

Leermechanismen in tweededialectverwerving

Kathy Rys, Walter Daelemans, Guy De Pauw, Steven Gillis

1. Inleiding

Binnen de dialectologie is er de afgelopen jaren een toenemende interesse waar te nemen voor onderzoek naar de verwerving van dialecten (cf. Berthele, 2002; Chambers, 1992; De Vogelaer, 2010; De Vogelaer & Rooze-Stouthamer, 2006; Kerswill, 1994; Payne, 1976, 1980; Roberts, 1994, 1997a, 1997b; Roberts & Labov, 1995; Rys, 2003, 2007; Rys & Bonte, 2006; Rys & De Valck, 2010; Siegel, 2010; Vousten & Bongaerts, 1990; Vousten, 1995). Dialect kan verworven worden als moedertaal (i.e. T1), maar gezien het feit dat dialect steeds minder als thuistaal gesproken wordt (cf. Hoppenbrouwers, 1990; Taeldeman, 1989, 1991), zijn kinderen voor de verwerving van het lokale dialect steeds vaker aangewezen op situaties buiten het gezin waarin nog dialect gesproken wordt, zoals het contact met grootouders of andere dialectsprekende familieleden en het contact met leeftijdgenoten die thuis wel nog dialect spreken. Onder bepaalde omstandigheden, waaronder een (vrij) positieve attitude tegenover het dialect en voldoende contact met het dialect, verwerven kinderen dan het lokale dialect als tweede taal. Het contact met leeftijdgenoten op school of in jeugd- en sportverenigingen heeft met name een grote invloed op het taalgebruik van kinderen en jongeren. Aangezien het dialect van kinderen die thuis wel nog het dialect hebben geleerd vaak vrij sterk genivelleerd is, is tweededialectverwerving vaak imperfecte verwerving (zg. ‘imperfect learning’, cf. De Vink, 2004).

In deze bijdrage bestuderen wij de verwerving van fonologische kenmerken van het dialect van Maldegem (Oost-Vlaanderen) door negen-, twaalf- en vijftienjarigen die in het Standaardnederlands zijn opgevoed. Meer bepaald focussen we op de vraag welke leermechanismen er ten grondslag liggen aan tweededialectverwerving: zijn er indicaties dat kinderen regels vormen bij het verwerven van een dialect als tweede taal, of leren zij de fonologische kenmerken van het dialect eerder op een woord-voor-woord basis? Onderzoek naar de leermechanismen van tweededialectverwerving is interessant omdat we te maken hebben met twee taalsystemen (i.e. T1 en T2) die typologisch nauw verwant zijn. In het geval van onze studie hebben we te maken met een standaardtaal als T1 en een dialect dat als het ware door die standaardtaal ‘overkoepeld’ wordt als T2. De nauwe verwantschap tussen de twee taalsystemen impliceert onder andere dat het lexicon grotendeels overlapt. De T1 en de T2 hebben dus grotendeels dezelfde woorden, maar een andere uitspraak. Bevordert deze situatie regelvorming of leidt het er net toe dat kinderen eerder woord voor woord gaan leren? Niettegenstaande dit een interessante onderzoeksvraag is, is er in de literatuur over tweededialectverwerving nauwelijks aandacht besteed aan leermechanismen. Voor zover er iets over de onderliggende leermechanismen van tweededialectverwerving gezegd wordt, is de heersende opvatting dat de verwerving van een dialect vooral regelgebaseerd is (cf. Chambers, 1992; Payne, 1980). In dit artikel onderzoeken wij of deze opvatting gehandhaafd kan blijven. Op die manier trachten wij in deze studie een bijdrage te leveren aan de actuele discussie tussen enerzijds de aanhangers van regelgebaseerde modellen van taalverwerving en anderzijds de aanhangers van woordgebaseerde modellen. Regelgebaseerde modellen (o.a. generatieve fonologie (Chomsky & Halle, 1968), de zogenaamde ‘dual-mechanism models’ (Pinker & Prince, 1988), en, toegepast op de dialectologie, bidialectale fonologie (Dressler & Wodak, 1982; Moosmüller, 1988; Auer, 1990, 1993; Taeldeman, 1993) nemen aan dat taalverwerfers generalisaties maken in de vorm van mentale regels. Volgens de aannames van de bidialectale fonologie, bijvoorbeeld, zal een verwerfer van het Maldegemse dialect

uiteindelijk een mentale regel abstraheren die bepaalt dat Standaardnederlands /ɛi/ correspondeert met dialect /e/ vóór een velaire of laryngale consonant (bijv. in *rijk*, *zwijgen*, etc.). Volgens woordgebaseerde modellen (o.a. usage-based learning (Bybee, 1995, 2001), memory-based language processing (Daelemans & Van den Bosch, 2005), analogical modeling (Skousen, Lonsdale & Parkinson, 2002)) worden woordvormen in het mentale lexicon opgeslagen zonder enige vorm van abstractie en het leren van nieuwe vormen is gebaseerd op analogie met opgeslagen woordvormen. Bijvoorbeeld, woorden met dialect /e/ vóór een velaire of laryngale consonant worden verworven op basis van hun overeenkomst met woorden die reeds in het mentale lexicon zijn opgeslagen.

In dit artikel wordt nagegaan of er aanwijzingen zijn voor het leren van de fonologische kenmerken van het Maldegemse dialect op een regelgebaseerde manier. Daarvoor wordt onderzocht of de scoregrafieken van de verschillende onderzochte dialectkenmerken een S-vormig patroon laten zien. Een dergelijk patroon wordt immers door Chambers (1992) beschouwd als een aanwijzing voor regelgebaseerd leren. Verder gaan we na of er aanwijzingen zijn voor lexicaal leren. Aan de hand van een geheugengebaseerde ('memory-based') classificatietask brengen we de meest gelijkende lexicale burens (i.e. de zogenaamde 'nearest neighbors') van woorden in kaart. Dit maakt het mogelijk de effecten van lexicale burens op de mate van verwerving van dialectkenmerken te onderzoeken.

2. Regelgebaseerd versus woordgebaseerd leren

2.1. S-curve als indicatie voor regelgebaseerd leren

Een vroege en toonaangevende studie binnen het domein van de tweededialectverwerving is Chambers (1992). Chambers onderzocht de verwerving van een aantal lexicale en fonologische kenmerken van het Brits-Engels door zes Canadese kinderen die naar Zuid-Engeland waren verhuisd. Chambers interviewde elk van zijn proefpersonen twee keer, met een tussenpoos van twee jaar. Op basis van zijn observaties stelt hij acht generalisaties over dialectverwerving voor. Een van die generalisaties is dat "phonological innovations are actuated as pronunciation variants" (Chambers, 1992: 164). Hiermee bedoelt hij dat de variabiliteit die geobserveerd kan worden in de verwerving van fonologische kenmerken het patroon van lexicale diffusie volgt. Het typerende patroon van lexicale diffusie is de S-curve (Wang, 1969): de theorie is dat een taalverandering zich langzaam en van morfeem tot morfeem doorzet in het lexicon (zie ook Wang & Cheng 1970; Chen & Wang, 1975; Chambers & Trudgill, 1980). Fonologische veranderingen worden op gang gebracht door het verwerven van specifieke gevallen van een nieuwe regel of een nieuw foneem, en ze worden pas regelmatig of systematisch nadat er een kritieke massa van woorden is verworven, wat tot uiting komt in een snelle verspreiding van de taalverandering doorheen het lexicon. Meestal vertraagt de verspreiding van de klankverandering weer als de verandering haar voltooiing nadert. In sommige gevallen blijft er een residu van woorden over die uiteindelijk niet aan de klankverandering onderhevig zijn. Deze verspreiding van een klankverandering doorheen de tijd en doorheen het lexicon wordt grafisch voorgesteld als een S-curve. Chambers & Trudgill (1980: 177-178) stellen de 'ideale' S-curve voor met een langzame spreiding doorheen de eerste 20% van het lexicon (d.w.z. 20% van alle morfemen/woorden die aan de structurele condities van de betreffende klankverandering voldoen)¹, gevolgd door een plotse en snelle verspreiding doorheen de volgende 60% van beschikbare gevallen in het lexicon, en afgesloten door een langzame verspreiding doorheen de laatste 20% van beschikbare

¹ Vaak zijn het morfemen/woorden met een hoge token frequentie die het eerst door een klankverandering worden getroffen.

gevallen.² Het is uiteraard onwaarschijnlijk dat een reële klankverandering zich precies volgens deze 20-80% grenzen zal ontwikkelen (cf. Devitt 1989: 35).

In zijn studie naar tweededialectverwerving stelde Chambers vast dat het patroon van die ideale S-curve zich aftekent in de scoregrafieken van de door hem onderzochte kenmerken. Bij de verwerving van fonologische verschijnselen in het kader van tweededialectverwerving komt dit patroon er op neer dat kinderen ofwel onder de 20% scoren op de dialectrealisatie van een bepaalde variabele, ofwel boven de 80%, maar dat er nauwelijks scores liggen tussen 20 en 80%. Volgens Chambers wijst dit patroon erop dat dialectverwerfers voor de eerste 20% van de gevallen die aan de condities van een bepaald fonologisch kenmerk voldoen, de nieuwe uitspraak (i.e. de T2-uitspraak) woord voor woord leren. Nadat er echter een kritische massa van woorden – meer bepaald, instanties van het betreffende kenmerk – verworven is, wordt er een regel gegeneraliseerd, die meteen wordt toegepast op de volgende 60% van de betrokken woorden, wat tot uiting komt in het feit dat er geen scores tussen 20 en 80% zitten. Chambers observeert dit patroon bij de verwerving van drie van de onderzochte fonologische variabelen: ‘T-voicing’, ‘Vowel Backing’ en ‘Low Vowel Merger’. Bij T-voicing moeten dialectverwerfers een kenmerk uit hun T1 leren te onderdrukken, bij Vowel Backing dienen zij een nieuw (T2) foneem te leren gecombineerd met een andere distributie van een foneem dat ze al uit hun T1 kennen, en bij Low Vowel Merger moeten zij twee (T2) fonemen, die in hun T1 samengevallen zijn, leren te onderscheiden. Hieronder worden deze fonologische kenmerken verder toegelicht.

‘T-voicing’ wijst op het fenomeen waarbij in het Canadees-Engels een mediale /t/ stemhebbend wordt gemaakt (dus [d]) tussen een vocaal of /r/ en een onbeklemtoonde syllabe. Door T-voicing komen er in het Canadees-Engels homofone woordparen voor die in het Brits-Engels een verschillende uitspraak hebben, zoals *putting/pudding*, *hearty/heardy*. Canadese kinderen die zich aanpassen aan het Brits-Engels moeten leren om T-voicing te onderdrukken. Van Chambers’ proefpersonen (N = 6) waren er drie die in meer dan 80% van de gevallen T-voicing onderdrukten, en drie die dat in minder dan 20% van de gevallen deden.

‘Vowel Backing’ komt voor in het Brits-Engels dat in Zuid-Engeland gesproken wordt (‘Southern-England-English’) en wijst op het verlengen en velariseren van Middelenlengse /ã/. Dit fenomeen maakte pas opgang in Londen in de 18de eeuw, waardoor het Canadees-Engels het niet onderging. Vowel Backing van de zogenaamde ‘short a’ doet zich voor in twee fonologische contexten: (1) vòòr stemloze anterieure fricatieven (bijv. in *plaster*, *bath*, *laughing*, *class*) en (2) vòòr clusters bestaande uit /n/ + obstruent (bijv. in *dancing*, *branch*, *France*, *plant*). Deze woorden worden in het Canadees-Engels met [æ] uitgesproken en in het Brits-Engels met [ɑ]. Er zijn een aantal uitzonderingen die ook in het Brits-Engels met [æ] uitgesproken worden, zoals *cafeteria*, *classic*, *ant*, *cancer*. Chambers (1992: 155) stelt: “Vowel Backing occurs mainly in frequent words and offers dialect learners ample opportunity to infer a rule (...) from the available evidence”. Hiermee stelt Chambers duidelijk dat fonologische kenmerken van een dialect geleerd worden door middel van regelvorming. Ook met betrekking tot Vowel Backing ziet Chambers een S-vormig patroon in de verwervingsdata: één van de zes proefpersonen scoort 100%, vijf van de zes scoren minder dan 20%, opnieuw zitten er geen data tussen 20 en 80%.

Low Vowel Merger is een kenmerk van het Canadees-Engels: de twee open achtervocalen /ɒ/ en /ɔ:/ zijn versmolten tot één vocaal, meestal de ongeronde /ɑ/. Daardoor ontstaan er homofone woordparen zoals *tot/taught*, *offal/awful*. In het Brits-Engels worden deze vocalen wel gedifferentieerd: woorden als *hot*, *blond*, *knot* worden uitgesproken met /ɒ/, *water*, *lawn*, *Paul*, *strawberry* worden uitgesproken met /ɔ:/. Om zich aan te passen aan het

² Vaak gaat het bij deze laatste 20% om morfemen/woorden met een zeer lage token frequentie. Soms ‘ontsnappen’ deze infrequente woorden dan ook aan de betreffende klankverandering.

Engels van Zuid-Engeland moeten de Canadese proefpersonen in Chambers' studie een onderscheid maken tussen deze twee vocalen en moeten zij Low Vowel Merger onderdrukken. Dit lukte alleen de twee jongste proefpersonen, zij scoorden beiden meer dan 80%. De vier anderen scoorden minder dan 20%.

Op basis van deze observaties argumenteert Chambers dat er na een periode van woord-voor-woord leren een regel gevormd wordt. De verwerving van fonologische kenmerken is dus een combinatie van lexicaal en regelgebaseerd leren. Het lexicaal leren staat echter in functie van de regelvorming: het komt erop neer voldoende gevallen van het betreffende fenomeen te leren om uiteindelijk tot regelvorming te komen.

In deze bijdrage zullen wij nagaan of het patroon van een S-curve ook te zien is in de scoregrafieken van verschillende fonologische kenmerken van het Maldegemse dialect. Op die manier willen wij nagaan of de verwerving van het Maldegemse dialect op regelgebaseerde dan wel op woordgebaseerde leermechanismen gegrond is. In tegenstelling tot Chambers (1992), gaat het hier over een grootschalig onderzoek, waarin de antwoorden van 128 proefpersonen op een woordenlijst van 167 woorden betrokken waren.

2.2. Neighborhood-effecten als indicatie voor woordgebaseerd leren

Als wij in de scoregrafieken van de door ons onderzochte kenmerken S-curves zouden observeren, dan zou dit een aanwijzing vormen voor regelgebaseerd leren: een plotse versnelling van de verwerving na het leren van een kritische massa woorden is volgens Chambers immers een indicatie dat er voor een bepaald kenmerk een regel gevormd is. Als wij daarentegen zogenoemde 'neighborhood effecten' (i.e., effecten van lexicale burens) zouden observeren, dan zou dit een aanwijzing vormen voor het feit dat tweededialectverwerving vooral een woordgebaseerd analogisch proces is. Als de verwerving van een specifiek dialectkenmerk in een specifiek woord beïnvloed wordt door de lexicale burens van dat woord (i.e., woorden in het mentale lexicon die de grootste overeenkomst met het betreffende woord vertonen), dan is dit immers een indicatie voor het feit dat een nieuw woord wordt geclassificeerd op basis van overeenkomst met reeds opgeslagen woorden. Dit wil zeggen dat de dialectuitspraak van de best gelijkende woorden in het mentale lexicon wordt overgenomen voor de uitspraak van het nieuwe woord. We illustreren dit met een voorbeeld uit het Maldegemse dialect: als een kind dat het Maldegemse dialect leert het woord *dijk* wil uitspreken in het dialect, dan zal het in zijn mentale lexicon 'op zoek gaan' naar woorden die op *dijk* lijken, zoals bijvoorbeeld *lijk*, *rijk*, *strijk*, en vervolgens de dialectuitspraak van die woorden, namelijk met de vocaal [e] in plaats van [ɛi], overnemen voor de uitspraak van *dijk*. Dus terwijl regelgebaseerde modellen van taalverwerving abstracties in de vorm van regels aannemen, worden dergelijke abstracties in analogiegebaseerde modellen niet gemaakt: een nieuw woord wordt niet geclassificeerd door gebruik te maken van een of andere abstracte regel, maar door de classificatie (of het 'class label'; in ons geval de dialectuitspraak) van de best gelijkende woorden die in het geheugen zijn opgeslagen (i.e., de zogenaamde 'nearest neighbors') te extrapoleren naar het betreffende woord.

Tussen de nearest neighbors kunnen zich zowel 'friendly neighbors' als 'enemy neighbors' bevinden. Friendly neighbors hebben dezelfde dialectuitspraak als het betreffende woord, enemy neighbors hebben een andere classificatie. Toegepast op de verwerving van de fonologie van het Maldegemse dialect: *eik* [ɛik] wordt in het dialect uitgesproken als [iə̯k]. In een classificatietaak waarin we de dialectuitspraak van de vocaal in woorden willen voorspellen, zal het woord *eik* het class label [iə̯] hebben. Mogelijke neighbors van *eik* zijn

onder andere *rijk, lijk, kijk*.³ Al deze neighbors hebben echter de dialectuitspraak [e], respectievelijk [ˈreʔə], [lek] en [kek]. Met andere woorden, wat betreft de dialectuitspraak van de vocaal, zijn het enemy neighbors van het woord *eik*. Als een nieuw woord wordt geclassificeerd op basis van overeenkomst met een enemy neighbor, treedt er overgeneralisatie op. Dit betekent dat in het bovenstaande voorbeeld er sprake zou zijn van een overgeneralisatie als de dialectuitspraak van de enemy neighbors (nl. het class label [e]) geëxtrapoleerd wordt naar *eik*, met als resultaat *[ek] in plaats van het correcte [iək].

Effecten van (enemy) neighbors op de verwerving van dialectvarianten en op eventuele overgeneralisaties zouden een sterke aanwijzing kunnen vormen voor woordgebaseerd leren in tweededialectverwerving. In deze studie onderzoeken wij dergelijke effecten. Daarvoor brengen we eerst de lexicale burens in kaart van woorden die we opvroegen bij onze Maldegemse proefpersonen. Dit gebeurt met behulp van een classificatietask in TiMBL, die in de volgende paragraaf besproken wordt.

2.3. Geheugengebaseerde taalverwerking

Analogiegebaseerd leren, zoals hierboven beschreven, is kenmerkend voor geheugengebaseerde taalverwerking ('memory-based language processing'). Daelemans & Van den Bosch (2005) ontwikkelden een computationele implementatie van woordgebaseerd analogisch leren, die uitgebreid werd voor natuurlijke taalverwerking (NLP), namelijk TiMBL (Tilburg Memory-Based Learner). In deze studie gebruiken we TiMBL voor een classificatietask, waarbij Standaardnederlandse woordvormen als test items aan het model gepresenteerd worden en het model de Maldegemse dialectvarianten moet voorspellen.

Geheugengebaseerd leren zoals dat in TiMBL wordt geïmplementeerd, is gebaseerd op het opslaan van voorbeelden (of instanties) in het geheugen en het bepalen van de meest gelijkende voorbeelden door middel van een zogenoemde similariteitsmetriek. Zoals in de vorige paragraaf al beschreven werd, wordt de classificatie of het class label van die nearest neighbors overgenomen als classificatie van een nieuw item. Op die manier wordt in een classificatietask in TiMBL rechtstreeks gebruik gemaakt van opgeslagen voorbeelden in plaats van eerst generalisaties – bijvoorbeeld in de vorm van regels – af te leiden uit die voorbeelden.

TiMBL gaat als volgt te werk voor de classificatietask in deze studie: in de leerfase of trainingfase worden Standaardnederlandse woordvormen met hun correcte dialectuitspraak (de zogenoemde training items) opgeslagen in het systeem, zonder enige vorm van abstractie, selectie of herstructurering. In de testfase of prestatiefase worden de training items gebruikt als basis om nieuwe input (namelijk de test items) om te zetten in output, namelijk door de juiste dialectuitspraak toe te kennen aan de test items. Dit gebeurt op basis van overeenkomst (i.e., similariteit). Het aantal nearest neighbors dat betrokken wordt in deze procedure wordt weergegeven als k . De waarde van k kan op voorhand worden vastgelegd. Voor de classificatietask in deze studie werd de waarde van k op 5 gebracht,⁴ wat wil zeggen dat de dialectuitspraak van de vijf best gelijkende training items in rekening werd gebracht bij het bepalen van de dialectuitspraak van een test item. Figuur 1 representeert bijvoorbeeld de vijf training items (namelijk: *buis, sluis, huid, kruis, pluïs*) die – wat betreft de dialectuitspraak

³ De gegeven voorbeelden zijn allemaal zogenoemde 'rhyme neighbors': de woorden vertonen overeenkomst wat betreft hun rijm. De best gelijkende burens van een woord hoeven echter niet noodzakelijk 'rhyme neighbors' te zijn.

⁴ De gekozen waarde van k is een goede trade-off tussen accuraatheid van het model en aantal nearest neighbors. Deze waarde werd enkel op basis van de trainingitems vastgesteld, zodat er geen overfitting op de testitems was.

van de vocaal – het meest overeenkomen met het test item *huis*, dat in het Maldegemse dialect wordt uitgesproken als [ø̯s].

```

huis h ø̯ s class: ø predicted: ø { ø 0.8, ø̯ 0.2 }
# k=1, 1 Neighbor(s) at distance: 0.096
#      b ø̯ s { ø 1.0 }
# k=2, 1 Neighbor(s) at distance: 0.106
#      sl ø̯ s { ø 1.0 }
# k=3, 1 Neighbor(s) at distance: 0.110
#      h ø̯ t { ø̯ 1.0 }
# k=4, 1 Neighbor(s) at distance: 0.116
#      kr ø̯ s { ø 1.0 }
# k=5, 1 Neighbor(s) at distance: 0.125
#      pl ø̯ s { ø 1.0 }

```

Figuur 1: Nearest neighbors van *huis* voor classificatie van de vocaal in *huis* (k = 5)

Uit Figuur 1 kan het volgende worden afgeleid: het test item *huis* werd door TiMBL correct geclassificeerd wat betreft de uitspraak van de vocaal: TiMBL voorspelt [ø̯] (zie Fig. 1: predicted: ø̯), wat overeenkomt met de Maldegemse uitspraak (zie Fig. 1: class: ø̯). Voor die classificatie is er 80% evidentie, terwijl er 20% evidentie is voor een classificatie als [ø̯] (zie Fig. 1: { ø̯ 0.8, ø̯ 0.2 }). De 80% evidentie voor de voorspelling [ø̯] is afkomstig van vier nearest neighbors die elk de dialectvocaal [ø̯] hebben, namelijk *buis* (zie Fig. 1: b ø̯ s { ø̯ 1.0}), *sluis* (zie Fig. 1: sl ø̯ s { ø̯ 1.0 }), *kruis* (zie Fig. 1: kr ø̯ s { ø̯ 1.0 }) en *pluis* (zie Fig. 1: pl ø̯ s { ø̯ 1.0 }). De 20% evidentie voor de voorspelling [ø̯] is afkomstig van één nearest neighbor die de dialectvocaal [ø̯] heeft, namelijk *huid* (zie Fig. 1: h ø̯ t { ø̯ 1.0 }). De dialectvocaal van het merendeel van de nearest neighbors wordt dus overgenomen voor de voorspelling of classificatie van het testwoord, wat in dit geval de correcte dialectuitspraak oplevert. In het voorbeeld in Figuur 1 zijn de vier neighbors die de correcte classificatie aanleveren zogenoemde friendly neighbors, terwijl de neighbor *huid* een enemy neighbor is, namelijk een woord dat een andere dialectvocaal heeft dan het testwoord *huis*.

De training items in de classificatietaak zijn Standaardnederlandse woordvormen en hun dialectequivalenten die gerepresenteerd worden als gesyllabificeerde reeksen van segmenten en die met elkaar gealligeneerd zijn (zie Tabel 1). De Standaardnederlandse woordvormen hebben nooit meer dan twee syllaben (zie paragraaf 3.4), maar door een regel van sjwa-toevoeging aan het wordeinde in het Maldegemse dialect kunnen de dialectvormen uit drie syllaben bestaan. Syllaben worden telkens opgedeeld in de eigenschappen onset, nucleus en coda. De waarden van die eigenschappen zijn dan de specifieke fonetische segmenten. Tabel 1 geeft bijvoorbeeld de Standaardnederlandse vorm en dialectvorm van het woord *vogel* weer zoals die in de trainingdata zijn gerepresenteerd. Het symbool '=' betekent dat de waarde van een specifieke eigenschap oningevuld blijft; S.N. staat voor Standaardnederlands, DIA voor dialect.

Eigenschappen	Waarden	
	S.N. vogel	DIA vogel
Onset eerste syllabe (onset ₁)	v	v
Nucleus eerste syllabe (nucleus ₁)	o	ø
Coda eerste syllabe (coda ₁)	=	=
Onset tweede syllabe (onset ₂)	ɣ	h
Nucleus tweede syllabe (nucleus ₂)	ə	ə
Coda tweede syllabe (coda ₂)	l	=
Onset derde syllabe (onset ₃)	=	l
Nucleus derde syllabe (nucleus ₃)	=	ə
Coda derde syllabe (coda ₃)	=	=

Tabel 1. Representatie van de training items in TiMBL: codering van de Standaardnederlandse vorm en dialectvorm van het woord *vogel*

De classificatietaak in TiMBL bestond erin om bij woorden die geen deel uitmaakten van de trainingdata voor elk Standaardnederlands segment het Maldegemse segment te voorspellen. Die omzettingen gebeurden dus op segmentniveau. Enkel de omzettingen die relevant waren voor de fonologische kenmerken die we beschouwen (zie par. 3.4), werden geanalyseerd.

Zoals hierboven beschreven werd, gebeurt de classificatie in TiMBL op basis van gelijkenis tussen het test item en de nearest neighbors of meest gelijkende woordvormen in de trainingdata. In ons geval werden de vijf best gelijkende neighbors van het test item in beschouwing genomen. Voor het bepalen van die neighbors maakt TiMBL gebruik van een zogenoemde ‘overlap metric’: de overeenkomst tussen een test item en alle opgeslagen items wordt berekend door de som te maken van het aantal overlappende waarden. Volgens deze berekening heeft *vogel* (v ø = h ə = l ə =) bijvoorbeeld 7 overlappende waarden met het woord *kogel*, dat in het dialect wordt uitgesproken als [kohələ] en dat in de trainingdata gerepresenteerd wordt als (k o = h ə = l ə =) (nl. enkel de waarden van de twee eerste eigenschappen verschillen). Het woord dat de grootste overlap vertoont met het test item is de nearest neighbor van dat item. Van alle training items vertoont de nearest neighbor de kleinste afstand tot het test item. Omdat echter niet alle eigenschappen even relevant zijn voor de classificatietaak (bijv. het rijm (i.e. nucleus + coda) kan meer bepalend zijn voor de classificatietaak dan onset + nucleus), maken we gebruik van Information Gain (IG) Weighting (zie Daelemans & Van den Bosch, 2005: 29-31). Door deze functie toe te voegen aan het algoritme dat de overlap metric berekent, wordt elke eigenschap (bijv. onset₁, nucleus₁, coda₁, etc.) afzonderlijk geëvalueerd, om op die manier te schatten hoeveel informatie die eigenschap bijdraagt aan de voorspelling van de correcte dialectuitspraak. Op basis van deze overlap metric met weegfunctie, werden in de classificatietaak in deze studie per test item telkens de vijf ($k = 5$) best gelijkende training items geselecteerd. Vervolgens kende het model de dialectuitspraak die bij die vijf nearest neighbors het frequentst voorkwam toe aan het test item (zie Figuur 1).

3. Methodologie

In deze paragraaf bespreken we de methodologie van onze studie. We bespreken achtereenvolgens de onderzoekslocatie, de selectie van proefpersonen, de opnameprocedure, de woordenlijst, fonologische kenmerken en afhankelijke variabelen, de training en test items en tenslotte de manier waarop neighborhood effecten gemeten werden en waarop we de scoregrafieken onderzochten op het voorkomen van S-curves.

3.1. Onderzoekslocatie

De gegevens werden verzameld in Maldegem. Dit is een gemeente in het uiterste noordwesten van Oost-Vlaanderen (België). Het dialect van Maldegem is een overgangsdialect tussen West- en Oost-Vlaamse dialecten, met een aantal eigen fonologische kenmerken (zie Rys, 2007: 174-180). Het Maldegemse dialect wordt nog door vrij veel mensen gesproken in de dagelijkse communicatie.

3.2. Proefpersonen

Op basis van een enquête ingevuld door de ouders werden 128 kinderen geselecteerd om deel te nemen aan het onderzoek. Deze kinderen hadden altijd in Maldegem gewoond en gingen er naar school, maar ze werden opgevoed in het Standaardnederlands door hun ouders. Er werden kinderen van drie leeftijdsgroepen (in het schooljaar 2003-2004) geïnterviewd, namelijk negen-, twaalf- en vijftienjarigen.

3.3. Procedure

Aan de proefpersonen werd gevraagd een aantal plaatjes te benoemen en een aantal zinnen aan te vullen, en daarbij te proberen de Maldegemse dialectvariant te geven. De antwoorden werden geregistreerd op minidisc met een SONY MZ-N707 draagbare recorder en met een SONY ECM-ZS90 Electret condenser microfoon. Vervolgens werden de opnames gedigitaliseerd en gesampled op 44 kHz, 16-bit stereo. Tijdens de opnames sprak de interviewer (i.e. de eerste auteur van dit artikel) steeds zelf het Maldegemse dialect (als moedertaalspreker) om op die manier een zo informeel mogelijke situatie te creëren waarin de proefpersoon zich vrij zou voelen om zelf dialectvormen te produceren. De opnames werden integraal fonetisch getranscribeerd.

3.4. Woordenlijst, fonologische kenmerken en afhankelijke variabelen

De benoemingstaak en zinsvervolgding werden aangewend om een woordenlijst van 167 woorden op te vragen. De woordenlijst bevat frequente en minder frequente mono- en disyllabische woorden die representatief zijn voor een twintigtal fonologische kenmerken van het Maldegemse dialect. Het gaat om 115 substantieven, 32 werkwoorden (infinitief) en 20 adjectieven. Alle woorden zijn cognaten: de Standaardnederlandse (verderop S.N.) vorm en de Maldegemse dialectvariant zijn fonologische varianten van hetzelfde lexem (bijv. *krijt*: S.N. [krɛit] vs. dialect [krøt]).

De volgende fonologische dialectkenmerken werden in ons onderzoek opgenomen:

- (1) Deletie van /l/ en compensatorische rekking van de voorafgaande vocaal
 - (a) vóór een pauze; bijv. *bal*: S.N. [bɑl] vs. dialect [bɑ:]
 - (b) vóór een consonant; bijv. *melk*: S.N. [mɛlk] vs. dialect [mæ:k]
- (2) Deletie van /n/ en compensatorische rekking en nasalisatie van de voorafgaande vocaal vóór een alveolaire fricatief; bijv. *spons*: S.N. [spɔns] vs. dialect [ˈspõ:sə]
- (3) Glottalisatie van /k/ tussen een beklemtoonde en een onbeklemtoonde vocaal; bijv. *kijken*: S.N. [kɛikən] vs. dialect [ˈkɛʔə̃]
- (4) Deletie van /r/ vóór een alveolaire fricatief in een beperkt aantal monomorfematische woorden; bijv. *worst*: S.N. [wɔrst] vs. dialect [wɔst]
- (5) S.N. /ɛi/ vs. dialect /e/ vóór een velaire of laryngale consonant (verderop de zogenaamde *rijk*-variabele); bijv. S.N. [rɛik] vs. dialect [ˈreʔə]
- (6) S.N. /ɛi/ vs. dialect /ø/ vóór een anterieure consonant (*wijn*-variabele); bijv. *wijn*: S.N. [wɛin] vs. dialect [wøn]
- (7) S.N. /ɛi/ vs. dialect /iə/ (*geit*-variabele); bijv. *geit*: S.N. [ɣɛit] vs. dialect [ˈhiətə]
- (8) S.N. /ɛi/ vs. dialect /æi/ (*kei*-variabele); bijv. *kei*: S.N. [kɛi] vs. dialect [kæi]
- (9) S.N. /œy/ vs. dialect /ø/ in alle posities behalve op het wordeinde (*duim*-variabele); bijv. *duim*: S.N. [dœym] vs. dialect [døm]
- (10) S.N. /o/ vs. dialect /ø/ (*zoon*-variabele); bijv. *zoon*: S.N. [zon] vs. dialect [ˈzønə]
- (11) S.N. /o/ vs. dialect /uə/ (*roos*-variabele); bijv. *roos*: S.N. [ros] vs. dialect [ˈruəzə]
- (12) S.N. /ʌ/ vs. dialect /ɛ/ (*put*-variabele); bijv. *put*: S.N. [pʌt] vs. dialect [pɛt]
- (13) S.N. /ɔ/ vs. dialect /ɛ/ (*pop*-variabele); bijv. *pop*: S.N. [pɔp] vs. dialect [ˈpɛpə]
- (14) S.N. /e/ vs. dialect /iə/ (*been*-variabele); bijv. *been*: S.N. [ben] vs. dialect [biən]
- (15) S.N. /e/ vs. dialect /ɛ/ vóór /r/ (*peer*-variabele); bijv. *peer*: S.N. [per] vs. dialect [ˈpɛrə]
- (16) S.N. /au/ vs. dialect /ai/ (*kous*-variabele); bijv. *kous*: S.N. [kaus] vs. dialect [ˈkaisə]
- (17) S.N. /a/ vs. dialect /ɔ:/ (*maan*-variabele); bijv. *maan*: S.N. [ma:n] vs. dialect [ˈmɔ:nə]
- (18) S.N. /a/ vs. dialect /ɛ/ vóór /r/ + alveolaire consonant (*paard*-variabele); bijv. *paard*: S.N. [part] vs. dialect [pɛrt]
- (19) S.N. /ɪ/ vs. dialect /æ/ (*rib*-variabele); bijv. *rib*: S.N. [rɪp] vs. dialect [ˈræbə]

Van de bovenstaande fonologische kenmerken zijn (1), (2), (3), (4), (5), (6), (9), (15) en (18) fonologisch geconditioneerd; dit wil zeggen dat er een fonologische context gedefinieerd kan worden (bijv. ‘vóór laryngale of velaire consonant’) die bepalend is voor het betreffende kenmerk. De overige kenmerken zijn lexicaal gedetermineerd; dit wil zeggen dat het op basis van de fonologische context onvoorspelbaar is of een specifiek kenmerk al of niet van toepassing is. Met andere woorden, of het kenmerk wordt toegepast, hangt af van woord tot woord.

In ons onderzoek zijn er twee afhankelijke variabelen betrokken. Enerzijds is er de vraag of een proefpersoon voor een specifiek woord de correcte Maldegemse variant realiseert of niet (i.e., de variabele *dialectrealisatie*); met andere woorden: wordt een specifiek dialectkenmerk correct toegepast of niet. Anderzijds bestaat de mogelijkheid dat een proefpersoon in plaats van de dialectvariant een andere vorm realiseert. Die andere vorm kan onder andere een overgeneralisatie zijn. De vraag of er een overgeneralisatie werd geproduceerd of niet, is de andere afhankelijke variabele in deze studie (i.e., de variabele *overgeneralisatie*). Met overgeneralisatie bedoelen wij (1) de toepassing van een fonologisch kenmerk (bijv. van (9) S.N. /œy/ vs. dialect /ø/) in een woord dat wel voldoet aan de

fonologische condities van dat kenmerk (namelijk ‘alle posities behalve wordeinde’), maar dat een lexicale uitzondering vormt (bijv. *spuit* → dialect [‘spæitə] of [‘spiətə], niet *[‘spøtə]), en (2) de toepassing van een kenmerk (bijv. (6) S.N. /ɛi/ vs. dialect /ø/) in een woord (bijv. *rijk*) dat niet voldoet aan de fonologische condities van dat kenmerk (namelijk ‘vóór een anterieure consonant’).

3.5. Training items en test items

Voor de training items in de classificatietaak werd geput uit een database van Standaardnederlandse woorden en hun ‘vertaling’ in het Maldegemse dialect. Deze database kwam tot stand op basis van de CELEX lexicale database voor het Nederlands (Baayen, Piepenbrock & Gulikers, 1995). De CELEX database bevat 42.380.000 woorden en is gebaseerd op een corpus van geschreven taal. Uit deze database selecteerden we de mono- en disyllabische monomorfematische woorden die bekend waren bij 2/3 van de 39 deelnemers in een grootschalig lexicale decisie-experiment (Keuleers, Diependaele & Brysbaert, 2010). Op die manier hielden we 3524 woorden over, die we vervolgens omzetten in Maldegemse dialectwoorden. Daarbij hielden we rekening met de Maldegemse fonologie zoals die beschreven is in Taeldeman (1976) en Versieck (1989) en met de kennis van het Maldegemse dialect van de eerste auteur van deze bijdrage. Omdat we in het geval van homografie enkel de eerst voorkomende woordvorm in de database behielden, werden de trainingdata nog verder gereduceerd tot 2435 woorden, wat overeenkwam met 9867 segmenten.

Zowel de Standaardnederlandse woorden als de dialectwoorden werden fonetisch gerepresenteerd en in syllabes ingedeeld. Op die manier werden de Standaardnederlandse woorden gealligeneerd met hun dialectvarianten (zie par. 2.3). Classificatie in TiMBL gebeurde op segmentniveau: bij de classificatie van het woord *huis* [h œy s], bijvoorbeeld, werd door het model een voorspelling gemaakt voor het eerste segment [h], vervolgens voor het tweede segment [œy] en tenslotte voor het derde segment [s]. De vijf nearest neighbors van *huis* bij de classificatie van het eerste segment verschillen dan ook van de nearest neighbors bij de classificatie van het tweede of derde segment. Omdat de meerderheid van de onderzochte variabelen vocalisch is (behalve variabelen (1) – (4); zie paragraaf 3.4), is voor deze studie meestal enkel de classificatie van de vocaal van belang.

De classificatietaak werd drie maal uitgevoerd: met 1000 segmenten, met 5000 segmenten, en met 9867 segmenten (of 2435 woorden)⁵ als trainingdata. De gegevens over aantal friendly vs. enemy neighbors van de testwoorden werden enkel verzameld op basis van de laatste classificatietaak, namelijk die met de volledige database als trainingdata.⁶

De test items, i.e., de woorden waarvan TiMBL de dialectvariant moest voorspellen, waren dezelfde woorden als de woorden die bij de Maldegemse proefpersonen waren opgevraagd (zie par. 3.4).

3.6. Meten van neighborhood effecten

Zoals we hierboven beschreven hebben, willen we in deze studie nagaan of er zich neighborhood effecten voordoen in tweedialectverwerving. Meer bepaald onderzoeken we het effect van aantal enemy neighbors op de mate van verwerving van dialectkenmerken, alsook op de mate van overgeneralisatie van kenmerken. Zoals in Figuur 1 geïllustreerd werd,

⁵ 9867 segmenten was het maximum aantal beschikbare trainingdata.

⁶ Uiteraard maakten de testwoorden geen deel uit van de trainingdata.

genereerde TiMBL voor elk test item vijf nearest neighbors. De nearest neighbors die dezelfde dialectuitspraak vertonen als de dialectvariant van het testwoord zijn de zogenoemde friendly neighbors; de neighbors die een andere dialectuitspraak vertonen, zijn enemy neighbors van het testwoord. Voor elk testwoord werd telkens het aantal enemy neighbors geteld. Vervolgens werd het effect van ‘aantal enemy neighbors’ op het al of niet realiseren van de correcte dialectvariant (i.e., de afhankelijke variabele *dialectrealisatie*) en op het al of niet produceren van overgeneralisaties (i.e., de afhankelijke variabele *overgeneralisatie*) nagegaan door middel van binaire logistische regressie-analyses.

3.7. Vaststellen van S-curves in scoregrafieken

Om te kunnen nagaan of de verwerving van de Maldegemse fonologische kenmerken een S-vormig patroon vertoont, hebben we per kenmerk (zie paragraaf 3.4) en per leeftijdsgroep (resp. negen-, twaalf- en vijftienjarigen) scoregrafieken gemaakt waaruit duidelijk werd hoeveel proefpersonen een specifieke score (in %) behaalden. Vervolgens werd het mogelijk om, net als Chambers, na te gaan of de scores vooral verdeeld zijn over de uiteinden (namelijk minder dan 20% of meer dan 80%) of dat er ook scores in het midden zitten (namelijk tussen 20 en 80%).

4. Resultaten

In deze paragraaf gaan we eerst na of we in de verdeling van de scores voor elk van de onderzochte fonologische kenmerken S-vormige patronen vinden (paragraaf 4.1). Vervolgens worden de resultaten van de classificatietaak besproken (paragraaf 4.2) en tenslotte bespreken we de effecten van het aantal enemy neighbors op de mate van dialectrealisatie (paragraaf 4.3.1) en op de mate van overgeneralisatie (paragraaf 4.3.2).

4.1. Verdeling van de scores

Tabel 2 representeert per fonologisch kenmerk en per leeftijdsgroep de verdeling van de scores op de afhankelijke variabele *dialectrealisatie*. Er wordt onderscheid gemaakt tussen het aantal scores van minder dan 20%, het aantal scores tussen 20 en 80% en het aantal scores van meer dan 80%. In Tabel 2 wordt verder nog aangeduid of een kenmerk fonologisch geconditioneerd (weergegeven als ‘F’) of lexicaal bepaald (‘L’) is.

Fonologisch Kenmerk	Fonologisch geconditioneerd (F) of lexicaal bepaald (L)	Aantal scores (N)		
		-20%	20%-80%	+80%
Leeftijd: 9 jaar				
(N = 55)				
(1) l-deletie	F	25	27	3
(2) n-deletie	F	24	12	19
(3) k-glottalisatie	F	9	26	20
(4) r-deletie	F	28	26	1
(5) 'rijk'-variabele	F	14	32	9
(6) 'wijn'-variabele	F	10	24	21
(7) 'geit'-variabele	L	38	17	0
(8) 'kei'-variabele	L	34	17	4
(9) 'duim'-variabele	F	7	18	30
(10) 'zoon'-variabele	L	22	28	5
(11) 'roos'-variabele	L	22	28	5
(12) 'put'-variabele	L	43	12	0
(13) 'pop'-variabele	L	36	18	1
(14) 'been'-variabele	L	33	16	6
(15) 'peer'-variabele	F	32	23	0
(16) 'kous'-variabele	L	32	9	14
(17) 'maan'-variabele	L	15	18	22
(18) 'paard'-variabele	F	20	28	7
(19) 'rib'-variabele	L	7	48	0
Leeftijd: 12 jaar				
(N = 51)				
(1) l-deletie	F	13	23	15
(2) n-deletie	F	2	13	36
(3) k-glottalisatie	F	5	15	31
(4) r-deletie	F	15	36	0
(5) 'rijk'-variabele	F	6	25	20
(6) 'wijn'-variabele	F	2	13	36
(7) 'geit'-variabele	L	12	38	1
(8) 'kei'-variabele	L	20	23	8
(9) 'duim'-variabele	F	2	8	41
(10) 'zoon'-variabele	L	1	35	15
(11) 'roos'-variabele	L	10	27	14
(12) 'put'-variabele	L	29	20	2
(13) 'pop'-variabele	L	17	32	2
(14) 'been'-variabele	L	11	11	29
(15) 'peer'-variabele	F	5	30	16
(16) 'kous'-variabele	L	15	8	28
(17) 'maan'-variabele	L	2	16	33
(18) 'paard'-variabele	F	3	36	12
(19) 'rib'-variabele	L	2	48	1

Leeftijd: 15 jaar (N = 22)				
(1) l-deletie	F	4	11	7
(2) n-deletie	F	1	2	19
(3) k-glottalisatie	F	0	5	16
(4) r-deletie	F	5	13	4
(5) 'rijk'-variabele	F	0	6	16
(6) 'wijn'-variabele	F	1	7	14
(7) 'geit'-variabele	L	4	18	0
(8) 'kei'-variabele	L	11	7	4
(9) 'duim'-variabele	F	1	5	16
(10) 'zoon'-variabele	L	1	12	9
(11) 'roos'-variabele	L	3	13	6
(12) 'put'-variabele	L	5	16	1
(13) 'pop'-variabele	L	6	15	1
(14) 'been'-variabele	L	3	6	13
(15) 'peer'-variabele	F	3	10	9
(16) 'kous'-variabele	L	4	5	13
(17) 'maan'-variabele	L	2	4	16
(18) 'paard'-variabele	F	2	15	5
(19) 'rib'-variabele	L	1	19	2

Tabel 2. Verdeling van de scores per kenmerk op de afhankelijke variabele *dialectrealisatie*.

Tabel 2 laat zien dat er voor elk fonologisch kenmerk bij elke leeftijdsgroep scores tussen 20 en 80% zitten. Voor de meeste kenmerken is dit aantal zelfs relatief hoog. In tegenstelling tot Chambers, die voor T-voicing, Vowel Backing en Low Vowel Merger geen scores vond tussen 20 en 80%, vinden wij dus voor alle kenmerken wel scores die in het midden zitten. Er zijn dus geen duidelijke patronen – meer bepaald S-curves – te onderscheiden die in de richting wijzen van regelgebaseerd leren. In het geval van de *kous*-variabele, is er in zekere mate een S-curve te onderscheiden, in die zin dat er bij de negen- en twaalfjarigen aanzienlijk meer scores aan een van de uiteinden vallen, dan in de middelste groep. Toch heeft ook dit kenmerk steeds een aantal scores tussen 20 en 80%.

Er zijn geen opvallende verschillen waar te nemen tussen fonologisch geconditioneerde (F) en lexicaal bepaalde (L) kenmerken: bij beide typen kenmerken valt een aanzienlijk deel van de scores tussen 20 en 80%. In een regelgebaseerd model zou je verwachten dat er wel regelvorming optreedt bij de fonologisch geconditioneerde kenmerken (dus wel een S-curve), maar niet bij de kenmerken die men woord voor woord moet leren (geen S-curve). Een dergelijk onderscheid komt echter niet uit Tabel 2 naar voren.

4.2. Accuraatheid in de classificatietask in TiMBL

In Tabel 3 wordt het aantal door TiMBL correct geclassificeerde test items per fonologisch kenmerk en de daarbij horende accuraatheidsscores ('accuracy scores') gepresenteerd. Deze gegevens zijn gebaseerd op drie classificatietaken met respectievelijk 1000, 5000 en 9867 segmenten als trainingdata. In de rechterkolom staan de percentages die de negen-, twaalf- en vijftienjarige tweedialectverwerwers behaalden voor het betreffende fonologische kenmerk. In de tweede kolom wordt aangegeven of een kenmerk fonologisch geconditioneerd (F) is of

lexicaal bepaald (L). De derde kolom geeft de typefrequentie van de fonologische kenmerken. De typefrequentie werd berekend op basis van het aantal voorkomens van een kenmerk in de database van 3524 Maldegemse woorden die we samenstelden op basis van de CELEX database (zie paragraaf 3.5).

Fonologisch kenmerk	F of L	Type-freq.	Aantal test items	Percentage correct geclassificeerd ('accuracy score') door TiMBL bij verschillend aantal segmenten als trainingdata:			Percentage correcte dialectrealisatie bij tweededialectverwervers per leeftijdsgroep:		
				1000	5000	9867	9 jaar	12 jaar	15 jaar
(1) l-deletie	F	294	10	0%	50%	70%	49%	59%	61%
(2) n-deletie	F	51	3	0%	33%	67%	52%	87%	91%
(3) k-glottalisatie	F	78	2	0%	100%	100%	63%	79%	87%
(4) r-deletie	F	16	7	0%	57%	57%	22%	42%	57%
(5) 'rijk'-variabele	F	18	4	0%	75%	25%	39%	62%	80%
(6) 'wijn'-variabele	F	79	9	78%	89%	89%	61%	83%	78%
(7) 'geit'-variabele	L	7	3	0%	0%	0%	14%	42%	57%
(8) 'kei'-variabele	L	15	5	0%	0%	0%	27%	49%	53%
(9) 'duim'-variabele	F	91	8	100%	100%	100%	69%	88%	87%
(10) 'zoon'-variabele	L	16	4	0%	0%	0%	37%	67%	74%
(11) 'roos'-variabele	L	145	8	100%	88%	100%	37%	65%	67%
(12) 'put'-variabele	L	34	6	0%	0%	0%	12%	36%	50%
(13) 'pop'-variabele	L	10	7	0%	0%	0%	21%	45%	49%
(14) 'been'-variabele	L	94	6	0%	50%	67%	35%	71%	81%
(15) 'peer'-variabele	F	10	3	0%	0%	0%	24%	68%	80%
(16) 'kous'-variabele	L	5	5	0%	0%	0%	37%	66%	76%
(17) 'maan'-variabele	L	294	4	100%	100%	100%	65%	75%	74%
(18) 'paard'-variabele	F	9	4	0%	0%	0%	37%	69%	72%

Tabel 3. Percentage correcte classificaties in de classificatietaak in TiMBL op basis van 1000, 5000 en 9867 segmenten als trainingdata en percentage correcte dialectrealisaties bij negen-, twaalf- en vijftienjarige tweededialectverwervers.

Tabel 3 laat zien dat TiMBL voor sommige dialectkenmerken een accuraatheid heeft van 100%, zelfs bij een vrij beperkte set aan trainingdata (namelijk 1000 segmenten). Dit betekent dat het model voor alle testwoorden de correcte dialectuitspraak voorspelt. Meer bepaald gaat het om de 'duim'-variabele, de 'roos'-variabele en de 'maan'-variabele. Opvallend is dat deze kenmerken steeds een relatief hoge typefrequentie hebben: de 'duim'-variabele kwam in 91 woorden van de Maldegemse CELEX voor, de 'roos'-variabele in 145 woorden, en de 'maan'-variabele in 294 woorden. De relatief hoge typefrequentie impliceert dat er in de trainingdata een groot aantal instanties van deze kenmerken voorkomt. Hierdoor zullen er – zelfs bij een relatief kleine trainingset – vrij homogene neighborhoods ontstaan, die de woorden waarop deze kenmerken van toepassing zijn, clusteren. Als een groot aantal neighbors dezelfde dialectuitspraak heeft als het test item, bevordert dat de juiste voorspelling

van de uitspraak van het test item (cf. Fig. 1). Uit een Pearson correlatie analyse bleek dat er inderdaad een significante samenhang is tussen type frequentie en de accuraatheidsscores bij een trainingset van respectievelijk 1000, 5000 en 9867 segmenten (resp. $r = 0,509^*$, $r = 0,557^*$ en $r = 0,656^{**}$).⁷

Tabel 3 laat verder zien dat sommige kenmerken totaal niet ‘geleerd’ worden door TiMBL, zelfs niet bij een maximale trainingset. Het gaat om de ‘geit’-variabele, de ‘kei’-variabele, de ‘zoon’-variabele, de ‘put’-variabele, de ‘pop’-variabele, de ‘peer’-variabele, de ‘kous’-variabele en de ‘paard’-variabele. Van deze acht variabelen zijn er zes lexicaal bepaald. Er lijkt dus een samenhang te zijn tussen de accuraatheid van het model en de vraag of een kenmerk lexicaal bepaald of fonologisch geconditioneerd is. Het gebrek aan een conditionerende fonologische context bij lexicaal bepaalde kenmerken impliceert dat de woorden waarop een dergelijk kenmerk van toepassing is niet omgeven zijn door gelijkkluidende neighbors die dezelfde dialectuitspraak als classificatie hebben. Het is immers zo dat de woorden waarop dergelijke kenmerken van toepassing zijn een zeer verschillende fonologische structuur kunnen hebben, waardoor ze niet samenclusteren in een homogene neighborhood. Daardoor wordt het voorspellen van de correcte dialectuitspraak bemoeilijkt. Bovendien hebben de acht kenmerken die niet ‘geleerd’ worden een relatief lage type frequentie. Bij een aantal van deze kenmerken was het aantal test items ten opzichte van de totale type frequentie waarschijnlijk dusdanig hoog dat er na het verwijderen van de test items uit de trainingset slechts een beperkt aantal instanties van het kenmerk overbleven.

Samenvattend kunnen we stellen dat de combinatie van lexicale determinatie en een lage type frequentie ervoor zorgt dat TiMBL geen juiste voorspellingen kan doen voor bepaalde kenmerken. Dit is vooral te wijten aan het feit dat er in het geval van deze kenmerken geen neighborhood kan ontstaan van woorden die allemaal in de richting van dezelfde dialectvariant wijzen. Aangezien ons model classificeert op basis van de dialectuitspraak die het frequentst voorkomt binnen de vijf nearest neighbors, ligt het voor de hand dat er verkeerd geclassificeerd wordt als de set van nearest neighbors erg heterogeen is.

Bij een aantal kenmerken presteert het model duidelijk beter naarmate er meer trainingdata zijn (bijv. l-deletie, n-deletie, ‘been’-variabele). Dit impliceert dat er een minimum aantal instanties van een specifiek kenmerk noodzakelijk is om het kenmerk goed te kunnen voorspellen. Met andere woorden, het ontstaan van een grotere homogene neighborhood bevordert de juiste voorspelling van een specifiek kenmerk. Bij de ‘rijk’-variabele doet TiMBL echter 75% goede voorspellingen met een trainingset van 5000 segmenten, terwijl het slechts 25% goede voorspellingen doet met een trainingset van 9867 segmenten. Dit lijkt het resultaat te zijn van toeval. Ook bij 5000 segmenten is de foute vorm vaak even frequent aanwezig bij de nearest neighbors als de juiste vorm. Toevallig kiest het algoritme in die gevallen steeds de correcte vorm. Bij de grotere trainingset begint de foute vorm te overheersen en wordt dan ook steeds gekozen.

Wanneer we tenslotte de accuraatheidsscores van TiMBL vergelijken met de gemiddelde scores van de proefpersonen, vallen de volgende observaties op. Ten eerste zien we dat de scores voor alle kenmerken progressie vertonen tussen negen en vijftien jaar. In tegenstelling tot de resultaten van het model, is het nooit zo dat een kenmerk helemaal niet geleerd wordt. Ten tweede doen zich bij de scores van de proefpersonen plafondeffecten voor: er is nog maar nauwelijks progressie tussen de leeftijd van twaalf en vijftien jaar in het geval van l-deletie, n-deletie, de ‘wijn’-, ‘kei’-, ‘duim’-, ‘roos’-, ‘pop’-, ‘maan’- en ‘paard’-variabele. Enkel in het geval van de ‘wijn’-variabele zien wij een zelfde plafondeffect bij de TiMBL-scores. Tenslotte hebben we ook correlatie analyses uitgevoerd om de samenhang te testen tussen type frequentie en de scores van de proefpersonen. Enkel in het geval van de negenjarigen

⁷ Significantie: * betekent $p < 0,05$, ** betekent $p < 0,01$.

was er een significante correlatie ($r = 0,560^*$). Globaal genomen, gedragen de scores van de proefpersonen zich dus anders dan de TiMBL-scores. Dit zou echter te wijten kunnen zijn aan het feit dat het maximum aantal trainingdata voor de classificatietask in TiMBL nog steeds relatief klein is. Met andere woorden, een lexicon van 2435 woorden (9867 segmenten) is niet representatief voor het lexicon van vijftienjarige kinderen. Om een betere vergelijking te kunnen maken tussen de resultaten van een geheugengebaseerde classificatietask in TiMBL en de resultaten van de Maldegemse proefpersonen zou in toekomstig onderzoek gebruik moeten worden gemaakt van een grotere database.

4.3. Neighborhood effects

In deze paragraaf bespreken we de effecten van *aantal enemy neighbors* op de afhankelijke variabelen *dialectrealisatie* en *overgeneralisatie*.

4.3.1. Effect van *aantal enemy neighbors* op *dialectrealisatie*

Om het effect van het aantal enemy neighbors op het al of niet realiseren van de dialectvariant in een specifiek woord te onderzoeken, voerden wij een binaire logistische regressie-analyse uit met *dialectrealisatie* als binaire afhankelijke variabele en met *aantal enemy neighbors* als onafhankelijke variabele (of predictor). Het resultaat hiervan wordt in Tabel 4 gepresenteerd.

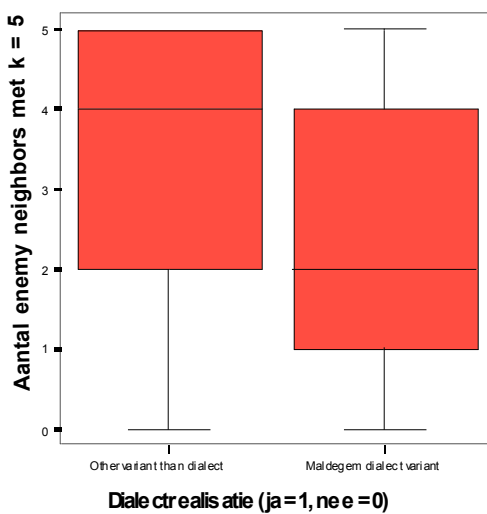
Predictor	B	S.E.	Exp(B)	Significantie
N enemy neighbors	-0,232	0,009	0,793	0,000
Constante	0,514	0,032	1,672	0,000
-2Loglikelihood Model chi-square	18448,175 628,018 (df = 1, p = 0,000)			

Tabel 4. Effect van *aantal enemy neighbors* op *dialectrealisatie* bij verwervers van het Maldegemse dialect

Tabel 4 laat zien dat er een hoogst significant negatief effect is van *aantal enemy neighbors* op *dialectrealisatie* ($B = -0,232$). Dit betekent dat naarmate een woord meer enemy neighbors heeft, de kans afneemt dat de correcte dialectvariant van dat woord gerealiseerd wordt. De logit⁸ neemt af met 0,232 per eenheid toename van *aantal enemy neighbors*, dat in deze studie een reikwijdte van 1 tot 5 heeft. Dat *aantal enemy neighbors* een negatief effect heeft op *dialectrealisatie* ligt in de lijn der verwachting, aangezien enemy neighbors woorden zijn die sterk gelijken op het testwoord, maar die een andere dialectuitspraak hebben. Doordat er dus geen eenduidig patroon is tussen een woord en zijn neighbors, in die zin dat niet alle neighbors in dezelfde richting wijzen, wordt het voor een kind dat het dialect leert moeilijker om het correcte patroon van een specifiek woord te verwerven.

⁸ De afhankelijke variabele in een logistische regressie-analyse wordt de *logit* genoemd: logit = natuurlijke logaritme van de *odds*; odds = de ratio van de kans op succes ($Y = 1$) ten opzichte van de kans op mislukking ($Y = 0$).

De boxplot in Figuur 2 visualiseert de distributie van *aantal enemy neighbors* voor incorrecte dialectrealisaties (cf. linker box) versus correcte dialectrealisaties (cf. rechter box). Uit deze figuur kunnen we afleiden dat de mediaan voor *aantal enemy neighbors* hoger ligt bij de incorrecte dialectrealisaties (namelijk: $x = 4$) dan bij de correcte dialectrealisaties ($x = 2$). Dit betekent dat voor de incorrecte dialectrealisaties, de helft van de gerealiseerde vormen 0 tot 4 *enemy neighbors* heeft, en de andere helft er 4 tot 5 heeft. Voor de correcte dialectrealisaties daarentegen, heeft de helft van de gerealiseerde vormen 0 tot 2 *enemy neighbors* en de andere helft 2 tot 5. Hieruit kunnen we besluiten dat er in het geval van de correcte dialectrealisaties over het algemeen minder *enemy neighbors* zijn (cf. Tabel 4: negatief effect van *aantal enemy neighbors*).



Figuur 2: Distributie van aantal enemy neighbors voor incorrecte vs. correcte dialectrealisaties

4.3.2. Effect van aantal enemy neighbors op overgeneralisatie

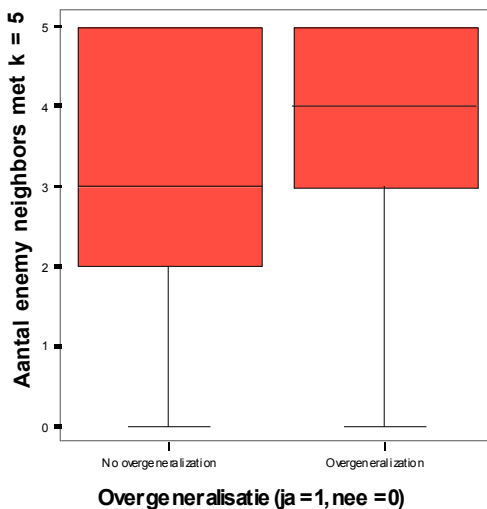
Om het effect van het aantal enemy neighbors op het al of niet produceren van een overgeneralisatie na te gaan, werd een binaire logistische regressie-analyse uitgevoerd met *overgeneralisatie* als binaire afhankelijke variabele en met *aantal enemy neighbors* als onafhankelijke variabele (of predictor). Tabel 5 toont het resultaat van deze analyse.

Predictor	B	S.E.	Exp(B)	Significantie
N enemy neighbors	0,315	0,026	1,370	0,000
Constante	-3,730	0,107	0,024	0,000
-2Loglikelihood Model chi-square	5804,445 162,486 (df = 1, p = 0,000)			

Tabel 5. Effect van *aantal enemy neighbors* op *overgeneralisatie* bij verwervers van het Maldegemse dialect

Tabel 5 laat zien dat er een hoogst significant positief effect is van *aantal enemy neighbors* op *overgeneralisatie*. Dit betekent dat naarmate een woord meer enemy neighbors heeft, de kans toeneemt dat er in dat woord een overgeneralisatie geproduceerd wordt. De logit neemt toe met 0,315 per eenheid toename van *aantal enemy neighbors*. Ook dit resultaat ligt in de lijn der verwachting: naarmate er meer woorden zijn die sterk gelijken op het testwoord, maar die een andere dialectuitspraak hebben, wordt de kans groter dat de verkeerde dialectuitspraak naar het testwoord wordt geëxtrapoleerd. Overgeneralisatie wordt in woordgebaseerde modellen immers verklaard als het overnemen van het class label van een enemy neighbor.

De boxplot in Figuur 3 toont de distributie van *aantal enemy neighbors* voor ‘geen overgeneralisatie’ (cf. linker box) versus ‘overgeneralisatie’ (cf. rechter box). De mediaan voor *aantal enemy neighbors* ligt lager bij die gevallen waar geen overgeneralisatie was (namelijk: $x = 3$) dan bij die gevallen waar er wel een overgeneralisatie was ($x = 4$). Dit betekent dat voor de gevallen zonder overgeneralisatie, de helft van de gerealiseerde vormen 0 tot 3 enemy neighbors heeft, en de andere helft er 3 tot 5 heeft. Voor de gevallen met overgeneralisatie heeft de helft van de gerealiseerde vormen 0 tot 4 enemy neighbors en de andere helft 4 tot 5. In het geval van overgeneralisatie zijn er dus over het algemeen meer enemy neighbors (cf. Tabel 5: positief effect van *aantal enemy neighbors*).



Figuur 3: Distributie van aantal enemy neighbors voor ‘geen overgeneralisatie’ vs. ‘overgeneralisatie’

5. Discussie

De bedoeling van deze studie was na te gaan of er in het geval van tweededialectverwerving (vooral) sprake is van regelgebaseerde dan wel woordgebaseerde leermechanismen. Tot nu toe werd er in de literatuur over tweededialectverwerving nauwelijks aandacht besteed aan de vraag welke leermechanismen er ten grondslag liggen aan de verwerving van een dialect als tweede taal. Een uitzondering is Chambers (1992), die beweert dat de fonologische kenmerken van een dialect aanvankelijk woord voor woord geleerd worden, totdat er voldoende woorden verworven zijn om een regel te kunnen vormen. Dit idee is gebaseerd op zijn observatie dat er zich in de scoregrafieken van een aantal fonologische kenmerken een S-vormig patroon aftekent. Deze S-curve duidt op een plotse versnelling in de acquisitie van een kenmerk door regelvorming. Chambers baseerde zich echter op de data van slechts zes proefpersonen en hij observeerde de S-curve niet bij alle door hem onderzochte kenmerken. Chambers observeerde immers enkel een S-curve in het geval van T-voicing, Vowel Backing en Low Vowel Merger. Voor twee andere onderzochte variabelen, namelijk ‘R-lessness’ (i.e., de deletie van niet-prevocalische /r/ in woorden als *summer*, *water*) en ‘Intrusive /r/’ (i.e., epenthesis van [ɹ] tussen vocalen aan een woord- of morfeemgrens, zoals in *sofa*[ɹ] and *couch*, *raw*[ɹ] *eggs*, of *draw*[ɹ]*ing*), vond Chambers wél percentages tussen 20 en 80%: voor aanwezigheid van R-lessness scoorde een van de zes proefpersonen 30% en voor aanwezigheid van Intrusive /r/ scoorde een van de proefpersonen 40%. Hoewel het hier telkens slechts om één proefpersoon ging, lijken deze gegevens ons niet onbelangrijk, gezien het toch al geringe aantal proefpersonen die in Chambers’ studie betrokken waren.

In tegenstelling tot Chambers, hebben wij niet enkel een veel groter aantal proefpersonen (nl. 128) in onze studie betrokken, maar ook het aantal onderzochte fonologische kenmerken (nl. 19) was groter. We hebben nagegaan of er zich in de scoregrafieken van 19 fonologische kenmerken van het Maldegemse dialect – dat door onze proefpersonen als tweede taal geleerd werd – S-vormige patronen aftekenen. Wij vinden echter dat er voor elk kenmerk ook scores

liggen tussen 20 en 80%, wat het idee van een S-curve tegenspreekt. Onze bevindingen wijzen erop dat fonologische dialectkenmerken geleidelijk en woord voor woord geleerd worden. Vandaar dat een groot aantal dialectverwervers voor de verwerving van een specifiek kenmerk een score behaalt die tussen 20 en 80% ligt. Als zij immers na het leren van een aantal instanties van een kenmerk een regel zouden vormen, dan zouden ze die regel meteen op alle resterende instanties moeten kunnen toepassen, wat tot uiting zou komen in een plotselinge stijging van hun leercurve (nl. een S-curve). Op basis van onze data hebben wij dus geen aanwijzingen dat de verwervers van het Maldegemse dialect regels vormen. Integendeel, er zijn eerder aanwijzingen voor het verwerven van het dialect op basis van lexicale of woordgebaseerde leermechanismen.

Om de hypothese dat standaardtalige kinderen het dialect verwerven op een woordgebaseerde manier verder te onderzoeken, hebben wij een geheugengebaseerde classificatietaak uitgevoerd in TiMBL, waarin het model op basis van standaardnederlandse woorden de correcte dialectvorm moest voorspellen. Deze classificatietaak leverde ons niet enkel informatie op over de accurateheid waarmee een woordgebaseerd analogisch model de juiste dialectvormen van een aantal testwoorden kan voorspellen, maar ook over wat de nearest neighbors van de testwoorden zijn en hoeveel daarvan enemy neighbors zijn. De accurateheidsscores bij een model dat gebaseerd is op een trainingset van respectievelijk 1000, 5000 en 9867 segmenten, laten zien dat niet alleen de omvang van de trainingdata van belang is voor een correcte classificatie, maar ook de aanwezigheid van een conditionerende fonologische context (i.e., tegenstelling tussen fonologisch geconditioneerde en lexicaal bepaalde kenmerken) en de typefrequentie van de dialectkenmerken. Woorden waarop lexicaal bepaalde kenmerken van toepassing zijn, worden niet gekenmerkt door een eenduidige fonologische structuur, waardoor ze niet geclusterd zitten in een homogene fonologische neighborhood. De woorden waardoor ze worden omringd, zijn over het algemeen woorden waarop andere kenmerken van toepassing zijn, met andere woorden, ze worden omringd door enemy neighbors. Ook type frequentie is een factor die een rol speelt in de vorming van (al of niet homogene) neighborhoods: kenmerken die op een groot aantal woorden van toepassing zijn, zullen over het algemeen gerelateerd zijn aan een grote homogene neighborhood. Aangezien de classificatietaak in TiMBL gebaseerd is op het toekennen van de meest frequente classificatie uit een set van vijf nearest neighbors aan het test item, zal het model beter presteren naarmate de set van nearest neighbors homogener is en vooral uit friendly neighbors van het test item bestaat.

Niet alleen blijkt neighborhood een belangrijke rol te spelen in de voorspelling van de correcte dialectuitspraak in de classificatietaak in TiMBL, maar neighborhood blijkt ook een belangrijk concept te zijn in de verwerving van fonologische dialectkenmerken door standaardtalige kinderen. Wij hebben namelijk de effecten van het aantal enemy neighbors op het al of niet realiseren van de correcte dialectvariant en op het al of niet produceren van een overgeneralisatie onderzocht bij de 128 proefpersonen, die het Maldegemse dialect als tweede taal leren. Een groter aantal enemy neighbors blijkt een negatief effect te hebben op de correcte dialectrealisatie van woorden en een positief effect op het produceren van overgeneralisaties. Beide resultaten tonen aan dat de verwerving van fonologische dialectkenmerken beïnvloed wordt door neighborhood effecten: naarmate een woord omringd wordt door een groter aantal woorden met een andere dialectuitspraak (i.e., enemy neighbors), zal de verwerving van de dialectuitspraak van het betreffende woord moeilijker zijn en zullen er vaker overgeneralisaties in dat specifieke woord voorkomen. Deze resultaten passen volledig in een analogisch woordgebaseerd taalverwervingsmodel. In een dergelijk model worden nieuwe woorden immers geclassificeerd op basis van hun gelijkenis met reeds opgeslagen woorden. Naarmate er zich tussen de nearest neighbors meer woorden bevinden die een andere classificatie (in ons geval, een andere dialectuitspraak) krijgen, met andere

woorden, naarmate een neighborhood heterogener is, wordt de classificatie van een nieuw item bemoeilijkt. Verder kan de aanwezigheid van enemy neighbors overgeneralisaties verklaren: op basis van gelijkenis met een enemy neighbor wordt er ten onrechte een bepaalde dialectuitspraak aan een specifiek woord toegekend.

We kunnen dus besluiten dat de verwerving van een dialect als tweede taal, door kinderen die in de standaardtaal werden opgevoed, grotendeels gestuurd wordt door woordgebaseerde leermechanismen. Deze bevinding is in strijd met de theorie van bidialectale fonologie (o.a. Auer, 1993; Taeldeman, 1993), waarin beweerd wordt dat bidialectale taalgebruikers correspondentieregels vormen tussen segmenten van hun T1 (bij Auer: Swabisch dialect; bij Taeldeman: Oost-Vlaams dialect) en equivalenten van hun T2 (bij Auer: Hoogduits; bij Taeldeman: Standaardnederlands). Het feit dat het lexicon tussen een dialect en de standaardtaal grotendeels overlapt, zou inderdaad kunnen doen vermoeden dat bidialectale sprekers correspondenties maken tussen overeenkomstige woorden van beide variëteiten. Onze bevindingen laten echter zien dat, als er dergelijke correspondenties gevormd worden, deze correspondenties geen regelkarakter hebben: we hebben geen aanwijzingen dat er op een gegeven moment een (correspondentie)regel gevormd wordt, waardoor de mate van verwerving van kenmerken plotseling versnelt. Er zijn daarentegen sterke aanwijzingen dat de verwerving van dialectkenmerken woord voor woord gebeurt en op basis van gelijkenis met reeds opgeslagen woorden.

Bibliografie

- Auer, Peter
 1990 *Phonologie der Alltagssprache. Zur Beschreibung von Variation am Beispiel des Konstanzer Stadtdialekts* (Studia Linguistica Germanica 28). Berlin & New York: Mouton de Gruyter.
 1993 Zweidimensionale Modelle für die Analyse von Standard/Dialekt-Variation und ihre Vorläufer in der deutschen Dialektologie. In Wolfgang Viereck (red.), *Verhandlungen des internationalen Dialektologenkongresses, Bamberg 1990*, 3-22. Stuttgart: Franz Steiner Verlag.
- Baayen, Harald, R. Piepenbrock en Leon Gulikers
 1995 *The CELEX Lexical Database* (CD-ROM). Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.
- Berthele, Raphael
 2002 Learning a second dialect: A model of idiolectal dissonance. *Multilingua* 21, 327-344.
- Bybee, Joan
 1995 Regular morphology and the lexicon. *Language and Cognitive Processes* 10, 425-455.
 2001 *Phonology and language use*. Cambridge: Cambridge University Press.
- Chambers, Jack K.
 1992 Dialect acquisition. *Language* 68, 673-705.
- Chambers, Jack K. en Peter Trudgill
 1980 *Dialectology*. Cambridge: Cambridge University Press.
- Chen, Matthew Y. en William S.-Y. Wang
 1975 Sound change: Actuation and implementation. *Language* 51, 255-81.
- Chomsky, Noam en Morris Halle
 1968 *The Sound Pattern of English*. New York: Harper & Row.

- Daelemans, Walter en Antal van den Bosch
 2005 *Memory-Based Language Processing*. Cambridge: Cambridge University Press.
- De Vink, Leendert
 2004 *Dialect en dialectverandering in Katwijk aan Zee*. Delft: Eburon.
- Devitt, Amy J.
 1989 *Standardizing written English: Diffusion in the case of Scotland, 1520-1659*. Cambridge: Cambridge University Press.
- De Vogelaer, Gunther
 2010 (Not) acquiring grammatical gender in two varieties of Dutch. In Dirk Geeraerts, Gitte Kristiansen en Yves Peirsman (red.), *Advances in cognitive sociolinguistics*, 167-190. Berlin: Mouton de Gruyter.
- De Vogelaer, Gunther en Clazien Rooze-Stouthamer
 2006 Taalcontact of onvolledige verwerving: casusverlies bij de Zeeuwse pronomina. *Tijdschrift voor Nederlandse Taal- en Letterkunde* 122, 348-368.
- Dressler, Wolfgang en Ruth Wodak
 1982 Sociophonological methods in the study of sociolinguistic variation in Viennese German. *Language in society* 11, 339-370.
- Hoppenbrouwers, Cor
 1990 *Het regiolect. Van dialect tot Algemeen Nederlands*. Muiderberg: Dick Coutinho.
- Keuleers, Emmanuel, Kevin Diependaele en Marc Brysbaert
 2010 Practice effects in large-scale visual word recognition studies: A lexical decision study on 14,000 Dutch mono- and disyllabic words and nonwords. *Frontiers in Psychology* doi: 10.3389/fpsyg.2010.00174.
- Kerswill, Paul
 1994 *Dialects converging: Rural speech in urban Norway*. Oxford: Oxford University Press.
- Moosmüller, Sylvia
 1988 Sociophonology. In Peter Auer en Aldo di Luzio (red.), *Variation and convergence*, 76-93. Berlin: Walter de Gruyter.
- Payne, Arvilla C.
 1976 *The acquisition of the phonological system of a second dialect*. PhD dissertation, University of Pennsylvania.
 1980 Factors controlling the acquisition of the Philadelphia dialect by out-of-state children. In William Labov (red.), *Locating language in time and space*, 143-178. New York & London: Academic Press.
- Pinker, Steven en Alan Prince
 1988 On language and connectionism: analysis of a parallel distributed processing model of language acquisition. *Cognition* 29, 73-193.
- Roberts, Julie
 1994 Acquisition of variable rules: (-t,d) deletion and (ing) production in preschool children. PhD dissertation, University of Pennsylvania.
 1997a Hitting a moving target: Acquisition of sound change in progress by Philadelphia children. *Language variation and change* 9, 249-266.
 1997b Acquisition of variable rules: A study of (-t,d) deletion in preschool children. *Journal of child language* 24, 351-372.
- Roberts, Julie en William Labov
 1995 Learning to talk Philadelphian: Acquisition of short a by preschool children.

- Language variation and change* 7, 101-112.
- Rys, Kathy
 2003 *Secundaire verwerving van fonologische elementen van een dialect. Taal en tongval* 55, 68-108.
 2007 *Dialect as a second language: Linguistic and non-linguistic factors in secondary dialect acquisition by children and adolescents.* Proefschrift, Universiteit Gent.
- Rys, Kathy en Dries Bonte
 2006 The role of linguistic factors in the process of second dialect acquisition. In Frans Hinskens (red.), *Language variation – European perspectives*, 201-215. Amsterdam/Philadelphia: John Benjamins.
- Rys, Kathy en Katrien De Valck
 2010 Overgeneralisatie in tweede dialectverwerving: het belang van typefrequentie voor de productiviteit van kenmerken en voor het proces van dialectverandering. In Johan De Caluwe en Jacques Van Keymeulen (red.), *Voor Magda. Artikelen voor Magda Devos bij haar afscheid van de Universiteit Gent*, 525-540. Universiteit Gent, Vakgroep Nederlandse Taalkunde/Academia Press, Gent.
- Siegel, Jeff
 2010 *Second dialect acquisition.* Cambridge: Cambridge University Press.
- Skousen, Royal, Deryle Lonsdale en Dilworth B. Parkinson (red.)
 2002 *Analogical Modeling. An exemplar-based approach to language.* Amsterdam: John Benjamins.
- Taeldeman, Johan
 1976 *De klankstructuur van het Kleitse dialect.* Proefschrift Universiteit Gent.
 1989 De taaltoestand in Vlaanderen. *Neerlandica Wratislaviensia* IV, 77-90.
 1991 Dialect in Vlaanderen. In Herman Crompvoets & Ad Dams (red.), *Kroesels op de bozzem: het dialectenboek*, 34-52. Waalre.
 1993 Dialectresistentie en dialectverlies op fonologisch gebied. In Frans Hinskens, Cor Hoppenbrouwers en Johan Taeldeman (red.), *Dialectverlies en regiolectvorming (Taal en tongval* 46), 102-119.
- Versieck, Sabina
 1989 *Het Maldegemse klanksysteem in het heden en honderd jaar geleden.* Masterscriptie, Universiteit Gent.
- Vousten, Rob en Theo Bongaerts
 1990 Acquiring a dialect as L2: The case of the dialect of Venray in the Dutch province of Limburg. Paper presented at the International Congress of Dialectologists, Bamberg, 1990.
- Vousten, Rob
 1995 *Dialect als tweede taal. Linguïstische en extra-linguïstische aspecten van de verwerving van een Noordlimburgs dialect door standaardtalige jongeren.* Proefschrift, Universiteit Nijmegen.
- Wang, William S.-Y.
 1969 Competing changes as a cause of residue. *Language* 45, 9-25.
- Wang, William S.-Y. en Chin-Chuan Cheng
 1970 Implementation of phonological change: the Shuang-Feng case. *POLA*, 2.10, 1-9.

