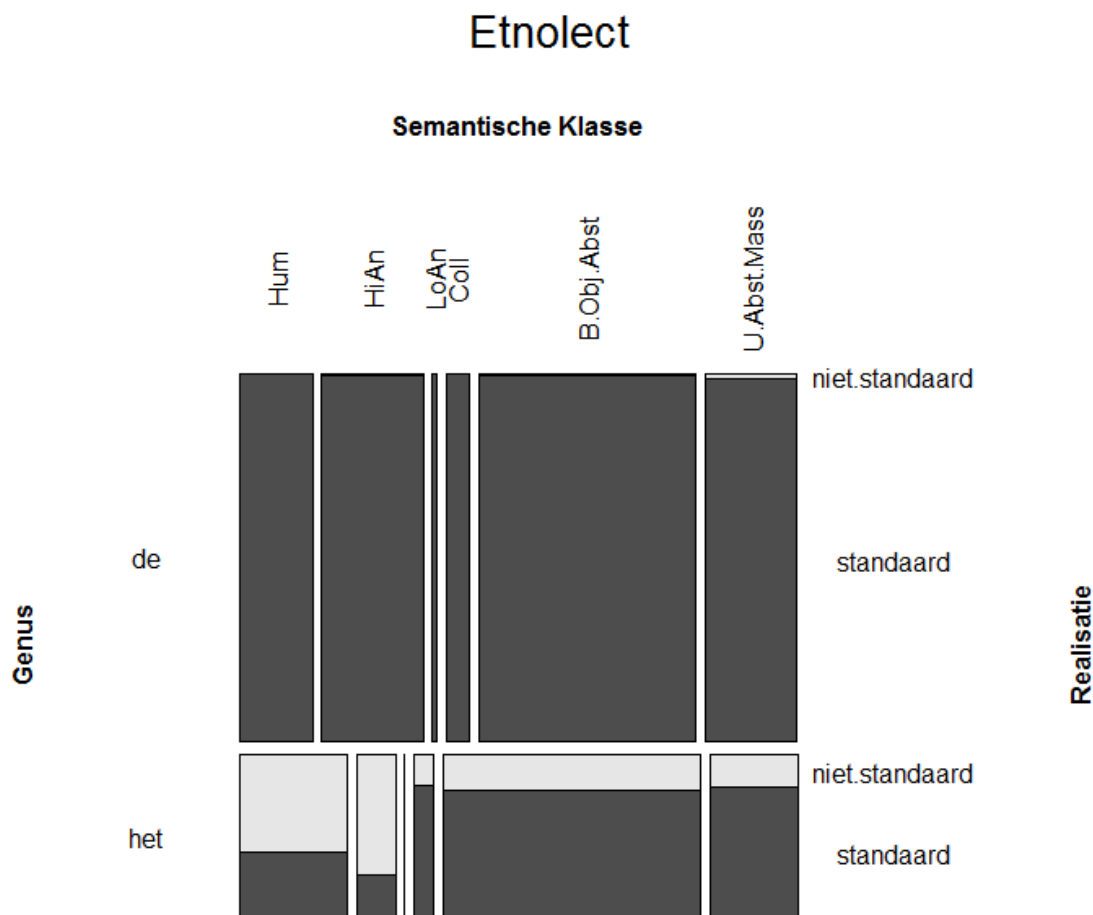
Percentage standaardrealisatie	
100%	aap, adem, appel, bakker, bal, bank, beer, brandweer, chocolade, duif, fiets, geit, haan, hemd, hond, insect, juffrouw, kakkerlak, kast, kever, koe, koffie, koorts, leeuw, lente, lepel, melk, mier, moeder, mug, muis, mus, olifant, ouderdom, pasta, pijn, politie, raad, ramp, rijst, rok, slak, slang, soep, spin, stoel, tante, tijger, vader, vis, vlinder, vondst, vriendschap, worm, zangeres
80% - 100%	vaas, vork, raam, kind, boek, glas, goud, wesp, huis, geluk, hout, verdriet, suiker, dak, leger, boter, probleem, water, kleur, vorm, haat, geheim, sleutel, kip, broer, saus, winter, ezel, broek, liefde, wind, tafel, dokter, kat, auto, deur
60% - 80%	varken, korps, paard, broertje, zand
40% - 60%	konijn, brood, meisje, zusje
20% - 40%	schaap
<20%	/

Tabel 7: Percentage standaardrealisatie van de geselecteerde woorden (corpusonderzoek)

Als we in de nieuwe dataset (dat wil zeggen: de dataset inclusief alle woorden die een variabel genus vertonen in verschillende dialecten, behalve *raam*) nagaan of er in de etnolectische taalvariëteiten een significante interactie optreedt tussen *Genus* en *Klasse*, blijkt zich grotendeels hetzelfde patroon voor te doen als in de eerste dataset: enerzijds zien we dat het effect van niet-standaardrealisatie zich hoofdzakelijk voordoet bij het lidwoord *het* in de hoogst-geïndividualiseerde semantische klassen, anderzijds blijkt het effect van de niet-standaardvorm binnen de *de-woorden* zeer miniem. De Fischer Exact Test wees deze keer echt wel op een significante invloed van semantische klasse op standaardrealisatie binnen de *de-woorden* (bovenste helft van de *mosaicplot* op de volgende pagina;  $p=0.02074$ ). In de laagst-geïndividualiseerde klasse *unbounded abstract/unspecific mass* kunnen we met andere woorden in de etnolectische taalvariëteiten een lichte tendens tot hersemantisering van *het* vaststellen.

De vraag rijst daarom hoe we dat divergente resultaat (geen hersemantisering als de we woorden uit de lijst van Pauwels en het WNT niet meetellen, wel hersemantisering als we ze erin laten) kunnen verklaren. In principe kunnen we hier twee kanten uit: ofwel hebben we te maken met dialectinterferentie en is het geobserveerde hersemantiseringseffect slechts een fata morgana, ofwel doet de hersemantisering bij de *de-woorden* zich net voor bij die woorden die deel uitmaken van de lijst van Pauwels en het WNT. Het probleem met de eerste verklaring is echter dat hetzelfde effect dan ook waarneembaar zou moeten zijn in het *Condiv*-corpus, wat niet het geval is (op het woord

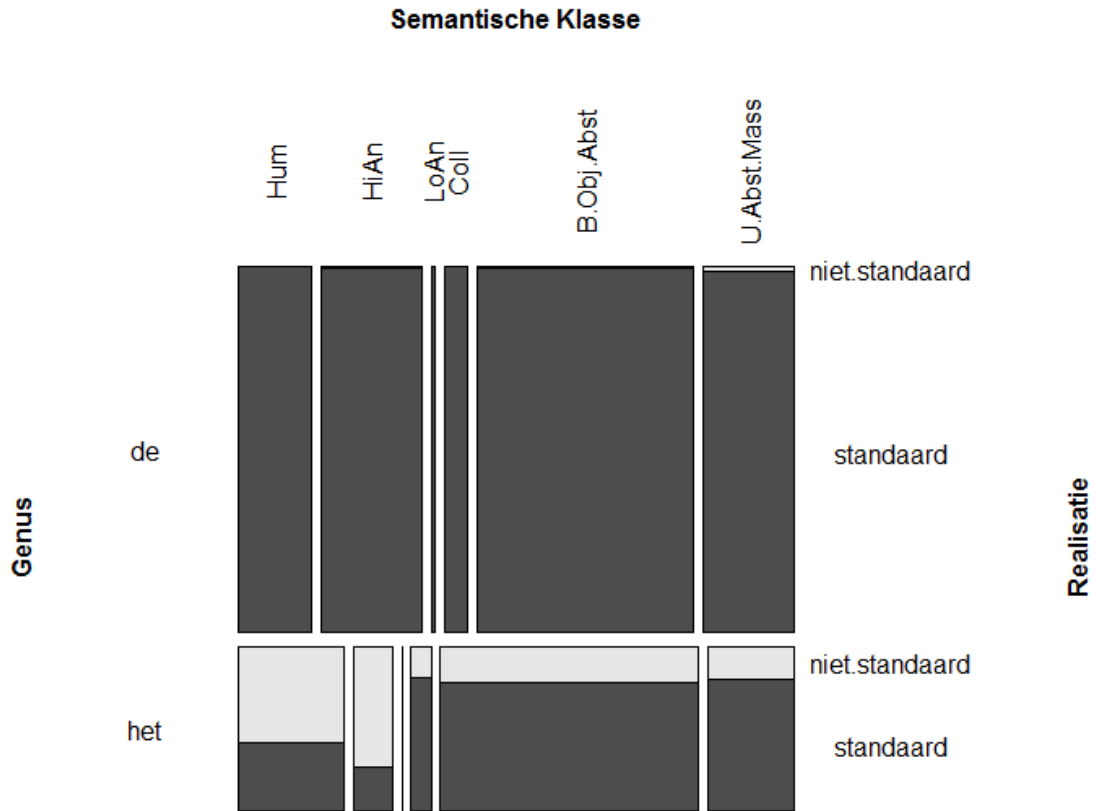
*raam* na, maar dat hebben we uit beide datasets gelaten). Om die reden is de tweede verklaring plausibeler. Het is allicht geen toeval dat uitgerekend die woorden op de lijst van Pauwels en het WNT het sterkst bijdragen aan het semantiserings-effect. Het genusgedrag van die woorden in onze etnolectische data is dan geen geval van dialectinterferentie, maar een hersemantisering die zich min of meer parallel voordoet in sommige dialecten en in de etnolecten. Door die woorden uit onze dataset te filteren kregen we met andere woorden hoogstwaarschijnlijk een onderschatting van dat geringe hersemantiserings-effect in etnolecten te zien.



Figuur 6: Standaard- en niet-standaardrealisatie van grammaticale *de*- en *het*-woorden in de verschillende semantische klassen in etnolecten (corpusonderzoek)

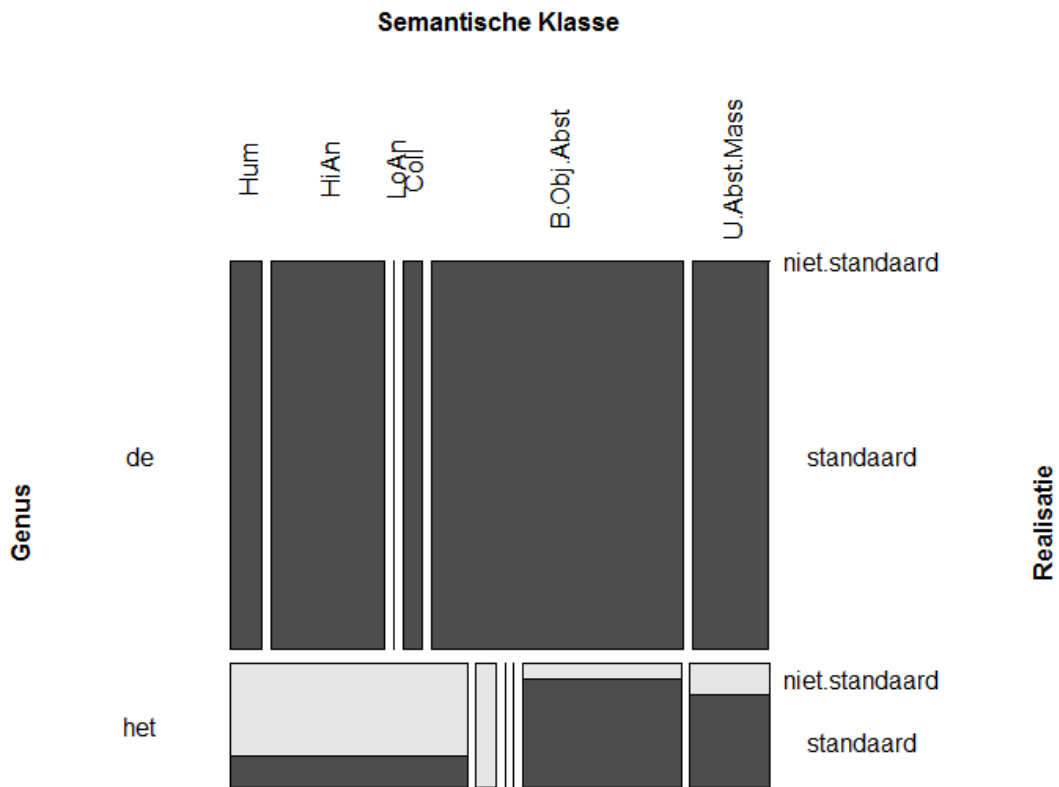
Wanneer we ten slotte naar de aparte *mosaicplots* van beide etnolecten kijken (zie Figuur 7 en 8), zien we dat het zwakke signaal tot hersemantisering van *het* in de laagst-geïndividualiseerde semantische klassen enkel te vinden is in taaldata uit het *Morocccorp*. In het *Citétaalcorpus* zou daarentegen enkel een overgeneralisering van *de* optreden, die zich in de hoogst-geïndividualiseerde klassen *human* en *high animate* wel sterker manifesteert dan in het *Morocccorp*. We moeten daarbij echter voor ogen houden dat we veel meer taal materiaal uit het *Citétaalcorpus* nodig hebben om die bevindingen te bevestigen. Door de geringe omvang van het *Citétaalcorpus* krijgen we mogelijk een vertekend beeld.

# Moroccorp



Figuur 7: Standaard- en niet-standaardrealisatie van grammaticale *de-* en *het-*woorden in de verschillende semantische klassen in het *Moroccorp* (corpusonderzoek)

# Citétaal



Figuur 8: Standaard- en niet-standaardrealisatie van grammaticale *de-* en *het-*woorden in de verschillende semantische klassen in het *Citétaalcorpus* (corpusonderzoek)



### 3.2.4. Logistische regressieanalyse

In een volgende stap van de analyse werd een regressiemodel opgesteld. De opname van verklarende variabelen is gebaseerd op een *bidirectionele stapsgewijze variabelenselectie-procedure* met het Akaike Informatie Criterium als scheidsrechter. De analyse laat ons toe om de sterkte van de impact van de verschillende variabelen ten opzichte van elkaar af te wegen en om te beoordelen welke variabelen een significante impact uitoefenen.

De regressie-output wordt samengevat in Tabel 8. Daaruit blijkt dat binnen de variabele *Klasse* de niveaus *high animate* en *low animate* geen significant verschil in standaardrealisatie van de determinator opleveren ten opzichte van het referentieniveau *human*. Anders gezegd: de scheidingslijn ligt na de bezielde entiteiten. De niet-bezielde referenten laten allemaal hogere *odds* optekenen voor de niet-standaardrealisatie van de determinator, als we controleren voor alle andere variabelen van het model.

De partiële effecten van de andere variabelen bevestigen allemaal de marginale bivariate analyses: afwijking van de standaardtaalversie van de determinator is meer waarschijnlijk bij *het-woorden* en in de etnolecten. We zien verder ook dat het aanwijzend voornaamwoord gemakkelijker afwijking toelaat van het standaardtalige genus dan het lidwoord. De sterkste effecten zijn die voor het genus (*de* vs. *het*) en voor het etnolect.

Coëfficiënts	Estimate	Pr(> z )	Sig.	95% confidence interval	
				2.5%	97.5%
(Intercept)	6.16	<2e-16	***	5.7363031	6.6024972
De.hethet	-4.27	<2e-16	***	-4.5878209	-3.9742351
Klasse.KHiAn	-0.25	0.1039		-0.5579051	0.0517651
Klasse.KLoAn	0.83	0.4893		-1.0635911	3.9263066
Klasse.KColl	1.08	0.0002	***	0.5394957	1.6868159
Klasse.KB.Obj.Abst	0.88	1.88e-14	***	0.6557399	1.1072216
Klasse.KU.Abst.Mass	1.05	2.5782e-12	***	0.7615823	1.3524918
TypeDeterminatorLidwoord	1.58	<2e-16	***	1.3788392	1.7748878
EtnolectEtnolect	-2.89	<2e-16	***	-3.1986791	-2.5948645

Tabel 8: Overzicht output regressiemodel (corpusonderzoek)

### 3.2.5. 'Random forest' en 'Conditional inference tree'

In de laatste stap van de analyse maken we ten slotte gebruik van de statistische toepassing *random forests*, die ontwikkeld werd door Breiman (2001) en die toegankelijk is via het softwarepakket R (Strobl et al. 2007,2008; Hothorn et al. 2006). Aangezien we tijdens de statistische analyse van het veldonderzoek (zie 4.2.2.) dezelfde analysemethode hanteren, zullen we in de finale discussie (zie Hoofdstuk 5) de resultaten van beide onderzoeken gemakkelijk met elkaar kunnen vergelijken.

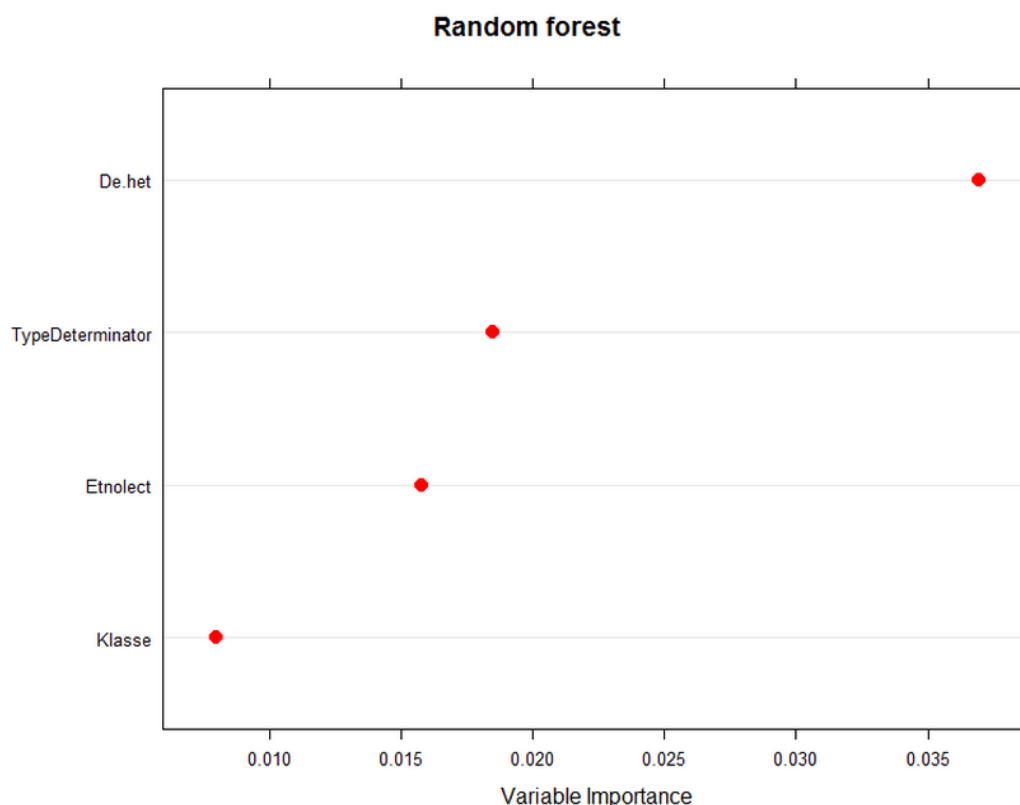
Net zoals logistische modellen, proberen *random forests* aan de hand van een set onafhankelijke variabelen te voorspellen welke van de alternatieven (*standaard, niet\_standaard*) van de responsvariabele het meest waarschijnlijk is (Tagliamonte & Baayen 2012: 22). In tegenstelling tot logistische modellen, worden de data bij *random forests* echter geleidelijk aan doorlopen en wordt er

aan de hand van *trial and error* vastgesteld of een bepaalde variabele een bruikbare predictor is of niet.

Aan de hand van het basialgoritme van *random forests* worden *conditional inference trees* opgesteld, waarin schattingen over de waarschijnlijkheid van de waarden van de responsvariabele *realisatie (standaard, niet\_standaard)* worden berekend op basis van een reeks binaire vragen over de waarden van de onafhankelijke variabelen (Tagliamonte & Baayen 2012: 22). Op basis van die binaire vragen worden in het sample binaire opdelingen (*splits*) gemaakt, waardoor er verschillende subsets worden gecreëerd. Wanneer er meerdere invloedrijke predictoren zijn, wordt eerst de sterkste predictor in rekening gebracht.

Op basis van complexere algoritmes zorgt de techniek van de *random forests* er vervolgens voor dat de gemaakte splits keer op keer opnieuw gesampled worden en dat de voorspellende kracht van de predictoren meerdere malen wordt getest. Dat maakt dat de techniek van de *random forests* iets betrouwbaarder is dan die van de *conditional inference tree*.

Figuur 9 toont de *random forest* waarin de relatieve invloed van de predictoren *De.het (Genus)*, *TypeDeterminator*, *Etnolect* en *Klasse* wordt beschreven aan de hand van *conditional permutation-based variable importance* (Tagliamonte & Baayen 2012: 24). Alle predictoren blijken een significante invloed uit te oefenen op de responsvariabele *Realisatie (standaard, niet\_standaard)*.



Figuur 9: Conditional permutation variable importance voor de random forest met de predictoren *De.het*, *TypeDeterminator*, *Etnolect* en *Klasse* (corpusonderzoek)

Figuur 9 bevestigt ten eerste wat de logistische regressieanalyse eerder aantoonde: de variabele *Genus (De.het)* blijkt duidelijk de meest invloedrijke predictor. Het grammaticale woordgeslacht – of meer bepaald het grammaticale lidwoord (*de* of *het*) van de referent – oefent met andere woorden een sterke invloed uit op de keuze voor een standaardvorm of een niet-standaardvorm. Omdat de overgeneralisering van het algemeen lidwoord *de* (niet-standaardrealisatie bij *het*-woorden) doorheen de voorgaande analyses statistisch bevestigd werd, weten we dat het grammaticale lidwoord *het* een sterk effect uitoefent op niet-standaardrealisatie en dat er bij grammaticale *de-woorden* vaker geopteerd wordt voor de standaardvorm.

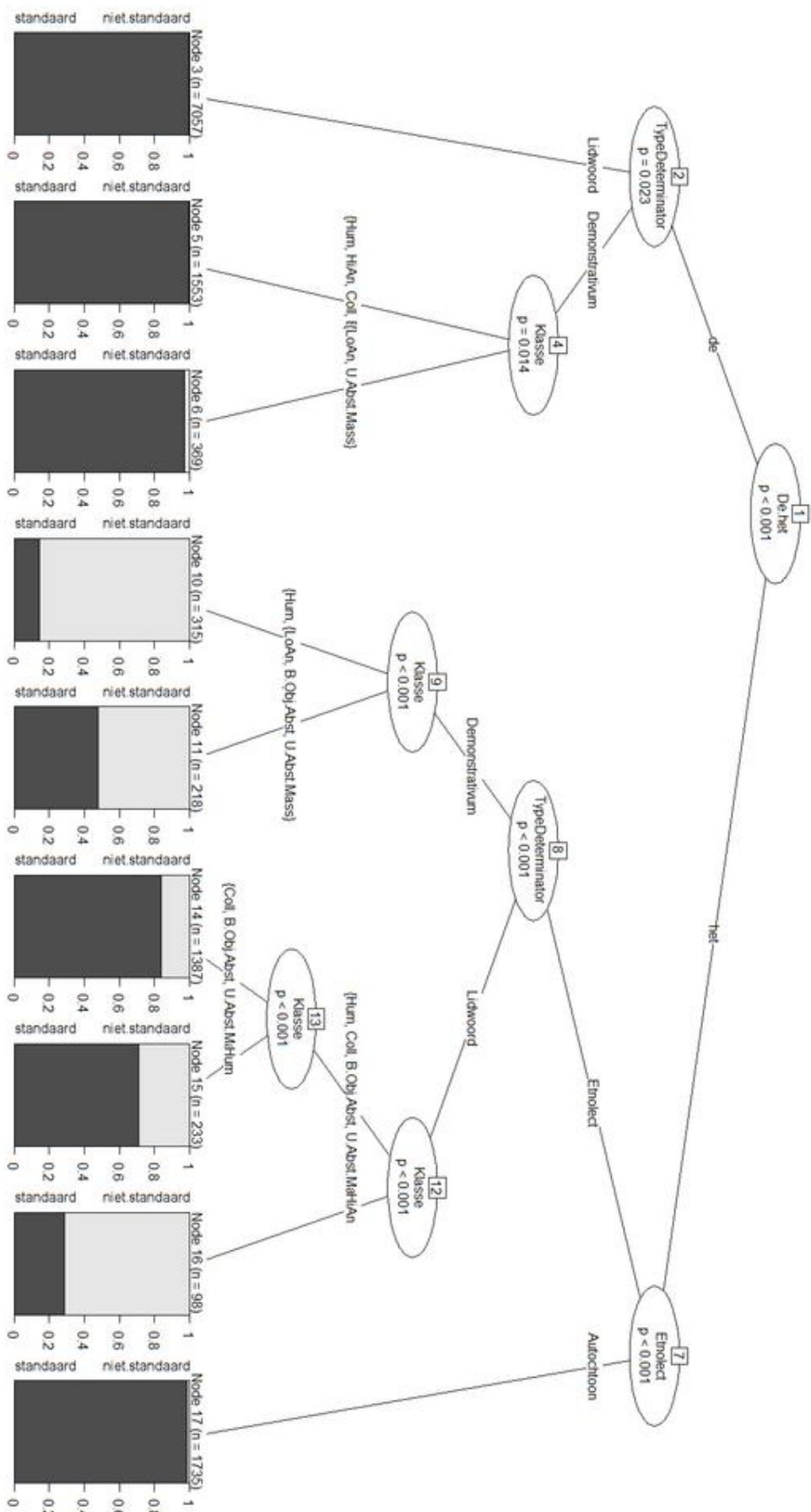
De variabele die daarna de belangrijkste invloed uitoefent op de responsvariabele, is de predictor *Typedeterminator*. Tijdens de logistische regressieanalyse kwam eerder aan het licht dat aanwijzende voornaamwoorden gemakkelijker afwijking van de standaardvorm toelaten dan definitieve lidwoorden. Het belang van de variabele *Typedeterminator* wordt hier bevestigd.

Uit de regressieanalyse bleek echter dat niet *Typedeterminator* maar *Etnolect* een sterke predictor was. In de *random forest* wordt de variabele *Etnolect* slechts als derde sterkste predictor aangeduid. De minst sterke predictor blijkt volgens deze analyse *Klasse*.

Nu we weten welke predictoren een significante invloed uitoefenen op de responsvariabele *Realisatie (standaard, niet\_standard)*, onderzoeken we in een volgende stap van de analyse hoe de betrokken predictoren samenwerken en elkaar beïnvloeden. Daarvoor bekijken we de *conditional inference tree* van de dataset (zie Figuur 10). Hoewel die techniek iets minder betrouwbaar is dan de *random forest*, komt de complexe interactie van de predictoren op de responsvariabele in de *conditional inference tree* duidelijk naar voor.

De *conditional inference tree* (zie Figuur 10) toont meteen het belang van de predictor *Genus (De.het)*, aangezien de dataset op het hoogste niveau meteen wordt opgesplitst in een subset 'de' en een subset 'het'. Binnen de subset met *de-woorden* blijkt *Typedeterminator* een belangrijke rol te spelen. Gemeenschappelijke lidwoorden vertonen in die subset nagenoeg geen afwijkingen van de standaardvorm, terwijl we die bij gemeenschappelijke aanwijzende voornaamwoorden wel vinden binnen de semantische klassen *low animate en unbounded abstract/unspecific mass*. Overgeneralisering van *het* in die laatste semantische klasse bevestigt, zoals reeds eerder aangehaald, de hersemantiserings-theorie.

Binnen de subset met *het-woorden* blijken de etnolectische corpora vervolgens significant meer niet-standaardvormen te bevatten dan de autochtone corpora. Binnen de etnolectische corpora blijken ook de variabelen *Typedeterminator* en *Klasse* een rol te spelen. Bij de aanwijzende voornaamwoorden blijken de semantische klassen *human, high animate* en *collective nouns* de meeste niet-standaardvormen uit te lokken, terwijl dat bij de lidwoorden eerst de klasse *high animate* is. Pas op een lager niveau lukt de klasse *human* meer niet-standaardvormen uit dan de resterende semantische klassen. Opnieuw zien we dus dat overgeneralisering van *de* voornamelijk optreedt in etnolecten binnen de hoogst-geïndividualiseerde klassen.



Figuur 10: Conditional inference tree (corpusonderzoek)

### 3.3. Conclusie

Tijdens dit corpusonderzoek werden taaldata met elkaar vergeleken, die afkomstig waren uit drie corpora: het *Moroccorp*, het *Citétaalcorpus* en het *Condiv*. Binnen het *Condiv* maakten we nog een onderscheid tussen het Vlaamse deel van het *IRC*-deelcorpus en het Nederlandse deel. De taaldata uit de corpora waren afkomstig uit gesprekken die door Vlaamse en Nederlandse jongeren in informele omstandigheden (op chatkanalen of in spontane interacties onder vrienden) werden gevoerd. De jongeren uit het *Moroccorp* en het *Citétaalcorpus* waren van allochtone afkomst en de sprekers uit het *Condiv* waren moedertaalsprekers van het Nederlands. Voor het corpusonderzoek werden 101 woorden geselecteerd op basis van hun semantische eigenschappen. Die woorden werden vervolgens onderverdeeld in 6 semantische klassen: *human*, *high animate*, *low animate*, *collective nouns*, *bounded object/abstract* en *unbounded abstract/unspecific mass*. Op die manier konden we een antwoord formuleren op de eerste twee onderzoeksvragen:

In de eerste onderzoeksvraag wilden we nagaan of etnolecten op het vlak van genustoekenning afwijken van het taalgebruik van reguliere moedertaalsprekers en of er in etnolecten inderdaad sprake is van een overgeneralisering van het algemeen lidwoord *de*, zoals in de literatuur vaak wordt aangehaald (Cornips 2008; Marzo & Ceuleers 2011; Nortier & Dorleijn 2008; Cornips & De Vogelaer 2009; Cornips, Van der Hoek & Verwer 2006). De resultaten van het corpusonderzoek bevestigden de hypothese. De Chi-kwadraattesten en de logistische regressieanalyse wezen immers uit dat de etnolectische corpora significant meer niet-standaardvormen van het definitieve lidwoord bevatten dan de autochtone corpora en dat grammaticale *het-woorden* significant vaker met de niet-standaardvorm voorkomen dan grammaticale *de-woorden* (wat duidt op de overgeneralisering van het algemeen lidwoord *de*). Dat was zowel het geval in de dataset zonder woorden met een alternerend lidwoordgebruik als in de dataset met integratie van die woorden (exclusief *raam*). Uit de logistische regressieanalyse bleek bovendien dat de effecten voor *Genus* (*de* vs. *het*) en *Etnolect* de sterkste waren. Aanwijzende voornaamwoorden zouden wel significant meer niet-standaardvormen toelaten dan definitieve lidwoorden. Ook de *random forest* en de *conditional inference tree* bevestigden de overgeneralisering van het lidwoord *de* in etnolecten.

De tweede onderzoeksvraag peilde naar de rol van het hersemantiseringsproces in de genustoekenning van adnominale elementen in etnolecten en autochtone variëteiten. We verwachtten dat we in beide variëteiten hersemantisering zouden kunnen vaststellen, maar dat het hersemantiseringsproces in de etnolectische variëteiten al duidelijker zichtbaar was. De datasets met en zonder woorden met een alternerend lidwoordgebruik in (supra-)regionale variëteiten leverden op dat vlak andere resultaten op.

Uit de dataset zonder de twaalf woorden uit de lijst van Pauwels en het WNT konden we eerst en vooral afleiden dat er voor alle taaldata uit de vier corpora samen een significante invloed bestaat van semantische klasse op standaardrealisatie. De semantische klassen *human* en *high animate* bleken immers vaker voor te komen met een niet-standaardvorm van de bijhorende determinator dan woorden uit de vier andere semantische klassen.

Om vervolgens na te gaan of er een hersemantiseringsproces aan de gang is, keken we of er significante interactie bestond tussen de variabelen *Genus* en *Klasse*. Niet-standaardrealisatie van de

hoog-geïndividualiseerde referenten zou in geval van hersemantisering voornamelijk moeten optreden bij grammaticale *het-woorden*, terwijl niet-standaardrealisatie van laag-geïndividualiseerde woorden hoofdzakelijk zou moeten voorkomen bij grammaticale *de-woorden*. De statistische analyse van de etnolectische taaldata wees echter uit dat het effect van niet-standaardrealisatie enkel aanwezig was bij *het-woorden*, wat alleen de overgeneralisering van *de* in etnolecten bevestigde en dus niet pleitte in het voordeel van een hersemantiseringsproces.

Wanneer we de woorden met een alternerend lidwoordgebruik (behalve *raam*) wel in de dataset betrokken, wezen statische testen opnieuw uit dat niet-standaardrealisatie in etnolectische variëteiten zich hoofdzakelijk voordoet bij grammaticale *het-woorden* van de hoogst-geïndividualiseerde semantische klassen. De Fischer Exact Test wees deze keer wel op een significante invloed van semantische klasse op standaardrealisatie binnen de grammaticale *de-woorden*. In de laagst-geïndividualiseerde klasse *unbounded abstract/unspecific mass* konden we met andere woorden in de etnolectische taalvariëteiten een lichte tendens tot hersemantisering van *het* vaststellen. Ook de *random forest* en de *conditional inference tree* toonden het effect van semantische klasse op standaardrealisatie.

Die laatste resultaten bevestigden deels onze hypothese. Aangezien we aan de twee uiterste polen van de *Individuation Hierarchy* in de etnolectische taalvariëteiten een lichte tendens tot hersemantisering konden vaststellen, is het mogelijk dat het hersemantiseringsproces zijn intrede doet in het adnominale domein van etnolecten. Wel moeten we daarbij opmerken dat dat hersemantiseringseffect enkel waarneembaar is wanneer we de woorden uit de lijst van Pauwels en het WNT meetellen. Het lijkt er daarom op dat hersemantisering van *de-woorden* zich net voordoet bij die woorden die deel uitmaken van die lijst. Eerder dan dat het genusgedrag van de woorden met lidwoordalternantie uit de etnolectische taaldata een geval is van dialectinterferentie, zouden we te maken kunnen hebben met een hersemantisering die zich min of meer parallel voordoet in sommige dialecten en in de etnolecten. Net door die woorden tijdens onze eerste analyse uit de dataset te filteren, kregen we hoogstwaarschijnlijk een onderschatting van het geringe hersemantiseringseffect in etnolecten te zien. In de finale discussie (zie Hoofdstuk 5) komen we nog op deze kwestie terug.

Om enerzijds het hersemantiseringsproces in het adnominale domein van etnolecten nauwkeuriger te onderzoeken en om anderzijds na te gaan of andere factoren zoals taalverwervingsproblemen of groepslidmaatschap een rol spelen bij de overgeneralisering van het algemeen lidwoord *de*, werd een bijkomend veldonderzoek uitgevoerd. De resultaten worden besproken in het volgende hoofdstuk.

# Hoofdstuk 4: Veldonderzoek

Aangezien er in de literatuur (Cornips & De Vogelaer 2009; Cornips 2008; Cornips, Van der Hoek & Verwer 2006) geregeld wordt aangehaald dat allochtone sprekers vaak moeilijkheden ondervinden met de verwerving van het Nederlandse lidwoordensysteem en zelfs op volwassen leeftijd nog variatie zouden vertonen in hun keuze voor definiete lidwoorden, werd een bijkomend veldonderzoek uitgevoerd naar het gebruik van lidwoorden bij sprekers van allochtone afkomst in Limburg. Enerzijds willen we nagaan of allochtone sprekers effectief vaak een ander lidwoord kiezen dan de standaardvorm voorschrijft (onderzoeksvraag 1), anderzijds willen we onderzoeken of er verschillen bestaan tussen leeftijdsgroepen (onderzoeksvraag 3). Wijken kinderen met andere woorden vaker af van de standaardvorm dan adolescenten of jongvolwassenen? Ten slotte willen we ook nagaan of de conclusies uit het voorgaande corpusonderzoek inzake hersemantisering in het adnominale domein doorheen dit veldonderzoek bevestigd worden (onderzoeksvraag 2).

## 4.1. Methode

Voor dit veldonderzoek werden taaldata verzameld aan de hand van een combinatie tussen de apparent-time methode en taalverwervingsonderzoek. Met die methode sluiten we ons aan bij voorgaand onderzoek van De Vos & De Vogelaer (2011), waarin vijf leeftijdsgroepen uit het Oost-Vlaamse Moerzeke door middel van een schriftelijke invultest ondervraagd werden over het geslacht van voornaamwoorden bij pronominale verwijzing. Omwille van haalbaarheidsredenen en de keuze voor een apparent time-onderzoek beperken we ons in dit veldonderzoek tot het taalgebruik van sprekers van allochtone afkomst uit Limburg en laten we het taalgebruik van autochtone en allochtone sprekers uit Nederland buiten beschouwing.

### 4.1.1. Beschrijving van de dataverzameling

Net zoals in de studie van De Vos en De Vogelaer (2011) werden er tijdens het veldonderzoek verschillende leeftijdsgroepen ondervraagd: een groep kinderen (7-8 jaar), een groep adolescenten (14 jaar) en een groep jongvolwassenen (18-21 jaar). Aan de hand van een schriftelijke test werd nagegaan in welke mate de respondenten opteerden voor het grammaticale definiete lidwoord of aanwijzend voornaamwoord van de opgegeven testwoorden of dat ze eerder een voorkeur vertoonden voor het semantische woordgeslacht<sup>15</sup>.

---

<sup>15</sup> In wat volgt, worden de termen *grammaticaal genus*, *semantisch woordgeslacht*, *standaardvorm* en *niet-standaardvorm* door elkaar gebruikt. De termen *grammaticaal genus* en *standaardvorm* drukken synoniemen uit. De term *niet-standaardvorm* omvat daarentegen zowel de *semantische lidwoorden* als de *resterende niet-standaardvormen*. Een eenvoudig voorbeeld maakt de terminologie duidelijk: Wanneer een respondent bij het onzijdige testwoord *meisje* het lidwoord *het* invult, is dat lidwoord het *grammaticale lidwoord* of de *standaardvorm*. Wanneer hij daarentegen *de* invult, hebben we te maken met een *niet-standaardvorm* die het *semantische lidwoord* uitdrukt. Volgens de hersemantiseringstheorie zouden menselijke referenten immers gecombineerd worden met het lidwoord *de*. Wanneer een respondent echter het testwoord *vader* combineert met het lidwoord *het*, hebben we ook te maken met een *niet-standaardvorm*, maar die vorm is in dit geval niet semantisch gemotiveerd.

#### 4.1.1.1. Proefpersonen en afdnameplaats

De keuze voor de drie geselecteerde leeftijdsgroepen gebeurde grotendeels op basis enkele bevindingen uit de literatuurstudie. Ten eerste bleek uit taalverwervingsonderzoek (zie 1.3.) dat moedertaalsprekers het grammaticale genussysteem van het Nederlands op 6-jarige leeftijd volledig onder de knie hebben en dat afwijkingen van het grammaticale woordgeslacht op latere leeftijd zouden wijzen op een toegenomen belang van semantisering. Aangezien de bevroegde kinderen in dit veldonderzoek bijna allemaal een thuistaal hebben die anders is dan het Nederlands, kunnen we ervan uitgaan dat ze op een leeftijd van 7-8 jaar het Nederlandse genussysteem nog niet volledig onder de knie hebben en mogelijk vaker semantische regels toepassen (zie 1.3.2.). De keuze voor kinderen van 7-8 jaar liet ons bovendien toe om data te verzamelen aan de hand van een schriftelijke test, aangezien hun lees- en schrijfvaardigheid al voldoende ontwikkeld is.

De leeftijd van de twee oudste groepen werd hoofdzakelijk gekozen door de sterke associatie van *Citétaal* met jongeren, zoals die in de literatuur geregeld aan bod komt (Marzo & Ceuleers 2001: 459; Marzo 2015: 4). Niet alleen wordt de overgeneralisering van het algemeen lidwoord *de* vanuit sociolinguïstisch standpunt voornamelijk in verband gebracht met jongeren voor wie sociale factoren als groepslidmaatschap een belangrijke rol spelen (zie 1.2.1.), maar allochtone jongeren zouden volgens taalverwervingsonderzoek bovendien tot op late leeftijd grote moeilijkheden met het Nederlandse grammaticale genussysteem ervaren en mogelijk zelfs in een staat van fossilisatie terecht komen (zie 1.3.2.). Om die reden menen we dat het van belang is om zowel adolescenten (14 jaar) als jongvolwassenen (18-21 jaar) in ons onderzoek te betrekken.

Aangezien we ons in dit onderzoek enkel focussen op het etnolect *Citétaal*, werden alle testen afgenomen in de provincie Limburg. Zo bevonden de deelnemende basisschool en middelbare school zich beide net buiten het stadscentrum van Genk. Zoals werd aangehaald in het eerste hoofdstuk (zie 1.2.3.), wordt het gebruik van *Citétaal* immers sterk in verband gebracht met de stad Genk en met lokale woonwijken waar (allochtone) jongeren opgroeien (Marzo & Ceuleers 2001: 459; Marzo 2015: 4). In tegenstelling tot de middelbare school, die ongeveer een gelijk aantal leerlingen van autochtone en allochtone afkomst aantrok, was ongeveer 90% van de leerlingen uit de deelnemende basisschool van allochtone afkomst. Dat hoge percentage is opvallend, zeker omdat de school in een eerder welstellende buurt van Genk was gelegen.

De testafname van de jongvolwassenen vond daarentegen plaats in een hogeschool in Hasselt. Aangezien er in Genk geen hogescholen zijn, stromen alle jongeren die verder studeren, immers door naar Hasselt en Diepenbeek. Doordat Genk, Diepenbeek en Hasselt zeer nabij gelegen zijn en er reeds tekenen van spreiding van *Citétaal* zijn aangetoond (Marzo & Ceuleers 2011: 454-455), menen we dat de ligging van de bevroegde hogeschool geen grote invloed zal uitoefenen op de resultaten.

In totaal werden 74 kinderen (35 jongens en 39 meisjes) uit het tweede en derde leerjaar basisschool, 21 adolescenten (6 jongens en 15 meisjes) uit het tweede middelbaar (ASO en TSO) en 60 jongvolwassenen (35 jongens en 25 meisjes) uit het eerste jaar hogeschool (studierichting *Financie- en verzekeringswezen*) ondervraagd. Dat maakte dat er op vlak van hoeveelheid proefpersonen sterk afgeweken werd van het oorspronkelijke design van De Vos en De Vogelaer (2011), dat slechts acht mannelijke en acht vrouwelijke sprekers per leeftijdscategorie voorschreef.



Aangezien we ervoor gekozen hadden om telkens de volledige klasgroep te ondervragen (om de leerlingen van allochtone afkomst niet te stigmatiseren), waren we echter genoodzaakt om dat aantal nog te reduceren. Bovendien bleken een aantal testen onleesbaar of onvolledig ingevuld. Uiteindelijk werden de resultaten van 60 kinderen (28 jongens en 32 meisjes), 14 adolescenten (4 jongens en 10 meisjes) en 13 jongvolwassenen (8 mannen en 5 vrouwen) in het onderzoek betrokken, wat een totaal van 87 respondenten van allochtone afkomst opleverde, die het Marokkaans, Turks of Italiaans als thuistaal hadden.

Met die aantallen komen we natuurlijk slechts in zekere mate tegemoet in de poging om de hoeveelheid mannelijke en vrouwelijke respondenten gelijk te houden. Omdat uit eerder onderzoek (De Vos 2009) echter is gebleken dat er voor verandering in het Nederlandse genussysteem voorlopig weinig of geen significante verschillen tussen mannen en vrouwen zijn waargenomen, menen we dat de ongelijke man-vrouwverdeling in de groep adolescenten (4 jongens versus 10 meisjes) niet problematisch is.

Socio-demografische gegevens over de respondenten werden ten slotte verkregen door middel van een vragenlijst die voor aanvang van de test door de deelnemers werd ingevuld. De vragen peilden voornamelijk naar leeftijd, geslacht, studierichting, thuistalen en sociale achtergrond.

#### 4.1.1.2. Beschrijving van de taak

De schriftelijke test werd afgenomen in de vorm van een leestekst waarin de respondenten 31 legevelden moesten invullen door telkens een keuze te maken tussen twee aangegeven woorden. Bij 18 van die velden moesten ze het juiste lidwoord of aanwijzend voornaamwoord<sup>16</sup> van een aantal testwoorden bepalen, terwijl de overige 13 invulvelden afleiders bevatten om te voorkomen dat de respondenten begrepen waarover de test ging. De respondenten wisten dat ze deelnamen aan een taaltest, maar ze waren niet op de hoogte welk aspect van hun taalvaardigheid juist getest werd. Onderstaande voorbeelden bevatten twee zinnen uit de leestekst waarin respectievelijk het juiste lidwoord van een testwoord en een afleider moesten worden ingevuld.

(13) Voorbeeldzin uit de leestekst

De scène waarin de Grote Boze Wolf \_\_\_\_\_ **(de/het)** meisje opeet, zou kinderen nachtmerries bezorgen.

(14) Afleider uit de leestekst

Ongeveer de helft van de ouders vindt *Rapunzel* en *Repelsteeltje* te beangstigend, omdat het hoofdpersonage daarin ontvoerd \_\_\_\_\_ **(word/wordt)**.

---

<sup>16</sup> De keuze tussen de aanwijzende voornaamwoorden *die/dat* werd geïntegreerd om variatie te brengen in de keuze tussen de bepaalde lidwoorden *de/het*. Net als bij de afleiders probeerden we zo te vermijden dat de deelnemers doorhadden waarover de test ging. De alternantie tussen *die/dat* wordt immers ook aangestuurd door genus.

Om ervoor te zorgen dat de respondenten een zo spontaan mogelijk antwoord gaven, werd de test aan een gemiddeld leestempo voorgelezen door de desbetreffende klasleerkracht. De deelnemers mochten enkel schrijven wanneer de leerkracht aan een invulveld was gekomen en moesten na afloop hun test meteen omdraaien en afgeven. Aan de leerkracht werd gevraagd om de invulvelden op de volgende manier voor te lezen:

(15) Voorbeeldzinnen (13) en (14) uit de leestekst, voorgelezen door de leerkracht

“De scène waarin de Grote Boze Wolf de of het meisje opeet, zou kinderen nachtmerries bezorgen”

“Ongeveer de helft van de ouders vindt *Rapunzel* en *Repelsteeltje* te beangstigend, omdat het hoofdpersonage daarin ontvoerd word zonder –t of wordt met –t”

Op die manier kon hetzelfde leestempo voortdurend worden aangehouden en kregen de deelnemers toch voldoende tijd om hun antwoord te noteren.

Aangezien het leeftijdsverschil tussen de drie groepen respondenten betrekkelijk groot was, werden er twee verschillende leesteksten opgesteld: één voor de jongste groep (kinderen) en één voor de twee oudste groepen (adolescenten en jongvolwassenen). Niet alleen kon er zo rekening gehouden worden met de moeilijkheidsgraad van de afleiders, grammaticale structuren en woordenschat, maar kon er ook een tekstonderwerp worden bepaald dat min of meer in het interessegebied van de deelnemers lag. De jongste groep kreeg een aangepaste versie van het sprookje van de *Bremer stadsmuzikanten* voorgeschoteld, terwijl de twee oudste groepen een informatieve tekst kregen over het nut van het voorlezen van sprookjes aan jonge kinderen.

De keuze voor die onderwerpen werd uiteraard ook in zekere mate uitgelokt door de selectie van de testwoorden. Alle testwoorden waren afkomstig uit de woordenlijst die werd opgesteld voor het corpusonderzoek en werden grotendeels geselecteerd op basis van enkele opmerkelijke bevindingen uit draaitabellen die voor een eerste verkenning van de resultaten van het corpusonderzoek werden opgesteld<sup>17</sup>. Zo werden er woorden gekozen die in het *Moroccorp* zeer vaak gecombineerd werden met een niet-standaardvorm van het bijhorende lidwoord of aanwijzend voornaamwoord of werden er woorden geselecteerd die door de beperkte omvang van het *Citétaaltaalcorpus* niet in dat corpus voorkwamen en daardoor aan grondiger onderzoek moesten worden onderworpen.

Net zoals tijdens het corpusonderzoek waren de testwoorden verdeeld over zes semantische klassen: *human*, *high animate*, *low animate*, *collective nouns*, *bounded object/abstract*, *unbounded abstract/unspecific mass*. Om te vermijden dat de test te lang werd en de aandacht van de respondenten zou verzwakken, werd er aanvankelijk voor gekozen om slechts één *de-woord* en één *het-woord* per semantische klasse te bevragen. De klassen *bounded object/abstract* en *unbounded abstract/unspecific mass* werden later echter nog verder onderverdeeld in subklassen, wat per subklasse nog een extra *de-woord* en een extra *het-woord* opleverde. Ook de klassen *Human* en *High animate* werden nog uitgebreid met een extra *het-woord*. De keuze om die twee extra *het-woorden*

---

<sup>17</sup> Zie Bijlage 4 voor de draaitabellen van het corpusonderzoek

in de test te integreren lag redelijk voor de hand. Ten eerste verschillen de onzijdige testwoorden *meisje* en *zusje* uit de klasse *human* van elkaar door hun mate van lexicalisering. In tegenstelling tot *zusje*, waarvan het suffix *-je* kan fungeren als een sterk morfologisch signaal dat het verkleinwoord onzijdig is, gaat het bij *meisje* om een gelexicaliseerde vorm, die daardoor mogelijk sterker onderhevig is aan semantische verwijzing. Ten tweede leidde het respectievelijk hoge en lage percentage in het gebruik van de niet-standaardvorm van het lidwoord bij *konijn* (96,55%) en *paard* (48,15%) in het *Moroccorp* tot de keuze om beide *het-woorden* uit de semantische klasse *high animate* te integreren in de test.

Onderstaande tabel bevat een overzicht van de 18 testwoorden die beide leesteksten bevatten:

Semantische klasse	De	Het
Human	<i>vader</i>	<i>meisje, zusje</i>
High animate	<i>kat</i>	<i>konijn, paard</i>
Low animate	<i>mier</i>	<i>insect</i>
Collective nouns	<i>politie</i>	<i>leger</i>
Bounded object	<i>vaas</i>	<i>huis</i>
Abstract	<i>ramp</i>	<i>probleem</i>
Unbounded abstract	<i>pijn</i>	<i>geluk</i>
Unspecific mass	<i>suiker</i>	<i>brood</i>

Tabel 9: Geselecteerde woorden veldonderzoek

Tijdens het opstellen van beide leesteksten werd er rekening mee gehouden dat de zinnen geen informatie bevatten waaruit het grammaticale geslacht van de testwoorden zou kunnen worden afgeleid. Om die reden werden de hoofdpersonages uit de leestekst van de jongste groep *Paard, Konijn, Haan, Kat* en *Mier* (met hoofdletter, zonder lidwoord) genoemd. Door het herhaaldelijk voorkomen van die personages liepen we immers de kans dat het grammaticale lidwoord doorheen de test verraden zou worden. Het gebruik van hoofdletters zorgde er echter wel voor dat de dieren in de perceptie van de deelnemers mogelijk sterker geïndividualiseerd werden en daardoor op de *Individuation Hierarchy* tussen de klassen *human* en *high animate* terecht kwamen. Om dat probleem op te lossen werd bij elk dier telkens eerst naar het juiste lidwoord gevraagd, vooraleer het personage met een hoofdletter in de tekst voorkwam.

Bovendien werd ook rekening gehouden met de kans op invloed van lidwoordalternantie in supra-regionale taalvariëteiten. Aan de hand van het werk van Pauwels (1938) en het WNT (De Vries & Te Winkel 1864-1998) werd duidelijk dat slechts 2 van de 18 geselecteerde testwoorden lidwoordvariatie vertoonden: *leger* en *suiker*. Aangezien we eerder tijdens het corpusonderzoek hadden aangetoond dat slechts 5 woorden uit die lijst (*boek, kleur, raam, winter* en *zand*) in de autochtone corpora voorkwamen met de niet-standaardvorm van hun bijbehorende determinator, werd de integratie van de woorden *leger* en *suiker* in de test niet als problematisch beschouwd. De kans dat afwijkingen op die woorden in etnolecten aan dialectinterferentie te wijten zijn, is immers zeer klein, wanneer de woorden zelfs in de autochtone corpora niet met een niet-standaardtalige determinator voorkomen.

Ten slotte werd ook op het einde van elke test aan de deelnemers gevraagd hoe moeilijk ze de test vonden. Alle deelnemers waren het er unaniem over eens dat het om een gemakkelijke test ging.

#### 4.1.2. Pilootstudie

Om de kwaliteit van de test te bepalen en om na te gaan of het niveau van de afleiders, de woordenschat en de gebruikte grammaticale structuren op maat van de verschillende leeftijdsgroepen was, werd vooraf een pilootstudie uitgevoerd bij de twee jongste groepen van deelnemers.

De pilootstudie voor de leestekst van de kinderen (7-8 jaar) vond plaats in een basisschool in Leuven, die door haar ligging in een lagere sociale woonbuurt net buiten het stadscentrum voornamelijk leerlingen van allochtone afkomst aantrekt. Om zowel rekening te kunnen houden met het leeftijdsniveau als met het feit dat de eigenlijke deelnemers van de test het Nederlands niet als moedertaal zouden hebben, werd de piloottest afgenomen bij één moedertaalspreker (1 meisje) en bij vier leerlingen (1 jongen en 3 meisjes) van allochtone afkomst die thuis Frans, Koerdisch of Somalisch spraken. Alle kinderen zaten op het moment van de testafname in het tweede leerjaar. In tegenstelling tot de meisjes, die tussen de 7 en 8 jaar oud waren, was de jongen al 10 jaar oud. Door zijn recente aankomst in België en door zijn betrekkelijk lage niveau van het Nederlands was er door de directie en het leerkrachtenteam besloten om hem dat schooljaar in het tweede leerjaar te houden.

De piloottest werd door 4 vrijwilligers afgenomen tijdens de naschoolse huiswerkbegeleiding. Zij kregen - net zoals de leerkrachten tijdens de eigenlijke test - de opdracht om de tekst aan een gemiddeld leestempo voor te lezen. Wanneer een kind duidelijk maakte dat het een woord of een zinsstructuur niet begreep, moesten de vrijwilligers dat aanduiden in de tekst. Bij eventueel tijdsgebrek werd hen gezegd dat het belangrijker was dat ze de hele tekst voorlezen en de moeilijke woorden aanduiden dan dat de kinderen de volledige test invulden.

Hoewel de moedertaalspreker weinig moeite ondervond met woordenschat of grammaticale structuren en bovendien een hoge score behaalde op zowel de testwoorden (16/18) als de afleiders (9/13), ervaarden de vier tweedetaalverwervers meer moeilijkheden. Ze vertelden na afloop van de test dat ze bij het invullen van de lege velden vaak hadden gegokt en een aantal zinnen en woorden niet goed hadden begrepen. Ook bleek geen enkel kind in staat om de inhoud van het verhaal gedetailleerd na te vertellen. Om die reden werd er beslist om een aantal zinsstructuren te vergemakkelijken<sup>18</sup> en om de moeilijkste woorden en afleiders<sup>19</sup> te vervangen door eenvoudigere alternatieven. Ondanks al die moeilijkheden bleken de 18 testwoorden wel door alle kinderen gekend.

---

<sup>18</sup> Een voorbeeld van een zinsstructuur die vergemakkelijkt werd, is: *“En de rovers? Die werden voor altijd opgesloten en ze bleven zich die dag herinneren als \_\_\_\_\_ (de/het) grootste ramp uit hun leven.”* In de definitieve test stond er: *“En de rovers? Die werden voor altijd opgesloten. Ze waren heel verdrietig. Het was \_\_\_\_\_ (de/het) grootste ramp uit hun leven.”*

<sup>19</sup> Moeilijke woorden en afleiders die vervangen werden, waren o.a. *molenaar* (door *baas*), *veulen* (door *paard*), *boordevol* (door *vol lekkere*), *uitstekend* (door *goed*) en *hijgen* (door *hulen*).

De pilootstudie voor de leestekst van de adolescenten (14 jaar) en jongvolwassenen (18-21 jaar) werd vervolgens afgenomen bij twee moedertaalsprekers (1 jongen en 1 meisje) uit de middenklasse. De jongen was op het moment van de testafname 13 jaar oud en zat in het tweede middelbaar (ASO), terwijl het meisje bijna 15 jaar oud was en in het vierde middelbaar (ASO) zat. Net zoals bij de eerste piloottest, werd de test luidop voorgelezen en werd aan de deelnemers gevraagd om te moeilijke zinsstructuren en woordenschat aan te duiden. Niet alleen bleek het niveau van de test relatief gemakkelijk voor de moedertaalsprekers, maar de jongen en het meisje behaalden ook zeer hoge scores voor de testwoorden (respectievelijk 18/18 en 17/18) en voor de afleiders (respectievelijk 11/13 en 10/13). Daarom werd er besloten dat de test uitdagend genoeg zou zijn voor anderstalige adolescenten en dat ook de jongvolwassenen weinig moeilijkheden zouden ondervinden met het niveau van de test. Een piloottest voor de oudste groep deelnemers leek daarom ook overbodig.

## 4.2. Analyse en resultaten

De analyse van de verkregen dataset valt vervolgens uiteen in twee delen. In een eerste exploratief gedeelte gaan we aan de hand van verkennende grafieken en enkele eenvoudige statistische testen (T-test, ANOVA-test en Bonferroni post hoc test) na of er sprake is van een hersemantiseringsproces in het adnominale domein van het onderzochte etnolect en of er tussen de drie leeftijdsgroepen significante verschillen in genustoekenning bestaan. Daarvoor maken we gebruik van het programma *Excel* en het softwarepakket *SPSS* (zie 4.2.1.). In het tweede deel (zie 4.2.2.) worden vervolgens de resultaten van een diepgaandere statistische analyse besproken, die werd uitgevoerd met behulp van het softwarepakket *R*.

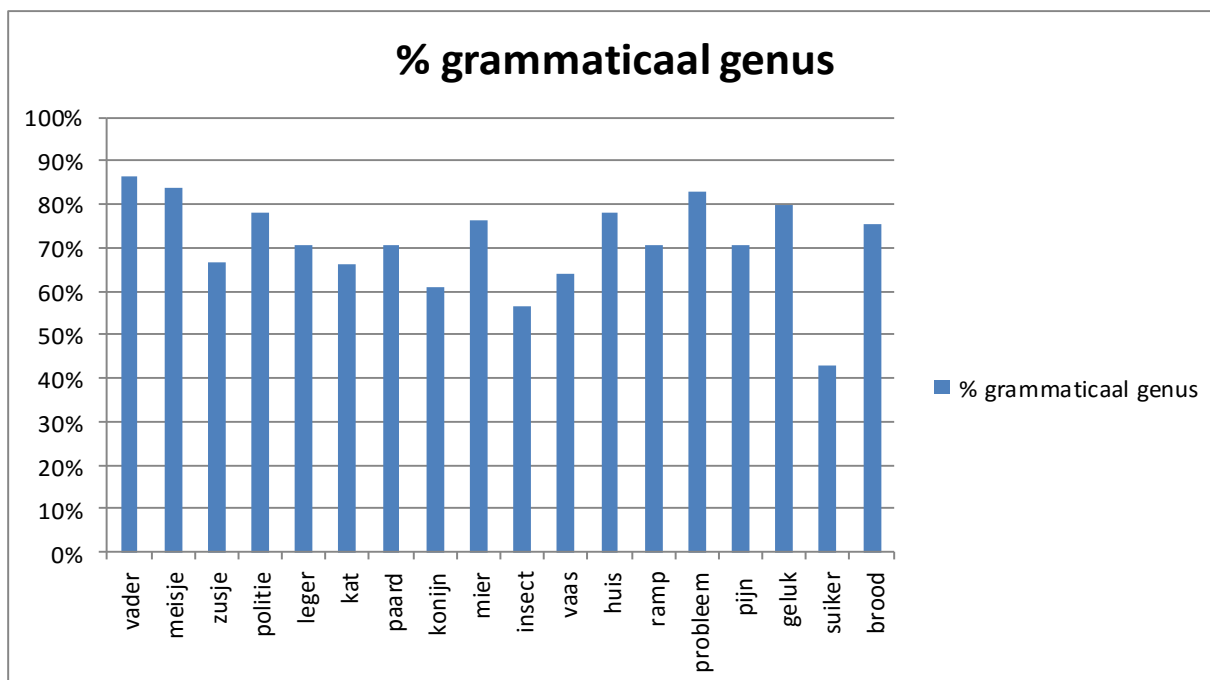
### 4.2.1. Exploratie van de verkregen data

In een eerste fase van de analyse werden de resultaten van de taaltesten verwerkt in het programma *Excel*. Dat gebeurde door per respondent een score toe te kennen op de verschillende testwoorden. Om gemakkelijk zulke scores te kunnen berekenen, werd er gewerkt met de codes *0* en *1*. Wanneer een respondent koos voor het grammaticale lidwoord (standaardvorm) van een opgegeven testwoord, kreeg hij voor dat woord de score *1*. Wanneer hij echter opteerde voor een niet-standaardvorm, werd de score *0* toegekend. Een totaalcijfer op 18 wees ten slotte uit hoe goed de respondent scoorde op ‘grammaticaal genus’ en hoe vaak hij dus koos voor de standaardvorm van het definiëte lidwoord of aanwijzend voornaamwoord voor alle testwoorden samen.

#### 4.2.1.1. Verkenning van de totaalscores ‘grammaticaal genus’

Na het toekennen van de codes, kwamen we eerst en vooral te weten dat de totaalscores ‘grammaticaal genus’ op de taaltesten van alle leeftijdsgroepen samen varieerden van 7/18 tot 18/18, met een gemiddelde score van 11,98/18 (65%). Omdat die scores ons slechts een globaal beeld geven van lidwoordvariatie, vroegen we ons af welke testwoorden het vaakst gecombineerd werden met een grammaticaal lidwoord of aanwijzend voornaamwoord en bij welke testwoorden de respondenten eerder opteerden voor de niet-standaardvorm. Op Figuur 11 staat af te lezen dat de hoogste scores voor ‘grammaticaal genus’ behaald werden bij de woorden *vader*, *meisje* en *probleem*

(respectievelijk 87%, 84% en 83%). Het woord *suiker* scoorde daarentegen betrekkelijk laag op 'grammaticaal genus' (43%).



Figuur 11: Percentages 'grammaticaal genus' per testwoord (veldonderzoek)

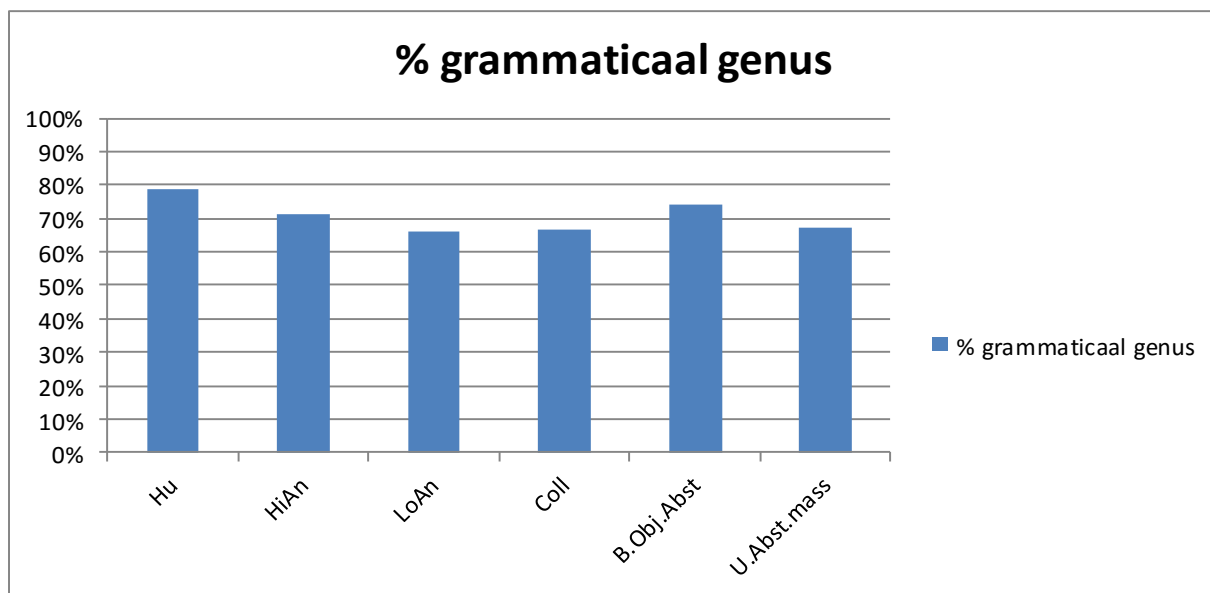
Wanneer we de totaalscores van de taalttest vervolgens vergeleken met de percentages 'standaardvorm' uit de twee allochtone corpora van het corpusonderzoek, merkten we meteen enkele verschillen (zie Tabel 10). Eerst en vooral valt de hoge score van het woord *meisje* op de taalttest op, aangezien dat woord in de twee onderzochte allochtone corpora (*Moroccorp* en *Citétaalcorpus*) zeer frequent gecombineerd werd met de niet-standaardvorm van het bijhorende lidwoord of aanwijzend voornaamwoord. Ook het woord *konijn*, dat in beide corpora opvallend laag scoorde op 'standaardvorm', behaalt in de taalttest een veel hogere score. De scores 'grammaticaal genus' van de woorden *insect* en *suiker*, die in de corpora betrekkelijk vaak gecombineerd werden met een standaardvorm, waren in de taalttest daarentegen zeer laag.

	Moroccorp	Citétaalcorpus	Taalttest
vader	100%		87%
<b>meisje</b>	<b>36,39%</b>	<b>18,18%</b>	<b>84%</b>
zusje	39,02%		67%
kat	98,72%	100%	66%
<b>konijn</b>	<b>1,28%</b>	<b>0%</b>	<b>61%</b>
paard	51,85%		71%
mier	100%		77%
<b>insect</b>	<b>100%</b>		<b>57%</b>
politie	100%	100%	78%
leger	85%		71%
vaas	75%		64%
huis	78,43%	100%	78%
ramp	100%		71%

probleem	85,91%	100%	83%
pijn	100%		71%
geluk	90,91%		80%
<b>suiker</b>	<b>78,95%</b>	<b>100%</b>	<b>43%</b>
brood	44%		76%

Tabel 10: Percentages 'grammaticaal genus' per testwoord uit het *Moroccorp*, *Citétaalcorpus* en de taaltest

Wanneer we alle testwoorden vervolgens onderbrachten in hun bijhorende semantische klassen<sup>20</sup>, bleek er op het eerste gezicht een contrast te bestaan op het vlak van genustoekenning. Zo valt uit onderstaande figuur af te leiden dat de testwoorden uit de klasse *human* net iets vaker gecombineerd worden met hun grammaticale determinator (79%) dan de testwoorden uit de andere semantische klassen.



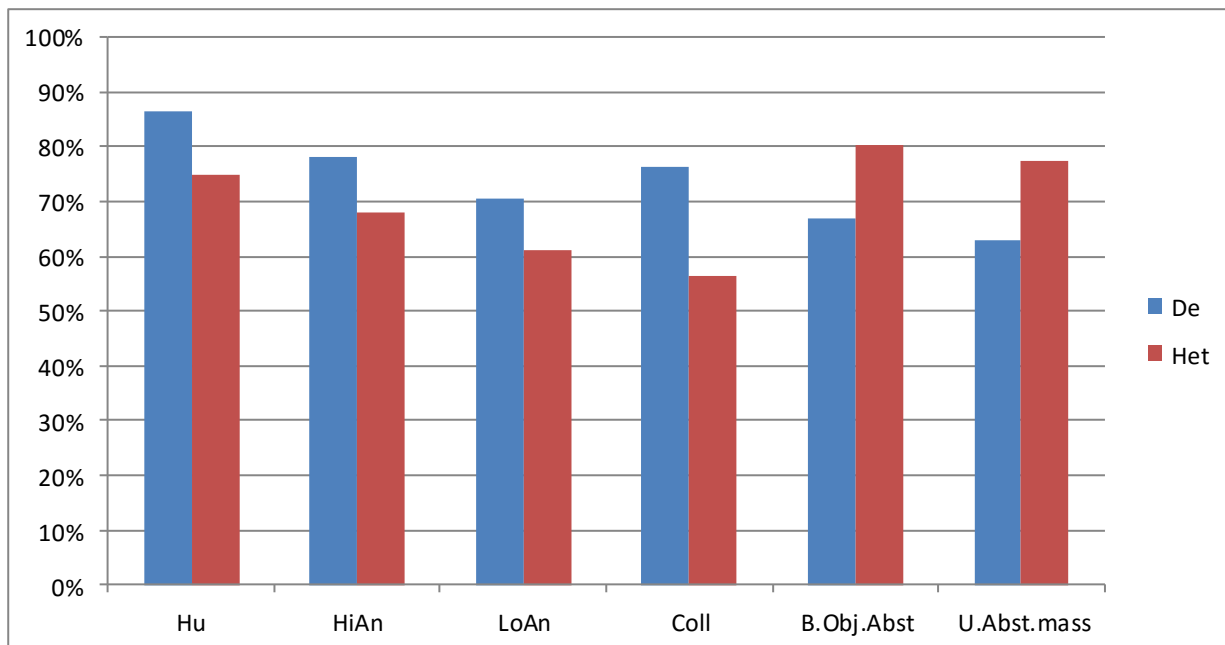
Figuur 12: Totaalscores 'grammaticaal genus' per semantische klasse (veldonderzoek)

Die vaststelling doet de vraag rijzen of niet-standaardrealisatie binnen de semantische klassen wordt uitgelokt door het gebruik van niet-standaardvormen van het gemeen lidwoord *de* of net door niet-standaardvormen van het onzijdig lidwoord *het*. Wanneer we Figuur 12 gedetailleerder bekijken door binnen de semantische klassen een opsplitsing te maken tussen grammaticale *de-woorden* en grammaticale *het-woorden* (zie Figuur 13), merken we dat niet-standaardrealisatie zich binnen de vier hoogst-geïndividualiseerde semantische klassen (*Hu*, *HiAn*, *LoAn* en *Coll*) hoofdzakelijk voordoet bij grammaticale *het-woorden*, wat duidt op een overgeneralisering van het gemeen lidwoord *de*. Bij de twee laagst-geïndividualiseerde klassen *B.Obj.Abst* en *U.Abst.mass* zien we daarentegen dat niet-standaardrealisatie voornamelijk optreedt bij grammaticale *de-woorden* en dat er zich bijgevolg een overgeneralisering van het onzijdig lidwoord *het* manifesteert. Die resultaten lijken dus op het eerste gezicht te pleiten voor een hersemantiseringsproces in de richting van de *Individuation Hierarchy* van Audring (2006): hoe hoger geïndividualiseerd de referent is, hoe groter de kans dat er

<sup>20</sup> De afkortingen voor de semantische klassen die op de horizontale as in Figuur 12 staan weergegeven, zijn dezelfde als tijdens het corpusonderzoek (zie Tabel 6 in Hoofdstuk 3).

voor een algemeen lidwoord wordt gekozen; hoe abstracter en lager geïndividualiseerd een item daarentegen is, hoe vaker er met een onzijdig lidwoord naar wordt verwezen.

Toch moeten we daarbij opmerken dat de *Individuation Hierarchy* van Audring (2006) niet alle variatie verklaart. Zo vinden we bij de hoogst-geïndividualiseerde klassen ook niet-standaardrealisaties van *de-woorden*. In de klassen *Hu* en *HiAn* treffen we bijvoorbeeld ‘slechts’ scores van respectievelijk 87% en 78% op ‘grammaticaal genus’ bij *de-woorden* aan. De resterende 13% en 22% worden met andere woorden ingevuld door realisaties als *het vader* of *het kat*, beide niet-standaardvormen die niet semantisch verklaard kunnen worden. Hetzelfde geldt voor de laagst-geïndividualiseerde klassen: de hoge scores op ‘grammaticaal genus’ van *het-woorden* (in de klassen *B.Obj.Abst* en *U.Abst.mass* respectievelijk 81% en 78%) nemen niet weg dat er bijvoorbeeld in de klasse *U.Abst.mass* constructies als *de brood* of *de geluk* optreden, die geen enkele semantische motivering hebben.



Figuur 13: Percentage ‘grammaticaal genus’ per semantische klasse, per woordgeslacht (veldonderzoek)

#### 4.2.1.2. Verkenning van de variabele ‘leeftijdscategorie’

Na een globale verkenning van de totaalscores op ‘grammaticaal genus’ van alle leeftijdsgroepen samen willen we in een tweede stap van de exploratie van de dataset onderzoeken of er verschillen bestaan tussen de scores van de drie ondervraagde leeftijdsgroepen. Het doel van dit apparent time-onderzoek is immers om na te gaan of er effectief sprake is van een ‘taalverwervingsprobleem’ bij de allochtone respondenten en of dat mogelijke ‘probleem’ evolueert naarmate de allochtone kinderen ouder worden. Op basis van eerdere studies naar tweedetaalverwerving (Cornips & De Vogelaer 2009; Cornips 2008; Cornips, Van der Hoek & Verwer 2006), die stellen dat allochtone sprekers tot op late leeftijd moeilijkheden ondervinden met de verwerving van het Nederlandse lidwoordensysteem, veronderstellen we dat de bevroegde kinderen vaker zullen afwijken van de standaardvorm van het definiete lidwoord of aanwijzend voornaamwoord dan de bevroegde adolescenten of jong-



volwassenen. De adolescenten zouden volgens die redenering op hun beurt ook vaker afwijken van de standaardvorm dan de jongvolwassenen.

Een ANOVA-test wees uit dat er inderdaad een significant verschil bestaat tussen de scores van de drie leeftijdsgroepen ( $p=0.000$ ). Een bijkomende post hoc-test (Bonferroni-correctie) toonde aan dat enkel het verschil tussen leeftijdscategorie 1 (kinderen) en leeftijdscategorie 2 (adolescenten) enerzijds en het verschil tussen leeftijdscategorie 1 (kinderen) en leeftijdscategorie 3 (jongvolwassenen) anderzijds significant waren ( $p=0.000$ ). Het verschil tussen de scores van de adolescenten en de jongvolwassenen bleek met andere woorden niet significant ( $p=1.000$ ).

#### ANOVA

total

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	349,092	2	174,546	29,628	,000
Within Groups	494,862	84	5,891		
Total	843,954	86			

#### Post Hoc Tests

##### Multiple Comparisons

Dependent Variable: total

Bonferroni

(I) Leeftijdscategorie	(J) Leeftijdscategorie	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	-4,29524 <sup>*</sup>	,72041	,000	-6,0551	-2,5354
	3	-4,36667 <sup>*</sup>	,74253	,000	-6,1806	-2,5528
2	1	4,29524 <sup>*</sup>	,72041	,000	2,5354	6,0551
	3	-,07143	,93486	1,000	-2,3552	2,2123
3	1	4,36667 <sup>*</sup>	,74253	,000	2,5528	6,1806
	2	,07143	,93486	1,000	-2,2123	2,3552

\*. The mean difference is significant at the 0.05 level.

Tabel 11: Resultaten Anova-test en Bonferroni-correctie (veldonderzoek)

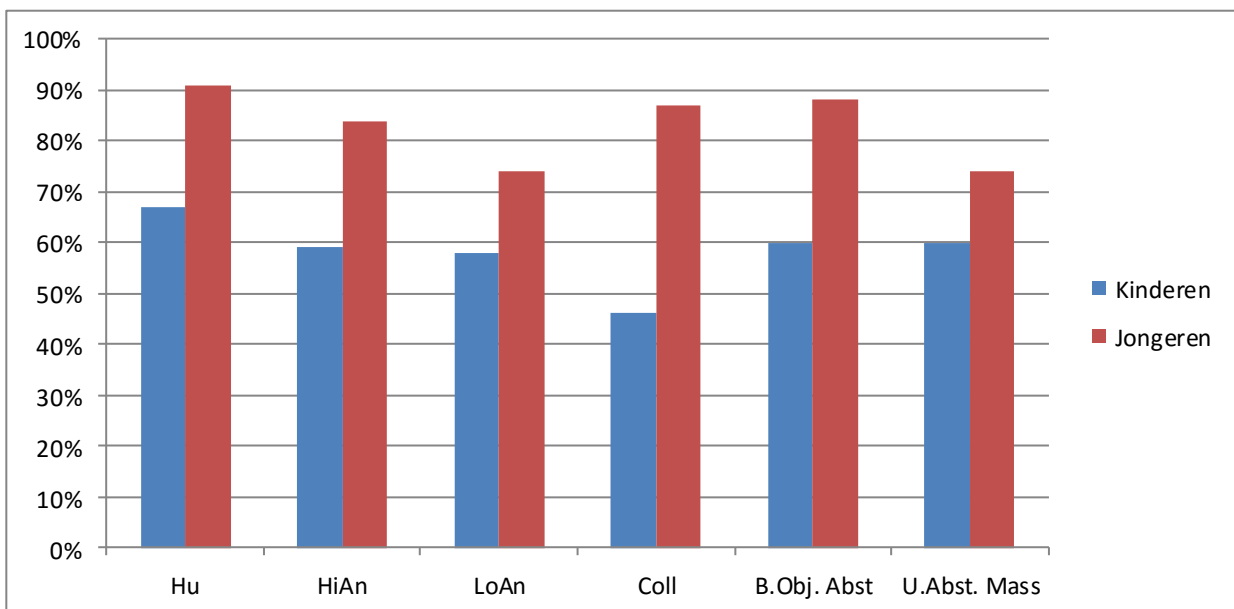
Die resultaten bevestigen grotendeels onze veronderstelling, en dus bijgevolg ook de bevindingen uit eerder taalverwervingsonderzoek (Cornips & De Vogelaer 2009; Cornips 2008; Cornips, Van der Hoek & Verwer 2006). Kinderen van allochtone afkomst blijken inderdaad moeilijkheden te vertonen met de verwerving van het Nederlandse lidwoordensysteem en scoren significant lager op 'grammaticaal genus' van de opgegeven testwoorden dan adolescenten en jongvolwassenen. We hadden echter verwacht dat ook de bevroegde adolescenten significant lager zouden scoren op 'grammaticaal genus' dan de jongvolwassen, maar dat bleek niet het geval te zijn.

Door het gebrek aan significantie tussen de scores van de adolescenten en de jongvolwassen en door het feit dat de gemiddelde scores van beide leeftijdsgroepen enorm dicht bij elkaar liggen (respectievelijk 14,9286/18 en 15/18), werd er beslist om de groep adolescenten ( $N=14$ ) en de groep

jongvolwassenen (N=13) in één groep samen te brengen. In de verdere analyse wordt er dus gewerkt met twee leeftijdsgroepen: de kinderen (N=60) en de jongeren (N=27). Nu we weten dat er effectief verschillen bestaan tussen die twee leeftijdsgroepen, zullen we in wat volgt onderzoeken waar die verschillen juist optreden (zie 4.2.1.3.).

#### 4.2.1.3. Vergelijking tussen de leeftijdsgroepen 'kinderen' en 'jongeren'

Om na te gaan waar de verschillen tussen de leeftijdsgroepen 'kinderen' en 'jongeren' juist optreden, werden een aantal statistische testen uitgevoerd en werden enkele grafieken opgesteld. Een T-test wees eerst en vooral uit dat de jongeren een significant betere gemiddelde totaalscore 'grammaticaal genus' behaalden dan de kinderen, met 14,9630/18 tegenover 10,6333/18 ( $p=0.000$ ). Uit onderstaand staafdiagram kunnen we bovendien afleiden dat de jongeren over alle semantische klassen hoger scoren dan de kinderen. Die verschillen bleken bovendien voor alle semantische klassen significant, behalve voor de klasse *low animate (LoAn)* ( $p=0.054$ ).



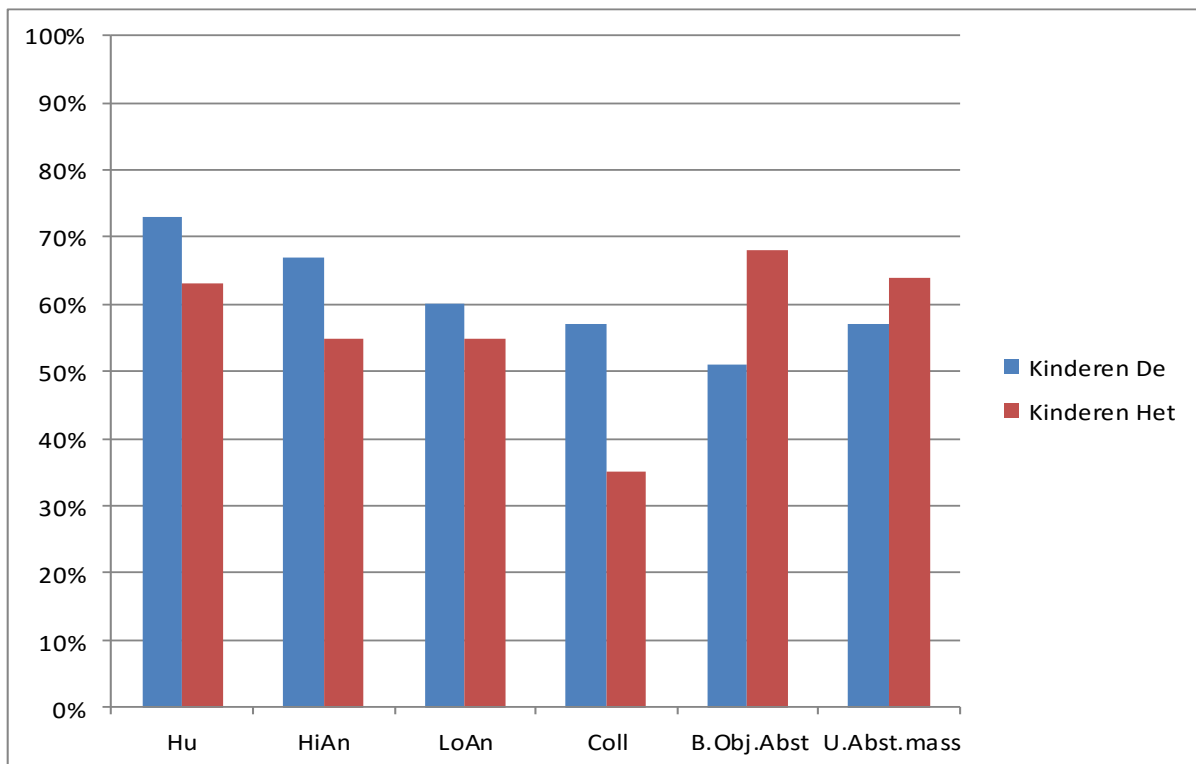
Figuur 14: Scores 'grammaticaal genus' per semantische klasse, per leeftijdscategorie (veldonderzoek)

Bij een gedetailleerdere analyse van Figuur 14 valt voornamelijk het verschil tussen de scores op de klasse *collective nouns (Coll)* op. Hoewel de jongeren daar gemiddeld zeer hoog scoren voor 'grammaticaal genus' (87%), behalen de kinderen slechts een score van 46%. Aangezien de klasse *collective nouns* in de leestekst bestond uit de testwoorden *politie (de)* en *leger (het)*, kunnen we met behulp van Figuur 15 afleiden dat vooral het woord *leger* bij de kinderen vaak gecombineerd werd met een niet-standaardvorm. Slechts 35% van de kinderen gebruikte bij het woord *leger* het grammaticale lidwoord. De jongeren bleken daarentegen veel minder problemen te ondervinden met dat woord: in 78% van de gevallen werd het grammaticale lidwoord *het* gebruikt (zie Figuur 14).

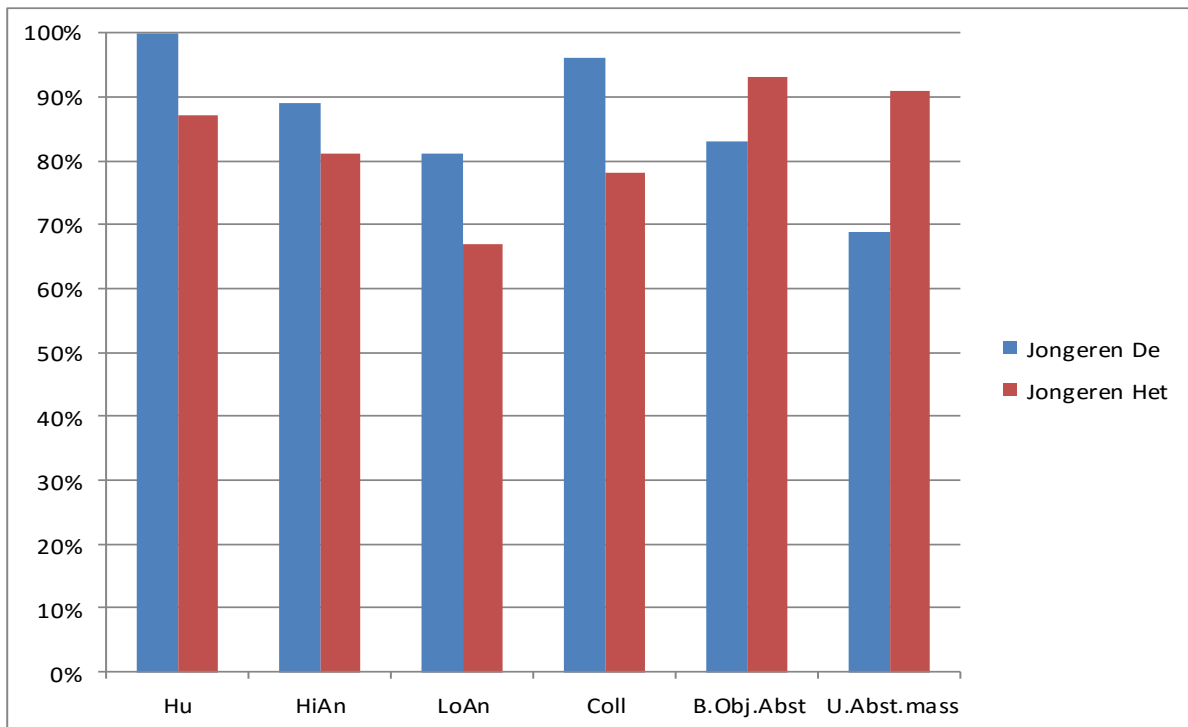
Vervolgens valt uit Figuur 14 ook af te leiden dat zowel de kinderen als de jongeren de testwoorden uit de klasse *human* het vaakst combineerden met een standaardvorm (respectievelijk 67% en 91%). De algemene tendens, die we reeds hadden opgemerkt in Figuur 12 en die stelde dat testwoorden het vaakst gecombineerd worden met de standaardvorm van de bijhorende determinator in de

klasse *human*, wordt nu dus bevestigd doorheen de verschillende leeftijdsgroepen. Zoals eerder aangehaald, is dat opmerkelijk in vergelijking met de resultaten uit het corpusonderzoek en de vaak aangehaalde voorbeelden uit de literatuur, die stellen dat een woord als *meisje* zeer vaak gecombineerd wordt met een niet-standaardvorm.

Wanneer we vervolgens in Figuur 15 en Figuur 16 een onderscheid maken in *de-woorden* en *het-woorden* per semantische klasse en per leeftijdscategorie, zien we in de grafieken van beide leeftijdsgroepen nog een tweede tendens terugkomen. In Figuur 13 hadden we namelijk al eerder besproken dat we voor alle leeftijdsgroepen samen bij de vier hoogst-geïndividualiseerde klassen een tendens tot overgeneralisering van het commun lidwoord *de* opmerkten en bij de twee laagst-geïndividualiseerde klassen een tendens tot overgeneralisering van *het*. Ook bij de kinderen en de jongeren afzonderlijk zien we die tendens optreden. In de vier hoogst-geïndividualiseerde klassen (*Hu*, *HiAn*, *LoAn* en *Coll*) zien we immers dat niet-standaardrealisatie zich hoofdzakelijk voordoet bij grammaticale *het-woorden*, terwijl niet-standaardrealisatie bij de twee laagst-geïndividualiseerde klassen (*B.Obj.Abst* en *U.Abst.mass*) voornamelijk voorkomt bij grammaticale *de-woorden*.



Figuur 15: Scores 'grammaticaal genus' per semantische klasse, per woordgeslacht van de leeftijdscategorie 'kinderen' (veldonderzoek)

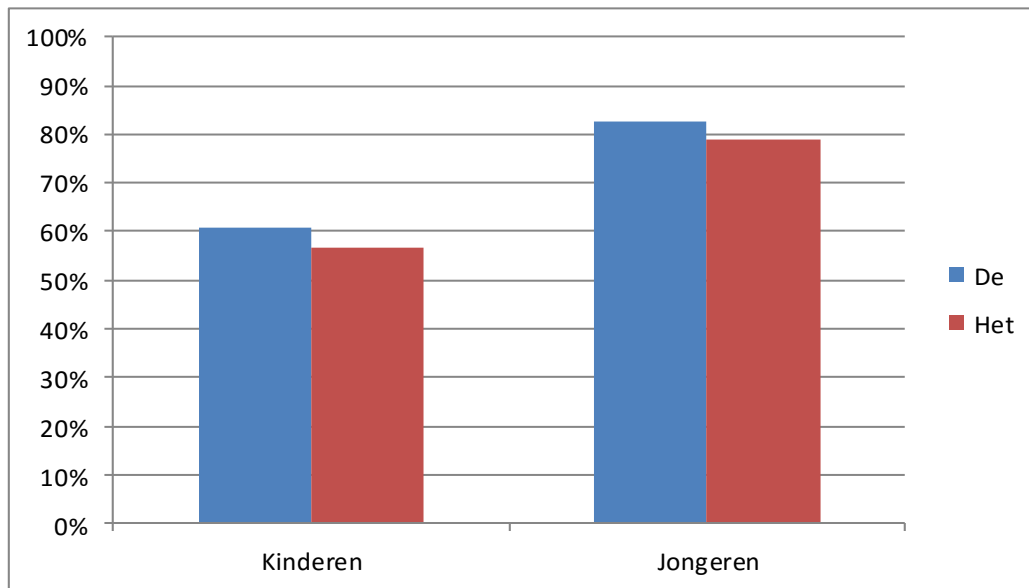


Figuur 16: Scores 'grammaticaal genus' per semantische klasse, per woordgeslacht van de leeftijdscategorie 'jongeren' (veldonderzoek)

#### 4.2.1.4. Vergelijking scores op 'de-woorden' en 'het-woorden'

In een derde stap van de exploratie van de dataset onderzochten we vervolgens de scores op grammaticale *de-woorden* en *het-woorden* van de kinderen en de jongeren. Uit Figuur 17 kunnen we eerst en vooral afleiden dat de jongeren – net zoals bij de verschillende semantische klassen – op zowel de *de-woorden* als de *het-woorden* een hogere score behaalden dan de kinderen. Het verschil tussen de scores op *de-woorden* enerzijds en de scores op *het-woorden* anderzijds bleek bovendien significant (beide  $p=0.000$ ).

Ten tweede zien we dat beide leeftijdsgroepen hogere scores behalen op grammaticale *de-woorden* dan op grammaticale *het-woorden*, ondanks het feit dat de scores erg dicht bij elkaar liggen. De kinderen scoren op *de-woorden* 61% en op *het-woorden* 57%. De jongeren behalen daarentegen een score van respectievelijk 83% en 79%. De statistische analyse wees echter uit dat er niet significant hoger gescoord wordt op grammaticale *de-woorden* dan op grammaticale *het-woorden* ( $p=0.678$ ). De overgeneralisering van het algemeen lidwoord *de* wordt met andere woorden in de taaltest statistisch gezien niet bevestigd.



Figuur 17: Totaalscore 'grammaticaal genus' per woordgeslacht, per leeftijdscategorie (veldonderzoek)

Voor een verklaring voor de afwezigheid van een overgeneralisering van het algemeen lidwoord *de* in de taaltest, kunnen we verschillende kanten uit. Een eerste mogelijkheid betreft de kans dat de deelnemers zich min of meer bewust zijn van de overgeneralisering van het lidwoord *de*, waardoor ze er bij het invullen van een formele taaltest rekening mee houden en mogelijk zelfs het lidwoord *het* overgeneraliseren ter compensatie. De jongeren zouden zich met andere woorden in zekere mate bewust zijn van de stigmatiserende oordelen die met de overgeneralisering van het lidwoord *de* in hun taalgebruik gepaard gaan, terwijl de kinderen (7-8 jaar) ook al in zekere mate rekening kunnen houden met de normcorrecties van hun leerkrachten. Een tweede mogelijkheid is daarentegen dat de tendens tot overgeneralisering van het onzijdig lidwoord *het* in de twee laagstgeïndividualiseerde klassen leidt tot een afwezigheid van significante verschillen tussen de scores op *de*-woorden en *het*-woorden. Eerder dan dat er enkel een overgeneralisering van *de* optreedt, zouden we te maken kunnen hebben met een overgeneralisering van beide lidwoorden, afhankelijk van de semantische klasse waartoe de woorden behoren.

#### 4.2.1.5. Vergelijking scores op individuele woorden

In een laatste stap van de exploratie van de dataset bekijken we ten slotte de scores op de 18 individuele testwoorden om zo een gedetailleerd beeld te krijgen van waar variatie in het lidwoordensysteem van de allochtone sprekers juist optreedt. Aan de hand van die analyse kunnen we mogelijk een antwoord vinden op de vraag welke van de twee bovenstaande verklaringen voor de afwezigheid van een significante overgeneralisering van het algemeen lidwoord *de* in de taaltest het meest plausibel lijkt.

Tabel 12 bevat een overzicht van de percentages 'grammaticaal genus' van de 18 testwoorden per leeftijdscategorie. Wanneer een score lager was dan 55%, werd dat in het grijs aangeduid. De opmerkelijkste bevindingen worden hieronder besproken.

Zoals eerder aangehaald tijdens deze analyse, vallen de hoge scores ‘grammaticaal genus’ op de testwoorden *meisje* en *konijn* eerst en vooral op, aangezien die woorden in de allochtone corpora net zeer frequent gecombineerd werden met een niet-standaardvorm. Ook de zeer lage score op het woord *suiker* werd reeds aangehaald. Zowel de kinderen als de jongeren behalen voor dat woord slechts respectievelijk 53% en 33%. Opvallend genoeg blijkt *suiker* bovendien het enige woord te zijn waarvoor de jongeren een lagere score behalen dan de kinderen. Verder vallen ook de zeer lage scores van de kinderen voor de woorden *leger* (35%) en *paard* (48%) op, zeker omdat de jongeren op die woorden een relatief hoge score behalen (respectievelijk 78% en 93%).

	Kinderen	Jongeren
vader (de)	73%	100%
meisje (het)	75%	93%
zusje (het)	52%	81%
kat (de)	67%	89%
konijn (het)	62%	70%
paard (het)	48%	93%
mier (de)	60%	81%
insect (het)	55%	67%
politie (de)	57%	96%
leger (het)	35%	78%
vaas (de)	50%	78%
huis (het)	67%	89%
ramp (de)	52%	89%
probleem (het)	70%	96%
pijn (de)	60%	81%
geluk (het)	67%	93%
suiker (de)	53%	33%
brood (het)	62%	89%

Tabel 12: Percentages ‘grammaticaal genus’ per testwoord, per leeftijdscategorie (veldonderzoek)

Een eerste opvallende bedenking bij die resultaten betreft de woorden *suiker* en *leger*. Tijdens de selectie van de testwoorden hadden we immers opgemerkt dat die twee woorden volgens het werk van Pauwels en het WNT lidwoordvariatie vertonen in (supra-)regionale variëteiten. Omdat we tijdens het corpusonderzoek echter hadden aangetoond dat de woorden *leger* en *suiker* in de autochtone corpora niet voorkwamen met de niet-standaardvorm van hun bijhorende determinator, meenden we dat de integratie van die twee woorden in de test niet problematisch was. Nu merken we echter dat de scores van die woorden net de opvallendste wendingen nemen. Niet alleen behalen *leger* en *suiker* de laagste scores uit de taaltest (35% op *leger* en 33% op *suiker*), maar we zien ook dat wanneer een bepaalde leeftijdscategorie opvallend laag scoort, de andere leeftijdscategorie een relatief goede score behaalt.

Net zoals we in het corpusonderzoek concludeerden (zie 3.2.3.), is het hoogstwaarschijnlijk geen toeval dat juist de woorden op de lijst van Pauwels en het WNT het sterkst bijdragen tot hersemantisering en dat die hersemantisering zich min of meer parallel zou voordoen in sommige

dialecten en etnolecten. Eerder dan dat er bij de bevroegde Limburgse respondenten sprake zou zijn van dialectinterferentie (de jongeren en kinderen van allochtone afkomst in Limburg spreken en horen immers amper dialect; zie ook 3.2.3), zouden moedertaalsprekers door lidwoordvariatie bij de woorden op de lijst van Pauwels en het WNT wel verwarring kunnen uitlokken en zo de lidwoordkeuze van allochtone sprekers beïnvloeden. In de finale discussie (zie Hoofdstuk 5) komen we nog op deze kwestie terug. De vraag blijft echter wel waarom dat hersemantiserings-effect slechts bij één bepaalde leeftijdsgroep optreedt en bovendien voor de twee woorden verschillend is. Een onderzoek met meer data zou eventueel een antwoord op die vraag kunnen bieden.

Een tweede opmerking gaat vervolgens over de keuze om de klassen *human* en *high animate* uit te breiden met een extra *het-woord*. Eerst en vooral hadden we het woord *zusje* in de klasse *human* extra geselecteerd omwille van de duidelijke diminutiefuitgang, die in het woord *meisje* reeds gelexicaliseerd is. We hadden daarom verwacht dat *zusje* vaker gecombineerd zou worden met het grammaticale lidwoord dan *meisje*. Uit de resultaten van de taaltest blijkt echter dat zowel de kinderen als de jongeren een hogere score ‘grammaticaal genus’ behaalden op het woord *meisje*. Een verklaring voor dat onverwachte resultaat kunnen we wel vinden bij Audring (2006: 105). Eerder dan dat we het verschil tussen de twee woorden taalintern moeten zoeken bij lexicalisatie van de diminutiefuitgangen, stelt Audring (2006) dat er tussen *meisje* en *zusje* een semantisch verschil bestaat in *verwantschap* (*kinship*). De *Animacy Hierarchy* van Silverstein (1976), waarop de *Individuation Hierarchy* van Audring (2006) is gebaseerd, bevat immers ook de subklasse *kin*, die net iets hoger staat dan alle andere menselijke referenten. Omdat het woord *zusje* tot de subklasse *kin* behoort, zou dat woord gemakkelijker aanzetten tot semantische referentie dan *meisje*. De spreker kent een *zusje* immers persoonlijk, terwijl een *meisje* verwijst naar een abstracter, onbekender individu. Bijgevolg zou *zusje* net iets gemakkelijker gecombineerd worden met het commuun lidwoord *de* dan *meisje*. Enerzijds biedt die redenering ons een plausibele verklaring voor het frequente gebruik van een niet-standaardvorm van de determinator bij het testwoord *zusje*, maar anderzijds verklaart ze het verschil in de scores ‘grammaticaal genus’ bij het testwoord *meisje* in het corpusonderzoek en het veldonderzoek niet volledig. Hoe kan het dat de scores op ‘grammaticaal genus’ voor *meisje* in de taaltest zo hoog zijn, terwijl het corpusonderzoek uitwees dat *meisje* net zeer vaak gecombineerd wordt met een niet-standaardvorm? Het lijkt haast alsof de twee leeftijdsgroepen zich min of meer bewust zijn van de overgeneralisering van het lidwoord *de* bij *meisje* in hun etnolect en er daarom tijdens het invullen van een formele taaltest meer rekening mee houden dan bij een woord als *zusje*. De eerste verklaring (jongeren zijn zich min of meer bewust van stigmatiserende oordelen, kinderen van normcorrecties) die we hierboven gaven voor de afwezigheid van een overgeneralisering van het commuun lidwoord *de* in de taaltest, lijkt daarom plausibel. In de finale discussie (zie Hoofdstuk 5) komen we hier ook nog op terug.

Ten tweede hadden we voor de onzijdige woorden uit de klasse *high animate* zowel *paard* als *konijn* geselecteerd. *Konijn* bleek in de allochtone corpora immers vaak gecombineerd te worden met een niet-standaardvorm, terwijl dat bij *paard* minder het geval was. Uit de taaltest blijkt dat de jongeren inderdaad hoger te scoren op ‘grammaticaal genus’ bij het woord *paard* (93%) dan bij het woord *konijn* (70%). De kinderen behalen daarentegen een opmerkelijk lage score voor het woord *paard* (48%) en scoren veel hoger bij het woord *konijn* (62%). Ook in de semantische klasse *high animate* lijkt er dus iets aan de hand te zijn.

## 4.2.2. 'Random forest' en 'Conditional inference tree'

In het tweede deel van de analyse maken we ten slotte gebruik van de statistische toepassing *random forests*, die toegankelijk is via het softwarepakket *R* (Strobl et al. 2007,2008; Hothorn et al. 2006) en die ook reeds tijdens het corpusonderzoek aan bod kwam (zie 3.2.5.). Deze analyse laat ons toe om globalere conclusies uit het veldonderzoek te trekken door de invloed van verschillende onafhankelijke variabelen te meten. Vooraleer we de statistische toepassing en de resultaten gedetailleerder bespreken (zie 4.2.2.2.), geven we eerst een overzicht van alle variabelen en hun bijhorende niveaus die in de analyse worden betrokken (zie 4.2.2.1.).

### 4.2.2.1. Overzicht van de variabelen

Onderstaande tabel bevat een overzicht van alle onafhankelijk variabelen en hun bijhorende niveaus:

Variabelen	Niveaus
<b>Responsvariabele</b>	
- realisatie	<i>correct, niet correct</i>
<b>Verklarende variabelen</b>	
- lexeem	<i>vader, meisje, zusje, kat, paard, konijn, mier, insect, politie, leger, vaas, huis, ramp, probleem, pijn, geluk, suiker, brood</i>
- semantische klasse (semklasse)	<i>human, high animate, low animate, collective nouns, bounded object/abstract, unbounded abstract/unspecific mass</i>
- genus	<i>de, het</i>
- leeftijdsgroep	<i>kind, jongeren</i>
- geslacht	<i>M, V</i>
- thuistaal	<i>Turks, Arabisch, Nederlands, Ander</i>
- taalvriend	<i>Nederlands, Turks, Mixed</i>
- bestlang	<i>Turks, Arabisch, Nederlands, Mixed, Ander</i>

Tabel 13: Overzicht van de variabelen en hun bijhorende niveaus (veldonderzoek)

De responsvariabele *realisatie* drukt eerst en vooral het verschil uit tussen het correct gebruik (*correct*) of incorrect gebruik (*niet correct*) van het grammaticale lidwoord of aanwijzend voornaamwoord van de 18 opgegeven testwoorden. Het onderscheid tussen de niveaus *correct* en *niet correct* komt overeen met de scores 0 en 1 die tijdens de exploratie van de dataset gebruikt werden. De variatie tussen *correct* en *niet correct* trachten we te verklaren aan de hand van 7 predictoren.

De eerste predictor *lexeem* omvat 18 levels die de individuele testwoorden uitdrukken: *vader, meisje, zusje, kat, paard, konijn, mier, insect, politie, leger, vaas, huis, ramp, probleem, pijn, geluk, suiker* en *brood*. Op die manier kunnen we nagaan bij welke testwoorden niet-standaardrealisatie het vaakst optreedt.

De predictor *semantische klasse* bevat daarnaast 6 niveaus waarin de semantische klasse van de testwoorden wordt weergegeven. Opnieuw gebruiken we dezelfde semantische klassen als tijdens



het corpusonderzoek en de exploratie van de dataset. Wanneer er sprake zou zijn van een hersemantiseringsproces in het adnominale domein, zou onder andere de variabele *semklasse* een significante invloed moeten uitoefenen op de responsvariabele *realisatie*.

De derde predictor *genus* geeft vervolgens het grammaticale woordgeslacht van de 18 testwoorden weer: commuun (*de*) of onzijdig (*het*). Aan de hand van die variabele kunnen we de overgeneralisering van het commuun lidwoord *de* in etnolecten gedetailleerder onderzoeken. Tijdens de exploratie van de dataset kwamen we namelijk te weten dat er bij de testwoorden geen sprake was van een overgeneralisering van *de*. Ook de variabele *genus* zou bij een hersemantiseringsproces in het adnominale domein een significante invloed moeten uitoefenen op de responsvariabele.

De variabele *leeftijdsgroep* omvat de verschillende leeftijdscategorieën van de respondenten. Omdat we tijdens de exploratie van de dataset aan de hand van een statistische test te weten kwamen dat er geen significant verschil bestaat tussen de scores van de adolescenten en de jongvolwassenen, maken we ook hier enkel een onderscheid tussen twee niveaus: *kind* en *jongeren*.

De vijfde variabele *geslacht* bevat vervolgens twee niveaus waarin we een onderscheid maken tussen het geslacht van de respondenten: *M* (man) of *V* (vrouw).

Aan de hand van de zesde predictor *thuis taal* gaan we na of de thuis taal van de respondenten een significante invloed uitoefent op hun scores voor het grammaticale lidwoord van de opgegeven testwoorden. We maken een onderscheid tussen vier niveaus: *Turks*, *Arabisch*, *Nederlands* en *Ander*. Het niveau *Nederlands* bevat eerst en vooral de respondenten van allochtone afkomst die thuis tegen hun moeder, vader, broers, zussen of andere familieleden enkel Nederlands spreken. De niveaus *Turks* en *Arabisch* bevatten vervolgens de respondenten die tegen hun moeder, vader, broers, zussen of andere familieleden respectievelijk Turks of Arabisch (Marokkaans/Berber/Arabisch) spreken of die taal tegen bepaalde familieleden afwisselen met Nederlands. Het niveau *Ander* bevat ten slotte de respondenten die een thuis taal hebben die anders is dan het Nederlands, Turks of Arabisch en zij die die andere taal tegen bepaalde familieleden afwisselen met Nederlands.

Bij de zevende variabele *taalvriend* maken we daarnaast een onderscheid tussen de respondenten die tegen hun vrienden enkel Nederlands spreken, de respondenten die tegen hun vrienden alleen Turks spreken en de respondenten die tegen hun vrienden meerdere talen spreken. Zo bekomen we drie niveaus: *Nederlands*, *Turks* en *Mixed*. Logischerwijs houdt die categorisatie in dat de respondenten die als thuis taal *Arabisch* of *Ander* opgaven, zijn ingedeeld in de niveaus *Nederlands* of *Mixed*. Van de Turkse respondenten waren er dus sommigen die aangaven dat ze tegen hun vrienden enkel Turks spraken, terwijl niemand van de Marokkaanse respondenten aangaf tegen zijn vrienden enkel Arabisch (Marokkaans/Berber/Arabisch) te spreken.

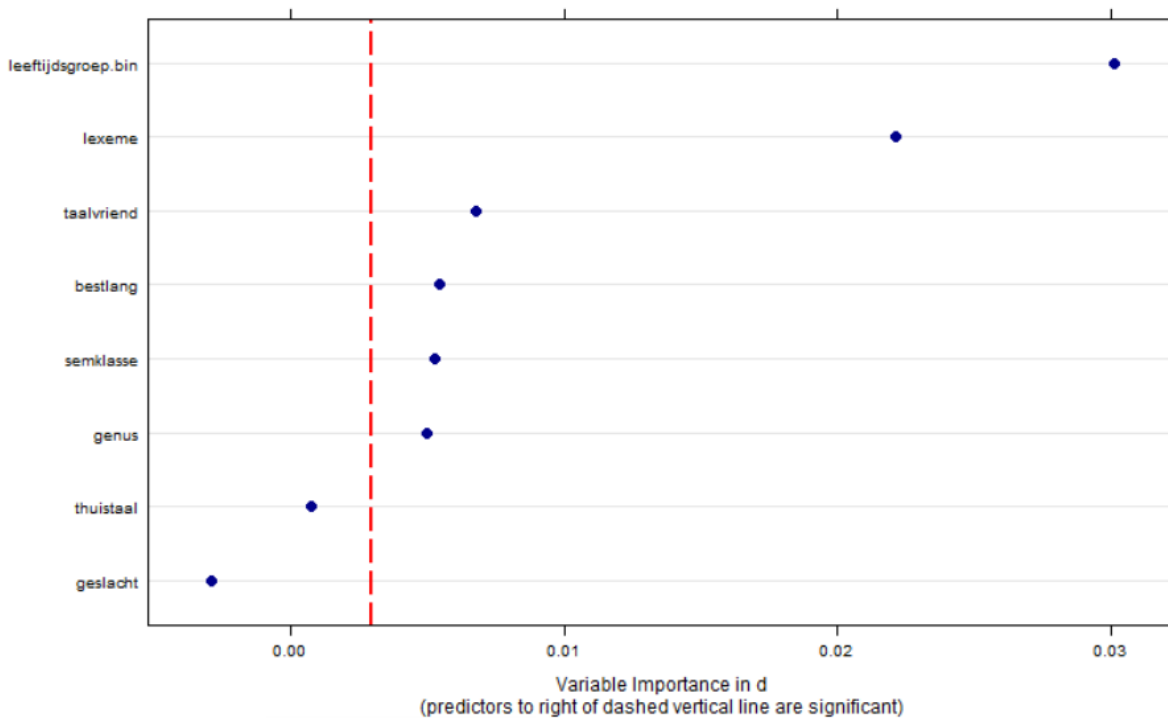
De predictor *best taal* geeft ten slotte de taal weer die de respondenten rapporteerden als de taal die ze volgens zichzelf het beste spraken. Hier maken we een onderscheid tussen vijf niveaus: *Turks*, *Arabisch*, *Nederlands*, *Mixed* en *Ander*. Het niveau *Mixed* bevat de respondenten die niet één bepaalde voorkeur hadden, maar naast het Nederlands nog een andere thuis taal opgaven.

#### 4.2.2.2. Bespreking van de 'random forest' en 'conditional inference tree'

Nu we alle betrokken variabelen overlopen hebben, zullen we in wat volgt een statistische analyse van de dataset bespreken, waarin we gebruik maken van de techniek van *conditional inference trees* en *random forests*, zoals die reeds tijdens het corpusonderzoek werd aangehaald (zie 3.2.5).

Tagliamonte & Baayen (2012: 23-24) halen verschillende voordelen aan van het gebruik van *random forests* als aanvulling op logistische modellen. Het belangrijkste voordeel voor het veldonderzoek betreft het feit dat *random forests* met *samples* van de predictoren werken en daardoor goed gebruikt kunnen worden bij problemen met meer variabelen dan observaties.

Figuur 18 toont de *random forest* waarin de relatieve invloed van alle betrokken predictoren wordt beschreven aan de hand van *conditional permutation-based variable importance* (Tagliamonte & Baayen 2012: 24). Predictoren aan de rechterkant van de rode lijn oefenen een significante invloed uit op de responsvariabele *realisatie* (*correct*, *niet correct*).

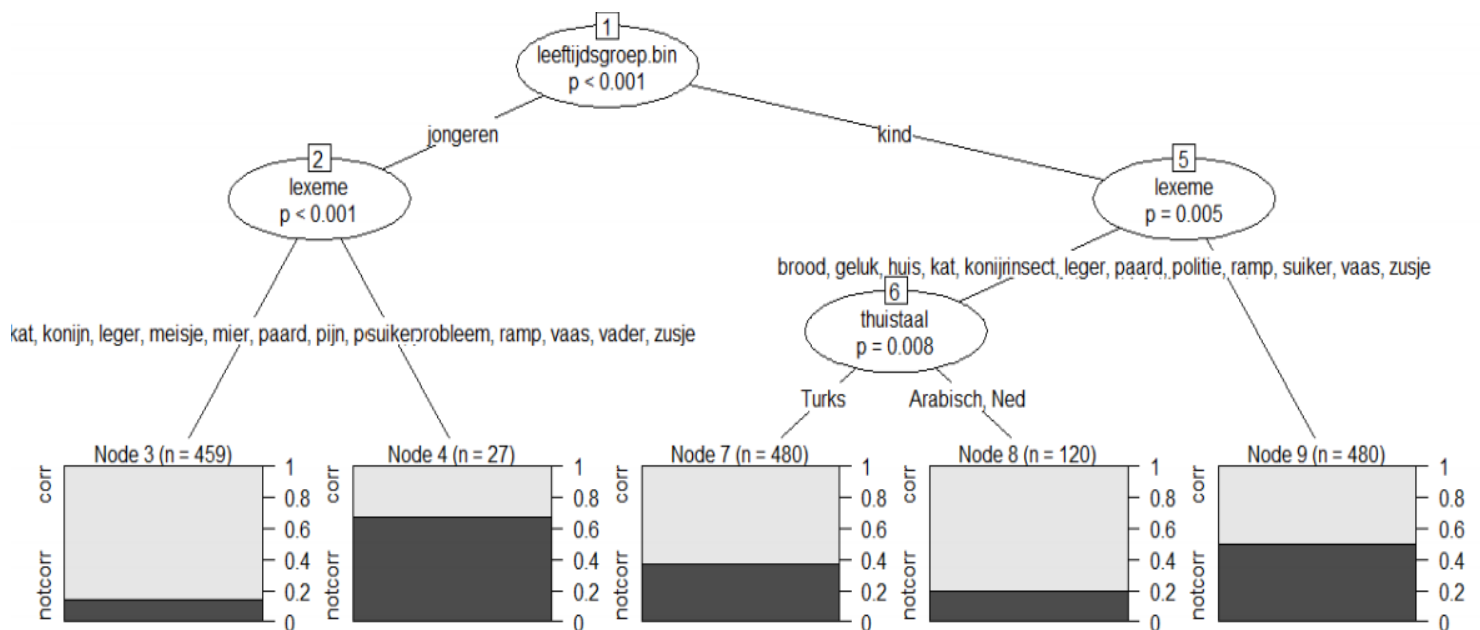


Figuur 18: Conditional permutation variable importance voor de random forest met alle predictoren (veldonderzoek)

Uit Figuur 18 kunnen we eerst en vooral afleiden dat *leeftijdsgroep* veruit de meeste invloedrijke predictor is. Die bevinding kwam ook reeds tijdens de exploratie van de dataset aan het licht (zie 4.2.1.), toen bleek dat de jongeren doorheen de semantische klassen, individuele testwoorden en *de- en het-woorden* steeds significant hoger scoorden dan de kinderen. De variabele die na *leeftijdsgroep* de belangrijkste invloed uitoefent op de responsvariabele, is *lexeme*. Eerder dan dat de semantische klasse (*semklasse*) of het grammaticaal geslacht van de testwoorden (*genus*) een invloed uitoefent op het gebruik van de standaardvorm van de determinator (beide predictoren zijn wel significant), blijken de individuele testwoorden (*lexeme*) een toonaangevende predictor te zijn. Wat de thuistalen

van de respondenten vervolgens betreft, zien we dat enkel de variabelen *taalvriend* en *bestlang* significant zijn. *Thuis taal* oefent daarentegen geen statistisch significant effect uit. Ook de predictor *geslacht* blijkt niet significant en is zelfs het minst invloedrijk. Bevindingen uit eerder onderzoek naar hersemantisering (o.a. De Vos 2009), die stelden dat er voor verandering in het Nederlandse genussysteem voorlopig weinig of geen significante verschillen tussen mannen en vrouwen zijn waargenomen, worden in dit onderzoek dus bevestigd.

Nu we weten welke predictoren een significante invloed uitoefenen op de responsvariabele *realisatie* (*correct*, *niet correct*), onderzoeken we in een volgende stap van de analyse hoe de betrokken predictoren samenwerken en elkaar beïnvloeden en bekijken we de *conditional inference tree* van de dataset (zie Figuur 19).



Figuur 19: Conditional inference tree (veldonderzoek)

Figuur 19 toont eerst en vooral het belang van de predictor *leeftijdsgroep*, aangezien de dataset op het hoogste niveau meteen wordt opgesplitst in een subset 'jongeren' en een subset 'kinderen'. Binnen die twee subsets blijkt ook de variabele *lexeme* een belangrijke invloed uit te oefenen, hoewel de toonaangevende testwoorden voor beide leeftijdsgroepen verschillend zijn. In de subset 'jongeren' wordt enerzijds een opsplitsing gemaakt tussen het testwoord *suiker* (rechts) en de 17 overige testwoorden (*vader, zusje, meisje, kat, paard, konijn, mier, insect, politie, leger, vaas, huis, ramp, probleem, pijn, geluk* en *brood*)(links). Het testwoord *suiker* lokt bij de jongeren duidelijk significant meer niet-standaardvormen uit dan de overige testwoorden. In de subset 'kinderen' wordt anderzijds een opsplitsing gemaakt tussen de testwoorden *brood, geluk, huis, kat, konijn, meisje, mier, pijn, probleem* en *vader* (links) en de testwoorden *insect, leger, paard, politie, ramp, suiker, vaas* en *zusje* (rechts). Die laatste groep testwoorden (rechts) wordt significant vaker gecombineerd met een niet-standaardvorm dan de woorden uit de andere subset (links). Binnen de linkse subset blijkt de variabele *thuis taal* bovendien wel relevant. Kinderen die het Turks als thuis taal hebben of die taal thuis met het Nederlands combineren, zouden bijgevolg significant lager scoren op 'grammaticaal genus' voor de testwoorden *brood, geluk, huis, kat, konijn, meisje, mier, pijn, probleem* en *vader* dan kinderen die thuis Arabisch (Marokkaans/Arabisch/Berber) en/of Nederlands

spreken. Het is echter moeilijk om aan de hand van deze analyse duidelijke uitspraken te doen over de variabele *thuis taal*, aangezien we door de complexe thuistaalsituaties van de respondenten verplicht waren om hun thuistalen te reduceren tot een beperkt aantal niveaus.

### 4.3. Conclusie

Gebaseerd op voorgaand onderzoek van De Vos en De Vogelaer (2011), werden in dit veldonderzoek drie leeftijdsgroepen ondervraagd over hun kennis van het grammaticale definiëte lidwoord of aanwijzend voornaamwoord van 18 opgegeven testwoorden. De kinderen (N=60) waren op het moment van de testafname tussen de 7 en 8 jaar oud, de adolescenten (N=14) waren 14 jaar oud en de jongvolwassenen (N=13) waren tussen de 18 en 21 jaar oud. Doorheen de analyse werden de twee oudste leeftijdsgroepen echter samengenomen in de leeftijdscategorie 'jongeren' (N=27). Alle respondenten woonden in de provincie Limburg en waren van allochtone afkomst. Aan de hand van dit apparent time-onderzoek wilden we een antwoord krijgen op verschillende onderzoeksvragen, hoewel de nadruk voornamelijk op taalverwerving lag.

Ten eerste wilden we aan de hand van dit onderzoek nagaan of sprekers van allochtone afkomst effectief vaak een ander lidwoord kiezen dan het Nederlandse genussysteem voorschrijft en of er sprake is van een overgeneralisering van het algemeen lidwoord *de*. De gemiddelde totaalscore van alle leeftijdsgroepen voor 'grammaticaal genus' in de taaltest bedroeg 11,98/18 (65%), wat aantoont dat er inderdaad heel wat niet-standaardvormen werden gebruikt. Hoewel de verkennende grafieken op het eerste gezicht aangaven dat er hoger werd gescoord op grammaticale *de-woorden* dan op grammaticale *het-woorden*, werd de overgeneralisering van het algemeen lidwoord *de* in de taaltest statistisch gezien niet bevestigd. De *random forest* toonde de variabele *genus* echter wel als een significante predictor, hoewel die een veel minder grote invloed bleek uit te oefenen dan alle andere significante predictoren.

Doorheen de bespreking van de resultaten werden twee verklaringen aangereikt voor de afwezigheid van een overgeneralisering van het algemeen lidwoord *de* in de taaltest. Een eerste mogelijkheid betrof de kans dat de respondenten zich min of meer bewust waren van de overgeneralisering van *de* in hun etnolect, waardoor ze er bij het invullen van een formele test rekening mee hielden of zelfs het lidwoord *het* overgeneraliseerden ter compensatie. Hoewel het onwaarschijnlijk lijkt dat ook kinderen zich al bewust zijn van etnolectisch taalgebruik, is het goed mogelijk dat zij de normcorrecties van hun leerkrachten gewaar worden. Een tweede mogelijkheid was daarentegen dat de tendens tot overgeneralisering van het lidwoord *het* in de twee laagst-geïndividualiseerde klassen de afwezigheid van significante verschillen tussen de scores op *de-woorden* en *het-woorden* in de hand werkte. De scores op enkele individuele testwoorden, zoals *meisje*, maakten de eerste verklaring echter plausibeler. In de formele taaltest werd er namelijk door alle leeftijdsgroepen betrekkelijk hoog gescoord op het testwoord *meisje*, terwijl *meisje* in het corpusonderzoek (chattaal en spontane interacties) vaak gecombineerd werd met een niet-standaardvorm. Tijdens de finale discussie komen we hier nog op terug (zie Hoofdstuk 5).

Ten tweede trachten we te onderzoeken of de bevindingen uit het corpusonderzoek inzake een hersemantiseringsproces in het adnominale domein doorheen dit veldonderzoek bevestigd werden. De verkennende grafiek van alle leeftijdsgroepen samen toonde ons dat niet-standaardrealisatie zich

binnen de vier hoogst-geïndividualiseerde klassen (*human, high animate, low animate* en *collective nouns*) hoofdzakelijk voordeed bij grammaticale *het-woorden* (overgeneralisering van *de*), terwijl niet-standaardrealisatie zich binnen de twee laagst-geïndividualiseerde klassen (*bounded object /abstract* en *unbounded abstract/unspecific mass*) grotendeels voordeed bij grammaticale *de-woorden* (overgeneralisering van *het*). Ook de verkennende grafieken van de leeftijdsgroepen 'kinderen' en 'jongeren' afzonderlijk toonden ons diezelfde tendens. Op het eerste gezicht leken die resultaten dus te pleiten in het voordeel van een hersemantiseringsproces volgens de *Individuation Hierarchy* van Audring (2006). Hoe hoger de referent zich op hiërarchie bevindt, hoe groter de kans wordt dat er voor het algemeen lidwoord *de* wordt gekozen. Hoe lager de referent zich daarentegen op de hiërarchie bevindt, hoe abstracter en minder geïndividualiseerd hij is en hoe vaker er met een onzijdig lidwoord *het* naar wordt verwezen. De statische analyse met de *random forest* en *conditional inference tree* toonde de variabelen *genus* en *semantische klasse* vervolgens wel als significante predictoren, maar ze bleken een relatief kleine invloed uit te oefenen op de responsvariabele *realisatie (correct, niet-correct)*, wanneer we vergeleken met de andere significante variabelen.

De variabele *lexeem*, die de afzonderlijke testwoorden representeerde, bleek daarentegen een grote invloed uit te oefenen op de responsvariabele. Eerder dan dat *genus* of *semantische klasse* niet-standaardrealisatie uitlokten, bleek het gebruik van een niet-standaardvorm in de taaltest afhankelijk van de individuele testwoorden. De woorden *leger* en *suiker* vertoonden daarbij opvallende scores. Wanneer de ene leeftijdsgroep relatief hoog scoorde op één van die twee testwoorden (de kinderen op *suiker* en de jongeren op *leger*), scoorde de andere leeftijdsgroep daarentegen opvallend laag. Opmerkelijk genoeg bleken *suiker* en *leger* de enige twee testwoorden die volgens het werk van Pauwels (1938) en het WNT lidwoordvariatie vertonen in (supra-)regionale variëteiten. Aangezien we tijdens het corpusonderzoek reeds hadden aangetoond dat die woorden in de autochtone corpora niet gecombineerd werden met een niet-standaardvorm van hun bijhorende determinator, meenden we dat de integratie van de woorden *leger* en *suiker* in de test niet problematisch was. Nu merken we echter dat de opvallende scores van *leger* en *suiker* in de taaltest de bevinding uit het corpusonderzoek bevestigen: woorden uit de lijst van Pauwels en het WNT lijken het sterkst bij te dragen tot hersemantisering. Die hersemantisering zou zich min of meer parallel voordoen in sommige dialecten en etnolecten. Eerder dan dat er bij de bevroegde respondenten sprake zou zijn van dialectinterferentie, zouden moedertaalsprekers door hun lidwoordvariatie de lidwoordkeuze van allochtone sprekers mogelijk beïnvloeden.

Ten derde wilden we aan de hand van dit apparent time-onderzoek nagaan of er effectief sprake is van een 'taalverwervingsprobleem' bij de ondervraagde allochtone respondenten en of de respondenten erin slagen om hoger te scoren op 'grammaticaal genus' naarmate ze ouder worden. Zijn er met andere woorden significante verschillen tussen de scores van de drie leeftijdsgroepen te vinden? Op basis van eerdere studies naar tweedetaalverwerving (Cornips & De Vogelaer 2009; Cornips 2008; Cornips, Van der Hoek & Verwer 2006) verwachtten we dat de kinderen vaker afwijkingen zouden vertonen van de standaardvorm dan de adolescenten en de jongvolwassenen en dat de adolescenten ook vaker niet-standaardvormen zouden gebruiken dan de jongvolwassenen. De statistische analyse toonde echter aan dat enkel het verschil tussen de scores van de kinderen en adolescenten enerzijds en het verschil tussen de scores van de kinderen en jongvolwassenen anderzijds significant waren. Aan de ene kant bevestigde dat onze hypothese en de bevindingen uit

eerdere studies, die stelden dat kinderen heel wat moeilijkheden ondervinden met de verwerving van het Nederlandse genussysteem en het zelfs op latere leeftijd niet volledig onder de knie hebben. De kinderen bleken immers significant lager te scoren op 'grammaticaal genus' dan de oudere leeftijdsgroepen en ook de oudere leeftijdsgroepen vertoonden – ondanks hun significant hogere score – nog relatief veel afwijkingen van de standaardvorm. Aan de andere kant bleken de scores van adolescenten niet significant te verschillen van de scores van de jongvolwassenen, wat tot de beslissing leidde om de adolescenten en jongvolwassenen samen te nemen in de leeftijdscategorie 'jongeren'. Een mogelijke verklaring daarvoor zou kunnen zijn dat de leeftijd van de twee groepen (14 jaar en 18-21 jaar) net iets te dicht bij elkaar lag om significante verschillen te bekomen. Wanneer de groep jongvolwassenen tussen de 25 en 30 jaar was geweest, hadden we hoogstwaarschijnlijk andere resultaten verkregen.

In het vijfde en laatste hoofdstuk zullen we vervolgens de bevindingen uit het corpusonderzoek en het veldonderzoek met elkaar vergelijken om te achterhalen in welke mate de onderzoeken methodologisch complementair zijn en hoe die complementariteit – en de multimethodologische onderzoeksopzet in het algemeen – nieuwe inzichten kan bieden en zo nieuwe perspectieven kan openen voor verder onderzoek naar overgeneralisering en hersemantisering in etnolecten.

## Hoofdstuk 5: Conclusie en discussie

In deze paper trachtten we een antwoord te formuleren op drie onderzoeksvragen. Ten eerste vroegen we ons af of etnolecten (*Marokkaans Nederlands* en *Citétaal*) op het vlak van genustoekenning afwijken van het taalgebruik van reguliere moedertaalsprekers en of er in etnolecten effectief sprake is van een overgeneralisering van het algemeen lidwoord *de* (onderzoeksvraag 1). Wanneer dat het geval zou zijn, wilden we ten tweede nagaan hoe groot de rol is van een mogelijk hersemantiseringsproces in het adnominale domein van etnolecten en autochtone variëteiten (onderzoeksvraag 2). Ten derde vroegen we ons af wat het aandeel is van taalverwerving. Wordt verandering in genustoekenning in etnolecten voortgestuwd door taalverwervers die het grammaticale genussysteem van het Nederlands moeilijk onder de knie krijgen of spelen andere (sociale) factoren eerder een rol (onderzoeksvraag 3)?

Om een antwoord te kunnen formuleren op bovenstaande onderzoeksvragen en om een omvattend inzicht te kunnen krijgen in de kwestie hadden we nood aan een multimethodologische onderzoeksopzet. Daarom werden zowel een corpusonderzoek als een taalverwervingsexperiment (veldonderzoek) uitgevoerd. Aan de hand van het corpusonderzoek probeerden we hoofdzakelijk een antwoord te formuleren op de twee eerste onderzoeksvragen, terwijl het taalverwervingsexperiment eerder focuste op de derde onderzoeksvraag. Dat nam niet weg dat de resultaten uit beide onderzoeken elkaar konden aanvullen en op die manier nieuwe inzichten konden bieden. Bovendien bood de combinatie van de twee onderzoeksmethodes ons de mogelijkheid om de etnolectische taalvariëteiten in verschillende onderzoekssettings te bestuderen. De taaldata uit het corpusonderzoek waren namelijk afkomstig uit gesprekken die door allochtone jongeren in informele omstandigheden (op chatkanalen of in spontane interacties onder vrienden) werden gevoerd, terwijl we de data uit het veldonderzoek aan de hand van een formele taaltest verzamelden.

### 5.1. Eerste onderzoeksvraag

Op basis van voorgaande sociolinguïstische studies naar variatie in het Nederlandse genussysteem en op basis van taalverwervingsonderzoeken (Cornips 2008; Nortier & Dorleijn 2008; Marzo & Ceuleers 2011; Cornips, Van der Hoek & Verwer 2006; Cornips & De Vogelaer 2009) stelden we als hypothese voor de eerste onderzoeksvraag dat etnolecten (*Marokkaans Nederlands* en *Citétaal*) op het vlak van genustoekenning afwijken van het taalgebruik van reguliere moedertaalsprekers en dat er sprake is van een overgeneralisering van het algemeen lidwoord *de*. Doorheen het corpusonderzoek werd die hypothese bevestigd. De etnolectische corpora (*Moroccorp* en *Citétaalcorpus*) bleken immers significant meer niet-standaardvormen te bevatten dan de autochtone corpora (*Condiv*) en niet-standaardrealisatie bleek significant vaker op te treden bij grammaticale *het-woorden* dan bij grammaticale *de-woorden*, wat duidt op een overgeneralisering van het lidwoord *de* in etnolecten. Uit de logistische regressieanalyse van het corpusonderzoek bleek bovendien dat de variabelen *Genus* (*de* vs. *het*) en *Etnolect* de sterkste invloed uitoefenden op de responsvariabele. Ook de *random forest* en *conditional inference tree* van het corpusonderzoek bevestigden de overgeneralisering van *de* in etnolecten. Wel bleek dat aanwijzende voornaamwoorden (bv. *die meisje*) significant meer niet-standaardvormen toelaten dan definitieve lidwoorden (bv. *de meisje*).

Doorheen het veldonderzoek gingen we vervolgens na of de bevindingen uit het corpusonderzoek bevestigd werden. Ten eerste bleek dat de bevroegde respondenten van allochtone afkomst effectief vaak een ander lidwoord kozen dan het Nederlandse genussysteem voorschrijft. De gemiddelde totaalscore op de taalttest bedroeg immers ‘slechts’ 65%. De overgeneralisering van het algemeen lidwoord *de* werd doorheen het taalverwervingsexperiment echter niet bevestigd. De bevroegde respondenten bleken namelijk niet significant vaker te opteren voor een niet-standaardvorm bij grammaticale *het-woorden* dan bij grammaticale *de-woorden*. Dat resultaat was eerder onverwacht, aangezien de verkennende grafieken die tendens (overgeneralisering van *de*) op het eerste gezicht wel toonden en ook de *random forest* van het veldonderzoek de variabele *genus* als een significante predictor aangaf. Daarbij moeten we wel opmerken dat *genus* volgens de *random forest* de minst invloedrijke significante predictor was.

Doorheen het veldonderzoek werden verschillende verklaringen aangereikt voor het afwezigheid van significantie van de overgeneralisering van het algemeen lidwoord *de* in de taalttest. Bij het bekijken van de scores op de individuele testwoorden kwamen we tot de conclusie dat het erg plausibel leek dat de bevroegde respondenten zich (bij sommige woorden) min of meer bewust waren van een overgeneralisering van het lidwoord *de* en er daardoor tijdens de formele taalttest rekening mee hielden. Jongeren van allochtone afkomst zouden zich met andere woorden min of meer bewust kunnen zijn van de stigmatiserende oordelen over hun taalgebruik en kinderen van allochtone afkomst zouden in zekere mate de normcorrecties van hun leerkrachten gewaar kunnen worden.

Onderstaande tabel bevat een overzicht van de scores ‘grammaticaal genus’ uit het *Morroccorp* en *Citétaalcorpus* (corpusonderzoek) enerzijds, en van de scores ‘grammaticaal genus’ uit de taalttest van de kinderen en de jongeren (veldonderzoek) anderzijds. Aangezien de sprekers uit het *Morroccorp* en *Citétaalcorpus* jongeren van allochtone afkomst zijn, kunnen we hun percentages ‘grammaticaal genus’ in zekere mate vergelijken met de percentages van de jongeren uit de taalttest. Een opvallende vaststelling bij het bekijken van de individuele testwoorden is dat wanneer de jongeren uit het *Morroccorp* of *Citétaalcorpus* voor bepaalde woorden in informele omstandigheden een score van minder dan 50% op ‘grammaticaal genus’ behalen (*meisje*, *zusje*, *konijn*, *brood*), de jongeren in de taalttest op diezelfde woorden net een zeer hoge score blijken te behalen. Het testwoord *meisje* behaalde in het corpusonderzoek bijvoorbeeld slechts een score van 36,39% (*Morroccorp*) en 18,18% (*Citétaalcorpus*) op ‘grammaticaal genus’, terwijl 93% van de jongeren het woord *meisje* in de taalttest combineerde met zijn grammaticaal lidwoord. De woorden *meisje*, *zusje*, *konijn* en *brood* blijken bovendien opmerkelijk genoeg allemaal onzijdige *het-woorden* te zijn (overgeneralisering van *de* in de corpora), wat doet vermoeden dat de overgeneralisering van *de* in etnolecten toch niet zo’n zwart-wit kwestie is als op het eerste gezicht lijkt. De resultaten uit het veldonderzoek suggereren immers dat de jongeren in een formele taalttest wel ‘weten’ dat die woorden het lidwoord *het* nemen. Zo dadelijk komen we nog op deze kwestie terug (zie 5.3.).

	Morroccorp	Citétaalcorpus	Taalttest kinderen	Taalttest jongeren
vader	100%		73%	100%
<b>meisje</b>	<b>36,39%</b>	<b>18,18%</b>	<b>75%</b>	<b>93%</b>
<b>zusje</b>	<b>39,02%</b>		<b>52%</b>	<b>81%</b>
kat	98,72%	100%	67%	89%
<b>konijn</b>	<b>1,28%</b>	<b>0%</b>	<b>62%</b>	<b>70%</b>



paard	51,85%		48%	93%
mier	100%		60%	81%
insect	100%		55%	67%
politie	100%	100%	57%	96%
leger	85%		35%	78%
vaas	75%		50%	78%
huis	78,43%	100%	67%	89%
ramp	100%		52%	89%
probleem	85,91%	100%	70%	96%
pijn	100%		60%	81%
geluk	90,91%		67%	93%
suiker	78,95%	100%	53%	33%
<b>brood</b>	<b>44%</b>		<b>62%</b>	<b>89%</b>

Tabel 14: Percentages 'grammaticaal genus' per testwoord uit het *Morocccorp*, *Citétaalcorpus* en de taaltest (kinderen en jongeren)

## 5.2. Tweede onderzoeksvraag

De tweede onderzoeksvraag peilde vervolgens naar rol van het hersemantiseringsproces in de genustoekenning van adnominale elementen in etnolecten en autochtone variëteiten. Daarbij vroegen we ons af of de mogelijkheid bestaat dat de overgeneralisering van *de* in etnolecten deel uitmaakt van een groter, zich ontwikkelend proces van hersemantisering volgens de *Individuation Hierarchy* van Audring (2006), zoals dat ook in het Noord-Nederlandse pronominale systeem is aangetoond. Wanneer er sprake zou zijn van zo'n hersemantiseringsproces in het adnominale domein, zou overgeneralisering van *de* hoofdzakelijk moeten optreden in de hoogst-geïndividualiseerde klassen van de *Individuation Hierarchy*, terwijl er in de laagst-geïndividualiseerde klassen bij de meer abstracte referenten eerder een overgeneralisering van *het* zou moeten plaatsvinden.

Uiteenlopende bevindingen uit verschillende taalkundige perspectieven leidden ertoe dat we bij deze onderzoeksvraag twee hypotheses formuleerden. Ten eerste stelden we dat zowel etnolecten als autochtone variëteiten mogelijk onderhevig zijn aan een hersemantiseringsproces in het adnominale domein, maar dat de hersemantisering in etnolecten al duidelijker zichtbaar zou zijn. De universele *Agreement Hierarchy* van Corbett (1979) en onderzoek van Fernández-Ordóñez (2009) deden immers vermoeden dat het hersemantiseringsproces zich ook in het Nederlands vanuit het pronominale domein verder zou uitbreiden naar andere woordklassen, zoals de definiëte lidwoorden. Van der Sijs (2014) haalde bovendien aan dat contactvariëteiten en etnolecten een belangrijke rol spelen bij taalverandering, wat het plausibel maakt dat ze daardoor gemakkelijker onderhevig zijn aan een hersemantiseringsproces dan autochtone variëteiten.

Ten tweede stelden we dat de mogelijkheid bestaat dat de overgeneralisering van het algemeen lidwoord *de* een inherent kenmerk is van de twee bestudeerde etnolecten en dat er weinig of geen sprake is van een hersemantiseringsproces in het lidwoordensysteem van *Citétaal* en *Marokkaans Nederlands*. Op basis van taalverwervingsonderzoek (Van der Velde 2003; Cornips & Hulk 2008; Cornips & De Vogelaer 2009), waarin de overgeneralisering van het algemeen lidwoord *de* bij eerste- en tweedetaalverwerfers wordt toegeschreven aan de grote moeilijkheden die met de verwerving

van het Nederlandse lidwoordensysteem gepaard gaan, kunnen we immers stellen dat de keuze voor definiete lidwoorden en aanwijzende voornaamwoorden in etnolecten eerder willekeurig verloopt, met een voorkeur voor het commuun geslacht (*de, die*) wanneer het grammaticale geslacht van de referent niet is gekend.

Om een antwoord te formuleren op de tweede onderzoeksvraag en om na te gaan welke van de twee hypothesen correct blijkt, baseerden we ons hoofdzakelijk op de resultaten uit het corpusonderzoek. Daaruit bleek dat de integratie van woorden met een alternerend lidwoord in (supra-)regionale variëteiten andere resultaten opleverde dan wanneer we die woorden uit de dataset weglieten. Uit de dataset zonder de twaalf woorden die volgens de studie van Pauwels (1938) en het WNT (De Vries & Te Winkel 1864-1998) lidwoordvariatie vertonen in (supra-)regionale variëteiten, konden we afleiden dat er voor alle taaldata uit de vier corpora samen een significante invloed bestond van semantische klasse op standaardrealisatie. De hoogst-geïndividualiseerde semantische klassen *human* en *high animate* bleken namelijk significant vaker niet-standaardrealisatie uit te lokken dan de vier andere semantische klassen (*low animate, collective nouns, bounded object/abstract, unbounded abstract/unspecific mass*). De statistische analyse van de etnolectische taaldata wees echter uit dat het effect van niet-standaardrealisatie enkel aanwezig was bij *het-woorden*, wat alleen de overgeneralisering van *de* in etnolecten bevestigde en dus niet pleitte in het voordeel van een hersemantiseringsproces.

Dat deed vervolgens de vraag rijzen of de integratie van de twaalf woorden met een alternerend lidwoordgebruik een invloed zou uitoefenen op de resultaten. Daarom gingen we na welke van die twaalf zelfstandige naamwoorden in de autochtone corpora voorkwamen met een niet-standaardvorm van de bijhorende determinator. Dat bleek slechts om vijf woorden te gaan: *boek, kleur, raam, winter* en *zand*. Enkel het woord *raam* bleek bovendien in het Nederlandse deel van het *Condiv* met een niet-standaardvorm gecombineerd te worden. Daarom werd er besloten om een nieuwe dataset aan te maken zonder het woord *raam*, maar inclusief de 11 andere woorden met alternerend lidwoordgebruik. De redenering was dat wanneer de woorden uit de lijst van Pauwels en het WNT in de autochtone corpora niet voorkwamen met een niet-standaardvorm, variatie in lidwoordtoekenning in de allochtone corpora bijna onmogelijk kon worden toegeschreven aan dialectinterferentie.

Opnieuw wezen de statistische testen uit dat niet-standaardrealisatie in de etnolectische corpora hoofdzakelijk optrad bij grammaticale *het-woorden* van de hoogst-geïndividualiseerde klassen. De Fischer Exact Test wees deze keer echter wel op een significante invloed van semantische klasse op standaardrealisatie binnen de grammaticale *de-woorden*. In de laagst-geïndividualiseerde klasse *unbounded abstract/unspecific mass* konden we met andere woorden in de etnolectische corpora een lichte tendens tot hersemantisering van *het* vaststellen. Aangezien we nu aan beide polen van de *Individuation Hierarchy* in de etnolectische variëteiten een tendens tot hersemantisering konden waarnemen (overgeneralisering van *de* in de hoogste klassen en overgeneralisering van *het* in de laagste klasse), is het mogelijk dat het hersemantiseringsproces zijn intrede doet in het adnominale domein van etnolecten. Wel moeten we daarbij opmerken dat dat hersemantiseringseffect enkel waarneembaar is wanneer we de woorden uit de lijst van Pauwels en het WNT meetellen. Net door die woorden tijdens de eerste analyse uit de dataset te filteren, kregen we mogelijk een onderschatting van het geringe hersemantiseringseffect in etnolecten te zien.

Doorheen het veldonderzoek trachtten we vervolgens na te gaan of bovenstaande bevindingen uit het corpusonderzoek bevestigd werden. Zowel de verkennende grafiek van alle leeftijdsgroepen samen als die van de kinderen en jongeren afzonderlijk toonden aan dat niet-standaardrealisatie zich binnen de vier hoogst-geïndividualiseerde klassen (*human, high animate, low animate* en *collective nouns*) hoofdzakelijk voordeed bij grammaticale *het-woorden* (overgeneralisering van *de*), terwijl niet-standaardrealisatie zich binnen de twee laagst-geïndividualiseerde klassen (*bounded object /abstract* en *unbounded abstract/unspecific mass*) hoofdzakelijk voordeed bij grammaticale *de-woorden* (overgeneralisering van *het*). Ook de resultaten van het veldonderzoek leken dus te pleiten in het voordeel van een hersemantiseringsproces volgens de *Individuation Hierarchy* van Audring (2006). De statistische analyse (*random forest* en *conditional inference tree*) bevestigde bovendien de variabelen *genus* en *semantische klasse* als significante predictoren, hoewel ze in vergelijking met de andere significante variabelen een relatief kleine invloed uitoefenden op de responsvariabele *realisatie (correct, niet-correct)*.

De variabele die volgens de *random forest* van het veldonderzoek meer invloed uitoefende op de responsvariabele dan de predictoren *genus* en *semantische klasse*, was de variabele *lexeem*, die de afzonderlijke testwoorden representeerde. Het gebruik van niet-standaardvormen in de taaltest werd dus met andere woorden hoofdzakelijk uitgelokt door de individuele testwoorden. Twee woorden die daarbij opmerkelijke scores vertoonden, waren de woorden *leger* en *suiker*. Zoals in Tabel 14 duidelijk te zien is, behaalden de kinderen een zeer lage score voor het testwoord *leger*, terwijl de jongeren – in lijn met hun score uit het corpusonderzoek – eerder een hoge score behaalden. De jongeren scoorden daarentegen opmerkelijk laag op het testwoord *suiker*, wanneer we vergeleken met de score van de kinderen en met de score uit het corpusonderzoek. Opvallend genoeg bleken de testwoorden *leger* en *suiker* net twee woorden die volgens de studie van Pauwels (1938) en het WNT (De Vries & Te Winkel 1864-1998) lidwoordalternantie vertonen in regionale variëteiten, maar in de autochtone corpora niet voorkwamen met een niet-standaardvorm. Omdat we ook doorheen het corpusonderzoek een soortgelijke tendens vaststelden, lijkt het erop dat de woorden uit de lijst van Pauwels en het WNT het sterkst bijdragen tot hersemantisering. Aangezien dialectinterferentie in de etnolectische taalvariëteiten vrij onwaarschijnlijk lijkt (cfr. supra), zou die hersemantisering zich min of meer parallel voordoen in sommige dialecten en etnolecten. Doorheen het veldonderzoek reikten we de mogelijkheid aan dat de lidwoordkeuze van allochtone sprekers bij woorden met lidwoordvariatie in regionale variëteiten beïnvloed zou worden door de variatie van moedertaalsprekers (taalexterne/sociale beïnvloeding). Een andere verklaring is echter dat die woorden door bepaalde, inherente eigenschappen zelf lidwoordvariatie uitlokken in dialecten en etnolecten afzonderlijk (taalinterne beïnvloeding). Om na te gaan welke verklaring correct blijkt of om eventuele andere hypothesen te testen, is verder onderzoek naar woorden met lidwoordalternantie in (supra-)regionale variëteiten noodzakelijk.

Uit het corpusonderzoek en het veldonderzoek kunnen we samenvattend concluderen dat we in etnolecten een lichte tendens tot hersemantisering kunnen vaststellen, die voornamelijk wordt uitgelokt door woorden met een alternerend lidwoordgebruik in (supra-)regionale variëteiten. Hoewel de eerste hypothese van deze onderzoeksvraag dus in zekere mate bevestigd werd, moeten we voorzichtig omspringen met deze conclusie. Een belangrijke opmerking betreft immers het feit dat ook lang niet alle lidwoordvariatie in etnolecten verklaard kan worden aan de hand van de *Individuation Hierarchy* van Audring (2006). Zo bleken er in het taalverwervingsexperiment immers

constructies voor de komen als *het vader, het kat, de brood* of *de geluk*, die in principe geen enkele semantische motivering hebben. In de hoogst-geïndividualiseerde klassen (*vader* en *kat*) verwachten we volgens de hersemantiserings-theorie immers het lidwoord *de* en in de laagst-geïndividualiseerde klassen (*brood* en *geluk*) het lidwoord *het*. Het voorkomen van zulke constructies roept ook op tot verder onderzoek.

### 5.3. Derde onderzoeksvraag

In de derde onderzoeksvraag onderzochten we ten slotte in welke mate taalverwerving een rol speelt bij lidwoordvariatie in etnolecten. Concreet vroegen ons af of verandering in genustoekenning in etnolecten wordt voortgestuwd door tweedetaalverwerpers die het Nederlandse genussysteem moeilijk onder de knie krijgen en of andere, sociale factoren (bv. groepsaffiliatie) ook een rol spelen. Op basis van literatuur uit de sociolinguïstiek en taalverwervingsonderzoek (Van der Velde 2003; Cornips & Hulk 2008; Cornips & De Vogelaer 2009; Cornips 2008) verwachtten we dat zowel taalverwervingsaspecten als sociale factoren mee aan de basis zouden liggen van de overgeneralisering van het algemeen lidwoord *de* in etnolecten. We verwachtten dat kinderen van allochtone afkomst effectief moeilijkheden zouden ervaren met de verwerving van het Nederlandse lidwoordensysteem, maar dat die problemen gedeeltelijk zouden verdwijnen naarmate ze ouder worden. Jongeren van allochtone afkomst zouden het grammaticale lidwoordensysteem van het Nederlands vervolgens nog niet volledig onder de knie hebben, maar ze zouden zich wel steeds bewuster worden van het feit dat buitenstaanders hen op basis van de overgeneralisering van het lidwoord *de* als 'allochtoon' bestempelen. Daarom zouden ze het lidwoord *de* in bepaalde contexten (bewust) overgeneraliseren om op die manier hun groepslidmaatschap te benadrukken.

Om een antwoord te kunnen formuleren op deze onderzoeksvraag werd een taalverwervings-experiment uitgevoerd, waarin we gebruik maakten van de apparent time-methode. Drie verschillende leeftijdsgroepen werden bevraagd: kinderen (7-8 jaar), adolescenten (14 jaar) en jongvolwassenen (18-21 jaar). De statistische analyse (*random forest* en *conditional inference tree*) toonde aan dat de scores 'grammaticaal genus' van de kinderen op de taalttest significant lager waren dan de scores van de adolescenten en de jongvolwassenen. Het verschil tussen de scores van de adolescenten en de jongvolwassenen bleek daarentegen niet significant.

Het eerste deel van de hypothese werd met andere woorden grotendeels bevestigd. Doordat de kinderen significant lager bleken te scoren dan de oudere leeftijdsgroepen (doorheen de verschillende semantische klassen, individuele testwoorden en *de-* en *het-woorden*), lijkt het vrij waarschijnlijk dat een taalverwervingsprobleem bij de jongste groep aan de basis ligt van hun lidwoordvariatie. Zoals Cornips & Hulk (2008) in hun studie aanhaalden, lijken inputfactoren dus wel degelijk een rol te spelen bij het aanleren van genusonderscheiding in het Nederlandse lidwoordensysteem. Het belang van duur van input werd doorheen dit onderzoek bevestigd, aangezien de oudere leeftijdsgroepen de hoogste scores behaalden op 'grammaticaal genus'. Of woordfrequentie een invloed uitoefent op de snelheid waarmee het grammaticale lidwoord wordt verworven (Brouwer et al. 2008), werd in het veldonderzoek niet expliciet getest. Woordfrequentie zou echter wel een mogelijke verklaring kunnen bieden voor de lage scores van de kinderen op niet-frequente testwoorden als *leger* of *ramp* (zie Tabel 14). Ook daarvoor is verder onderzoek noodzakelijk.

Wat vervolgens doorheen het onderzoek ook bevestigd werd, was het feit dat jongeren van allochtone afkomst het grammaticale lidwoordensysteem van het Nederlands nog steeds niet volledig onder de knie bleken te hebben. Hoewel de twee oudste leeftijdsgroepen significant hoger scoorden op 'grammaticaal genus' dan de kinderen, vertoonden ook zij nog relatief veel afwijkingen van de standaardvorm. Een bijkomend onderzoek met een oudere leeftijdsgroep (25-30 jaar) zou kunnen uitwijzen of er effectief sprake is van een fossilisatie-effect bij sprekers van allochtone afkomst, zoals Cornips (2008: 109) in haar studie aanhaalt. Bovendien zou een bijkomende vergelijking met de testresultaten van autochtone jongeren kunnen aantonen of autochtone jongeren effectief significant hoger scoren op 'grammaticaal genus' dan allochtone respondenten.

Op het laatste onderdeel van de onderzoeksvraag (oefenen sociale factoren een invloed uit op de lidwoordvariatie van sprekers van allochtone afkomst?) kunnen we aan de hand van onze data geen sluitend antwoord bieden, omdat we daarvoor ook andere interactionele data nodig hebben. Wel hebben we, zoals reeds aangegeven bij de bespreking van de eerste onderzoeksvraag, een aantal opmerkelijke bevindingen gedaan tijdens de vergelijking van de resultaten uit het corpusonderzoek en het veldonderzoek. Wanneer de jongeren uit het *Moroccorp* of *Citétaalcorpus* voor bepaalde onzijdige woorden in informele omstandigheden een betrekkelijk lage score behaalden op 'grammaticaal genus', scoorden ze voor die individuele testwoorden daarentegen opmerkelijk hoog in de formele taaltest. Voornamelijk de score van het woord *meisje* viel daarbij op (36,99% (*Moroccorp*) en 18,18% (*Citétaalcorpus*) in het corpusonderzoek en 93% in het veldonderzoek), aangezien dat woord door de media en literatuur vaak wordt aangehaald als stereotype van etnolectisch taalgebruik. We haalden daarom aan dat het lijkt alsof de jongeren in een formele taaltest wel 'weten' dat die woorden het onzijdig lidwoord *het* nemen, maar dat ze in bepaalde informele omstandigheden (al dan niet bewust) kiezen voor het lidwoord *de*.

Aangezien een taaltest steeds een redelijk formele vorm van dataverzameling is, waarin de respondenten een soort van meta-talige oefening maken en zich bijgevolg erg 'bewust' zijn van hun taalgebruik, konden we aan de hand van deze onderzoeksmethode nagaan wat de respondenten doen in een dergelijke formele context: kiezen ze voor het semantische genus van de opgegeven testwoorden of verkiezen ze net de grammaticale determinator? Doordat de verschillen tussen de resultaten uit de formele taaltest (veldonderzoek) en de informele interacties (corpusonderzoek) groot bleken, zouden sociale factoren als groepslidmaatschap een rol kunnen spelen.

Hoewel we doorheen deze paper een antwoord konden formuleren op het merendeel van de onderzoeksvragen, blijven nog heel wat kwesties onbeantwoord. Daarom willen we voornamelijk aansporen tot verder onderzoek waarin het volledige lidwoordensysteem van etnolecten in kaart wordt gebracht. De exploratie van andere onderzoeksmethodes is daarbij belangrijk. Niet alleen kan interactioneel onderzoek naar het lidwoordgebruik (werken met authentieke, interactionele data) een duidelijk inzicht bieden, maar ook eventueel ander experimenteel onderzoek, waarbij jongeren in een formele spreek situatie worden aangespoord om bepaalde lidwoorden te gebruiken, kan duidelijk maken in welke contexten de respondenten meestal kiezen voor het grammaticale genus. Ook zouden we willen aansporen tot uitbreiding van het *Citétaalcorpus* en tot een taalverwervings-experiment waarin ook oudere en autochtone leeftijdsgroepen worden bevraagd. Op die manier zouden we een beter inzicht kunnen verwerven in zowel het lidwoordensysteem van allochtone variëteiten als dat van autochtone variëteiten.



## Referenties

- Audring, J. (2006). Pronominal gender in spoken Dutch. *Journal of Germanic Linguistics*, 18(2), 85-116. doi: 10.1017/S1470542706000043
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi: 10.18637/jss.v067.i01
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32. doi: 10.1023/A:1010933404324
- Brouwer, S., Cornips, L., & Hulk, A. (2008). Misrepresentation of Dutch neuter gender in older bilingual children? In G. Haznedar, & E. Gavrusseva (Eds.), *Current trends in child second language acquisition: a generative perspective* (pp. 83-96). Amsterdam: John Benjamins.
- Brugman, M. (2007). *Een onderzoek naar de verwerving van grammaticaal geslacht en telbaarheid bij eentalige en meertalige kinderen van vier, vijf en zes jaar*. Universiteit Utrecht, Faculteit Geesteswetenschappen.
- Corbett, G. G. (1979). The agreement hierarchy. *Journal of Linguistics*, 15(2), 203-224. doi: 10.1017/S0022226700016352
- Cornips, L. (2002). Etnisch Nederlands. In H. Bennis, G. Extra, P. Muysken, & J. Nortier (Eds.), *Een buurt in beweging: talen en culturen in het Utrechtse Lombok en Transvaal* (pp. 285-302). Amsterdam: Askant.
- Cornips, L. (2005). Het Surinaams-Nederlands in Nederland. In N. Van der Sijs (Eds.), *Wereldnederlands: oude en jonge variëteiten van het Nederlands* (pp. 131-147). Den Haag: SDU.
- Cornips, L. (2008). Loosing grammatical gender in Dutch: the result of bilingual acquisition and/or an act of identity? *International Journal of Bilingualism*, 12(1&2), 105-124. doi: 10.1177/13670069080120010701
- Cornips, L., & De Vogelaer, G. (2009). Variatie en verandering in het Nederlandse genus: een multidisciplinair perspectief. *Taal & Tongval*, 61(1), 1-12. doi: 10.5117/TET2009.1.CORN

- Cornips, L., & Hulk, A. (2008). Factors of success and failure in the acquisition of grammatical gender in Dutch. *Second Language Research*, 24(3), 267-295. doi: 10.1177/0267658308090182
- Cornips, L., Van der Hoek, M., & Verwer, R. (2006). The acquisition of grammatical gender in bilingual child acquisition of Dutch (by older Moroccan and Turkish children): the definite determiner, attributive adjective and relative pronoun. In B. Los, & J. van de Weijer (Eds.), *Linguistics in The Netherlands* (pp. 40-51). Amsterdam: John Benjamins.
- Den Boon, T. (editor in chief), Geeraerts, D., & Van der Sijs, N. (2005). *Van Dale Groot Woordenboek van de Nederlandse Taal (veertiende, herziene uitgave)*. Utrecht/Antwerpen: Van Dale Lexicografie.
- De Paepe, J., & De Vogelaer, G. (2008). Grammaticaal genus en pronominale verwijzing bij kinderen: een taalverwervingsperspectief op een eeuwenoud grammaticaal probleem. *Neerlandistiek.nl*, 8(2), 1-23.
- De Rooij, V. (2005). *De studie van etnolecten: overdenkingen vanuit een etnografisch perspectief [The study of etnolects. Reflections from an ethnographic perspective]*. Talk presented at the Workshop Etnisch Nederlands, Utrecht University, Utrecht, The Netherlands.
- De Vogelaer, G. (2006). Pronominaal genus bij 'Zuid-Nederlandse' taalverwervers: grammaticaal of semantisch systeem? In M. Hüning, A. Verhagen, U. Vogl, & T. van der Wouden (Eds.), *Nederlands tussen Duits en Engels* (pp. 89-102). Leiden: Stichting Neerlandistiek Leiden.
- De Vogelaer, G. (2009). Changing pronominal gender in Dutch: transmission or diffusion? In S. Tsiplakou, M. Karyolemou, & P. Pavlou (Eds.), *Language variation, European perspectives II: selected papers from the 4th International Conference on Language Variation in Europe* (Vol. 5, pp. 71-80). Amsterdam & Philadelphia: John Benjamins.
- De Vogelaer, G. (2010). (Not) acquiring grammatical gender in two varieties of Dutch. In D. Geeraerts, G. Kristiansen, & Y. Peirsman (Eds.), *Advances in cognitive sociolinguistics* (pp. 167-190). Berlijn: Mouton de Gruyter.



- De Vos, L. (2009). De dynamiek van hersemantisering. *Taal & Tongval*, 61(1), 82-110. doi: 10.5117/TET2009.1.DEVO
- De Vos, L., & De Vogelaer, G. (2011). Dutch gender and the locus of morphological regularization. *Folio Linguistica*, 45, 245-281. doi: 10.1515/flin.2011.011
- De Vries, M., & Te Winkel, L. A. (1864-1998). *Woordenboek der Nederlandsche Taal*. 's Gravenhage: Nijhoff.
- Donaldson, B. (1993). *A grammar of Afrikaans*. Berlijn: Mouton de Gruyter.
- Fernández-Ordóñez, I. (2009). The development of mass/count distinctions in Indo-European languages. In V. Bubenik, J. Hewson, & S. Rose (Eds.), *Grammatical change in Indo-European languages* (pp. 55-68). Amsterdam/Philadelphia: John Benjamins.
- Fought, C. (2006). *Language and ethnicity*. Cambridge: Cambridge University Press.
- Franco, K. (2012). *'Het pronominale probleem': een corpusonderzoek naar het pronominale genus in de spreektaal in Vlaanderen en Nederland* (Unpublished master's thesis). Katholieke Universiteit Leuven, Faculteit Letteren.
- Franco, K. (2013). *Gender puzzles. Adnominal and pronominal agreement with neuter nouns in Southern Dutch* (Unpublished master's thesis). Katholieke Universiteit Leuven, Faculteit Letteren.
- Geeraerts, D. (1992). Pronominale masculiniseringsparameters in Vlaanderen. In H. Bennis, & J. W. De Vries (Eds.), *De binnenbouw van het Nederlands: een bundel artikelen voor Piet Paardekoper* (pp.73-84). Dordrecht: ICG Publications.
- Geeraerts, D. (2001). Een zondagspak? Het Nederlands in Vlaanderen: gedrag, beleid, attitudes. *Ons Erfdeel*, 44(3), 337-344.
- Geerts, G. (1966). *Genus en geslacht in de Gouden Eeuw: een bijdrage tot de studie van de nominale klassifikatie en daarmee samenhangende adnominale flexievormen en pronominale verschijnselen in Hollands taalgebruik van de zeventiende eeuw*. Brussel: Belgisch Interuniversitair Centrum voor Neerlandistiek.

- Grondelaers, S., Deygers, K., Van Aken, H., Van den Heede, V., & Speelman, D. (2000). DigiTaal: het CONDIV-corpus geschreven Nederlands. *Nederlandse Taalkunde*, 5(4), 356-363.
- Haeseryn, W., Romijn, K., Geerts, G., De Rooij, J. & Van Den Toorn, M. (1984). *Algemene Nederlandse Spraakkunst*. Groningen: Martinus Nijhoff/Deurne: Wolters Plantijn.
- Hoppenbrouwers, C. (1983). Het genus in een Brabants regiolect. *TABU (Bulletin voor Nederlandse taalkunde)*, 13, 1-25.
- Hothorn, T., Buehlmann, R., Dudoit, S., Molinaro, A., & Van Der Laan, M. (2006). Survival ensembles. *Biostatistics*, 7(3), 355-373. doi: 10.1093/biostatistics/kxj011
- Klom, J., & De Vogelaer, G. (To appear, 2016). Semantic agreement competing with syntactic agreement: the case of Dutch pronouns. *Leuvense bijdragen – Leuven Contributions in Leuven Contributions in Linguistics and Philology*.
- Köpcke, K., & Zubin, D. (1996). Prinzipien für die Genuszuweisung im Deutschen. In E. Lang, & G. Zifonun (Eds.), *Deutsch – typologisch* (pp.473-491). Berlin: Mouton de Gruyter.
- Kotsinas, U. (2001). Pidginization, creolization, and creoloids in Stockholm, Sweden. In N. Smith, & T. Veenstra (Eds.), *Creolization and contact* (pp. 125-156). Amsterdam: John Benjamins.
- Kraaikamp, M. (2012). The semantics of the Dutch gender system. *Journal of Germanic Linguistics*, 24(3), 193-232. doi: 10.1017/S1470542712000074
- Labov, W. (1994). *Principles of linguistic change 1: internal factors*. Oxford: Blackwell.
- Maillart, C. (2003). *Origine des troubles morphosyntaxiques chez les enfants dysphasiques*. Université Louvain-la-Neuve, Faculté de psychologie et des sciences de l'éducation.
- Marzo, S. (2015). Exploring the social meaning of contemporary urban vernaculars: perceptions and attitudes about Citétaal in Flanders. *International Journal of Bilingualism*, doi: 10.1177/1367006914566831
- Marzo, S., & Ceuleers, E. (2011). The use of Citétaal among adolescents in Limburg : the role space appropriation in language variation and change. *Journal of Multilingual and Multicultural Development*, 32(5), 451-464. doi: 10.1080/01434632.2011.597560

- Marzo, S., & Svendsen, B.A. (2015). A 'new' speech style is born: the omnipresence of structure and agency in the life of semiotic registers in heterogeneous urban spaces. *EuJAL*, 3(1), 1-39.
- Meyer, D., Zeilis, A., & Hornik, K. (2015). *vcd: Visualizing Categorical Data*. R package version 1.4-1. Available online: <https://cran.r-project>
- Mills, A. E. (1986). *The acquisition of gender: a study of English and German*. Berlijn: Springer.
- Muysken, P. (2001). The origin of creole languages: the perspective of second language learning. In N. Smith, & T. Veenstra (Eds.), *Creolization and contact* (pp. 157-173). Amsterdam: John Benjamins.
- Nortier, J., & Dorleijn, M. (2008). A Moroccan accent in Dutch: a sociocultural style restricted to the Moroccan community? *International Journal of Bilingualism*, 12(1/2), 125-142. doi: 10.1177/13670069080120010801
- Pauwels, J. L. (1938). *Bijdrage tot de kennis van het geslacht der substantieven in Zuid-Nederland*. Tongeren: Michiels-Broeders.
- Plevoets, K. (2008). *Tussen spreek- en standaardtaal: een corpusgebaseerd onderzoek naar de situationele, regionale en sociale verspreiding van enkele morfosyntactische verschijnselen uit het gesproken Belgisch-Nederlands*. Katholieke Universiteit Leuven, Faculteit Letteren.
- Pijpops, D., & Van de Velde, F. (To appear, 2016). Ethnolect speakers and Dutch partitive adjectival inflection: a corpus analysis. *Taal en Tongval*, 67(2), 343-371.
- Quist, P. (2008). Sociolinguistic approaches to multiethnolect: language variety and stylistic practice. *International Journal of Bilingualism*, 12(1/2), 43-61. doi: 10.1177/13670069080120010401
- R Core team. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. Available online: <http://www.r-project.org>
- Reijers, C. (2007). *The acquisition of grammatical gender in Dutch: a comparison between Dutch monolingual children and Spanish-Dutch bilingual children* (Unpublished master's thesis). Universiteit Leiden.

- Roodenburg, J., & Hulk, A. (2009). 'Gender' puzzels. *Taal & Tongval*, 61(1), 143-164. doi: 10.5117/TET2009.1.ROOD
- Ruette, T., & Van de Velde, F. (2013). Moroccorp: tien miljoen woorden uit twee Marokkaans-Nederlandse chatkanalen. *Lexikos*, 23(1), 456-475.
- Semplicini, C. (2012). Dutch double gender nouns: arbitrary or motivated agreement? *Journal of Germanic Linguistics*, 24(2), 133-186. doi: 10.1017/S1470542712000025
- Siemund, P. (2008). *Pronominal gender in English: a study of English varieties from a crosslinguistic perspective*. New York: Routledge.
- Silverstein, M. 1976. Hierarchy of features and ergativity. In R. Dixon (Eds.), *Grammatical categories in Australian languages* (pp. 112-171). Canberra: Australian Institute of Aboriginal Studies.
- Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeilis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, 9:307. doi: 10.1186/1471-2105-9-307
- Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC Bioinformatics*, 8:25. doi: 10.1186/1471-2105-8-25
- Tabouret-Keller, A. (1997). Language and identity. In F. Coulmas (Eds.), *The handbook of sociolinguistics* (pp. 315-326). Oxford: Blackwell.
- Tagliamonte, S. A., & Baayen, R. H. (2012). Models, forests and trees of York English: was/were variation as a case study for statistical practice. *Language Variation and Change*, 24(2), 135-178. doi: 10.1017/S0954394512000129
- Unsworth, S. (2006). Age and input in early child bilingualism: the acquisition of grammatical gender in Dutch. In A. Belikova, L. Mernoi, & M. Umeda (Eds.), *Galana 2: proceedings of the conference on generative approaches to language acquisition North America 2* (pp. 448-458). Somerville: MA: Cascadilla Press.

- Unsworth, S., Argyro, F., Cornips, L., Hulk, A., Sorace, A., & Tsimpli, I. (2014). The role of age of onset and input in early child bilingualism in Greek and Dutch. *Applied Psycholinguistics*, 35(4), 765-805. doi: 10.1017/S0142716412000574
- Van Berkum, J. J. A. (1996). *The psycholinguistics of grammatical gender: studies on language comprehension and production* (Unpublished doctoral dissertation). Universiteit Nijmegen, Max Planck Institute for Psycholinguistics.
- Van der Sijs, N. (2014). Systematisch onderzoek naar Nederlandse contactvariëteiten. *Taal & Tongval*, 66(2), 117-142. doi: 10.5117/TET2014.2.SIJS
- Van der Velde, M. (2003). *Déterminants et pronoms en néerlandais et en français: syntaxe et acquisition* (Unpublished master's thesis). Université Paris 8, Vincennes-Saint Denis.
- Wickham, H. (2015). *stringr: simple, consistent wrappers for common string operations*. R package version 1.0.0. Available online: <https://CRAN.R-project.org/package=stringr>
- Wickham, H. & Francois, R. (2015). *dplyr: a grammar of data manipulation*. R package 0.4.3. Available online: <https://CRAN.R-project.org/package=dplyr>



## Bijlagen

### Bijlage 1: Geattesteerde woorden per corpus

	Mannelijk	Vrouwelijk	Onzijdig
Human	bakker broer dokter vader	juffrouw moeder tante zangeres	broertje kind meisje zusje
High animate	aap beer ezel haan hond leeuw olifant tijger vis	duif geit kat kip koe muis mus slang	konijn paard schaap varken
Low animate	kakkerlak kever vlinder worm	mier mug slak spin wesp	insect
Collective nouns	raad	brandweer politie	korps leger
Bounded object/Abstract	appel auto bal fiets lepel rok sleutel stoel  vorm wind	bank broek deur kast tafel vaas vork  kleur lente ramp vondst	boek dak glas hemd huis raam  geheim probleem
Unbounded abstract, Unspecific mass	haat pijn adem ouderdom wind  chocolade koffie pasta rijst suiker	liefde vriendschap koorts  boter melk saus soep	geluk verdriet  brood goud hout water zand

Tabel 1: Geattesteerde woorden *Morocccorp*

	Mannelijk	Vrouwelijk	Onzijdig
Human	broer dokter vader	juffrouw moeder tante	kind meisje zusje
High Animate	hond leeuw olifant vis	kat kip muis	konijn
Low Animate	/	/	/
Collective nouns	/	politie	/
Bounded object/Abstract	auto bal fiets sleutel  winter	broek deur kast tafel vork  kleur	dak glas huis raam  probleem
Unbounded abstract/unspecific mass	haat pijn adem wind  chocolade koffie pasta rijst suiker	liefde vriendschap  saus soep	/  brood water zand

Tabel 2: Geattesteerde woorden *Citëtaalcorpus*



	Mannelijk	Vrouwelijk	Onzijdig
Human	bakker broer dokter vader	juffrouw moeder tante zangeres	broertje kind meisje zusje
High animate	aap beer ezel haan hond leeuw olifant tijger vis	duif geit kat kip koe muis mus slang	konijn paard schaap varken
Low animate	kever vlinder worm	mier mug spin	/
Collective nouns	raad	brandweer politie	leger
Bounded Object/Abstract	appel auto bal fiets lepel rok sleutel stoel  vorm winter	bank broek broer deur kast tafel vork  kleur lente ramp	boek dak glas hemd huis raam  geheim probleem
Unbounded abstract/unspecific mass	haat pijn adem ouderdom wind  chocolade koffie pasta rijst suiker	liefde vriendschap  boter melk saus soep	geluk verdriet  brood goud hout water zand

Tabel 3: Geattesteerde woorden *IRC* (Nederland)

	Mannelijk	Vrouwelijk	Onzijdig
Human	bakker broer dokter vader	juffrouw moeder tante zangeres	broertje kind meisje zusje
High animate	aap beer ezel haan hond leeuw olifant tijger vis	duif geit kat kip koe muis mus slang	konijn paard schaap varken
Low animate	kever vlinder worm	mier mug slak spin wesp	insect
Collective nouns	raad	brandweer politie	leger
Bounded Object/Abstract	appel auto bal fiets lepel rok sleutel stoel  vorm winter	bank broek broer deur kast tafel vork  kleur lente ramp	boek dak glas hemd huis raam  geheim probleem
Unbounded abstract/unspecific mass	haat pijn adem ouderdom wind  chocolade koffie pasta rijst suiker	liefde vriendschap koorts  boter melk saus soep	geluk verdriet  brood goud hout water zand

Tabel 4: Geattesteerde woorden IRC (Vlaanderen)

## Bijlage 2: Reguliere expressies

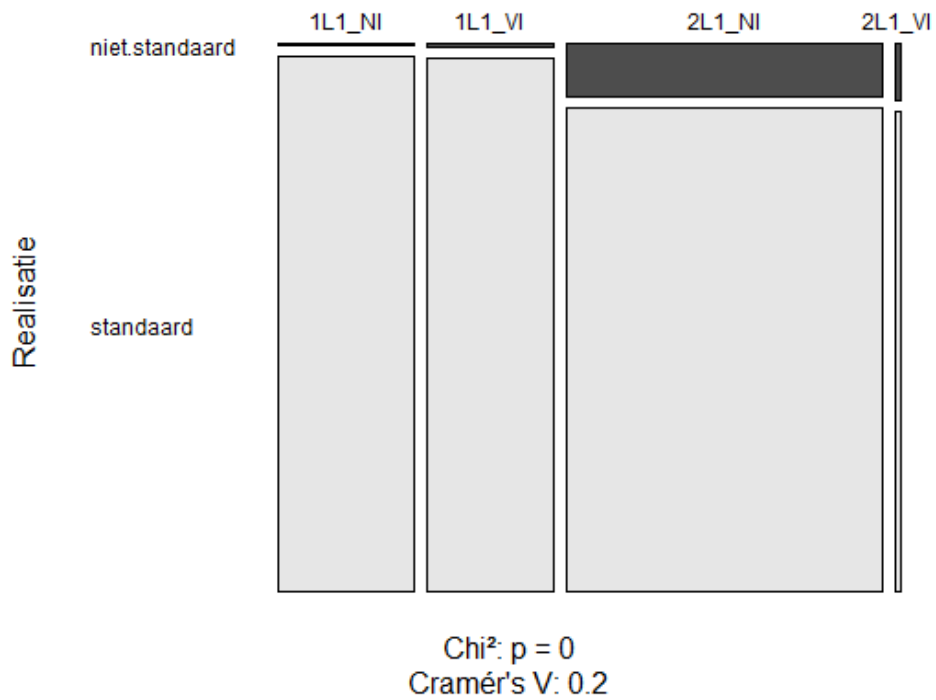
**AAP:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\baap\b  
**ADEM:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\badem\b  
**APPEL:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bapp?el\b  
**AUTO:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bauto\b  
**BAKKER:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bbakk?er\b  
**BAL:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bbal\b  
**BANK:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bbank\b  
**BEER:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bbeer\b  
**BOEK:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bboek\b  
**BOTER:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bboter\b  
**BRANDWEER:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bbrandweer\b  
**BROEK:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bbroek\b  
**BROER:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bbroer\b  
**BROERTJE:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bbroertje\b  
**BROOD:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bbrood\b  
**CHOCOLADE:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bchocolade\b  
**DAK:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bdak\b  
**DEUR:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bdeur\b  
**DOKTER:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bdo(c|k)ter\b  
**DUIF:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bduif\b  
**EZEL:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bezel\b  
**FIETS:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bfiets\b  
**GEHEIM:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bgeheim\b  
**GEIT:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bgeit\b  
**GELUK:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bgeluk\b  
**GLAS:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bglas\b  
**GOUD:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bgoud\b  
**HAAT:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bhaat\b  
**HEMD:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bhem(d|t)\b  
**HOND:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bhon(d|t)\b  
**HOUT:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bhout\b  
**HUIS:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bhuis\b  
**INSECT:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\binsect\b  
**JUFFROUW:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bjuff?rouw\b  
**KADAVER:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bkadaver\b  
**KAKKERLAK:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bkakkerlak\b  
**KAST:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bkast\b  
**KAT:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bkat\b  
**KEVER:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bkever\b  
**KIND:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bkin(d|t)\b  
**KIP:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bkip\b  
**KLEUR:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bkleur\b  
**KOE:** (?i)(t|\b(d(i?e|at)|h?e?t))(\w+){0,3}+\bkoel\b

**KOFFIE:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\bkoffie\b  
**KONIJN:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\bkonijn\b  
**KOORTS:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\bkoorts\b  
**LEEUEW:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\bleeuw\b  
**LEGER:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\bleger\b  
**LENTE:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\blente\b  
**LEPEL:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\blepel\b  
**LIEFDE:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\bliefde\b  
**LUIPAARD:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\bluipaar(d|t)\b  
**MEISJE:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\bmeisje\b  
**MELK:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\bmelk\b  
**MIER:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\bmier\b  
**MOEDER:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,2} +\bmoeder\b  
**MUG:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\bmug\b  
**MUIS:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\bmuis\b  
**MUS:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\bmus\b  
**OLIFANT:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\bolifant\b  
**OUDERDOM:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\bouderdom\b  
**PAARD:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\bpaar(d|t)\b  
**PASTA:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\bpasta\b  
**PIJN:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\bpijn\b  
**POLITIE:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\bpolitie\b  
**PROBLEEM:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\bprobleem\b  
**RAAD:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\braad\b  
**RAAM:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\braam\b  
**RAMP:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\bramp\b  
**RIJST:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\brijst\b  
**ROK:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\brok\b  
**SAUS:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\bsaus\b  
**SCHAAP:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\bschaap\b  
**SLAK:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\bslak\b  
**SLANG:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\bslang\b  
**SLEUTEL:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\bsleutel\b  
**SOEP:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\bsoep\b  
**SPIN:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\bspin\b  
**STOEL:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\bstoel\b  
**SUIKER:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\bsuiker\b  
**TAFEL:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\btafel\b  
**TANTE:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\btante\b  
**TIJGER:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\btijger\b  
**VAAS:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\bvaas\b  
**VADER:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,2} +\bvader\b  
**VARKEN:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\bvarken\b  
**VERDRIET:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\bverdriet\b  
**VIS:** (?i)(t|b(d(i?e|at)|h?e?t))( +\w+){0,3} +\bvis\b

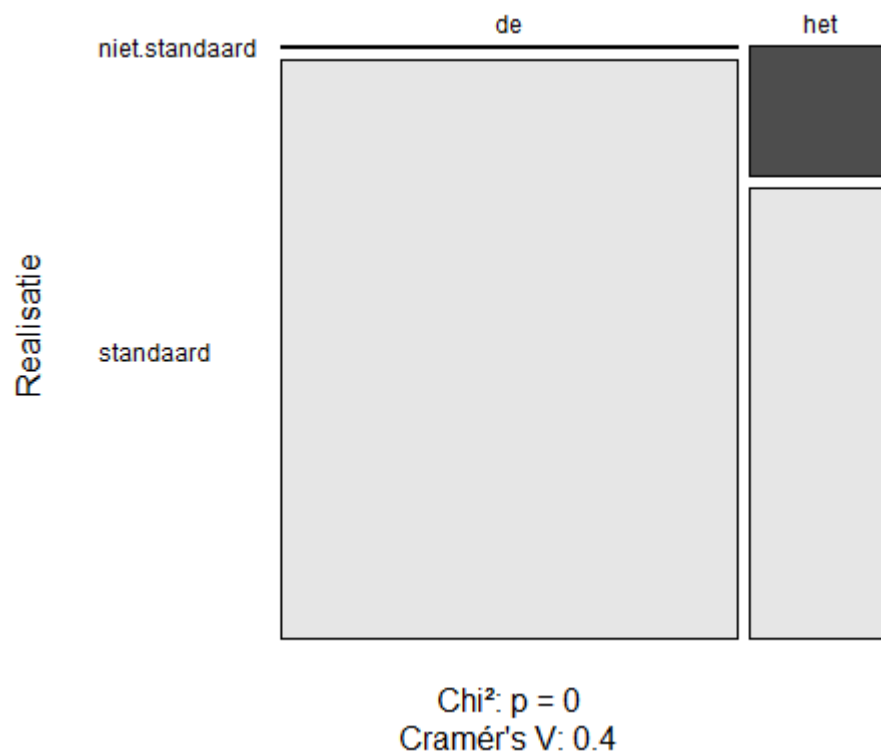
**VLINDER:** (?i)( 't | \b(d(i?e|at)|h?e?t) ) ( +\w+ ){0,3} +\bv linder\b  
**VONDST:** (?i)( 't | \b(d(i?e|at)|h?e?t) ) ( +\w+ ){0,3} +\bvondst\b  
**VORK:** (?i)( 't | \b(d(i?e|at)|h?e?t) ) ( +\w+ ){0,3} +\bvork\b  
**VORM:** (?i)( 't | \b(d(i?e|at)|h?e?t) ) ( +\w+ ){0,3} +\bvorm\b  
**VRIENDSCHAP:** (?i)( 't | \b(d(i?e|at)|h?e?t) ) ( +\w+ ){0,3} +\bvriendschap\b  
**WATER:** (?i)( 't | \b(d(i?e|at)|h?e?t) ) ( +\w+ ){0,3} +\bwater\b  
**WESP:** (?i)( 't | \b(d(i?e|at)|h?e?t) ) ( +\w+ ){0,3} +\bwesp\b  
**WIND:** (?i)( 't | \b(d(i?e|at)|h?e?t) ) ( +\w+ ){0,3} +\bwind\b  
**WINTER:** (?i)( 't | \b(d(i?e|at)|h?e?t) ) ( +\w+ ){0,3} +\bwinter\b  
**WORM:** (?i)( 't | \b(d(i?e|at)|h?e?t) ) ( +\w+ ){0,3} +\bworm\b  
**ZAND:** (?i)( 't | \b(d(i?e|at)|h?e?t) ) ( +\w+ ){0,3} +\bzand\b  
**ZANGERES:** (?i)( 't | \b(d(i?e|at)|h?e?t) ) ( +\w+ ){0,3} +\bzangeres\b  
**ZUSJE:** (?i)( 't | \b(d(i?e|at)|h?e?t) ) ( +\w+ ){0,3} +\bzusje\b



### Bijlage 3: Verkennende plots corpusonderzoek



Figuur 1: Standaard en niet-standaardrealisatie in de verschillende corpora



Figuur 2: Standaard en niet-standaardrealisatie van grammaticale *de-woorden* en *het-woorden* in de vier corpora





## Bijlage 4: Draaitabellen corpusonderzoek

### Bijlage 4A: Moroccorp

Woord	Som van Observatie?	Som van niet-standaardvorm	Som van standaardvorm	Percentage niet-standaardvorm
aap	89	3	86	3,37%
adem	4	0	4	0%
appel	18	0	18	0%
auto	415	2	413	0,48%
bakker	104	0	104	0%
bal	69	0	69	0%
bank	156	0	156	0%
beer	12	0	12	0%
boek	203	47	156	23,15%
boter	3	0	3	0%
brandweer	19	0	19	0%
broek	53	0	53	0%
broer	47	0	47	0%
broertje	21	8	13	38%
brood	39	22	17	56%
chocolade	11	0	11	0%
dak	40	10	30	0,25
deur	448	1	447	0%
dokter	127	1	126	0,79%
duif	6	0	6	0%
ezel	135	2	133	1,48%
fiets	175	1	174	0,57%
geheim	43	0	43	0%
geit	20	0	20	0%
geluk	44	4	40	9,09%
glas	22	10	12	45,45%
goud	33	5	28	15,15%
haan	11	0	11	0%
haat	117	5	112	4,27%
hemd	3	0	3	0%
hond	276	0	276	0%
hout	3	1	2	33,33%
huis	255	55	200	21,57%
insect	1	0	1	0%
juffrouw	6	0	6	0%
kadaver	0	0	0	0%
kakkerlak	9	0	9	0%
kast	184	1	183	0,54%

kat	78	1	77	1,28%
kever	1	0	1	0%
kind	73	29	44	39,73%
kip	93	2	91	2,15%
kleur	171	2	169	1,17%
koe	86	0	86	0%
koffie	162	0	162	0%
konijn	29	28	1	96,55%
koorts	3	0	3	0%
korps	6	2	4	33,33%
leeuw	40	0	40	0%
leger	80	12	68	15,00%
lente	11	0	11	0%
lepel	6	0	6	0%
liefde	262	3	259	1,15%
meisje	327	208	119	63,61%
melk	31	0	31	0%
mier	4	0	4	0%
moeder	251	0	251	0%
mug	8	0	8	0%
muis	34	0	34	0%
mus	1	0	1	0%
olifant	8	0	8	0%
ouderdom	3	0	3	0%
paard	27	13	14	48,15%
pasta	5	0	5	0%
pijn	101	0	101	0%
politie	204	0	204	0%
probleem	298	42	256	14,09%
raad	5	0	5	0%
raam	265	85	180	32,08%
ramp	10	0	10	0%
rijst	6	0	6	0%
rok	1	0	1	0%
saus	32	1	31	3,13%
schaap	65	58	7	89,23%
slak	2	0	2	0%
slang	13	0	13	0%
sleutel	39	2	37	5,13%
soep	63	0	63	0%
spin	10	0	10	0%
stoel	45	0	45	0%
suiker	19	4	15	21,05%

tafel	92	0	92	0%
tante	10	0	10	0%
tijger	5	0	5	0%
vaas	4	1	3	25,00%
vader	153	0	153	0%
varken	48	27	21	56,25%
verdriet	5	1	4	20,00%
vis	77	0	77	0%
vlinder	1	0	1	0%
vondst	13	0	13	0%
vork	7	0	7	0%
vorm	71	4	67	5,63%
vriendschap	2	0	2	0%
water	216	24	192	11,1%
wesp	3	0	3	0%
wind	69	1	68	1,45%
winter	112	2	110	1,79%
worm	1	0	1	0%
zand	40	11	29	27,50%
zangeres	10	0	10	0%
zusje	41	25	16	60,98%
<b>Eindtotaal</b>	<b>7179</b>	<b>766</b>	<b>6413</b>	<b>10,67%</b>

*Bijlage 4B: Citétaalcorpus*

Rijlabels	Som van Observatie?	Som van niet-standaardvorm	Som van standaardvorm	Percentage niet-standaardvorm
adem	0	0	0	0,00%
auto	11	0	11	0,00%
bal	6	0	6	0,00%
broek	0	0	0	0,00%
broer	0	0	0	0,00%
brood	0	0	0	0,00%
chocolade	1	0	1	0,00%
dak	1	0	1	0,00%
deur	8	0	8	0,00%
dokter	2	0	2	0,00%
fiets	3	0	3	0,00%
glas	0	0	0	0,00%
haat	1	0	1	0,00%
hond	10	0	10	0,00%
huis	2	0	2	0,00%
juffrouw	1	0	1	0,00%
kast	3	0	3	0,00%

kat	1	0	1	0,00%
kind	1	0	1	0,00%
kip	2	0	2	0,00%
kleur	1	0	1	0,00%
koffie	0	0	0	0,00%
konijn	1	1	0	100,00%
leeuw	0	0	0	0,00%
liefde	1	0	1	0,00%
meisje	11	9	2	81,82%
moeder	2	0	2	0,00%
muis	2	0	2	0,00%
olifant	1	0	1	0,00%
pasta	2	0	2	0,00%
pijn	0	0	0	0,00%
politie	3	0	3	0,00%
probleem	3	0	3	0,00%
raam	2	1	1	50,00%
rijst	1	0	1	0,00%
saus	1	0	1	0,00%
sleutel	2	0	2	0,00%
soep	2	0	2	0,00%
suiker	1	0	1	0,00%
tafel	4	0	4	0,00%
tante	0	0	0	0,00%
vader	0	0	0	0,00%
vis	2	0	2	0,00%
vork	1	0	1	0,00%
vriendschap	0	0	0	0,00%
water	3	1	2	33,33%
wind	2	0	2	0,00%
winter	1	0	1	0,00%
zand	1	0	1	0,00%
zusje	0	0	0	0,00%
<b>Eindtotaal</b>	<b>103</b>	<b>12</b>	<b>91</b>	<b>11,65%</b>

*Bijlage 4C: Condiv*

Rijlabels	Som van Observatie?	Som van niet-standaardvorm	Som van standaardvorm	Percentage niet-standaardvorm
aap	15	0	15	0,00%
adem	4	0	4	0,00%
appel	15	0	15	0,00%
auto	354	1	353	0,28%
bakker	48	0	48	0,00%

bal	53	0	53	0,00%
bank	61	0	61	0,00%
beer	17	0	17	0,00%
boek	80	3	77	3,75%
boter	10	1	9	10,00%
brandweer	18	0	18	0,00%
broek	59	1	58	1,69%
broer	52	2	50	3,85%
broertje	16	0	16	0,00%
brood	9	0	9	0,00%
chocolade	3	0	3	0,00%
dak	85	1	84	1,18%
deur	514	0	514	0,00%
dokter	112	0	112	0,00%
duif	3	0	3	0,00%
ezel	7	0	7	0,00%
fiets	174	0	174	0,00%
geheim	14	1	13	7,14%
geit	16	0	16	0,00%
geluk	29	0	29	0,00%
glas	35	0	35	0,00%
goud	1	0	1	0,00%
haan	39	0	39	0,00%
haat	7	0	7	0,00%
hemd	3	0	3	0,00%
hond	293	0	293	0,00%
hout	5	0	5	0,00%
huis	143	2	141	1,40%
insect	0	0	0	0,00%
juffrouw	18	0	18	0,00%
kast	219	0	219	0,00%
kat	170	0	170	0,00%
kever	0	0	0	0,00%
kind	85	0	85	0,00%
kip	45	1	44	2,22%
kleur	66	10	56	15,15%
koe	101	0	101	0,00%
koffie	144	0	144	0,00%
konijn	32	3	29	9,38%
koorts	4	0	4	0,00%
korps	1	0	1	0,00%
leeuw	29	0	29	0,00%
leger	85	0	85	0,00%

lente	42	0	42	0,00%
lepel	5	0	5	0,00%
liefde	242	1	241	0,41%
meisje	139	3	136	2,16%
melk	17	0	17	0,00%
mier	4	0	4	0,00%
moeder	77	0	77	0,00%
mug	7	0	7	0,00%
muis	49	0	49	0,00%
mus	2	0	2	0,00%
olifant	8	0	8	0,00%
ouderdom	12	0	12	0,00%
paard	31	0	31	0,00%
pasta	9	0	9	0,00%
pijn	57	0	57	0,00%
politie	88	0	88	0,00%
probleem	451	2	449	0,44%
raad	8	0	8	0,00%
raam	174	9	165	5,17%
ramp	3	0	3	0,00%
rijst	2	0	2	0,00%
rok	10	0	10	0,00%
saus	23	0	23	0,00%
schaap	21	1	20	4,76%
slak	2	0	2	0,00%
slang	3	0	3	0,00%
sleutel	34	0	34	0,00%
soep	58	0	58	0,00%
spin	8	0	8	0,00%
stoel	56	0	56	0,00%
suiker	18	0	18	0,00%
tafel	95	1	94	1,05%
tante	2	0	2	0,00%
tijger	6	0	6	0,00%
vaas	1	0	1	0,00%
vader	36	0	36	0,00%
varken	21	0	21	0,00%
verdriet	4	0	4	0,00%
vis	40	0	40	0,00%
vlinder	11	0	11	0,00%
vork	6	2	4	33,33%
vorm	50	1	49	2,00%
vriendschap	3	0	3	0,00%

water	230	0	230	0,00%
wesp	4	1	3	25,00%
wind	96	0	96	0,00%
winter	56	1	55	1,79%
worm	0	0	0	0,00%
zand	23	2	21	8,70%
zangeres	28	0	28	0,00%
zusje	18	2	16	11,11%
<b>Eindtotaal</b>	<b>5683</b>	<b>52</b>	<b>5631</b>	<b>0,92%</b>





## Bijlage 5: Leesteksten veldonderzoek

### Bijlage 5A: Leestekst kinderen (7-8 jaar)

**KU LEUVEN**

#### Algemene gegevens

Ik heet .....

Ik ben ..... jaar

Ik ben een JONGEN / MEISJE (omcirkel)

Ik zit in het 2<sup>de</sup> / 3<sup>de</sup> leerjaar (omcirkel)

Welke taal/talen spreek je tegen je mama? .....

.....

Welke taal/talen spreek je tegen je papa? .....

.....

Welke taal/talen spreek je tegen je broer/zus?.....

.....

Welke taal/talen spreek je tegen de rest van je familie? .....

.....

Welke taal/talen spreek je tegen je vrienden?.....

.....

Welke taal spreek je het best?.....

Wat is het beroep van je mama? .....

Wat is het beroep van je papa? .....

*Opgave: In deze tekst zijn een aantal woorden weg. Kies één van de twee woorden en schrijf het juiste woord op de lijn. De juf of meester leest de tekst voor.*

## **Leestekst: De Bremer stadsmuzikanten**

Er was eens heel lang geleden een oud paard. \_\_\_\_\_ **(die/dat)** paard had heel zijn leven voor zijn baas gewerkt. Op een dag zei de baas: "Paard, het wordt tijd dat ik \_\_\_\_\_ **(jouw/jou)** vervang. In plaats daarvan koop ik een jong paard." Toen dacht Paard: "Weet je wat, ik ga naar Bremen om er stadsmuzikant te worden." Onderweg zag Paard een dier dat zat te \_\_\_\_\_ **(huillen/ huilen)** langs de kant van de weg. Het was een konijn. "Wat is er met jou aan de \_\_\_\_\_ **(hant/hand)?",** vroeg Paard. "Ik word veel te oud en mijn baas wil me opeten", snikte \_\_\_\_\_ **(de/het)** konijn. "Ga dan samen met mij naar Bremen om stadsmuzikant te worden!", zei Paard. Samen vertrokken ze. Even later kwamen ze een haan en een kat tegen. "Ga mee met ons naar Bremen!", zei Paard. Dat vond Haan, die heel mooi kon zingen, een goed idee. Onderweg zongen de vier dieren samen een \_\_\_\_\_ **(lied/liet).**

Toen het nacht werd, besloten de dieren om een slaappleats te zoeken in het bos. Plots zagen ze \_\_\_\_\_ **(ligt/licht)** branden in een huisje. Toen \_\_\_\_\_ **(de/het)** kat naar binnen had gekeken, wisten de dieren dat er een gevaarlijke roversfamilie woonde. "Ik zag een moeder, een vader, een broertje en een zusje", zei Kat. \_\_\_\_\_ **(die/dat)** vader zag er heel eng uit. Hij droeg een lapje voor zijn ogen en \_\_\_\_\_ **(de/het)** meisje was een stuk taart aan het eten!" Kat had gelijk. Heel de tafel stond vol lekkere dingen. Al \_\_\_\_\_ **(de/het)** brood was vers gebakken en rook heerlijk. Toen Paard al \_\_\_\_\_ **(die/dat)** suiker zag, liep het water hem in de \_\_\_\_\_ **(mond/mont).** Ook Konijn keek naar een glas dat vol lekkere limonade zat.

"Wij lusten ook wel iets!" zei Haan. "Ik denk dat ik een oplossing weet voor \_\_\_\_\_ **(het/de)** probleem", zei Konijn. Hij legde zijn plannetje uit: "Als we een \_\_\_\_\_ **(hooge/hoge)**

toren bouwen en heel hard zingen, dan worden de rovers bang en lopen ze weg". Zo \_\_\_\_\_ (**gezegd/gezegt**), zo gedaan. Kat en Konijn klommen op de rug van Paard en Haan ging helemaal vanboven zitten. Dat zag er eng uit in het donker! Eerst stootte Paard \_\_\_\_\_ (**de/het**) vaas om op de vensterbank. Toen begonnen de dieren heel \_\_\_\_\_ (**luid/luit**) te zingen. Door al het lawaai schrokken de rovers heel erg! Toen ze door het raam keken, zagen ze een grote gedaante die wel op een monster leek! Ook de insecten uit het bos kwamen een handje helpen. \_\_\_\_\_ (**het/de**) mier kroop in de trui van de rover en bezorgde hem ontzettend veel jeuk. \_\_\_\_\_ (**enkele/enkelen**) van de rovers deden bijna in hun broek van de angst! Toen ze naar buiten liepen, beet Paard in de billen van \_\_\_\_\_ (**de/het**) zusje. Omdat \_\_\_\_\_ (**die/dat**) pijn zo erg was, gilte ze zo hard dat \_\_\_\_\_ (**het/de**) politie een kijkje kwam nemen. Eindelijk hadden ze de gevaarlijke rovers gevonden! Ze moesten meteen naar de gevangenis. Paard, Kat, Konijn en Haan gingen naar binnen en smulden de \_\_\_\_\_ (**heele/hele**) avond van al het lekkers. Ook \_\_\_\_\_ (**de/het**) leger van insecten mocht mee-eten. Alle dieren gaven een groot applaus, want ze vonden dat Mier \_\_\_\_\_ (**de/het**) dapperste insect was. Iedereen at zijn buikje \_\_\_\_\_ (**ront/rond**). Na het eten besloten de dieren om voor altijd in \_\_\_\_\_ (**die/dat**) huis te blijven wonen. Al \_\_\_\_\_ (**gauw/gouw**) wilden ze niet meer naar Bremen. \_\_\_\_\_ (**de/het**) geluk kon niet meer op!

En de rovers? Die werden voor altijd opgesloten. Ze waren heel verdrietig. Het was \_\_\_\_\_ (**de/het**) grootste ramp uit hun leven. Maar het was ook hun verdiende loon!

Bron: <http://www.efteling.com/NL/Over-de-Efteling/Sprookjes-bestaan/Eftelingen-Sprookjes/Lees-hier-alle-Sprookjes/De-Bremer-stadsmuzikanten.html>

**Algemene gegevens**

**KU LEUVEN**

Voornaam:.....

Naam:.....

Geslacht: M / V

Geboortedatum:...../...../.....

2<sup>DE</sup> / 3<sup>DE</sup> middelbaar

Studierichting:.....

ASO / TSO / BSO / KSO

Welke taal/talen spreek je tegen je moeder?.....

Welke taal/talen spreek je tegen je vader?.....

Welke taal/talen spreek je tegen je broer/zus?.....

.....

Welke taal/talen spreek je tegen je vrienden?.....

.....

Welke taal/talen spreek je tegen andere familieleden?.....

.....

Welke van die talen spreek je het best?.....

Wat is het beroep van je moeder?.....

Wat is het beroep van je vader?.....

**Algemene gegevens**

Voornaam:.....

Naam:.....

Geslacht: M / V (*omcirkel*)

Geboortedatum:...../...../.....

Hogeschool / universiteit (*omcirkel*)

Studierichting:.....

Studierichting middelbaar: ASO / TSO / BSO / KSO (*omcirkel*)

Welke taal/talen spreek je thuis?.....

Welke taal/talen spreek je tegen je moeder?.....

Welke taal/talen spreek je tegen je vader?.....

Welke taal/talen spreek je tegen je broer/zus?.....

.....

Welke taal/talen spreek je tegen je vrienden?.....

.....

Welke taal/talen spreek je tegen andere familieleden?.....

.....

Welke van die talen spreek je het best?.....

Welke van die talen spreek je het liefst?.....

Wat is het beroep van je moeder?.....

Wat is het beroep van je vader?.....

Opgave: In onderstaande tekst zijn een aantal woorden weggelaten. Vul ze aan door telkens een keuze te maken uit één van de twee vetgedrukte woorden tussen haakjes. De tekst wordt voorgelezen.

## Leestekst:

### Wie is er bang van de Grote Boze Wolf? – Het voorlezen van sprookjes

Er was eens heel lang geleden in een land hier ver vandaan... Die zin kennen we allemaal. Al generaties lang worden sprookjes voorgelezen en dromen kinderen weg bij het horen van verhalen over prinses, prinsessen en kastelen. Uit een onderzoek van *Libelle* in 2012 bleek echter dat steeds meer ouders \_\_\_\_\_ (**weigeren/wijgeren**) om sprookjes voor te lezen aan hun kinderen. Ongeveer de helft van de ouders vindt *Rapunzel* en *Repelsteeltje* te beangstigend, omdat het hoofdpersonage daarin ontvoerd \_\_\_\_\_ (**word/wordt**). Ook *Roodkapje* is geen geliefd sprookje. De scène waarin de Grote Boze Wolf \_\_\_\_\_ (**de/het**) meisje opeet, zou kinderen nachtmerries bezorgen. Eén op de drie ouders beweert zelfs dat *Roodkapje* hun kinderen al deed huilen. \_\_\_\_\_ (**die/dat**) pijn willen ouders hun kinderen besparen.

Maar het zijn niet altijd de horrorscenario's die ouders tegenhouden om sprookjes voor te lezen. \_\_\_\_\_ (**sommige/sommigen**) vinden sprookjes niet meer van deze tijd. Zo zou *Sneeuwitje* een te negatief beeld vormen van de stiefmoeder en zou \_\_\_\_\_ (**de/het**) *Gelaarsde kat* de kinderen aanzetten tot leugentjes om bestwil. Bovendien zouden sprookjes een te idealistisch beeld vormen van de werkelijkheid en jonge meisjes hopeloos doen zoeken naar hun prins op \_\_\_\_\_ (**de/het**) witte paard.

Toch \_\_\_\_\_ (**biedt/bied**) het voorlezen van sprookjes vele voordelen. In elk sprookje zit eerst en vooral een moraal verborgen, die ons leert wat wel of niet kan. Neem nu bijvoorbeeld *Hans en Grietje*. Wanneer de kinderen het peperkoekenhuisje ontdekken, weten ze niet dat achter al \_\_\_\_\_ (**die/dat**) suiker de woning van een heks schuil gaat. Gulzigheid drijft hen vervolgens in haar \_\_\_\_\_ (**klouwen/klauwen**).

Ten tweede sporen sprookjes kinderen aan om creatief naar oplossingen te zoeken. Nadat \_\_\_\_\_ **(de/het)** vader van Hans en Grietje zijn kinderen heeft achtergelaten in het bos, besluit \_\_\_\_\_ **(het/de)** zusje van Hans om broodkruimels te strooien, zodat ze \_\_\_\_\_ **(dank zij/dankzij)** dat spoor de weg naar huis kunnen vinden. Dat \_\_\_\_\_ **(de/het)** brood uiteindelijk wordt opgegeten door de vogels, vertellen we er even niet bij.

Ten derde helpen sprookjesfiguren als heksen en feeën kinderen bij de ontwikkeling van hun fantasie. Een koning die met de hulp van \_\_\_\_\_ **(de/het)** leger een kudde vuurspuwende draken gaat verslaan, \_\_\_\_\_ **(paddestoelen/paddenstoelen)** die plots kunnen zingen, een tovenaer die \_\_\_\_\_ **(de/het)** huis van de keizer onzichtbaar maakt, een heks die \_\_\_\_\_ **(de/het)** vaas van haar vijand vult met dodelijk gif... In Sprookjesland kan het allemaal. Als je er \_\_\_\_\_ **(ten minste/tenminste)** maar zelf in gelooft. Bovendien spelen ook sprekende dieren in sprookjes als *De wolf en de zeven geitjes*, de *Bremer stadsmuzikanten* en de *Kikkerkoning* een grote rol. Daarmee vormen ze de basis van een traditie die \_\_\_\_\_ **(al lang/allang)** aan de gang is en die dieren met duidelijke karaktertrekken prominent op de voorgrond plaatst. Denk maar aan andere verhalen zoals *Alice in Wonderland* waarin \_\_\_\_\_ **(de/het)** witte konijn met zijn zakhorloge haastig heen en weer loopt. Of aan de fabels van Jean de la Fontaine, waarin \_\_\_\_\_ **(de/het)** mier de harde werker blijkt te zijn. Opmerkelijk genoeg komen \_\_\_\_\_ **(sommigen/sommige)** dieren, die nochtans duidelijk tot de fantasiewereld van kinderen behoren, nauwelijks in sprookjes voor. Zo \_\_\_\_\_ **(gebeurd/gebeurt)** het zelden dat de vlinder of het lieveheersbeestje een hoofdrol krijgt, hoewel \_\_\_\_\_ **(die/dat)** insect vaak wel tot de verbeelding spreekt.

Ten vierde leren sprookjes ons om het op te nemen tegen het kwade. In Sprookjesland, waar de personages heel zwart-wit worden voorgesteld, is het het ultieme doel van de held om de gemene

koningin of de Grote Boze Wolf te verslaan. \_\_\_\_\_ **(de/het)** politie, die in de reële wereld de dieven de cel in zou gooien, blijkt in Sprookjesland niet te bestaan en is door de superkrachten of slimme plannetjes van het hoofdpersonage ook volstrekt overbodig.

Ten slotte eindigen sprookjes altijd met een happy end, waardoor ze kinderen een gelukkig gevoel bezorgen. Ook al is het doorheen het verhaal niet altijd \_\_\_\_\_ **(rozegeur/rozengeur)** en \_\_\_\_\_ **(manenschijn/maneschijn)**, toch wordt er voor \_\_\_\_\_ **(de/het)** probleem een oplossing gevonden. Zelfs \_\_\_\_\_ **(de/het)** allergrootste ramp kan worden overwonnen. Steeds staat \_\_\_\_\_ **(de/het)** geluk aan de kant van de hoofdpersonages en daardoor ook aan de kant van de lezer. En zo leefde iedereen nog lang en gelukkig. Misschien moeten we toch meer sprookjes voorlezen?

Bron: <https://www.libelle.be/vrije-tijd/wie-is-er-bang-van-de-grote-boze-wolf/>  
[http://www.rekels.nl/leesvoer/er-was-eens\\_184.html](http://www.rekels.nl/leesvoer/er-was-eens_184.html)