t&t

T&T 2015, Vol. 67, No. 2

www.taalentongval.eu

Amsterdam University Press

Inhoudsopgave

Themanummer
Constructions in Variation and Change –
Variation and Change in Constructions
Gastredactie: Timothy Colleman en Freek Van de Velde

Variatie en verandering in constructies

Op het snijvlak van de constructiegrammatica en de variatielinguïstiek

Timothy Colleman & Freek Van de Velde

MEM 67 (2): 135–148
DOI: 10.5117/TET2015.2.COLL

Abstract

This introductory paper outlines the different trends and movements that have in recent years led to a marked increase in the number of linguistic studies that combine a construction-based theoretical outlook on grammar with a dedicated interest in issues of synchronic and/or diachronic language variation. In addition, it gives an overview of the papers included in this thematic issue and links them with broader tendencies in the fields of cognitive linguistics and construction grammar.

Keywords: construction grammar, language variation, sociolinguistics, historical linguistics, grammaticalization, Cognitive Sociolinguistics, language contact research

1 Taalvariatie en –verandering als blinde vlek in het vroege constructiegrammaticale onderzoek

Toen op het eind van de jaren '80 en in de eerste helft van de jaren '90 van de vorige eeuw een aantal publicaties verschenen waarin voor het eerst de contouren werden geschetst van een 'constructionele' benadering van de grammatica – zie o.m. Lakoff (1987), Fillmore (1988), Fillmore, Kay en O'Connor (1988), Goldberg (1992, 1995) – werd daarin meteen sterk beklemtoond dat die ontluikende theorie in de grammaticale beschrijving die ze voorstond streefde naar *volledigheid*: de *hele* grammatica kan worden beschreven als een gestructureerd netwerk van duizenden en duizenden *constructies*, in de specifieke betekenis die daaraan in het constructiegrammaticale framework wordt gegeven, nl. een geconventionaliseerd vorm/betekenis-paar. In Goldbergs (2003) wat latere overzicht van de fun-

damentele uitgangspunten die worden gedeeld door de verschillende constructiegrammaticale benaderingen – die m.b.t. allerlei andere aspecten een behoorlijk heterogene groep vormen – wordt dat samengevat onder de slogan 'It's constructions all the way down' (p. 223), of, iets omstandiger geformuleerd: 'The totality of our knowledge of language is captured by a network of constructions' (p. 219). In datzelfde artikel wordt opgemerkt dat de representatie van een constructie in het mentale "constructicon" ook informatie kan bevatten over *taalvariatie:* 'Facts about the use of entire constructions, including register (e.g. formal or informal), dialect variation and so on, are stated as part of the construction as well' (Goldberg, 2003, p. 221). Echter, in de praktijk is er in de eerste pakweg 15 jaar van het constructiegrammaticale onderzoek nauwelijks aandacht besteed aan kwesties van taalvariatie en taalverandering: de empirische focus lag grotendeels op de beschrijving van individuele grammaticale fenomenen uit het hedendaagse (Amerikaanse) Engels, en vanaf de tweede helft van de jaren 90 ook uit een groeiend aantal andere talen, waarbij het bovengenoemde streven naar volledigheid zich vooral uitte in de grote aandacht voor (i) allerlei fenomenen die in andere grammaticale benaderingen werden gezien als te afwijkend van de 'gewone' grammaticale regels om met goed fatsoen tot het studie-object van de grammatica te kunnen worden gerekend (zie bv. Fillmore, Kay en O'Connor, 1988 over de *let alone*-constructie), en (ii) de kenmerkende *semantische* en *pragmatische* eigenschappen van allerlei grammaticale patronen (zie bv. Goldberg, 1992, 1995 over de semantiek van de ditransitieve constructie en andere Engelse argumentstructuurconstructies). Met betrekking tot die beide aspecten onderscheidde de opkomende nieuwe benadering zich van – en zette ze zich expliciet af tégen – de formeel georiënteerde benaderingen die de grammaticale theorievorming van de jaren zeventig en tachtig in belangrijke mate hadden beheerst. Tegelijk werd, ondanks dat streven naar volledigheid, een belangrijk deel van de taalrealiteit vrijwel volledig buiten beschouwing gelaten: een sporadische zijdelingse opmerking niet te na gesproken, werd volledig abstractie gemaakt van eventuele regionale, sociale, stilistische, etc. variatie in de formele of inhoudelijke eigenschappen van de onderzochte constructies, net zoals er nauwelijks aandacht was voor de processen van (voortschrijdende) taalverandering waarvan die constructies het voorlopige resultaat vormen.[1] Op dat vlak onderscheidde de vroege constructiegrammatica zich trouwens niet van het bredere paradigma van de cognitieve linguïstiek, waaruit ze grotendeels was voortgekomen en waarmee verscheidene stromingen in het hedendaagse constructiegrammaticale onderzoek nog altijd een belangrijke theoretische en methodolo-

gische affiniteit vertonen. Volgens Geeraerts, Kristiansen & Peirsman (2010, pp. 5-6) vertoonde de cognitieve linguïstiek als onderzoeksveld tot voor kort een merkwaardige paradox: aan de ene kant had sinds Bybee (1998), Langacker (2000), Tomasello (2003) en andere sleutelpublicaties uit die periode de fundamentele theoretische aanname veld gewonnen dat de grammatica in essentie 'usage-based' is, d.w.z. dat het taalsysteem een dialectische relatie onderhoudt met het taalgebruik. Aan de andere kant was er in de praktijk van het onderzoek relatief weinig aandacht voor de *variatie* die inherent is aan het taalgebruik en richtten veel studies zich op fenomenen uit 'het' Engels, 'het' Nederlands, enzovoort, met andere woorden op 'language situated taxonomically at an almost Chomskyan level of abstraction' (p. 6).

2 De kentering in de jaren 2000: Taalvariatie en –verandering als belangrijk aandachtsdomein

Naarmate de constructiegrammatica een vastere positie verwierf als een omvattende nieuwe benadering van de grammatica met een eigen gezicht, verbreedde ook haar aandachtsdomein – een ontwikkeling die we net zo goed zouden kunnen beschrijven als het gaandeweg doordringen van constructiegrammaticale noties en inzichten in andere subdisciplines van de taalwetenschap. Een van de opvallendste klemtoonverschuivingen is de duidelijke toename in het aantal studies waarin vanuit een constructioneel perspectief wordt ingegaan op allerlei aspecten van *intralinguale* variatie en verandering, die zich vooral vanaf de late jaren 2000 begint te manifesteren: zie Barđdal (2007), Siewierska en Hollmann (2007), Hoffmann en Mukherjee (2007), Grondelaers, Speelman en Geeraerts (2008), Bergs en Diewald (2009), Fried (2009), Szmrecsanyi (2010), Gries en Hilpert (2010), Rutten (2010), Colleman en De Clerck (2011), Traugott en Trousdale (2013), Hilpert (2013), De Smet, Ghesquière en Van de Velde (2013) en Barđdal *et al.* (2015) voor een greep van representatieve artikelen, monografieën en verzamelbundels.

Boogaart, Colleman en Rutten (2014, pp. 6-7) brengen die evolutie in verband met drie bredere tendensen in het hedendaagse (cognitief-functionele) taalwetenschappelijke onderzoek. Ten eerste heeft de op dat moment net voorzichtig aan de oppervlakte komende *diachrone* constructiegrammatica een sterke impuls gekregen doordat een aantal invloedrijke stemmen uit de grammaticalisatietheorie in de late jaren 2000 uitgebreid zijn gaan reflecteren over de rol van constructies (als vorm/betekenis-

paren) en van constructionele netwerken in grammaticalisatieprocessen (zie bv. Traugott, 2008a,b, 2015; Diewald, 2008; Trousdale, 2010), wat uiteindelijk zelfs tot een zekere verstrengeling heeft geleid van het diachrone constructiegrammaticale onderzoek en een belangrijk deel van het onderzoek naar grammaticalisatie (zie ook Noël, 2007 voor een vroege theoretische reflectie op de relatie tussen de grammaticalisatietheorie en diachrone toepassingen van de constructiegrammatica). Vrijwel gelijktijdig was er, als tweede tendens, de opkomst van de zgn. Cognitieve Sociolinguïstiek, een benadering die een sterkere integratie bepleit van vraagstellingen en concepten uit enerzijds de cognitieve linguïstiek en anderzijds de variationele sociolinguïstiek, zie Kristiansen en Dirven (2008), Geeraerts, Kristiansen en Peirsman (2010) en Geeraerts en Kristiansen (2014) voor een uitgebreide toelichting over de theoretische en descriptieve meerwaarden die zo'n integratie voor beide benaderingen zou kunnen opleveren. Een van de primaire onderzoeksdomeinen van de Cognitieve Sociolinguïstiek is 'lectale' (d.w.z. regionale, sociale, stilistische, generationele, enz.) variatie in het gebruik van (lexicale of grammaticale) constructies, dus precies het soort betekenisdragende talige items dat in de variationele sociolinguïstiek vaak buiten beeld blijft, vanuit de overweging dat bij sociolinguïstische variabelen beter geen betekeniskwesties betrokken worden en dat het enige interessante register gesproken taal in spontane conversaties of sociolinguïstische interviews is (zie Rickford, 2014 voor een bespreking). In de cognitieve linguïstiek en constructiegrammatica staat betekenis juist centraal en is een uitgebreid analytisch apparaat ontwikkeld voor de studie van de semantische structuur van lexicale en grammaticale items. Voorbeelden van case studies die intralinguale variatie integreren in zo'n constructiegrammaticale analyse zijn Grondelaers, Speelman en Geeraerts (2008), Szmrecsanyi (2010) en Colleman (2010). De derde constituerende tendens, die samenhangt met de tweede, is het toenemende belang dat in een belangrijk deel van het hedendaagse grammaticale onderzoek, inclusief de cognitieve linguïstiek en de constructiegrammatica, wordt gehecht aan *empirische ondersteuning*, op basis van data uit corpora, enquêtes, experimenten, enz. Zoals wordt beklemtoond in Geeraerts (2006, p. 30), is variatie inherent aan het taalgebruik, zodat onderzoekers die gebruik maken van reëel geobserveerde taaldata zich bijna per definitie verplicht zien om lectale variabelen op te nemen in hun onderzoeksdesign; zie ook Gries (2013) voor een illustratieve bespreking van de verschillende bronnen van potentiële variatie in de data waarmee in corpusgebaseerd onderzoek rekening moet worden gehouden.

Een tendens die in Boogaart, Colleman en Rutten (2014) nog niet wordt

genoemd, ten slotte, is dat er naast of bovenop de toenemende interesse in lectale variatie recent ook meer aandacht is gekomen voor de invloed van taalcontact op variatie en verandering. Om de structuur en variatie van het Nederlands – of enige andere taal – goed te begrijpen, moet de interactie met andere talen in beeld gebracht worden. De grote rol van taalcontact in taalvariatie en -verandering is van meet af aan erkend in de sociolinguïstiek (Weinreich, Labov en Herzog, 1968), maar is het afgelopen decennium nog eens uitvoerig voor het voetlicht gebracht door onder meer Heine en Kuteva (2005, 2006), Drinka (2010), Trudgill (2011), Kortmann en Szmrecsanyi (2012) en Lucas (2015). Binnen de cognitieve linguïstiek heeft dat geleid tot wat men 'Cognitieve contactlinguïstiek' zou kunnen noemen (Wolf en Polzenhagen, 2009; Winter-Froemel, 2011; Zenner, 2013; Zenner en Kristiansen, 2014; Backus, 2014; Soares da Silva, 2015). Voorbeelden van case studies naar verandering door taalcontact vanuit een expliciet constructiegebaseerd perspectief zijn Doğruöz en Backus (2009), Pietsch (2010), Höder (2012, 2014), Colleman en Noël (2014) en Colleman (in voorb.).

De optelsom van die verschillende tendensen is dat in een aanzienlijk deel van de huidige constructiegrammaticale wetenschappelijke productie een belangrijke plaats wordt ingeruimd voor kwesties van taalvariatie en -verandering. Symptomatisch daarvoor is, om maar één voorbeeld te noemen, dat een van de vijf grote delen in het recente *Handbook of Construction Grammar* van Hoffmann en Trousdale (2013) volledig is gewijd aan 'Language Variation and Change', met vijf bijdragen over synchrone en/of diachrone intralinguale variatie. Hoewel het constructiegrammaticale onderzoek daarnaast ook nog in allerlei andere beloftevolle richtingen is uitgewaaierd – zoals een snelle blik op de andere vier delen in Hoffmann en Trousdale (2013) leert – zijn het uiteraard vooral de ontwikkelingen op het snijvlak van de constructiegrammatica, de historische taalkunde en de variatielinguïstiek die voor het lezerspubliek van *Taal en Tongval* interessant zijn. Het onderhavige nummer biedt een staalkaart van die ontwikkelingen.

3 De bijdragen in het themanummer

De zeven resterende artikelen in dit themanummer vormen de schriftelijke neerslag van lezingen gepresenteerd op het Taal en Tongvalcolloquium over 'Constructions in variation and change/Variation and change in constructions', dat plaatsvond op 6 december 2013 in de gebouwen van de KANTL te Gent.

Het nummer opent met drie artikelen met een historisch-taalkundige focus. De bijdrage van **Evie Coussé** gaat over de opkomst van dubbele-modaal-constructies in het Nederlands, d.w.z. constructies waarin een modaal een andere modaal in zijn bereik heeft, van het type *Waarom zou een man niet mogen huilen?* met *zou … mogen* als dubbele modaal. Hoe natuurlijk dat soort zinnen sprekers van het hedendaags Nederlands ook in de oren klinken, ze hebben niet altijd bestaan. Coussé gaat na hoe deze constructie ontstaan is in het Middelnederlands en zich verder ontwikkeld heeft in het Nieuwnederlands. Ze argumenteert dat de toenemende mogelijkheden van modale combinaties zich goed laten vatten in een diachroon-constructionele beschrijving, als die zich tenminste niet beperkt tot de veranderingen in 'schematiciteit', 'productiviteit' en 'compositionaliteit' die in de constructiegrammaticale theorie van taalverandering van Trousdale en Traugott (2013) als cruciaal naar voren worden geschoven – maar die hier niet goed van toepassing blijken te zijn – maar daarnaast ook 'complexiteit' als dimensie in rekening brengt. Coussé ziet de opkomst en uitbreiding van gestapelde modalen door het in elkaar schuiven van verschillende constructies en door analogische uitbreiding naar nieuwe gevallen als een toename in constructionele complexiteit.

In de volgende bijdrage, van de hand van **Alexandra N. Lenz**, wordt uitgebreid ingegaan op de ontstaansgeschiedenis van resultatieve (en resultatief-modale) constructies met de Duitse werkwoorden *kriegen* en *bekommen* – de auteur presenteert een synthese van haar bestaand onderzoek naar de constructies in kwestie, dat nog niet in het Engels beschikbaar was, en ze last een contrastieve uitstap in naar vergelijkbare constructies met Nederlands *krijgen* en Afrikaans *kry*. Aan de hand van data uit een brede waaier van zowel historische corpora als moderne dialectcorpora, aangevuld met relevante gegevens uit dialectwoordenboeken, worden de grammaticalisatiepaden gereconstrueerd die tot het moderne resultatieve gebruik hebben geleid. De theoretische achtergrond van Lenz' bijdrage is de grammaticalisatietheorie. Zoals in paragraaf 2 al werd beklemtoond, is het hedendaagse onderzoek naar grammaticalisatie echter in belangrijke mate verstrengeld met de diachrone constructiegrammatica: de affiniteit van Lenz' benadering met een constructiegebaseerd theoretisch perspectief blijkt onder meer uit de grote aandacht die ze besteedt aan de specifieke lexicale en syntactische omgevingen waarin de combinatie van *kriegen* of *bekommen* met bepaalde types van aanvullingen (bv. *etwas in den Griff/die Finger kriegen/bekommen*) voor het eerst aanleiding kon geven tot een resultatieve interpretatie – of, veel later, waarin uit dat resultatieve gebruik een modale interpretatie kon ontstaan (de zgn. 'bridging' of 'criti-

cal contexts', zie daarover o.m. Heine, 2002; Diewald, 2002, 2008; Traugott 2008).

De bijdrage van **Tim Geleyn en Timothy Colleman** verschilt van de twee vorige artikelen doordat de focus niet zozeer ligt op het ontstaan van een 'nieuwe' constructie maar op semantische verandering in een bestaande constructie. In het licht van de belangrijke impuls die de integratie van constructionele noties in het grammaticalisatie-onderzoek heeft gegeven aan de ontwikkeling van een constructiegrammaticale benadering van taalverandering, is het weinig verrassend dat de grote meerderheid van de bestaande diachroon-constructiegrammaticale case studies zich buigt over de vraag hoe 'nieuwe' constructies precies ontstaan ('constructionalisering' in termen van Traugott en Trousdale 2013). Zoals in meer detail wordt beargumenteerd in Colleman en De Clerck (2011) is dat ook een fascinerende kwestie, maar is het van belang dat de diachrone constructiegrammatica zich daar niet toe *beperkt*, maar ook andere types verandering vanuit constructioneel perspectief probeert te duiden (de zgn. '(gewone) constructionele veranderingen' in termen van Traugott en Trousdale, 2013; zie bv. ook Hilpert, 2013, pp. 8-17 voor discussie van wat constructionele verandering zoal kan inhouden). Terwijl in de diachrone lexicologie al uitgebreid onderzoek is verricht naar de verschillende semasiologische en onomasiologische veranderingsprocessen waaraan *lexicale* betekenissen onderhevig zijn (zie bv. Geeraerts, 1997; Allen, 2009 en verscheidene bijdragen in Winters et al., 2010) is nog weinig bekend over betekenisverandering in schematischere, grammaticale constructies. Geleyn en Colleman presenteren de resultaten van een semasiologisch onderzoek naar de semantische reikwijdte van de prepositionele datiefconstructie met aan (bv. *Jan heeft een boek aan Piet gegeven*) in het 17de- en 20ste-eeuwse Nederlands.

Op de drie historisch-taalkundig georiënteerde artikelen volgt nog een tweeluik van artikelen over *taalcontact*. Zoals hierboven is opgemerkt, is in de constructiegrammatica en cognitieve linguïstiek pas recent aandacht gekomen voor (verandering door) taalcontact, een verschijnsel dat overigens in het algemeen grotendeels buiten beschouwing is gebleven in de grammaticale theorievorming (zie bv. Nicolaï, 2007 voor een poging tot verklaring voor die karige aandacht voor contactfenomenen). De bijdrage van **Eline Zenner en Dirk Geeraerts** bestudeert Engelse invoegingen die uit meer dan één woord bestaan in het spontane gesproken taalgebruik van deelnemers aan het Nederlands-Vlaamse realityprogramma *Expeditie Robinson*, van het type *free at last* of *fuck the world*. Zulke formaties zijn op het eerste gezicht moeilijk in te delen volgens traditionele typologieën van

contactverschijnselen waarin een scherp onderscheid wordt gemaakt tussen lexicale ontlening aan de ene en code-switching aan de andere kant (zie echter al Matras, 2009 voor een genuanceerder beeld). Een dergelijk dichotomisch onderscheid laat zich moeilijk rijmen met de notie van het *lexicon-syntaxiscontinuüm*, d.w.z. er wordt in de constructiegrammatica van uitgegaan dat de grammatica bestaat uit een gestructureerd netwerk van duizenden constructies met sterk wisselende graden van complexiteit en schematiciteit, veeleer dan dat er een tweedeling zou bestaan tussen 'lexicale' en 'grammaticale' constructies – zie bv. Goldberg (2006, p. 5) en Croft en Cruse (2004, p. 255) (zie ook het verwante concept van de *lexicality-schematicity hierarchy*, bv. in Barðdal en Gildea 2015, pp. 27-28). Zenner en Geeraerts komen op basis van drie diagnostische criteria tot de conclusie dat de meeste Engelse invoegingen in het corpus een sterke mate van lexicale gefixeerdheid vertonen – en dus veeleer aan de lexicale kant van het continuüm thuishoren. Daarnaast is er een kleiner aantal gevallen waarvoor aannemelijk (of op zijn minst *mogelijk*) is dat ze zijn gebaseerd op een wat schematischer patroon, met een of meer open slots (bv. een microconstructie [*fuck* NP], waarmee constructs kunnen worden gebouwd als *fuck the world*, *fuck the others*, enz.). Zenner en Geeraerts brengen het sterke overwicht van volledig gelexicaliseerde expressies in verband met de relatief *indirecte* aard van het taalcontact Nederlands-Engels.

De bijdrage van **Ad Backus** is wat programmatischer van aard: de auteur reflecteert over de meerwaarde die een gebruiksgebaseerde ('usage-based') benadering kan opleveren voor het sociolinguïstische onderzoek – of anders gezegd, de meerwaarde van de integratie van sociolinguïstische en historisch-taalkundige vraagstellingen in de constructiegrammatica en verwante gebruiksgebaseerde benaderingen (zie ook paragraaf 2 over de opkomst van de Cognitieve Sociolinguïstiek). Backus zoomt meer bepaald in op het onderzoek naar (voortschrijdende) taalverandering door taalcontact. Hij keert terug naar de vijf kwesties die in Weinreich, Labov en Herzog (1968) werden geïdentificeerd als de cruciale vragen die de toen ontluikende discipline van de sociolinguïstiek zou moeten beantwoorden met het oog op de ontwikkeling van een omvattende theorie over *diachrone* variatie en laat aan de hand van een bespreking van geselecteerde fenomenen uit het Nederturks (d.w.z. het Turks zoals dat gesproken wordt in Nederland door nakomelingen van migranten uit Turkije die daarnaast ook over een zeer gevorderde (moedertaal)kennis van het Nederlands beschikken) zien dat die vragen even pertinent zijn voor verandering als gevolg van taalcontact als voor veranderingen als gevolg van taalinterne fac-

toren. Backus beargumenteert dat de opkomst van gebruiksgebaseerde benaderingen kansen biedt voor de ontwikkeling van een theorie over taalverandering die *alle vijf* de vragen van Weinreich, Labov en Herzog (1968) ernstig neemt en geeft een aantal theoretisch-methodologische voorzetten voor onderzoek naar taalcontact dat aan die ontwikkeling kan bijdragen.

Het themanummer besluit met twee bijdragen over lectale variatie in het hedendaagse Nederlands. Het artikel van **Jocelyne Daems, Kris Heylen en Dirk Geeraerts** kijkt naar de onomasiologische variatie in het Belgisch Nederlands en Nederlands Nederlands in kledingtermen, en bouwt zo verder op het pionierswerk over lectale variatie in kledingstermen in Geeraerts, Grondelaers en Speelman (1999) en Grondelaers et al. (2001). Door het vervolgonderzoek zowel methodologisch als inhoudelijk nauw aan te laten sluiten bij het eerdere onderzoek kunnen de resultaten goed vergeleken worden, en kunnen de diachrone trends in het eerdere onderzoek, dat materiaal bevatte van 1950 tot 1990, doorgetrokken worden tot in 2012. De belangrijkste conclusies zijn (i) dat de lexicale toenadering tussen het Belgisch Nederlands en het Nederlands Nederlands tot stilstand gekomen is, en dat de beide variëteiten nu weer van elkaar weg lijken te drijven; (ii) dat de afstand tussen de regionale informele variëteit en de supraregionale formele variëteit in België nog steeds groter is dan in Nederland. De beide variëteiten in Nederland liggen zelfs dichter bij elkaar dan de bovengewestelijke formele variëteiten in Nederland en België onderling; (iii) dat het Nederlands Nederlands een grotere reserve ten opzichte van het Engels vertoont dan het Belgisch Nederlands, en dat er een afname is van Franse invloed op beide taalniveaus in beide variëteiten. Het lectometrische onderzoek laat goed zien hoe de lectale, diachrone, en empirische tendensen in de cognitieve linguïstiek, waarop hogerop ingegaan is, hun beslag krijgen in het neerlandistische onderzoek. Ligt de focus in het artikel van Daems en collega's op lexicale variatie, de gehanteerde methodologie zou uiteraard even goed kunnen worden ingezet voor de studie van lectale variatie in complexere en schematischere constructies.

Dirk Pijpops en Freek Van de Velde, ten slotte, gaan in hun bijdrage aan dit themanummer in op adjectiefmorfologie in het Marokkaanse Nederlandse etnolect. Door vergelijking van omvangrijke corpora van informeel chatmateriaal zoeken de auteurs uit of Marokkaans-Nederlandse vroege tweedetaalverwervers zich anders gedragen dan reguliere eerstetaalverwervers bij het realiseren van de partitieve genitiefs-s in patronen als *iets interessants*. Door via regressiemodellen de twee variëteiten te vergelijken, stellen ze vast dat er geen verschil optreedt. Dat is opvallend, want andere adjectiefmorfologie laat juist wél een duidelijk verschil zien,

zoals uit eerder onderzoek blijkt. Pijpops en Van de Velde gaan in op de aard van dit verschil en zoeken een verklaring in de mate van transparantie van het morfeem. Die speelt volgens hen een grote rol in de overlevingskansen van het morfeem in taalcontact. De hardnekkigheid van het partitieve genitiefsuffix is gelegen is zijn transparante functie als herkenbare markeerder van een specifieke constructie. Wat eruit ziet als een complexe aansturing van het voorkomen van het *s*-suffix, is het gevolg van interfererende constructies, wat ten gronde wordt onderzocht in Pijpops en Van de Velde (ter perse). Met hun analyse sluiten de auteurs aan bij de constructiegrammaticale analyse van de Nederlandse partitieve genitief die voorgesteld is in Booij (2010).

Noot

1. Een uitzondering die hier niet onvermeld mag blijven is de studie van Israel (1996) over de geschiedenis van de zgn. "*way*-construction" in het Engels (bv. *He elbowed his way to the door*), die een interessante zij het nog enigszins embryonale visie bevat over de manier waarop "nieuwe" constructies kunnen ontstaan, maar die terugblikkend als de spreekwoordelijke ene zwaluw moet worden beschouwd.

Bibliografie

Allan, Kathryn, *Metaphor and Metonymy: A Diachronic Approach*. Chichester: Wiley-Blackwell, 2009.

Backus, Ad, 'A usage-based approach to borrowability'. In: Eline Zenner & Gitte Kristiansen (red.), *New Perspectives on Lexical Borrowing: Onomasiological, Methodological and Phraseological Innovations*. Berlin & New York, Mouton de Gruyter, 2014, pp. 19-39.

Barðdal, Jóhanna, 'The Semantic and Lexical Range of the Ditransitive Construction in the History of (North) Germanic', *Functions of Language*, 14 (2007), pp. 9-30.

Barðdal, Jóhanna, Smirnova, Elena, Sommerer, Lotte & Gildea, Spike (red.), *Diachronic Construction Grammar*. Amsterdam: John Benjamins, 2015.

Barðdal, Jóhanna & Gildea, Spike, 'Diachronic Construction Grammar: Epistemological context, basic assumptions and historical implications'. In: Jóhanna Barðdal et al. (red.), *Diachronic Construction Grammar*. Amsterdam: John Benjamins, 2015, pp. 1-49.

Bergs, Alexander & Diewald, Gabriele (red.), *Constructions and Language Change.* Berlin & New York: Mouton de Gruyter, 2008.

Boogaart, Ronny, Colleman, Timothy & Rutten, Gijsbert, 'Constructions all the way everywhere: Four new directions in constructionist research'. In: Ronny Boogaart, Timothy Colleman & Gijsbert Rutten (red.), *Extending the Scope of Construction Grammar*. Berlin & New York: Mouton de Gruyter, 2014, pp. 1-14.

Booij, Geert, *Construction morphology*. Oxford: Oxford University Press, 2010.

Bybee, Joan, The emergent lexicon. *Chicago Linguistic Society*, 34 (1998), pp. 421-435.

Colleman, Timothy, 'Lectal variation in Constructional Semantics: Benefactive ditransitives in

Dutch'. In: Dirk Geeraerts, Gitte Kristiansen & Yves Peirsman (red.), *Advances in Cognitive Sociolinguistics.* Berlin & New York: Mouton de Gruyter, 2010, pp. 191-221.

Colleman, Timothy, 'Distributional assimilation in constructional semantics: On contact-related semantic shifts in Afrikaans three-argument constructions'. In: Steffen Höder & Hans C. Boas (red.), *Constructions in contact* (werktitel), in voorbereiding.

Colleman, Timothy & De Clerck, Bernard, 'Constructional semantics on the move: On semantic specialization in the English double object construction', *Cognitive Linguistics,* 22 (2011), pp. 183-209.

Colleman, Timothy & Noël, Dirk, 'Tracing the history of deontic NCI patterns in Dutch: A case of polysemy copying'. In: Irma Taavitsainen, Andreas H. Jucker & Jukka Tuominen (red.), *Diachronic Corpus Pragmatics*. Amsterdam: John Benjamins, 2014, pp. 213-235.

Croft, William & Cruse, D. Alan, *Cognitive Linguistics*. Cambridge: Cambridge University Press, 2004.

De Smet, Hendrik, Ghesquière, Lobke & Van de Velde, Freek (red.), 'Multiple source constructions in language change'. Special issue of *Studies in Language*, 37 (2013).

Diewald, Gabriele, 'A model for relevant types of contexts in grammaticalization.' In: Ilse Wischer & Gabriele Diewald (red.), *New Reflections on Grammaticalization*. Amsterdam: John Benjamins, 2002, pp. 103-120.

Diewald, Gabriele, 'The catalytic function of constructional restrictions in grammaticalization.' In: Elisabeth Verhoeven, Stavros Skopeteas, Yong-Min Shin, Yoko Nishina & Johannes Helmbrecht (red.), *Studies on Grammaticalization*, Berlin & New York: Mouton de Gruyter, 2008, pp. 219-240.

Doğruöz, A. Seza & Backus, Ad, 'Innovative constructions in Dutch Turkish: An assessment of ongoing contact-induced change', *Bilingualism: Language and Cognition*, 12 (2009), pp. 41-63.

Drinka, Bridget, 'Language contact'. In: Silvia Luraghi & Vit Bubenik (red.), *The Continuum companion to historical linguistics*. London: Continuum, 2010, pp. 325-345.

Fillmore, Charles J., 'The mechanisms of "Construction Grammar"', *Proceedings from the annual meeting of the Berkeley Linguistics Society*, 14 (1988), pp. 35-55.

Fillmore, Charles J., Kay, Paul & O'Connor, Mary C., 'Regularity and idiomaticity in grammatical constructions: The case of *let alone'. Language*, 64 (1988), pp. 501-538.

Fried, Mirjam, 'Construction Grammar as a tool for diachronic analysis', *Constructions and Frames,* 1 (2009), pp. 261-290.

Geeraerts, Dirk, *Diachronic Prototype Semantics.* Oxford: Clarendon Press, 1997.

Geeraerts, Dirk, Grondelaers, Stefan & Speelman, Dirk, *Convergentie en divergentie in de Nederlandse woordenschat: een onderzoek naar kleding- en voetbaltermen*. Amsterdam: P.J. Meertens-Instituut, 1999.

Geeraerts, D. & Kristiansen, G., 'Cognitive linguistics and linguistic variation'. In: John R. Taylor & Jeanette Littlemore (red.), *The Bloomsbury Companion to Cognitive Linguistics*. London: Bloomsbury Academic, 2014, 202-217.

Geeraerts, Dirk, Kristiansen, Gitte & Peirsman, Yves, 'Introduction: Advances in Cognitive Sociolinguistics'. In: Dirk Geeraerts, Gitte Kristiansen & Yves Peirsman (red.), *Advances in Cognitive Sociolinguistics*. Berlin & New York: Mouton de Gruyter, 2010, pp. 1-20.

Goldberg, Adele E., 'The inherent semantics of argument structure: The case of the English ditransitive construction', *Cognitive Linguistics*, 3 (1992), pp. 37-74.

Goldberg, Adele E., *Constructions: A Construction Grammar approach to argument structure*. Chicago: University of Chicago Press, 1995.

Goldberg, Adele E., 'Constructions: A new theoretical approach to language', *Trends in Cognitive Science* 7 (2003), pp. 219-224.

Goldberg, Adele E., *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press, 2006.

Gries, S.Th, 'Sources of variability relevant to the cognitive sociolinguist, and corpus- as well as psycholinguistic methods and notions to handle them', *Journal of Pragmatics,* 52 (2013), pp. 5-16.

Gries, Stefan Th. & Hilpert, Martin, 'Modeling diachronic change in the third person singular: a multi-factorial, verb- and author-specific exploratory approach', *English Language and Linguistics*, 14 (2010), pp. 293-320.

Grondelaers, Stefan, Speelman, Dirk & Geeraerts, Dirk, 'National variation in the use of *er* "there": Regional and diachronic constraints on cognitive explanations'. In: Gitte Kristiansen & René Dirven (red.), *Cognitive Sociolinguistics: Language Variation, Cultural Models, Social Systems*. Berlin & New York: Mouton de Gruyter, 2008, pp. 153-203.

Grondelaers, Stefan, Van Aken, Hilde, Speelman, Dirk & Geeraerts, Dirk, 'Inhoudswoorden en prepositie als standaardiseringsindicatoren. De diachrone en synchrone status van het Belgische Nederlands', *Nederlandse Taalkunde,* 6 (2001), pp. 179-202.

Heine, Bernd, 'On the role of context in grammaticalization'. In: Ilse Wischer & Gabriele Diewald (red.), *New Reflections on Grammaticalization*. Amsterdam: John Benjamins, 2002, pp. 83-101.

Heine, Bernd & Kuteva, Tania. *Language Contact and Grammatical Change*. Cambridge: Cambridge University Press, 2005.

Heine, Bernd & Kuteva, Tania. *The Changing Languages of Europe*. Oxford: Oxford University Press, 2006.

Hilpert, Martin, *Constructional change in English: Developments in allomorphy, word formation, and syntax*. Cambridge: Cambridge University Press, 2013.

Höder, Steffen, 'Multilingual constructions: A diasystematic approach to common structures'. In: Kurt Braunmüller, & Christoph Gabriel (red.), *Multilingual Individuals and Multilingual Societies* (241-257). Amsterdam: John Benjamins, 2012, pp. 241-257.

Höder, Steffen, 'Constructing diasystems: Grammatical organisation in bilingual groups'. In: Tor A. Åfarli & Britt Mæhlum (red.), *The sociolinguistics of grammar*. Amsterdam: John Benjamins, 2014, pp. 137-152.

Hoffmann, Sebastian & Mukherjee, Joybrato, 'Ditransitive verbs in Indian English and British English: A corpus-linguistic study', *Arbeiten aus Anglistik und Amerikanistik*, 32 (2007), pp. 5-24.

Hoffmann, Thomas & Trousdale, Graeme (red.), *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press, 2013.

Israel, Michael, 'The *Way* Constructions grow'. In: Adele E. Goldberg (red.), *Conceptual Structure, Discourse and Language*. Stanford: CSLI Publications, 1996, pp. 217-230.

Kortmann, Bernd & Szmrecsanyi, Benedikt (red.), *Linguistic Complexity: Second Language Acquisition, Indigenization, Contact*. Berlin & New York: Mouton de Gruyter, 2012.

Kristiansen, Gitte & Dirven, René, 'Introduction: Cognitive Sociolinguistics: Rationale, methods and scope'. In: Gitte Kristiansen & René Dirven (red.), *Cognitive Sociolinguistics: Language variation, cultural models, social systems*. Berlin & New York: Mouton de Gruyter, 2008, pp. 1-17.

Lakoff, George, *Women, fire and dangerous things: What categories reveal about the mind.* Chicago: University of Chicago Press, 1987.

Langacker, Ronald W., 'A dynamic usage-based model'. In: Michael Barlow & Suzanne Kemmer (red.), 24-63. *Usage-based Models of Language*. Stanford: CSLI publications, 2000, pp. 24-63.

Lucas, Christopher, 'Contact-induced language change'. In: Claire Bowern & Bethwyn Evans (red.), *The Routledge Handbook of Historical Linguistics*. London: Routledge, 2015, pp. 519-536.

Matras, Yaron, *Language contact*. Cambridge (MA): Cambridge University Press, 2009.

Nicolaï, Robert, 'Language contact: A blind spot in "things linguistic"', *Journal of Language Contact,* 1 (2007), pp. 11-21.

Noël, Dirk, 'Diachronic construction grammar vs. grammaticalization theory', *Functions of Language,* 14 (2007), pp. 177-202.

Pietsch, Lukas, 'What has changed in Hiberno-English: Constructions and their role in contact-induced change', *Sprachtypologie und Universalienforschung* 63 (2010), pp. 118-145.

Pijpops, Dirk & Van de Velde, Freek, 'Constructional contamination. A case study.', *Folia Linguistica*, ter perse.

Rickford, John R., 'Situation: Stylistic Variation in Sociolinguistic Corpora and Theory', *Language and Linguistics Compass* 8 (2014), pp. 590-603.

Rutten G.J., 'Vroegmoderne relativa: naar een diachrone constructiegrammatica', *Nederlandse Taalkunde,* 15 (2010), pp. 1-32.

Siewierska, Anna & Hollmann, Willem, 'Ditransitive clauses in English with special reference to Lancashire dialect'. In: Mike Hannay & Gerard J. Steen (red.), *Structural-functional studies in English grammar*. Amsterdam: John Benjamins, 2007, pp. 83-102.

Soares da Silva, Augusto, 2015, 'Cultural cognitive models of language variation'. In: Jocelyne Daems, Eline Zenner, Kris Heylen, Dirk Speelman & Hubert Cuyckens (red.), *Change of Paradigms – New Paradoxes*. Berlin & New York: Mouton de Gruyter, 2015, pp. 253-274.

Speelman, Dirk & Geeraerts, Dirk Causes for causatives: the case of Dutch 'doen' and 'laten' In: Ted Sanders & Eve Sweetser (red.), *Causal Categories in Discourse and Cognition*. Berlin & New York: Mouton de Gruyter, 2009, pp. 173-204.

Szmrecsanyi, Benedikt, 'The English genitive alternation in a cognitive sociolinguistics perspective'. In: Dirk Geeraerts, Gitte Kristiansen & Yves Peirsman (red.), *Advances in Cognitive Sociolinguistics*. Berlin & New York: Mouton de Gruyter, 2010, pp. 141-166.

Tomasello, Michael, *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge: Harvard University Press, 2003.

Traugott, Elizabeth C., 'The grammaticalization of *NP of NP* patterns'. In: Alexander Bergs & Gabriele Diewald (red.), *Constructions and Language Change*. Berlin & New York: Mouton de Gruyter, 2008a, pp. 23-45.

Traugott, Elizabeth C., 'Grammaticalization, constructions and the incremental development of language: Suggestions from the development of degree modifiers in English'. In: Regine Eckhardt, Gerhard Jäger & Tonjes Veenstra (red.), *Variation, Selection, Development: Probing the Evolutionary Model of Language Change*. Berlin & New York: Mouton de Gruyter, 2008b, pp. 219-250.

Traugott, Elizabeth C., 'Toward a coherent account of grammatical constructionalization'. In: Jóhanna Barðdal et al. (eds), *Diachronic Construction Grammar*. Amsterdam: Benjamins, 2015, pp. 51-79.

Traugott, Elizabeth C. & Trousdale, Graeme, *Constructionalization and Constructional Changes*. Oxford: Oxford University Press, 2013.

Trousdale, Graeme, 'Issues in constructional approaches to grammaticalization in English'. In: Katerina Stathi, Elke Gehweiler & Ekkehard König (red.), *Grammaticalization: Current views and issues*. Amsterdam: John Benjamins, 2010, pp. 51-71.

Trudgill, Peter, *Sociolinguistic Typology: Social Determinants of Linguistic Complexity*. Oxford: Oxford University Press, 2011.

Weinreich, Uriel, Labov, William & Herzog, Marvin, 'Empirical foundations for a theory of language change'. In: Winfred P. Lehmann & Yakov Malkiel (red.), *Directions for Historical Linguistics: A Symposium*. Austin: University of Texas Press, 1968, pp. 95-195.

Winter-Froemel, Esme, *Entlehnung in der Kommunikation und im Sprachwandel. Theorie und Analysen zum Französischen*. Berlin & New York: Mouton de Gruyter, 2011.

Winters, Margaret, Tissari, Heli & Allan, Kathryn (red.), *Historical Cognitive Linguistics*. Berlin & New York: Mouton de Gruyter, 2010.
Wolf, Hans-Georg & Polzenhagen, Frank. *World Englishes. A Cognitive Sociolinguistic approach*. Berlin & New York: Mouton de Gruyter, 2009.
Zenner, Eline, Cognitive Contact Linguistics. The macro, meso and micro influence of English on Dutch. Proefschrift KU Leuven, 2013.
Zenner, Eline & Kristiansen, Gitte (red.), *New Perspectives on Lexical Borrowing: Onomasiological, Methodological and Phraseological Innovations*. Berlin & New York: Mouton de Gruyter, 2014.

Over de auteurs

Timothy Colleman, Universiteit Gent, Ghent University research team on Linguistic Meaning & Structure, Blandijnberg 2, B-9000 Gent. E-mail: timothy.colleman@UGent.be

Freek Van de Velde, Katholieke Universiteit Leuven, onderzoeksgroep Quantitative Lexicology and Variational Linguistics, Blijde-Inkomststraat 21, B-3000 Leuven. E-mail: freek.vandevelde@kuleuven.be

Constructional complexification

The rise of double modal constructions in Dutch

Evie Coussé

MEM 67 (2): 149–176
DOI: 10.5117/TET2015.2.COUS

Abstract

This articles investigates the rise of double modal constructions in Dutch. Double modal constructions combine two modal auxiliaries with one lexical verb (e.g. *zal moeten gaan* 'shall have to go'). Little is known on why and how such complex verb constructions came into being. This article presents historical data for the earliest double modal constructions in Middle Dutch (13th century) drawing on both previous corpus studies and on new empirical observations. The historical data is then analyzed from a diachronic construction grammar perspective. The main theoretical point of this article is that the emergence of double modal constructions is a case of 'constructional complexification', i.e. the process in which constructions become increasingly larger in size.

Keywords: grammaticalization, construction grammar, auxiliary, modality, complexity

1 Introduction

The development of verb constructions has received a great deal of attention within historical linguistics in recent decades. The emergence of modal verbs, for instance, has featured as a prime example of grammaticalization and of related phenomena such as subjectification and intersubjectification (Plank, 1984; Traugott, 1989; Diewald, 1999; Traugott and Dasher, 2002; Byloo and Nuyts 2014, among many others). The focus of that research has recently shifted from studying the development of auxiliaries as such to the wider syntactic context in which they grammaticalize (Diewald, 2006; Hilpert, 2008; Coussé 2013, 2014), reflecting a growing alli-

ance between grammaticalization research and construction grammar (Traugott, 2008; Hilpert, 2008, 2013; Traugott and Trousdale, 2013). The diachronic study of verb constructions has recently moved beyond the level of individual constructions. Diewald (2009) and Diewald and Smirnova (2012), for instance, have argued that the integration of auxiliaries into a verbal paradigm should be seen as an integral part of their grammaticalization process.

A topic that has been largely ignored is the tendency of verb constructions to start combining into more elaborate assemblies after their initial establishment in the language. This process can be illustrated with the historical development of the Dutch verbal system. The earliest verb constructions in Old Dutch and Early Middle Dutch are two-verb constructions consisting of one auxiliary verb and one non-finite main verb. Longer verb constructions stacking multiple auxiliaries are still very rare in earlier Dutch. It was only in the course of the Middle Dutch period and onwards that such longer constructions were introduced into the language. This process has been mentioned in passing in the literature, but no systematic corpus data are available, let alone a more theoretical account of the process and its underlying motivations. This article presents an exploratory study of this elaboration process in the framework of diachronic construction grammar.

The central idea of this article is that the innovative long verb constructions are more complex than the two-verb constructions that precede them. Croft (2001, p.17) and Croft and Cruse (2004, p. 255) point out that constructions may vary in their degree of complexity (atomic – complex) and schematicity (substantive – schematic). Studies in diachronic construction grammar typically focus on changes in schematicity in constructions and constructional networks (and other dimensions such as productivity and compositionality, e.g. Trousdale and Traugott, 2013, p. 13). Changes in the dimension of complexity are generally overlooked. This article argues that the emergence of long verb clusters is a case of 'constructional complexification', i.e. the process in which constructions become increasingly larger in size.[1] The present study more specifically looks into the motivating factors leading to the creation of long verb constructions, taking the usage-based approach advocated by Von Mengden and Coussé (2014). It also fleshes out how two-verb constructions serve as source constructions for complex verb constructions, relating to the work published in De Smet, Ghesquière and Van de Velde (2013) on multiple source constructions in constructional change. The article furthermore builds on the constructionist framework laid out in Traugott and Trousdale

(2013), detailing how constructional networks change and how new constructions are introduced in the network.

This present study addresses in particular the rise of double modal constructions in Dutch – one of the few complex verb constructions that have been studied in some detail from a diachronic perspective (Coupé, 2009, 2015; Coupé and Van Kemenade, 2009). The following examples illustrate how two modal verbs may be combined in present-day Dutch.

(1) *De burger **zal** zijn pensioenprobleem zelf **moeten oplossen**.*[2]
The citizen shall.3SG his pension.problems self must.INF solve.INF
'People will have to solve their own pension problems themselves.'

(2) *Rode Kruis **wil** duizenden Nederlanders om hulp **kunnen vragen**.*
Red Cross will.3SG thousands Dutchmen for help can.INF ask.INF
'The Red Cross wants to be able to ask thousands of Dutchmen for help.'

(3) *Waarom **zou** een man niet **mogen huilen**?*
Why should.3SG a man not may.INF cry.INF
'Why would a man not be allowed to cry?'

(4) *Een grapje op het werk **moet** toch **kunnen**?*
A joke on the job must.3SG surely can.INF
'Surely it must be allowed to joke at work?'

There is little research discussing the double modal construction in its own right, apart from the seminal paper of Jędrzejowski and Van de Vate (2013). A defining formal feature of the construction is that the first modal is finite and the second embedded modal has the form of an infinitive. As such the construction differs from the well-studied double modal construction in some Southern American varieties of English, Northern British English and Scots, where two finite modals are combined (Di Paolo, 1989; Battistella, 1995; Hasty, 2012). Jędrzejowski and Van de Vate (2013) show that the double modal construction in Dutch is compatible with a wide range of dynamic, deontic and epistemic meanings.

The article is structured as follows. Section 2 reviews the historical data for the earliest double modal constructions in Middle Dutch, drawing on both previous corpus studies and on my own empirical observations. The focus will be on three-verb constructions combining two modals with one lexical infinitive (thus ignoring cases like (4), reflecting the preponderance of this construction type in the historical data and the secondary literature. Section 3 complements the available historical observations with an exploratory semantic analysis of double modals in the earliest chancery texts

for Dutch (i.e. early 13th century). These texts are translations from Latin originals, allowing us to enrich the historical study of modal meanings with methods taken from parallel corpus linguistics. Section 4 addresses the historical development of double modals from a diachronic construction grammar perspective, discussing both possible motivations and mechanisms of constructional complexification in language change. Section 5 brings together the most important empirical and theoretical findings of this study.

2 Historical observations

The first double modal constructions date from the 13th century. Coupé (2009, 2015) and Coupé and Van Kemenade (2009) systematically searched the Corpus Gysseling and the Corpus Van Reenen-Mulder, containing chancery texts from the 13th and 14th century respectively, for double modals with an additional lexical infinitive. They found the first attestation in a chancery text from the city administration of Brussels, dated 1277, presented in (5).

(5) *soe dat deen sonder den andren niet daer towe en* ***sal moghen gaen***
(Coupé and Van Kemenade 2009:261)
so that the.one without the other not there to not shall.3SG may.INF go.INF
'so that the former one shall not be able/allowed to go there without the other one'

The first double modal construction, however predates this by at least a couple of decennia. I was able to identify ten double modals in chancery texts written before 1277. Some of these earlier attestations are given in (6) to (8).

(6) *Nochdanne neghen uan den mestren ne* ***salmogen*** *prouende* ***gheuen*** *iemene no sieken no gansen in dit hus; het ne si bi der schepenen wetene. ende bi haren wille.* (1236 Statutes)
nonetheless none of the masters not shall.3PL.may.INF ration give.INF someone nor sick nor healty in this house it not be by the aldermen's knowledge and by their will
'Nonetheless, none of the masters shall be able/allowed to give a ration to anyone, neither the sick, nor the healthy inhabitants of this

house, except with the knowledge of the aldermen and with their approval.'

(7) *Te desen dinghen die hir sin uorseit, ne**sullen** scepnen no portren niet **mogen** toe **doen**. no wandelen. no beteren, hensi bi scrauen wille.* (1237 Charter)
to these things that here are aforesaid not.shall.3PL aldermen nor citizens not may.INF to do.INF nor change.INF nor correct.INF it.-not.be by the.count will
'Neither the aldermen nor the citizens shall be able/allowed to add, change or correct these things, except with the approval of the count.'

(8) *dit soude wesen grote urome ende bate den lande. ende des **soude** men wel **moghen** wl **comen*** (Gent 1240-1260)
this should be great profit and benefit the country and it should.3SG one well may.INF full come.INF
'This should be of great profit and benefit to the country and one should be able/allowed to accomplish it.'

Most of the earlier examples were found in the *Statuten van de Gentse leprozerie* 'Statutes of the Ghent leper house' (from 1236) and the *Gentse keurenvertaling* 'Ghent charter translation' (from 1237). Both texts represent the oldest chancery documents written in the Dutch language. It is not impossible that the emergence of double modal constructions actually occurred even before these first attestations.

Coupé (2009, 2015) and Coupé and Van Kemenade (2009) observe that the finite modal in early double modal constructions is always a variant of *zullen* 'shall'. This generalization also holds for the ten earlier examples found in this study. The non-finite modal is *mogen* 'may' (first observed in 1277), *willen* 'will' (1281) and *moeten* 'must' (1292). It appears that *zullen* 'shall' and *kunnen* 'can' are not used as embedded modals. Coupé and Van Kemenade (2009) do not draw attention to the fact that *mogen* is used much more frequently than *moeten* and *willen* in their data (39 attestations versus 4 and 2 according to table 15.3 on page 261). In my own ten attestations, embedded *mogen* is also strongly represented, as is *moeten* (5 times each).

The historical observations for Dutch show interesting parallels to the development of double modal constructions in English. Double modals appears to be lacking in Old English. Nagle (1993, p. 366) does find some combinations of modals in the Toronto Microfiche Concordance of Old English, but on closer inspection it turns out that the embedded modals are used as ordinary main verbs with a lexical meaning. Visser (1963-1973,

vol. III, §1685, §1649, §2134) reports numerous examples of double modals in Middle English, the earliest ones dating from the 13th century. Some examples are given in (9) to (11).

(9) *þatt mannkinn* ***sholde muȝhenn*** *well* ***Upp cumenn*** *inntill heoffne* (c. 1200 Orm 3944)
that mankind should.3SG must.INF well up come.INF into heaven
'that mankind should be able to come into heaven'

(10) *þatt I* ***shall cunnen cwemenn*** *Godd.* (c. 1200 Orm 2958)
that I shall.1SG can.INF please.INF God
'that I shall be able to please God'

(11) *He* ***muste kunne evacuen*** *him þat is ful of yuel humouris* (c. 1400 Lanfranc)
he must.3SG can.INF relieve.INF him that is full of yellow bile
'He needs to be able to relieve him who is full of yellow bile'

Visser (1963-1973, vol. III, §2134) mentions explicitly that the finite modal in these early attestations is almost exclusively *shall* or *should.* As non-finite embedded modals, *may, can* and *will* are reported. Nagle (1993, p. 367) gives additional evidence that double modal constructions in Middle English are predominantly headed by *shall.* He did not find any double modals headed by *may, must* or *might* in the concordances to *Gawain* or to the works of Chaucer. He finds, moreover, that *shall may* is the most common modal combination. Ogura (1993) presents some additional examples of *shall may* plus infinitive in the Wycliffite Gospels dating from the end of the 14th century.

The combined observations taken from Middle Dutch and Middle English suggest that the double modal construction emerged at the beginning of the 13th century. The earliest attestations show very limited combinatory possibilities: the finite modal is always *zal/shall*; the embedded modal is usually *mogen/may* and to a lesser extent also *moeten* in Early Middle Dutch. These findings suggest that the double modal construction started off in a very specific context, or in more constructionist terms, that it goes back to two micro-constructions (more details are given in section 4.3).

Most of the studies on early double modals focus on their formal properties. Little is known about their meaning. Coupé (2009) suggests that finite *zullen* in early double modals is used as a future auxiliary and that the embedded modal expresses meanings such as obligation and ability.

The *Vroegmiddelnederlands Woordenboek* (VMNW) 'Dictionary of Early Middle Dutch' downplays the future interpretation of *zullen* in combination with *mogen* and *moeten*, emphasizing instead its interpretations of obligation and permission; see (12).

(12) In combinatie met een ander werkwoord met de betekenis 'verplicht zijn, schuldig zijn, moeten, mogen'. Een futuraal aspect is hier niet geheel uit te sluiten. Bij het gebruik van deze formuleringen in keuren e.d. lijkt echter in eerste instantie sprake te zijn van oneigenlijk modaal gebruik. Alleen aangetroffen met *moeten* en *moghen.* (VMNW lemma *sullen* §I.2)
'In combination with another verb with the meaning 'be obliged, owe, must, may'. A future aspect is not to be ruled out entirely here. When used in formulations in chancery texts and the like, however, dynamic modal usage seems to be of prior importance. Only encountered with *moeten* and *mogen.*'

In order to get a better insight into the meaning of early double modals, section 3 presents an exploratory semantic study of the earliest cases found in the statutes of the Ghent leper house from 1236 (henceforth the '1236 Statutes') and the Ghent charter translations from 1237 (the '1237 Charter').

3 An exploratory semantic analysis

The 1236 Statutes and the 1237 Charter are among the oldest chancery texts written in Dutch.

They go back to an original written in Latin. This fact is usually seen as a problem in historical corpus studies due to the risk of direct translations from the original. The double modals found in the 1236 Statutes and 1237 Charter thus run the risk of being an artifact of the translation process rather than being an authentic part of the language. This issue can be relatively easily resolved by checking the originals of both texts. We are fortunate that the originals have not only stood the test of time but also are easily accessible in published form (Koch, 1960; Gysseling, 1963). The preservation of the translations together with their originals also presents us with an opportunity to approach both texts as a small parallel corpus, allowing us to come to grips with the elusive meaning of modals. Latin is known to express some tense and modal categories in a more synthetic way (notably by verbal conjugation and some participial constructions)

than Middle Dutch. These Latin inflectional markings may help us determine what semantic meanings are expressed in double modal constructions in Middle Dutch.[3]

3.1 Corresponding constructions in the Latin original

Let us first consider the cases with *zal mogen* plus an infinitive. The 1236 Statutes contain one attestation of this modal combination (given in 13) and the 1237 Charter provides three attestations (given in 14 to 16). Both the double modal constructions in Dutch and the corresponding sequences in Latin are marked in bold. The English glosses and translations pertain to the Dutch version, as the focus of this study is on double modals in Dutch.

(13) *Nullus autem magistrorum* ***conferre poterit*** *prebendam in dicta domo alicui sive sano sive leproso; sine conscientia & assensu scabinorum.*
Nochdanne neghen uan den mestren ne ***salmogen*** *prouende* ***gheuen*** *iemene no sieken no gansen in dit hus; het ne si bi der schepenen wetene. ende bi haren wille.* (1236 Statutes)
nonetheless none of the masters not shall.3PL.may.INF ration give.INF someone nor sick nor healthy in this house it not be by the aldermen knowledge and by their will
'Nonetheless, none of the masters shall be able/allowed to give a ration to anyone, neither the sick, nor the healthy inhabitants of this house, except with the knowledge of the aldermen and with their approval.'

(14) *Quocienscumque uero super huiusmodi falcitate summoniti fuerint, nullatenus* ***poterunt contradicere****, quin diem a comite sibi prefixum teneant ubicumque comes uoluerit in Flandria.*
*ende so welken tiet so si uan duschedaenre ualsheit werden geboden; so ne****sul****si niet* ***mogen wederseghen****. si nemoten houden den dach die hem degraue heft gheset war so hi wille in ulandren.* (1237 Charter)
and so what time so they from such falsehood became summoned so no.shall.3PL.they not may.INF contradict.INF they not.must hold the day that them the.count has assigned where so he wants in Flanders
'And during the time they are being summoned of such falsehood, they shall not be able/allowed to contradict, they rather must hold the day that the count has assigned to them wherever he wants in Flanders'

(15) *Ad hec nec scabini nec burgenses aliquid* ***addere, mutare*** *nec* ***corrigere poterunt*** *nisi per consensum comitis*
*Te desen dinghen die hir sin uorseit, ne****sullen*** *scepnen no portren niet*

mogen *toe* ***doen****. no wandelen. no beteren, hensi bi scrauen wille.* (1237 Charter)
to these things that here are aforesaid not.shall.3PL aldermen nor citizens not may.INF to do.INF nor change.INF nor correct.INF it.-not.be by the.count will
'Neither the aldermen nor the citizens shall be able/allowed to add, change or correct these things, except with the approval of the count.'

(16) ***arrestari facere*** *et* ***stare potest****, donec ille qui possessor est bonorum, plegios sufficientes dederit*
so ***sal*** *hi* ***mogen doen hahten****. ende* ***doen staen****. onthir ende die houdre uan dien goede heuet ghegeuen genoegende borghen.* (1237 Charter)
thus shall.SG he may.INF do.INF arrest and do.INF stand.INF until and the owner of the property has given enough pledges
'Thus, he shall be able/allowed to make them stop and let them be arrested until the owner of the property has given enough pledges.'

Inspection of (13) to (16) shows that *zal mogen* plus infinitive corresponds to a periphrastic verb construction in Latin consisting of the finite modal verb *possum* and a lexical verb in the infinitive form. *Possum* is inflected for future tense in (13) to (15) and for present tense in (16). This type of close structural correspondence has also been observed in Middle English. Ogura (1993) found that all cases of *shall may* plus infinitive in the Wycliffite Gospels correspond to future *possum* plus infinitive in the Latin original.

Now let us turn to the attestations of *zal moeten* plus infinitive. All instances of this micro-construction are found in the 1236 Statutes. Two of these cases correspond to a passive subjunctive in Latin, given in (17) and (18), with the corresponding verbs marked in bold.

(17) *Sic minuti aliquantulum commodius* ***procurentur****.*
Die dus geblotlat sin; hem ***sal****men* ***moten*** *bed****don*** *ichtewat dan degemene prouende.* (1236 Statutes)
those thus bloodletted are them shall.3SG.one must.INF better.-do.INF somewhat than the.ordinary ration
'One will have to provide more to those who have undergone bloodletting than the ordinary ration.'

(18) *Qui uero existenti in penitentia aliquid ad uictum pertinens absque magistri licentia contulerit; conuictus in illius penitentiam* ***retrudatur****.*
Sowie danne so dengonen die in penitentien es íet geuet dat tespisen

behort sonder desmesters orlof. ende dies bedregen werdet; men ***sal****ne* ***moten steken*** *inds gons penitentie.* (1236 Statutes)
so.who then so the.one who in punishment is something gives that to.eat belongs without the.master permission and it convicted became one shall.3SG.him must.INF put.INF in that punishment
'In case someone gives someone who is being punished something to eat without the permission of the master, and this person is found guilty of it, then one will have to punish him accordingly.'

The passive subjunctives in Latin are turned into an active construction by inserting the generic subject *men* 'one' in Dutch. Thus, the passive sequence *minuti procurentur* 'those who have undergone bloodletting should be taken care of' in (17) corresponds to the active sequence *hem salmen moeten beddon* 'one will have to take care of them better'.

A third example, given in (19), correspondes to two ablative absolute constructions, i.e. a noun and a modifying participle in the ablative case. More specifically, the sequence *salmoten sueren* integrates elements from the ablative absolutes *juramento addito* 'an oath was added' and *fidelitate conservanda* 'loyalty should be observed' and turns them into a finite verb construction.

(19) *Juramento nichilominus addito super fidelitate domui conservanda*
Oc ne mot niet achter bliuen hine ***salmoten sueren*** *getrowe tesine den hus* (1236 Statutes)
also not must not behind remain he.not shall.3SG.must.INF swear.-INF loyalty to.be the house
'It should also not be neglected that he will need to swear loyalty to the house.'

The last two instances of *zal moeten* plus infinitive do not have a corresponding form in the Latin original. The sequence *hi sal moten gedogen de pine* in (20) elaborates on the shorter nominal sequence *sub pena* 'under punishment', and for (21) no correspondence at all was found in the original.

(20) *non sub uoto. sed sub pena inferius annotata*
mesgripe hi dran; nemar hi ***sal moten gedogen*** *de pine die hierachter ghescreuen es.* (1236 Statutes)
violate he there.on not.but he shall.3SG must.INF endure.INF the punishment that hereafter written is

'If he violates this, he will have to endure the punishment that is written hereafter'

(21) [No Latin correspondence]
Die dis iet doet; hi ***sal moten gedogen*** *depine die men sculdech es den ouerhoregen.* (1236 Statutes)
who this something does he shall.3SG must.INF endure.INF the.punishment that one culpable is the offender
'If he does something against it, he will have to endure the punishment that one is obliged to assign the offender'

In summary, there turns out to be a wide range of corresponding forms in the Latin orginals. This suggests that double modals should not be seen as a direct translation of one particiular form in Latin. This is particularly true for *zal moeten* plus infinitive, where there is little structural correspondence between the original and the translation. Two attestations have no correspondence at all. The cases of *zal mogen* plus infinitive correspond more closely to the original construction with (future) *possum* plus infinitive. However, it should be pointed out that only a fraction of all cases of future *possum* plus infinitive are translated with *zal mogen* plus infinitive. An informal investigation of all instances of future *possum* plus infinitive in the 1236 Statutes (6 attestations) and the 1237 Charter (15 attestations) shows that they are most often translated with finite *mogen* plus infinitive in Dutch, as in (22).

(22) *Carnibus autem per totum annum excepta quadragesima etiam per aduentum tribus diebus licite* ***uesci poterunt*** *in septimana.*
Flesch ***mogen*** *si* ***eten*** *al dat iar. ende oc in aduent drie dage in de weke;* (1236 Statutes)
meat may.3PL they eat.INF all that year and also in Advent three days in the week
'They can/may eat meat all year round and also three days a week during Advent.'

The wide range of forms in Latin also comes with a large variety in meaning. The future tense inflection of *possum* overtly marks future time reference. The verb *possum* itself is used with modal meanings ranging from ability, capacity and permission, according to both the *Oxford Latin Dictionary* (covering Classical Latin before 200 AD) and the *Lexicon Latinitatis Nederlandicae Medii Aevi* (covering Medieval Latin in the Northern Nether-

lands). The passive subjunctive is used to express possibility, intention, necessity, and statements contrary to the fact. Ablative absolutes express general circumstances under which the action of a sentence occurs, which can be time, condition or cause, but the exact interpretation needs to be understood from the context (Morwood, 1999). This wealth of potential meanings in the Latin original suggests that the double modals of Middle Dutch are also able to cover this broad spectrum of meaning. However, it is possible that not all of the meaning potential of these forms is expressed in actual usage. The next subsection will therefore have a closer look at the meaning of the double modals in the context of their usage.

3.2 Contextual analysis

The double modals under investigation are taken from two chancery documents. The purpose of these documents is concisely stated in excerpts (23) and (24).

(23) *ombe de gemene nutschepe der gonre die nu sin en die hier na wesen sullen; de seden die men int hus gehouden heuet tote nu. ende de gesette dinge die men uord ward mer houden sal; die sin in desen brief gescreuen. bede bi der cracht des biscops uan dornoke. ende bi den goden wille der schepenen uan ghent* (1236 Statutes)
'For the common benefit of the people who are living now and who will be living hereafter, the customs that one has kept in the house until now and the decreed things that one will keep from now on, these are written in this charter, both by the power of the bishop of Tournai, and by the approval of the aldermen of Ghent.'

(24) *Dit sin de pointe die de graue geboet te houdene ouer al sin lant.* (1237 Charter)
'These are the points that the count orders to be kept in the whole of his land.'

The documents describe current rules and regulations (*de seden die men inte hus gehouden heuet tot nu*) and decree new ones (*de gesette dinge die men uord ward meer houden sal, de pointe die de graue geboet tehoudene*). Modal meanings such as obligation, permission and prohibition may be expected to prevail in this type of regulatory context. Nuyts, Byloo en Diepeveen (2010, p. 24) subsume these notions under directive modality, i.e. 'the agent participant in the SoA [State of Affairs] is instructed – with some degree of strength: advised, obliged, interdicted – or permitted to do what is involved in the SoA, on behalf of some source (the speaker him/

herself, or some other willful being or institution)'. A distinctive property of directive modality is the possibility of specifying the source of obligation or permission in context. Inspection of the double modals in the previous section shows that a number of attestations do have such an explicit source indicator, i.e. *bi der schepenen wetene ende bi haren wille* 'with the knowledge of the aldermen and their approval' in (13), *bi scrauen wille* 'with the approval of the count' in (15), and in (18) *sonder desmesters orlof* 'without permission of the master'. This points to a directive reading of these double modal constructions.

However, it should be noted that the directive source may also be broader than just the authorities specified in (23) and (24). Nuyts, Byloo en Diepeveen (2010, p. 24) argue that some directive usages of modals may be informed by a deontic assessment of the degree of moral acceptability of the proposition. This also seems to be the case for some of the double modals under investigation. In (16), the instruction to give some patients a larger ration may be ultimately guided by moral (religious) considerations such as compassion and charity. This moral dimension is also made explicit in (23) by referring to the *gemene nutschepe* 'common benefit'. But situational considerations, typically the realm of dynamic modality,[4] may serve as the ultimate driving force for the directive use of double modals. The execution of a punishment in (18) depends on the local circumstances specified by the context. Byloo and Nuyts (2014, p. 26) point out that dynamic-imposed modality is often ambiguous with directive modality in Dutch and that there might be a diachronic development from dynamic-imposed to directive modality.

The excerpt in (23) also makes future time reference explicit by *die hier na wesen sullen* 'who will be living hereafter' and *die men uord ward mer houden sal* 'that one will keep from now on'. However, none of the double modals under investigation is accompanied by such an explicit future reference. This may be a consequence of the explicit marking at the beginning of the document which makes other references redundant. It may also have to do with the fact that future time reference is implied by modal meanings such as obligation, permission and prohibition. Instructing the agent of the clause to do something implies 'a prediction on the part of the speaker that the situation in the proposition, which refers to an event taking place after the moment of speech, will hold', which is the focal use of futures, according to Bybee, Perkins and Pagliuca (1994, p. 244). This future implication is widely assumed to have been grammaticalized as part of the meaning of *zullen*, but is in principle also compatible with a directive usage of *moeten* or *mogen*, as is explicitly mentioned for *moeten* in

the VMNW (lemma *moeten* §II) and for *mogen* in Byloo and Nuyts (2011, p. 56).

3.3 Interim summary

The exploratory semantic study of early double modals shows that they are compatible with a wide variety of modal readings. It turns out that double modals translate a broad range of source constructions in Latin, expressing modal meanings such as ability, capacity, permission, necessity, etc. Contextual analysis of these double modals narrowed down these potential interpretations to mainly directive modality (obligation, permission and prohibition), with a flavor of deontic and dynamic-imposed modality also possible. The double modals under investigation also show some evidence of future time reference: a number of attestations translate explicit future tense constructions in Latin, and chancery documents in general are made to decree rules and regulations for the future.

4 A diachronic constructionist account

With the historical data in place, it is now time to integrate these observations into a coherent scenario of language change. A couple of questions are relevant with respect to constructional complexification. The double modal construction was claimed to be a multiple source construction emerging out of already existing 'single modal' constructions. A first major question is why language users start combining the available single modal constructions into more complex assemblies. Section 4.2 explores a usage-based motivation for this complexification process. In order to sketch the historical background for this process, section 4.1 first discusses the development of single modal constructions in Dutch. Another important question is how language users combine modal source constructions into a more complex verb construction. Section 4.3 elaborates on how combinatory principles from construction grammar can be put to work to arrive at a more complex construction. A third central question is what place the new complex modal construction takes in the constructional network. Section 4.4 analyses how complexification affects the degree of schematicity, productivity, compositionality and complexity of constructions. Finally, section 4.5 explores some further developments from early double modals to the double modals constructions we know today.

4.1 The development of modal source constructions

The emergence of double modal constructions was preceded by a stage with 'single' modal constructions consisting of a finite modal and a lexical infinitive. What do we know about the development of these modal constructions? There has been quite some work on the grammaticalization of the modals in these constructions. Much earlier work (Conradie, 1987; Duinhoven, 1997; Booij, Los and Rem, 2006; Coupé, 2009) considers Dutch modals to have developed in a chain-like fashion, as summarized by Coupé (2009, p. 99) in table 1.

Table 1 Semantic shifts in modal verbs

	zullen	*moeten*	*mogen*	*kunnen*
(Early) Middle Dutch	obligation	permission	ability	(lexical)
Modern Dutch	future/irrealis	obligation	permission	ability

Coupé (2009) suggests that the semantic shift in the modal system may have been triggered either by the development of a future/irrealis meaning of *zullen* (a pull chain) or by the introduction of *kunnen* as a modal verb expressing ability (a push chain).

One problem with table 1 is that it overtly simplifies the semantic meaning of modals. Section 3 illustrates that double modals are compatible with a wide range of meanings. The same picture also emerges from the empirical study of the 'single modals' *moeten, mogen* and *kunnen* by Byloo and Nuyts (2014) – summarizing a decade of meticulous semantic analysis of modals in historical sources. Their analysis first of all demonstrates that the core modals are polysemous throughout the history of Dutch. *Moeten*, for instance, is used with dynamic, directive, deontic and evidential meanings from the earliest sources onwards. Moreover, these meanings are not easy to distinguish in context but are often ambiguous. Byloo and Nuyts (2014) report a large number of ambiguous cases where two or more readings are possible for the same modal in context. Finally, each modal verb was found to have its own semantic development that goes beyond a systematic push or drag chain.

Let us have a closer look at the modal meanings for each of the modals attested in the period around the emergence of double modals, i.e. Old Dutch and Early Middle Dutch. Byloo and Nuyts (2014) reveal that the core modals are mainly used with dynamic meanings in the earliest sources. The modal *moeten* is also used to express directive meanings in Old Dutch, and in Early Middle Dutch deontic meaning and volition as well. All these meanings are already common for *mogen* in Old Dutch and

are complemented with epistemic, concessive and conditional meanings in Early Middle Dutch. The modal *kunnen* is the only modal that is still used with its lexical meaning in Old Dutch. It should however be noted that dynamic meanings already largely prevail for this modal in Old Dutch and Early Middle Dutch.

Another problem with table 1 is that modal meanings are suggested to be part of a 'modal system'. This is a structuralist notion par excellence. Modals are considered to be part of a paradigm and their meaning is defined in terms of oppositions. There have been recent attempts to reconcile the concept of 'paradigm' and 'semantic opposition' with a constructionist approach to grammaticalization (Nørgård-Sørensen, Heltoft and Schøsler, 2011; Diewald and Smirnova, 2012). I have serious doubts as to whether these structural concepts can successfully be integrated into construction grammar, especially within the cognitive usage-based branch of the framework. However, independent of this essentially theory-internal discussion, it is rather uncontroversial to assume that in the early stages of grammaticalization, modal paradigms are yet not in place. Diewald and Smirnova (2012), for instance, argue that paradigmatic integration should be considered the last stage of grammaticalization. First, constructions develop rather independently from each other, embedded in their own context and following their own grammaticalization path. It is only in later stages of their development that constructions are associated with each other and that their meanings may become constrained by paradigmatic oppositions.

4.2 Usage-based motivation for double modal constructions

The preceding subsection argues that single modal constructions can be used with a wide range of meanings from the earliest sources onwards, in a similar fashion to the double modals in section 3. This raises the question of why double modal constructions emerge when 'single modal' constructions seem to be perfectly able to express the same range of meanings. A complicating factor is that not every modal construction starts combining randomly. The historical data in section 3 clearly show that *zal mogen* and *zal moeten* were the first possible modal combinations in Dutch (and English).

The *Woordenboek der Nederlandsche Taal* (WNT) 'Dictionary of the Dutch Language' suggests that the first double modals are the result of reinforcement, or more specifically, the reinforcement of the directive *zal* by another modal also expressing obligation or prohibition.

(25) Versterkt met andere ww. in de bet. 'verplicht zijn', 'moeten', of, met ontkenning, '(niet)mogen'. Dit gebruik komt voor in ambtelijke teksten uit de 16de en 17de e. en kan voortkomen uit het verlies van de bet. 'verplicht zijn' van *zullen. Zullen* kan in deze zinnen een temporaal bet.-aspect hebben. (WNT lemma *zullen* §I.1.b)
'Reinforced with other verbs expressing the meaning 'to be obliged', 'must', or, with negation, 'may (not)'. This usage occurs in chancery texts from the 16th and 17th centuries and may arise from the loss of the meaning 'be obliged' of *zullen. Zullen* may have a temporal meaning aspect in these sentences.'

Reinforcement provides a compelling reason as to why two modals are used instead of only one. On a general level, reinforcement means adding new material to existing structures for reasons of emphasis. A typical example of reinforcement is *not at all*, in which the negative marker *not* is emphasized by the additional material *at all.* Reinforcement is a usage-based motivation for change, as it appeals to the socio-psychological behavior of the speaker. Von Mengden and Coussé (2014, p. 5) clarify that 'It is not that much the speaker's aim to achieve any semantic effects, like additions or specifications, but innovative usage is triggered for instance by the mere wish to achieve the highest possible attention for the point a speaker intends to make'.

In the specific case of double modals, a directive modal is added to *zal* in order to counterbalance its ongoing loss of directive meaning. Although this is not made entirely explicit in (25), this meaning loss is related to the grammaticalization of *zullen* as a future auxiliary. Indeed, it is well-known that speakers resort to reinforcement in the context of grammaticalization when the grammaticalizing element starts losing its original lexical force. A case in point is the negation marker *ne* in French, which came to be reinforced by the lexical element *pas* in the course of its grammaticalization. Similarly, the ongoing loss of directive force in *zullen* as the result of its grammaticalization is counterbalanced by adding another directive modal.

4.3 The integration of modal source constructions

Let us now have a closer look at how the modal source constructions are combined into the first double modal constructions. I assume a model of constructional integration within the framework of usage-based construction grammar (Langacker, 1987; Goldberg, 1995; Croft, 2001). Modal constructions are semi-schematic constructions with the modal as a substan-

tive element and a schematic position for the infinitive. The structure of the relevant modal source constructions is rendered in (26), with the syntactic pole of the construction to the left of the double arrow and the semantic pole to the right, following the notional conventions of Booij (2010) and Traugott and Trousdale (2013).

(26) $zal_i\ V_k \quad \leftrightarrow \quad \textbf{FUT}/DIR_i\ (PRED_k)$
$mag_i\ V_k \quad \leftrightarrow \quad FUT/\textbf{DIR}_i\ (PRED_k)$
$moet_i\ V_k \quad \leftrightarrow \quad FUT/\textbf{DIR}_i\ (PRED_k)$

The schematic position for predicates in these constructions was originally elaborated by lexical verbs. It may in principle also be elaborated with more schematic elements, such as another semi-schematic modal construction, if both component structures show enough conceptual overlap. One of the overlapping fragments in all modal constructions in (26) is the schematic position for a predicate. There is no reason to assume that modal constructions impose different selectional restrictions on the types of predicates elaborating the schematic predicate position. Moreover, all modals in (26) were argued to express directive meaning, although this meaning may be on its way out in *zullen*. The modals are also claimed to be compatible with future time reference, as a pragmatic implicature in *mogen* and *moeten*, and grammaticalized as part of its meaning in *zullen.* These differences in meaning salience are rendered in (25) using bold face.

I consider the innovative integration (or embedding, in more traditional terms) of a modal construction into another modal construction as a case of host-class expansion. Himmelmann (2004, p. 32) defines host-class expansion as construction-internal expansion of the class of elements with which the grammaticalizing element is in construction. Usually, host-class expansion is illustrated with a class of substantive elements, such as the nouns co-occurring with demonstratives being grammaticalized into articles (Himmelmann 2004, p. 32) or the past participles combined with the grammaticalizing perfect auxiliaries *have* or *be* (Coussé, 2014). Coussé and Van de Velde (2014), however, argue that the integration of a modal construction into a perfect construction (yielding a three-verb constructions of the type *heeft kunnen werken* 'has been able to work') is a further step in the host-class expansion of perfects observed in Coussé (2014). Conceptually, this host-class expansion represents an expansion in scope of the modal construction from modifying actual propositions to modifying a modal evaluation of this proposition. The integration of the modal constructions in (26) into the earliest double modal constructions is given in (27).

(27) $zal_i\ mogen_j\ V_k \leftrightarrow$ **FUT**/DIR$_i$ (FUT/**DIR**$_j$ (PRED$_k$))
$zal_i\ moeten_j\ V_k \leftrightarrow$ **FUT**/DIR$_i$ (FUT/**DIR**$_j$ (PRED$_k$))

The discussion has until now focussed on the conceptual side of the integration of modal constructions. An equally important aspect is the formal integration of the source constructions. I suggest that double modal constructions iconically reflect the embedding semantic structure elaborated in (27). One way of signalling embedding is to mark the embedded verb with a non-finite form. Indeed, the embedded modal in double modals has the form of an infinitive from the earliest attestations onwards – which is argued to be an innovation by Coupé (2009). She found that modal infinitives first occur in double modal constructions. Modals only appear with finite morphology in older sources in West Germanic (Old English, Old High German, Old Saxon), which can be related to their origin as preterit-present verbs. How did speakers then arrive at a modal infinitive in double modal constructions? In simple modal constructions, the embedded predicate is the infinitive of a lexical verb, represented in (28).

(28) $zal_{vf}\ V_{inf}$ ↔ **FUT**/DIR (PRED)
$mag_{vf}\ V_{inf}$ ↔ FUT/**DIR** (PRED)
$moet_{vf}\ V_{inf}$ ↔ FUT/**DIR** (PRED)

The embedded modal in the *zal* modal construction may likewise be coded as an infinitive. As opposed to ordinary lexical verbs, the infinitive form of modals in Old West Germanic cannot be available as an entrenched form that might be retrieved as a whole from memory. Instead, the infinitive form of modals needs to be composed in situ, as is represented by the square brackets in (29).

(29) $zal_{vf}\ [mag]_{inf}\ V_{inf} \leftrightarrow$ **FUT**/DIR (FUT/**DIR** (PRED))
$zal_{vf}\ [moet]_{inf}\ V_{inf} \leftrightarrow$ **FUT**/DIR (FUT/**DIR** (PRED))

Modal infinitives are thus a case of embedded productivity, a term originally used with regard to complex word formation by Booij (2010, p. 47), but which also turns out to be relevant for complex syntactic constructions.

Another way of coding embedding iconically is through word order. Double modals today stand out among the three-verb clusters by their dominant ordering: finite modal – modal infinitive – lexical infinitive (Barbiers et al., 2008). A historical corpus study by Coupé (2015) shows

that this word order also dominated in legal texts from the 14th to 16th century.

4.4 Schematicity and complexity in constructionalization

The integration of the modal source constructions described above gives rise to two new micro-constructions in the verbal constructional network (i.e. the constructions in 29). Most striking is their innovative form: a finite modal and a lexical verb in the infinitive are supplemented by a newly created modal infinitive. Both constructions also have a new semantic structure: modal meanings are combined in an asymmetric scope relation.[5] Traugott and Trousdale (2013, p. 22) argue that the creation of a new construction, what they call 'constructionalization', is accompanied by changes in schematicity, productivity and compositionality. These three dimensions do not seem to apply very well to the earlier described process of complexification.

The new modal constructions do not differ essentially in schematicity from other modal constructions in the network. They can be characterized as semi-schematic with two substantive elements and one open slot for lexical verbs. The productivity of the new micro-constructions also seems to be rather unremarkable with regard to other modal constructions. There is no reason to assume that double modals impose other selectional restrictions on the lexical verbs in the open slot than the single modals. Only the dimension of compositionality can be argued to differ between the new micro-constructions and the modal source constructions. Traugott and Trousdale (2013, p. 19) define compositionality as 'the extent to which the link between form and meaning is transparent'. It was pointed out in section 4.1 that modals are highly polysemous in the source constructions as a result of their ongoing grammaticalization. The link between form and meaning is more transparent in the new complex modal constructions. I argued that the morphological form and word order of the elements in the syntactic structure reflect their semantic embedding iconically. This isomorphic linking between form and meaning can be made more visible by only representing the most salient meaning components in both constructions, as in (30).

(30) $zal_i\ mogen_j\ V_k \leftrightarrow FUT_i\ (DIR_j\ (PRED_k))$
$zal_i\ moeten_j\ V_k \leftrightarrow FUT_i\ (DIR_j\ (PRED_k))$

Indeed, each syntactic element in the syntactic pole is linked to one semantic component in the semantic pole, as is indicated by the indexes *i, j*

and *k* in (30). The modal source constructions do not display such clear one-to-one correspondences between form and meaning. Thus, double modal constructions can be thought to be more compositional than their source constructions.

The dimensions of schematicity, productivity and (to a lesser extent) compositionality turn out to be fairly uninteresting in the initial stage of emerging complex modal constructions[6]. Does this imply that we are not dealing with the formation of new constructions? No. The 'problem' with the diachronic application of these dimensions is that they are mainly informed by research on grammaticalization and lexicalization. Grammaticalization, for instance, typically involves an increase in productivity and schematicity and a decrease in compositionality (Traugott and Trousdale, 2013, p. 112). The process of complexification, however, first and foremost affects the dimension of 'complexity' in the constructional network, which has remained beneath the radar in diachronic construction grammar. This can be illustrated in figure 1, representing the constructional network for modal constructions at the time when the first double modal constructions emerge.

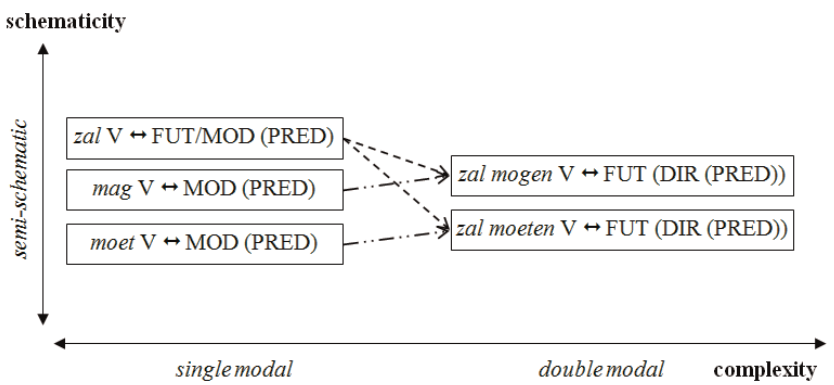


Figure 1 Modal constructional network

The multidimensional constructional network is projected onto two dimensions. The vertical dimension represents the degree of schematicity of constructions with fully substantive constructions at the bottom and fully schematic constructions on top (not shown in the diagram due to space limitations). Both the single and double modal constructions are at the same level of schematicity somewhere in the middle, being semi-schematic constructions with one open slot for lexical verbs. The horizontal dimension represents the degree of complexity of constructions with

atomic constructions to the left (not shown) and complex ones to the right. The double modal constructions cluster more to the right than the single modals.

Figure 1 demonstrates that the emergence of two new micro-constructions leads to network growth. Instead of expanding into the dimension of higher schematicity, a type of constructional growth that has been identified before, the new constructions venture into the dimension of higher complexity. Growth in the constructional network also comes with an increase in network density. Figure 1 shows how multiple new inheritance links are created between the single and double modal constructions. The links from the *zal* construction can be classified as (metaphorical) extension links relating to the host-class expansion in the *zal* construction, which is needed for the integration of other modal construction. The links from the *mag/moet* constructions can be seen as subpart links. I hypothesize that these inheritance links are a first step towards a tighter association between modal constructions as one family of semantically and formally related constructions. Note in this respect that figure 1 does not draw any inheritance links between the single modal constructions. I see no compelling evidence in the historical data of the time to assume a more schematic construction of the type 'schematic modal verb plus infinitive'.[7]

4.5 Towards modern double modal constructions

After the emergence of *zal mogen/moeten* plus infinitive, other double modal combinations start to be used in the course of the Dutch language history. Unfortunately, there is little historical data on what combinations were introduced when, exactly. We know from the data in Coupé and Van Kemenade (2009, p. 261) that *zullen* occasionally is combined with the volitional modal *willen* in the 13th century. Later on, it may be reasonably assumed that other modals also start to embed modal constructions, leading up to the wide range of modal combinations we know today. How can these broad developments be accounted for in constructionist terms?

A first issue is why language users start making new modal combinations. The strategy of reinforcement is very specific to the combination of *zullen* and *mogen/moeten* and cannot easily be extended to other modal combinations. My suggestion is that language users instead started 'exploiting' the constructional potential of the early double modals, combining *zal* with *mogen/moeten*. These early double modal constructions more specifically provide a way of combining modal meanings in an asymmetric scope relation. This property of double modals was argued to be the result

of the constructional integration of modals used to reinforce each other. However, once the first double modal micro-constructions are in place, their form and semantic structure may be exploited by language users in order to combine other modals, quite independently of the original usage-based motivations for modal integration.

This brings us to the question of how this constructional exploitation comes about. I assume that the exploitation of the early double modal construction is a gradual process guided by analogical thinking. In view of the historical data, language users have first expanded the range of embedded modals in double modals headed by *zullen*. Repetitive embedding of other modals in these early double modal constructions ultimately gives rise to a more schematic construction, as in (31).

(31) $zullen_{vf}\ [MV]_{inf}\ V_{inf} \leftrightarrow$ FUT/DIR (LOWER SCOPE (PRED))

This schematic construction has an open slot for modal verbs in the form pole corresponding to a very general modal meaning in the semantic structure. The only restriction imposed on the embedded modal is that it should have lower semantic scope than the meaning of *zullen*. The establishment of an open slot for modals in double modal constructions is not trivial. It indicates that modals both formally and semantically behave as one category. This observation is relevant in the light of my earlier remark that there is no reason for assuming that modals in the single modal constructions form a category. The creation of a double modal construction with an open slot for modals gives us good evidence that such a separate modal category has been formed.

The diversification of modal combinations with *zullen* may have spurred on the further proliferation of modal combinations, this time not only extending the range of embedded modals but also the finite ones. The incremental introduction of new modal combinations will ultimately lead to the establishment of the present-day double modal construction, given in (32).

(32) $MODAL_{vf}\ [MODAL]_{inf}\ V_{inf} \leftrightarrow$ HIGHER SCOPE (LOWER SCOPE (PRED))

This construction allows, in the most general way, the combination of two modals in an asymmetric scope relation. It is a fully schematic construction that should be situated high on the vertical schematicity scale in figure 1. Below this fully schematic construction an intricate network of subsche-

mas and micro-constructions could be drawn, if we only had more diachronic data to substantiate them.

5 Conclusion

This study investigated the rise of double modal constructions in Dutch as a case of constructional complexification. This section brings together the most important empirical and theoretical findings of the study.

The empirical study of double modals revealed that the earliest attestations of the construction date from the 13th century. These first observations show very limited modal combinatory possibilities: the finite modal is always *zullen* and the modal infinitive is either *mogen* or *moeten.* An exploratory semantic analysis of the earliest double modals indicated that these cases predominantly express directive meanings with possibly deontic and dynamic-imposed modal overtones. There is also evidence of future time reference in some of the observations. Moreover, future time reference was argued to be available as a pragmatic implication of the directive meanings in all attestations.

One important theoretical question was why language users start combining modal constructions in the 13th century. It was argued that reinforcement is a powerful usage-based motivation for adding new material to already existing structures. Applied to early double modals, this gave rise to the hypothesis that *zullen* is reinforced by the directive modals *mogen* or *moeten* in order to counterbalance its ongoing loss of directive force as a result of grammaticalization.

The study also addressed the question of how exactly two modal constructions were combined into one complex construction. It was proposed that one modal construction was embedded into the schematic slot of the other. This embedding involves a number of innovations. First, the schematic slot of the embedding construction undergoes host-class expansion from incorporating only lexical predicates to more schematic types of predicates. This host-class expansion also implies an expansion in scope for the entire construction, from modifying only actual propositions to modifying a modal evaluation of this proposition. Furthermore, the embedding leads to the creation of modal infinitives, reflecting the embedded status of the second modal.

The embedding of modal constructions gives rise to two new micro-constructions in the modal constructional network. The study showed that the creation of these two new nodes in the network does not correlate

with changes in schematicity and productivity, two defining criteria of constructionalization given by Traugott and Trousdale (2013). Complexification instead gives rise to new nodes in the network that are higher in complexity, a network dimension that has largely remained beneath the radar in diachronic construction grammar. The creation of the new double modals also involves an increase in network density through the establishment of multiple inherence links between the source and target constructions.

This study has only scratched the surface of why and how construction complexification may manifest itself in language. The phenomenon is more pervasive than this small case study is able to do justice to. The findings for double modals in Dutch need be related to the development of other complex verb constructions in Dutch – such as the integration of modals in perfect constructions explored in Coussé and Van de Velde (2014). The rise of long verb constructions also shows interesting parallels to the continuous extension of the nominal phrase in the history of Dutch, as detailed in Van de Velde (2009). These developments are of course not unique to Dutch but also have parallels in other languages. A wealth of case studies lies ahead of us waiting to be explored.

Notes

1. It should be noted that the notion of complexity in this article (i.e. basically construction size) is not intended to relate directly to issues of complexity in language processing – although larger size typically comes with more embedding and thus potentially higher processing costs.
2. Examples (1) to (4) are authentic examples found on the internet. Example (1) was found on http://trends.knack.be/economie/nieuws/finance/de-burger-zal-zijn-pensioenprobleem-zelf-moeten-oplossen/article-4000572814603.htm, example (2) on http://www.ad.nl/ad/nl/1012/Nederland/article/detail/3636997/2014/04/17/Rode-Kruis-wil-duizenden-Nederlanders-om-hulp-kunnen-vragen.dhtml, (3) on http://plazilla.com/page/4295015312/waarom-zou-een-man-niet-mogen-huilen, consulted 5 May 2014, and (4) on http://www.bureaugelijkebehandeling.nl/nieuws/een-grapje-op-het-werk-moet-toch-kunnen. All pages were consulted on 5 May 2014.
3. I am indebted to Catharina Peersman at the University of Sheffield for her help with identifying and analyzing the Latin correspondences in both chancery texts.
4. Defined by Nuyts, Byloo en Diepeveen (2010, p. 17) as 'indicating abilities or needs of the agent participant in a SoA [state of affairs], as expressed by the modal auxiliary in *John can eat like a lion*, or of possibilities or necessities inherent in a SoA as a whole, as involved in the auxiliary in *it can rain here in winter*'.
5. Asymmetric scope relations in the modal domain have been thoroughly examined from both a formal-generative and a semantic-functional perspective (e.g. Cinque, 1999 and Byloo and Nuyts, 2013, respectively).

The formal-generative approach in particular seems to take the hierarchy of modal meanings (or 'functional projections' in their specific terminology) as a given fact of language, whereas the present article rather takes an emergent perspective. Asymmetric modal scope relations also relate to the so-called epistemic non-finiteness gap in long verb constructions, discussed in Abraham (2001, 2002) and Reis (2001) for German, excluding an epistemic reading of embedded modals.It is beyond the scope of the article to go deeper into these matters.

6. Note that I am not addressing later developments here. Reviewer Graeme Trousdale points out that the incorporation of a wider range of modals in both the first and second position (as indicated by contemporary examples 1-4) does imply an increase in schematicity. These developments are described in more detail in section 4.5.
7. Evidence for this is given in section 4.6.

References

Abraham, Werner, 'Modals: towards explaining the "epistemic non-finiteness gap"'. In: Reimar Müller & Marga Reis (eds.), *Modalität und Modalverben im Deutschen.* Hamburg: Buske, 2001, pp. 7-36.

Abraham, Werner, 'Modal verbs. Epistemics in German and English'. In: Sjef Barbiers, Frits Beukema & Wim van der Wurff (eds.), *Modality and its interaction with the verbal system*. Amsterdam: John Benjamins, 2002, pp. 19-50.

Barbiers, Sjef, van der Auwera, Johan, Bennis, Hans, Boef, Eefje, De Vogelaer, Gunther & van der Ham, Margreet, *Syntactic Atlas of the Dutch Dialects Volume II*. Amsterdam: Amsterdam University Press, 2008.

Battistella, Edwin L., 'The syntax of the double modal constructions', *Linguistica Atlantica*, 17 (1995), pp. 19-44.

Booij, Geert, *Construction morphology*. Oxford: Oxford University Press, 2010.

Booij, Geert, Los, Bettelou & Rem, Margit, 'De oorsprong van *ghe-* als negatief-polair element in het Middelnederlands', *Taal en Tongval*, 58 (2006), pp. 3-27.

Bybee, Joan, Perkins, Revere & Pagliuca, William, *The evolution of grammar. Tense, aspect, and modality in the languages of the world.* Chicago: University of Chicago Press, 1994.

Byloo, Pieter & Nuyts, Jan, *The diachrony of Dutch* mogen, Antwerp Papers in Linguistics 113 (2011).

Byloo, Pieter & Nuyts, Jan, 'Modal auxiliaries and tense: The case of Dutch'. In: Kasia Jaszczolt & Louis de Saussure (eds.), *Time: Language, Cognition, and Reality*. Oxford: Oxford University Press, 2013, pp. 73-97.

Byloo, Pieter & Nuyts, Jan, 'Meaning change in the Dutch core modals. (Inter)subjectification in a grammatical paradigm', *Acta Linguistica Hafniensia*, 46 (2014), pp. 85-116.

Cinque, Guglielmo, *Adverbs and functional heads: A cross-linguistic perspective.* Oxford: Oxford University Press, 1999.

Conradie, Jac, 'Semantic change in modal auxiliaries as a result of speech act embedding'. In: Martin Harris & Paola Ramat (eds.), *Historical development of auxiliaries*. Berlin: Mouton de Gruyter, 1987, pp. 171-180.

Coupé, Griet, 'Modal verbs in long clusters. An innovation in Early Modern Dutch'. In: Stavroula Tsiplakou, Marilena Karyolemou & Pavlos Pavlou (eds.), *Language variation – European perspectives II.* Amsterdam: John Benjamins, 2009, pp. 59-70.

Coupé, Griet, *Syntactic extension. The historical development of Dutch verb clusters.* Leiden: LOT, 2015.

Coupé, Griet & van Kemenade, Ans, 'Grammaticalization of modals in Dutch: Uncontingent change'. In: Paola Crisma & Giuseppe Longobardi (eds.), *Historical syntax and linguistic theory.* Oxford: Oxford University Press, 2009, pp. 250-270.

Coussé, Evie, 'The grammaticalization of the *have* perfect in Dutch. A corpus study of contextual extension and semantic generalization', *Language Sciences*, 36 (2013), pp.103-112.

Coussé, Evie, 'Lexical expansion in the HAVE and BE perfect in Dutch. A constructionist prototype account', *Diachronica*, 31 (2014), pp. 159-191.

Coussé, Evie & Van de Velde, Freek, 'Hulpwerkwoordselectie in drieledige perfecta met een modaal. Een alternatieve historische verklaring'. In: Freek Van de Velde, Hans Smessaert, Frank Van Eynde & Sara Verbrugge (eds.), *Patroon en argument*. Leuven: Leuven University Press, 2014.

Croft, William, *Radical construction grammar. Syntactic theory in typological perspective*. Oxford: Oxford University Press, 2001.

Croft, William & Cruse, D. Alan, *Cognitive linguistics.* Cambridge: Cambridge University Press, 2004.

De Smet, Hendrik, Ghesquière, Lobke & Van de Velde, Freek (eds.), *On multiple source constructions in language change*. Special issue in *Studies in Language* 37 (2013).

Di Paolo, Marianna, 'Double modals as single lexical items', *American Speech*, 64 (1989), pp. 195-224.

Diewald, Gabriele, *Die Modalverben im Deutschen: Grammatikalisierung und Polyfunktionalität*. Tübingen: Niemeyer, 1999.

Diewald, Gabriele, 'Context types in grammaticalization as constructions', *Constructions* (2006), Special Volume 1.

Diewald, Gabriele 'Konstruktionen und Paradigmen', *Zeitschrift für Germanische Linguistik*, 37 (2009), pp. 445-468.

Diewald, Gabriele & Smirnova, Elena, 'Paradigmatic integration: the fourth stage in an expanded grammaticalization scenario'. In: Kristin Davidse, Tine Breban, Lieselotte Brems & Tanja Mortelmans (eds.), *Grammaticalization and Language Change. New reflections.* Amsterdam: John Benjamins, 2012, pp. 111-133.

Duinhoven, A.M., *Middelnederlandse syntaxis. Synchroon en diachroon. Deel 2. De werkwoordgroep.* Groningen: Nijhoff, 1997.

Goldberg, Adele E., *Constructions. A construction grammar approach to argument structure*. Chicago: University of Chicago Press, 1995.

Gysseling, Maurits, 'De statuten van de Gentse leprozerie van 1236', *Studia Germanica Gandensia*, 5 (1963), pp. 9-43.

Hasty, J. Daniel , 'We might should oughta take a second look at this: A syntactic re-analysis of double modals in Southern United States English', *Lingua*, 122 (2012), pp. 1716-1738.

Hilpert, Martin, *Germanic future constructions. A usage-based approach to language change.* Amsterdam: John Benjamins, 2008.

Hilpert, Martin, *Constructional change in English. Developments in allomorphy, word formation, and syntax.* Cambridge: Cambridge University Press, 2013.

Himmelmann, Nikolaus P., 'Lexicalization and grammaticization: Opposite or orthogonal?'. In: Walter Bisang, Nikolaus P. Himmelmann & Björn Wiemer (eds.), *What makes grammaticalization. A look from its components and its fringes.* Berlin: Mouton de Gruyter, 2004, pp. 21-42.

Jędrzejowski, Łukasz & van de Vate, Marleen, 'How might we want to explain them or how to deal with double modal verbs in Dutch and German?', Paper presented at *A Germanic Sandwich 2013*, 11 -12 January 2013, Leuven.

Koch, A.C.F., *Vroeg Middelnederlands ambtelijk proza. Gentse keuren van vóór 1240.* Groningen: Wolters, 1960.

Langacker, Ronald, W., *Foundations of cognitive grammar. Volume 1. Theoretical prerequisites.* Stanford: Stanford University Press, 1987.

Morwood, James, *A Latin grammar.* Oxford: Oxford University Press, 1999.

Nagle, Stephen J., 'Double modals in Early English'. In: Henk Aertsen & Robert Jeffers (eds.), *Historical linguistics 1989*, Amsterdam: John Benjamins, 1993, pp. 363-370.

Nørgård-Sørensen, Jens, Heltoft, Lars & Schøsler, Lene, *Connecting grammaticalization.* Amsterdam: John Benjamins, 2011.

Nuyts, Jan, Byloo, Pieter & Diepeveen, Janneke, 'On deontic modality, directivity, and mood: The case of Dutch *mogen* and *moeten*', *Journal of Pragmatics*, 42 (2010), pp. 16-34.

Ogura, Michiko, '*Shal* (*not*) *mowe*, or double auxiliary constructions in Middle English', *The review of English studies*, 44 (1993), pp. 539-548.

Plank, Frans ,'The modal story retold', *Studies in Language*, 8 (1984), pp. 305-364.

Reis, Marga, 'Bilden Modalverben im Deutschen eine syntaktische Klasse?'. In: Reimar Müller & Marga Reis (eds.), *Modalität und Modalverben im Deutschen*. Hamburg: Buske, 2001, pp. 287-318.

Traugott, Elizabeth C., 'On the rise of epistemic meanings in English: An example of subjectification in semantic change', *Language*, 65 (1989), pp. 31-55.

Traugott, Elizabeth C. 'Grammaticalization, constructions and the incremental development of language. Suggestions from the development of degree modifiers in English'. In: Regine Eckardt, Gerhard Jäger & Tonjes Veenstra (eds.), *Variation, selection, development. Probing the evolutionary model of language Change*. Berlin: Mouton de Gruyter, 2008, pp. 219-250.

Traugott, Elizabeth C. & Dasher, Richard, *Regularity in semantic change.* Cambridge: Cambridge University Press, 2002.

Traugott Elizabeth C. & Trousdale, Graeme, *Constructionalization and constructional changes*. Oxford: Oxford University Press, 2013.

Van de Velde, Freek, *De nominale constituent: structuur en geschiedenis.* Leuven: Universitaire Pers Leuven, 2009.

Visser, F.Th., *An historical syntax of the English language*. Leiden: Brill, 1963-1973.

Von Mengden, Ferdinand & Coussé, Evie, 'The role of change in usage-based conceptions of language'. In: Evie Coussé & Ferdinand von Mengden (eds.), *Usage-based approaches to language change*. Amsterdam: John Benjamins, 2014, pp. 1-19.

Over de auteur

Evie Coussé, University of Gothenburg. E-mail: evie.cousse@gu.se

On the resultative-modal grammaticalisation pathway of German GET verbs – with an outlook on Dutch and Afrikaans

Alexandra N. Lenz

MEM 67 (2): 177–209
DOI: 10.5117/TET2015.2.LENZ

Abstract

This contribution focuses on grammaticalisation pathways of German resultative and modal constructions using the auxiliaries *kriegen* 'to get, to receive' and *bekommen* 'to get, to receive'. In order to illustrate the complex grammaticalisation pathways of the two GET verbs and the subtypes of resultative-modal constructions in which they are used, first, historical and second, (more) contemporary data from varieties of German are analysed. Third, additional evidence is provided by means of already existing findings on Dutch and Afrikaans to complement the hypotheses derived from the analyses of German varieties.

The contribution provides answers to the following questions: Where do the resultative-modal constructions with *kriegen* and *bekommen* come from and what about their future potential? What are the main differences and parallels between the *kriegen* and the *bekommen* constructions which both use an auxiliary from the semantic network of German GET verbs? What do the synchronic data tell us about the present-day variation of the constructions, but also – on the basis of an "apparent time" hypothesis – about their diachronic grammaticalisation pathways? How do "real time" data correspond to synchronic variation and vice versa? Which parallels or differences can a language comparison reveal? To what extent do the *kriegen* and *bekommen* results mirror what we find with regard to Dutch *krijgen* ('to get, to receive') and Afrikaans *kry* ('to get, to receive')?

Keywords: grammaticalization, auxiliarization, language change, resultative constructions, modal constructions, GET verbs, language comparison, German, Dutch, Afrikaans

1 Introduction

This contribution focuses on grammaticalisation pathways of German resultative and modal constructions using the auxiliaries *kriegen* ('to get, to receive') and *bekommen* ('to get, to receive'). According to Hopper/Traugott (1993, p. 2), grammaticalisation is a type of linguistic change, 'through which a lexical item in certain uses becomes a grammatical item, or through which a grammatical item becomes more grammatical'. With regard to auxiliaries, it is important to add that grammaticalisation 'often does not so much involve single, isolated word forms but rather entire constructions of more than one word form' (Kuteva, 2001, p. 1). In order to illustrate the complex grammaticalisation pathways of the two GET verbs and the subtypes of resultative-modal constructions in which they are used, first, historical and second, (more) contemporary data from varieties of German will be analysed. Third, additional evidence will be provided by means of already existing findings from Dutch and Afrikaans to complement the hypotheses derived from the analyses of German varieties.

The paper is structured as follows: In a first step (section 2), the phenomena concerned are presented. The general overview of resultative-modal constructions with *kriegen* and *bekommen* is followed by analyses of Low and High German texts in comprehensive historical corpora from the 14th century onwards (section 3.1). In section 3.2, the historical analyses are supplemented by research on (more or less) contemporary data that represent oral and written varieties of present-day German, including dialects, intermediate varieties (so called *regiolects*) and standard varieties. The results of 'real time' and 'apparent time' data will be used to sketch the grammaticalisation pathways of resultative-modal *kriegen* and *bekommen* constructions in German (section 4). Additional evidence for the hypotheses sketched will be provided by means of findings detectable in the research literature on Dutch *krijgen* and Afrikaans *kry* (section 5).

The following research questions will guide the discussion: where do the resultative-modal constructions with *kriegen* and *bekommen* come from and what about their future potential? What are the main differences and parallels between the *kriegen* and the *bekommen* constructions which both use an auxiliary from the semantic network of German GET verbs? What do the synchronic data tell us about the present-day variation of the constructions, but also – on the basis of an 'apparent time' hypothesis – about their diachronic grammaticalisation pathways? How do 'real time' data correspond to synchronic variation and vice versa? Which parallels or differences can a language comparison reveal? How do the *kriegen* and

bekommen results parallel what we find with regard to Dutch *krijgen* ('to get, to receive') and Afrikaans *kry* ('to get, to receive')?

Parts of the analyses presented below (see section 3) build upon a previously published German-language article, viz. Lenz (2013b). To complement and extend the results presented there, this paper both discusses findings on modal constructions and gives an outlook on resultative-modal GET verbs in related languages.

2 Resultative-modal constructions with *kriegen* and *bekommen*

It is unquestionable that *kriegen* and *bekommen* are 'special' verbs in German, whose syntacto-semantic variation gives rise to a broad range of research possibilities. Over the last ten years in particular, numerous studies on the two verbs have been published (cf. Lenz, 2013c for a review of this literature). A brief look at these studies reveals that two of the various *kriegen/bekommen* constructions have been particularly common subjects of linguistic research: first, the so called *Dativpassiv* (dative passive) and second, the construction that has a resultative meaning, which, on the surface level, appears to be identical to the dative passive.[1]

(1) **German dative passive ('recipient passive')**
   a. *Alle Eltern von Neugeborenen des Jahres 2009 aus Rennerod* ***bekommen/kriegen/erhalten*** *von der Feuerwehr einen Rauchmelder* ***geschenkt***.[2]
      literal: *All parents of newborn babies* […] *get a smoke detector presented by the fire brigade.*
      'In 2009, all parents of newborns in R. received a smoke detector from the fire brigade as a gift.'
   b. *Katharina* ***kriegt/bekommt/*erhält*** *den Führerschein* ***entzogen.***
      literal: *Katharina gets the driving licence taken away.*
      'Katharina loses her driving licence.'

(2) **Participial resultative constructions ('participial RCs') with *kriegen/bekommen***
   Franzl ***kriegt/bekommt*** *das Problem* (*nicht*) ***gelöst.***
   literal: *Franzl* (*doesn't*) *get*(*s*) *the problem solved.*
   'Franzl (doesn't) manage(s) to solve the problem.'

In both the dative passive and in the 'participial RC', a GET verb (usually *kriegen* or *bekommen*) is used as an auxiliary together with the past participle of a full verb. While the subject referent in the dative passive takes the semantic role of a prototypical dative referent (usually recipient, addressee, beneficiary, or maleficiary), the subject referent of the RC is actively or causatively involved in the action. The result of the RC action is strongly intended by the subject referent and the subject referent aspires to reach the resultative state, even if the 'hindrance specialisation' of the RC implies that the action requires a certain amount of effort (cf. McIntyre, 2005, p. 402). The meaning of the participial RC can therefore very likely be paraphrased as 'accomplish/achieve the direct object referent to be affected by the verbalized action'.

Another *kriegen* construction with resultative meaning that is closely connected with the participial RC is the construction with resultative adjective phrases shown in (3). This construction refers to a 'state' of the object referent that is reached through the verbalised process. The intended condition is verbalised in (3) through the adjectival resultative attributes *trocken* and *gesund* ('dry' and 'fit/healthy', resp.), which both function as secondary predicates (alternatively *object predicate, predicative complement*) with a resultative reading (cf. Kaufmann, 1995, p. 135; Halliday, 1967; Zifonun et al., 1997). According to the Duden grammar, adjectival RCs – with regard to adjectival object predicates – have 'a strong tendency to be integrated into the predicate. The main accent then lies on the predicate. It is, however, often unclear if the construction is still a phrase or only the 'Nebenkern' of the verb, i.e. a verb particle.' (Duden-Grammatik, 2009[8], p. 858; my own translation, ANL)[3]

(3) **Adjectival RCs with *kriegen*/*bekommen***
   a. *Ich **kriege** meine Haare nicht **trocken**.* (E-Valbu: *kriegen*, type 18)
      'I can't get my hair dry.'
   b. *Wir werden dich schon wieder **gesund bekommen**.* (Duden-Universal, 2010ff.: *bekommen*)
      'We will surely get you fit/healthy again.'

A third type of German RC, which is often found in the literature and is considerably less frequent than participial and adjectival RCs, will be classified here as 'adverbial RC' (cf., for example, Duden-Grammatik, 2009[8], p. 791; E-VALBU: *kriegen*). Adverbial RCs are RCs with a directive adverbial complement (illustrated in (4)), the meaning of which can be paraphrased as 'to accomplish/achieve that something/somebody reaches a certain lo-

cation or state'. This includes the ability for the directive complement to refer to a specific concrete destination of the object movement (cf. (4a) and (4b)) or to a more abstract state that the object referent is put in. (cf. (4c)): *etwas in den Griff kriegen/bekommen*, literal: *to get sth. in the grasp* 'to get control over something'). The directive adverbial can be verbalised in the form of a directive prepositional phrase or a directive adverb (cf. (4d)). Goldberg (1995) denotes the construction type with directive adverbial as a 'caused-motion' construction, distinguishing it from adjectival RCs. We, however, follow Boas (2011), who analyses both types of constructions as RCs, since they have a parallel interpretable meaning, i.e. they refer to a final condition of a post-adverbial object (cf. Boas, 2011, p. 11, fn. 4).

(4) **Adverbial RCs with *kriegen*/*bekommen***

a. *das Klavier* ***durch die Tür kriegen****; den Ball* ***ins Tor, Netz kriegen*** (Duden-Universal, 2010ff.: *kriegen*)
'to get the piano through the door, the ball into the goal, net'

b. *Wir haben das schwere Bett nicht* ***an die Wand bekommen***. (E-Valbu: *bekommen*-II, type 15)
'We could not get the heavy bed to the wall.'

c. *eine Seuche, die wirtschaftlichen Probleme* ***in den G*[*riff*]* *bekommen/kriegen*** (Duden-Universal, 2010ff.: *Griff*)[4]
literal: *to get an epidemic / economic problems in the grasp*
'to get an epidemic or economic problems under control'

d. ***abbekommen****:* […] *3. etwas fest Haftendes, fest Aufgeschraubtes o. Ä. lösen* (Duden-Universal, 2010ff.: *abbekommen*)
'to get something that sticks, is screwed on tightly etc. to come off' / 'to loosen something that sticks, is screwed on tightly etc.'

While the dative passive (cf. (1)) can only be built with three auxiliaries from the semantic network of GET verbs (mostly *kriegen* and *bekommen*, sometimes *erhalten*; cf. Lenz, 2013a), a broad range of verbs appear in the three RC types described above (examples are provided in (5)). However, *kriegen* and *bekommen* are characterised by an above-average frequency of use in RCs and their various possibilities of combination with lexically differently filled participles, adjectives, and directive complements.

(5) **RCs with other resultatively used verbs than *kriegen* and *bekommen***

a. Adverbial RCs

a.1 *jemanden **unter den Tisch trinken*** (Duden-Universal, 2010ff.: *Tisch*)
'to drink someone under the table'

a.2 *Gwyneth Paltrow: Ich **ess** die Männer **unter den Tisch.*** (http://www.shape.de/fitness/workout/a-53833/gwyneth-paltrow-ich-ess-die-maenner-unter-den-tisch-video.html)
literal: *I eat the men under the table*
'I am able (and willing) to eat much more than men can.'

b. Adjectival RCs

b.1 *Unfallprävention: Ampeln sollen Kreisverkehre **sicherer machen**.* (http://www.der-postillon.com/2014/10/unfallpravention-ampeln-sollen.html)
literal: *Accident prevention: traffic lights should make roundabouts safer*
'Accident prevention: traffic lights are to make roundabouts safer.'

b.2 *Was soll das denn bringen, erst mit Absatzschuhen die Knochen **kaputt laufen**, und das dann mit Barfuss laufen ausgleichen?* (http://www.gutefrage.net/frage/jeden-tag-absatzschuhe-tragen-mit-barfuss-laufen-ausgleichen-sodass-die-fuesse-nicht-kaputt-gehen-)
literal: […] *to run the bones broken in high heels*
'What benefit would there be in first running in high heels until one's bones are damaged and then compensating this by running barefoot?'

c. Participial RCs

Carbon schaut ganz gut aus, an einigen Stellen könnte man den [*in Carbon-Farbe erhältlichen*] *Wagen aber etwas **verwischt machen**.* (http://www.pagenstecher.de/Car-Fakes/Work-in-Progress/t131633p1/Forfour.html)
literal: *Carbon* [*here referring to the paint colour of a car*] *looks pretty good, on some spots the car could be made a little bit blurred.*
'Carbon looks pretty good, but could do with being blurred a bit in some spots.'

Even intransitive verbs such as *laufen* (s. (5b.2)) can enter the construction scheme. It is, however, not the verb *laufen* by itself that refers to 'object movement' from a before-state (*ganz/heil sein* 'to be complete; not broken') to an after-state (*kaputt sein* 'to be broken'). Rather, 'it is more reasonable and economical if we accept the idea that [the] construction itself contributes to meaning and has its own argument structure. The surface form […] is the result of composition of verbal argument structure and con-

structional argument structure. This approach keeps simple the argument structure of a verb, and explains productivity found among similar patterns.' (Huang 2006: 20)

We suggest using a joint analysis of the various RCs with *kriegen/bekommen*, due to the following features inherent to the constructions: In all the RC examples mentioned in (2), (3) and (4), we are dealing with an agentive/causative *kriegen* or *bekommen*, and the subject of the sentence refers to the entity which either manages to get the direct object to a certain location or into a certain state, or causes these changes to come about. Seen from a syntactic perspective, the light verbs (or auxiliaries) *kriegen* and *bekommen* are connected to a directive complement (in the form of a PP or AdvP) or an adjective phrase (AdjP) or a past participle.

As we will see from the historical analyses, the RCs with *kriegen* and *bekommen* are rather young phenomena in German language history (cf. Lenz, 2012). Nevertheless they show a very successful career, which is indicated by an increasing spread over the language area and its varieties from the base dialects up to the standard varieties. If we take a look at the core area of the RCs, i.e. West Low German (see below), another feature that has, as of yet, hardly been discussed in connection to the RCs can be detected, namely participial modal constructions ('MC') with *kriegen* as illustrated in (6) (an example from the dialect dictionary of Schleswig-Holstein which is situated in the Northern part of Germany and hence in the Low German area).

(6) **Example of a participial MC with *kriegen***
du hesst woll ni ***utslapen kregen*** (SchlHWb, 1927-1935, vol. 3: 322)
literal: *you seem to not have gotten well rested* (*slept enough*)
'you seem to not have had a good night's sleep.'

The potential for the participial MC to lead to an areal-horizontal as well as to a vertical-social spread in German is illustrated by the online example in (7), in which a transitive verb (*einkaufen* 'to do shopping') occurs without realising the direct object overtly. Cases like this might provide a bridging context in grammaticalisation processes from participial RC to participial MC. (It is, as of yet, unclear, whether the participial MC can also occur with *bekommen* or not.)

(7) **Example for "*kriegen* + *eingekauft*" (literal: *to get shopped* 'it is possible (in terms of time or for other reasons) to do the shopping')**
Ach ja. Vor 12-Stunden-Tagen sieht man sich zum Frühstück, während

dessen man von den kleinen Erfolgen des Tages berichtet: wieder ein paar Vokabeln gelernt, wieder was von der großen möglichen Chance und überhaupt wieder was vom anderen gehört. Dann sitzt man schweigend da und beißt in altbackene Brote. Nur noch ein paar letzten Minuten vor dem Dienst am Kunden! Man legt man [*sic!*] *ein paar Münzen für den geplanten Einkauf zusammen. Wir müssen irgendwie noch* ***eingekauft kriegen!***
(http://www.stupidedia.org/stupi/Goodbye_Deutschland!_Die_Auswanderer)
literal: […] *We have to somehow get shopped.*
'We have to somehow get the shopping done.'

Table 1 summarises the construction types discussed above and their syntacto-semantic characteristics schematically. In all the RCs mentioned, we are dealing with an agentive/causative *kriegen* or *bekommen*. The subject of the sentence refers to an entity which either manages to get the accusative NP (the direct object) to a certain location or into a certain state, or causes these changes to come about, thus achieving (or causing to achieve) the intended movement of the object or its change of state in spite of an obstacle that has to be overcome (cf. 'hindrance specialisation', McIntyre, 2005). This hindrance reading is also present in the participial MC, in which, however, no object referent experiences a transition to a new resulting state. Instead, the subject referent puts itself in a certain state or condition, or intends to cause the changes that enable it to act.

Table 1 Phenomena in focus: *kriegen* and *bekommen* RCs and MCs (syntactic patterns and semantics)

Construction	Syntactic pattern	Meaning
Adverbial RC	*kriegen/bekommen* + $(PP/AdvP)_{+dir}$ + NP_{ACC}	'to accomplish/achieve that something/somebody gets to a certain location or reaches a certain state'
Adjectival RC	*kriegen/bekommen* + AdjP + NP_{ACC}	'to accomplish/achieve that something/somebody gets a certain attribute'
Participial RC	*kriegen/bekommen* + past part. + NP_{ACC}	'to accomplish/achieve that something is done with something/somebody'
Participial MC	*kriegen/?bekommen* + past $part._{intrans.}$	'(making an effort) to put oneself in the state to be able to do something'

3 Corpus analyses (cf. Lenz, 2013b)

The corpus analyses in Lenz (2013b) provide evidence for the nexus of the *kriegen*/*bekommen* constructions illustrated above and for the postulated scenario of grammaticalisation. In order to make up for the lack of diachronic evidence in the historical corpus analyses (cf. section 3.1), synchronic data from Present Day German will be analysed (cf. section 3.2).

3.1 Analyses of historical corpora

In this section, the question of the emergence and the dynamics of RCs with *kriegen* and *bekommen* is investigated on the basis of language-historical data from 1350 to 1925. Low German as well as High German texts are taken into account and will be systematically compared with each other. The broad range of texts analysed for the corpus analysis of Middle Low German (14th to 17th century) is listed in the bibliography at the end of the article.[5] The point of reference for New Low German is the analysis of the first two episodes of Fritz Reuter's *Läuschen und Rimels* (1853/1858). The basis for the analysis of Early Modern High German texts is the Bonner Frühneuhochdeutsch Korpus (1350-1700) as well as the Münsteraner Hexen Korpus (1565-1656). These corpora are complemented by the analysis of emigrant letters from the corpus of Elspaß (2005), which were mainly written between 1800 and 1925.

Table 2 Absolute numbers of *kriegen* and *bekommen* occurrences in the historical data (cf. Lenz 2013b)

Time frame	Language area	Corpus size (words)	Number of occurrences (absolute)			
			kriegen		*bekommen*	
			all	RCs	all	RCs
1350 - 1700	Low German	ca. 932.000	399	16	101	0
	High German	ca. 802.000	46	1	149	5
1853/1858	Low German	ca. 80.000	193	24	4	0
1800 - 1925	High German	ca. 385.000	19	1	323	3

The analysed corpus, which contains a total of over 2 million words, was first manually searched for all verbal *kriegen* and *bekommen* occurrences. The 657 *kriegen* and 577 *bekommen* occurrences detected were then classified syntacto-semantically (cf. Lenz, 2013c for further details). The analysis of the text material shows a total of 50 occurrences of RCs with *kriegen* and *bekommen*. However, the vast majority of the occurrences (namely 42) are

with *kriegen* (cf. Table 2). First evidence of the different types of *kriegen* RCs can be found during the following time frames: the first instance of an adverbial RC occurs in the second half of the 15th century (1470), the first instance of an adjectival RC is from 1598 (see (8a) and (8b)). The first participial RC in the corpus is found in 1853 in Reuter's texts, occurring at a considerably later point in time than the other constructions (s. (8c))). There are no occurrences of the participial MC in the analysed texts.

(8) **First occurrences of *kriegen* RCs in the historical corpora**

a. **Adverbial RC**
[…] *vnde he sande eme vele volckes vppe dat he mochte Jherusalem* ***in syne macht krijghen***. (LübHist, Low German, 1470)
literal: […] *and he sent him many people so he could get Jerusalem in his power*
'and he sent many people to him in order to get Jerusalem under his control.'

b. **Adjectival RC**[6]
Sagt. der hundt sei Ir nahekom[*end*] *lauffen biß An die Portz, sie wisse nitt ob der hondt hinckendt word*[*en*] *od*[*er*] *nitt, wo d*[*er*] *hondt etwas krieg*[*en*] *das muß An vnd zwisch*[*en*] [*INT*] |*d*[*er*]| *Portzen, die sie nitt woll* ***zukrieg*****[*****en*****]** *kunnen, gescheh*[*en*] *sein.*
(HexK, text 60, West Central German (Erkelenz), 1598)
literal: *says: the dog had run after her to the door, she did not know if the dog started to limp or not or where the dog had gotten his injury, it must have happened at and between the door that she could not have gotten closed*
'[…] and between the door which she could not get closed.'

c. **Participial RC**
Nu fehlt man blot, ick [...] **kreg** den Strimer nich **verköfft**. (Reut-I, Low German, 1853)
literal: *now one only misses, I could not get the cow sold*
'The only thing that's missing is, I was not able to sell the cow.'

With regard to the frequency of the subtypes of *kriegen* RCs (cf. Table 3), the dominance of constructions with a directive adverbial phrase is particularly striking. However, this only applies to the Low German language area, in which 40 of all *kriegen* RC occurrences can be found. Only two occurrences can be found in the High German language area despite the fact that the High German corpora contain 230,000 more words than the Low German texts.

Table 3 **Absolute numbers of *kriegen* RCs and MCs in both time frames and in both language areas**

kriegen	Low German		High German	
	1350-1700	1853/1858	1350-1700	1800-1925
Adverbial RC	12	19	0	1
Adjective RC	4	0	1	0
Participial RC	0	5	0	0
Participial MC	0	0	0	0

As a first interim result, the following thesis (still formulated cautiously because of the generally low number of occurrences) can be derived from the results in the historical corpora with regard to the *kriegen* RCs: First, the quantitative dominance of the adverbial RCs might indicate that this construction constitutes the beginning of the resultative grammaticalisation pathway, while adjectival and participial *kriegen* RCs embody later grammaticalisation stages. Second, an inter-regional comparison suggests that the RCs start in the Low German language area.

With regard to RCs with *bekommen*, the available data are considerably less conclusive (cf. Table 4). The *bekommen* RCs found in the analysed material amount to a mere total of eight occurrences, seven of which represent the syntactic pattern with directive adverbials, whose first occurence can be dated to 1650 (cf. (9a)). Additionally, there is a single participial *bekommen* RC in 1668 (cf. (9b)). Despite the low numbers of occurrences, the formal nature of the construction variants supports the thesis that the adverbial RC constitutes the beginning of grammaticalisation, also with regard to *bekommen*. Concerning the areal distribution of the RC, the *bekommen* data show the opposite picture to the *kriegen* data: RCs with *bekommen* only occur in the High German language area of the analysed material, but they do not occur in Low German.

Table 4 **Absolute frequency of RCs with *bekommen* in both time frames and in both language areas**

bekommen	Low German		High German	
	1350-1700	1853/1858	1350-1700	1800-1925
Adverbial RC	0	0	4	3
Adjectival RC	0	0	0	0
Participial RC	0	0	1	0

(9) **First occurrences of *bekommen* RCs in the historical material**

a. **Adverbial RC**

[…] *vnd in allem diesem tieffen Nachsinnen, fallen sie in verdeckte Gruben,* ***darauß*** *man sie mit grosser mühe kaum widerum kan* ***bekommen.*** (BFNhD, text 237, Alsatian, 1650)

literal: […] *they fall into hidden caverns, out of which they can only be got with the greatest effort*

'[…] out of which it is hard to rescue them.'

b. **Participial RC**

[…] *daselbst hat er des Bischoffs Schwester-Sohn* […] ***gefangen bekommen.*** (BFNhD, text 137, East Franconian, 1668)

literal: *there he got the son of the bishop's sister captured*

'Over there, he managed to have the son of the bishop's sister arrested.'

3.2 Analyses of (more) contemporary data

As we are dealing with quite recent phenomena, the historical analyses only offer limited insight into the hypotheses formulated in this paper. In order to solve these problems, or at least to optimise the interpretation base, recent data are analysed in Lenz (2013b). As these analyses will reveal, the various subtypes of *kriegen/bekommen* RCs and MCs show very different distributions in different regional subareas as well as in different varieties along the 'vertical' axis of the dialect-standard repertoire (i.e. dialects, regiolects, regional standards; cf. Lenz, 2010). The dialect analyses are mainly based on 650 dialect recordings of the so called Zwirner corpus. They include eight different dialect regions covering the three main dialect groups of the German language area, i.e. Low, Central and Upper German (from north to south).[7] The results of the dialect analyses are summarised in Table 4 and Figure 1.

While RCs with *bekommen* do not occur in the analysed dialect recordings of the Zwirner corpus, 189 occurrences of *kriegen* RCs can be found, unevenly regionally distributed (cf. Table 5). The highest frequencies are reached by the RCs in (West Low German) 'Holsteinish' and in (West Central German) 'Central Franconian'. Here, they amount to 13% of all *kriegen* occurrences. In terms of the frequency of *kriegen* RCs, these regions are followed by East Central German (cf. Upper Saxonian and Silesian recordings) and East Low German (East Pommeranian), while the Upper German data (in the Alemannic and Bavarian areas) include distinctly less *kriegen* RCs. Also, the formal nature of the occurrences reveals interesting information (cf. Table 5): The adverbial *kriegen* RCs constitute the biggest

part of the RCs found and they occur in all six investigated dialect areas. In the frequency hierarchy they are followed by the adjectival RCs that are realised in the Low and Central German recordings, while those constructions are missing in Upper German. Participial RCs only occur in dialect recordings of Low German and West Central German.

Table 5 ***Kriegen*** **occurences in the Zwirner corpus and frequencies of RCs in all *kriegen* occurrences**

Dialect area	Corpus size (number of recordings)	*kriegen* occurrences		
		all	RCs	
			absolute	%
Holsteinish	100	488	68	**13,9**
East Pommeranian	106	356	25	**7,0**
Central Franconian	119	504	66	**13,1**
Silesian + Upper Saxon	60 + 22	176 + 43	20	**9,1**
Low + High Alemannic	69 + 26	105 + 53	6	**3,8**
Central Bavarian	148	310	4	**1,3**

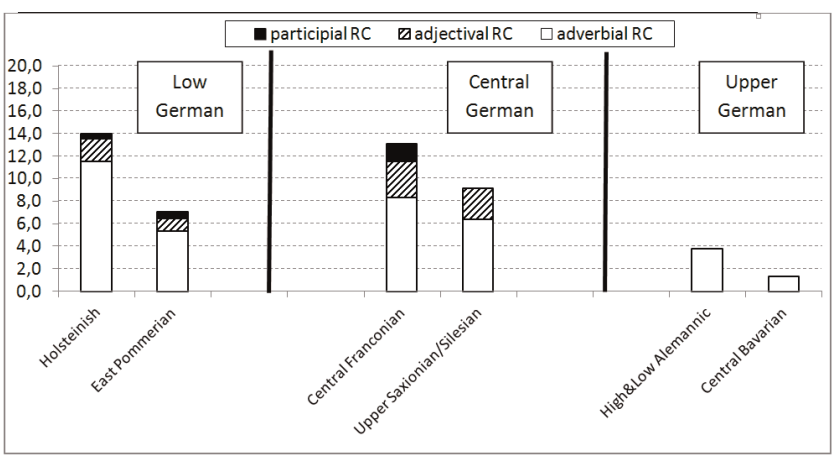


Figure 1 Frequencies (in %) of adverbial, adjectival and participial RCs in all kriegen *occurrences in the Zwirner corpus*

Just like participial MCs were absent from the historical corpus texts, there are no occurrences of participial MCs in the dialect recordings of the Zwirner corpus either (which itself is already 60 years old). However, traces of this construction, at least for Schleswig-Holstein and therefore for the dia-

lects of West Low German, can be detected in dialect dictionaries, as illustrated in the examples in (10).

(10) **Participial *kriegen* MCs in Low German dialect dictionaries**
   a. *du hesst woll ni* ***utslapen kregen*** (SchlHWb, 1927-1935, vol. 3: 322)
      literal: *you seem to not have gotten well rested/slept*
      'You seem to not have had a good night's rest.'
   b. *ehr he* ***kriggt snuuft*** *und* ***będen,*** *is de Kark ut* (SchlHWb, 1927-1935, vol. 3: 322)
      literal: *before he has gotten breathed out and prayed, church is over*
      'Before he has had time to breath out and pray, church is over.'

In addition, regiolects (i.e. 'intermediate' (colloquial) varieties 'between' the dialects and the standard language (cf. Lenz, 2010)) were analysed in Lenz (2013b) using the so-called Pfeffer corpus, as its 398 regiolect recordings provide an optimal basis for comparison with the dialect recordings of the Zwirner corpus.[8] The results from the 'colloquial' Pfeffer recordings are summarised in Table 6 and Figure 2. Although *bekommen* is generally more frequently realised than *kriegen*, the Pfeffer recordings include more *kriegen* RCs than *bekommen* RCs. The *kriegen* RCs occur 35 times, which is 12% of all *kriegen* occurrences. RCs with *bekommen* are not represented in the dialectal Zwirner corpus; they do, however, occur in the regiolectal Pfeffer recordings (24 times, which is 4.1 % of all *bekommen* occurrences). A comparison of the three different linguistic areas (Low, Central and Upper German) shows the well-known decline in occurrence from north to south for both resultative verbs *kriegen* and *bekommen*: RCs have the highest frequency in the Pfeffer recordings from the Low German language area and the lowest frequency in Upper German. In addition, the Pfeffer data reflect the previously observed frequency hierarchy with regard to the subtypes of the constructions: It is the adverbial RC who reaches the highest number of occurrences in all three areas with both resultative verbs. RCs with adverbials are followed by adjectival RCs – which occur at least sporadically –, while participial RCs are not represented in the Pfeffer recordings at all.

It would be premature, however, to assume that participial RCs do not occur in "intermediate" varieties on the vertical dialect-standard axis. This becomes evident when additional regiolectal (and more current) data, such as those in the "AdA" maps in Figures 3 and 4, are considered. The maps illustrate that the phraseme *etwas gebacken kriegen/bekommen* (literal: *to get sth. baked*), which builds on the resultative reading of 'to master

something; to be able to do something', is classified as 'completely uncommon' in the Alltagssprache ('vernacular') by informants in the mapped regions in Austria and Switzerland – independently of the auxiliary *kriegen* or *bekommen*. In Germany, however, the assessment of the phraseme depends to a high degree on the specific auxiliary: while *gebacken kriegen* is classified as occurring 'every now and then' by only few informants in the Bavarian and Franconian language area (in the South of Germany), the *kriegen* stimulus seems to be (very) common in the rest of Germany. The stimulus *gebacken bekommen*, however, is only classified as being vernacular 'every now and then' – also by informants in the middle and the north of Germany.

Table 6 ***Kriegen/bekommen* occurrences in the Pfeffer corpus and frequencies of RCs in all *kriegen/bekommen* occurrences**

Dialect area	Corpus size (number of recordings)	*kriegen* occurrences			*bekommen* ocurrences		
		all	RCs		all	RCs	
			absolute	%		absolute	%
Low German	151	117	22	**18,8**	219	13	**5,9**
Central German	131	64	6	**9,4**	223	9	**4,0**
Upper German	137	111	7	**6,3**	148	2	**1,4**

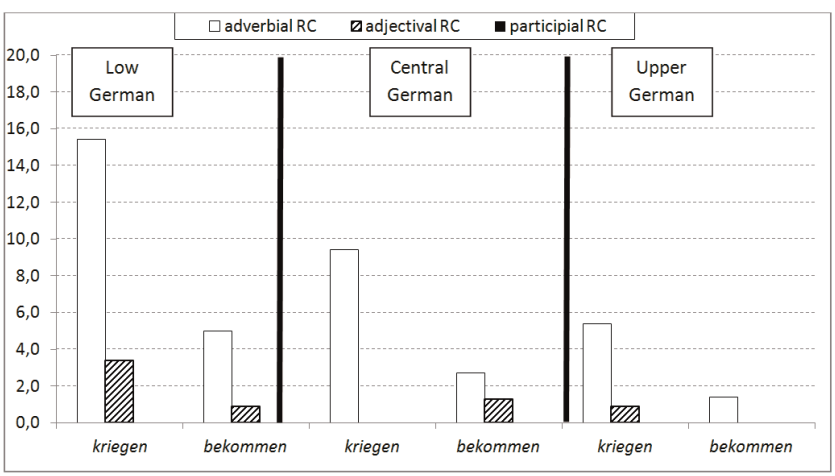


Figure 2 Frequencies (in %) of adverbial, adjectival and participial RCs in all kriegen and bekommen *occurrences in the Pfeffer corpus*

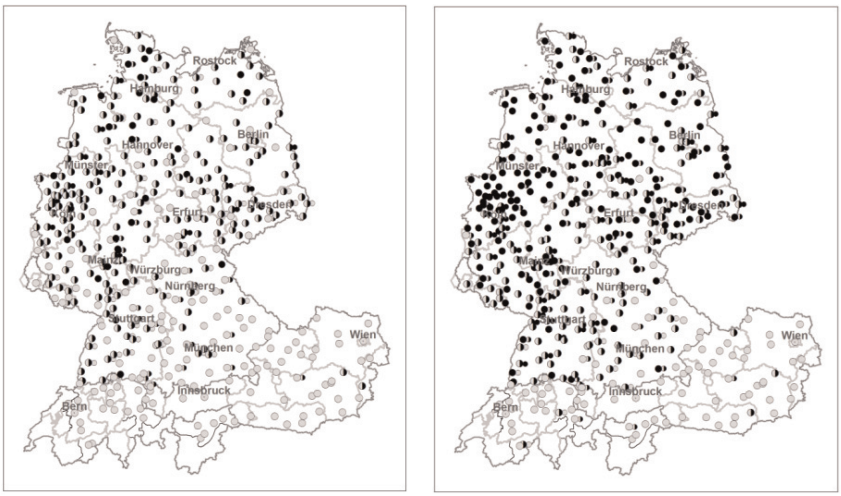


Figure 3 AdA maps gebacken kriegen (*3a*) *and* gebacken bekommen (*3b*) (*literal:* to get sth. baked *'to manage/master sth.'*)

[re-drawn in black and white: black symbols = 'very common', black-grey symbols = 'occurs every now and then', grey symbols 'completely uncommon'] (<http://www.atlas-alltagssprache.de/runde-4/f21a-b/>)

In addition to the (more or less) oral speech data taken from dialects and regiolects, data from written corpora were also taken into consideration. The first source of written data is provided by a corpus of urban chat forums and the second by journals.

The impression that the regiolectal Pfeffer corpus generally contains linguistic material that is relatively close to the standard varieties (despite the initially intended 'vernacular' character; cf. also Lenz, 2007), is supported by direct comparison with a corpus that is conceptually oral but medially written and which is comprised of recordings of Internet Relay Chats (IRC). The chat corpus contains a total of 1.7 million words from 845 chat sessions that were collected from various North, Central, and South German urban chat forums in 2006. The corpus is a highly valuable resource for research into RCs. This becomes immediately evident when taking a look at the frequency of the relevant variants (cf. Table 7): there are 73 occurrences of *kriegen* RCs and 112 occurrences of *bekommen* RCs, which makes up 19.1% and 13.3% of all *kriegen* and *bekommen* occurrences, respectively. Although *bekommen* is the overall most frequent verb of the two in the chat data, the relative number of RCs per verb is higher for *kriegen* than for *bekommen*. This means the probability to be realised as an RC is higher for *kriegen* occurences than for *becommen.*

Furthermore, a closer look at the formal nature of the occurrences in

the chat corpus (Figure 4) shows that the participial RCs occur in all three areas with both resultative verbs (cf. e.g. (11)). This resultative subtype is, however, still ranked third on the frequency hierarchy, while the adverbial RCs represent the dominant construction type in the chat data as well. The adverbial RCs are followed by the adjectival RCs. Even though chat data come with their own set of corpus-linguistic problems, for example with regard to the correct regional localisation of the data (cf. Lenz, 2007, 2013c), the grouped bar diagram in Figure 4 for the chat data shows similar inter-regional differences as observed in the data analysed above. In chats from northern city channels, RCs occur relatively most frequently – i.e., in relation to all *kriegen* and *bekommen* variants.

Table 7 ***Kriegen/bekommen*** **occurences in the chat corpus and frequencies of RCs in all *kriegen/bekommen* occurrences**

City chats	Corpus size (words)	*kriegen* occurrences			*bekommen* occurrences		
		all	RCs		all	RCs	
			absolute	%		absolute	%
Northern Germany	560.841	109	26	**23,9**	308	50	**16,2**
Central Germany	730.570	176	34	**19,3**	333	41	**12,3**
Southern Germany	425.799	98	13	**13,3**	202	21	**10,4**

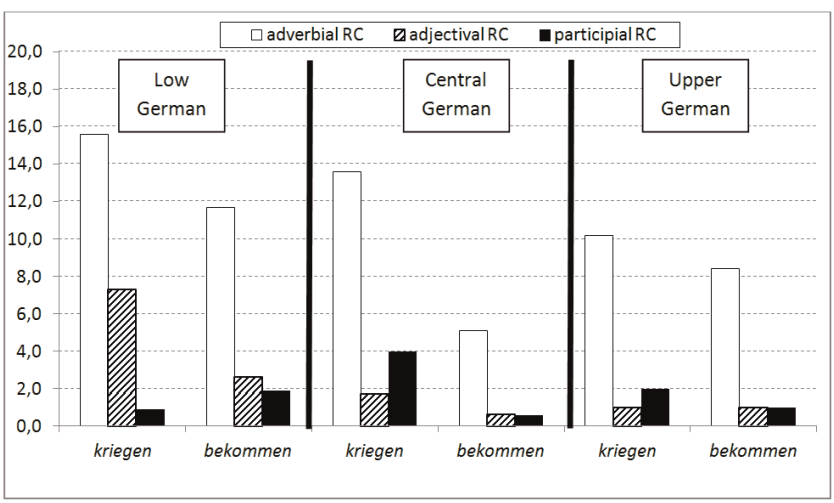


Figure 4 Frequencies (in %) of adverbial, adjectival and participial RCs in all kriegen *and* bekommen *occurrences in the chat corpus*

(11) **Chat example of participial *kriegen* RC** (bold print A.N.L.)
IRC channel #muenchen, 21th February 2006
[21:37] * _dreamer has joined #muenchen
[…]
[21:37] <_dreamer> *bah bah.*
[21:37] <_dreamer> *ich* ***krieg*** *dieses dreckssc*annerding *nicht* ***installiert****. bah.*
'I can't **get** this shitscannerthing **installed**. ugh.'

The chat data were complemented by findings from five newspapers included in the German Reference Corpus (*Deutsches Referenzkopus*, DeReKo) of the Institute for German Language (IDS), which represent a register of written language closer to the standard language than the chat data. The five newspapers (each of them were analysed for the year 2009) are the (North German) *Hannoversche Allgemeine*, the (West Central German) *Rhein-Zeitung*, the (South German) *Nürnberger Nachrichten*, the (Swiss) *St. Galler Tagblatt*, and the (Austrian) *Niederösterreichische Nachrichten*. As anticipated, the verb *bekommen* is dominant in the newspaper data, while *kriegen* constitutes in-between 3.6 and 7.9 per cent of all occurrences only (cf. Table 8).[9]

Concerning the formal nature of the *kriegen* and *bekommen* variants and in particular of the RCs, Figure 5 shows the frequencies of the RCs of all *kriegen* and *bekommen* occurrences out of the first 300 occurrences of each verb in 2009. The analysis shows that approximately every third *kriegen* variant is an RC in the newspaper texts. This percentage is considerably lower for *bekommen*. Put differently, if a *kriegen* occurrence is found in a newspaper text, the probability that it is realised as an RC is 30%. In the case of *bekommen*, this probability lies at only 4 to 10%.

Table 8 Absolute and relative numbers of all *kriegen* and *bekommen* occurrences in the newspaper corpus (year: 2009) in comparison

Newspapers (year 2009)	Absolute numbers		Relative numbers (%)		
	kriegen	*bekommen*	*kriegen*	*bekommen*	Σ
Hannoversche Allgemeine	450	6.866	6,2	93,8	100
Rhein-Zeitung	1.189	18.028	6,2	93,8	
Nürnberger Nachrichten	470	5.454	7,9	92,1	
St. Galler Tagblatt	379	5.849	6,1	93,9	
Niederösterr. Nachrichten	460	12.328	3,6	96,4	

The adverbial RC is clearly dominant among the RC types in German newspapers in Germany, Austria and Switzerland; it occurs the most fre-

quently in all regional/national subcorpora for both resultative verbs. The reason for the high frequency of adverbial RCs includes the prefix verb *hinkriegen* ('to master/accomplish') as well as highly frequent phrasemes such as *etwas in den Griff bekommen/kriegen* (literal: *to get sth. in the grip* 'to get a grip on sth.') and *etwas auf die Reihe bekommen/kriegen* (literal: *to get sth. on the line* 'to handle/master sth.'). The adverbial RC is followed by the adjectival RC, which is considerably less frequent but is still represented by both resultative verbs in all five subcorpora. The participial RCs can be found with the auxiliary *kriegen* in four subcorpora (they do not occur within the first 300 *kriegen* instances in the Austrian *Niederösterreichische Nachrichten*). The only occurrence of a participial RC with *bekommen* auxiliary is found in the *Hannoversche Allgemeine* (North Germany) in the form of the phraseme *etwas gebacken bekommen* (literal: *to get baked sth.* 'to master/manage sth.'). Examples from the newspaper corpora are listed in (12).

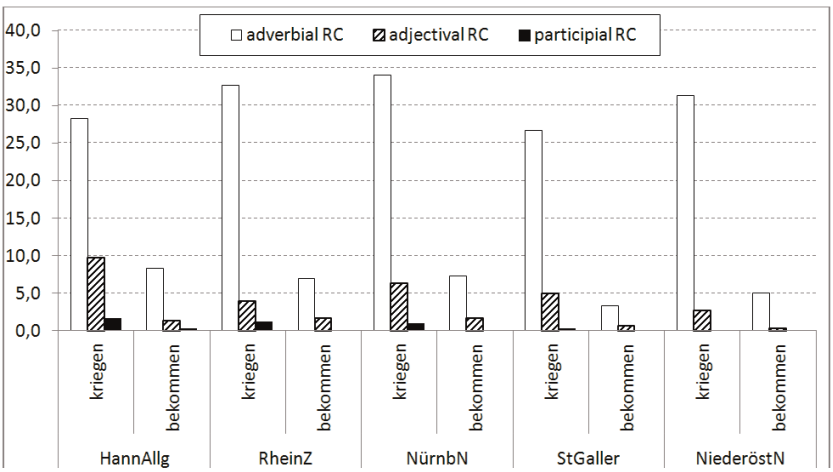


Figure 5 Frequencies (in %) of RCs in the first 300 occurrences per verb of every newspaper in the newspaper corpus (year 2009)

(12) **Newspaper occurrences of *kriegen/bekommen* RCs (examples)**

a. **Adverbial *bekommen* RC**

Am Anfang konnte Fredi Niederer noch einige Rennen gewinnen. Dann ***bekamen*** *Jürgen Stiefken und Michael Sorger ihre Fahrzeuge immer besser* ***in den Griff****.*

(*St. Galler Tagblatt*, 08/01/2009)

literal: *In the beginning Fredi Niederer could still win a few races.*

AUP – 156 x 234 – 3B2-APP flow
<TET1502_03_LENZ_1Kv36_proef1 • 25-03-16 • 12:40>
Pag. 0195

Then Jürgen Stiefken and Michael Sorger got a better and better grip on their vehicles.
'[…] Then, J.S. and M.S. started to get their vehicles under better control.'

b. **Adverbial *kriegen* RC**
*'Damit ist aber noch nicht das Projekt Nikimax gescheitert', betonte Thomas Beushausen, Direktor des Kinderkrankenhauses auf der Bult. 'Gescheitert ist nur, dass wir in der von uns gedachten Zeit eine gemeinsame Kinderklinik bauen. Das **kriegen** wir zum jetzigen Zeitpunkt nicht **in Gang**.*
(*Hannoversche Allgemeine*, 10/01/2009)
literal: […] *We cannot get it into gear at this point in time.*
'"The project Nikimax has not failed yet," Thomas Beushausen, head of the children's hospital in Bult emphasised. "The only thing that has failed is the plan to build a joint children's hospital in the estimated time. We can't get this started right now."'

c. **Adjectival *kriegen* RC**
*Der technischen Nanoisierung ging jedoch die sprachliche voraus. Vor allem eine ganz bestimmte Silbe hat sich damit einen Namen gemacht, selbst die dicksten Sprachbrocken **kleinzukriegen**. Der Eindampfer unter den Nomensuffixen heißt* -chen. (*Hannoversche Allgemeine*, 12/09/2009)
literal: […] *Above all, a very specific syllable has made a name for itself by getting even the thickest language chunks small.* […]
'The technical nanoisation was, however, preceded by the linguistic one. Above all, one particular syllable became known to be able to shrink even the thickest language chunks. This vaporiser among noun suffixes is known as *-chen*.'

d. **Participial *kriegen* RC**
*Gerade zur Osterzeit, wenn die katholischen Erstkommunionkinder zum ersten Mal zur Beichte gehen, haben auch viele Erwachsene den Wunsch, das mit dem Himmel irgendwie wieder **hingebogen** zu **kriegen**. Ein erster Schritt kann da das Seelsorgegespräch sein.* (*Rhein-Zeitung*, 08/04/2009)
literal: *Especially around the time of Easter, when the catholic First Communion children go to confession for the first time, many adults also have the wish to get "this* [*thing*] *with Heaven" somehow sorted out again.* […]
'Especially around the time of Easter, when Catholic children go to confession for the first time during First Communion, many

adults have the desire to somehow work out that "thing with Heaven" again, too.'

e. **Participial *bekommen* RC**
*Der 18-Jährige ist stolz, mithalten zu können, und er ist überzeugt: Egal, was auf ihn zukommt in der Ausbildung, er wird es **gebacken bekommen**.* (*Hannoversche Allgemeine*, 02/01/2009)
literal: […] *No matter what comes at him in his education, he will get it baked.*
'The 18-year-old is proud to be able to keep up. He is convinced that no matter what challenges arise during his education – he will manage to meet them.'

4 Synopsis: historical and contemporary findings

This section provides a synopsis of the historical and (more) contemporary synchronic findings presented above. The latter were built on the basis of comprehensive areal-linguistic data that represent diverse varieties/registers positioned along various horizontal-vertical dimensions of the complex areal-linguistic variation in German.

The earliest occurrences of RCs for both *kriegen* and *bekommen* are adverbial RCs in the form of directive prepositional or adverbial phrases that point in the direction of the control domain of the subject referent (for example *etwas/jemanden in seine Gewalt kriegen/bekommen* ('to get someone/something hostage/in prison etc.'), *etwas/jemanden in seine Macht kriegen/bekommen* ('to get power over something/somebody'), *etwas/jemanden in sein Lager kriegen/bekommen* ('to get something/somebody to come into one's own camp'). Thus, the intended result of the verbalised action is the movement of an object. This movement is characterised by the following two properties: first, agentive-causative movement is induced by the subject referent (or supposed to be induced by the subject referent); second, the end location of the object movement is the subject referent itself and his/her immediate control domain (viz. their hand, fingers, grasp, regime etc.).

A transitive *kriegen* or *bekommen* clause with an active-transferential reading ('to obtain/to get hold of the object referent (with effort)') in combination with an optional directive phrase can be seen as the source for RCs with directive adverbial phrases. The optional specification provides additional information on 'where exactly' the subject referent 'gets' the object referent (e.g. in his hand/fingers etc.). The semantic pathways towards becoming an adverbial RC, in which the directive phrase is no longer

optional but has the function of a directive adverbial complement (C_{adv}), are illustrated in (13). A crucial component in the semantic development illustrated is the cognitive-semantic closeness of the concepts HAVE and BE,[10] which also become apparent in the succession of the conversational implicatures in (13).

(13) **Emergence of the adverbial *kriegen*/*bekommen* RC**

kriegen/bekommen + NP_{ACC} $[(+ PP/AdvP)_{+dir}]_{facultative}$ =
'to get referent NP_{ACC} (in one's own possession) with effort'
>
'to achieve (with effort) that referent NP_{ACC} is moving into the control domain of the subject referent'
>
'to achieve (with effort) that referent NP_{ACC} moves in the direction of a GOAL' =
kriegen/bekommen + $[(PP/AdvP)_{+dir}]_{Cadv}$ + NP_{ACC}

The fact that RCs with directive adverbials pointing to the immediate control domain of the subject referent developed first (e. g. *etwas in den Griff/die Finger kriegen/bekommen,* literal: *to get sth. in the grasp/fingers* 'to get sth. under control') is supported by the results of all of the regional-linguistic analyses: this construction type is clearly dominant for both resultative verbs. In the next step of grammaticalisation, the GOAL can also refer to a location outside of the direct control domain (e. g. *das Klavier durch die Tür kriegen/bekommen* 'to get the piano through the door'), to a less concrete objective and even to an abstract state of being (e.g. *etwas wieder in Ordnung kriegen/bekommen,* literal: *to get sth. in order* 'to clear up a matter'). The de-concretisation of the intended objective is a crucial semantic bridge. From a syntactic point of view, this bridge paves the way for (1) RCs with prepositional phrases to become (2) RCs with adverbs and then to become (3) RCs with adjectives and finally with past participles. The suggested chronology of grammaticalisation steps is supported by both the historical corpora and the more contemporary data. In every case, the hierarchy frequency goes from adverbial RCs (most frequent) to adjectival RCs to participial RCs (least frequent).

With regard to the temporal origin of the resultative grammaticalisation pathways of both verbs, the historical data point to the 'pioneering role' of *kriegen*. The earliest occurrences of *kriegen* in RCs occur as early as the second half of the 15th century, while parallel constructions with *bekom-*

men were not detected until the 17th century. Even though the participial RC with *kriegen* does not occur until the mid-19th century, several characteristics suggest a considerably earlier creation of this construction type. First, the earliest occurrences of adjectival *kriegen* RCs can be found at the end of the 16th century (there are no adjectival occurrences for *bekommen* in the entire historical material). These adjectival *kriegen* RCs are syntacto-semantically related to the participial RCs. Second, the parallel participial RC with the *bekommen* auxiliary occurs as early as the 17th century, even though resultative *bekommen* shows a slower grammaticalisation process than resultative *kriegen*.

The significance of the regional-linguistic material for the analysis of the construction is further highlighted by the fact that RCs with adjective and participial phrases are very rare in the historical material (i.e. they do not show up until the 20th century). Consequently, the analyses of synchronic variation of the more contemporary data not only support the hypotheses on the chronology of the progression, they also provide evidence for the inter-regional differences between the grammaticalisation pathways. The first occurrences of RCs can be found in the Low German area; and RCs are still most frequent in the dialect and regiolect data for Low German as well. In terms of frequency, the Low German area is followed by the Central German area, with the fewest occurrences to be found in Upper German. With regard to the 'vertical' distribution of the *kriegen* RCs on the dialect-standard axis, its status is closely connected to changing attitudes towards the general use of the verb *kriegen* since the 18th century (cf. Lenz, 2013c).[11] Still, the contemporary data support the thesis that the RCs are particularly 'friendly' contexts in the sense that even in higher registers and written language data a particularly high number of *kriegen* RCs can be found. Furthermore, such examples frequently occur without being embedded in direct or indirect speech contexts and are thus not explicitly marked in written contexts (e. g. as 'vernacular' or 'oral').

Concerning *bekommen* RCs, the earliest occurrences in history suggest the High German language area as point of origin. In general, however, the *bekommen* RCs in High German occur significantly later than their *kriegen* counterparts in Low German. Not only does *bekommen* start its resultative grammaticalisation pathway at a later time, it also moves slower and in a less consequent manner than *kriegen*. Furthermore, when *bekommen* occurs in the contemporary data, it is less frequently realised as an RC than *kriegen* in all regions and varieties on the vertical dialect-standard axis. While no *bekommen* RCs can be found in the dialect data at all, they do appear in 'higher' registers. In these registers (i.e. regiolectal and written

language data), they also show the same decline from north to south as *kriegen*, with the highest frequencies and the strongest degrees of grammaticalisation in the north.

5 Evidence from language comparison: Dutch and Afrikaans

Evidence for the postulated chronology of events (adverbial RC > adjectival RC > participial RC) can not only be derived from analyses of variation in German, but also from comparative analyses of German with Dutch and Afrikaans. At least this is the case for the verb *kriegen*, which has an equivalent in both languages.

With regard to the Dutch *krijgen* RCs, we can draw on the language-historical analysis of Landsbergen (2006). According to his corpus analyses, the origin of the participial RC ('new *krijgen*' in terms of Van der Horst, 2002) can be found in adverbial RCs, which he traces back to the 14th century.

(14) **Occurrences of adverbial *krijgen* RCs (according to Landsbergen 2006: 156)**

a. *Op dat ic Gelloene mach enechsijns* ***te campe gecrigen****.*
(1340 – Roman der Lorreinen)
literal: On that I in some way I can get G. at a fight.
'So that in some way I can get Gelloene to fight me.'

b. *Mocht sine* ***in haren arm ghecrighen****, hine souts haer niet swighen.*
(1350 – Borchgravinne van Vergi)
'Should she get him in her arms, he would not conceal anything from her.'

The examples in (14) show adverbial RCs whose PP complements include concrete locations as the goal of the movement of the object (*fight* and *arms*). PP complements with abstract locations, which increasingly appear in the corpus (over time), represent the transition to adjectival complements decribing states. Landsbergen (2006) finds evidence of adjectival RCs for Dutch *krijgen* in his corpus for the first time in 1569 (cf. 15).

(15) **First occurrence of adjectival *krijgen* RCs (according to Landsbergen, 2006, p. 156)**
Ende so daer de eene Sluetel niet op en paste, sy mochte eene andere versoecken […] *dat sy het Slot* ***op crege****.*
(1569 – Marnix, Byenc)
'And because the key did not fit, she tried another one, so that she got the lock open.'

Landsbergen (2006) postulates that the bridging of the gap between adjectival and lastly participial *krijgen* RCs can be found in ambiguous examples such as the one in (16): "[T]heir meaning can be described both as 'to get a person/object in a V-ed state' and 'to V a person/object'. In the latter interpretation, the main activity described in the examples is reduced to that of the complement, which is a first indication of a reanalysis of *krijgen* from main verb to auxiliary." (Landsbergen, 2006, p. 156)

(16) **Occurrence of *krijgen* + adjectival part./past part. (according to Landsbergen, 2006, p. 156)**
Desen dach waren eenige ruyteren uyt Heusden naer des viants leger gereden ... ende rencontreerden eenige voeragiers, daeraf sij eenigen ***gevangen cregen****.*
(1600 – Duyck, Journ)
literal: On this day, some horsemen had ridden from Heusden towards the enemy's army, and encountered some freebooters, of which they got some caught.
'[…] of which they took some prisoner.'

If we compare Landsbergen's results (2006) with our *kriegen* analyses, strong parallels between Dutch *krijgen* and German *kriegen* become evident, pointing to a parallel succession of their grammaticalisation process from (1) adverbial RC to (2) adjectival RC and (3) participial RC.[12] A comparison in terms of first occurrences indicates that the grammaticalisation steps occur slightly earlier in Dutch, as can be seen in Landsbergen (2006).

The resultative grammaticalisation pathway of *kriegen* does not end with the participial RC. In some Low German dialects, it has already developed into a modal construction (MC) that can occur not only without an accusative nominal phrase, but also with intransitive verbs. Examples can be found in the dictionary of Schleswig Holstein, see example (6) above: *du hesst woll ni utslapen kregen* (SchlHWb, 1927-1935, vol. 3: 322) (literal: *you seem to not have gotten well rested* 'you seem to not have had a good night's

sleep'). As the corpus analyses above have shown, only few traces of the participial *kriegen* MC could be found in German. But its features concur with the features of its counterparts in Afrikaans. They are described by Molnárfi (1997, p. 20) as follows:

a. In contrast to the participial RC, the participial MC is not causative (see, for example for Afrikaans **Ek kry reg dat iemand anders slaap.* 'I achieve that someone else sleeps').
b. The embedded past participle is no longer transitive and attribution is no longer possible (see, for example, Afrikaans **die geslaapte kind* 'the slept children').
c. The theta and casus assignment inside the construction is solely determined by the embedded participle (Afrikaans *kry* as well as *kriegen* lose their transitivity).
d. The event-structural description changes its direction to a durative one-phase structure.

Above all, according to Molnárfi (1997), c. and d. provide clear evidence for the syntactic change of Afrikaans *kry* from transitivity to intransitivity as well as for changes in the aspectual structure of the construction (transition from a terminative small clause to a durative auxiliary). Similar to Afrikaans *kry* in *ek kry geslaap* ('I get to sleep'), *kriegen* in *ich kriege geschlafen* has been auxiliarised to a modal auxiliary.

The comparison of German, Dutch und Afrikaans shown in Figure 6 provides evidence for the parallel pathways of grammaticalization of the RCs and MCs of German *kriegen*, Dutch *krijgen* und Afrikaans *kry*.

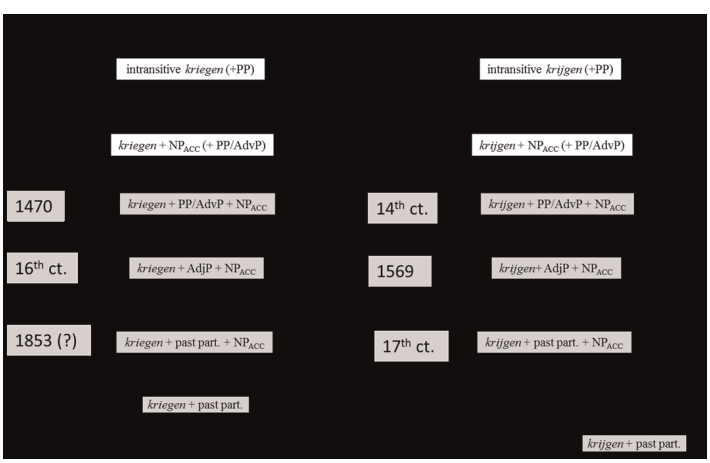


Figure 6 Resultative-modal grammaticalization pathways of German kriegen*, Dutch* krijgen *and Afrikaans* kry (*and their constructions*)

6 Summary

The focus of this contribution was on the grammaticalisation pathways of German resultative and modal constructions using the auxiliaries *kriegen* ('to get, to receive') and *bekommen* ('to get, to receive'). Both pathways, which have been simplified and summarised in Figure 7, have different origins. This is due to the different linguistic histories of the two verbs (cf. in more detail Lenz, 2013c). While *bekommen* was originally an intransitive movement verb with COME semantics, the intransitive *kriegen* has its origin in semantic variants with meanings relating to FIGHT. Both intransitives brought forth transitive variants, which even now carry the meaning 'to achieve through effort' (among other meanings) and which are the source of resultative auxiliaries.

Although they have different points of departure, the resultative-modal constructions of *kriegen* and *bekommen* are apparently the result of parallel pathways of grammaticalization: out of one transitive variant with the semantics 'to get by effort', there first crystallized adverbial resultative constructions (RCs), which were further grammaticalized to adjectival RCs and then to participial RCs. At least for *kriegen*, the resultative grammaticalisation pathway does not end with a participial RC. In some dialects of the Low German area, there are (rarely attested) traces of a modal construction (MC) – with CAN semantics – that can occur without an accusative nominal phrase and even with intransitive verbs.

In order to illustrate the complex grammaticalisation pathways of the two GET verbs and the subtypes of resultative-modal constructions in which they are used, first, historical and second, (more) contemporary data from varieties of German were analysed. Third, additional evidence was provided by means of already existing findings from Dutch and Afrikaans to complement the hypotheses derived from the analyses of German varieties.

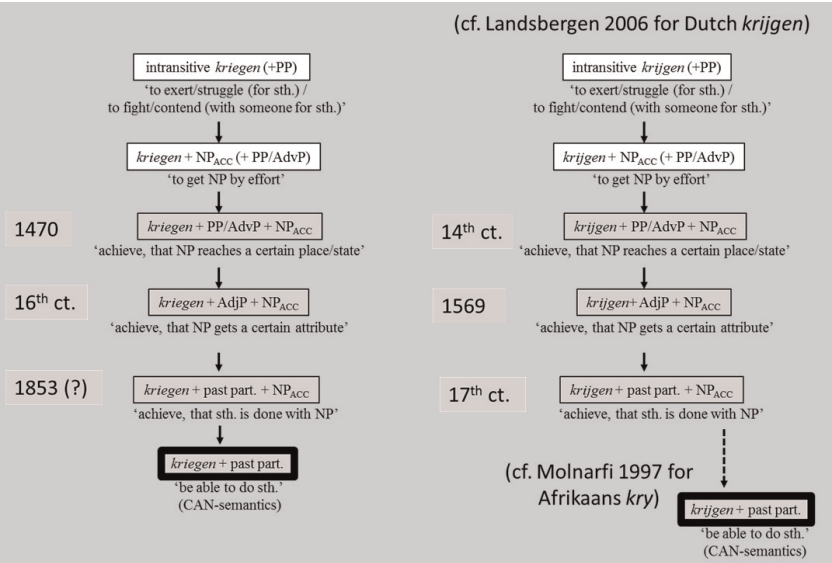


Figure 7 Resultative-modal grammaticalisation pathways of kriegen/bekommen (*and their constructions*)*; first occurrences provided in grey boxes*

Notes

1. For literature on the German dative passive, see, e.g., Leirbukt (1997), Askedal (2005), Abraham (1985), Molnárfi (1998), Glaser (2005), Lenz (2007, 2009, 2011, 2012, 2013a); for literature on resultative constructions in German in general and on the *kriegen* type specifically, see, e.g., Duden-Grammatik (20098, pp. 790ff.), Abraham (1985), Welke (2011, pp. 223-231), Boas (2003, 2011), Chang (2003), Lenz (2013b,c). For Dutch RCs with *krijgen*, see, e. g. Hoekstra (1984) and Van der Horst (2002). On Afrikaans *kry*, see Molnárfi (1995). For English, see Goldberg (1995), Molnárfi (1997), Goldberg and Jackendoff (2004), Boas (2003), Ettlinger (2005).
2. The *erhalten* example is a slightly adapted example from the *Rhein-Zeitung* (found on 03.01.2009; emphasis A.N.L.).
3. Duden-Grammatik (2009, p. 858): '… eine starke Tendenz zur Integration ins Prädikat; die Hauptbetonung liegt dann auf dem Prädikativ. Dabei ist oft unklar, ob überhaupt noch eine Phrase vorliegt oder nur noch ein Nebenkern des Verbs, also eine Verbpartikel'.
4. Duden-Zitate (20072, p. 381): 'Eine Sache im Griff haben bedeutet "das Betreffende [manuell] beherrschen, unter Kontrolle haben"'. ('*eine Sache im Griff haben* means to have (manual) control over a thing' [own translation; A.N.L.])
5. Cf. also the extensive overview in Lenz (2013c).
6. For evidence of the adjectival character of *zu* 'closed' in the Ripuarian locality Erkelenz, cf. RheinWb (1928-1971, vol. 9: p. 837).
7. The corpus, the audio recordings of which were mainly collected in the years between 1950 and 1960 under the supervision of Eberhard Zwirner, is archived and edited in the

German Language Archive (*Deutsches Spracharchiv* (DSAv)) at the Institute for German Language in Mannheim (*Institut für deutsche Sprache* (IDS)). It comprises 5857 dialect audio recordings with an average length of twelve minutes. For more detail cf. Lenz 2007 and <http://dsav-wiss.ids-mannheim.de/DSAv/KORPORAB.HTM>.

8. The Pfeffer corpus was compiled in 1961 by Alan Pfeffer and Walter F. W. Lohnes, who collected recordings of 403 speakers of 'colloquial German' in 57 cities in the German-speaking area. The speakers' dialogues were supposed to represent the 'urban vernacular'. For more details, cf. Lenz (2007) and <http://dsav-wiss.ids-mannheim.de/DSAv/KORPORAB.HTM>.
9. It might be assumed that the *kriegen* occurrences only occur as variants in the context of the representation of oral speech. However, this is not, or only partly, the case: a certain amount (namely inter-regionally divergent between 33% and 62%) of all *kriegen* occurrences are indeed found in such contexts, but the many remaining occurrences are realised without any labelling.
10. '[*A*]*voir* n'est rien autre qu'un *être-à* inversé' (Benveniste, 1966, p. 197). For a comprehensive discussion, see Lenz (2013c) as well as Wunderlich (2006, 2012).
11. Cf. e.g. DWb (1854-1961, volume 5, p. 2232).
12. According to the results of Lenz (2012), the emergence of the *kriegen* passive in German seems to occur in a different way than Landsbergen (2006) assumes is the case for the *krijgen* passive in Dutch. Landsbergen (2006) derives the Dutch *krijgen* passive from the Dutch participial RC with *krijgen*. The results reported in Lenz (2012), however, support the hypothesis that the German *kriegen* passive has not developed from the resultative construction with a past participle, but has gone through an independent grammaticalization channel.

References

Abraham, Werner, 'Grammatik von *kriegen* und *bekommen*', *Osnabrücker Beiträge zur Sprachtheorie* 30 (1985), pp. 142-165.

AdA = Elspaß, Stephan & Möller, Robert, *Atlas zur deutschen Alltagssprache*. 2003ff. http://www.atlas-alltagssprache.de/

Askedal, John Ole (2005). 'Grammatikalisierung und Persistenz im deutschen "Rezipienten-Passiv" mit *bekommen/kriegen/erhalten*'. In: Torsten Leuschner, Tanje Mortelmans & Sarah De Groodt (eds.), *Grammatikalisierung im Deutschen*. Berlin & New York: Mouton De Gruyter, 2005, pp. 211-228.

Benveniste, Émile, *Problèmes de linguistique générale*. Paris: Gallimard, 1966.

Boas, Hans C., *A constructional approach to resultatives*. Stanford: CSLI Publications, 2003.

Boas, Hans C., 'Zum Abstraktionsgrad von Resultativkonstruktionen'. In: Stefan Engelberg, Anke Holler & Kristel Proost (eds.), *Sprachliches Wissen zwischen Lexikon und Grammatik*. Berlin & New York: Mouton de Gruyter, 2011, pp. 37-69.

Chang, Linling, *Resultativkonstruktionen im Deutschen mit einem Exkurs zu chinesischen Resultativkonstruktionen*. München: Herbert Utz, 2003.

Duden-Grammatik = *Duden. Die Grammatik. Unentbehrlich für richtiges Deutsch. Herausgegeben von der Dudenredaktion. 8., überarbeitete Auflage*. Vol. 4. Mannheim & Zürich: Dudenverlag, 2009[8].

Duden-Universal = *Duden. Deutsches Universalwörterbuch*. Mannheim: Langenscheidt, 2010ff. Electronic resource.

Duden-Zitate = *Duden. Das große Buch der Zitate und Redewendungen. Über 15000 klassische und moderne Zitate und feste Wendungen. 2., überarbeitete und aktualisierte Auflage. Herausgegeben von der Dudenredaktion*. Mannheim: Dudenverlag, 2007[2].

DWb = *Deutsches Wörterbuch (1854-1961). Von Jacob und Wilhelm Grimm. 16 Bde. in 32 Teilbänden. Quellenverzeichnis Leipzig 1971*. Leipzig: Hirzel, 1854-1961.

E-VALBU = *Elektronisches Valenzwörterbuch deutscher Verben.* Institut für deutsche Sprache (IDS) in Mannheim. http://hypermedia2.ids-mannheim.de/evalbu/index.html

Elspaß, Stephan. *Sprachgeschichte von unten. Untersuchungen zum geschriebenen Alltagsdeutsch im 19. Jahrhundert*. Tübingen: Niemeyer, 2005.

Ettlinger, Marc, '*Constructional compositionality and the English resultative'. Proceedings of the 41st Annual Meeting of the Chicago Linguistics Society*. Chicago: Chicago Linguistics Society, 2005.

Glaser, Elvira, 'Krieg und kriegen: zur Arealität der BEKOMMEN-Periphrasen'. In: Ulla Kleinberger Günther, Annelies Häcki Buhofer & Elisabeth Piirainen (eds.), *"Krieg und Frieden". Auseinandersetzung und Versöhnung in Diskursen*. Tübingen: Francke, 2005, pp. 43-64.

Goldberg, Adele E., *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press, 1995.

Goldberg, Adele E. & Jackendoff, Ray, 'The English Resultative as a Family of Constructions', *Language* 80 (2004), pp. 532-568.

Halliday, Michael A.K., 'Notes on transitivity and theme in English, Part I', *Journal of Linguistics* 3 (1967), pp. 37-81.

Hoekstra, Teun. 'Krijgen'. In: Hans Bennis & William van Lessen Kloeke (eds.). *Linguistics in the Netherlands 1983*. Dordrecht, Foris, 1984, pp. 65-72.

Hopper, Paul & Traugott, Elizabeth C., *Grammaticalization*. Cambridge: Cambridge University Press, 1993.

Horst, Joop van der, 'Het nieuwe *krijgen*. Hoe er een hulpwerkwoord bij komt', *Onze Taal* 71 (2002), pp. 174-176.

Huang, Han-Chun, 'A constructional approach to argument realization of Chinese resultatives', *UTA Working Papers in Linguistics* 2 (2006), pp. 13-31.

Kaufmann, Ingrid, *Konzeptuelle Grundlagen semantischer Dekompositionsstrukturen. Die Kombinatorik lokaler Verben und prädikativer Elemente*. Tübingen: Niemeyer, 1995.

Kuteva, Tania, *Auxiliation. An Enquiry into the Nature of Grammaticalization*. Oxford: Oxford University Press, 2001.

Landsbergen, Frank, 'Getting a sense of *krijgen*. A diachronic study of its polysemy and grammaticalization'. In: Jeroen van de Weijer & Bettelou Los (eds.), *Linguistics in the Netherlands 2006*. Amsterdam: John Benjamins, 2006, pp. 150-161.

Leirbukt, Oddleif, *Untersuchungen zum bekommen-Passiv im heutigen Deutsch*. Tübingen: Niemeyer, 1997.

Lenz, Alexandra N., 'Zur variationslinguistischen Analyse regionalsprachlicher Korpora'. In: Werner Kallmeyer & Gisela Zifonun (eds.), *Sprachkorpora. Datenmengen und Erkenntnisfortschritt. IDS-Jahrbuch 2006.* Berlin & New York: Mouton de Gruyter, 2007, pp. 169-202.

Lenz, Alexandra N., 'On the Perspectivization of a Recipient Role - Crosslinguistic Results from a Speech Production Experiment'. In: Marc Fryd (ed.), *The passive in Germanic Languages. Groninger Arbeiten zur germanistischen Linguistik* (*GAGL*), 49 (2009), pp. 125-144. http://gagl.eldoc.ub.rug.nl/root/2009-49/2009-49-06/

Lenz, Alexandra N., 'Emergence of Varieties through Restructuring and Reevaluation'. In: Peter Auer & Jürgen Erich Schmidt (eds.), *Language and Space. An International Handbook of*

Linguistic Variation. Volume 1: Theories and Methods. Berlin & New York: Mouton de Gruyter, 2010, pp. 295-315.

Lenz, Alexandra N., 'Zum *kréien*-Passiv und seinen "Konkurrenten" im schriftlichen und mündlichen Luxemburgischen'. In: Peter Gilles & Melanie Wagner (eds.), *Linguistische und soziolinguistische Bausteine der Luxemburgistik*, Frakkfurt: Peter Lang, 2011, pp. 5-27.

Lenz, Alexandra N., 'On the genesis of the German recipient passive – Two competing hypotheses in the light of current diaelct data'. In: Gunther de Vogelaer & Guido Seiler (eds.), *The Dialect Laboratory. Dialects as a testing ground for theories of language change*. Amsterdam: John Benjamins, 2012, pp. 121-138.

Lenz, Alexandra N., 'Three "competing auxiliaries" of a non-canonical passive – On the German GET passive and its auxiliaries'. In: Artemis Alexiadou & Florian Schäfer (eds.), *Non-canonical passives*. Amsterdam: John Benjamins, 2013a, pp. 63-94.

Lenz, Alexandra N. Lenz, 'Synchrone Variation als Spiegel diachroner Pfade - Resultativkonstruktionen in „real" und „apparent time"'. In Petra Vogel (ed.), *Jahrbuch der Gesellschaft für Germanistische Sprachgeschichte: Sprachwandel im Neuhochdeutschen*. Berlin & New York: Mouton de Gruyter, 2013, pp. 144-169.

Lenz, Alexandra N. *Vom* kriegen *und* bekommen*. Kognitiv-semantische, variationslinguistische und sprachgeschichtliche Perspektiven*. Berlin & New York: Mouton de Gruyter, 2013c.

Mcintyre, Andrew, 'The Semantic and Syntactic Decomposition of *get*. An Interaction between Verb Meaning and Particle Placement', *Journal of Semantics* 22 (2005), pp. 401-438.

Molnárfi, László, 'Wie kry wat in 'iets gedoen kry'? Oor die resultatiewe *kry*+Part II-konstruksies in die huidige Afrikaans', *Tydskrif vir Geesteswetenskappe* 35 (1995), pp. 109-127.

Molnárfi, László, '*He gets the problem solved.* On the Functional Grammaticalization of *get* and *kry* in English and Afrikaans', *South African Journal of Linguistics* 15 (1997), pp. 18-26.

Molnárfi, László, 'Kasusstrukturalität und struktureller Kasus – zur Lage des Dativs im heutigen Deutsch', *Linguistische Berichte* 176 (1998), pp. 535-580.

RheinWb = *Rheinisches Wörterbuch. Auf Grund der Vorarbeiten von J. Franck begonnen, von allen Kreisen des Rheinischen Volkes unterstützten Sammlung. Bearb. und hrsg. von Josef Müller* [*Bd. 7 und 8 hrsg. von Karl Meisen*]. *8 Volumes*. Bonn: Klopp, 1928-1971.

SchlHWb = Mensing, Otto (ed.), *Schleswig-Holsteinisches Wörterbuch*. Neumünster: Wachholtz, 1927-1935.

Welke, Klaus, *Valenzgrammatik des Deutschen. Eine Einführung*. Berlin & New York: Mouton de Gruyter, 2011.

Wunderlich, Dieter, 'Towards a structural typology of verb classes'. In: Dieter Wunderlich (ed.), *Advances in the Theory of the Lexicon*. Berlin & New York: Mouton de Gruyter, 2006, pp. 57-166.

Wunderlich, Dieter, 'Lexical decomposition in grammar'. In: Markus Werning, Wolfram Hinzen & Edouard Machery (eds.), *The Oxford Handbook of Compositionality*. Oxford: Oxford University Press, 2012, pp. 307-327.

Zifonun, Gisela, Hoffmann, Ludger & Strecker, Bruno, *Grammatik der deutschen Sprache. Volumes 1-3*. Berlin & New York: Mouton de Gruyter, 1997.

Historical corpora and texts

BFNhd = *Bonner Frühneuhochdeutschkorpus* (1350-1700). http://www.korpora.org/Fnhd/

Br19Jh = Auswandererbriefe (1800-1925); cf. Elspaß, Stephan, *Sprachgeschichte von unten. Untersuchungen zum geschriebenen Alltagsdeutsch im 19. Jahrhundert*. Tübingen: Niemeyer, 2005.

BugBib = Bugenhagen-Bibel [1.-5. Buch Mose] = *Bibel, übersetzt von Martin Luther, ndd.* Lübeck, Ludwig Dietz, 1 April 1534 (Titel page reads 1533).

ChrOlde = *Chronik des Johann Oldecop* (*1500*). Karl Euling (ed.) (1891). Tübingen.

GoslStr = *Das Stadtrecht von Goslar* (*um 1350*). Wilhelm Ebel (ed.) (1968). Göttingen, Vandenhoeck & Ruprecht.

HexK = Münsteraner Hexen-Korpus (1565-1656); cf. Macha, Jürgen, Elvira Topalović, Iris Hille, Uta Nolting & Anja Wilke (eds.), *Deutsche Kanzleisprache in Hexenverhörprotokollen der Frühen Neuzeit. Band 1: Auswahledition*. Berlin & New York: Mouton de Gruyter, 2005.

LaurSch = *Niederdeutsche Scherzgedichte von Johann Lauremberg. Nachdruck der Ausgabe von 1652. Mit Einleitung, Anmerkungen und Glossar von Wilhelm Braune.* (1979). Halle (Saale), Niemeyer.

LectFlor = Lectulus Noster Floridus (1486). *Unser Blumenbettchen. Eine devot-mystische Schrift des 15. Jahrhunderts. Niederdeutsch von Johannes Veghe*. Heinrich Rademacher (ed.) (1938). Hiltrup, Herz-Jesu-Missionshaus.

Lubbeke = *Die Aufzeichnungen des Wewelsflether Kirchspielvogts Daniel Lubbeke aus den Jahren 1599-1609*. Klaus-J. Lorenzen-Schmidt (ed.) (2000). Neumünster, Wachholtz.

LübBib = Lübecker Bibel [1.-5. Buch Mose] = *Biblia, mit Glossen nach den Pastillen des Nicolaus Lyra. Lübeck*. Steffern Arndes, 19 November 1494.

LübHist = *Die Lübecker Historienbibel* (*um 1470*). *Die niederdeutsche Version der Nordniederländischen Historienbibel.* Margarete Andersson-Schmitt et al. (eds.) (1994). Köln, Böhlau.

RdV = *Reinke de Vos. Nach der ältesten Ausgabe* (*Lübeck 1498*). *Mit Einleitung, Anmerkungen und einem Wörterbuche. Von August Lübbe*. (1867). Oldenburg, Stalling.

Reut-I = *Läuschen un Rimels* (*1853*). *Von Fritz Reuter. Plattdeutsche Gedichte heiteren Inhalts in mecklenburgisch-vorpommerscher Mundart. Erster Teil*. Wismar, Hinstorff Verlag. http://gutenberg.spiegel.de/buch/1386/1.

Reut-II = *Läuschen un Rimels* (*1858*). *Von Fritz Reuter. Plattdeutsche Gedichte heiteren Inhalts in mecklenburgisch-vorpommerscher Mundart. Zweiter Teil*. Wismar, Hinstorff Verlag. http://gutenberg.spiegel.de/buch/1393/2.

StRechte = *Slesvig stadsret / Flensborgs stadsret* (*1400/1431/1450*). In: *Danmarks gamle Købstadlovgivning.* Erik Kroman (ed.) (1951). Vol. 1. Sønderjylland. Kopenhagen, Rosenkilde og Bagger.

VerlS = *Der verlorene Sohn, ein Fastnachtspiel von Burkard Waldis. Nachdruck der Ausgabe von 1527. Hrsg. von Gustav Milchsack. Nachdruck der Ausgabe von 1527.* (1881). Halle (Saale), Niemeyer.

Newspapers

(cf. Cosmas IIweb (version 1.6.2) http://www.ids-mannheim.de/cosmas2/)

HannAllg = *Hannoversche Allgemeine*

NieÖ = *Niederösterreichische Nachrichten*

NüN = *Nürnberger Nachrichten*

RhZ = *Rhein-Zeitung*

StGal = *St. Galler Tagblatt*

Over de auteur

Alexandra N. Lenz, University of Vienna, Institute for Germanistics, Universitätsring 1, A-1010 Wien.
E-mail: alexandra.lenz@univie.ac.at

De *aan*-constructie in het 17de-eeuwse Nederlands

Een semasiologische studie

Tim Geleyn & Timothy Colleman

MEM 67 (2): 211–245
DOI: 10.5117/TET2015.2.GELE

Abstract

The Dutch *aan*-construction (e.g. *Hij gaf een bos bloemen aan zijn vrouw* 'He gave a bouquet to his wife'), the prepositional alternative for the double object construction (*Hij gaf zijn vrouw een bos bloemen* 'He gave his wife a bouquet'), is a post Middle Dutch innovation (i.e. after 1500 AD). The precise details of the rise of the *aan*-construction remain, however, understudied. It is for example unclear whether the construction really breaks through in the 17th century, as Weijnen and Gordijn (1970) argue on the basis of a small corpus of farces, and what its semantic range was in those early days. In this paper we try to shed more light on these issues. On the basis of a self-compiled corpus of literary Dutch, we firstly show that the construction was not only already frequently attested in the language use in the 17th century, but also covered a remarkably wide semantic range at that time. Next, via a detailed comparison with data for the 20th century, we show that there have been interesting changes concerning the semantic evolution of the *aan*-construction. The structural weight of a cluster of 'do'- and 'send'-verbs for example declines over time and at more general level there seems to be a trend towards more abstract uses of the *aan*-construction. A diachronic collostructional analysis (Hilpert, 2006) and Configural Frequency Analysis (von Eye, 2002) lends a statistical underpinning to our observations.

Keywords: construction grammar, corpus linguistics, Dutch, *aan*-dative, semantic changes, diachronic collostructional analysis, Configural Frequency Analysis

1 Inleiding[1]

Een bekend grammaticaal verschijnsel in het hedendaagse Nederlands is de zgn. *datiefalternantie*: veel ditransitieve werkwoorden komen zowel voor in een constructie met twee nominale objecten (*Hij gaf zijn vrouw een bos bloemen*; in wat volgt: de *dubbelobjectconstructie* of afgekort DOC) als in een constructie met een direct object en een omschreven indirect object met het voorzetsel *aan* (*Hij gaf een bos bloemen aan zijn vrouw*; in wat volgt: de *prepositionele datiefconstructie met aan*, of kortweg de *aan-constructie*), zonder dat met dat grammaticaal verschil een duidelijk betekenisverschil correspondeert (zie over de relatie tussen die beide indirect objectconstructies o.m. Van Belle & Van Langendonck, 1996; Duinhoven, 2003; Colleman 2006, 2009, 2012). De dubbelobjectconstructie komt van oudsher voor in het Nederlands, met dien verstande dat de nominale objecten in oudere taalfasen formeel gemarkeerd werden als resp. accusatief- en datiefobject – zie het Middelnederlandse voorbeeld in (1) met datiefmarkering *-n* bij het lidwoord en het adjectief in het mannelijk meervoud.[2]

(1) Ic hebbe ghenomen grote have den riken luden van haren goede.
'Ik heb de rijke mensen een groot deel van hun bezit ontnomen.'
(*Karel ende Elegast,* ± 1270)

De *aan*-constructie is van recenter datum: pas als het naamvallensysteem vanaf het Vroegnieuwnederlands (vanaf 1500) sterk is afgezwakt (van der Horst, 2008, pp. 795-803), zien we de geleidelijke opkomst van het locatieve voorzetsel *aan* om de indirect objectfunctie uit te drukken in ditransitieve contexten. Voorbeeldzin (2) is een vroeg 17de-eeuws voorbeeld:

(2) [D]at […] het lichaem wederom sij aen den vrinden gaeven.
(P.C. Hooft - *Achilles en Polyxena*, 1614)

In verscheidene andere Germaanse talen heeft zich grosso modo dezelfde evolutie voorgedaan: zo is in het Engels het allatieve voorzetsel *to* gerekruteerd om het indirect object te markeren; blijkens de corpusdata van Zehentner (2014) gebeurde dat overigens al in de loop van de Middelengelse periode (zie o.m. ook nog Visser 1963, pp. 606-635; Allen, 1995, pp. 28-29; McFadden, 2002). Rohdenburg (2007, pp. 219-220) merkt over de opkomst van het prepositionele indirect object in het Engels en de bijbehorende veranderingen in het semantische bereik van de dubbelobjectconstructie op dat het hier gaat om 'a vast and complex area of change which has been

barely touched upon now'. Dat dat terrein nog grotendeels braak ligt, geldt zeker ook voor het Nederlands: het mag dan wel algemeen aangenomen worden dat prepositionele constructies zoals de *aan*-constructie ontstaan zijn om een lacune te vullen die was ontstaan door erosie van de naamvalsonderscheidingen (zie bv. Duinhoven 2003, pp. 218-219, van der Horst, 2008, p. 578), de details van dat proces zijn nog nauwelijks onderzocht. Het enige bestaande corpusonderzoek naar de opkomst van het prepositionele indirect object in het Nederlands is een al wat oudere verkennende studie van Weijnen & Gordijn (1970), waarin op basis van een klein corpus met kluchten vanaf het Middelnederlands t.e.m. de 18de eeuw wordt geconcludeerd dat de *aan*-constructie niet vroeger doorbreekt dan in de 17de eeuw. Het oudste voorbeeld[3] van de *aan*-constructie dat ze aantroffen in hun corpus is:

(3) Dat ghy niet wilt dat U geschiet,
En doet sulx aen een ander niet.
(G.A. Bredero - *De klucht van de molenaar*, 1613)

Weijnen & Gordijn (1970, p. 128) wijzen er ook op dat de omschrijving met *aan* als een volwaardig alternatief voor de datiefnaamval wordt gepresenteerd in de *Waernemingen op de Hollandsche Tael* (1635-1638), een reeks opmerkingen en observaties in verband met de Nederlandse taal toegeschreven aan P.C. Hooft.[4] Nauwelijks enkele decennia na het oudste voorbeeld dat Weijnen & Gordijn aantroffen in hun (weliswaar beperkt) corpus, is de *aan*-constructie dus blijkbaar al voldoende ingeburgerd om te worden opgenomen in een prescriptief grammaticaal naslagwerk. Vanaf dan lijkt er dus zeker sprake te zijn geweest van een datiefalternantie in het Nederlands.

Over de precieze manier waarop de *aan*-constructie haar intrede deed in het Nederlands, bestaat nog veel onduidelijkheid. In deze paper willen we dan ook enkele kwesties in meer detail onderzoeken. Ten eerste is uit de enkele 17de-eeuwse voorbeelden die worden aangehaald in Weijnen & Gordijn (1970) en in naslagwerken zoals van der Horst (2008) niet duidelijk welk semantisch domein de *aan*-constructie precies bestreek in de vroegste fasen van het Nieuwnederlands. Op basis van uitgebreid corpusonderzoek zullen we het semantische bereik van de *aan*-constructie in de periode 1600-1650 gedetailleerd in kaart brengen, m.a.w. in de periode waarin de constructie lijkt te zijn doorgebroken. De eerste vraag daarbij is al of er inderdaad van een doorbraak sprake was: komt de *aan*-constructie voldoende vaak – en met een voldoende verschillend aantal werkwoorden

– voor om te bevestigen dat het om een goed ingeburgerde constructie gaat, of blijkt het gebruik ervan toch nog relatief sporadisch te zijn geweest? Interessant hierbij is om na te gaan of de constructie in de vroegste periode al meteen dienst deed als prepositioneel alternatief voor een breed gamma van ditransitieve werkwoorden en situatietypes, dan wel of ze in de 17de eeuw nog vooral beperkt was tot prototypische 'geven'-situaties, d.w.z. situaties waarbij er sprake is van een concrete overdracht van een fysiek voorwerp tussen twee personen. Dat laatste zou kunnen worden verondersteld vanuit de metaforische origine van de *aan*-constructie: het locatieve voorzetsel *aan* (met basisbetekenis 'contact') wordt ingezet in ditransitieve contexten via de metafoor 'bezit is contact', zie daarover o.m. Colleman en De Clerck (2009), Colleman (2010a) en verder, paragraaf 3. Vanuit dat opzicht kan er in de vroegste periode nog een overwicht verwacht worden van *aan*-zinnen die een overdracht van een concreet object tussen twee bezielde medespelers beschrijven.

Ten tweede willen we aan de hand van een vergelijking met data voor de eerste helft van de 20ste eeuw nagaan hoe de constructie zich door de tijd heen ontwikkeld heeft. In welke opzichten verschilt het semantische bereik van de *aan*-constructie in de periode 1900-1950 van dat in de periode 1600-1650, zowel kwalitatief (ontstaan/verdwijnen van gebruikswijzen) als kwantitatief (verschuivingen in het relatieve gewicht van de verschillende gebruikswijzen)? Om onze bevindingen statistisch te onderbouwen maken we gebruik van twee technieken: diachrone collostructionele analyse (Hilpert, 2006) en *Configural Frequency Analysis* (von Eye, 2002). Die vergelijking van twee perioden van 50 jaar is te beschouwen als de eerste aflevering van een grootschaliger diachroon onderzoek dat het semantische bereik van de *aan*-constructie in de gehele tussenliggende periode gedetailleerd in kaart tracht te brengen. Om te verduidelijken waarin precies de toegevoegde waarde schuilt van een dergelijk longitudinaal onderzoek naar de semantiek van één welbepaalde grammaticale constructie, gaan we in de volgende paragraaf in op de huidige stand van het constructiegrammaticale onderzoek naar betekenisverandering in grammaticale constructies. Nadien volgt een paragraaf waarin de vorm- en betekeniskenmerken van de Nederlandse *aan*-constructie verder worden toegelicht. Paragraaf 4 beschrijft de methodologie van het onderzoek en in paragrafen 5 en 6 worden de resultaten gepresenteerd van resp. de inventarisatie van het semantische bereik van de constructie in de 17de eeuw en de vergelijking met de 20ste eeuw. Conclusies worden in de zevende paragraaf geformuleerd.

2 Theoretische achtergrond: diachrone constructionele semasiologie

Het is bekend dat het taalsysteem vanuit constructiegrammaticaal oogpunt beschouwd wordt als een *constructicon*, d.w.z. een gestructureerd netwerk van duizenden en duizenden constructies, waarbij constructies gedefinieerd worden als geconventionaliseerde vorm/betekenis-paren. Constructies vertonen sterke onderlinge verschillen in complexiteit en schematiciteit: van morfemen en enkelvoudige woorden over complexe woorden en meerwoordspatronen tot abstracte, grotendeels of volledig schematische constructies zoals de dubbelobjectconstructie of de passieve constructie (zie o.m. Goldberg, 2003 voor een kort overzicht van de belangrijkste uitgangspunten van de constructiegrammatica, en de papers in Hoffmann en Trousdale, 2013 voor een mooie illustratie van de veelzijdigheid van het hedendaagse constructiegrammaticale onderzoek). De *diachrone* constructiegrammatica houdt zich dan uiteraard bezig met de studie van verandering in het constructicon. In hun invloedrijke recente monografie stellen Traugott en Trousdale (2013, p. 39) zich uitdrukkelijk tot doel om een overkoepelende theorie uit te werken voor het diachrone constructiegrammaticale onderzoek, waarin onder meer een cruciaal onderscheid gemaakt wordt tussen constructionalisatie aan de ene kant en constructionele veranderingen aan de andere. Onder constructionalisatie (*constructionalization*, Traugott en Trousdale, 2013, pp. 20-30 en passim) wordt de creatie van een nieuwe knoop in het constructionele netwerk verstaan, m.a.w. een nieuw vorm/betekenis-paar. Dat is een gradueel proces dat gepaard gaat met veranderingen in schematiciteit, productiviteit en compositionaliteit. Als voorbeeld van constructionalisatie bespreken de auteurs onder meer de evolutie van Engelse binominale partitieven (bv. *a lot/bit/shred of a N*) tot grammaticale hoeveelheidsmarkeerders. Op betekenisvlak zijn die partitieven gradueel hun lexicale betekenis (bv. *a lot of land* als 'een perceel grond') verloren en duiden ze enkel nog een hoeveelheid aan ('veel grond'). Congruentieverschijnselen (vgl. *This lot of books was never read* met *A lot of books were never read*) wijzen erop dat er daarnaast cruciaal ook *vorm*verandering heeft plaatsgevonden: het tweede nomen is gereanalyseerd (of, in termen van Traugott en Trousdale, 'geneoanalyseerd') als het syntactische hoofd van de nominale constituent. We kunnen dan ook spreken van een reeks *nieuwe* constructies, die een nieuwe vorm koppelen aan een nieuwe betekenis (Traugott en Trousdale, 2013, pp. 23-26). Veel diachroon constructiegrammaticaal werk beschrijft veranderingen van dat type; de focus ligt m.a.w. vaak op het ontstaan van nieuwe

constructies – of, in mindere mate, ook wel op het verdwijnen van constructies (zie ook Colleman en De Clerck, 2011, pp. 201-205 voor discussie en voorbeelden). In dergelijke studies wordt ook vaak de link gemaakt met inzichten uit de grammaticalisatietheorie (zoals bv. *semantic bleaching* of *gradualness*) of wordt de vraag gesteld hoe diachrone constructiegrammatica en de grammaticalisatietheorie zich tot elkaar verhouden of waarin ze elkaar aanvullen. Noël (2007) is een van de eersten die de twee expliciet tegenover elkaar zet; na hem wordt de kwestie nog in veel andere papers ter sprake gebracht, zie o.m. de bijdragen in Bergs en Diewald (2008), Hilpert (2008), Fried (2009), Traugott (2008), enz.

Naast constructionalisatie zijn er 'gewone' constructionele veranderingen (*constructional changes*, Traugott en Trousdale 2013, pp. 20-30 en passim), die slechts betrekking hebben op één interne dimensie van een constructie, d.w.z. de vorm *of* de betekenis ervan. Het gaat hier met andere woorden om veranderingen binnen een bestaande constructie, die niet leiden tot het ontstaan een nieuwe knoop in het constructionele netwerk. Aan dergelijke veranderingen is verhoudingsgewijs veel minder aandacht besteed in de diachrone constructiegrammatica. Nochtans is makkelijk te beargumenteren dat ook de studie daarvan van groot belang is voor de verdere uitbouw van een constructiegrammaticale theorie van taalverandering. In het *lexicon* is verandering uiteraard niet beperkt tot het ontstaan van nieuwe woorden en het verdwijnen van oude: daarnaast zijn ook de vorm en betekenis van bestaande woorden vatbaar voor verandering. Wat de betekenis betreft, zijn er uit het diachrone lexicologische onderzoek verschillende processen van semantische verandering bekend, zoals generalisatie, specialisatie, metonymie, enz. (zie o.m. Geeraerts, 1997; Geeraerts, Grondelaers en Bakema, 1994; Geeraerts, Grondelaers en Speelman, 1999). Aangezien de constructiegrammatica ervan uitgaat dat er geen fundamenteel onderscheid is tussen woorden (of, in de constructiegrammaticale terminologie, *substantiële* constructies) en (grotendeels) schematische grammaticale constructies zoals de dubbelobjectconstructie, de *aan*-constructie, enz., mag verwacht worden dat ook die laatste onderhevig zijn aan allerlei processen van betekenisverandering. Als zou blijken dat hun betekenissen *niet* op een vergelijkbare manier veranderen als die van lexicale items, dan zou dat een ernstig argument zijn *tegen* de fundamentele theoretische gelijkschakeling van substantiële en schematische constructies.

Om meer zicht te krijgen op die kwestie zijn diachrone casestudies nodig waarin de semasiologische structuur van gevestigde schematische constructies gedurende lange tijd gevolgd wordt. In Colleman (2011) en

Colleman en De Clerck (2011) wordt uitgebreid ingegaan op veranderingen in het semantische bereik van de Nederlandse en Engelse dubbelobjectconstructies. In beide talen lijkt er zich bij de dubbelobjectconstructie in de loop van de laatste drie à vier eeuwen een proces van betekenisspecialisatie te hebben voorgedaan: voor bepaalde situatietypes was de dubbelobjectconstructie vroeger wel een geschikt uitdrukkingsmiddel maar nu niet meer. Voor het Engels (Colleman en De Clerck, 2011, pp. 193-201) gaat het daarbij onder meer om het gebruik van de constructie om 'verbanning' te benoemen (met werkwoorden zoals *banish*, *discharge*, *dismiss*), het gebruik ervan om privatieve bezitsoverdrachten te benoemen (met onder meer *rob, deprive, bereave*) en het gebruik om benefactieve situaties te benoemen waarin geen beoogde overdracht in het geding is (als in ??*She opened me the door*, een geval van zgn. *substitutive benefaction*, waarbij de dubbelobjectconstructie in het hedendaagse Engels in het beste geval een zeer gemarkeerde optie is). Voor de DOC in het Nederlands maakt Colleman (2011, pp. 403-405) een verkennende vergelijking tussen 19de-eeuws en hedendaags Nederlands. Het lijkt erop dat bepaalde semantische klassen in de loop van die periode minder gebruikelijk zijn geworden in de DOC. Dat geldt bijvoorbeeld voor benefactieven en malefactieven (*Ik kocht hem een boek*[5]), en voor bepaalde gebruikswijzen met privatieve werkwoorden (bv. *ontnemen, afnemen*). Daarbij is het doorgaans niet zo dat die gebruikswijzen geheel uit het Nederlands verdwenen zijn, wel is de frequentie ervan afgenomen. Vergelijkbaar onderzoek naar de DOC vinden we tot slot bij Barðdal, Kristoffersen en Sveen (2011), waarin het semantische bereik van de ditransitieve constructie met objecten in accusatief en datief in het Moderne IJslands en Faeröers wordt beschreven en waarin wordt vastgesteld dat dat semantische bereik uitgebreider is dan in het hedendaagse Noors, dat een DOC zonder formeel gemarkeerde objecten heeft.

Het huidige artikel over de semantiek van de Nederlandse *aan*-constructie in de 17de en in de vroege 20ste eeuw past binnen die ontluikende subdiscipline van de diachrone constructionele semasiologiestudies naar constructionele veranderingen: we bekijken in detail de evolutie van een constructie waarvan beweerd wordt dat ze sinds de 17de eeuw goed ingeburgerd is en door grammatici aanvaard werd als volwaardig alternatief voor de dubbelobjectconstructie. Het is in dit artikel dus uitdrukkelijk *niet* de bedoeling om meer licht te werpen op de eigenlijke constructionalisatie van de *aan*-constructie: we zullen ons niet buigen over de bronnen van de *aan*-constructie in de periode vóór 1600, noch zullen we de vraag proberen te beantwoorden vanaf wanneer de *aan*-constructie precies als een afzonderlijke argumentstructuurconstructie mocht gelden. De focus ligt op het se-

mantische bereik van de constructie in twee verschillende subfasen van haar evolutie. In de volgende paragraaf bespreken we als aanloop op de presentatie van de resultaten van het diachrone de onderzoek eerst de formele afbakening en semantiek van de *aan*-constructie in het hedendaagse Nederlands.

3 De Nederlandse *aan*-constructie

In navolging van Colleman (2010a) beschouwen we de *aan*-constructie formeel als een argumentstructuurconstructie met vier slots: subject, werkwoord, direct object en *aan*-object, zie voorbeeldzinnen (4) en (5).

(4) Hij gaf een bos bloemen aan zijn vrouw.
SU V DO *aan*-object

(5) Hij besteedde nauwelijks aandacht aan dat probleem.
SU V DO aan-object

Merk op dat we, in tegenstelling tot bijvoorbeeld Van Belle en Van Langendonck (1996), zowel *aan*-zinnen met indirecte objecten (zin 4), parafraseerbaar door de DOC (*Hij gaf zijn vrouw een bos bloemen*), als *aan*-zinnen met voorzetselvoorwerpen (zin 5), niet parafraseerbaar door de DOC (**Hij besteedde dat probleem nauwelijks aandacht*) tot de *aan*-constructie rekenen. We willen in deze studie immers de *aan*-constructie op zich beschouwen, niet in relatie tot de DOC, vandaar dat we beide types *aan*-objecten meenemen. Een zin zoals (6a) is dan weer geen voorbeeld van de *aan*-constructie, omdat hierbij geen sprake is van een object maar van een complement (cf. vervangbaarheid van *aan* door andere voorzetsels in 6b, en de verwijzing met *waar*/*daar*[6] in 6c):

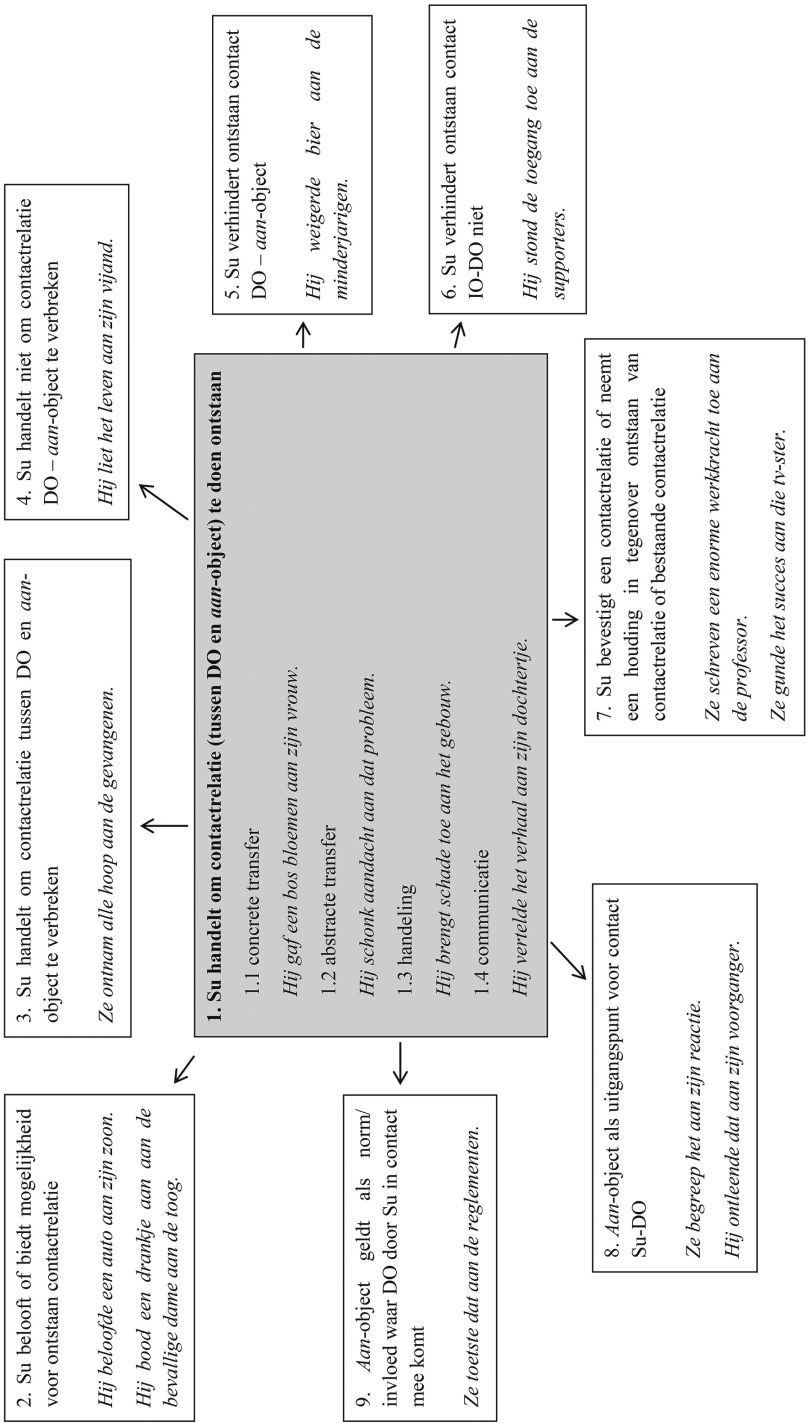
(6) a. Hij zette zijn schoenen aan de deur.
b. Hij zette zijn schoenen bij/naast/... de deur.
c. Waar zette hij zijn schoenen? Aan de deur.

De semantiek van de *aan*-constructie is in het bestaande onderzoek vooral ter sprake gekomen in het kader van studies van de datiefalternantie, d.w.z. er is vooral ingezoomd op de subtiele semantische *verschillen* tussen de *aan*-constructie en de DOC. Van Belle en Van Langendonck (1996) stellen bijvoorbeeld dat de *aan*-constructie vooral ingezet wordt bij concrete, materiële transfers omdat *aan* het ruimtelijke aspect van de beschreven overdracht onderstreept, terwijl de DOC de sterkere betrokkenheid van het

indirect object in de verf zet. Ook in andere pogingen tot explicitering van het semantische contrast spelen noties als 'ruimtelijke overdracht' en '(actieve of passieve) betrokkenheid' of 'geaffecteerdheid' vaak een cruciale rol, al verschillen natuurlijk de details (zie onder meer Kirsner, Verhagen en Willemsen, 1985; Janssen, 1997; Duinhoven, 2003). Colleman (2010a, pp. 274-275) merkt echter op dat het aandacht verdient om de semantiek van de *aan*-constructie ook *op zich* te beschrijven, m.a.w. niet in relatie tot de DOC. In Colleman (2006, 2010a) worden de verschillende betekenissen/gebruikswijzen van de *aan*-constructie van nabij bekeken. De centrale betekenis van de constructie wordt omschreven als 'Subject veroorzaakt: er ontstaat contact tussen direct object en *aan*-object' (of kort gezegd 'caused contact', zie Colleman, 2010a, p. 292), ingegeven door de 'contact'-betekenis van het voorzetsel *aan* zoals beschreven in o.m. Beliën (2002), Cuyckens (1991), Schermer-Vermeer (1991, 2001). Die contactnotie kan letterlijk zijn (zin 7), maar ook abstract (zin 8) of communicatief (zin 9):

(7) Ze koppelden enkele wagons aan de locomotief.
(8) Ik schenk mijn volle aandacht aan dat probleem.
(9) Ze vertelde de waarheid aan de politie.

In Figuur 1 hebben we, net zoals bijvoorbeeld Goldberg (1995) dat doet voor de Engelse DOC, een radiaal semantisch netwerk opgesteld voor de Nederlandse *aan*-constructie: er wordt een semantische kern onderscheiden van waaruit verschillende extensies mogelijk zijn die op verschillende manieren van die semantische kern afwijken. Als semantische kern van de *aan*-constructie kan de eerste betekenis gelden, en meer specifiek betekenis 1.1: 'Subject handelt om een contactrelatie te doen ontstaan tussen een concreet direct object en het *aan*-object'. Vanuit die semantische kern zijn er ten eerste metaforische extensies mogelijk: het direct object is niet concreet maar abstract (betekenissen 1.2 en 1.3) of communicatief (betekenis 1.4).[7] Daarnaast zijn er nog andere, metonymische extensies die betrekking hebben op de variatie m.b.t. de 'handelen'-component van de centrale betekenis (bv. betekenis 2 waarin het subject niet actief handelt maar enkel de belofte van een overdracht doet), de 'richting'-component (betekenis 3: *aan*-object functioneert niet als doel maar als bron), enz. Figuur 1 biedt een overzicht van de belangrijke extensies vanuit de centrale betekenis van de *aan*-constructie. Een van de vragen die we met dit onderzoek willen beantwoorden, is of de *aan*-constructie in de 17de eeuw al een vergelijkbaar breed semantisch terrein bestreek: zoals al in de inleiding werd vermeld, heeft die kwestie nog geen aandacht gekregen in de historisch-taalkundige beschrijving van het Nederlands.



Figuur 1 Betekenisnetwerk van de Nederlandse aan*-constructie*

4 Corpus en methodologie

Voor dit onderzoek is gebruik gemaakt van een zelf gecompileerd corpus literair Nieuwnederlands. Dat corpus omvat de periode 1600-1950, opgedeeld in zeven subperiodes van 50 jaar, met ongeveer 1.5 tot 2 miljoen woorden tekst (proza, drama en non-fictie) per subperiode. Verder zijn er restricties op het aantal woorden per auteur (max. 100.000) en het geboortejaar van de auteur (voor teksten uit de subperiode 1600-1650 bijvoorbeeld moet de auteur geboren zijn tussen 1570 en 1620; dezelfde restrictie wordt gehanteerd in het vergelijkbare Engelse CLMET-corpus, zie De Smet, 2005). Tot 1800 zijn alle teksten van de hand van auteurs uit de Noordelijke Nederlanden. Vanaf 1800 is er ook telkens een (kleiner) deel Zuidelijk (dus Belgisch) Nederlands opgenomen zodat regionale vergelijking mogelijk is.[8]

Het corpus is niet verrijkt met taalkundige annotatie, wat het automatisch opsporen van bepaalde grammaticale constructies uiteraard enigszins bemoeilijkt, zeker als het gaat om volledig schematische constructies als de DOC. Voor de *aan*-constructie liggen de kaarten iets beter: we konden zoeken op het voorzetsel *aan*, rekening houdend met de verschillende spellingsvarianten (*aan, aen, ane, aene, an, aenden, aanden, aende, aande* enz.). Dat leverde ongeveer 10.000 treffers op voor de periode 1600-1650 en 12.000 voor de periode 1900-1950.

Uiteraard waren niet alle zoekresultaten voorbeelden van de *aan*-constructie zoals we die hier afbakenen. Bij de manuele filtering hanteerden we een syntactisch filter, d.w.z. enkel zinnen die vormelijk tot de *aan*-constructie gerekend kunnen worden, dus zinnen met twee objecten, werden meegenomen, inclusief passieve zinnen en zinnen waarin het direct object een zin is i.p.v. een nominale constituent. Verbindingen met o.m. *hangen*, *knopen* en *leggen* hebben we niet meegenomen (zie voorbeeldzin 10), omdat *aan* in zulke contexten vervangbaar is door andere locatieve voorzetsels of bijwoorden.

(10) Wampie trok met een zwaai haar hoed af, en hing hem aan de tuit van een koperen gieter.
(A. den Doolaard - Wampie, 1938)

Er moest dus sprake zijn van een 'vast' voorzetsel *aan*. Uiteraard is er in de praktijk geen dichotomisch onderscheid tussen syntactische contexten met een vast voorzetsel – en dus een *aan*-object – en syntactische contexten met een ruimtelijk complement met *aan*; zie ook de literatuur over de

verre van onproblematische syntactische afbakening van het voorzetselvoorwerp (Schermer-Vermeer, 1990, 2006; Broekhuis, 2004; Vandeweghe, 2011 enz.). Twijfelgevallen op de rand van de categorie werden in een eerste analyseronde als zodanig gemarkeerd en in tweede instantie door de twee auteurs samen besproken – waar mogelijk werd er een beroep gedaan op de lexicografische beschrijving in het WNT om meer zicht te krijgen op de 'vastheid' van het voorzetsel. Uiteindelijk werd een relatief brede opnamepolitiek gehanteerd: gevallen als *iets koppelen aan iets* of *iets verbinden aan iets* bijvoorbeeld werden wel degelijk meegenomen, hoewel ze relatief dicht bij 'locatieve' gebruikswijzen als *iets hangen aan iets* staan en hoewel het WNT aangeeft dat de werkwoorden in kwestie behalve met *aan* ook met *met* voorkomen. Bij sommige werkwoorden kan er nu eenmaal competitie zijn tussen *aan* en een ander vast voorzetsel. Merk op dat er ook een direct object aanwezig moet zijn in de onmiddellijke syntactische context. Ook bij divalente werkwoorden is *aan* opgekomen ter omschrijving van de datief, zie (11) voor een voorbeeld. Het betreft hier echter een andere argumentstructuurconstructie, met slechts twee argumentrollen: we richten ons in dit artikel op de *aan*-constructie met drie argumentrollen, d.w.z. op de constructie waarvan het semantische bereik deels overlapt met dat van de dubbelobjectconstructie. De tweeplaatsige constructie in (11) 'concurreert' niet met de dubbelobjectconstructie maar met constructies met één nominaal (direct of indirect) object.

(11) De grootste val gheschiet ghemeen'lijck aen de grooten.
(S. Coster - *Ithys*, 1615)

Verder hebben we ook zinnen met een reflexief direct object geweerd (voorbeeldzin 12). In dergelijke zinnen verwijst het direct object naar dezelfde referent als het subject en is er dus geen sprake van een op zich staande entiteit, naast het subject, waarmee iets gebeurt. Tot slot zijn de *aan*-constructies bij een kleine set van werkwoorden als *herinneren* en *helpen* weggevallen (13), omdat ook daarvoor geargumenteerd kan worden dat ze een andere argumentstructuurconstructie vertegenwoordigen, waarin het *aan*-object als 'eerste object' geldt in plaats van als 'tweede object' (lees: waarin de *aan*-constituent als een prepositionele tegenhanger van het direct object geldt die wordt gecombineerd met een IO-achtig nominaal object; zie De Schutter, 1974, pp. 223-227 en Colleman, 2006, pp. 414-417 voor een nadere bespreking van dat onderscheid tussen prepositioneel eerste versus tweede object). Andere grensgevallen hebben we geprobeerd zo consistent mogelijk te behandelen.

(12) Maar eenmaal thuis, de schrik in haar sterke lichaam overwonnen, kon zij zich in eenvoudige grootheid van ziel kalm overgeven aan dit noodlot.
(I. Boudier-Bakker - *De klop op de deur*, 1931)

(13) Een kerel, die mijn zaak oplicht, die mij aan een bankroet helpt, als het hem meeloopt?
(M. ter Braak - *De pantserkrant*, 1935)

Uiteindelijk hielden we 1530 voorbeelden van de *aan*-constructie over voor de periode 1600-1650, en 2129 voor 1900-1950. Dat betekent dat alvast bevestigd kan worden dat de *aan*-constructie ook in de vroegst onderzochte periode al goed ingeburgerd was: de constructie komt omgerekend 996 keer voor per miljoen woorden tekst, in de periode 1900-1950 is die genormaliseerde frequentie 1070 per miljoen woorden tekst. Hierna bekijken we eerst het bereik van de *aan*-constructie in 1600-1650 aan de hand van het semantisch netwerk (paragraaf 5), daarna maken we een vergelijking met de periode 1900-1950 en voeren we een diachrone collostructionele analyse uit (subparagrafen 6.1 en 6.2). Tot slot werden 1000 willekeurige zinnen van elke periode geannoteerd op vijf semantische variabelen (causatietype, richting overdracht, aard Subject/DO/*aan*-object), om op die manier bottom-up een semasiologisch profiel van de *aan*-constructie te kunnen opstellen aan de hand van een *Configural Frequency Analysis* (subparagraaf 6.3).

5 Semantisch bereik van de *aan*-constructie in 1600 - 1650

In de eerste fase werden de 1530 *tokens* van de *aan*-constructie gelabeld aan de hand van het betekenisnetwerk uit Figuur 1. We zien dat alle betekenissen die we op basis van Colleman (2006, 2010a) voor het hedendaagse Nederlands kunnen onderscheiden ook al vertegenwoordigd zijn in de eerste helft van de 17de eeuw, zie het overzicht in Tabel 1.

Tabel 1 Verhouding tussen de betekenissen van de *aan*-constructie in 1600-1650 (n = 1530)

Betekenisrelatie	Aantal
1. Su handelt om contactrelatie (tussen DO en *aan*-object) te doen ontstaan	1214 (79.3%)
1.1 Concrete transfer	342 (28.2%)
1.2 Abstracte transfer	282 (23.2%)
1.3 Handeling	241 (19.9%)
1.4 Communicatieve overdracht	349 (28.7%)
2. Su belooft of biedt mogelijkheid voor ontstaan contactrelatie t ussen DO en *aan*-object	41 (2.7%)
3. Su handelt om contactrelatie tussen *aan*-object en DO te verbreken	8 (0.5%)
4. Su handelt niet om contactrelatie tussen *aan*-object en DO te verbreken	8 (0.5%)
5. Su verhindert ontstaan contact tussen DO en *aan*-object	12 (0.8%)
6. Su verhindert ontstaan contact DO – *aan*-object niet	7 (0.5%)
7. Su bevestigt een contactrelatie of neemt een houding in tegenover ontstaan van contactrelatie of bestaande contactrelatie	22 (1.4%)
8. *Aan*-object geldt als uitgangspunt voor contact Su – DO	163 (10.7%)
9. *Aan*-object geldt als norm/invloed waar DO door Su in contact mee komt	12 (0.8%)
10. Restcategorie [9]	43 (2.8%)

Die relatieve 'rijpheid' van het betekenisnetwerk van de *aan*-constructie is opvallend te noemen in het licht van haar nog jonge leeftijd: de constructie moet immers ergens in de loop van de voorafgaande eeuw ontstaan zijn. Als we Tabel 1 in meer detail bekijken, dan zien we dat de eerste groep betekenissen overduidelijk het frequentst is (79.3%). Het gaat om zinnen waarin het subject actief handelt om een contactrelatie tussen DO en *aan*-object te doen ontstaan. Binnen die groep maakten we nog een nader onderscheid tussen concrete transfers[10] (voorbeeldzinnen 14-15), abstracte transfers (zinnen 16-17), op de referent van het *aan*-object gerichte handelingen[11] (18-19) en communicatieve overdrachten (20-21), waarbij 1.1 als de centrale betekenis kon gelden. De vier groepen komen allemaal vrij frequent voor, met percentages van 19 tot 28 procent. Er lijkt met andere woorden niet meteen sprake te zijn van een overwicht van of beperking tot concrete transfers in deze periode zoals misschien verwacht mocht worden vanuit de hypothese dat de *aan*-constructie vanuit haar metaforische basis oorspronkelijk vooral ingezet zou zijn bij materiële transfers. Het valt uiteraard niet uit te sluiten dat het aandeel concrete transfers in de periode voor 1600-1650, m.a.w. bij het echte ontstaan van de constructie, nog aanzienlijker was.

(14) Ende comende te sterven gaf hy eenen rinck aen elck een der voorseyde kinderen.

(G.H. van Breughel - *De tweede vijftigh lustige historien ofte nieuwigheeden Johannis Boccatii*, 1644)

(15) Daer nae wierd' aen mijn heer een gouden Cop gheschoncken.
(I. van Damme - *M. Accii Plauti Amphitruo*, 1635)

(16) En wat brenght ouderdom aen meenigh groot gemoedt Met sijne swackheyt, Heer een swaren teghenspoedt.
(J. van Heemskerck - *De* [*verduytste*] *Cid*, 1641)

(17) Alle dese seer heerlicke victorien heeft God aende vereenigde Landen verleent in vijf maenden tijts.
(A. Valerius - *Nederlandtsche gedenck-clanck*, 1626)

(18) Verwonnen is de stadt [...] die soo veel boosheyt heeft aen onse stadt bedreven.
(I. van Damme - *M. Accii Plauti Amphitruo*, 1635)

(19) Wy sullen ons mede stil houden sonder eenige schade aen de Stadt te doen.
(J.J. Orlers - *Beschrijvinge der stad Leyden*, 1614)

(20) [D]at de Heer van Jenlis, als zelf geweest oover 't beleydt van 't inneemen dier stadt, deeze voorspoedt, door den Ammiraal, aan die Majesteit, ginge verkundighen.
(P.C. Hooft - *Nederlandsche Historien*, 1642-1647)

(21) Nochtans een Vrou heeft my ghegheven om te toonen, O grootste Coningin, aen u dit kostlijck werck.
(S. Coster - *Iphigenia*, 1617)

Ook alle andere betekenissen zijn vertegenwoordigd, maar meestal slechts met een tiental attestaties. De enige groep betekenissen die nog met een aanzienlijke frequentie voorkomt (163 attestaties, of 10.7%), zijn zinnen waarin het *aan*-object geldt als uitgangspunt/bron voor contact tussen het subject en het direct object (betekenis 8). Hoewel dergelijke constituenten in de marge van de categorie 'object' zitten, beschouwen we ze hier wel degelijk nog als objecten en niet als bepalingen, in navolging van o.m. De Schutter (1974) en Colleman (2006). Ze zijn immers referentieel en moeilijk weglaatbaar zonder dat de betekenis van de zin drastisch verandert. Het gaat vooral om de werkwoorden *hebben*[12] (28 keer), *zien* (22 keer) en *kennen* (19 keer).[13] Onder 22 tot 25 worden hiervan enkele voorbeeldzinnen gegeven.

(22) [E]n de Romeynen sagen dat sy teghen de Belgen, Duytschen, Brittoenen, jae teghen de gheheele werelt een groot behulp souden hebben aen de Bataviers.

(H. de Groot - *Tractaet vande oudtheyt vande Batavische nu Hollandsche republique*, 1610)

(23) Noch seit hy, sien wy dit uytdrucklijck aen de Vogeltjens, die nimmermeer ophouden het Lof Godes te singen.
(D.P. Pers - *Cesare Ripa's Iconologia of Uytbeeldinghen des Verstants*, 1644)

(24) 'T is, ghelijck-men zeght, den ezel kent-men aen zijn ooren.
(J. De Brune - *Emblemata of Zinne*-werck, 1636)

(25) Siet hier u onderdaan, die noyt aan u verdiende Het minste van u jonst.
(G.A. Bredero - *Griane*, 1616)

De acht andere betekenissen komen veel minder vaak voor, van minimaal 7 (0.5%) tot maximaal 41 (2.7%) attestaties (zie zinnen 26-32 voor voorbeeldzinnen van elke betekenis). Dit zijn m.a.w. meer perifere betekenissen van de *aan*-constructie.

(26) *betekenis 2*: Het hof van Saturnijn Heb ik ten dienst van u belooft aan Prozerpijn.
(J. Vos - *Aran en Titus of Wraak en weerwraak: treurspel*, 1641)

(27) *betekenis 3*: [D]at ghy in't sin hebt die twee Dochters aende armen Ridder te ontweldighen.
(G.H. van Breughel - *De tweede vijftigh lustige historien ofte nieuwigheden Johannis Boccatii*, 1644)

(28) *betekenis 4*: Tot eynd'lijck op de vlucht met yle voeten vlugh Den vyandt sich begheeft, en laet de grootste glory, Aen Maurits.
(G. van Hogendorp - *Truer-spel van de moordt, begaen aen Wilhelm by der gratie Gods, prince van Oraengien*, 1617)

(29) *betekenis 5*: Nimmermeer weygerde hy iets aen die geene, die 't van hem versochten.
(D.P. Pers - *Cesare Ripa's Iconologia of Uytbeeldinghen des Verstants*, 1644)

(30) *betekenis 6*: De burgherwacht zeid'er neen toe; en verwillighde aan hem en zyn' aadelborsten, mits dat hun geweer daar bleeve, naauwlyx ingank.
(P.C. Hooft - *Nederlandsche Historien*, 1642-1647)

(31) *betekenis 7*: Op die voorwaarde; dat ghy gantschelyck an myn, myn ongehoorsaamheyt … sult vergeven.
(G.A. Bredero - *Griane*, 1616)

(32) *betekenis 9*[14]: Gelyk nu de volkrykste plaatsen, aan de zwaarste beweeghenissen, onderworpen zyn.
(P.C. Hooft - *Nederlandsche Historien*, 1642-1647)

In de 1530 *tokens* van de *aan*-constructie troffen we 291 verschillende werkwoorden aan. Bij deze *types* waren er 144 hapaxen. Dat wijst erop dat de constructie met een zeer diverse groep werkwoorden gecombineerd kon worden en dus duidelijk productief was. Het lijkt er zelfs op dat de constructie nog iets breder toepasbaar was dan vandaag: we vinden immers ook enkele benefactieven[15] (33) en possessieven (34). In het hedendaagse Nederlandse is de *aan*-constructie in die contexten onmogelijk; andere voorzetsels worden dan gebruikt (resp. *voor* en *bij*). Het gaat echter maar om een klein aantal voorbeelden, dus erg ingeburgerd zijn deze betekenissen niet.

(33) O liefde die de poorte doed open aen mijn wensch, en laet mijn driften in.[16]
(J. van Heemskerck - *De* [*verduytste*] *Cid*, 1641)

(34) Laet my toe te trouwen die aen Rodrigo sal de kop af konnen houwen.
(J. van Heemskerck - *De* [*verduytste*] *Cid*, 1641)

Daarnaast zijn er talloze voorbeelden van zinnen waarvoor de *aan*-constructie in het hedendaagse Nederlands geen optie meer is, maar die in tegenstelling tot de benefactieven en possessieven wel duidelijk in een van de negen onderscheiden betekeniscategorieën passen, bijvoorbeeld *ontdekken* in de betekenis 'laten blijken' (zin 35), *bidden* in de betekenis 'met aandrang vragen' (zin 36) en *verzoeken* in combinatie met een direct object in de vorm van een nominale constituent (zin 37). Veranderingen op het lexicale niveau zorgen ervoor dat die werkwoorden niet langer op die manier in de *aan*-constructie gebruikt kunnen worden – zie Colleman (2011) over het verschil tussen lexicale en constructionele verschuivingen, en ook verder, 6.2.1 en 6.2.2.

(35) Lodewijck heeft aen Vrouw Beatricks zijn Meestersse de liefde ontdeckt die hy haer droech.
(G.H. van Breughel - *De tweede vijftigh lustige historien ofte nieuwigheden Johannis Boccatii*, 1644)

(36) [A]en welcken Heere, Martelin, nae dat hy voor hem gecomen was, gebeden heeft, dat hy hem uyt sonderlingen ghenaede wilde toch laten wech gaen.
(G.H. van Breughel - *De tweede vijftigh lustige historien ofte nieuwigheden Johannis Boccatii*, 1644)

(37) Den Spaenschen Koning Ferdinandus versocht door-tocht voor zijn leeger aen Koninck van Navarre Jean d'Albert.
(A. Valerius - *Nederlandtsche gedenck-clanck*, 1626)

6 Semantische verschuivingen: vergelijking met de 20ste eeuw

6.1 Semantisch netwerk

Om kwantitatieve verschuivingen op te sporen in het betekenisnetwerk van de *aan*-constructie, werden de 2129 *tokens* van de *aan*-constructie uit de periode 1900-1950 op dezelfde manier gelabeld volgens de verschillende onderscheiden betekenissen, zie de rechterkolom in Tabel 2. Voor de 20ste eeuw werden er 286 verschillende werkwoorden in de constructie aangetroffen, waarvan 132 hapaxen. Om de betrouwbaarheid van de semantische labeling te testen, werd een random selectie van 500 zinnen uit elke periode ook door de tweede auteur gelabeld, uiteraard zonder inzage in de labeling door de eerste auteur. Er was een inter-rater agreement van κ=0.79, wat wijst op een uitstekende overeenkomst (zie Carletta, 1996 voor meer informatie over de kappascore).[17] De cijfers die in Tabel 2 gepresenteerd worden, zijn de uiteindelijke cijfers bekomen na individueel overleg over de zinnen die door beide auteurs aanvankelijk verschillend waren geanalyseerd.

Tabel 2 Verhouding tussen de betekenissen van de *aan*-constructie in 1600-1650 vs. 1900-1950

Betekenisrelatie	1600-1650 (n = 1530)	1900-1950 (n = 2129)
1. Su handelt om contactrelatie (tussen DO en *aan*-object) te doen ontstaan	1214 (79.3%)	1486 (69.8%)
1.1 Concrete transfer	342 (28.2%)	216 (14.5%)
1.2 Abstracte transfer	282 (23.2%)	670 (45.1%)
1.3 Handeling	241 (19.9%)	172 (11.6%)
1.4 Communicatieve overdracht	349 (28.7%)	428 (28.8%)
2. Su belooft of biedt mogelijkheid voor ontstaan contactrelatie t ussen DO en *aan*-object	41 (2.7%)	41 (1.9%)
3. Su handelt om contactrelatie tussen *aan*-object en DO te verbreken	8 (0.5%)	37 (1.7%)
4. Su handelt niet om contactrelatie tussen *aan*-object en DO te verbreken	8 (0.5%)	14 (0.7%)
5. Su verhindert ontstaan contact tussen DO en *aan*-object	12 (0.8%)	16 (0.7%)
6. Su verhindert ontstaan contact DO – *aan*-object niet	7 (0.5%)	2 (0.1%)

Betekenisrelatie	1600-1650 (n = 1530)	1900-1950 (n = 2129)
7. Su bevestigt een contactrelatie of neemt een houding in tegenover ontstaan van contactrelatie of bestaande contactrelatie	22 (1.4%)	44 (2.1%)
8. *Aan*-object geldt als uitgangspunt voor contact Su – DO	163 (10.7%)	355 (16.7%)
9. *Aan*-object geldt als norm/invloed waar DO door Su in contact mee komt	12 (0.8%)	45 (2.1%)
10. Restcategorie	43 (2.8%)	89 (4.2%)

Een chikwadraattest toont aan dat het verschil in distributie van de verschillende betekenissen over de twee periodes significant is (χ^2 = 63.54, df = 8, p<.001). Op basis van de residuen[18] zien we dat er met name drie categorieën zijn die sterk aan dat overkoepelende verschil bijdragen: betekenissen 8, 3 en 9 komen significant vaker voor in de periode 1900-1950. De daling bij de eerste (centrale) betekenis is secundair, d.w.z. een gevolg is van een algemene stijging van de frequentie van andere betekenissen. De genormaliseerde frequentie per miljoen woorden is niet significant veranderd (χ^2 = 2.06, df = 1, p>.1). De verhouding tussen de verschillende betekenissen *binnen* de eerste groep verschilt wel significant als we de twee perioden tegenover elkaar plaatsen (χ^2 = 180.58, df = 3, p<.001). Bij de communicatieve transfers is er geen significant verschil, bij de andere groepen wel: concrete transfers en handelingen komen vaker voor in 1600-1650 dan verwacht, abstracte transfers minder vaak. Hoewel de *aan*-constructie in de eerste helft van de 17de eeuw geenszins beperkt was tot concrete transfers, blijkt het aandeel van die concrete transfers wel duidelijk af te nemen doorheen de tijd. Dat vormt dus enigszins bevestiging voor de hypothese uit de inleiding dat de *aan*-constructie gaandeweg voor meer abstracte betekenissen gebruikt zou worden.

Een volgende stap in de analyse is om na te gaan of we de hierboven aangehaalde verschillen kunnen linken aan specifieke (groepen van) werkwoorden die zich anders zijn gaan gedragen in de loop van de drie eeuwen. Hiervoor kunnen we gebruikmaken van een diachrone collostructionele analyse (Hilpert 2006), waarvan we de resultaten in paragraaf 6.2 rapporteren.

6.2 Diachrone collostructionele analyse

Diachrone collostructionele analyse (Hilpert, 2006) is een aanpassing op de distinctieve collexeemanalyse, een statistische techniek ontwikkeld door Gries en Stefanowitsch (2004) die het mogelijk maakt om de 'collostructionele' voorkeuren van individuele lexemen te identificeren, d.w.z. om na te gaan welke lexicale items een significante voorkeur hebben voor een van

de twee (of meer) concurrerende grammaticale constructies boven de andere. Zo werd de methode door hen onder meer ingezet om te bepalen welke (groepen van) werkwoorden significant werden aangetrokken tot de Engelse DOC dan wel tot de *to-dative*, of welke woorden de *s-genitive* dan wel de genitiefconstructie met *of* verkiezen. De methode werd door Hilpert (2006) uitgebreid naar de vergelijking van historische perioden: de bedoeling is dan om de voorkeur van bepaalde lexicale elementen voor een periode te bepalen in een vergelijking van twee of meer historische perioden, in plaats van voor een bepaalde constructie binnen één periode. Een significante voorkeur voor één periode kan een indicatie zijn van een semantische verandering: 'Systematic differences in the collocational preferences can be interpreted as an ongoing change in the constructional semantics' (Hilpert, 2006, pp. 243). Op basis van een diachrone collostructionele analyse van de werkwoorden die voorkomen bij het Engelse hulpwerkwoord *shall* in drie verschillende periodes (1500-1640, 1640-1780 en 1780-1920), concludeert Hilpert (2006, pp. 251-254) dat *shall* een proces van subjectivering heeft ondergaan, iets wat eerder ook al in de literatuur geponeerd werd door o.m. Traugott (1989).[19]

Tabel 3 hieronder presenteert de resultaten van een diachrone collostructionele analyse[20] voor 1600-1650 vs. 1900-1950 en geeft dus aan welke werkwoorden significant vaker voorkomen in de *aan*-constructie in een van beide periodes. De verhouding tussen haakjes naast elk werkwoord geeft aan hoeveel keer dat werkwoord in elke periode geattesteerd is; voor *doen* is dat bijvoorbeeld 118 keer in 1600-1650 tegenover slechts 24 keer in 1900-1950. Het getal ernaast (in het geval van *doen*: 24.17) is een maat voor de collostructionele sterkte[21] van dat werkwoord: hoe hoger het getal, hoe distinctiever dat werkwoord voor die periode is. In totaal zijn er 44 collexemen distinctief voor 1600-1650 (met 95% zekerheid), en 37 voor 1900-1950. Tabel 3 hieronder geeft voor elke periode de collexemen met een collostructionele sterkte groter dan 3, d.w.z. die distinctief zijn voor die periode met 99.9% zekerheid.

Tabel 3 Resultaten van een diachrone collostructionele analyse van de *aan*-constructie

1600-1650 (n = 1530)				1900-1950 (n = 2129)			
doen (118:24)	24.17	*bewijzen* (21:5)	4.26	*(een einde) maken* (0:55)	13.06	*onttrekken* (0:22)	5.19
verzoeken (42:5)	10.98	*zoeken* (13:1)	4.00	*overlaten* (1:60)	12.84	*vertellen* (7:39)	4.01
zenden (48:12)	8.90	*schikken* (10:0)	3.80	*hebben* (42:166)	10.99	*veranderen* (0:15)	3.54
slaan (27:2)	8.12	*begaan* (14:2)	3.66	*wijden* (0:45)	10.66	*opdragen* (2:21)	3.25
schenden (18:0)	6.84	*betonen* (12:1)	3.65	*vragen* (10:83)	10.52	*toevertrouwen* (2:20)	3.05
bevelen (17:0)	6.46	*schrijven* (49:31)	3.51	*ontlenen* (0:42)	9.95		
vereren (18:1)	5.78	*leveren* (15:3)	3.43	*danken* (0:39)	9.23		
plegen (16:1)	5.07	*kennen* (19:6)	3.28	*toevoegen* (2:48)	9.02		
bedrijven (12:0)	4.56	*antwoorden* (8:0)	3.03	*toekennen* (0:32)	7.57		

Aan de hand van deze tabel kunnen we enkele vaststellingen doen omtrent semantische verschuivingen in de *aan*-constructie, die we ook kunnen linken aan de verschillen die in de labeling a.d.h.v. het semantisch netwerk naar voren zijn gekomen. Meer specifiek kunnen we twee types semantische verschuivingen onderscheiden, die hierna besproken worden in 6.2.1 en 6.2.2.

6.2.1 Shift in lexicale semantiek

Een van de distinctieve collexemen voor de eerste periode (1600-1650) is *schikken* (zie voorbeeld 38). Dat werkwoord werd tien keer aangetroffen in de *aan*-constructie in die periode, maar werd niet meer als dusdanig gebruikt in 1900-1950.

(38) Ten welke eynde zy beslooten hadden, binnen elf oft twaalf daaghen, wat min oft meer, ontrent vyftienhondert mannen van waapenen, uit de hunne aan haare Hoogheit te schikken. (P.C. Hooft - *Nederlandsche Historien*, 1642-1647)

Het werkwoord *schikken* wordt hier gebruikt in de betekenis 'zenden/sturen', een betekenis die niet meer standaardtalig is in het huidige Nederlands. We zien dan ook dat in de 20ste eeuw *schikken* niet meer voorkomt

in de *aan*-constructie. Dat is niet omdat de constructie niet meer compatibel is met 'zenden'-werkwoorden (*zenden* zelf wordt bijvoorbeeld nog 12 keer gebruikt met de *aan*-constructie in 1900), maar omdat het werkwoord *schikken* die 'zenden'-betekenis heeft verloren en in zijn andere, resterende betekenissen niet verenigbaar is met de *aan*-constructie.

6.2.2 Shift in constructionele semantiek

Bij het tweede, vanuit constructiegrammaticaal perspectief interessanter type van betekenisverschuiving, doet er zich wel degelijk een verandering voor in de semantische klassen van werkwoorden die in de constructie voorkomen. In Tabel 3 zien we een cluster van werkwoorden met als centrale betekenis 'doen/bedrijven' die consistent significant vaker voorkomen in de *aan*-constructie in de periode 1600-1650 dan in de periode 1900-1950: *doen* (118:24), *plegen* (16:1), *bedrijven* (12:0), *bewijzen* (21:5) en *begaan* (14:2), zie 39-41 voor enkele voorbeelden. De mogelijkheid om in de *aan*-constructie te verschijnen is voor die werkwoorden in de hedendaagse standaardtaal veelal beperkt tot een paar vaste combinaties: *afbreuk/onrecht doen aan*, *een voorstel/oproep/suggestie doen aan, dank/eer bewijzen aan* enz., zie voorbeeldzinnen 42-43. Dat bevestigt de vaststelling uit paragraaf 5.1: een chikwadraattest bij Tabel 2 toonde aan dat betekenis 1.3 (de handelingen) frequenter was in de periode 1600-1650 dan in de periode 1900-1950.

(39) [B]eveelende den soldaat, de punten der krystucht naa te koomen, zonder eenighen ooverlast aan geestelyken, oft weirlyken, te pleeghen.
(P.C. Hooft - *Nederlandsche Historien*, 1642-1647)

(40) En hier door vinden sich de Koningen en Princen des Werrelts bewogen, om gehoorsaemheyt te bewijsen aen haeren Roomschen Opper-Priester.
(D.P. Pers - *Cesare Ripa's Iconologia of Uytbeeldinghen des Verstants*, 1644)

(41) De schennis en 't geweld, begaen, in Ammons kamer, Aen d'allerkuischte ziel, mijn eige dochter, Thamer.
(J. van den Vondel - *Gebroeders*, 1640)

(42) 'k Weet niet aan wien hij die belofte deed, maar ik geloof [...].
(M. Böhtlingk - *Astrid*, 1920)

(43) [I]n het bijzonder waar het geldt eer te bewijzen aan een lid van het Oranjehuis.
(A.C.J. de Vrankrijker - *Vier eeuwen Nederlandsch studentenleven*, 1939)

Een tweede semantische klasse met een distinctieve voorkeur voor 1600-1650 wordt gevormd door vier werkwoorden met als centrale betekenis 'zenden/sturen': *zenden* (48:12) en *schikken* (10:0), allebei met p<.001, en *afvaardigen* (7:0) en *dragen* (6:0) met p<.01.[22] Net zoals bij de 'doen'-klasse is hier sprake van een kwantitatieve tendens: het gewicht van deze cluster van werkwoorden is afgenomen in de 20ste eeuw ten opzichte van de 17de eeuw. Terwijl dat bij de 'doen'-klasse gepaard gaat met een inperking van de mogelijke direct objecten, is die inperking er bij de 'zenden'-klasse niet. De afzwakking van 'zenden'-werkwoorden in de *aan*-constructie kan misschien in verband worden gebracht met het feit dat *aan* vroeger nog veel dichter bij de voorzetsels *naar* en *tot* stond qua semantiek (cf. WNT, s.v. *aen*: 'In het Mnl. strekte deze tweede beteekenis [i.e. de richtingsbetekenis, TG&TC] zich verder uit dan tegenwoordig, zóó zelfs dat *aen* bijna geheel met *naar* of *tot* gelijkstond'). Mogelijk heeft deze groep werkwoorden ook zijn aandeel in het overwicht van concrete transfers in 1600 ten opzichte van 1900. Vaak is het direct object bij deze werkwoorden immers een persoon of tastbaar voorwerp (zie voorbeeldzinnen 44 en 45).

(44) […] dat zy een stucksken van haer vingher gesneden hebben, en het zelve, in den brief gesloten, aen haer Iuffrouw ghezonden hebben.
(J. de Brune - *Emblemata of Zinne-werck*, 1636)

(45) Draeght dees schoone goude Beecker Aen Alcumeen u Vrou.
(I. Van Damme - *M. Accii Plauti Amphitruo*, 1635)

Tot slot bevestigt de collexeemanalyse nog enkele andere observaties die we bij de vergelijking a.d.h.v. het semantisch netwerk al opmerkten. Zo zijn de werkwoorden *hebben*, *danken* en *ontlenen* distinctief voor 1900-1950, wat het grotere aandeel zinnen in betekenis 8 (*aan*-object als uitgangspunt voor contact tussen subject en direct object) verklaart. *Onttrekken* (0:22) zorgt er dan weer voor dat de derde betekenis frequenter is in 1900-1950. Dat werkwoord komt in 1600-1650 nog uitsluitend in de DOC voor (zie 46), maar wordt drie eeuwen later al frequent in de *aan*-constructie gebruikt.[23]

(46) Op hoop dat s' uyt mee-doogh de schicht ruckt uyt dit hart, Die my de siel ontreckt.
(K.V. Dusart - *Spaensche heydin*, 1644)

Op die manier ontstaat een gemengd beeld over de semantische evolutie van de *aan*-constructie. Aan de ene kant breiden bepaalde gebruikswijzen zich uit, vooral de abstractere gebruikswijzen zoals betekenissen 8, 3 en 9

uit het betekenisnetwerk. Aan de andere kant daalt het aandeel van concrete transfers in de *aan*-constructie, en ook andere (groepen van) werkwoorden komen minder vaak voor in de 20ste eeuw i.v.m. de 17de eeuw, meest opvallend de 'doen'- en 'zenden'-klasse.

6.3 Configural Frequency Analysis

In deze laatste paragraaf willen we de semantiek van de *aan*-constructie nog gedetailleerder beschrijven aan de hand van een bottom-up semasiologisch profiel. Hiervoor bekijken we niet het volledige semantische bereik van de *aan*-constructie, maar focussen we op het deel dat overlapt met de DOC, m.a.w. waarin de rol van het *aan*-object vergelijkbaar is met de rol van het IO in de DOC: *Recipient/Addressee/Beneficiary/Possessor/Experiencer* (De Schutter, 1974, 1993). De zinnen die vallen onder groep 8 en 9 uit het betekenisnetwerk vallen daardoor weg. Die beperking is vooral pragmatisch gemotiveerd: op deze manier konden we voor dit verkennend onderzoek maximaal voortbouwen op de set van semantische paramaters die in Colleman (2006) werden geformuleerd voor de dubbelobjectconstructie. Door die gedetailleerde labeling zullen we ook zicht krijgen op de contexten waarin *aan* specifiek de datieffunctie in de DOC is komen te vervangen. *Configural Frequency Analysis* (CFA) is een techniek om combinaties van kenmerken (zgn. configuraties) te identificeren die samen vaker of minder vaak voorkomen dan verwacht op basis van toeval. We spreken dan respectievelijk van types en antitypes. Zie von Eye (2002) voor een bespreking van de techniek, en o.m. Stefanowitsch en Gries (2005, 2008), Schmidtke-Bode (2009) en Hilpert (2009) voor enkele taalkundige casestudies.

Voor deze analyse labelden we 1000 willekeurige zinnen uit elke periode op vijf semantische parameters, zie Tabel 4.

Tabel 4 Overzicht semantische parameters voor de *Configural Frequency Analysis*

Semantische parameter	Omschrijving
1. Causatie	Causale relatie tussen handeling van subject op het direct object en het ontstaan/verloren gaan van contactrelatie tussen het DO en het *aan*-object
1.1 Direct	Handeling subject zorgt voor ontstaan van contactrelatie (tussen DO en *aan*-object of subject en DO) bv. *geven*, *zeggen*
1.2 Indirect	Handeling subject volstaat niet voor ontstaan van contactrelatie bv. *bieden* (*aan*-object kan overdracht weigeren), *beloven* (toekomstige transfer)
1.3 Benefactief/malefactief	Subject treft voorbereidingen voor overgang of doet iets waarbij *aan*-object als begunstigde geldt bv. *voorschieten, goedmaken*
1.4 Passief	Subject is een eerder passieve medespeler bv. *toelaten*, *toestaan*
1.5 Attributioneel/attitudineel	Subject bevestigt een contactrelatie of neemt een houding in tegenover ontstaan van contactrelatie of bestaande contactrelatie bv. *toeschrijven*, *gunnen*
1.6 Preventie	Subject verhindert overdracht (van subject naar *aan*-object of van *aan*-object naar subject) bv. *weigeren*, *laten*
2. Contact	
2.1 Ontstaat	Er ontstaat contact tussen DO en *aan*-object
2.2 Verbroken	Contact tussen DO en *aan*-object wordt verbroken
3. Aard subject	
3.1 Bezield	Personen, dieren, organisaties, politieke partijen ...
3.2 Onbezield	Zaken
4. Aard direct object	
4.1 Bezield	Personen, dieren, organisaties, politieke partijen ...
4.2 Concreet onbezield	Tastbaar (bv. *een boek*)
4.3 Abstract onbezield	Niet-tastbaar (bv. *aandacht*)
4.4 Communicatief	Boodschappen (bv. *het verhaal*)
4.5 Deverbale handeling	bv. *een kus geven aan iemand*
5. Aard *aan*-object	
5.1 Bezield	Personen, dieren, organisaties, politieke partijen ...
5.2 Onbezield	Zaken

We onderwierpen de 2000 gelabelde zinnen aan een CFA.[24] Als we kijken naar configuraties met alle vijf semantische variabelen, en een extra variabele 'tijd' om diachrone vergelijking mogelijk te maken, dan vinden we dertien types en drie antitypes, zie Tabel 5 voor een overzicht, gerangschikt volgens dalende *effect size Q* (zie Gries 2009, p. 249). De drie antitypes staan cursief.

Tabel 5 Types en antitypes van de CFA voor 1600-1650 en 1900-1950

Configuratie							Freq.	Freq.
#	Causatie	Contact	Aard subject	Aard DO	Aard *aan*-object	Tijd	(geobs.)	(verw.)
1	*direct*	*ontstaan*	*bezield*	*abstract*	*bezield*	*1900*	104	207
2	direct	ontstaan	bezield	comm.	bezield	1900	169	84
3	direct	ontstaan	bezield	comm.	bezield	1600	155	84
4	direct	ontstaan	bezield	abstract	onbezield	1900	151	97
5	direct	ontstaan	bezield	concreet	bezield	1600	128	82
6	direct	ontstaan	onbezield	abstract	onbezield	1900	58	11
7	direct	ontstaan	bezield	bezield	bezield	1600	83	41
8	*direct*	*ontstaan*	*bezield*	*comm.*	*onbezield*	*1600*	4	40
9	passief	verbroken	bezield	abstract	onbezield	1900	24	0
10	direct	ontstaan	onbezield	abstract	onbezield	1600	27	11
11	*direct*	*verbroken*	*bezield*	*abstract*	*bezield*	*1600*	0	12
12	passief	verbroken	bezield	abstract	bezield	1600	10	1
13	preventie	verbroken	bezield	abstract	bezield	1600	7	0
14	passief	verbroken	bezield	abstract	bezield	1900	7	1
15	passief	verbroken	onbezield	abstract	onbezield	1900	6	0
16	preventie	verbroken	bezield	concreet	bezield	1900	4	0

Allereerst dient opgemerkt dat *Configural Frequency Analysis* vooral een exploratieve techniek is: alle mogelijke combinaties van de variabelen worden getest op de aan- of afwezigheid van een effect en dat zorgt ervoor dat het niet mogelijk is om bij alle (anti)types zinvolle veralgemeningen te maken. De geobserveerde frequentie van bepaalde configuraties (in ons geval: configuraties 8 t.e.m. 16) is soms (erg) laag, maar wijkt toch significant af van de verwachte frequentie, waardoor ze als (anti)type gelden. Configuratie 11 bijvoorbeeld geldt als antitype in de eerste periode: ze komt er in het geheel niet voor, en dat is een significant verschil met de verwachte 12 attestaties ($p<.05$).[25] De configuratie komt eenmaal voor in 1900-1950 (zie voorbeeldzin 47), maar geldt voor die periode niet als antitype, omdat het verschil met het aantal verwachte attestaties niet significant is. Verdere conclusies kunnen hier dan ook niet aan vastgeknoopt worden.

(47) Wie zulk een tweegevecht begeert, wil aan een ander iets ontnemen, namelijk zijn eer.
(J. Huizinga, *Herfsttij der Middeleeuwen*, 1919)

Daarnaast, doordat de verwachte frequentie wordt berekend aan de hand van de totalen van alle variabelen (zie eindnoot 25 voor meer uitleg en een uitgewerkt voorbeeld), dus inclusief de variabele tijd, is het mogelijk dat een bepaalde configuratie een type of antitype is voor beide periodes. Dat geldt bijvoorbeeld voor configuraties 2 en 3, de communicatieve overdracht tussen twee personen, die in beide periodes frequent voorkomt en daarom als type geldt. De echt interessante configuraties zijn dus die die als type/antitype gelden in één periode maar dat niet zijn in de andere periode. Daarvan lichten we er in wat volgt enkele in meer detail toe.

Configuratie 1 is een antitype voor de periode 1900-1950; we vinden deze configuratie 104 keer in 1900-1950, en 201 keer in 1600-1650. Deze combinatie van kenmerken komt statistisch gezien minder vaak voor in de 20ste eeuw dan de verwachte 207 attestaties ($p<.001$). Het gaat om zinnen zoals:

(48) Terwijl Egmont van zijn verblijf in Spanje vertelde, besteedde de prins zijn aandacht aan Antonius Perrenot.
(M. Dekker - *Oranje en de opstand der Nederlanders*, 1935)

(49) [E]lk ingeschrevene moest [...] toezeggen aan Rector, Curatoren en Senaat alle eer en onderdanigheid te zullen bewijzen.
(A.C.J. de Vrankrijker - *Vier eeuwen Nederlandsch studentenleven*, 1939)

Het zijn vooral de werkwoorden *geven* (32 keer), *doen* (30 keer), *verzoeken* (18 keer), *plegen* (12 keer), *bewijzen* (12 keer) en *betonen* (9 keer) die ervoor zorgen dat die configuratie zoveel frequenter is in 1600-1650. We zien dus dat een groep werkwoorden, nl. de 'doen'-klasse, waarvan we al bij het betekenisnetwerk en de diachrone collostructionele analyse aangetoond hebben dat ze frequenter voorkomen in 1600-1650, er ook hier mede voor zorgt dat we een antitype in 1900-1950 krijgen.

Configuraties 4, 5, 6 en 7 kunnen in verband worden gebracht met de observatie in paragraaf 6.1 dat de *aan*-constructie oorspronkelijk vaker werd ingezet bij concrete transfers. Er zijn twee 'concrete' types voor de periode 1600-1650 met een overdracht van een concreet onbezield (configuratie 5, voorbeeldzin 50) of bezield (configuratie 7, voorbeeldzin 51) direct object tussen twee bezielde personen. Daarnaast zijn er voor de tweede periode twee 'abstractere' configuraties terug te vinden: configuratie 4 met een abstract direct object en een onbezield *aan*-object (vaak min of meer vaste combinaties als *een einde maken aan iets*, *geloof/uiting geven aan iets*, *waarde hechten aan iets*, zie voorbeeldzin 52), en configuratie 6[26] waarbij alle drie de medespelers abstracta zijn (voorbeeldzin 53).

(50) […] houdende in de rechter hand een cristallijne schaele vol roode wijn, die zy met een soete bevalligheyt en aengenaeme zeedigheydt, aen een ander overlevert.
(D.P. Pers - *Cesare Ripa's Iconologia of Uytbeeldinghen des Verstants*, 1644)

(51) De Paus confirmeert wederom sijn voorige Conspiratie, ende twee Ambassadeus aen Lowisa gesonden hebbende, biet haer aen alle hulpe ende bystant.
(P. Nootmans - *Van den bloedigen slach van Pavyen*, 1627)

(52) Desondanks (zoo constateerde de synode) hadden de deputaten der synode-Duursema geweigerd, medewerking te verleenen aan een poging tot wederzijdsche, summiere, ondubbelzinnige, schriftelijke situatieteekening.
(K. Schilder - *Jaaroverzicht*, 1946-49)

(53) Maar dit standpunt levert inderdaad de waarheid over aan de subjectieve stellingkeuze van de afvallige persoonlijkheid.
(H. Dooyeweerd - *De wijsbegeerte der wetsidee*, 1935)

Tot slot lijkt de infrequentie van configuraties voor zinnen met een preventief of passief betekenismoment, en de afwezigheid van zinnen met een indirect, benefactief/malefactief of attributioneel/attitudineel betekenismoment te suggereren dat er op dat gebied weinig lijkt te zijn veranderd doorheen de eeuwen.

7 Besluit

Het doel van dit artikel was tweeledig: via een gedetailleerde corpusstudie wilden we (i) de stelling in de literatuur toetsen dat de *aan*-constructie echt doorbreekt als alternatief voor de dubbelobjectconstructie in de 17de eeuw, en (ii) de semantische evolutie van de constructie in kaart brengen a.d.h.v. een vergelijking met data voor de 20ste eeuw.

In eerste instantie werden alle attestaties van de *aan*-constructie uit de periode 1600-1650 opgezocht. De constructie was op dat moment duidelijk al frequent (genormaliseerde frequentie 997 p.m.w. tekst) en werd met een groot aantal verschillende werkwoorden gebruikt (291 *types*). Alle *aan*-constructies werden vervolgens gelabeld a.d.h.v. negen onderscheiden betekenissen uit het betekenisnetwerk. Die analyse toonde aan dat de *aan*-constructie aan het begin van de zeventiende eeuw opmerkelijk genoeg al het volledige semantische terrein van de hedendaagse *aan*-constructie be-

streek. Het lijkt er dus op dat de doorbraak al vroeger op gang is gekomen dan gesuggereerd in de literatuur. Er bleek in de periode 1600-1650 ook duidelijk geen overwicht van of beperking tot concrete transfers. De hypothese dat de *aan*-constructie vanuit haar metaforische origine oorspronkelijk vooral ingezet zou zijn bij transfers van een fysiek voorwerp tussen twee bezielde medespelers, om vandaaruit geleidelijk uit te breiden naar abstractere betekenissen, zal verder getoetst moeten worden in het 16de-eeuwse Nederlands, m.a.w. de periode van het eigenlijke ontstaan van de constructie.

Daarnaast leverde een vergelijking met de periode 1900-1950 een gemengd beeld op over de semantische evolutie van de *aan*-constructie sinds de 17de eeuw. Hoewel het niet zo is dat er bepaalde betekenissen uit het semantische netwerk volledig verdwenen zijn, zijn de mogelijkheden binnen enkele klassen wel sterk ingeperkt (er is m.a.w. een *kwantitatieve* tendens). Een diachrone collostructionele analyse leerde dat veel werkwoorden uit de 'doen'- en 'zenden'-klasse veel gebruikelijker waren in de *aan*-constructie aan het begin van de 17de eeuw. Aan de andere kant zien we dat bepaalde betekenissen dan weer frequenter zijn geworden in de loop der tijd, namelijk het gebruik om standen-van-zaken te benoemen waarin het *aan*-object als uitgangspunt (betekenis 8) of norm/invloed (betekenis 9) geldt voor het contact tussen subject en direct object, en privatieve transfers (betekenis 3). Algemeen gesteld lijkt er een toename van meer abstracte betekenissen te zijn. Het betekenisnetwerk liet zien dat we inderdaad relatief gezien minder concrete transfers aantroffen in de periode 1900-1950 dan in 1600-1650, en uit een *Configural Frequency Analysis* bleek dat er voor de 20ste eeuw meer abstracte configuraties te vinden zijn.

Noten

1. Beide auteurs maken deel uit van de onderzoeksgroep GLIMS (*Ghent University research group on Linguistic Meaning and Structure*) van de vakgroep Taalkunde aan de Universiteit Gent; het hier gerapporteerde onderzoek is tot stand gekomen binnen het onderzoeksproject 'Variation and Change in constructional semantics: Argument structure constructions in varieties of Dutch', gefinancierd door het Bijzonder Onderzoeksfonds van de Universiteit Gent (BOF-project nr. 01N00912). We bedanken graag de twee anonieme beoordelaars voor hun nuttige commentaar bij een eerdere versie van dit artikel.
2. Voor een overzicht van de naamvalsvormen, zie bijvoorbeeld van der Horst (2008, p. 1353), gebaseerd op Moonen (1706).
3. Voor de 16de eeuw troffen Weijnen en Gordijn (1970) in hun materiaal slechts enkele

discutabele voorbeelden aan van de *aan*-constructie, zoals zinnen (i) en (ii) hieronder, waarin de status van de *aan*-constituent als object naar hun zeggen nog twijfelachtig is.
(i) Dus wil ick de wijssheit gaen besorgen
Aen onssen prochiaen tegen morgen
(*Een esbatement vande Schuyfman*, 1524)
(ii) Ick salt best vraegen aent volck, dat hier staat.
(*Een batement van den Preecker*, s.d.)

4. Er bestaat enige onzekerheid over zowel de datering als het auteurschap van de *Waernemingen*. Het werk wordt gedateerd tussen 1635 en 1638, maar verschijnt pas in 1723 als bijlage in de *Aenleiding tot de kennisse van het verhevene deel der Nederduitsche sprake* van Lambert ten Kate Hz. (1674-1731). Het valt voorlopig niet uit te sluiten dat Hooft, als hij inderdaad de auteur van de *Waernemingen* is, een constructie heeft gepropageerd die op dat moment nog *niet* algemeen ingeburgerd was. We laten het echter over aan toekomstig onderzoek om na te gaan in hoeverre de omschrijving met *aan* ook in andere 17de en vroeg-18de-eeuwse grammaticale geschriften al expliciet genoemd wordt als alternatief voor het nominale indirect object.
5. De mogelijkheden van die groep werkwoorden om in de dubbelobjectconstructie gebruikt te worden zijn in het hedendaagse Belgisch Nederlands en bepaalde dialecten van het Nederlands Nederlands breder, zie o.m. Colleman (2010b).
6. Naar objecten wordt gevraagd met 'wie' of 'wat', cf.:
Aan wie gaf hij een bos bloemen? Aan zijn vrouw.
Bij niet-animate objecten vindt er nog een transformatie tot voornaamwoordelijk bijwoord plaats:
Waaraan (< *'aan wat'*) *koppelden ze enkele wagons? Aan de locomotief.*
7. In het betekenisnetwerk werden dergelijke metaforische extensies enkel voor de eerste betekenis expliciet onderscheiden. Uiteraard zijn die ook relevant binnen de andere betekenissen, bv. *Hij beloofde een auto aan zijn zoon* (concreet) vs. *Hij beloofde een mooie toekomst aan zijn dochter* (abstract) voor de tweede betekenis (toekomstige transfers). Veel zinnen wijken m.a.w. in meerdere opzichten tegelijk af van de centrale betekenis; dat gegeven wordt geïntegreerd in zgn. *multidimensionale* voorstellingen van semantische structuren, zie bijvoorbeeld Geeraerts (1998) voor zo'n multidimensionale beschrijving van het Nederlandse indirect object, waarin ook de verschillende *soorten* betekenisextensies nog explicieter gemaakt worden.
8. Het corpus is beschikbaar voor onderzoeksdoeleinden; geïnteresseerden kunnen zich wenden tot de eerste auteur.
9. Er waren 43 zinnen (2.8%) waaraan geen semantisch label kon worden toegekend wegens (i) moeilijk te interpreteren, of (ii) moeilijk in te passen binnen een van de negen onderscheiden categorieën. Dat de restgroep relatief uitgebreid lijkt komt vooral door een kleine set van min of meer vaste combinaties zoals *de handen schenden aan iets*, *deel hebben aan iets* enz. die frequent in het corpus geattesteerd zijn. Omdat het om een beperkte set van combinaties gaat lijkt het ons gerechtvaardigd hiervoor geen extra betekenis in het betekenisnetwerk te voorzien.
10. Tot de concrete transfers werden gerekend: (i) zinnen waarin een menselijk *aan*-object het direct object letterlijk in handen (*een boek geven aan iemand*) of in zijn/haar macht krijgt (*een stad overgeven aan iemand*), en (ii) zinnen met een niet-menselijk *aan*-object waarin een situatie van letterlijk contact ontstaat (*een wagon koppelen aan de locomotief*).
11. Zowel de DOC als de *aan*-constructie kunnen gebruikt worden om situaties uit te drukken waarin het indirect object/*aan*-object het doelwit noemt van een (fysieke)

handeling, via de metafoor waarin acties gericht op een persoon begrepen worden als entiteiten die overgedragen worden naar die persoon (zie Goldberg, 1995, p. 149 voor de beschrijving van die metafoor en enkele Engelse voorbeelden). In feite kan deze categorie beschouwd worden als een subtype van de categorie 'abstracte transfer'.

12. Voor combinaties met *hebben* waarin het *aan*-object duidelijk een uitgangspunt/bron noemt, is een parafraseproef mogelijk: *iem. heeft een vriend aan iem. ~ iem. is een vriend voor iem.*, *iem. heeft plezier aan iets ~ iets is een plezier voor iem.*/*iets veroorzaakt plezier bij iem*. Die parafrases zijn onmogelijk voor combinaties waarin het *aan*-object nog veel moeilijker als een uitgangspunt/bron gezien kan worden, zoals *deel hebben aan iets*, *lak hebben aan iets* enz. Dergelijke gevallen werden dan ook niet in categorie 8 maar in de restcategorie ondergebracht.
13. Een reviewer merkt op dat in sommige van die contexten ook andere voorzetsels mogelijk zijn: bv. *Wij zien dit uitdrukkelijk bij/in/met de vogeltjes* en heeft dan ook bedenkingen bij de status van de *aan*-groep bij bv. *zien* als object. Echter, al willen we niet ontkennen dat we hier aan de grenzen van de categorie object zitten, er lijkt ons toch nog een verschil in inherentie tussen de *aan*-groep bij *zien* en de voorzetselgroepen met *bij*, *in* en *met*, die veeleer als locatieve bepaling lijken te mogen gelden (bv. *Waar zien we dat? Bij de vogeltjes.*). In de ontleding van de *aan*-groep bij *zien*, *horen*, *hebben*, enz. als object volgen we De Schutter (1974, pp. 204-207).
14. Bij deze categorie werden twee types werkwoorden ondergebracht, in navolging van Colleman (2006): (i) werkwoorden zoals *aanpassen* en *toetsen*, waarbij het *aan*-object geldt als (ideale) norm (zie ook De Schutter 1974, p. 206), en (ii) werkwoorden zoals *onderwerpen* en *blootstellen*, waarbij het direct object een invloed ondervindt van het *aan*-object.
15. Ook malefactieven zijn wellicht mogelijk in de *aan*-constructie op dat moment. We vonden in onze data weliswaar geen ditransitief voorbeeld, maar wel de volgende zin:
    (i) Alle lichten des hemels wil ick aen u laten doncker werden, ende ic wil eene duysternisse in uwen lande maken, spreeckt de Heere.
    (J. Cats - *Aenmerckinghe op de tegenwoordige steert-sterre*, 1618)
    Bovenstaande zin werd niet meegenomen als voorbeeld van de *aan*-constructie omdat het niet om de juiste syntactische context gaat. Het is echter aannemelijk dat sprekers die een *aan*-object konden gebruiken in een context als (i) dat ook konden in combinatie met transitieve werkwoorden: bv. *iets vernietigen aan iemand*, *een strik spannen aan iemand*, enz. Verder onderzoek moet die hypothese bevestigen.
16. Een reviewer vraagt zich af of (i) *aen myn wensch* hier niet geïnterpreteerd moet worden als 'in overeenstemming met mijn wens' en (ii) of het hier gezien de weglaatbaarheid wel om een benefactief gaat. Wat de tweede deelvraag betreft: een belanghebbend voorwerp is per definitie weglaatbaar, omdat het niet wordt opgeroepen door de valentie van het zinswerkwoord (bv. *Hij kocht* (*mij*) *een boek*, *Ze bakte* (*haar moeder*) *een taart*, enz.). Wat de eerste deelvraag betreft: de alternatieve interpretatie 'in overeenstemming met mijn wens' lijkt onwaarschijnlijk op grond van de context. Het is namelijk niet zo dat de liefde letterlijk een poort doet opengaan in overeenstemming met de wens van de spreekster (de Infante), de zin is veeleer metaforisch te interpreteren: de liefde opent de weg voor de driften en voor de wens van de spreekster om samen te zijn met haar verboden geliefde Don Rodrigo.
17. De agreement tussen de eerste auteur en een derde annotator die niet bij voorafgaand overleg over de semantische rubricering betrokken was, was κ=0.66, m.a.w. een goede overeenkomst.
18. Plus- en mintekens geven aan dat geobserveerde waarde respectievelijk groter of klei-

ner is dan de verwachte waarde; hoe groter het getal, hoe groter het verschil.

	Bet_1	Bet _2	Bet _3	Bet _4	Bet _5	Bet _6	Bet _7	Bet _8	Bet _9
1600-1650	2.2427	1.0933	-2.5190	-0.4187	0.0568	1.6456	-1.1044	-3.7482	-2.4543
1900-1950	-1.9148	-0.9334	2.1507	0.3575	-0.0485	-1.4050	0.9429	3.2001	2.0954

19. Die diachrone toepassing van de distinctieve collexeemanalyse is zeker niet onproblematisch. Stefanowitsch (2006) stelt dat, in tegenstelling tot grammaticale constructies, tijdsperiodes geen psychologisch autonome categorieën zijn waartussen een spreker kan kiezen. Dat was precies het uitgangspunt van een distinctieve collexeemanalyse. Bovendien kan een andere periodisering (bijvoorbeeld per 25 of per 100 jaar) de resultaten drastisch veranderen. De diachrone collostructionele analyse wordt daarom best aangevuld met een gewone collexeemanalyse per periode, of een andere methode die eventuele veranderingen in de semantiek van een constructie kan nagaan (Stefanowitsch, 2006, p. 259). In dit artikel vergelijken we de resultaten van de diachrone collostructionele analyse dan ook zorgvuldig met verschuivingen in het radiale betekenisnetwerk en met de CFA.
20. De analyse werd uitgevoerd in R, aan de hand van het script Coll.Analysis 3.2 (Gries 2007).
21. De collostructionele sterkte is een negatieve logaritmische transformatie van de p-waarde van de Fisher-Yates Exact Test die gebruikt wordt om de distinctiviteit van de collexemen te bepalen, cf. Gries (2007).
22. Ook *brengen* (22:21) heeft een voorkeur voor de periode 1600-1650, maar die voorkeur is statistisch niet significant. Toch lijken de mogelijkheden om in de *aan*-constructie gebruikt te worden voor *brengen* in de loop der tijd beperkter te worden. Zes van de 21 voorbeelden uit 1900-1950 hebben 'bezoek(je)' als direct object - dat is bij geen enkele zin in 1600-1650 zo, waar we nog veel vaker een concreet direct object aantreffen (zoals een beeldje, drank, een geschenk, een brief, wapens enz.). Die combinatie met een concreet direct object wordt infrequenter, maar aan de andere kant staat dat de combinatie met 'bezoek(je)' frequenter wordt, wat als nettoresultaat heeft dat het werkwoord *brengen* op zich niet significant vaker voorkomt in een van beide periodes. Dergelijke verschillen illustreren dat het lonend kan zijn om in collexeemanalyses meer rekening te houden met verbale polysemie, d.w.z. meer oog te hebben voor verschillen in constructionele voorkeur tussen de onderscheiden gebruikswijzen van een en hetzelfde werkwoord (zie daarover ook Bernolet en Colleman, ter perse).
23. Er lijkt ook lexicaal-semantisch iets aan de hand met dit werkwoord. De combinatie in (46) zou vandaag maar moeilijk in de *aan*-constructie passen: ??*de ziel aan iemand onttrekken*. Ook voor andere zinnen uit 1600-1650 geldt dat: ??*een gunst onttrekken aan*, ??*een oorzaak onttrekken aan*, enz.
24. Hiervoor werd het R-script HCFA 3.2 (Gries, 2004) gebruikt.
25. De verwachte frequentie wordt berekend door het product van alle totalen van een cel te delen door naantal variabelen - 1. Voor deze configuratie is dat 1768 (totaal aantal directe causatie) * 109 (verbreken contact) * 1457 (bezield subject) * 999 (abstract direct object) * 1361 (bezield *aan*-object) * 1000 (1600-1650) gedeeld door 20006-1.
26. Die configuratie is weliswaar ook een type voor de periode 1600-1650 (zie configuratie 10), maar het effect is daar minder sterk.

Bibliografie

Allen, Cynthia L, *Case marking and reanalysis. Grammatical relations from Old to Early Modern English.* Oxford: Oxford University Press, 1995.

Barðdal, Jóhanna, Kristoffersen, K.E. & Sveen, Andreas, 'West Scandinavian Ditransitives as a family of constructions: With a special attention to the Norwegian 'V-REFL-NP' construction'. *Linguistics,* 49 (2011): pp. 53-104.

Beliën, Maaike, 'Force Dynamics in Static Prepositions: Dutch *Aan*, *Op*, and *Tegen*'. In: Hubert Cuyckens & Gunter Radden (red.), *Perspectives on Prepositions.* Tübingen: Niemeyer, 2002, pp. 195-209.

Bergs, Alexander & Diewald, Gabriele, *Constructions and Language Change.* Berlin & New York: Mouton de Gruyter, 2008.

Bernolet, Sarah & Colleman, Timothy, 'Sense-based and lexeme-based alternation biases in the Dutch dative alternation'. In Jiyoung Yoon & Stefan Gries (red.), *CxG beyond English: current corpus-based approaches.* Amsterdam: John Benjamins, ter perse.

Broekhuis, Hans, 'Het voorzetselvoorwerp', *Nederlandse Taalkunde*, 9 (2004): pp. 97-131.

Carletta, Jean, 'Assessing agreement on classification task: the kappa statistic'. *Computational linguistics,* 22 (1996), pp. 249-254.

Colleman, Timothy, De Nederlandse datiefalternatie: een constructioneel en corpusgebaseerd onderzoek. Proefschrift Universiteit Gent, 2006.

Colleman, Timothy, 'Verb disposition in argument structure alternations: a corpus study of the dative alternation in Dutch', *Language Sciences,* 31 (2009), pp. 593-611.

Colleman, Timothy, 'Beyond the dative alternation: the semantics of the Dutch aan-dative'. In: Dylan Glynn & Kerstin Fischer (red.), *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches.* Berlin & New York: Mouton de Gruyter, 2010a, pp. 271-303.

Colleman, Timothy, 'Lectal variation in Constructional Semantics: Benefactive ditransitives in Dutch'. In: Dirk Geeraerts, Gitte Kristiansen & Yves Peirsman (red.), *Advances in Cognitive Sociolinguistics.* Berlin & New York: Mouton de Gruyter, 2010b, pp. 191-221.

Colleman, Timothy, 'Ditransitive verbs and the ditransitive construction: a diachronic perspective', *Zeitschrift für Anglistik und Amerikanistik,* 59 (2011): pp. 387-410.

Colleman, Timothy, 'Ditransitieve constructies in het Nederlands: semasiologische en onomasiologische kwesties', *Nederlandse Taalkunde,* 17 (2012): pp. 345-361.

Colleman, Timothy & De Clerck, Bernard, 'Caused motion? The semantics of the English to-dative and the Dutch aan-dative', *Cognitive Linguistics,* 20 (2009): pp. 5-42.

Colleman, Timothy & De Clerck, Bernard, 'Constructional semantics on the move: On semantic specialization in the English double object construction', *Cognitive Linguistics,* 22 (2011): pp. 183-209.

Cuyckens, Hubert. The semantics of spatial prepositions in Dutch. A cognitive-linguistic exercise. Proefschrift UIA, Antwerpen, 1991.

De Schutter, Georges, *De Nederlandse zin: poging tot beschrijving van zijn struktuur.* Brugge: De Tempel, 1974.

De Schutter, Georges, 'Towards a cognitive analysis of the NP-PP variation in the indirect object in Dutch'. *Leuvense Bijdragen,* 82 (1993): pp. 485-499.

De Smet, Hendrik, 'A corpus of Late Modern English texts', *ICAME-Journal,* 29 (2005): pp. 69-82.

Duinhoven, A.M., 'Actieve en passieve indirect objecten. Een kwestie van analyse en synthese', *Nederlandse Taalkunde,* 8 (2003): pp. 199-230.

Fried, Mirjam, 'Construction Grammar as a tool for diachronic analysis'. *Constructions and Frames,* 1 (2009): pp. 261-290.

Geeraerts, Dirk, *Diachronic Prototype Semantics.* Oxford: Clarendon Press, 1997.
Geeraerts, Dirk, 'The semantic structure of the indirect object in Dutch'. In: Willy Van Langendonck & William Van Belle (red.), *The Dative. Volume II: Theoretical and Contrastive Studies.* Amsterdam: John Benjamins, 1998, pp. 185-210.
Geeraerts, Dirk, Grondelaers, Stefan & Bakema, Peter, *The Structure of Lexical Variation. Meaning, Naming, and Context.* Berlin & New York: Mouton de Gruyter, 1994.
Geeraerts, Dirk, Grondelaers, Stefan & Speelman, Dirk, *Convergentie en divergentie in de Nederlandse woordenschat. Een onderzoek naar kleding- en voetbaltermen.* Amsterdam: Meertens Instituut, 1999.
Goldberg, Adele E., *Constructions: a construction grammer approach to argument structure.* Chicago: University of Chicago Press, 1995.
Goldberg, Adele E., 'Constructions: a new theoretical approach to language'. *TRENDS in Cognitive Sciences,* 7 (2003): pp. 219-224.
Gries, Stefan Th., HCFA 3.2. A program for R for Windows 2.x., 2004.
Gries, Stefan Th., Coll.analysis 3.2. A program for R for Windows 2.x, 2007.
Gries, Stefan Th., *Statistics for linguistics with R. A practical introduction.* Berlin & New York: Mouton de Gruyter, 2009.
Gries, Stefan Th. & Stefanowitsch, Anatol, 'Extending collostructional analysis: a corpus-based perspective on "alternations"', *International Journal of Corpus Linguistics,* 9 (2004): pp. 97-129.
Hilpert, Martin, 'Distinctive collexeme analysis and diachrony'. *Corpus Linguistics and Linguistic Theory,* 2 (2006), pp. 243-256.
Hilpert, Martin, *Germanic future constructions: A usage-based approach to language change.* Amsterdam: John Benjamins, 2008.
Hilpert, Martin, 'The German *mit*-predicative construction'. *Constructions and Frames,* 1 (2009), pp. 29-55.
Hoffmann, Thomas & Trousdale, Graeme, *The Oxford Handbook of Construction Grammar.* Oxford: Oxford University Press, 2013.
Horst, Joop van der, *Geschiedenis van de Nederlandse Syntaxis.* Leuven: Universitaire Pers Leuven, 2008.
Janssen, Theo, 'Giving in Dutch: an intra-lexematical and inter-lexematical description'. In: John Newman (red.), *The Linguistics of Giving.* Amsterdam: John Benjamins, 1997, pp. 267-306.
Kirsner, Robert S., Verhagen, Arie & Willemsen, Mariëtte, 'Over PP's, transitiviteit en het zgn. indirekt objekt', *Spektator,* 14 (1985): pp. 341-347.
McFadden, Thomas, 'The rise of the *to*-dative in Middle English'. In: David Lightfoot (red.), *Syntactic effects of morphological change.* Oxford: Oxford University Press, 2002, pp. 107-123.
Moonen, Arnold, *Nederduitsche Spraekkunst.* Amsterdam, 1706.
Noël, Dirk, 'Diachronic construction grammar vs. grammaticalization theory', *Functions of Language,* 14 (2007), pp. 177-202.
Rohdenburg, Günter, 'Functional constraints in syntactic change: The rise and fall of prepositional constructions in Early and Late Modern English', *English Studies,* 88 (2007), pp. 217-233.
Schermer-Vermeer, E.C., 'Het voorzetselvoorwerp en het begrip "vorm"', *De nieuwe Taalgids,* 83 (1990), pp. 238-247.
Schermer-Vermeer, E.C., *Substantiële versus formele taalbeschrijving : het indirect object in het Nederlands.* Amsterdam: University of Amsterdam, 1991.
Schermer-Vermeer, E.C., 'Grammatica, lexicon en de dubbelobject-constructie in het Nederlands en het Engels'. *Nederlandse Taalkunde,* 6 (2001), pp. 22-37.
Schermer-Vermeer, E.C., 'Worstelen met het voorzetselvoorwerp', *Nederlandse Taalkunde,* 11 (2006), pp. 146-167.

Schmidtke-Bode, Karsten, 'Going-to-V and gonna-V in child language: A quantitative approach to constructional development'. *Cognitive Linguistics,* 20 (2009), pp. 509-553.

Stefanowitsch, Anatol, 'Distinctive Collexeme Analysis and Diachrony: A comment', *Corpus Linguistics and Linguistic Theory,* 2 (2006), pp. 257-262.

Stefanowitsch, Anatol & Gries, Stefan Th., 'Covarying collexemes', *Corpus Linguistics and Linguistic Theory,* 1 (2005), pp. 1-43.

Stefanowitsch, Anatol & Gries, Stefan Th., 'Register and constructional meaning: A collostructional case study'. In: Gitte Kristiansen & René Dirven (red.), *Cognitive Sociolinguistics: Language variation, cultural models, social systems.* Berlin & New York: Mouton de Gruyter, 2008, pp. 129-152.

Traugott, Elizabeth C., 'On the rise of epistemic meanings in English: An example of subjectification in semantic change', *Language,* 57 (1989), pp. 33-65.

Traugott, Elizabeth C., '"All that he endeavoured to prove was...": On the emergence of grammatical constructions in dialogic contexts'. In: Robin Cooper & Ruth Kempson (red.), *Language in flux: Dialogue coordination, language variation, change and evolution.* London: Kings College Publications, 2008, pp. 143-177.

Traugott, Elizabeth C. & Trousdale, Graeme, *Constructionalization and Constructional Changes.* Oxford: Oxford University Press, 2013.

Van Belle, William & Van Langendonck, Willy, 'The indirect object in Dutch'. In: William Van Belle & Willy Van Langendonck (red.), *The Dative. Volume 1: Descriptive studies.* Amsterdam: John Benjamins, 1996, pp. 217-250.

Vandeweghe, Willy, 'Het voorzetselvoorwerp en de hiërarchie der objecten', *Nederlandse Taalkunde,* 16 (2011), pp. 88-101.

Visser, F.Th., *An historical syntax of the English language.* Leiden: Brill, 1963.

Von Eye, Alexander, *Configural Frequency Analysis. Methods, Models, and Applications.* Mahwah, New Jersey: Lawrence Erlbaum, 2002.

Weijnen, A.A. & Gordijn, A.M.J., 'Het binnendringen van het voorzetsel in het indirekt object in het Nederlands', *Leuvense Bijdragen,* 59 (1970), pp. 125-130.

WNT = De Vries, Matthias, L.A. Te Winkel et al. (red.), *Woordenboek der Nederlandsche Taal.* 's Gravenhage: Martinus Nijhoff, 1882-1998.

Zehentner, Eva. Of harming, helping, and handing: the case of Middle English ditransitives. Paper presented at the 47th Annual Meeting of the Societas Linguistica Europaea (SLE). Poznan, 2014.

Over de auteurs

Tim Geleyn, Universiteit Gent, Ghent University research team on Linguistic Meaning & Structure, Blandijnberg 2, B-9000 Gent.
E-mail: tim.geleyn@UGent.be

Timothy Colleman, Universiteit Gent, Ghent University research team on Linguistic Meaning & Structure, Blandijnberg 2, B-9000 Gent.
E-mail: timothy.colleman@UGent.be

I'm queen of the world!

(*Semi-*)*fixed English expressions and constructions in Dutch*

Eline Zenner & Dirk Geeraerts

MEM 67 (2):
DOI: 10.5117/TET2015.2.ZENN

Abstract

This paper analyses English multi-word insertions found in a Dutch corpus of naturally occurring spoken conversation. Specifically, the corpus is based on manual morphophonological transcriptions of three seasons of the reality TV show *Expeditie Robinson* (known as 'Survivor' in the English-speaking world) and contains 10,000 utterances produced by 52 Belgian Dutch and Netherlandic Dutch speakers. Our analysis zooms in on all English insertions found in the data that contain more than one word, such as *alive and kicking* and *you're almost there* (300 instances for 187 types). Three different diagnostic tests are presented that measure the degree of fixedness and conventionality of these multi-word insertions, namely lexicographical treatment, raw frequency and paradigmatic modifiability. Results reveal that the English multi-word inclusions are typically highly conventional fixed expressions, copied as a whole from English and inserted in Dutch much like traditional loanwords. Constructions containing open slots (e.g. *NOUN of the day*) and prototypical codeswitches (e.g. *someone's got to do all the work*) are rare. These findings will be linked to the nature of the English-Dutch contact setting. Overall, our study contributes to research on contact-induced variation and change by paying attention to the use of foreign (semi-)fixed phrases and constructions, which form an interesting but often neglected grey-zone on the continuum from borrowing to codeswitching. Additionally, it opens up the restricted focus on monolingual contexts of early research on phraseology and Construction Grammar.

Keywords: contact-induced variation and change, borrowing, borrowing-codeswitching continuum, Construction Grammar, phraseology, English, Dutch

1 Introduction: Grey-zones on the borrowing-codeswitching continuum

Contact linguistic research has repeatedly debated the distinction between borrowing and codeswitching, particularly concerning the status of single-word switches. More recently, it has been proposed to let go of the idea of a sharp dichotomy between both phenomena and instead consider them as endpoints on a continuum (e.g. Matras, 2009). Single-word switches should then be considered as a grey-zone on this continuum: varying on the context of use and the profile of the user, they are more like borrowing or more like codeswitching.

This paper draws attention to another grey-zone on the continuum from borrowing to codeswitching, namely (semi-)fixed foreign expressions and constructions (e.g. English *as good as it gets* occurring in Dutch). Instead of lumping all multi-word foreign insertions together as codeswitching, we study variation in the level of fixedness and schematicity of these phrases, aiming to reveal insights in the nature of the contact situation under scrutiny. As such, our study builds on Doğruöz & Backus (2009), who present a study on the borrowability of constructions in Dutch-Turkish contact (see also Heine & Kuteva, 2005; Sakel, 2010). However, where their analysis focused on the ways in which foreign language patterns are rebuilt with language-own material (indirect borrowing, or pattern replica), we propose an analysis the of direct insertion of foreign language material.

More specifically, our analysis scrutinizes all English insertions containing more than one word that are used by participants to the Dutch reality TV show *Expeditie Robinson* (known as 'Survivor' in the English-speaking world). Before discussing the English-Dutch contact setting, the game format of Expeditie Robinson and the database, we first frame the position of our study on foreign phraseology and (semi-)fixed constructions against the general background of research on borrowing and codeswitching.

1.1 The borrowing-codeswitching continuum

Regardless of the motivation (e.g. migration, colonization, globalization), contact between two languages typically results in change in one or both of the languages involved (Thomason, 2001, p. 10). The language exerting influence is often referred to as the source language (SL), with the language being influenced referred to as the receptor language (RL). The most widely and frequently attested type of influence in language contact concerns the borrowing of words.

Lexical borrowing can be defined as the complete or partial transfer of a form-meaning pair from a source language to a receptor language. As can be seen in Table 1, this approach theoretically allows for four different types of borrowing (see Zenner, 2013).

Table 1 **Borrowing form-meaning pairs**

	SL meaning	**RL meaning**
SL form	direct borrowing	pseudo borrowing
RL form	indirect borrowing	(language-internal)
mixed form	hybrid borrowing	

Direct lexical borrowing concerns the transfer of both a content word and (one of) its meaning(s) from the SL to the RL. Examples include the use of the Italian noun *espresso* or the English adjective *gay* in various languages. Indirect borrowing concerns the semantic extension of an existing RL form based on an SL form-meaning pair, such as the use of Dutch *muis* ('mouse') to refer to a small pointing device attached to personal computers, based on a parallel metaphorical extension of English *mouse*. When RL meanings are attached to an existing or newly coined SL form, this is typically referred to as pseudo borrowing (e.g. the German use of the English word *Handy* as referring to mobile phones) (see Furiassi and Gottlieb, 2015). Hybrid forms consist of both RL and SL material (e.g. the Dutch compound *computertafel*, which has the Dutch noun *tafel* ('table') as head and the anglicism *computer* as modifier). In this paper, indirect borrowings, where SL structures are re-built by means of RL forms, are not included: only direct insertions of SL material are considered.

Existing research on lexical borrowing has focused on creating taxonomies of borrowing (Duckworth, 1977), drafting borrowability clines (also including direct grammatical borrowing[1], see Field, 2002; Curnow, 2006) and charting different strategies for morpho-phonological adaptation of loanwords to the structure of the receptor language (Onysko, 2007; Winter-Froemel, 2013). One of the most popular topics so far however concerns the distinction between borrowing and codeswitching. Put simplistically, codeswitching includes the process of switching between different grammatical systems within a single conversation (see example 1, taken from the title of Poplack's well-known 1980 study) (also see Gumperz, 1982; Myers-Scotton, 1993; Muysken, 2000 for standard works). Though the distinction between borrowing and codeswitching may seem straightforward at first sight, especially the status of single source language content words occurring in the receptor language has proven to be con-

troversial. Should a distinction be made between single-word switches and loanwords, and if so, how? Concerning the first question, Myers-Scotton (2002, p. 153) stresses that there is basically no need for a synchronous distinction between borrowing and codeswitching. Most other researchers do believe that, relying on a number of criteria, a distinction needs to be made.

(1) sometimes I'll start a sentence in Spanish *y termino en español* (Poplack, 1980)

First, recourse is occasionally taken to the proficiency of speakers: if monolingual RL speakers use an SL element in speaking the RL, it is considered safe to conclude that we are dealing with lexical borrowing (Thomason, 2001, p. 133). However, proficiency cannot be used equally straightforwardly when trying to classify single-word insertions occurring in the language use of only one bilingual speaker. A second criterion, the morphophonemic adaptation of the word, can then be turned to: if a form is adapted to the morphophonological rules of the RL, it is a loanword (Poplack, 1980; Appel and Muysken, 1988, pp. 172-173). However, this test also has its shortcomings (but see statistically advanced implementations in e.g. Cacoullos and Aaron, 2003). Sometimes a borrowed word needs no adaptation to follow receptor language rules (e.g. if SL and RL are closely related): the test only provides proof for those items that are adapted. Yet another test relies on the 'listedness' of the items (i.e. inclusion in standard lexicography of the receptor language) (see Muysken, 2000, p. 71 on DiSciullo and Williams, 1989). Again, the practical implementation has been criticized (Deuchar, Muysken and Wang, 2007, p. 311), but the biggest controversies on diagnostic tests no doubt concern the use of frequency as criterion. The idea is that if a word occurs only once, it is considered a switch (Jake and Myers-Scotton, 2002). The question of course begs itself what 'just once' actually means (once in what timeframe for how many speakers). More crucially, Sankoff, Poplack and Vanniarajan (1990) and Poplack and Meechan (1995) have strongly opposed the basic idea of using frequency as a factor, instead claiming the existence of nonce borrowing (i.e. a loanword that occurs only once).

An alternative to these diagnostic tests is to let go of the idea of a sharp dichotomy between borrowing and codeswitching, and to consider both contact-induced phenomena as related points on a continuum (Thomason, 2001, p. 133; Matras, 2009, p. 110). To this end, Matras (2009, pp. 111-113) joins together six different features (bilinguality, composition, functional-

ity, specificity, operationality, regularity and structural integration) to define the two end-points of the continuum. A given single-word source-language insertion can then either be considered as a prototypical instance of borrowing or codeswitching, or it can be placed on the grey-zone of the continuum, when it has features of both contact-related concepts.

Crucially, the six criteria used by Matras should each in themselves be conceived of as continua (Matras, 2009, p. 111). The bilinguality criterion for example entails a gradual cline from the monolingual speaker (prototypical borrowing) to the fully balanced bilingual (prototypical switching), the regularity criterion presupposes a continuum from hapax legomena (prototypical switching) to highly frequent occurring words and patterns (prototypical borrowing) and the criterion of composition implies a cline from single lexical items (prototypical borrowing) to elaborate utterances and phrases (prototypical switching) uttered in the source language.

It can be assumed that scrutinizing (different points on) these continua can provide new insights into contact-induced variation and change. In this paper, we focus on Matras's composition criterion (although the bilinguality and regularity criterion will also prove relevant). Linking the criterion to the lexicon-syntax continuum of Cognitive Linguistics, we analyze the types of multi-word English insertions found in Dutch (see also Backus, 1999).

1.2 Phraseological units and constructions as grey-zones on the continuum

For his composition criterion Matras considers a cline from single lexical items (linked to prototypical borrowing) to elaborate utterances and phrases (linked to prototypical codeswitching). This cline can be linked the lexicon-syntax continuum proposed in Cognitive Linguistics in general and Cognitive Grammar and Construction Grammar in particular (see also Doğruöz and Backus, 2009).[2] Neither of these frameworks makes a sharp distinction between lexicon and syntax; instead, they assume a continuum from the lexical pole of highly specific and conventionalized lexical items (e.g. single content words such as *computer* or completely fixed idioms such as *as good as it gets*) to the syntactic pole of highly schematic patterns based on grammatical roles and word order (e.g. *I called you* [S V O]). In between are partially schematic units which have both fixed lexicalized items and open slots that can be filled up with a variety of items, ranging from lowly schematic phraseological sequences (such as [pimp my N] for *pimp my ride*, *pimp my grandma*) to highly schematic phraseological sequences (such as the verb body part off-construction for *work my butt off*).

Foreign-language multi-word insertions can occur on any position on this lexicon-syntax continuum.[3] Each of the English examples listed above can for example be found in Dutch. Lumping all these foreign insertions containing more than one word together under the flag 'codeswitching' is however clearly oversimplifying matters. Most importantly, it ignores the fact that these phrases and constructions portray varying degrees of fixedness and conventionality, which affects their position on the borrowing/codeswitching-continuum. This is most obvious when linking constructions to Alison Wray's definition of phraseology as including 'all sequence[s], continuous or discontinuous, of words or other elements, […] stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar' (Wray, 2002, p. 9)

Just like phrases can be stored and retrieved whole from memory, they can also be copied whole from a source language and introduced as such in a receptor language (e.g. English *long time no see* in Dutch), without any internal parsing or creativity needed on behalf of the receptor language user. In this respect, certain phrases function much like traditional lexical borrowing (see also Backus, 1999). At the same time, constructions and phrases often have or acquire one or more open slots. Filling these up (with source language material[4]) does require some creativity on part of the speaker (*verb your noun off?*). In contact settings, such creativity in itself requires at least some level of bilingualism from the receptor language speaker, who needs enough understanding of the source language structure and lexicon to fill up the open slots in the construction adequately (*dance your socks off*). Of course, the more open slots are included in a phrase, the higher this level of creativity and proficiency demanded from the user will be, and the more the phrase moves up to the codeswitching pole of the borrowing-codeswitching continuum.

In this way, scrutinizing the level of fixedness and schematicity of the multi-word source language insertions found in a given receptor language corpus might be revealing of the nature and intensity of the contact setting involved. The weaker the influence of source on receptor language, the weaker the level of source language proficiency of the RL speakers will be, the more we can expect multi-word insertions to be restricted to completely fixed expressions. As intensity of language contact increases, speakers become more bilingual, and the schematicity of the expressions can be expected to rise and the degree of fixedness can be expected to drop. The way these different elements relate to each other is summarized in Figure 1.

example	*as good as it gets*	*- N of the day - V body part off*	*- S V O*
phrase type	fixed expression	- construction with open slot(s)	- new language material
degree of fixedness	fixed	→	free
speaker input	copying	→	creating
speaker proficiency	weakly bilingual	→	strongly bilingual
contact phenomenon	borrowing	→	codeswitching

Figure 1 Multi-word insertions in contact settings

The idea of relating contact setting to the type of multi-word units and constructions borrowed already formed the cornerstone of Doğruöz and Backus (2009). Their study focused on indirect borrowing (pattern loans) in an intense contact setting (Dutch-Turkish contact for second generation Turkish immigrants in the Netherlands). Focusing on Dutch Turkish constructions which sound unconventional to Turkish Turkish speakers, the analysis zooms in on constructions at the more lexical/fixed side (the copying of the Dutch collocation [N *doen*] to Turkish [N *yapmak*]), on the partially syntactic/fixed side (e.g. morphosyntactic issues such as the unwarranted omission of genitive markers), and on the maximally syntactic side of the continuum (word order in two non-finite clauses). Results show that the unconventional nature of the constructions results from contact with Dutch and that unconventional constructions turn out to be scattered along the lexicon-syntax continuum, but that more unconventional constructions are found at the lexical side than at the syntactic side of the continuum.

Where the analysis of Doğruöz and Backus (2009) zooms in the types and ways in which source language structures are re-built by means of receptor language forms, our study only considers direct insertions of foreign source language material. Additionally, we shift the focus from an intense contact setting (migration) to a weak contact setting (mediatized anglicization in the Low Countries). Specifically, we analyze all English insertions containing more than one word that occur in a corpus of spontaneous Dutch conversation. Our main interest is to see whether we only find English phrases that are maximally fixed (and can hence be more or less equated with single-word borrowing) or whether we also find constructions with open slots or freely created English language material (more prototypical codeswitches). To this end, three diagnostic tests are introduced that measure the degree of fixedness and conventionality of a given multi-word insertion. Results indicate to what extent source language creativity is at play in the English multi-word insertions used by

Dutch speakers. To allow for an optimal understanding of these results, we first sketch the English-Dutch contact situation in the next section.

2 Case: English-Dutch contact

English-Dutch contact mainly results from urbanization and (economic) globalization. Contact between the two languages is strongly asymmetrical: whereas contact with Dutch is rare for the English, the Dutch are in seemingly constant contact with English via mass media such as movies, popular music, television and – most importantly – the internet (see Thomason, 2001, pp. 2-3). Because of the popular position of English as *lingua franca* for international communication, most Belgian Dutch and Netherlandic Dutch speakers are able to have at least a basic conversation in English (Eurobarometer 2012; Zenner, Speelman en Geeraerts, 2015). However, direct face-to-face contact with native speakers of English is unusual in the Low Countries. The contact between English and Dutch can hence best be described as indirect, asymmetrical, and generally weaker than in typical contact settings resulting from immigration or colonization (see Onysko, 2009 for terminology).

Research on the outcomes of English-Dutch contact has mainly focused on loanwords: the admiration for the American liberator in the aftermath of World War II, the Anglo-Saxon orientation of the IT revolution and the Americanization of pop culture and the movie industry resulted in an increasing influx of English loanwords in the second half of the twentieth century. Far less attention has been paid to the occurrence of longer stretches of English in Dutch, which form the focus of the present study: we analyze the degrees of fixedness and conventionality of the longer stretches of English used by Netherlandic Dutch and Belgian Dutch participants to the reality TV show *Expeditie Robinson*. The game format, corpus and resulting database are described in more detail in the next section.

3 Data: English insertions in *Expeditie Robinson*

Our analysis zooms in on the use of English by Belgian Dutch and Netherlandic Dutch participants to the reality TV show Expeditie Robinson (known as 'Survivor' in the English-speaking world). Within the broader field of reality TV, the show belongs to the subgenre of the game doc. It is a social game where approximately sixteen to eighteen participants try to

survive on a (supposedly) deserted island. Meanwhile, participants have to compete in physical and intellectual challenges - initially per tribe, later on an individual basis (see Zenner, Geeraerts and Speelman, 2009; Zenner, Speelman and Geeraerts, 2015 for more details on the structure of the game). The main benefit of reality TV is that it provides highly accessible data than can easily be gathered on a large scale. This is specifically important for our study, considering how the weak and asymmetrical contact between English and Dutch gives us reasons to believe that longer English insertions will not occur very frequently. An additional advantage is that, in contrast to other reality TV shows such as *The Bachelor*, participants of *Expeditie Robinson* come from a wide variety of social backgrounds.

The present analysis zooms in on seasons 4, 5 and 6 of the show (broadcast in 2003, 2004 and 2005). The data were transcribed following the CHAT conventions of the CHILDES project (MacWhinney, 2000), resulting in a corpus of approximately 10,000 utterances for forty-eight participants (23 from the Netherlands and 25 from Flanders, the Dutch-speaking part of Belgium).[5]

All English elements occurring in the utterances are coded using XML-style mark-up tags. Across the data, 777 utterances contain English words, amounting to a total of 883 English inclusions that represent 492 different types. These inclusions mainly involve single word units, such as nouns and adjectives related to the game structure (e.g. *team, power, fair, heavy*) and discourse markers used to express negative and positive emotions (e.g. *shit, fuck, yes*). Nevertheless, we also find a number of longer stretches of English. Specifically, nearly 30% of all English inclusions contain more than one word: 249 longer English inclusions were found in the data, representing 187 different types. As can be seen in Table 2, the majority of these longer stretches are verb phrases and complementizer phrases (example 2) or noun phrases (example 3); adjective phrases and prepositional phrases are relatively rare (examples 4 and 5).[6]

Table 2 **Longer English insertions in *Expeditie Robinson***

phrase	types
verb phrase/complementizer phrase	114
noun phrase	59
other	14
total	187

The English inclusions occurring in the examples (*fuck the world, survival of the fittest, alive and kicking, out of the blue*) all appear to be instances of the

completely fixed and conventionalized units on the most lexical side of the lexicon-syntax continuum described above. These insertions seem routinized formulae that should not be considered as phrases that are coined creatively and *ad hoc* by the Dutch participants. Instead, they seem copied as a whole from English into Dutch. In this sense, the English phrases occurring in the Dutch sample appear to be located closer to prototypical lexical borrowing than to prototypical codeswitching. In the next section, we verify whether we can provide empirical support for this intuition that extends towards all phrases in our database.

(2) @Situation: <Y04.E03.F098.U01.MIC>[7]
**MIC:* misschien kom ik over a(l)s nen etterbak ma(ar) <ENG>fuck the world</ENG>

eh@fp zeg.
'Maybe I come across as a total asshole but **fuck the world**, you know.'[8]

(3) @Situation: <Y04.E01.F085.U01.MAT>
**MAT:* tis [: het is] (h)ier wel een <ENG>survival of the fittest</ENG> ma(ar) (i)k vind

nie(t) da(t) wij (n)et recht hebben om iemand het eiland te ontzeggen.
'It's a true **survival of the fittest** here but I don't believe we have the right to decline anyone access to the island.'

(4) @Situation: <Y03.E10.F043.U01.BJO>
**BJO:* ik geniet (h)ier gewoon # waar ik niet van genoten (h)eb vandaag was dat ik zo

stom geweest ben ~~om~~ euh@fp in men blote kont te willen rondlopen # ze was nog helemaal wit # nu is hij helemaal rood dus euh@fp mensen # eerst insmeren # bedankt en tot morgen # hoogste tijd # snaveltje toe welterusten iedereen # Bjorn is <ENG>alive and kicking</ENG>
'I'm just enjoying this - what I didn't enjoy today was that I was stupid enough to run around butt naked - it was still completely white and now it's all red, so people: use sunscreen... thanks and see you tomorrow. High time... off to bed and good night to everyone: Bjorn is **alive and kicking**!'

(5) @Situation: <Y04.E07.F090.U05.RON>
RON:* ja ik (h)et kwam voor mij een beetje <ENG>out of the blue**</ENG> moet ik

zegge(n) # (i)k (h)ad (h)et helemaal niet verwacht en euh@fp ik heb (h)em een hand gegeve(n) en euh@fp wat mij betreft is (h)et euh@fp zand derover en a(l)s hij zegt van ja ik wil liever niet euh@fp met jou prate(n) wat dan ook dan respecteer ik dat.
'yes for me it kind of came up **out of the blue** I must say and I shook his hand and as far as I'm concerned it's bygones and if he says "yeah, I'd rather not talk to you" or whatever, then I respect that.'

4 Method: Measuring fixedness and conventionality

Our analysis focuses on all longer stretches of English found in the Robinson database: all English insertions except single-word units are included. The main hypothesis is that these insertions are fixed expressions, prefabricated chunks with a high level of fixedness and conventionality in the source language, which are copied as a whole into the receptor language.

To confirm this hypothesis, we need to determine the level of conventionality and fixedness (known as 'unithood' in term extraction research; cf. Heylen and De Hertog, 2015) of each of the 187 longer English insertions in the data. To this end, we rely on a combination of several diagnostic tests. First, we assess the level of conventionality of the phrase in the source language by relying on lexicographical evidence. To assess the degree of fixedness/unithood, we turn to corpus-based evidence: we rely on raw frequency information for the entire phrase, and on a simplified version of the paradigmatic modifiability measure designed by Wermter and Hahn (2005). Below, each of these three tests is presented in some more detail. In Section 5, the three methods are combined in an aggregate score. Based on this score, a general discussion on the fixedness and conventionality of the multi-word insertions will follow.

4.1 Lexicographical treatment

Lexicographical treatment of a phrase primarily helps to assess its degree of conventionality in the source language, but simultaneously sheds light on its level of fixedness: phrases listed as lemma are conventionalized in the specific fixed form attested in the dictionary (see also Poplack, Sankoff en Miller, 1988). Of course, caution is needed. Dictionaries can be rather conservative and often lag behind on actual use. Additionally, due to size

restrictions, entry policies for multi-word units have typically been strict for print dictionaries. To accommodate these shortcomings, we complement this lexicographical measure with corpus-based techniques (cf. infra). Moreover, we combine information from different types of lexicographical sources.[9]

First of all, we looked up the phrases on the subscribers' online module of *Oxford English Dictionary*. Next, we relied on the free online module of *Collin's Cobuild*.[10] Next to these traditional descriptive dictionaries of the English language, we included *Van Dale Groot Woordenboek Engels-Nederlands*, an English-Dutch translation dictionary. To ensure maximal coverage for newer phrases, we also consulted two open-source volunteer-driven web dictionaries: *Wiktionary* and the slang dictionary *Urban Dictionary*.

Before discussing the treatment of our 187 phrases in these sources, a semantic comment is in order: we only coded a phrase as listed in a given dictionary when the semantic description of the lemma corresponds to the use of the phrase in our corpus. This proved specifically important for Urban Dictionary, where quite some of the expressions acquired highly 'specialized' meanings, describing all sorts of (peculiar) sexual activities. A well-known example concerns the phrase *happy ending*. Example (6) contains the phrase in its literal meaning, as used by Karen in season 4 of Expeditie Robinson. The only meaning attested in Urban Dictionary is provided in (7). For our purposes, we hence do not consider the phrase *happy ending* as attested in Urban Dictionary.

(6) @Situation: <Y03.E09.F011.U01.KAR>
**KAR:* (i)k (h)ad dit nooit wille(n) misse(n) en # (i)k (h)eb er heel veel uit geleerd en dat
ist [: is het] belangrijkste # a(l)s ik daar ee(n) paar vrienden aan over(h)ou die da(t) ik af en toe kan zien meer moe(t) da(t) nie (t) zijn voor mij # tis [: het is] geen <ENG>happy ending</ENG> ma(ar) ik bedoel ik ben zeker nie(t) verbitterd.
'I never would've wanted to miss out on this and I've learned a whole lot from it and that's what's most important; if I end up with some good friends that I can meet occasionally that's okay for me... It's not a **happy ending** but I mean I'm not bitter or anything.'

(7) When a masseuse feels inclined to finish your session with oral sex or manual release (usually for an extra twenty dollars)

The fact that the specialized meaning of the phrase is the only one provided by Urban Dictionary points to the limitations of free-content volunteer-driven dictionaries, but paying extra attention to the semantic load of the phrases helps us account for this shortcoming.

In all, 88 out of 187 of the n-grams attested in our corpus (47%) occur in at least one of the five consulted sources, 64 (34%) occur in at least two sources, 47 (25%) occur in at least three sources, 29 (15%) in four or more and 22 (11%) in all five. Given the tradition of strict entry policies for MWU's in descriptive dictionaries, these are surprisingly high numbers. Hence, we have a closer look at the distribution of the attested phrases over the types of dictionary in Table 3, mainly focusing on the distinction between traditional dictionaries (*OED, Collin's* and *Van Dale*) and volunteer-based web dictionaries (*Wiktionary* and *Urban Dictionary*).

Table 3 Lexicographic treatment for MWU's in Robinson

dictionaries	number	%
all	22	12%
combined traditional/free-content	33	17%
only free-conten	24	13%
only traditional	9	5%
none	99	53%

As can be seen in the table, 55 of the 88 attested phrases (62%) occur in both the web-based and in the more traditional 'closed-source' dictionaries. This means that the types of sources consulted are sufficiently comparable to allow for a ternary classification: in what follows, we contrast phrases included in all sources, with phrases included in some sources and those without a lemma (see Table 4).

Table 4 Lexicographic treatment for MWU's in Robinson: ternary

dictionaries	number	% of data
entry in all	22	12%
entry in some	66	35%
no entry	99	53%

The main question now is what we can say about the level of unithood of the 99 phrases that are not included in any of the six dictionaries. Examples (7) and (8) demonstrate that these phrases are not necessarily less fixed units. Below, we turn to two corpus-based measures to support this intuition.

(8) @Situation: <Y04.E01.F097.U02.MIC>
**MIC:* nu gaan weraan [: we eraan] beginnen eh@fp # <ENG>free at last</ENG>.
'Now we're gonna go do it, **free at last!**'

(9) @Situation: <Y05.E03.F012.U01.EST>
**EST:* kapitein Douwe # <ENG>permission to board the Papillon</ENG>.
'Captain Douwe, **permission to board the Pappilon**?'

4.2 Google frequencies

A first step that can be taken when looking for corpus-based support of unithood is to retrieve basic frequency information for a given phrase: the more frequent the phrase, the more fixed it is. This is obviously a rather rudimentary approach to fixedness, but it is relevant for our study (see Van Hout and Muysken, 1994 on the impact of source language frequency on borrowability). Because Dutch-English contact is most frequently channeled through the internet, we turn to Google frequencies. As the internet is most likely one of the prime sources where the phrases are initially picked up (internet memes), frequency information for the web is of prime interest to us. Of course, Google frequencies are known to be susceptible to fluctuations depending on the time and location of the query, and to be disturbed by semantic ambiguity (cf. *happy ending*). To accommodate this shortcoming we refrain from working with specific numeric information. Instead, we work with five broad frequency bands. Additionally, we complement this web-based analysis with a more complex corpus-based criterion (cf. infra). The results of our query (conducted in May 2013, limited to pages in English) are presented in Table 5.

Table 5 Google frequency bands

freq band	number	% of data
<=10	8	4%
[11-100,000]	18	10%
]100,000-1,000,000]	19	10%
]1,000,000-10,000,000]	47	25%
>10,000, 000	95	51%

Table 5 reveals that most of our phrases occur highly frequently on the internet: 142 phrases (more than 75%) have more than one million hits on Google. This provides us with some further support of the associative strength between the elements in the phrases, but as mentioned above

raw frequencies are not the most reliable type of information to assess unithood and fixedness. The best known corpus linguistic example to support this critical note are bi-grams: when studying a bi-gram that consists of two highly frequent elements (e.g. *good man*), the odds of the two elements occurring together are *by chance* relatively high: a high token count for *good man* does not necessarily mean that this bi-gram constitutes a fixed expression (cf. De Hertog, 2013, p. 17). More complex techniques can help us account for this observation. To this end, we turn to more stable corpora.

4.3 Paradigmatic modifiability

Within term extraction research and corpus linguistics, several methods have been suggested to include a probabilistic perspective in assessing unithood. For bi-grams, a well-known approach is to measure collocational strength by means of log-likelihood scores, Chi^2 and mutual information. These measures indicate whether two words occur together more than would be expected by chance (see Manning and Schütze, 1999 for an overview). For longer phrases, this is however not a very elegant solution, but corpus linguistics offers only few alternatives (see De Hertog, 2013, p. 16). In contrast, term extraction research has paid quite some attention to ways to establish unithood for n-grams. Most existing approaches rely either on language-specific linguistic information, on statistical information, or on a combination of both. Where language-specific techniques focus on information retrieved by means of part-of-speech tagging, statistical approaches rely on frequency and fixedness - the two main cross-linguistic properties of multi-word units (cf. Heylen and De Hertog, 2015). Given our focus on fixedness, it is natural that we turn to the latter type of approach.

Specifically, we propose an adaptation and simplification of Wermter and Hahn (2005)'s algorithm.[11] Their technique is meant to measure the reduced paradigmatic modifiability of a multi-word unit, which is an important property of fixed expressions. It is for example quite hard to replace the word *diesel* in the expression *diesel engine* by other nouns, whereas replacing *management* in *management course* is easy and straightforward. The procedure adopted by Wermter and Hahn (2005) to verify this reduced paradigmatic modifiability is summarized as follows by Heylen and De Hertog (2015):

> For each candidate multiword combination that has come out of an initial linguistic filtering step, they collect the frequencies of all word combinations that have the same length and share at least one word with the candidate, but

> that also have one or more constituting parts replaced by another word. The accumulated frequency of these modified versions is then compared with the frequency of the actual multiword term candidate. (Heylen and De Hertog, 2015, p. 210)

For this study, we apply a strongly simplified version of the algorithm, based on frequency information retrieved from the online search module of the COCA corpus.[12] Our target phrases consist of all 187 English n-grams found in *Expeditie Robinson*. For each of these, we measure the frequency of the different attested instantiations of the n-gram when varying on its right-most slot (including only those forms that have the same POS as in the original n-gram[13]). These different instantiations are then ranked according to their frequency of occurrence. The rank number of our target phrase is used as indication of its fixedness: the closer to the top, the more fixed the expression.[14] Below, we provide an example to illustrate the method.

(10) @Situation: <Y04.E04.F052.U01.PAT>
**PAT:* ja ik eet alles op maar.
@Situation: <Y04.E04.F052.U02.PET>
**PET:* ja langzaam.
@Situation: <Y04.E04.F052.U03.MIT>
**MIT:* <ENG>slow motion</ENG>

'Yes I'm eating it all, but.'
'Yes, slowly.'
'Slow motion.'

In season 5 of *Expeditie Robinson*, Mitchell uses the phrase *slow motion*. When using this phrase as input for the algorithm, the next step is to look for all instantiations of [*slow* N] in COCA. Next, all attested combinations are ranked according to frequency of occurrence. Table 6 provides this ranked frequency information for the ten most frequent instantiations of [*slow* N].

Table 6 [*slow* NOUN]

[slow + NOUN]	frequency	rank
slow motion	1039	1
slow pace	339	2
slow start	320	3
slow process	303	4
slow growth	292	5
slo w cooker	274	6
slow progress	188	7
slow death	137	8
slow food	119	9
slow speed	101	10

The table reveals that our target phrase, *slow motion*, is the most frequent instantiation of [*slow* N]. This can be considered a clear indication of unithood. We repeat this procedure for all 187 n-grams in our dataset, and compare the attested rank for all phrases. For ease of reference, we work with four different categories, as shown in Table 7.

Table 7 Rank information COCA

rank	number of cases	% of data
1	54	29%
[2-10]	43	23%
[11-50]	19	10%
over50	29	16%
NA	42	22%

We can see in Table 7 that approximately one out of three phrases is the top most frequent instantiation of the phrase, which provides clear support for fixedness. Only 22% of the phrases are not attested in the corpus.

5 Results: aggregating over the measures

In this section, we bring the measures presented above together. After presenting results from an inferential test for association between the three diagnostic tests, we introduce a simple aggregate measure that summarizes the results from the three tests in a score out of ten. Based on this measure, we can assess the average fixedness of the 187 n-grams in the Robinson database.

The three measures of fixedness presented above are only first suggestions to measure source language fixedness and conventionality (experimental techniques could for example also be used). Nevertheless, the measures can be used to acquire a first view on the degree of fixedness of the English multi-word inclusions found in our data. This is supported by the fact that the three measures are significantly associated ($p < 0.00001$ for Cochran-Mantel-Haenszel); we can safely bring the results together. We opt for a simple, but highly intuitive way to aggregate over the data, arriving at a score of fixedness/conventionality out of ten. Based on the lexicographical test, a multi-word insertion receives a score on three. For each of the two corpus-based measures, a maximum score of four can be attained. Table 8 provides an overview of the aggregation.

Table 8 Overview of aggregation

measure	observation	points assigned
lexicographic treatment	no lemma	0
	lemma in some sources	1
	lemma in all sources	2
G oogle frequencies	<=10	0
	[11-100,000]	1
	]100,000-1,000,000]	2
	]1,000,000-10,000,000]	3
	>10,000,000	4
COCA rank	NA	0
	over50	1
	[11-50]	2
	[2-10]	3
	1	4

By means of illustration, we have a look at the phrase *bad vibes* (example 11). The phrase is listed in one of our lexicographical sources (*Urban Dictionary*), which gives the phrase one point. Additionally, we found 1,220,000 hits on Google (three points). Finally, with a rank-number of 151 for paradigmatic modifiability, we assign the phrase one point.

(11) @Situation: <Y05.E06.F083.U02.MAX>
**MAX:* dan schiet er van die groep bijna niks meer van over he@fp # totaal verslonde(n)
door (h)un eige(n) <ENG>bad vibes</ENG>
'Then there's nearly nothing left of the group, they've been completely consumed by their own **bad vibes**.'

As such, *bad vibes* receives a combined score of five out of ten for fixedness. Looking across the database, we find fifteen phrases with a maximal score of ten (e.g. *survival of the fittest; ups and downs*) and eight phrases with a minimal score of zero (e.g. *we have three hotdogs, one burger*; *next step, next hurdle*).

Figure 2 summarizes the scores for the 187 phrases in a boxplot. The figure reveals how most phrases in our database have a high score for fixedness: more than half of the phrases have a score of seven or higher (99 out of 187), and only 25% of the phrases receive a score of three or less. For the majority of the longer stretches of English in our data, our measures indicate that we are dealing with highly fixed phrases. It seems most likely that these are copied as a whole from English into Dutch, without much creativity on behalf of the Dutch speakers, rather than being generated *ad hoc* by the Dutch language users.

Figure 2 Aggregate scores for fixedness for English multi-word inclusions in Dutch

As discussed in our introduction, this result can be linked to the type of contact under scrutiny: contact with English in the Low Countries is indirect, remote and primarily mediated through the media. Most Dutch speakers have a very reasonable understanding of English, but are not necessa-

rily proficient enough to codeswitch fluently. Adding the high social status of English as global international language and the constant stream of English via the Internet, television and radio, fixed phrases of all sorts can easily be propagated. This idea is supported by the specific types of phrases we find in the dataset. Amongst others, we find phrases from lyrics (*every little thing she does is magic*, The Police; *good vibrations*, The Beach Boys) or movies (*king of the world*, Titanic). In future research, we aim to flesh out the different sources for English fixed expressions occurring in Dutch in more detail (see also Zenner, Speelman and Geeraerts, 2013). These mediatized phrases are very occasionally used more creatively by the Dutch language users: in example 12, the traditional 'Mr. Bond' is replaced by 'my friend'.

(12) @Situation: <Y04.E09.F138.U03.MIC>
MIC:* <ENG>we meet again, my friend**</ENG>

The high use of highly fixed expressions is not unexpected against the background of English-Dutch contact. But what about the multi-word insertions that have lower scores for fixedness? A first question is how dispersed these phrases are across our different speakers. Table 9 contrasts the amount of speakers using phrases with a score of seven or more out of ten with phrases with a score lower than four out of ten. More than twice as many participants use highly fixed phrases as those using less fixed phrases: only fourteen speakers out of forty-eight use more creatively coined English insertions.

Table 9 Unithood and speakers

	number of speakers
highly fixed n-grams (scores >= 7)	31
less fixed n-grams (score <= 3)	14

When we zoom in on the amount of lowly fixed phrases used by each of these fourteen speakers (Table 10), we see that only three of them (Judge, Bjorn and Robin) are relatively frequent users of such more freely coined expressions. The other eleven participants use these multi-word insertions more sparsely.

Table 10 Speakers using lowly fixed phrases

speaker	number of low-fixed phrases
JUD	12
BJO	9
ROB	8
MAJ	4
MAX	4
MIC	4
MAT	3
KAR	2
CAR	1
EST	1
FAT	1
GIO	1
PET	1
RYA	1

Interestingly, the background of the three top users goes a long way in explaining their use of less-fixed phrases. Judge grew up in the United States and is hence near-native, which is also obvious from the higher syntactic complexity of his English insertions (examples 13 and 14).

(13) @Situation: <Y03.E07.F034.U07.JUD>

JUD:* <ENG>we know how**</ENG> # <ENG>**yes we do**</ENG> # <ENG>**we have tree hotdogs # one burger**</ENG>

(14) @Situation: <Y03.E10.F021.U01.JUD>
**JUD:* <ENG>somebody blew up like a motherfucker</ENG> ja (h)et ja ik ben gewoon helemaal blij.
"**somebody blew up like a motherfucker**, yeah, I'm over the moon"

Robin looks up to Judge and often mirrors his language use (see Zenner and Van de Mieroop, submitted for more details), including his use of English multi-word insertions. Finally, Bjorn is an international truck driver who is in contact with English (and *English as a lingua franca*) through his job. Together, these three speakers (Judge, Robin and Bjorn) account for 56% of the less-fixed phrases (29 out of 52).

Table 11 lists the 23 less-fixed phrases that are used by the other nine speakers, whose use of less-fixed English multi-word insertions cannot be

AUP – 156 x 234 – 3B2-APP flow
<TET1502_05_ZENN_1Kv36_proef1 • 25-03-16 • 13:51>
Pag. 0267

straightforwardly linked to their background. Four types of phrases are found in this table.

Table 11 Lowly fixed phrases

speaker	phrase
CAR	muscles no brain
EST	permission to board the papillon
FAT	in the spirit to win
GIO	fake guy
KAR	game, set and the match is over (n = 2)
MAJ	I'm so glad I m a woman
MAJ	if you sleep with the enemy, you're gonna die
MAJ	okay baby
MAJ	we love the mangrove
MAT	the battle has started (n = 3)
MAX	be strong like a lion (n = 2)
MAX	positive vibrations (n = 2)
MIC	fuck the others
MIC	Mitch the bitch
MIC	strong lady
MIC	we meet again, my friend
PET	game, set, over
RYA	I won this game

First, we again see some instances of media-related catchphrases. Although the scores for our diagnostic tests are rather low, their mediatized background makes it plausible that they were nevertheless copied as a prefabricated whole from English into Dutch (e.g. *we meet again*, James Bond).

Next, we see some instantiations of constructions with one open slot, such as *permission to board* as instantiation of [*permission to* V], *in the spirit to win* as instantiation of [*in the spirit to* V] and *fuck the others* as instantiation of [*fuck* NP]. However, it is unclear whether participants indeed creatively filled up the open slot or simply copied one specific instantiation of the construction. To be more certain that creativity on behalf of the RL speakers is indeed at play, we should be able to find several different instantiations of the same phrase in the language use of one speaker (e.g. *in the spirit to fight* and *in the spirit to win*). Alternatively, cases of self-correction in the database can be indicative of true receptor language speaker creativity. In example (15), Jutta starts off by using the Titanic

phrase *king of the world*, but then self-corrects to the more appropriate *queen of the world*.

(15) @Situation: <J03.A13.F054.U02.T1.E2.STDE.SP.CME.JUT>
**JUT:* oh ik voel mij echt supergelukkig # super super supergelukkig # tusse(n) mijn palmbomen (h)ier en mijn mijn hutje # (i)k voel mij echt #
<ENG>king of the world</ENG> # nee # <ENG>queen of the world</ENG>
'Oh I feel so super happy, super super super happy, amidst the palm trees and my little hut, I really feel like **king of the world**, no, **queen of the world**.'

Another dubious case of receptor language creativity concerns examples such as *positive vibrations* or *game, set and the match is over*. Here, it seems more as if the Dutch speakers attempt at reproducing a fixed and conventionalized English phrase (*good vibrations; game, set, match*) but do not completely succeed in successfully reconstructing the phrase.

Finally, we see a handful of syntactically very simple sequences that could have been generated on the spot (*fake guy; I won this game*). These are however very rare in the database.

6 Discussion and conclusions

This paper focused on the use and status of English insertions containing more than one word in the Dutch reality TV show *Expeditie Robinson*. Three diagnostic tests were introduced to measure the degree of fixedness and conventionality of these insertions, namely lexicographical treatment, Google frequencies and paradigmatic modifiability in COCA. After aggregating over these three measures, we found that the average degree of fixedness and conventionality of the attested multi-word insertions was very high. Overall, three types of insertions were frequently attested in the data.

The majority of multi-word insertions belonged to the first group, i.e. prefabricated, fixed expressions that can be assumed to be borrowed as a whole from English and inserted in Dutch without any creativity on behalf of the receptor language speaker (e.g. *ups and downs*). Next, some phrases with lower scores for fixedness can be considered as instantiations of constructions with one or more open slots (e.g. *free at last* as instantiation of

[A *at last*]). In these cases, receptor language creativity could be at play. However, as most of these constructions occurred only once in the database in one specific form, it is hard to say whether the RL speakers did effectively parse the expression and filled in the open slot with a self-selected appropriate form (going from [A *at last*] to *free at last*) or whether the speakers simply copied the specific instantiation of the construction (*free at last*) as a whole from English into Dutch. Third, some phrases were found that much resemble a specific fixed English expression, but are somewhat different in terms of word order or lexical choice (e.g. *game, set, over* instead of *game, set, match*). In these cases, it seems more likely that speakers are aiming for (but failing to reproduce) the SL expression than that they are deliberately and creatively altering the expression. Besides these three types, only a handful of freely created English multi-word insertions were found.

Overall, then, the level of RL creativity involved in the use of English multi-word insertions in this database is very low. English insertions are located much more on the lexical side of the lexicon-syntax continuum than on the syntactic side: as far as composition and creativity is concerned, these insertions behave much more like prototypical borrowing than like prototypical codeswitching. These findings can be linked to the type of contact between English and Dutch in the Low Countries. Because of the indirect and asymmetrical nature of the contact setting, most speakers are unbalanced bilinguals. They pick up repeated English expressions through the internet and television, but do not often create completely new English sequences on the spot (see also Androutsopoulos, 2012; Onysko, 2007; Stefanowitsch, 2002; Furiassi, Pulcini and Rodriguez-González, 2012). In order to fully grasp how this process works, future research will have to look into the reasons behind the copying of these expressions, and into their social and pragmatic functions in discourse (see Zenner and Van de Mieroop, submitted).

Several other avenues for future research can be thought of. First, we now only focused on source language forms occurring in Dutch (matter replica): no attention was paid to ways in which English patterns can be rebuilt through Dutch forms (pattern replica). Future research could verify how diverse such indirectly copied constructions from English into Dutch are in general, or could study the occurrence and spread of very specific constructions (e.g. scrutinizing whether *ik had zoiets van*, a construction older than *I was like*, saw a surge in popularity under influence of quotative *like*). Second, it is possible to re-run the analysis presented here for more intense contact settings, to further corroborate the link between type of

contact and types of foreign multi-word insertions found. Finally, on a methodological level, future research should pay more attention to different ways of measuring unithood. The techniques introduced here are only applicable to well-documented languages: to what extent can experimental techniques address this issue?

Future studies along these lines will undoubtedly attenuate the results we present here, but some interesting conclusions can already be drawn. Most importantly, this paper presents the benefits of imbedding insights from Cognitive Linguistics into the contact linguistic framework. On the one hand, our study breaks up foreign multi-word insertions into different categories, paying attention to grey-zones on Matras's composition continuum in particular and on the borrowing-codeswitching continuum in general (see also Backus, 1999). On the other hand, we address the monolectal fallacy that is indicative of early Cognitive Linguistic research by studying (semi-)fixed expressions and constructions in contact situations, as such emphasizing that 'a linguistic community is *never* homogeneous and hardly ever self-contained' (Weinreich, 1970, p. vii).

Notes

1. In syntax and grammar, we also find instances of direct and indirect borrowing. In this case, the distinction has been referred to as matter vs. pattern loans (Sakel, 2007), and as global vs. structural copying (Johanson, 2002)
2. It is definitely not the idea of this paper to revise the entire Construction Grammar framework in its different forms (e.g. Langacker, 1987; Croft, 2001; Goldberg, 1995). We refer the interested reader to Langacker (2005).
3. Please note again that this paper does not pay attention to pattern replication (i.e. the reconstruction of SL patterns based on RL material). For an analysis of pattern replication and the lexicon-syntax continuum, please see Doğruöz and Backus (2009).
4. Open slots in foreign constructions can also be translated, which starts a gradual process towards nativization of a construction (e.g. *pimp je grootje* 'pimp your nanny'; see Van de Velde and Zenner, 2009).
5. Four of the original fifty-two participants have been excluded from our analyses due to data sparsness. These four participants (Simon and Marlieke for the 2003 series, Johnny and Margriet for 2005) leave the island early on in the show and hence have too few utterances to rely on for measurements.
6. All insertions with a (finite) verb were included in the first category. In case of ambiguity concerning the switch-point from English to Dutch (e.g. in example 4), we adhered to a conservative classification (maximally Dutch, minimally English (e.g. *Bjorn is <ENG>alive and kicking</ENG>* instead of *<ENG>Bjorn is alive and kicking</ENG>*).
7. The situation-string is used to provide each utterance with a unique ID (<Year; Episode; Fragment; Utterance; Speaker>).
8. Words in bold face in the translation are English in the original Dutch version as well.

9. All online queries were conducted in April and May 2013.
10. http://www.collinsdictionary.com/dictionary/english
11. We would like to thank Dirk De Hertog and Kris Heylen for their ideas and assistance.
12. http://corpus.byu.edu/coca/, consulted in May 2013.
13. One exception needs to be noted: when imposing POS-restrictions led to a very low number of possible phrases, the part-of-speech of the right-most slot was ignored to acquire more reliable results. For example: as [*the middle of* ADVERB] has only 39 different combinations, we expanded the search to [*the middle of* __], which resulted in 1000 possible combinations.
14. Additionally, we also ran the algorithm varying on the left-most slot of the phrase. Results were highly comparable with the analysis we present in the paper (i.e. varying on the right-most slot).

References

Androutsopoulos, Jannis, 'English "on top": Discourse functions of English resources in German mediascapes', *Sociolinguistic studies*, 6 (2012), pp. 209-238.

Appel, René & Muysken, Pieter, *Language contact and bilingualism*. London: Arnold, 1988.

Backus, Ad, 'Evidence for lexical chunks in insertional codeswitching'. In: Bernt Brendemoen, Elizabeth Lanza & Else Ryen (red.), *Language encounters across time and space*. Oslo: Novus Press, 1999, pp. 93-109.

Cacoullos, Rena Torres & Aaron, Jessi E., 'Bare English-origin nouns in Spanish: Rates, constraints and discourse functions', *Language Variation and Change*, 15 (2003), pp. 289-328.

Croft, William, *Radical Construction Grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press, 2001.

Curnow, Timothy J., 'What language features can be "borrowed"?' In: Aleksandra Iur'evna Aïkhenval'd & Robert M.W. Dixon (eds), *Areal diffusion and genetic inheritance: Problems in comparative linguistics*. Oxford: Oxford University Press, 2006, pp. 412-436.

De Hertog, Dirk, TermWise Xtract. Automatic Term Extraction Applied to the Legal Domain. Ph.D. thesis, KU Leuven, 2013.

Deuchar, Margaret, Muysken, Pieter & Wang, Sung-Lan, 'Structured variation in codeswitching: Towards an empirically based typology of bilingual speech patterns', *The International Journal of Bilingual Education and Bilingualism*, 10 (2007), pp. 298-339.

DiSciullio, Anna-Maria & Williams, Edwin, *On the definition of word*. Cambridge (MA): MIT press, 1989.

Doğruöz, A. Seza & Backus, Ad, 'Innovative constructions in Dutch Turkish: An assessment of ongoing contact-induced change', *Bilingualism: Language and Cognition*, 12 (2009), pp. 41-63.

Duckworth, David, 'Zur terminologischen Grundlage der Forschung auf dem Gebiet der englisch-deutschen Interferenz. Kritische Übersicht und neuer Vorschlag'. In: Herbert Kolb & Hartmut Lauffer (eds), *Sprachliche Interferenz: Festschrift für Werner Betz zum 65. Geburtstag*. Tübingen: Niemeyer, 1977, pp. 35-65.

Eurobarometer, *Special Eurobarometer 386, Europeans and their languages*. Created by the European Commission, 2012. <http://ec.europa.eu/public_opinion/archives/ebs/ebs_386_en.pdf>

Field, Fredric W., *Linguistic borrowing in bilingual contexts*. Amsterdam: John Benjamins, 2002.

Furiassi, Cristiano & Gottlieb, Henrik, *Pseudo-English. Studies on False Anglicisms in Europe*. Berlin & Boston: Mouton de Gruyter, 2015.

Furiassi, Cristiano, Pulcini, Virginia & Rodriguez-González, Félix, *The anglicization of European lexis*. Amsterdam: John Benjamins, 2012.

Goldberg, Adele E., *Constructions: A Construction Grammar approach to argument structure*. Chicago: Chicago University Press, 1995.

Gumperz, John, 'Conversational code-switching'. In: John Gumperz (ed.), *Discourse Strategies*. Cambridge: Cambridge University Press, 1982, pp. 59-99.

Heine, Bernd & Kuteva, Tania. *Language contact and grammatical change*. Cambridge (MA): Cambridge University Press, 2005.

Heylen, Kris & De Hertog, Dirk, 'Automatic Term Extraction'. In: Hendrik J. Kockaert & Frieda Steurs (eds), *Handbook of Terminology: Volume 1*. Amsterdam: John Benjamins, 2015, pp. 199-219.

Jake, Janice L. & Myers-Scotton, Carol, 'Making a minimalist approach to codeswitching work: Adding the matrix language', *Bilingualism: Language and Cognition*, 5 (2002), pp. 69-91.

Langacker, Ronald W., *Foundations of Cognitive Grammar, Vol. I: Theoretical prerequisites*. Stanford (CA): Stanford University Press, 1987.

Langacker, Ronald W., 'Construction Grammars: Cognitive, radical, and less so'. In: Fransisco Ruiz de Mendoza Ibáñez & M. Sandra Peña Cervel (eds), *Cognitive Linguistics: Internal dynamics and interdisciplinary interaction*. Berlin & New York: Mouton de Gruyter, 2005, pp. 101-159.

MacWhinney, Brian, *The CHILDES project: Tools for analyzing talk*. Mahwah (NJ): Lawrence Erlbaum Associates, 2000.

Manning, Christopher & Schütze, Hinrich, *Foundations of statistical natural language processing*. Cambridge (MA): MIT Press, 1999.

Matras, Yaron, *Language contact*. Cambridge (MA): Cambridge University Press, 2009.

Muysken, Pieter, *Bilingual speech: A typology of code-mixing*. Cambridge (MA): Cambridge University Press, 2000.

Myers-Scotton, Carol, *Duelling Languages: Grammatical structures in codeswitching*. Oxford: Clarendon, 1993.

Myers-Scotton, Carol, *Contact linguistics: Bilingual encounters and grammatical outcomes*. Oxford: Oxford University Press, 2002.

Onysko, Alexander, *Anglicisms in German. Borrowing, lexical productivity and written codeswitching*. Berlin & New York: Walter de Gruyter, 2007.

Onysko, Alexander, 'Divergence with a cause? The systemic integration of anglicisms in German as an indication of the intensity of language contact'. In: Falco Pfalzgraf (ed.), *Englischer Sprachkontakt in den Varietäten des Deutschen / English in contact with varieties of German.* Frankfurt am Main: Peter Lang, 2009, pp. 53-74.

Poplack, Shana, 'Sometimes I'll start a sentence in Spanish y termino en español: Toward a typology of code-switching.' *Linguistics*, 18 (1980), pp. 581-618.

Poplack, Shana & Meechan, Marjory, 'Patterns of language mixture: Nominal structure in Wolof-French and Fongbe-French bilingual discourse'. In: Lesley Milroy & Pieter Muysken (eds), *One speaker, two languages. Cross-disciplinary perspectives on code-switching.* Cambridge (MA): Cambridge University Press, 1995, pp. 199-232.

Poplack, Shana, Sankoff, David & Miller, Chris, 'The social correlates and linguistic processes of lexical borrowing and assimilation', *Linguistics*, 26 (1988), pp. 47-104.

Sakel, Jeanette, 'Grammatical borrowing from Spanish/Portuguese in some native languages of Latin America', *Language Typology and Universals*, 36 (2010), pp. 65-78.

Sankoff, David, Poplack, Shana & Vanniarajan, Swathi, 'The case of the nonce loan in Tamil', *Language Variation and Change*, 2 (1990), pp. 71-101.

Stefanowitsch, Anatol, 'Nice to miet you: bilingual puns and the status of English in Germany', *Intercultural Communication Studies*, 11 (2002), pp. 67-84.

Thomason, Sarah Grey, *Language Contact*. Edinburgh: Edinburgh University Press, 2001.

Van de Velde, Freek & Zenner, Eline, 'Pimp my Lexis: het nut van corpusonderzoek in normatief taaladvies.' In: Els Hendrickx, Karl Hendrickx, Willy Martin, Hans Smessaert, William Van Belle & Joop van der Horst (eds), *Liever meer of juist minder? Over normen en variatie in taal*. Gent: Academia press, 2010, pp. 51-68.

Van Hout, Roeland & Muysken, Pieter, 'Modeling lexical borrowability', *Language Variation and Change*, 6 (1994), pp. 39-62.

Weinreich, Uriel, *Languages in contact*. The Hague/Paris: Mouton de Gruyter, 1970.

Wermter, Joachim & Hahn, Udo, 'Paradigmatic modifiability statistics for the extraction of complex multi-word terms'. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing* (*HLT/EMNLP*). Vancouver: Association for Computational Linguistics, 2005, pp. 843-850.

Winter-Froemel, Esme, 'Formal variance and semantic changes in borrowing: Integrating semasiology and onomasiology.' In: Eline Zenner & Gitte Kristiansen (eds), *New perspectives on lexical borrowing*. Berlin & New York: Mouton de Gruyter, 2013, pp. 65-100.

Wray, Alison, *Formulaic language and the lexicon*. Cambridge (MA): Cambridge University Press, 2002.

Zenner, Eline, Cognitive Contact Linguistics: The macro, meso and micro influence of English on Dutch. Ph.D. thesis, KU Leuven, 2013.

Zenner Eline, Geeraerts, Dirk & Speelman, Dirk, 'Expeditie Tussentaal: Leeftijd, identiteit en context in "Expeditie Robinson"', *Nederlandse Taalkunde*, 14 (2009), pp. 26-44.

Zenner, Eline, Speelman, Dirk & Geeraerts, Dirk, 'What makes a catchphrase catchy? Possible determinants in the borrowability of English catchphrases in Dutch'. In: Eline Zenner & Gitte Kristiansen (eds), *New perspectives on lexical borrowing*. Berlin & New York: Mouton de Gruyter, 2013, pp. 41-64.

Zenner, Eline, Speelman, Dirk & Geeraerts, Dirk, 'A sociolinguistic analysis of borrowing in weak contact situations: English loanwords and phrases in expressive utterances in a Dutch reality TV show', *International Journal of Bilingualism*, 19 (2015), pp. 333-346.

Zenner, Eline & Van De Mieroop, Dorien, 'The social and pragmatic function of English in weak contact situations: Ingroup and outgroup marking in the Dutch reality TV show Expeditie Robinson'. Submitted to *Journal of Pragmatics*.

About the authors

Eline Zenner, University of Leuven, RU Quantitative Lexicology and Variational Linguistics, FWO Flanders, Blijde-Inkomststraat 21, 3000 Leuven.
E-mail: eline.zenner@kuleuven.be

Dirk Geeraerts, University of Leuven, RU Quantitative Lexicology and Variational Linguistics, Blijde-Inkomststraat 21, 3000 Leuven.
E-mail: dirk.geeraerts@kuleuven.be

Rethinking Weinreich, Labov & Herzog from a usage-based perspective

Contact-induced change in Dutch Turkish

Ad Backus

MEM 67 (2): 275–306
DOI: 10.5117/TET2015.2.BACK

Abstract

In the late sixties, Weinreich, Labov & Herzog (1968) discussed five empirical problems that together defined the subject matter of sociolinguistics. The research program they initiated has been very influential in sociolinguistics, but is not well known outside of it. In the branches of linguistics primarily focused on describing and accounting for the linguistic knowledge possessed by the individual speaker, branches often referred to as 'general linguistics' or 'theoretical linguistics', issues of variation and change have not attracted much theoretical attention up until recently. This has changed now that the tradition of 'usage-based linguistics' has appeared on the scene. This article examines the commonalities between the concerns of usage-based linguistics and those of Weinreich, Labov & Herzog (1968). It concludes that combining the cognitive orientation of the usage-based approach and the social-interactional orientation of sociolinguistics holds promises for the development of an improved understanding of language change. This is illustrated with examples from language contact, more specifically using data from Dutch-Turkish contact, including the results of recent studies that have implemented some methodological innovations necessitated by the confrontation of usage-based with sociolinguistic frameworks.

Keywords: usage-based approach, sociolinguistics, language change, Dutch-Turkish contact, codeswitching

1 Introduction

Suppose you are sitting down in a restaurant with some friends, you have looked at the menu, one of your company asks you what you will have, you think about it some more, and then you announce that you will have the steak. Whichever language you are speaking will have many ways available for saying this. If you and your friends are speaking Dutch, for example, you may say any of the following: *ik neem de steak* ('I'll take the steak'), *ik denk de steak* ('I think the steak'), *ik denk dat ik het bij de steak hou* ('I think I'll stick with the steak'), or *ik ga voor de steak* ('I'll go for the steak'). And there are many other alternatives. The last of the mentioned options uses a construction 'go for X' that is presumably taken from English. If we compare the Dutch version with its English translation, we see various similarities:

1. Ik ga voor de steak
   I go.PRES.1sg for the steak

   English:
   I'll go for the steak
   I.FUT.1sg

Dutch, by the looks of it, has borrowed this construction to say that one chooses a particular option out of several alternatives, including the choice of verb and the preposition it subcategorizes for. The only detail it has not adopted is the overt marking of future tense, presumably because the Dutch present tense covers future reference.

I am not actually aware of any serious scholarship on the history of this particular construction, so quite possibly I am mistaken about the origin, and embarrassing as that would be, the veracity of the analysis is not the main point of this contribution. One would need to provide a diachronic analysis of how this construction came to be used in Dutch, and an exhaustive synchronic analysis of its current use, while here it is just meant to illustrate a problem. This problem is that we do not have a generally accepted theory of this kind of change that addresses all aspects of change in an integrated way, and no generally accepted method of investigating it. Various questions immediately come to mind when one considers this particular example. How did the construction come into the language? Even if we can establish that it was 'borrowed', this only identifies the ultimate source, but says little about what exactly was taken from English,

how the borrowed pattern was adapted to the receiving structure, and why it was borrowed. One also needs to ask how come it spread so successfully. Apparently, there was something attractive about the construction so that it was picked up by many, and became a relatively entrenched convention, but why is that? How did it diffuse through the speech community? What is its status now, and how is it perceived by speakers?

These are the kinds of questions that were asked already more than forty years ago in a paper that may now be seen as the founding text of sociolinguistics, or at least one of the pivotal texts that helped set the agenda for this now vibrant discipline. Weinreich, Labov & Herzog (1968), to which I will from here on refer to as 'WLH', was an ambitious attempt to link the exploits of historical linguistics, particularly its views on how languages change, with those of the then emerging field of sociolinguistics. At the time, sociolinguistics mostly studied synchronic variation in a language and its correlation with social stratification. The aim of the current paper is to argue that the advent of the 'usage-based approach' in linguistics now makes it possible to rethink the link between variation and change. Sections 2 and 3 review some of the issues raised in WLH regarding variation and change in general, and contact-induced variation and change in particular. Section 4 is a third theoretical section which describes the usage-based approach and identifies concerns that it shares with sociolinguistics. The rest of the paper (sections 5 to 9) is more empirical, and illustrates how these shared concerns find expression in concrete analyses of data from Dutch-Turkish language contact. The emphasis in these sections will not be on the findings of past research but rather on new questions raised by new theoretical proposals, and the new kinds of data that are required to answer these questions. Some recent work is presented in which such data were obtained. Section 10 presents the main conclusions.

2 WLH on variation and change

The main empirical interest for WLH was the linguistic variation that is easily observed to exist all around us. People clearly do not form a homogeneous set, since we do not speak exactly the same and we say different things all the time. In addition, it is clear that language is forever changing and that this change is not entirely predictable. On the other hand, all this fluidity is counteracted by all kinds of efforts, most of them carried out unconsciously, to reduce the degree of heterogeneity, and this is what leads to the equally obvious observation that we do not differ from one

another all that much: there are common tendencies and apparently we feel certain constraints. As a result, we keep our differences manageable enough so that we generally have little trouble understanding each other. Another interest of WLH was the reason for variation. Heterogeneity serves a function: it helps marking group membership or identity. It has been remarked often enough that language and culture go hand in hand. As cultural groups form, whether on the large scale of nations and tribes or on the small scale of subcultures and communities of practice, the ways in which they speak tend to become distinct, ultimately giving us languages, dialects, sociolects and jargons (e.g. Christiansen and Chater 2008; Clark, 1996b; Labov, 1972). One of the reasons why we have so many languages around the world, rather than a single one, might well be that language is used to mark the speaker as belonging to a particular cultural group.

While WLH have helped set in motion the development of sociolinguistics, some aspects of the research program they envisioned have still barely materialized. In this paper, I will concentrate on one of those aspects: in most linguistic research, synchrony (variation) and diachrony (change) have been treated separately. By and large, linguistic studies either focus on synchronic variation (as in Labovian sociolinguistics) or on diachronic development (as in historical linguistics), but not on an integrated combination of the two planes of description. The message I aim to convey in this paper is that the theoretical outlook of what is commonly called 'usage-based linguistics' (Barlow en Kemmer, 2000; Langacker, 2008; Bybee, 2010) makes this combination not just more feasible, but in fact demands it because of its basic hypothesis that synchronic usage immediately affects competence. This entails that competence is claimed to be in continuous flux.

Perhaps the two most central interesting facts about language are its synchronic variability or diversity, and its diachronic changeability. At the same time, all this heterogeneity is kept in check, since as speakers we do not have the feeling that others speak so differently from us that we do not understand them, and changes do not proceed so fast that we cannot understand our parents. WLH cite Hermann Paul as proclaiming that there are 'as many languages as there are individuals' (WLH, Section 1.1, p. 105). True as this may be as a general statement of language, it masks some issues of considerable complexity, which we must understand in order to understand why there is variability and changeability to begin with. If every individual has her own language, this means that no two idiolects are exactly the same. However, at least two tricky questions arise. The first is how different from each other idiolects really are. While the abstraction

of the ideal native speaker-hearer may seem to many to be too serious a distortion of the empirical facts, we also need to be wary of the claim that everyone's language is different: we need to document what those individual differences consist of, and how extensive they are. Second, it is not at all obvious how we should describe an idiolect, or whether there really is such a thing. Just like a language, an idiolect too will consist of co-existing registers. To what extent we should treat an individual speaker's mental representation as a unified whole is at this moment not obvious.

I propose we see the whole complex of concepts as an issue of norms. A speaker's mental representation is her internal norm; the cluster of relatively similar internal norms of speakers who form a speech community together is a cumulative norm (cf. Backus and Spotti, 2012), also referred to as 'language' or 'dialect' or 'variety' (Hermann Paul had a term for this as well: 'language custom'; WLH Section 1.1). Changes in idiolects are not changes in language yet: they only are if the same change goes on in many idiolects at more or less the same time, and these idiolects are spoken by people who are in regular contact with each other, directly or indirectly.

This sets up a duality between individual and community, and introduces a paradox. As linguists, we are interested in the mental knowledge of the individual speaker, but we are not interested in all idiosyncratic details. It is 'language', after all, that we aim to say something about. Therefore, we look at cumulative levels, and extrapolate from there to the knowledge of the individual speaker, but at the same time, claims are based on data taken from individual speakers that are generalized to the community. This only makes sense if the informants cluster around a prototype for the group for which we are making these claims. Crucially, this requires that individual differences stay manageable: the community in question should not be too diffuse.

If change is really a design feature of language, we may want to include a model of change in our general model of linguistic competence. Croft (2000) provides the outlines of such a model, by proposing that what a speaker does at the start of every selection event is choose between saying what she always says ('normal replication' in Croft's biological metaphor) and saying something new ('altered replication'). As far as lexical and grammatical choices are concerned, pretty much everything we ever say will be normal replication; at the sentence level, however, we mostly say things we have never said before. That is, linguistic creativity resides in the stringing together of overlapping chunks of speech which themselves are stored on the basis of previous usage. Often, however, there is more than

one conventional way of saying something: change happens when a newer variant is selected more and more often at the expense of an older variant. Both cases of selection represent normal replication. In fact, this is more relevant for language change than 'altered replication', since as soon as you have heard someone use a word or an expression you had not heard before, it may get stored in your memory. If subsequently you use it for the first time yourself, this is already an act of normal replication. What is crucial for language change, though, is that it is the more newly introduced variant that gets selected. The reasons for choosing the newish variant can be many, including the well-researched psycholinguistic phenomena of priming and alignment (e.g. Pickering and Garrod, 2004; Jaeger and Snider, 2013) and the equally well-documented sociolinguistic process of accommodation (cf. Bell, 2001).

WLH famously identified five problems that sociolinguistics needs to solve in order to construct a viable theory about linguistic variation. These problems were:

- Constraints: what is possible? It is clear that not just any kind of change happens, so there must be principles that underlie what kinds of innovations are possible, and what kinds of changes propagate easily.
- Transition: how does a change propagate in an idiolect and in a community?
- Embedding: what implications does a change have for the larger linguistic system in which it occurs? And how is it connected to different layers in society, i.e. who uses the new variant?
- Evaluation: what is the social meaning of a particular change? How do people view it?
- Actuation: why this change, and why now?

In what follows, I will argue that although these problems have been on the agenda for five decades, there has not as yet developed an integrated field that treats them all together. That is unfortunate, because there are still good arguments to support WLH's basic claim, that a full theory of language change needs to address all five problems at the same time. Language change is determined by what people do, and what they do is partly determined by cognitive constraints (addressed in the Constraints and Embedding problems, as these pertain to the linguistic system and the cognitive skills underlying it). It is also partly determined by social constraints (addressed in the Transition and Evaluation problems, as these pertain to the current status of linguistic forms, e.g. whether they are seen as conventional or not).

3 Contact linguistics and WLH

The study of multilingualism and language contact has been a constant in sociolinguistics, but mostly as a stand-alone topic. Contact issues have not been brought to bear much on the abovementioned cluster of problems identified by WLH. Most likely this is due to the fact that the set of five problems was initially established with internally induced change in mind, and contact-induced change is often portrayed as an entirely different kind of change. The present article builds on the premise that the two types of change are not as different as the dichotomy suggests. It is an attempt to rethink how codeswitching, lexical borrowing, loan translation, interference and structural borrowing can be interpreted in the light of the five problems.

3.1 Codeswitching and borrowing introduced

Codeswitching is the use of two languages as part of the same discourse. It comes in many different forms, and it is not immediately obvious that all these forms should indeed be seen as different instantiations of the same phenomenon. For example, when a speaker talks to one interlocutor in one language and then turns to another addressing that person in a different language, this is referred to as codeswitching, but the same term is also used to describe the use of a single noun from one language in an utterance that is otherwise in another language. The first kind is an extreme example of what is often called the alternational subtype of codeswitching, while the second kind is referred to as insertional codeswitching. However, it is important to know that bilingual speech often contains constant back-and-forth switching between the languages, with clauses or parts of clauses in one language, sometimes containing insertional codeswitches themselves, alternating with chunks in the other. In such cases of intense codeswitching, the borderline between what is insertion and what is alternation is blurred.

The literature on codeswitching has been positively obsessed with one particular type of insertion: the single-word insertion. The reasons for this obsession speak directly to our current concerns, since they have to do with language change. A much debated question is whether a given single-word codeswitch should be seen as indeed a switch to the other language for the duration of that word or as the use of a loanword, i.e. a word that once indeed was used as an insertional codeswitch, but which by now has been incorporated into the lexicon of the base language. In the latter

case, its synchronic use merely represents selecting the appropriate word, one that happens to be etymologically traceable to another language.

It now becomes interesting to see how we should see this transition from codeswitch to borrowing in terms of WLH's five problems. Let us examine this in a preliminary way with the example of a Dutch word inserted in a Turkish clause, such as the verb *lenen* ('borrow') in the next example.

2. Ben seninkisini *lenen* yapacaktım, *toen had ik ze al*
I yours *borrow* was.going.to.do, *then had I them already*
"I was going to *borrow* yours, *but then* [*it turned out*] *I already had them*"

Here is what could be said about this example in relation to the five problems:

- Constraints: the pattern is typical for insertional codeswitching and loanwords alike. The Dutch-origin verb is a content word, inserted into the utterance in a way that preserves the morphosyntactic system of the base language. Like other verbs, it is used in a special construction in which it co-occurs in its infinitival form with an inflected Turkish auxiliary verb *yapmak* ('do'), forming a compound verb.
- Transition: since the speaker uses the word, it is probably part of her idiolect, and most likely she estimates it to be part of her interlocutor's idiolect as well; however, by definition the utterance allows no further extrapolation about the transition issue, i.e. how the word spread through the community; all we know at this point is that it was used at least once.
- Embedding: again, it is likely that the word is part of the speaker's and hearer's idiolects, and most likely this holds for many other people in the community. However, studying the utterance tells us nothing about the degree to which it is part of these idiolects, let alone of the dialect spoken by the community (i.e. the degree to which it is an established loanword); for that we need information on who does and who does not use or know it. We also do not know from this example what its position in the Turkish lexicon is. To what extent, for example, has it pushed out any Turkish equivalents? Studying the exact usage of the word does tell us something about its exact meaning, though, and may provide us with clues about meaning specialization: here the informant was talking about borrowing a pair of shoes from the interlocutor, and it is possible that the Dutch word is only used for borrowing particular

types of things, not others (e.g. money). However, once more we will not be able to tell on the basis of the single example found in speech: we need information on other uses of *lenen* ('to borrow') in Dutch Turkish and of its Turkish equivalents.

– Evaluation: on the basis of the utterance we can say that this particular case of usage of the word passed without comment, and this is circumstantial evidence that the use of this particular word and perhaps of Dutch words in general in Turkish clauses is acceptable behavior for this pair of speakers. Whether or not there is community-wide acceptance of bilingual speech, and whether or not this serves the expression of a dual identity as is sometimes suggested (cf. Sebba and Wootton, 1998) is impossible to say on the basis of the utterance alone.
– Actuation: while there is no way of knowing whether this is the first or the millionth occurrence of the Dutch word in bilingual speech, we do know that it was used once in a Turkish clause for the first time, and we may speculate about why that first occurrence happened. For clues, we might as well examine this particular utterance, since it might well be representative of other uses. We could speculate that the actuation of the change, i.e. of the addition of *lenen* to the Turkish lexicon, had to do with the attractiveness of its specific meaning. This requires the same close comparison as advocated for the Embedding problem above with the uses of its semantic neighbors, i.e. of Turkish (and perhaps other Dutch) words that mean something similar.

Several of these considerations imply a view in which codeswitching is seen as instantiating change in progress. However, this perspective is not often taken in the codeswitching literature, and I think this is because the link between synchrony (codeswitching) and diachrony (borrowing) has not been clearly theorized, though WLH already contained the seeds for this theory. Instead, most efforts have gone into a search, in my view a fruitless search (cf. Backus, 2015), for criteria with which to distinguish the two phenomena (e.g. Poplack and Dion, 2012). In fact, as I will argue below, they should be linked, not distinguished.

3.2 Interference and structural borrowing

If we do this same analysis for an instance of contact-induced structural change, very similar considerations appear. In Dutch Turkish speech, one finds occasional examples of the use of the demonstrative pronoun *o* ('that') in ways that seem to indicate it is being used like a definite article. An example is the following:

3. O filmi görmedim
that film I.didn't.see
"I didn't see the film" OR "I didn't see that film"

In TR-Turkish, i.e. Turkish as spoken in Turkey, only the meaning 'that film' would be conveyed. A natural hypothesis then is that *o* is increasingly used like a definite article. This surely reflects a common grammaticalization path, and the contact literature shows there are many cases around the world where contact with a language that has gone through this grammaticalization process has triggered the same change in the borrowing language (Heine and Kuteva, 2005). Something like this may well be going on in Dutch Turkish, though the change will only be in its initial stages: examples such as the one above have not been encountered very often at this point.

The above analysis would qualify as an account of the Constraints problem for this example. In the light of Matras (2009), one could add that the easy 'extension' (cf. Langacker 2008, p. 170) from demonstrative to definite article makes the translation link between *o* and *de* an easy pivot. Moreover, in the particular contact situation with Dutch, the category of definite article might represent an attractive category (Johanson, 2002) because of the shade of meaning it makes possible. The Actuation problem could be tackled with the suggestion that at some point, speakers are so dominant in Dutch that they often start conceptualizing their utterances using Dutch frames (their 'Thinking for Speaking' is in Dutch, cf. Slobin, 1996). In utterances such as the one in (3) this would trigger the need for a definite article (also see Matras, 2009; Heine and Kuteva, 2005). The Embedding problem would be answered with a statement about how the Turkish system of reference tracking would be affected by the addition of a definite article, which goes hand in hand with the loss of deictic force of the demonstrative, at least in some circumstances.

Once more, for the Transition problem it might be argued that if the only data that are available are recordings of conversations from a small number of speakers at one moment in time, there is no way in which we can answer it. The likely hypothesis is that there is a gradual spread through the idiolect of this speaker and probably in enough idiolects to say that it is gradually spreading through the community: individual speakers use the 'article' more and more often, and more and more individuals do it. As for Evaluation, one may hypothesize that the definite article-like use will still be perceived by many as incorrect or unconventional, but not by all or not by all categorically. The Actuation problem is perhaps the

hardest one to crack: why would the usage have started at the particular point in time at which it did? Even if the Thinking for Speaking analysis suggested in the previous paragraph is correct, we still need to know why conceptual interference happened at the moment that it did. Perhaps the structure to be borrowed needs to have a particular degree of entrenchment in the memory of the borrowing speaker, but we know very little about this. In addition, though the particular type of interference the example represents is a cross-linguistically common one, we still need to explain why it *is* common. Explanations for contact-induced changes are relatively easy to give after the fact, but for every change that we do observe there are others that could have occurred but did not. As WLH (Section 3.4, p. 186) observe: '[P]redictive hypotheses are not readily available'.

4 Usage-based linguistics and sociolinguistics

Several of these questions that are deemed fundamental for sociolinguistics in WLH are of equally fundamental importance to usage-based linguistics. The central hypothesis of this approach is that linguistic competence is the result of mentally storing linguistic experience with the help of human cognitive skills (Bybee, 2010). To investigate this hypothesis, one needs to address the sorts of questions asked by WLH. At least four of the ingredients of a usage-based account of linguistic knowledge (cf. Barlow and Kemmer, 2000) are also central to the sociolinguistic paradigm as envisaged by WLH. First is the claim that a linguistic theory needs to relate synchronic and diachronic perspectives to each other (i.e. articulate the relationship between synchronic use and diachronic change; cf. Croft, 2009 and WLH: Sections 2.2 and 3.1, pp. 139-140 and 153-154). Second, both traditions wish to assess the impact of use on knowledge (cf. the reference to Bybee, 2010 and WLH: Section 0, p. 101). Third, both focus on usage data as primary research data, rather than, for example, grammatical intuition (e.g. Gries, Hampe and Schönefeld, 2005, p. 636; WLH point this out throughout their article). Finally, both traditions emphasize that meaning is as important for a theory of language as form (e.g. Geeraerts, Kristiansen and Peirsman, 2010; WLH: Section 3.21, p. 159). The latter point surfaces in sociolinguistics primarily through the emphasis on pragmatic and social meaning (or 'indexicality') rather than on semantics, but it is important to recognize that all of this is meaning too. In fact, it provides a much-needed complementarity to the study of meaning, which tends to separate the

domains of semantics and pragmatics (including sociolinguistics) rather strictly.

All this suggests that an integration of usage-based linguistics and sociolinguistics is profitable, as first argued in Croft (2009), who draws attention to the fact that much of Cognitive Linguistics is silent on how language use is determined and much of Sociolinguistics is equally silent on the implications of language use for mental representation. Since usage-based theories hold that mental representation is based on linguistic experience and sociolinguistics actually studies linguistic experience, joining forces seems a good idea. In my opinion, the cognitive emphasis of usage-based linguistics, especially its associated emphasis on storage, entrenchment and automatic unconscious linguistic behavior, helps correct for the emphasis on conscious behavior that governs much of sociolinguistics (e.g. the emphasis on identity marking as the motivation for what people say and how they say it). At the same time, sociolinguistics can help usage-based linguistics to actually study usage in more detail than it normally does, and with greater attention for the contexts of usage and their impact on what is said. Slowly but surely, studies are getting off the ground that explicitly identify themselves as a combination of these two disciplines, e.g. Kristiansen and Dirven (2008), Geeraerts, Kristiansen and Peirsman (2010), Zenner (2013), Zenner and Kristiansen (2014). In the final sections of this article, I will discuss some recent contact linguistic studies that likewise combine the two perspectives in order to tackle issues not addressed before in this field.

I see this as just one of the ways in which contact linguistics can be moved into the orbit of general linguistics, in which contact issues have never been a central concern. While there has been no shortage of contact linguistic theories, there have been relatively few attempts to merge contact linguistic theories with theories about language in general, i.e. about the competence or mental representation of speakers. To be sure, there have been contributions in which contact data are accounted for using the framework of a linguistic theory (for example MacSwans, 2005 analysis of codeswitching data in a Minimalist framework), but few attempts exist that construct a theory that gives equal importance to contact and non-contact data. The reason for this, as I see it, is that language change has not been a central concern for linguistic theory, and actually even only to part of contact linguistics. In contrast, usage-based linguistics and sociolinguistics both argue that change should be a central concern for any linguistic theory.

It follows from the basic usage-based assumption that change must be

endemic. If usage determines mental representation, then mental representation must be dynamic, in constant flux (Bybee, 2010). This entails that it makes no sense, strictly speaking, to ask whether or not change has occurred, or if it has, at what point it did (WLH also struggled with the difficulty of casting something that is continuous into a model that uses discrete categories, cf. WLH: Section 3.4, p. 184). Of course, in reality we make a pragmatic decision: we call something a change if it has gone far enough and has affected enough people. By the same token, we ignore minute changes in degree of entrenchment, since the cumulative picture for many linguistic elements is that they are very stable over time (such as basic word order, the core meaning of words, etc.).

As a usage-based definition of change, most would probably subscribe to something like this: change is a fluctuation in the degree of entrenchment of a form-meaning unit, in which entrenchment stands for the strength with which a linguistic unit is stored in an individual speaker's memory. Special cases of such fluctuation include the emergence of a new unit (change from zero entrenchment to some entrenchment) and perhaps the disappearance of a unit from the inventory of units a speaker commands (change to zero entrenchment), though it is unclear whether things are ever really, truly, lost. Entrenchment is determined, it is generally argued, by frequency of use (Bybee, 2010). Most likely there are other determinants too, which perhaps can be grouped together under the heading of salience, but by far the most efforts have gone into examining the role of frequency. Corpus frequencies are often employed as a proxy for the frequency with which individual lexical and structural units (i.e. words and constructions) occur in the input and output of an individual speaker. To check to what degree corpus frequencies indeed correlate with degrees of entrenchment, such frequencies are often compared with participants' scores on experimental measures that supposedly tap into degree of entrenchment, such as lexical decision, self-paced reading or repetition tasks (cf. Arnon and Snider, 2010; Backus and Mos, 2011; Bannard and Matthews, 2008; Caldwell-Harris and Morris, 2008; Ellis and Simpson-Vlach, 2009). Generally, these correlations are pretty good, which inspires some confidence that frequency measures are reliable predictors of entrenchment. This is not to say that frequency is the only determining factor of entrenchment. In particular, there are all kinds of salience-enhancing factors that play a role in communication, and if a linguistic unit is made salient, this most likely also raises entrenchment levels, since salience entails enhanced cognitive attention and thus improved uptake (see Blumenthal-Dramé, 2012).

This new approach forces us to ask some new questions. Since sociolinguistics boasts fifty years of experience investigating language change, it is useful to see to what extent these new research questions overlap with the questions asked by sociolinguists. What we see is that some questions are similar but formulated differently. These cases of overlap will be identified below. However, there are also some questions the usage-based account forces us to ask which are generally not addressed in sociolinguistics. Not surprisingly, keeping in mind Croft's (2009) criticism that sociolinguistics does not concern itself enough with what goes on in the mind, these have to do with cognition.

In recent years, I have been involved in several studies on Dutch Turkish that adopt a usage-based sociolinguistic approach. In the remainder of this paper, I will discuss some of this research. Rather than focusing on detailed empirical results, for which I refer to the actual research papers, I will point out how these studies help answering questions that a usage-based approach to contact data forces us to ask, and how it brings us closer to fulfilling the research agenda that WLH envisaged.

5 Dutch Turkish data

The Dutch Turkish community is a relatively well-studied one. Many publications have examined the degree to which the community has maintained Turkish, the degree to which its members use Dutch, what their Dutch proficiency and their acquisition patterns are like, and, beyond language proper, to what degree they maintain an ethnic identity or accommodate towards the majority culture (Backus, 2013). Until recently, the linguistic studies have mostly focused on codeswitching. All of this work was based on recorded natural conversation, and generally concluded that in ordinary conversations both languages are used in mixed fashion, with a shift towards using more and more Dutch for people born in The Netherlands. The last few years witnessed a shift to a focus on the structure of Turkish, as there were good reasons to expect that in addition to the lexical influence of Dutch manifested by codeswitching, there was also 'covert', i.e. structural influence on Turkish grammar. Doğruöz (2008) and Doğruöz and Backus (2009) found quite a bit of 'unconventional structure', i.e. ways of saying something that employed collocations and constructions that differ from how it would be said in Turkey. Core syntax, however, was not affected, confirming a fairly robust finding of contact linguistics: while change is possible in any linguistic subsystem, it requires highly intensive

contact settings of long duration to bring really new syntactic structures into a language, structures that would have been ungrammatical in the ancestral variety. The Constraints problem forces us to find out why this is, and there is a relatively rich literature in contact linguistics tackling such issues. In my view, this is a domain where contact linguistics has a lot to offer for usage-based linguistics.

The hypothesis forcefully made by the usage-based approach is that structural changes should be no different from lexical changes as far as the basic mechanisms are concerned. The only difference should reside in whether the source of the change is more lexical or more structural, and the difference between lexicon and syntax is claimed to be gradual. The actual change is in all cases the diachronic fluctuation in entrenchment levels: an increase in the case of borrowing and a decrease in the case of loss, and all based on what speakers produce and hear in synchronic interaction. This is schematized in Table 1.

Table 1 Language contact phenomena

	Synchronic (selection)	**Diachronic (change)**
Lexical – overt	Insertional codeswitch	Lexical borrowing
Lexical – covert	Loan translation	Semantic borrowing / Collocational borrowing
Structural	Interference/Transfer	Structural borrowing

If speakers of Language A repeatedly select the same word from Language B in their everyday utterances, this can synchronically be described as an insertional codeswitch, but once it happens often enough, the word has diachronically become an established loanword in Language A. The same holds for the use of a foreign-inspired collocation or semantic extension (i.e. the use of a word from Language A with a meaning inspired by the meaning of its translation equivalent in Language B): if the combination is selected often enough synchronically, it becomes a case of collocational or semantic borrowing diachronically. Finally, the same picture also holds for the relationship between interference and structural borrowing.

In what follows, I will first explore to what extent existing data, comprising recordings of natural conversations, can be fruitfully analyzed to answer the usage-based and WLH-type questions. Subsequently I will introduce the issue of propagation, important for both WLH and usage-based approaches in their attempts to account for language change. It will be argued that corpus data are not very well-suited for answering questions about propagation. The final two subsections discuss data from two recent

studies on Dutch Turkish that both attempted to tackle this issue by gathering data through additional methods.

6 Conventional data: Analyzing corpus data

Let us start with a simple example: the use of a Dutch word in a Turkish clause, i.e. a case of insertional codeswitching, similar to example (2) above.

4. Sonra bi çarşıya gidek ben para *wisselen* yapıyım
   Later one mall.DAT go.IMP.3sg I money *change* do.OPT.1sg
   "After that let's go to the mall, I need to *change* money"

In this example, the speaker reports on a recent trip to Turkey and says that she needed to go change some euros into Turkish liras. If we look at it from the perspective of language change, various questions can be asked. Perhaps the most basic one is why the speaker used the Dutch word here. The speaker obviously needed to convey the concept of 'changing money', and one way in which to do that would be the one observed in the example. However, other options were possible: one of several conventional Turkish expressions, a complete switch to a Dutch clause, or a more elaborate codeswitch. One hypothesis could be that the unit that was best entrenched for the speaker was the one that connects this meaning with the Dutch form *geld wisselen*. The reader may now raise one or two eyebrows, and rightly so: there is no way I can support this claim on the basis of this example. The question of what has motivated the use of this particular Dutch word, and by extension the motivation behind any insertional codeswitching, has not received a satisfactory answer so far in the literature, mostly because the question itself has more or less fallen in between the cracks. The usual explanation is that a new word fills a lexical gap, at least whenever there was an obvious gap to fill, but most of the time the question is simply not asked at all. The Constraints and Actuation problems demand that we ask it, though.

In relation to WLH, I maintain that the reason why the question has not been asked much is that codeswitching has not always been recognized as relevant to language change. If, however, we frame the motivation question in terms of change, it becomes clear that we need to know why the word was used at all, as it provides information on why the word was borrowed. On the face of the evidence, i.e. the occurrence of the word in

a randomly recorded Turkish sentence, it is likely that the word *wisselen* is part of the Dutch Turkish lexicon. Since it is obviously Dutch in origin, and therefore has not always been part of Turkish, it must have been borrowed: it is an instantiation of lexical change. Once we accept that, the whole battery of questions outlined by WLH becomes relevant. Why is it used the way it is (Constraints)? How has it spread through the community, if it has (Transition)? How has it been affected by the borrowing process and how has it affected the system into which it was borrowed (Embedding)? Who uses it in what social settings (also Embedding)? Is it generally accepted as an established loanword or is it evaluated as an unnecessary fad (Evaluation)? And why was it borrowed when it was (Actuation)? None of these questions can really be answered with the data at hand, though we can make educated guesses for the Constraints and Actuation issues. For example, we know from massive empirical evidence that Dutch verbs used in Turkish clauses are always used in their infinitival form and are directly followed by an inflected form of the verb *yapmak* ('do'). We also know that Turkish incorporates foreign verbs in other immigration settings in the same way, that all Turkic languages do this, and that it is a widespread phenomenon all around the world. Accordingly, suggestions have been made as to why this pattern is so prevalent, though the actual answer is not a concern for the present paper. As for Actuation, we might advance the hypothesis that changing money is a concept typical of public settings (rather than the home domain), that therefore it may more often be talked about in Dutch than in Turkish by people like our participant, and that therefore the Dutch expression *geld wisselen* may be more entrenched than any of its Turkish equivalents (my sources prevaricate between *para bozdurmak* ('change money') and *liralar almak* ('buy liras') as the default translation). However, we cannot know whether this is true without independent evidence on the degrees of entrenchment of these units for this speaker. In addition, there is the question why we get the form we see in the actual example. If *geld wisselen* is indeed the source, the question is why we get the partial loan translation *para wisselen yap-* rather than either a complete preservation of the Dutch form (i.e. *geld wisselen yap-*, or a complete switch to Dutch: *ik moet geld wisselen*) or a complete loan translation. This could be *para bozdurmak*, already a unit in Turkish and therefore most likely analyzed as not a loan translation at all, but just the continuation, perhaps reinforced by the Dutch unit, of a Turkish convention; or it could be *para değiştirmek*, with a more default Turkish equivalent of *wisselen*: this word is also attested in Turkish in the collocation with

para, further complicating the issue of deciding what is Dutch influence and what is not.

So, while a usage-based approach provides plenty of suggestions, and asks interesting questions that WLH would also have asked, we run into problems because of the types of data we have. Theoretical innovation, in this case, also necessitates methodological renewal.

7 Propagation and the limitations of corpus data

The informal observations above concerning the Dutch verb *wisselen* will have shown that it is easy enough to think of possible motivations for its use, but also that it is very difficult to find actual proof for any of the scenarios one may come up with. Things improve a little if we conceptualize meaning as consisting of more than semantics, since that provides us with more possible motivations for why the foreign word was used. Consider the following example, from the same recorded conversation as the previous one.

5. *Nee*, Türkiye *sowieso* Samsun'da var *echt heel veel*.
   No Turkey *anyway* Samsun.LOC there.is *really very much*
   "*No*, in Turkey, there are in Samsun *anyway*, really *a whole lot of them*"

Why would the speaker have selected the Dutch phrase *echt heel veel*? Referentially, she was referring to the abundance of shops in the shopping district of the Turkish city of Samsun. In Dutch, the expression conveys more than just this referential meaning, though: it emphasizes ('*echt*') that there really are a whole lot of them, perhaps more than you would expect. This is what we could call the pragmatic meaning of the expression. It can be related to the denotational meanings of the elements that make up the expression, but it is unlikely that Dutch speakers arrive at the pragmatic connotation through a compositional route. Instead, we rely on numerous past experiences in which we have used or heard the expression with more or less the same pragmatic impact, and have the whole unit represented in our memory, complete with all three words, with some information about its usual syntactic position, and with the pragmatic connotation as part of its meaning. This insight brings us closer to answering the Actuation question, but we are not there yet. Note that there is quite a bit of Dutch in the example, which also contains the discourse marker *nee* ('no') and the adverb *sowieso* ('anyway'). The utterance has been lifted from a transcript

that is full of rapid back-and-forth switching between Turkish and Dutch. Many of the Dutch bits that are found in syntactically Turkish clauses are expressions from the domain of informal conversation. One could argue that in addition to the referential and pragmatic meanings motivating the selection of the phrase, there is also a social motivation. As the phrase is indexical of urban Dutch street language, and the speaker is a teenage girl from The Hague talking to her best friend, she may well be emphasizing her allegiance to that culture by using phrases that are emblematic of its speech style. After all, there are Turkish phrases too that carry the same denotational and pragmatic meanings. Whether or not identity marking is behind the selection of this phrase can only really be found out through ethnographic study (e.g. by interviewing and observing the speaker).

In relation to example 4, I mentioned that the evidence we have has little bearing on most of WLH's questions. Methodological limitations especially hinder investigation of the Transition problem. How do we know whether *wisselen* and *echt heel veel* are established borrowings in Dutch Turkish? I have argued that you cannot tell on the basis of their one-off occurrence in a random recording, but some reflection makes a somewhat more measured answer possible. There are three kinds of circumstantial evidence that these Dutch units are indeed established borrowings. First, they have relatively specific meaning, and this is typical for loanwords. Languages rarely borrow basic vocabulary: this is why the comparative method relies on the Swadesh-200 list and indeed perusal of the codeswitching literature turns up few examples of basic words from the other language being inserted, and plenty with relatively specific meaning (Backus, 2001; the specificity of *echt heel veel* derives from its pragmatics). Second, both can be argued to be related to Dutch culture or life in Holland, which means they could be concepts often conveyed in their Dutch form in the daily life of the speakers, which helps entrenching those forms. Third, the very fact that they are used in this example, and are not accompanied by metalinguistic commentary, suggests that they are in common use. Chances are that their capture on tape in the one half hour of the speaker's life that happened to be recorded reflects general usage rather than a unique case of borrowing for the first and only time. If they are in regular use, they are established lexical elements. Note that we do not wonder whether any of the Turkish words in these utterances, for example *çarşı* 'center' in Example 4, are established words in the lexicon of the speakers. If the issue of borrowing status is re-interpreted as a question of entrenchment in the mental lexicon of individual speakers, however, it would apply to all words, not just to those with a foreign origin. The default

assumption should be that words that are used are probably entrenched for the speaker, and most likely for the interlocutor too.

My separation of codeswitching and borrowing as synchronic and diachronic aspects of the same phenomenon puts me on a collision course with the tradition in the codeswitching literature associated with the work of Shana Poplack (e.g. Poplack and Dion, 2012), in which a principled *synchronic* distinction is made between codeswitching and borrowing. However, I am confident that once we have a sound theory in place that links synchronic usage and diachronic development, this debate can be laid to rest.

The final type of circumstantial evidence mentioned above has some further interesting implications. In agreement with much of the usage-based literature, the argument is that what speakers say reflects their linguistic competence, and their competence is built up on the basis of their usage, both in the sense of active use and that of passive exposure (Bybee, 2010). However, speakers speak to other speakers, and this is significant. One reason why the speakers in our examples feel free to use the Dutch words they do (and the Turkish ones, for that matter) is that they trust that their interlocutors know those words. In other words, as Clark (1996a) explains, we constantly make educated guesses about the competence of others, and since most communication runs fluidly, especially between friends, those guesses are usually quite good. This is because language is a system of conventions, and unless we are in communicatively novel situations with strangers, we know that the other has the same conventions. We use particular units because those units are entrenched in our own mental representations, and because we assume or guess that they are similarly entrenched in the mental representations of others, and thereby instantiate social conventions in our speech community. This explains the use of established loanwords, but also of any other lexeme. As an interim conclusion, it seems that we can definitely draw some conclusions from the analysis of corpus data, but that the answers for many of the questions remain tentative and unproven. More direct evidence is called for. We will discuss some of this evidence below, but first we need to turn to two new questions that a reinterpretation of WLH in the light of the usage-based tradition calls for.

The first question is this: How conventional is any given foreign word? Presuming we can find an answer to this question, the second question is: what determines its degree of success? These questions tie in well with WLH's Problems, particularly with the Transition problem. They also resonate with what could be called the first usage-based reincarnation of WLH:

the evolutionary account of language change put forward in Croft (2000). In this theory, change is presented as involving two stages: a first and usually short stage of innovation and a phase of propagation that usually takes quite long, sometimes centuries. Innovation is when a new unit or feature is introduced into the system; propagation is how it conventionalizes as the new norm. While this idea is not new in itself (see for example Keller, 1994), the way in which it is then developed into a comprehensive model that attempts to link synchronic behavior directly to diachronic development is typically the result of applying a usage-based approach to the problem.

Synchronic behavior is what one does in response to some communicative task (Clark, 1996a). Every utterance instantiates such a response and every utterance contains numerous selection events, for words, collocations, constructions, syntactic templates, discourse styles etc. As mentioned earlier, faced with wanting to say something, one can say the conventional thing ('normal replication') or say something new ('altered replication'). If the new variant is successful, it can get selected time and time again, and that turns its selection more and more into a case of normal replication: the change is propagating. In sociolinguistics, we speak of 'ongoing change' in such cases. Part of the research agenda inspired by Croft (and I think less so by WLH) is to understand what determines selection (of words, pronunciation variants, syntactic constructions, discourse styles, etc.). We select what is semantically useful for the present communicative task, what is pragmatically useful for the pragmatic or social meaning we may wish to convey, what satisfies our creative needs, what has been primed by previous mention, and what is entrenched. The first three types of motivation are often the focus of conscious attention, and perhaps for that reason they are favored in sociolinguistic analysis, unlike the fourth and fifth, which involve selection events that are more under the radar. Priming is typically studied in psycholinguistic lab experiments; entrenchment, i.e. the strength of storage of a form-meaning unit in an individual's mental representation, is the bread and butter of usage-based linguistics. Once a unit is introduced through an initial act of codeswitching, it has embarked on its path of propagation. Every time it gets selected, its degree of entrenchment goes up. How entrenched it is at any given moment is impossible to say with conversational data, especially if they are collected just once. Exactly the same scenario can be envisioned for loan translation and structural borrowing. What starts as a case of interference (an innovation) ends up as a conventionalized, propagated new collocation or structure that has its origins in the other language: that is

why it is contact-induced. Contact has a lot to do with the innovation, i.e. with Actuation, but not necessarily with propagation, i.e. with Transition, Embedding and Evaluation.

If we want to deal with these questions, we run up against some serious methodological limitations as long as we rely only on corpus data. Detailed and reliable frequency measurements are not easily available in ongoing contact settings, given that corpora are never going to be of sufficient size (Backus, 2014). This holds more for lexical variables than for syntactic ones, so in a corpus of Dutch Turkish one will probably find enough instances of for example subordination, case marking, word order or pro-drop to investigate whether Dutch-influenced variants are used significantly more often by speakers who are exposed to Dutch more. However, even then, it is likely that there are problems with the corpus that make generalization to the level of the Dutch Turkish community tricky.

For example, we might want to know whether units that do not occur in a corpus indeed do not occur in the language use of the relevant speech community in general. If they do occur in the corpus, we want to know how frequent they are in daily life. Corpus frequencies tell us something, of course, but their value is limited if the corpus itself is not representative. There are no truly large corpora of spoken bilingual data, and they may be too expensive to build, so we cannot generalize easily from the variation and distribution that we find in the corpus to language use in the speech community at large. In addition, most corpora will be snapshots of one moment in time, so diachronic developments, i.e. actual propagation, cannot be read off the distributional analyses very well.

To exemplify this, let's say we have observed in our conversational data that Dutch Turks often resort to Dutch to frame reported speech, and that they often seem to use the construction [Pronoun + *zo*], as in example (6). The reported speech is quoted directly, the quote is introduced with the adverb *zo* ('like'), and *zo* is preceded by the subject, which is usually a personal pronoun. The Dutch construction is probably a calque, formed on the basis of the popular English phrase 'and I'm like'. If we now want to ask the question whether Dutch Turkish has changed its grammatical system for conveying reported speech, and in particular whether it has propagated the construction [Pronoun + *zo*], what data would we need?

6. *Ja man. Zij zo* kesin bana öğret *dit dat. Ik zo* öğretiriz.
   "*yeah man. She was like* you definitely have to teach me *and so on. I was like* I'll teach you."

First of all, we would need data from a sufficiently wide sample of people to see whether the change has propagated for the entire community or for only a segment, e.g. for all Dutch Turks or only for young people, young women, or young women in the metropolitan area of Western Holland (known in The Netherlands as the *Randstad*). These questions speak to the Transition problem. In addition, we would need such data from various periods, so as to follow diachronic developments. We would also like to have enough data to be able to tell whether all pronouns can occur in the slot before *zo*, and whether other words than pronouns can co-occur – e.g. *m'n zus zo* ('my sister's like'), *de directeur zo* ('the manager's like') – and how often they do. These questions are relevant for answering the Embedding problem. Similarly, it would be relevant to know to what degree rival constructions, such as *ze zei* ('she said'), *zij van* ('she's all'; combining the pronoun and the genitive preposition *van* 'of', another quotative construction used in colloquial Dutch, cf. Mazeland, 2006), also occur, and how often and in what contexts.

These are all data we could in theory get from corpus analyses, as long as our corpus is representative enough. However, since it is not realistic we will ever get such corpora for spoken language use of minority groups, we need to look at other methodological possibilities. Interviews could be used to just ask speakers how common they think particular phenomena are, judgment tasks can be devised to tap into these intuitions in less direct ways, and it is also possible to design experimental tasks that allow conclusions about propagation. The next section draws on a study in which this was attempted.

8 Converging evidence: the case of Dutch Turkish subordination

The remainder of this paper serves to illustrate some recent work on Dutch Turkish in which we have attempted to supplement corpus data with such other sources of evidence.

8.1 Subordinate clauses: conversational, experimental and judgment data

The study I will focus on the most is the one Pelin Onar Valk and I undertook on subordination (Onar Valk and Backus, 2013; Onar Valk, 2013, 2014, 2015). Immigrant Turkish seems to be changing its way of making subordinate clauses. A preference for non-finite clauses that precede the matrix

verb has changed to a preference for finite clauses that come after the matrix verb. These structures resemble how Dutch forms subordinate clauses.

Finite subordination with verb-initial structure is illustrated in example (6). Its non-finite and verb-final equivalent is given in example (7). The non-finite subordinated verb in (7) uses a nominalization marker to turn 'love' into a non-finite form: with the addition of a possessive it means something like 'your loving', and this whole complex is marked with accusative case because it functions as the direct object of 'see'.

6. gör-üyo-m [siz birbiriniz-i çok sev-iyo-nuz]
see-PROG-1SG [you each other-ACC very love-PROG-2PL]
"...I see [that you love each other a lot]..."

7. [siz birbiriniz-i çok sev-diği-niz-i] gör-üyo-m
[you each other-ACC very love-NMNL-2PL-ACC] see-PROG-1SG
"...I see [that you love each other a lot]..."

Onar Valk (2015) has amassed various kinds of evidence for the conclusion that Immigrant Turkish has indeed gone over to predominantly using finite and post-verbal subordinate clauses. These pieces of evidence include observations in a corpus of various types of natural speech, responses to an experimental task that elicited subordinate clauses, and responses to a judgment task in which various types of subordinate clause were judged for the degree to which they sounded 'normal'.

In the conversational data, NL-Turkish speakers significantly more often used finite and post-verbal clauses than a control group from Turkey. In elicited imitation data, in which speakers had to repeat a sequence of three or four sentences immediately after hearing them, NL-Turkish speakers significantly more often preserved finite subordinate clauses and post-verbal position if these had been present in the stimulus item than TR-Turkish speakers did, who instead tended to change such items to non-finite and verb-final structures. About a fourth of the time, NL-Turkish speakers even produced the Dutch-like structures if the stimulus item had contained the TR-Turkish non-finite and/or preverbal structure, something TR-Turkish speakers never did. Finally, when asked to rate the conventionality of stimulus items containing the various subordination structures, NL-Turkish speakers once again consistently rated finite subordination and verb-initial structures like the one in (6) as more conventional than TR-Turkish speakers did. Interestingly, for the conventional non-finite

and verb-final structures as exemplified in (7), the data did not support a contact effect, as Dutch Turks gave these structures the same ratings as the control group did. We will come back to this interesting detail below. By and large, though, we can say that the evidence from various sources converged: NL-Turkish speakers preferred the Dutch-like structures. The question is how to account for this. A usage-based interpretation would be that the finite and verb-initial structures are more entrenched for NL-Turkish speakers, and therefore activated more easily than the TR-Turkish equivalents. The higher entrenchment levels would reflect that Dutch structures are more entrenched in general for these speakers, consistent with the finding that Dutch Turks often report that they are better in Dutch than in Turkish (cf. Backus, 2013 for a review). Taking the usage-based explanation further, this could mean that sentence planning often goes through Dutch (recall the reference above to Slobin's concept of 'Thinking for Speaking'), or that this has happened in the past, leading to the current situation where these structures are simply entrenched well in the mental representations of these speakers. Abundant use of finite and verb-initial structures further increases their frequency, and thus their entrenchment in the competence of individual speakers.

It is encouraging that the three kinds of evidence converge. It means that the impression we would have formed about the propagation of the change in question, i.e. information pertaining to WLH's Transition problem, on the basis of just recorded conversation would actually have been accurate. This suggests we can be relatively confident after all about the generalizations we might wish to propose on the basis of such recordings.

8.2 Alternational codeswitching and changes in subordinate clauses

So far all codeswitching-related phenomena we looked at were cases of insertional codeswitching (though *echt heel veel* in example 5 is not simple insertion), and the link between this kind of mixing and lexical borrowing is conceptually obvious. However, even alternational codeswitching should be studied with an eye to its role in language change. Consider the following example:

8. O gün İlahe geldiydi ya *wij gingen zo chillen je weet toch* Esra'ya falan gittiydik. *Zij naar haar je weet toch?*
"That day İlahe had come, right, *we went chilling, remember,* we went to Esra and so. *Off to her she goes, remember?*"

We have not done the requisite quantitative analysis yet, but it seems from initial observation that the speakers we recorded often switch to Dutch for the phrase *we gingen* 'we went' to encode this kind of future aspect in the past tense. Perhaps this phrase can be said to have been borrowed into Dutch Turkish? If a minority language can be shown to borrow phrases like this, the boundary between Turkish and Dutch starts to break down, and this is not incompatible with what we see sometimes in the development of mixed languages, as the ultimate result of intensive language contact (Estigarribia, 2015). To see how conventional this particular construction has become, various aspects of the WLH paradigm would have to be investigated:

- In bilingual speech, how often is this phrase uttered in Dutch? (Transition)
- What is the relative distribution of this phrase and its Turkish equivalent(s)? Is the Turkish equivalent used at all? (Embedding)
- In bilingual speech, does it always follow a Turkish clausal unit? (Embedding)
- Do many people use it and do they always use it? (Embedding)
- What do people think about it? (Evaluation)

Contact linguistics has mostly been interested in the origin question, which in WLH terms relates to Actuation and Constraints. These can be investigated with some degree of success by analyzing recordings, as shown by much of the literature and my remarks above. The other issues, represented by the list of questions above, are more profitably addressed with the sorts of methods used by Onar Valk (2015). Elicited imitation, for example, allows one to control for many of the influences that play a role in ordinary conversation, and yields more data points so that statistical analysis becomes possible.

Alternational codeswitching is normally studied to see for what pragmatic or identity-related reason a person has switched from one language to the other. Consider one more example: in Turkish-Dutch contact data, we often see the Dutch phrase 'I think' (in various forms) co-occurring with Turkish phrases that convey the content of the reported thought, as in the following example (9). Fully Turkish forms with the same finite and verb-initial structure are rare in our data (hence I have added a constructed example in 10 that illustrates what it would look like).

9. *ik vind* kültürle, kültürle dini karıştırıyorlar... [attested]
"*I think* they are mixing up culture and religion ..."

10. düşünüyorum ki kültürle, kültürle dini karıştırıyorlar... [constructed] I.think that they are mixing up culture and religion ..."

What contact linguists are interested in, first of all, is why the speaker used this codeswitch. Framing the question that way means one is led towards a pragmatic analysis and perhaps one would conclude that, typically, Turkish-Dutch bilinguals use Dutch for discourse markers and Turkish for content. Once more, we do not need to be concerned at this point with whether or not that is the correct analysis: what interests me here is why this question is singled out as the single most interesting one. My impression is that this too has to do with the synchronic outlook of codeswitching studies: the motivations for what people say are located in the immediate environment of the communicative situation. Of course, there is nothing wrong with that, but it is not the only hypothesis one could entertain. Notably, if a usage-based approach is adopted, it is important to consider how synchronic behavior affects diachronic development. We might expect that the speaker says it this way simply because she always says it this way: using the Dutch discourse marker has become the convention, sheer frequency has entrenched it, and it is activated as soon as the intended meaning needs to be conveyed. Its origin is still to be explained in terms of pragmatic usefulness and attractiveness, but its synchronic selection may well be more automatic, not linked to any conscious pragmatic or social agency.

9 Judgment tasks: the case of loan translations

Several recent studies of Dutch Turkish have employed judgment tasks (e.g. Doğruöz & Gries, 2012; Onar Valk, 2015; Şahin, 2015). While judgment tasks tend to be associated with issues of grammaticality, we used the task to assess to what degree participants felt that a particular construction or collocation was heard often in everyday conversation. We argue that the answers say something about the degree of conventionality of a given phrase. Many aspects of the methodology need to be developed further, such as how exactly to interpret what participants are doing when they answer such questions. The jury is still out on what the best way is, or whether there even will turn out to be a best way. In studies on Immigrant Turkish we have tried out a few ways of asking:

– How common is this phrase in the language you hear around you?
– How common is it for you to use this phrase?

– To what degree is this the normal way of saying this?
– How well do you like this phrase?

One interesting result was found consistently in all studies so far. On the one hand, the judgments of NL-Turkish speakers tended to confirm their preference for Dutch-influenced collocations and structures (such as the Dutch-style subordination structures discussed above) by giving such structures high ratings. On the other hand, however, they often join Turkish control groups in giving the same high ratings to conventional TR-Turkish collocations and structures. What this suggests is that Transition is not just a matter of the distribution of forms in usage: the TR-Turkish units that are being displaced by the incoming Dutch-influenced forms may start to disappear from use, but they are still well entrenched in people's mental representations.

There were two exceptions to this pattern. Many observers of Immigrant Turkish have noted that the word *almak* ('to take') is used in co-occurrence with names for transport vehicles such as the bus, the train, the subway etc. TR-Turkish employs a dative-marked construction here with the verb *binmek* ('to get on'). Similarly, NL-Turkish speakers tend to say *piyano oynamak* (with the verb for 'play', modeled on Dutch *spelen*) rather than TR-Turkish *piyano çalmak* (with the verb for 'hit'), and do this with any other musical instrument, too. Most likely, of course, these are loan translations from Dutch, German or whatever other European language is spoken by the individual immigrant, as all these languages use the words for 'to take' and 'to play' in these combinations. A WLH-free contact linguistics would be content to show that these various calques exist, and maybe point out that they form families of structures. However, taking the WLH agenda seriously involves asking additional questions, which would, again, also be asked from a usage-based perspective: have the immigrant speakers just stored various combinations with *almak* and with *oynamak* in their mental representations, or have they extracted the partially schematic constructions N + *almak* and N + *oynamak*? This pertains to the Embedding problem. And of course there is again the perennial Transition problem-inspired question how the changes have propagated and how far along they are. Also, if we take Actuation seriously, just stating that these are calques is not enough. There are many other lexical combinations that *could* be calqued. What makes these particular ones attractive, or their TR-Turkish equivalent vulnerable? At what point did the Dutch origin start influencing the NL-Turkish lexicon, and why? Perhaps the answers have something to do with the frequency with which the concepts

are used in conversation, the cultural associations with the country of immigration, and the semantic basicness of the words involved (similar constructions are based on combinations of the names for school subjects and programs with the word for 'do'). However, none of these explanations seem promising for the combinations of musical instruments with the word for 'play'. Investigating these hypotheses requires work that so far is not really undertaken by contact linguists.

In our judgment tasks, these two loan translations were the exception to the rule that judgments of TR-Turkish conventions remained high. In these particular cases, participants even gave the NL-Turkish expressions much higher ratings than the inherited TR-Turkish ones, indicating that the changes have propagated to near completion (Şahin, 2015). In most cases, the putative NL-Turkish conventions tended to still get lower scores than their TR-Turkish equivalents, even where we observed exclusive use, sometimes by the same participants, of the NL-Turkish variant in conversational data. This suggests, as we saw for the data on subordination, that relying only on production data provides us with a limited picture of change.

10 Conclusion

Work within the usage-based approach has by now progressed far enough that we can sketch an account of how languages change within this framework. In the beginning, there are some first cases of interference in the form of codeswitching, loan translation or structural interference. The way in which this happens is relevant to an investigation of the Constraints problem, and identifying the reasons why a particular change got underway, and why it did when it did provides input for the Actuation problem. Further propagation is the topic of the Transition problem. These units get entrenched through repeated usage, and more and more people in the community keep increasing their entrenchment of them, until it has become the new conventional way of saying it. In the case of syntactic templates, high type frequency helps to entrench a pattern, and therefore to increase its productivity. The degree of conventionality can perhaps also be read off the responses to judgment data, perhaps even better than is possible with usage data. The degree to which this affects the larger system of the receiving language (Embedding problem) can be investigated by checking how alternative ways of saying the same thing are used, including the inherited words and phrases that are ostensibly being replaced. How

this can be undertaken was illustrated through our study on subordinate clauses in NL-Turkish. Metalinguistic data should be gathered along the way to see whether the new structure grows organically, through repeated use and ever higher frequency, or whether it catches people's attention, becoming a salient sociolinguistic marker (Evaluation problem). Knowing about this is important because it will affect the course of change. Interestingly enough, all of these issues are equally relevant from the viewpoints of WLH and of usage-based linguistics.

Expanding the methodological palette in ways suggested above holds the potential of increasing our insight into the processes of internally and externally induced change. Of course, focusing on variation is only germane to the investigation of changes that are in progress, when there is no way to know whether they will propagate to the logical endpoint. Much of this was already attended to in WLH. Given the affinity between usage-based and sociolinguistics it is not surprising that their research agendas for studying language change are relatively similar. The usage-based approach does focus on some different aspects, however, and brings in some methodological innovation that has the potential to rejuvenate sociolinguistics. It will be exciting to witness the further propagation of *that* change.

Bibliography

Arnon, Inbal & Snider, Neal 'More than words: Frequency effects for multi-word phrases', *Journal of Memory and Language*, 62 (2010), pp. 67-82.

Backus, Ad, 'The role of semantic specificity in insertional codeswitching: Evidence from Dutch Turkish'. In: Rodolfo Jacobson (ed.), *Codeswitching Worldwide* II. Berlin & New York: Mouton de Gruyter, 2001, pp. 125-154.

Backus, Ad, 'Turkish as an immigrant language in Europe'. In: Tej K. Bhatia & William C. Ritchie (eds), *The Handbook of Bilingualism and Multilingualism*, 2nd edition. London: Blackwell, 2013, pp. 770-790,

Backus, Ad, 'A usage-based approach to codeswitching: The need for reconciling structure and function'. In: Gerard Stell & Kofi Yakpo (eds), *Code-switching Between Structural and Sociolinguistic Perspectives*. Berlin & New York: Mouton de Gruyter, 2015, pp. 19-37.

Backus, Ad & Mos, Maria, 'Islands of (im)productivity in corpus data and acceptability judgments: Contrasting two potentiality constructions in Dutch'. In: Doris Schönefeld (ed), *Converging Evidence. Methodological and theoretical issues for linguistic research*. Amsterdam: John Benjamins, 2011, pp. 165-192.

Backus, Ad & Spotti, Massimiliano, 'Normativity and change: Introduction to the Special Issue on agency and power in multilingual discourse', *Sociolinguistic Studies*, 6 (2012), pp. 185-208.

Bannard, Colin, & Matthews, Danielle, 'Stored word sequences in language learning: The effect of

familiarity on children's repetition of four word combinations', *Psychological Science*, 19 (2008), pp. 241-248.

Barlow, Michael & Kemmer, Suzanne (eds.), *Usage based models of language*. Stanford: CSLI Publications, 2000.

Bell, Alan, 'Back in style: reworking audience design'. In: Penelope Eckert & John R. Rickford (eds), *Style and Sociolinguistic Variation*. Cambridge: Cambridge University Press, 2001, pp. 139-169.

Blumenthal-Dramé, Alice, *Entrenchment in Usage-Based Theories: What Corpus Data Do and Do not Reveal about the Mind.* Berlin & New York: Mouton de Gruyter, 2012.

Bybee, Joan, *Language, Usage and Cognition*. Cambridge: Cambridge University Press, 2010.

Caldwell-Harris, Catherine, & Morris, Alison, 'Fast Pairs: A visual word recognition paradigm for measuring entrenchment, top-down effects, and subjective phenomenology', *Journal of Consciousness and Cognition*, 17 (2008), pp. 1063-1081.

Christiansen, Morten H. & Chater, Nick, 'Language as shaped by the brain', *Behavioral & Brain Sciences* 31 (2008), pp. 489-558.

Clark, Herbert, *Using language*. Cambridge: Cambridge University Press, 1996a.

Clark, Herbert, 'Communities, commonalities, and common ground'. In: John Gumperz & Stephen Levinson (eds), *Rethinking linguistic relativity*. Cambridge: Cambridge University Press, 1996b, pp. 324-355.

Croft, William, *Explaining Language Change: An Evolutionary Approach*. Harlow: Longman, 2000.

Croft, William, 'Towards a social cognitive linguistics'. In: Vyvyan Evans & Stéphanie Pourcel (eds), *New Directions in Cognitive Linguistics*. Amsterdam: John Benjamins, 2009, pp. 395-420.

Doğruöz, Ayşe Seza, Synchronic variation and diachronic change in Dutch Turkish: A corpus based analysis. Ph.D. dissertation, Tilburg University, 2008.

Doğruöz, Ayşe Seza & Backus, Ad, 'Innovative constructions in Dutch Turkish: An assessment of on-going contact-induced change', *Bilingualism: Language and Cognition* 12 (2009), pp. 41-63.

Doğruöz, Ayşe Seza & Gries, Stefan Th., 'Spread of on-going changes in an immigrant language. Turkish in the Netherlands', *Review of Cognitive Linguistics,* 10 (2012), pp. 401-426.

Ellis, Nick, & Simpson-Vlach, Rita, 'Formulaic language in native speakers: Triangulating psycholinguistics, corpus linguistics and education', *Corpus Linguistics and Linguistic Theory*, 5 (2009), pp. 61-78.

Estigarribia, Bruno, 'Guaraní-Spanish Jopara Mixing in a Paraguayan novel. Does it reflect a third language, a language variety, or true codeswitching?' *Journal of Language Contact*, 8 (2015), pp. 183-222.

Geeraerts, Dirk, Kristiansen, Gitte & Peirsman, Yves, *Advances in Cognitive Sociolinguistics*. Berlin & New York, Mouton de Gruyter, 2010.

Gries, Stefan Th., Hampe, Beate & Schönefeld, Doris, 'Converging evidence: Bringing together experimental and corpus data on the association of verbs and constructions' *Cognitive Linguistics*, 16 (2005), pp. 635-676.

Heine, Bernd & Kuteva, Tania, *Language Contact and Grammatical Change*. Cambridge: Cambridge University Press, 2005.

Jaeger, T. Florian & Snider, Neil, 'Alignment as a consequence of expectation adaptation: syntactic priming is affected by the prime"s prediction error given both prior and recent experience', *Cognition*, 127 (2013), pp. 57-83.

Johanson, Lars, *Structural Factors in Turkic Language Contacts*. London: Curzon, 2002.

Keller, Rudi, *On language change: the invisible hand in language*. London: Routledge, 1994.

Kristiansen, Gitte & Dirven, René (eds.), *Cognitive Sociolinguistics: Language variation, cultural models, social systems*. Berlin & New York: Mouton de Gruyter, 2008.

Labov, William, *Sociolinguistic patterns*. Philadelphia: University of Philadelphia Press, 1972.

Langacker, Ronald W., *Cognitive Grammar. A Basic Introduction*. Oxford: Oxford University Press, 2008.

MacSwan, Jeff, 'Codeswitching and generative grammar: A critique of the MLF model and some remarks on "modified minimalism"', *Bilingualism: Language and Cognition*, 8 (2005), pp. 1-22.

Matras, Yaron, *Language Contact*. Cambridge: Cambridge University Press, 2009.

Mazeland, Harrie, '"VAN" as a quotative in Dutch: Marking quotations as a typification'. In: Tom Koole, Jacomine Nortier & Bert Tahitu (eds), *Artikelen voor de 5e Sociolinguïstische Conferentie*. Delft: Oburon, 2006, pp. 354-365.

Onar Valk, Pelin & Backus, Ad, 'Syntactic change in an immigrant language: From non-finite to finite subordinate clauses in Turkish', *Journal of Estonian and Finno-Ugric Linguistics*, 4 (2013), pp. 7-29.

Onar Valk, Pelin, 'Dutch Turkish diverging from Turkey-Turkish: A judgment task study on how Dutch Turkish employs subordination and word order', *Working Papers of the Linguistics Circle of the University of Victoria, Canada,* 23 (2013), pp. 158-176.

Onar Valk, Pelin, 'Convergent Developments on Dutch Turkish word order: A comparative study using 'elicited production' and 'judgment' data: Converging evidence?', *Applied Linguistics Review*, 5 (2014), pp. 351-372.

Onar Valk, Pelin, *Transformation in Dutch Turkish Subordination? Converging evidence of change regarding finiteness and word order in complex clauses*. Utrecht: LOT, Tilburg University Ph.D. dissertation, 2015.

Pickering, Martin J. & Garrod, Simon C., 'Toward a mechanistic psychology of dialogue', *Behavioral and Brain Science*, 27 (2004), pp. 169-226.

Poplack, Shana & Dion, Nathalie, 'Myths and facts about loanword development', *Language Variation and Change*, 24 (2012), pp. 279-315.

Şahin, Hülya, *Cross-linguistic influences: Dutch in contact with Papiamento & Turkish* Utrecht: LOT, Radboud University Ph.D. dissertation, 2015.

Sebba, Mark & Wootton, Tony, 'We, they and identity: Sequential versus identity-related explanation in code-switching'. In: Peter Auer (ed), *Code-Switching in Conversation: Language, Interaction and Identity*. London: Routledge, 1998, pp. 262-286.

Slobin, Dan, 'From "thought and language" to "thinking for speaking"'. In: John Gumperz & Stephen Levinson (eds), *Rethinking linguistic relativity*. Cambridge: Cambridge University Press, 1996, pp. 70-96.

Weinreich, Uriel, Labov, William & Herzog, Marvin, 'Empirical foundations for a theory of language change'. In: Winfried P. Lehmann & Yakov Malkiel (eds), *Directions for Historical Linguistics: A Symposium*. Austin: University of Texas Press, 1968, pp. 95-195.

Zenner, Eline, Cognitive Contact Linguistics. The macro, meso and micro influence of English on Dutch. Ph.D. dissertation, University of Leuven, 2013.

Zenner, Eline & Kristiansen, Gitte (eds.), *New Perspectives on Lexical Borrowing*. Berlin & New York: Mouton de Gruyter, 2014.

About the author

Ad Backus, Tilburg University, Tilburg School of Humanities, Department of Culture Studies.
E-mail: a.m.backus@tilburguniversity@edu

Wat dragen we vandaag: een hemd met blazer of een shirt met jasje?

Convergentie en divergentie binnen Nederlandse kledingtermen

Jocelyne Daems, Kris Heylen & Dirk Geeraerts

MEM 67 (2): 307–342
DOI: 10.5117/TET2015.2.DAEM

Abstract

What to wear today: *een hemd met blazer* ('a dress shirt with suit jacket') or *een shirt met jasje* ('a dress shirt with suit jacket')? Convergence and divergence in Dutch clothing terminology

This paper reports on a corpus-based investigation into naming preferences in Belgian Dutch and Netherlandic Dutch for fourteen clothing terms. The study is a follow-up of Geeraerts, Grondelaers and Speelman (1999), in which soccer and clothing terminology from 1950, 1970 and 1990 was analysed as an indicator of standardisation in Dutch. This study extends the clothing corpus with new, comparable data from 2012 collected from magazines and shop windows. A profile-based measure of linguistic uniformity quantifies the differences in naming preferences across the 14 concepts between different varieties of Dutch. The results shed new light on the current linguistic situation in the Low Countries. The diachronic convergence between Belgian Dutch and Netherlandic Dutch found in Geeraerts, Grondelaers and Speelman (1999) seems to have come to a halt in present-day Dutch. On the other hand, the recent data confirm that the distance between the language in the lower register shop windows and the standard language in magazines remains largest in Belgian Dutch.

Keywords: profile-based uniformity measure, lexical variation, diachronic variation, register variation, onomasiology, pluricentricity

1 Inleiding[1]

Eind vorige eeuw presenteerden Geeraerts, Grondelaers en Speelman (1999) een veelzijdige studie naar de verhouding tussen het Nederlands in België en dat in Nederland[2]. Lexicale variatie bij vijftien kledingconcepten en evenveel voetbalbegrippen werd kwantitatief bestudeerd om na te gaan of er diachroon sprake was van convergentie of divergentie tussen de twee taalvariëteiten. Hiertoe werd krant- en magazinemateriaal uit 1950, 1970 en 1990 verzameld. Tegelijk werd de contemporaine interne taalstructuur van het Belgisch Nederlands en het Nederlands Nederlands vergeleken door het bovengewestelijke schriftelijke standaardtaalgebruik, gerepresenteerd door de kledingtijdschriften, te plaatsen naast dat van het regionaal gekleurde substandaardtaalniveau, zoals te vinden op prijskaartjes en labels in winkeletalages. Een derde pijler van de studie vertrok van attitudinele gegevens om het effect van normatieve taalzorg na te gaan en onderzocht daarbij eveneens de waarde van geëliciteerde data naast feitelijk taalgebruik. Methodologisch bestond de uitdaging erin om een nieuwe maat in te voeren om de onderzoeksvragen op een kwantitatief eenduidige wijze te beantwoorden. De hiertoe ontwikkelde profielgebaseerde uniformiteitsmaat laat toe om van geobserveerd taalgebruik te vertrekken om de gestandaardiseerdheid van de taalsituatie in kwestie te onderzoeken. Het achterliggende idee is immers dat uit een gemeenschappelijke linguïstische norm, hier de Nederlandse standaardtaal, gemeenschappelijk taalgebruik volgt. Enerzijds steunt de profielgebaseerde uniformiteitsmaat op het idee dat de overlapping in woordkeuze bij het uitdrukken van een concept tussen twee taalvariëteiten aanduidt hoe uniform die laatste zijn. Anderzijds laat de maat ook toe om te aggregeren over verschillende concepten zodat algemenere uitspraken over lexicale variatie tussen taalvariëteiten mogelijk worden. De twee belangrijkste empirische vaststellingen die volgden uit Geeraerts, Grondelaers en Speelman (1999) zijn de diachrone lexicale convergentie tussen het Nederlands Nederlands en het Belgisch Nederlands, grotendeels op het conto van die laatste, en de synchroon grotere afstand tussen de onderzochte registers in Nederlandstalig België (zie de laatste alinea van sectie 2 voor een meer uitgebreid overzicht van de resultaten).

Sindsdien hebben verschillende studies voortgebouwd op deze basis. Grondelaers et al. (2001b) hernemen enerzijds het kledingtermenonderzoek en voegen een nieuw corpus van internetmateriaal toe als bijkomende vertegenwoordiging van het meer informele niveau. Anderzijds stappen de onderzoekers over van lexicale variatie bij inhoudswoorden

naar die bij functiewoorden. In Grondelaers et al. (2001a) wordt dieper ingegaan op het gebruik van internetmateriaal bij onderzoek naar lexicale standaardisering. Zenner, Speelman en Geeraerts (2012) gebruiken de uniformiteitsberekening om het succes van Engelse leenwoorden na te gaan bij persoonsaanduidende nomina. Daarnaast zijn er verscheidene studies waarbij de vergelijking van verschillende methodes centraal staat, en vorderingen in het onderzoek naar automatische synoniem- en betekenisbepaling laten toe dat de schaal van het onderzoek aanzienlijk vergroot wordt, mede door de toenemende beschikbaarheid van grote corpora. Speelman, Grondelaers en Geeraerts (2003; 2006) vergelijken aan de hand van vijftien profielen (een aangepaste versie van) de profielgebaseerde uniformiteitsmaat uit Geeraerts, Grondelaers en Speelman (1999)[3] met niet-profielgebaseerde methodes met als doel om taalvariëteiten zo goed mogelijk te classificeren. Met hetzelfde doel, maar om andere methodes te vergelijken, vertrekken Ruette, Speelman en Geeraerts (2011) van 1455 profielen met telkens een typisch Belgisch-Nederlands gemarkeerde variant en Heylen en Ruette (2013) van 218 semiautomatisch geselecteerde concepten.

Deze vervolgstudie keert terug naar de oorsprong en neemt twintig jaar later de draad van de kledingtermen weer op. Voor veertien kledingconcepten zijn opnieuw verscheidene Belgische en Nederlandse (mode)tijdschriften doorzocht en winkeletalages bezocht[4]. De diachrone dimensie wordt zo uitgebreid met een nieuwe sprong van een twintigtal jaar. De convergerende trend tussen Belgisch Nederlands en Nederlands Nederlands kan nu verder onderzocht worden met data uit het tweede decennium van de eenentwintigste eeuw. Daarbovenop, terwijl er in Geeraerts, Grondelaers en Speelman (1999) enkel synchrone data voorhanden waren om variatie binnen respectievelijk het Belgisch en het Nederlands Nederlands te onderzoeken, beschikken we met de toevoeging van het nieuwe etalagemateriaal ook over een diachroon stratificationeel perspectief tussen 1990 en 2012.

Zoals bij Geeraerts, Grondelaers en Speelman (1999) moet dit lexicale variatieonderzoek gezien worden in het licht van de taalpolitieke geschiedenis van het Nederlands in België en de positie die het daarbij inneemt ten opzichte van het Nederlands Nederlands. Voor een uitgebreide beschrijving van de historische en taalpolitieke achtergrond bij het standaardiseringsproces van het Belgisch Nederlands verwijzen we naar Geeraerts, Grondelaers en Speelman (1999, pp. 12-29) (maar zie ook Van Hoof en Jaspers 2012 voor een kritisch overzicht van de standaardiseringsacties in Vlaanderen in de periode 1950-1980). Het volstaat in deze context om te herhalen dat het Belgisch Nederlands een standaardiseringsachterstand

heeft opgelopen onder meer door de invloed van het Frans en dat de taalpolitiek vanaf de jaren zestig van de vorige eeuw bewust aansluiting met de Nederlands-Nederlandse norm nastreefde. Die integrationistische visie staat in contrast met het particularisme, waar er ruimte is voor een eigen Belgisch-Nederlandse standaardtaal. In deze context kan niet onvermeld blijven dat het idee van een standaardiseringsachterstand van het Belgisch Nederlands betwist wordt (Van Hoof & Jaspers 2012, pp. 97-98), of alleszins als problematisch beschouwd wordt (Grondelaers & van Hout 2011a, p. 200). Die laatsten wijzen er immers op dat de beschrijving van de standaardisering van het Belgisch Nederlands in het gedrang komt wanneer de Nederlands-Nederlandse situatie als ijkpunt genomen wordt, doordat beide nationale variëteiten historisch zeer verschillende standaardiseringsprocessen kennen (2011a, p. 200).

Het onderzoek van Geeraerts, Grondelaers en Speelman (1999) neemt de tweedeling, integrationisme tegenover particularisme, als vertrekpunt, maar houdt een descriptieve benadering aan om na te gaan bij welke visie de feitelijke data het best aansluiten. Hetzelfde geldt overigens voor dit artikel. Met de empirische bevindingen in deze studie beogen we niet om uitspraken te doen over de historische standaardiseringsprocessen van het Belgisch en Nederlands Nederlands. Wel gaan we van naderbij bekijken welke verschillen er tussen deze twee nationale variëteiten bestaan in de kledingwoordenschat. Zo zal onder meer blijken dat in de laatste twintig jaar er in het Belgisch Nederlands andere verschuivingen plaatsvinden dan in het Nederlands Nederlands.

In sectie 2 van dit artikel wordt het onderzoek van Geeraerts, Grondelaers en Speelman (1999) samengevat, in het bijzonder het deel met betrekking tot het kledingmateriaal. We lichten ook uitvoerig de profielgebaseerde uniformiteitsmaat toe. Daarna wordt in sectie 3 het huidige vervolgonderzoek voorgesteld en enkele methodologische aanpassingen toegelicht. Eveneens wordt er uitgebreid ingegaan op de huidige resultaten en wordt de vergelijking gemaakt met de originele studie. Sectie 4 is gewijd aan de conclusies en we bespreken verdere uitbreidingen in het studiedomein.

2 Corpusgebaseerd onderzoek naar convergentie en divergentie in de Nederlandse kledingwoordenschat

Het onderzoek in Geeraerts, Grondelaers en Speelman (1999) vertrekt ruwweg van twee basishypotheses. De diachrone convergentiehypothese

steunt op de verwachting dat in de loop van de standaardiseringsbeweging van het Belgisch Nederlands de twee nationale variëteiten convergeren als gevolg van de expliciete normatieve oriëntatie op het Nederlandse Nederlands, kortom de integrationistische visie. Daarnaast veronderstelt de synchrone stratificatiehypothese een grotere afstand tussen het regionale en supraregionale niveau in België door de veronderstelde standaardiseringsachterstand. Immers, de bovengewestelijke standaardtaal in het Belgisch Nederlands is gebaseerd op een exogene norm en is niet het gevolg van een natuurlijk proces dat vanuit een (uitverkoren) dialect begonnen is.

Om na te gaan in welke mate er uniformiteit is tussen en binnen het Belgisch Nederlands en het Nederlands Nederlands werd een gebalanceerd corpus samengesteld met een geografische, diachrone en stratificationele dimensie. Voor de synoniemen van vijftien kleding- en vijftien voetbalconcepten werden observaties uit Belgische en Nederlandse kranten en tijdschriften uit 1950, 1970 en 1990 verzameld. Daarbovenop werden voor het kledingmateriaal benamingen op labels en etiketten uit etalages in twee Belgische en twee Nederlandse steden opgenomen. Leuven en Leiden golden hierbij als vertegenwoordiging van respectievelijk de linguïstische centra Brabant en de Randstad, terwijl Kortrijk en Maastricht dan weer gelegen zijn in economische centra met een secundaire academische functie, en zich daarenboven bevinden in de periferie van het taalgebied. In tegenstelling tot de benamingen in tijdschriften, werden deze labels niet geredigeerd door taalprofessionals en is de relatie met de potentiële koper veel directer. Het lezerspubliek van het etalagemateriaal is bovendien ook beperkter dan dat van het tijdschriftenmateriaal, maar de ruime regionale commerciële functie van de geselecteerde steden belet dat we te maken hebben met zuiver dialectisch taalgebruik. Dit semiformele schriftelijk brontype vormt zo een aanvulling op het nationaal gerichte tijdschriftenmateriaal binnen het stratificationele continuüm van het Nederlands. De twee gekozen lexicale velden bestaan bovendien telkens uit een representatieve selectie van basisbegrippen binnen het veld. Een gedetailleerde omschrijving van de materiaalverzameling is te vinden in Geeraerts, Grondelaers en Speelman (1999, pp. 30-35; pp. 162-164), voor een summier overzicht verwijzen we naar onderstaande tabel:

Tabel 1 Kwantitatief overzicht per lexicaal veld, land en brontype in aantal woorden

			België	Nederland	totaal
Voetbal	1950	magazine/krant	3283	3220	
	1970		4529	1604	
	1990		4850	4215	21.801
Kleding	1950		2210	3555	
	1970		3858	2972	
	1990		2450	3427	18.472
centrum (Leuven/Leiden)		etalage	3289	1162	
periferie (Kortrijk/Maastricht)			3299	1694	9444
Totaal			27.768	21.849	49.617

Om de lexicale convergentie tussen het Belgisch Nederlands en het Nederlands Nederlands te meten, introduceerden Geeraerts, Grondelaers en Speelman (1999) de profielgebaseerde uniformiteitsmaat. Hierbij is het onomasiologisch profiel van een bepaald concept in een gegeven bron het geheel van alternatieve benamingen voor dat concept uitgedrukt in de frequentieverhouding in die bron. Onderstaande tabel[5] geeft de onomasiologische profielen weer voor het concept OVERHEMD (MAN) in het Belgische en Nederlandse standaardtaalmateriaal uit 1990 (respectievelijk B90 en N90 genoemd):

Tabel 2 Onomasiologische profielen voor OVERHEMD (MAN) in het Belgische en Nederlandse materiaal uit 1990

overhemd (man)	B90	N90
hemd	31%	17%
overhemd	69%	46%
shirt	0%	37%

De uniformiteitsmaat drukt uit in hoeverre twee onomasiologische profielen overeenkomen. We spreken van maximale uniformiteit wanneer twee taalvariëteiten, door een gedeelde norm, voor eenzelfde begrip dezelfde benaming gebruiken. Het is ook mogelijk dat verschillende benamingen gebruikt worden voor een gegeven concept, maar dan moet de distributie hiervan identiek zijn in de beide variëteiten. Wanneer dit niet het geval is, wanneer de alternatieve lexicaliseringen binnen een profiel slechts gedeeltelijk hetzelfde gedistribueerd zijn, is er sprake van partiële uniformiteit.

Om de uniformiteit voor een bepaald concept te berekenen, meten we

de overlapping in alternatieve benamingen tussen twee profielen. Anders gezegd, voor iedere alternatieve lexicalisering nemen we de kleinste relatieve frequentie van de twee profielen (gemarkeerd in grijs in Tabel 2) en hiervan berekenen we de som om tot de uniformiteitsmaat te komen. Voor het concept OVERHEMD (MAN) houdt dat in dat de keuze voor de variant *hemd* in 17% van de gevallen overlapt tussen België en Nederland. Voor *overhemd* delen Belgisch Nederlands en Nederlands Nederlands 46% voorkeur en doordat in België *shirt* niet gebruikt wordt om OVERHEMD (MAN) uit te drukken, is er daar geen overlapping. Kortom, voor het begrip OVERHEMD (MAN) in Belgische en Nederlandse standaardtalige bronnen uit 1990 vinden we (17% + 46% + 0% =) 63% uniformiteit. Omgezet in formulevorm ziet deze intuïtieve berekening er zo uit:

$$U_Z(Y_1, Y_2) = \sum_{i=1}^{n} \min\left(F_{Z,Y_1}(x_i), F_{Z,Y_2}(x_i)\right)$$

We kijken dus naar de mate van onomasiologische uniformiteit U tussen de bronnen, of deelverzamelingen, Y_1 en Y_2 voor concept Z dat uitgedrukt wordt door de alternatieve benamingen x_1 tot x_n. Deze formule drukt uit dat de graad van uniformiteit berekend wordt door het minimum te nemen van de relatieve frequentie van variant x_i voor concept Z in de materiaalverzameling Y_1, weergegeven als $F_{Z,Y_1}(x_i)$, en van dezelfde variant voor hetzelfde concept in de deelverzameling Y_2. Dezelfde berekening wordt herhaald voor alle varianten x_i die concept Z aanduiden en vervolgens worden alle minima opgeteld. Merk op dat voor mogelijke verschillen in conceptfrequentie tussen de deelverzamelingen gecontroleerd wordt door met de relatieve in plaats van de absolute frequenties te werken.

Om te aggregeren over verschillende concepten en te komen tot de gemiddelde uniformiteit voor een lexicaal veld wordt het gemiddelde genomen van de uniformiteitsgraad van alle concepten Z_i in het veld, zoals onderstaande formule illustreert:

$$U(Y_1, Y_2) = \frac{1}{n}\sum_{i=1}^{n} U_{Z_i}(Y_1, Y_2)$$

Tot slot is er nog een aanvulling op deze grondformule die het relatieve gewicht van het concept in het veld mee in rekening brengt. Immers, in communicatief opzicht is het logisch dat hoogfrequente concepten sterker doorwegen bij het meten van de uniformiteit van een lexicaal veld. De gewogen uniformiteit U' neemt de som van de uniformiteitswaarde van

elk concept vermenigvuldigd met de wegingsfactor G voor dat concept. Die laatste is de relatieve frequentie van concept Z_i in de gehele deelverzamelingen Y_1 en Y_2. De formule voor de gewogen uniformiteit wordt dan:

$$U'(Y_1,Y_2) = \sum_{i=1}^{n} U_{Z_i}(Y_1,Y_2) \cdot G_{Z_i}(Y_1 \cup Y_2)$$

Naast de onomasiologische uniformiteit introduceerden Geeraerts, Grondelaers en Speelman (1999) nog twee maten die toelaten een beter inzicht te krijgen in het materiaal. De interne uniformiteit drukt uit in hoeverre in een deelverzameling voor een bepaald concept eenzelfde benaming wordt gekozen. Twee factoren beïnvloeden de interne uniformiteit, het aantal alternatieve termen en de dominantie van een bepaalde term. De interne uniformiteit is de som van de relatieve frequentie in het kwadraat[6] van ieder van de alternatieve benamingen voor een concept:

$$I_Z(Y) = \sum_{i=1}^{n} F_{Z,Y}(x_i)^2$$

Voor het voorbeeld van OVERHEMD (MAN) uit Tabel 2 houdt dit in dat de interne uniformiteit voor het concept in het Belgisch Nederlands (0.31^2 + 0.69^2 =) 57,22% is en in het Nederlands Nederlands slechts (0.17^2 + 0.46^2 + 0.37^2 =) 37,74%. Deze maat drukt met andere woorden uit in hoeverre er een dominante term is om het lexicale concept uit te drukken. Om dan de interne uniformiteit van een lexicaal veld in een bepaalde bron te berekenen nemen we het gemiddelde van de interne uniformiteitswaarde van alle concepten in het veld:

$$I(Y) = \frac{1}{n}\sum_{i=1}^{n} I_{Z_i}(Y)$$

Net als bij de (externe) uniformiteitsmaat speelt de conceptfrequentie een rol en wordt de wegingsfactor geïntroduceerd om een accurater beeld te krijgen van de interne uniformiteit binnen het lexicale veld:

$$I'(Y) = \sum_{i=1}^{n} I_{Z_i}(Y) \cdot G_{Z_i}(Y)$$

Tot slot, om het relatieve aandeel van termen met een bepaald kenmerk (zoals 'normatief afgekeurd' of 'van Franse oorsprong') te berekenen in het

bronmateriaal, is de aandeelmaat A geïntroduceerd. Bij deze maat staat K voor de verzameling van alle lexicaliseringen met dat kenmerk. De wegingsfactor W, een waarde van 0 tot 1, bepaalt voor een individuele variant in welke mate het lidmaatschap tot K moet doorwegen. Zo is *jeans* nog steeds ontegensprekelijk herkenbaar als van Engelse oorsprong, terwijl bij *bloes* de afstand tot het Frans iets groter is. Dat onderscheid uit zich in verschillende wegingen bij de berekening van de aandeelmaat[7]. De onderstaande formule toont vervolgens dat de aandeelmaat berekend wordt door de som van de relatieve frequentie van iedere lexicalisering voor een concept vermenigvuldigd met zijn specifieke wegingsfactor:

$$A_{K,Z}(Y) = \sum_{i=1}^{n} F_{Z,Y}(x_i) \cdot W_{x_i}(K)$$

Als we dan opnieuw gaan kijken wat dat concreet betekent voor het concept OVERHEMD (MAN) uit Tabel 2, vinden we dat enkel de variant *shirt* een Engelse term is. Voor het Belgisch Nederlands is het aandeel voor het Engels dus 0%, want er zijn geen observaties van shirt, en voor het Nederlands-Nederlandse materiaal is het 37%. We berekenen vervolgens het gemiddelde om te komen tot het aandeel van een bepaald taalkundig kenmerk in een lexicaal veld:

$$A_K(Y) = \frac{1}{n} \sum_{i=1}^{n} A_{K,Z_i}(Y)$$

Ook hier is er een gewogen alternatief, waarbij de frequentie van het concept Z in de gehele deelverzameling Y meegerekend wordt:

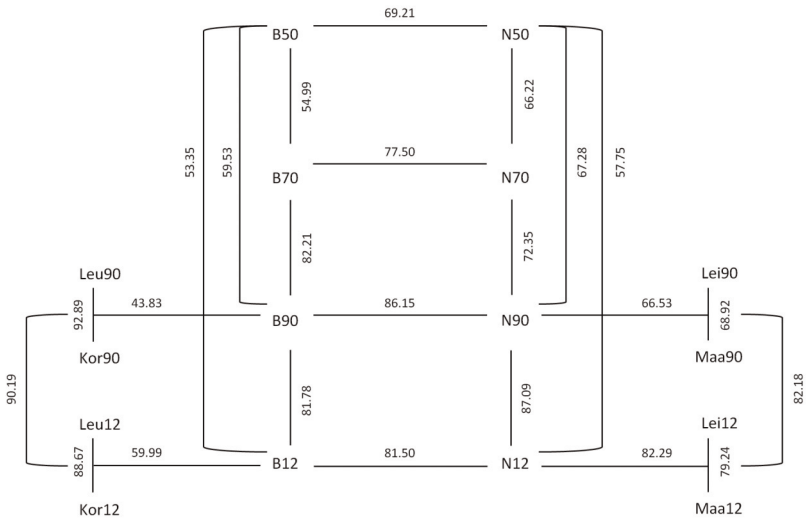
$$A'_K(Y) = \sum_{i=1}^{n} A_{K,Z_i}(Y) \cdot G_{Z_i}(Y)$$

Vanuit deze systematische methodologische onderbouw is het een kleine stap om de diachrone convergentiehypothese en de synchrone stratificatiehypothese te operationaliseren in termen van uniformiteit. De convergerende beweging tussen het Belgische en Nederlandse Nederlands manifesteert zich in een stijgende uniformiteit over de tijdspanne van 1950 tot 1990. De grotere synchrone afstand tussen standaardtaal en substandaardtaal in België laat zich kwantificeren als een lagere uniformiteit tussen het tijdschrift- en het winkeletalagemateriaal in België dan in Nederland. We

geven hierbij de voornaamste resultaten uit Geeraerts, Grondelaers en Speelman (1999) weer, relevant voor dit vervolgonderzoek.

De data in Figuur 1[8] tonen dat er in de tijdspanne van 1950 over 1970 tot 1990 onmiskenbaar sprake is van lexicale convergentie tussen Belgisch Nederlands en Nederlands Nederlands. De convergerende trend die waargenomen wordt, is bovendien voornamelijk te wijten aan een normatieve oriëntatie van het Belgisch Nederlands op het Nederlands Nederlands: typisch Nederlands-Nederlandse termen worden overgenomen ten koste van typisch Belgisch-Nederlandse uitdrukkingen. Verder geldt algemeen dat de interne uniformiteit van de concepten toeneemt: een dominante term binnen het profiel krijgt duidelijk de voorkeur op de onomasiologische alternatieven. Verder, zowel het Belgisch als het Nederlands Nederlands worden gekenmerkt door een toenemende invloed van Engelse kledingtermen. Voor het aandeel Franse termen wordt de ontwikkeling van het Belgisch Nederlands gekenmerkt door een duidelijk observeerbare afname als historische reactie tegen de Franse invloed, terwijl in het Nederlands-Nederlandse materiaal die afwijzing afwezig is. Opmerkelijk is wel dat het gebruik van Franse termen in Nederland onderhevig lijkt te zijn aan modegrillen: het aandeel "chique" Franse kledingtermen daalt aanzienlijk in 1970, wat toegeschreven kan worden aan de informele tijdsgeest van toen (Geeraerts en Grondelaers, 1999, p. 62). Tot slot, de veronderstelde vertraagde standaardisering van het Belgisch Nederlands wordt weerspiegeld door een beduidend grotere afstand tussen de supraregionale tijdschriftobservaties en het regionale winkeletalagemateriaal dan het geval is bij het Nederlands Nederlands.

Hoewel deze bevindingen steunen op een stevige empirische basis, blijft voorzichtigheid geboden bij de interpretatie ervan. Zo steunt het onderzoek op slechts twee lexicale velden en zijn de bevindingen dus niet zonder meer extrapoleerbaar naar andere woordvelden of woordsoorten. Ook de tweeledige stratificatie binnen het Nederlands in het onderzoek is slechts een beperkte representatie van de werkelijke registerdiversiteit (zie sectie 1 voor vervolgstudies die op deze twee punten uitbreidingen hebben voorgesteld). Tot slot, de omvang van het manueel aangelegde corpus laat niet toe om bepaalde statistische testen uit te voeren (met name de lineaire regressieanalyse). Voor de vergelijkingen van uniformiteitsindices wordt dus veiligheidshalve gewerkt met een foutenmarge van 5% alvorens met zekerheid gezegd kan worden of een uniformiteitswaarde groter al dan niet kleiner is dan een andere.



Figuur 1 Gewogen uniformiteitsverhoudingen binnen het kledingmateriaal
Bron: Geeraerts, Grondelaers en Speelman, 1999, p. 69

3 Convergentie en divergentie binnen Nederlandse kledingtermen anno 2012

Om de vraag te beantwoorden of de synchrone stratificatiehypothese en de diachrone convergentiehypothese twintig jaar later nog steeds standhouden, werd het initiële corpus uitgebreid. De originele samenstelling werd hierbij zo goed mogelijk gerespecteerd. Voor het standaardtaalmateriaal werd gezocht naar observaties van kledingtermen in mode- en lifestylemagazines uitgegeven in 2012. Twee aanpassingen waren hierbij nodig. Zo waren sommige van de oorspronkelijk gebruikte tijdschriften niet langer in de handel verkrijgbaar en waren niet alle tijdschriften van de maanden januari en februari 2012 verkrijgbaar (zie Appendix 3 voor een overzicht). Voor het etalagemateriaal werden opnieuw winkeletalages in dezelfde twee Belgische en twee Nederlandse universiteitssteden uitgekamd[9]. Anders dan in de oorspronkelijke studie hebben we het concept MANTELPAKJE (VROUW) achterwege gelaten omwille van dataschaarste. Dit houdt eveneens in dat voor het materiaal uit Geeraerts, Grondelaers en Speelman (1999) alle berekeningen opnieuw uitgevoerd zijn voor veertien in plaats van vijftien concepten. Tabel 3 toont de samenstelling in aantal observaties voor veertien kledingconcepten in alle deelverzamelingen:

Tabel 3 **Kwantitatief overzicht voor de kledingtermen per land en brontype in aantal woorden**

	België	Nederland	totaal
1950	1918	3114	
1970	3389	2824	
1990	2347	3412	**17.004**
centrum (Leuven/Leiden)	3240	1141	
periferie (Kortrijk/Maastricht)	3210	1652	**9243**
2012	2844	4363	**7207**
centrum (Leuven/Leiden)	504	355	
periferie (Kortrijk/Maastricht)	413	537	**1809**
totaal	**17.865**	**17.398**	**35.263**

Voor het testen van de hypotheses[10] in deze vervolgstudie vertrekken we van dezelfde methodes als Geeraerts, Grondelaers en Speelman (1999). Aan de basis ligt de uniformiteitsmaat en die wordt aangevuld met de interne uniformiteitsmaat en de aandeelmaat voor linguïstische kenmerken. We zullen onze interpretaties baseren op de gewogen uniformiteit U' omdat die maat de reële verhoudingen in het taalgebruik het best reflecteert.

We brengen in herinnering dat de statistische betrouwbaarheid van de berekeningen uit Geeraerts, Grondelaers en Speelman (1999) beperkt werd door de omvang van het corpus, maar dat die beperking opgevangen werd door een marge van 5% te hanteren bij het vergelijken van de resultaten. In Speelman, Grondelaers en Geeraerts (2003, p. 321; 2006, p. 192) wordt een bijkomende stap geïntroduceerd bij de uniformiteitsberekening, en die nemen wij in deze studie ook. Om na te gaan of een profiel uit de ene variëteit zonder meer vergeleken mag worden met het profiel voor hetzelfde concept in een andere variëteit, passen we eerst de *log likelihood ratio test*[11] toe. Deze toets test de aannemelijkheid dat de twee profielen significant van elkaar verschillen. Indien dit niet het geval is, moeten we aannemen dat de uniformiteit maximaal (100%) is. Dat wil zeggen dat we op basis van de beschikbare data moeten concluderen dat de verschillende (absolute) frequenties voor de lexicaliseringen in de twee profielen te wijten zijn aan het toeval. In het andere geval, als de twee profielen wel significant van elkaar verschillen, kunnen we uitgaan van de betrouwbaarheid van de uniformiteitsmaat. We hanteren het gebruikelijke significantieniveau van 5%[12]. Parallel hebben we ook de uniformiteitsberekeningen

gemaakt waarbij we deze voorafgaande aannemelijkheidsquotiënttest niet toepassen om na te gaan wat het effect ervan is. Hoewel de uniformiteit wat verschilt in waarde, zijn de waargenomen tendensen nagenoeg dezelfde. Waar er afwijkingen zijn, vermelden we die ook bij de resultaten. Tabel 6 in Appendix 2 geeft een overzicht van alle berekende uniformiteitswaarden.

Voor we tot de bespreking van de resultaten overgaan, verklaren we eerst de voorstellingswijze van de uniformiteitswaarden. Eerst geven we telkens schematisch de algemene tendens weer, om dan daaronder dezelfde vergelijking te kwantificeren. Helemaal rechts is tussen haken steeds een verwijzing te vinden naar het aantal concepten waarop de berekening gebaseerd is. Immers, in het materiaal van 1950 hebben we geen observaties van drie concepten, namelijk T-SHIRT (MAN/VROUW), LEGGING (VROUW) en JEANS (MAN/VROUW). Voor de diachrone convergentiehypothese vergelijken we daarom slechts uniformiteitswaarden gebaseerd op de gedeelde elf concepten. Wanneer we enkel data bekijken uit 1990 en 2012, zoals voor de stratificatiehypothese, gaan we uit van alle veertien kledingconcepten.

De convergentiebeweging die geconstateerd werd in Geeraerts, Grondelaers en Speelman (1999) zet zich niet voort in het nieuwe materiaal en een divergerende trend lijkt zich te manifesteren[13]:

(1)	U'(B50,N50)	<	U'(B70,N70)	<	U'(B90,N90)	>	U'(B12,N12)		
	69,21		77,50		86,15		81,50		(*11C*)

De onderlinge toenadering tussen het Belgisch Nederlands en het Nederlands Nederlands in woordkeuze is niet langer af te leiden uit deze dataverzameling. Bijkomend kan dan de vraag gesteld worden aan welke kant de meeste verschuivingen plaatsvinden. Als de uniformiteit tussen de periodes lager is in België kunnen we daaruit afleiden dat de ontwikkelingen daar sterker zijn. In Geeraerts, Grondelaers en Speelman (1999) was de verwachting dat de convergentie kwam door wijzigingen aan de Belgische kant, wat vooral bevestigd werd door de vergelijking op lange termijn in (2). Ondanks de recente divergerende ontwikkeling vinden we in (3) dat de verschuivingen over de periode van 1950-2012 nog altijd enigszins sneller gebeuren aan Belgische kant. Hoewel een interval van deze lengte andere effecten kan maskeren, toont (4) dat ook voor de laatste tijdsafstand de bevindingen hetzelfde zijn. Het is wel zo dat de ongewogen uniformiteitsberekeningen in (5) een licht ander beeld tonen, namelijk dat de uniformiteit in Belgisch Nederlands iets hoger ligt. Wanneer we de data van nader-

bij bekijken, vinden we dat die vertekening grotendeels veroorzaakt wordt door de uniformiteit voor het laagfrequente concept KOSTUUM (MAN). Hier zien we duidelijk het effect van het niet mee in rekening brengen van de frequentie van concepten, wat de keuze om onze interpretaties te baseren op de gewogen uniformiteitsberekeningen bekrachtigt. In grote lijnen kunnen we stellen dat het verschil in ontwikkelingssnelheid met het Nederlands Nederlands geringer wordt en lijkt te stagneren[14].

(2)	U'(B50,B90)	<	U'(N50,N90)	
	59,53		67,28	(*nC*)
(3)	U'(B50,B12)	<	U'(N50,N12)	
	53,35		57,75	(*nC*)
(4)	U'(B90,B12)	<	U'(N90,N12)	
	81,78		87,09	(*nC*)
(5)	U(B50,B12)	>	U(N50,N12)	
	49,94		45,39	(*nC*)

In Geeraerts, Grondelaers en Speelman (1999) werd bovendien met de nodige voorzichtigheid verondersteld dat standaardisering interne variatie reduceert. Immers, als alle taalgebruikers dezelfde benaming kiezen voor een bepaald begrip, is het taalgebruik in die situatie maximaal gestandaardiseerd. Hoewel het niet vaststaat hoeveel interne variatie normaal is voor een gestandaardiseerde taalsituatie, kan de Nederlands-Nederlandse situatie als ijkpunt genomen worden. De interne uniformiteit in (6) fluctueert maar in grote lijnen is er sprake van een toename in het Belgisch Nederlands. De voornaamste ontwikkeling gebeurde hierbij tijdens de eerste periode (1950-1970). Hoewel de interne uniformiteit van het Nederlands Nederlands in (7) over het algemeen hoog blijft, is deze ook onderhevig aan schommelingen. Ook in een gestandaardiseerde taal kunnen immers steeds onomasiologische verschuivingen optreden. Wanneer we kijken naar de interne uniformiteit voor alle veertien concepten bij het recentste materiaal in (8) vinden we zelfs een hogere interne uniformiteitswaarde voor het Belgisch Nederlands. Een mogelijke verklaring is dat het Belgisch Nederlands na de inhaalbeweging stopt met vernieuwen, aangezien er geen externe bron van variatie meer is, namelijk de exonormatieve oriëntatie. Deze stilstand kan dan een indicatie zijn van de mogelijke zoektocht naar een eigen Belgisch-Nederlandse norm, wat aansluit bij de particularistische visie.

(6)	I'(B50)	<	I'(B70)	>	I'(B90)	≅	I'(B12)		
	61,06		77,68		72,61		77,56	(*nC*)	
(7)	I'(N50)	≅	I'(N70)	<	I'(N90)	≅	I'(N12)		
	71,05		65,86		74,72		75,34	(*nC*)	
(8)	I'(B12)	>	I'(N12)						
	79,35		73,71	(*14C*)					

Algemeen wordt aangenomen dat de ontwikkeling van het Belgisch Nederlands gekenmerkt wordt door een reactie tegen de Franse invloed, een afkeuring die afwezig geacht wordt in het Nederlands Nederlands (Geeraerts en Grondelaers 2000, p. 54). Voor termen van Engelse oorsprong daarentegen wordt een toegenomen aandeel verwacht. Voor zowel het Nederlands in België (9) als in Nederland (10) vinden we dat het aandeel Franse termen daalt, een tendens die overigens nog sterker is bij het ongewogen materiaal waar de laagfrequente concepten harder doorwegen. Zoals we al eerder aanhaalden, kan het grillige patroon voor het Franse aandeel bij het Nederlands-Nederlandse materiaal gerelateerd worden aan de trendgevoeligheid van het Nederlands Nederlands.

Eveneens vanuit taalintern perspectief vinden we dat het aandeel Engelse termen in beide variëteiten in (11-12) stijgt. Bovendien is dit patroon nog sterker wanneer we alle veertien kledingconcepten beschouwen in (13), aangezien JEANS (MAN/VROUW), LEGGING (VROUW) en T-SHIRT (MAN/VROUW) net nieuwe concepten zijn waarvoor aanvankelijk de Engelse term werd overgenomen. In (13) stellen we eveneens vast dat de Engelse invloed minder sterk aanwezig is bij het Nederlands Nederlands dan bij het Belgisch Nederlands. Terwijl de taalzuivering in Nederlandstalig België wegens historische redenen vooral gallicismen in het vizier nam, lijkt de houding tegenover anglicismen permissiever geweest te zijn. Nederland lijkt de Engelse invloed van meet af aan meer te weren.

(9)	A'_{FR}(B50)	>	A'_{FR}(B70)	≅	A'_{FR}(B90)	>	A'_{FR}(B12)		
	29,00		14,90		17,63		11,83	(*nC*)	
(10)	A'_{FR}(N50)	<	A'_{FR}(N70)	<	A'_{FR}(N90)	>	A'_{FR}(N12)		
	39,39		20,78		31,71		19,39	(*nC*)	
(11)	A'_{EN}(B50)	≅	A'_{EN}(B70)	≅	A'_{EN}(B90)	<	A'_{EN}(B12)		
	2,87		3,29		7,83		15,18	(*nC*)	
(12)	A'_{EN}(N50)	≅	A'_{EN}(N70)	≅	A'_{EN}(N90)	≅	A'_{EN}(N12)		
	1,75		2,87		7,80		7,71	(*nC*)	
(13)	A'_{EN}(B12)	>	A'_{EN}(N12)						
	23,92		17,40	(*14C*)					

Samenvattend kunnen we zeggen dat de toenadering tussen Belgisch Nederlands en Nederlands niet langer volgt uit het materiaal van 2012. Om de standaardiseringsachterstand in te halen zijn de verschuivingen diachroon altijd sneller geweest aan de Belgische kant, maar recent wordt het verschil met het Nederlands Nederlands minder duidelijk. Ook voor de interne uniformiteit zijn beide variëteiten tot op hetzelfde niveau genaderd. Voor het Belgisch-Nederlandse kan er hier sprake zijn van een normverschuiving weg van het Nederlands Nederlands. Verder lijkt zich naast de algemene afname van de Franse invloed in beide variëteiten, in Nederland eveneens een meer terughoudende reactie te manifesteren ten opzichte van het Engels.

Naast de ontwikkelingen op het standaardtaalniveau, laat de opbouw van de dataverzameling ook toe om de taalsituatie binnen de nationale variëteiten zelf te bestuderen. Het etalagemateriaal vertegenwoordigt het regiolectische niveau en wordt in de bespreking aangeduid met de stadsnaam waarvan de observaties afkomstig zijn. De hoofdhypothese in Geeraerts, Grondelaers en Speelman (1999) luidde dat er door de veronderstelde standaardiseringsachterstand in het Belgisch Nederlands een grotere afstand zou zijn tussen het standaard- en substandaardtaalgebruik. Zoals de berekeningen in (14) en (15) tonen, werd die verwachting destijds bevestigd en blijft dat ook het geval in de huidige dataset. Wel vinden we zowel voor het Belgisch Nederlands (16) als voor het Nederlands Nederlands (17) dat de afstand tussen regionaal en supraregionaal taalgebruik is afgenomen in de periode 1990-2012.

(14)	U'(B90,LeuKor90)	<	U'(N90,LeiMaa90)		
	46,86		67,72		(*14C*)
(15)	U'(B12,LeuKor12)	<	U'(N12,LeiMaa12)		
	63,95		81,65		(*14C*)
(16)	U'(B90,LeuKor90)	<	U'(B12,LeuKor12)		
	46,86		63,95		(*14C*)
(17)	U'(N90,LeiMaa90)	<	U'(N12,LeiMaa12)		
	67,72		81,65		(*14C*)

Dat de afstand tussen de twee registers groter is in België dan in Nederland impliceert overigens niet dat de onderlinge variatie tussen de regionale meetpunten ook groter zou zijn in België. Net als in de oorspronkelijke studie in (18), vinden we daarvoor ook nu geen bewijs in (19). Bovendien bevestigen de interne uniformiteitswaarden in (20-23) dat het Belgisch Nederlands niet heterogener is dan het Nederlands Nederlands.

(18) U'(Leu90,Kor90) > U'(Lei90, Maa90)
93,56 87,33 (*14C*)
(19) U'(Leu12,Kor12) > U'(Lei12, Maa12)
90,05 82,06 (*14C*)
(20) I'(B90) ≅ I'(N90)
72,10 73,69 (*14C*)
(21) I'(LeuKor90) > I'(LeiMaa90)
86,25 73,20 (*14C*)
(22) I'(B12) > I'(N12)
79,35 73,71 (*14C*)
(23) I'(LeuKor12) > I'(LeiMaa12)
76,28 69,56 (*14C*)

Doordat deze vervolgstudie de opbouw van Geeraerts, Grondelaers en Speelman (1999) overneemt, beschikken we voor het substandaardtaalgebruik nu ook over een diachrone dimensie. Dit laat toe om na te gaan op welk stratificationeel niveau er de meeste verschuivingen zijn geweest de laatste twintig jaar. We herhalen dat een lagere uniformiteit een indicatie is voor meer ontwikkelingen. In Nederlandstalig België vinden we dat het bovengewestelijke niveau in (24) het meest in beweging is geweest, wat we kunnen linken aan de eerder geobserveerde divergerende tendens op dat niveau. Het Nederlands-Nederlandse materiaal in (25) toont een omgekeerde situatie, waarbij de verschuivingen vooral plaatsvinden op het regionale niveau. Voor de volledigheid signaleren we dat dit niet het geval is bij de ongewogen uniformiteit in (26), maar dat daar de waarden eerder vergelijkbaar zijn.

(24) U'(B90,B12) < U'(LeuKor90,LeuKor12)
81,57 87,71 (*14C*)
(25) U'(N90,N12) > U'(LeiMaa90,LeiMaa12)
88,00 81,93 (*14C*)
(26) U(N90,N12) < U(LeiMaa90,LeiMaa12)
81,34 83,59 (*14C*)

Verder was het een van de hypotheses in Geeraerts, Grondelaers en Speelman (1999) dat het standaardtalige Belgisch Nederlands dichter zou staan bij het Nederlands Nederlands dan bij het regionale taalgebruik in België. Deze verwachting werd niet alleen ingelost in Nederlandstalig België in (27), maar bleek ook te gelden voor het Nederlands Nederlands in (30). De *inter*nationale convergentie tussen de twee taalgebieden was met an-

dere woorden sterker dan de *intra*nationale uniformiteit. Voor de recentste data observeren we in (28) een toegenomen uniformiteit tussen het Belgische tijdschriften- en etalagemateriaal, maar het supraregionale niveau staat wel nog steeds dichter bij Nederlands Nederlands dan bij het regionale taalgebruik. De ongewogen uniformiteit in (29) laat echter een ander beeld zien onder invloed van laagfrequente concepten als COLBERT (MAN) en JACK (MAN/VROUW) die binnen het Belgische materiaal zeer uniform zijn, maar duidelijk andere lexicalisatievoorkeuren vertonen in Nederland. Tot slot merken we ook voorzichtig op dat in (31) het Nederlandse intranationale taalgebruik een licht sterkere uniformiteit vertoont.

(27) U'(B90,N90) > U'(B90,LeuKor90)
83,49 46,86 (*14C*)

(28) U'(B12,N12) > U'(B12,LeuKor12)
80,50 63,95 (*14C*)

(29) U(B12,N12) < U(B12,LeuKor12)
70,01 73,72 (*14C*)

(30) U'(N90,B90) > U'(N90,LeiMaa90)
83,49 67,72 (*14C*)

(31) U'(N12,B12) < U'(N12,LeiMaa12)
80,50 81,65 (*14C*)

Als laatste benaderen we de stratificatiehypothese vanuit taalintern perspectief. Voor de data van 2012 vinden we dat het Franse aandeel het grootst is in het regionale materiaal, wat vooral blijkt uit het ongewogen materiaal in (35) voor het Belgisch Nederlands en uit (37) voor het Nederlands Nederlands. Geeraerts, Grondelaers en Speelman (1999) brachten die observatie in verband met de commerciële functie van de winkeletalages, waar de appreciërende invloed van het Frans meer tot uiting komt. Daarnaast verwachtten Geeraerts, Grondelaers en Speelman (1999, p. 118; p. 126) dat het Franse aandeel het grootst zou zijn in de Leuvense en Kortrijkse etalages door de historische invloed van het Frans. De data wezen echter meer in de omgekeerde richting, wat als een mogelijke aanwijzing beschouwd werd dat de typisch Belgische puristische tendens ook tot het niveau van het minder formele taalgebruik doorgedrongen zou zijn. In navolging dan van de bevindingen in Geeraerts, Grondelaers en Speelman (1999, p. 126) kunnen we opnieuw verwachten dat de invloed van het Frans in (41) sterker is in Nederland dan in Vlaanderen. Daarnaast stellen we ten opzichte van 1990 in (38-39) een daling vast.

(32) A'_{FR}(LeuKor90) ≅ A'_{FR}(B90)
19,29 17,62 (*14C*)
(33) A'_{FR}(LeuKor12) ≅ A'_{FR}(B12)
12,21 10,60 (*14C*)
(34) A_{FR}(LeuKor90) > A_{FR}(B90)
33,04 17,81 (*14C*)
(35) A_{FR}(LeuKor12) > A_{FR}(B12)
25,49 12,10 (*14C*)
(36) A'_{FR}(LeiMaa90) ≅ A'_{FR}(N90)
33,76 28,94 (*14C*)
(37) A'_{FR}(LeiMaa12) > A'_{FR}(N12)
22,87 17,03 (*14C*)
(38) A'_{FR}(LeuKor90) > A'_{FR}(LeuKor12)
19,29 12,21 (*14C*)
(39) A'_{FR}(LeiMaa90) > A'_{FR}(LeiMaa12)
33,76 22,87 (*14C*)
(40) A'_{FR}(LeuKor90) < A'_{FR}(LeiMaa90)
19,29 33,76 (*14C*)
(41) A'_{FR}(LeuKor12) < A'_{FR}(LeiMaa12)
12,21 22,87 (*14C*)

Voor het Engels bevinden de grootste verschuivingen zich in het Belgisch Nederlands. Het aandeel lexicaliseringen van (deels) Engelse oorsprong kent een toename in de periode 1990-2012 in (47) en komt zelfs tot op dezelfde hoogte als het aandeel in het Nederlands Nederlands in (50). Het ongewogen Belgisch-Nederlandse materiaal in (44) vertoont een groter aandeel op standaardtaalniveau dan op het niveau van het meer regionaal gekleurde taalgebruik door een toename van Engelse termen als *skirt* en *blazer* bij respectievelijk de minder frequente concepten ROK (VROUW), en COLBERT (MAN) en COLBERT (VROUW).

(42) A'_{EN}(LeuKor90) > A'_{EN}(B90)
21,61 15,03 (*14C*)
(43) A'_{EN}(LeuKor12) ≅ A'_{EN}(B12)
27,48 23,92 (*14C*)
(44) A_{EN}(LeuKor12) < A_{EN}(B12)
34,83 41,72 (*14C*)
(45) A'_{EN}(LeiMaa90) > A'_{EN}(N90)
29,61 14,86 (*14C*)

(46) A'_{EN}(LeiMaa12) > A'_{EN}(N12)
29,71 17,40 (*14C*)

(47) A'_{EN}(LeuKor90) < A'_{EN}(LeuKor12)
21,61 27,48 (*14C*)

(48) A'_{EN}(LeiMaa90) ≅ A'_{EN}(LeiMaa12)
29,61 29,71 (*14C*)

(49) A'_{EN}(LeuKor90) < A'_{EN}(LeiMaa90)
21,61 29,61 (*14C*)

(50) A'_{EN}(LeuKor12) ≅ A'_{EN}(LeiMaa12)
27,48 29,71 (*14C*)

Ten slotte kijken we naar de invloed van taalnormering op het hedendaagse Belgische materiaal. We veronderstellen hier uitsluitend een synchroon perspectief omdat de laatste decennia het belang van taalzuiveringswerken is afgenomen (zie ook Van Hoof en Jaspers 2012, p. 112) en taaladviezen minder prescriptief zijn geworden (Hendrickx 2013, p. 26). Bovendien zijn de taalbeheersingsbronnen uit Geeraerts, Grondelaers en Speelman (1999) niet langer actueel. Vertrekkende van de afgekeurde termen in Theissen en Debrabrandere (2004) en Koops et al. (2005), bevestigt de aandeelmaat in (51) dat vooral het regionale, meer informele materiaal afgekeurde kledingbenamingen bevat. Een blik op de data toont dat dit voornamelijk komt door het concept JEANS (MAN/VROUW), waarvoor *spijkerbroek* volgens de normatieve werken de voorkeur heeft op jeans.

(51) A'_{TZ}(LeuKor12) > A'_{TZ}(B12)
23,83 10,20 (*14C*)

In het kort kunnen we stellen dat de oorspronkelijke stratificatiehypothese in grote lijnen bevestigd is voor het hedendaagse materiaal. De afstand tussen beide onderzochte registers blijft het grootst in het Belgisch Nederlands, maar dat maakt die variëteit intern niet heterogener dan het Nederlands Nederlands. In Nederlandstalig België vonden in de periode 1990-2012 de meeste verschuivingen plaats op het standaardtaalniveau, terwijl in Nederland dat het geval was op regionaal niveau. Het Nederlandse Standaardnederlands staat nu zelfs dichter bij het lagere register dan dat het bij het Standaard Belgisch Nederland staat. Ook de resultaten voor de aandeelmaten stemmen overeen met de traditionele verwachtingen. Het aandeel Franse termen is afgenomen en is groter in het regionale register, vooral dan in de Nederlands-Nederlandse data. Voor het Engels vinden we een inhaalbeweging bij het Belgisch Nederlands, waardoor het tot op de-

zelfde hoogte als het Nederlands Nederlands komt wat dit aandeel betreft. Ten slotte merken we de meeste afgekeurde termen op in het etalagemateriaal in België, maar kunnen we tegelijk een vraagteken plaatsen bij de huidige relevantie van taalzuiveringswerken.

4 Discussie en conclusies

De vervolgstudie die in dit artikel voorgesteld wordt, bouwt voort op de basis van Geeraerts, Grondelaers en Speelman (1999).Voor veertien kledingconcepten hebben we onderzocht of er voor de periode 1950-2012 sprake is van convergentie of divergentie tussen het Belgisch Nederlands en Nederlands Nederlands en of die eerste variëteit bovendien een grotere afstand vertoont tussen het regionale en supraregionale taalregister als gevolg van de exogene normgerichtheid. Hiertoe gaan we uit van een corpus van reëel taalgebruik dat op drie niveaus gelaagd is. De geografische dimensie weerspiegelt het pluricentrische karakter van het Nederlands, waarbij zowel de Nederlands-Nederlandse als de Belgische-Nederlandse variëteit vertegenwoordigd zijn. Het oorspronkelijke corpus bestond uit observaties van standaardtaalgebruik zoals aangetroffen in tijdschriften in 1950, 1970 en 1990 en daar is voor dit artikel een nieuwe deelverzameling aan toegevoegd met gelijkaardig materiaal uit 2012. Daarbovenop zijn er ook observaties opgenomen van kledingbenamingen uit winkeletalages in vier universiteitssteden in Vlaanderen en Nederland als weergave van het meer informele, regionaal gekleurde taalgebruik.

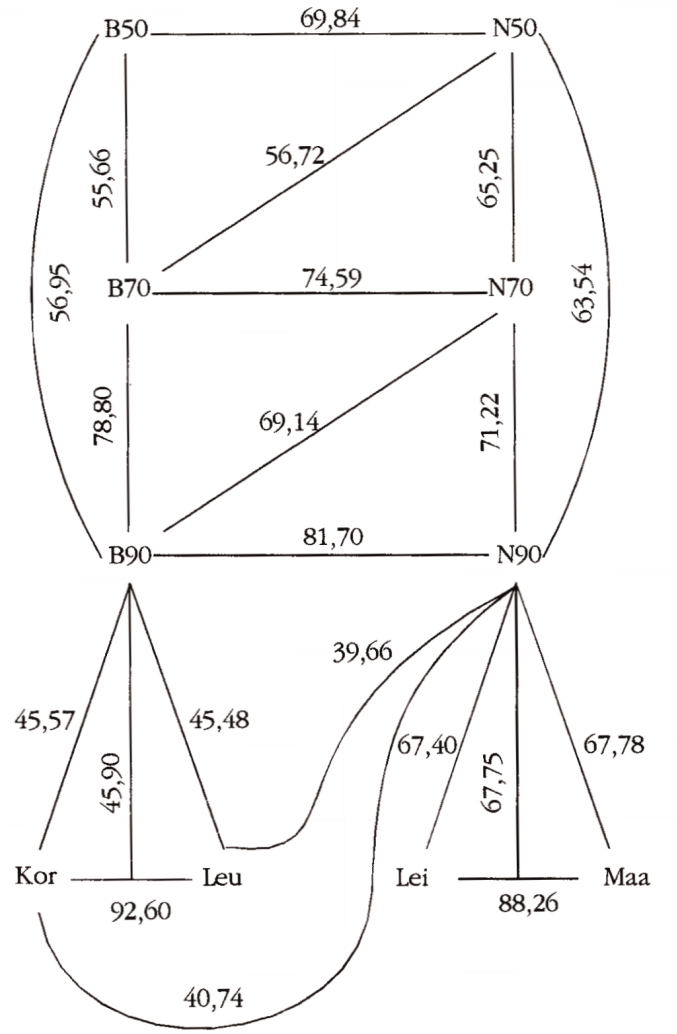
De methodologische vernieuwing uit Geeraerts, Grondelaers en Speelman (1999), met name de introductie van de profielgebaseerde uniformiteitsmaat, reikt een eenduidige berekeningswijze aan om de overlapping tussen de verschillende taalvariëteiten te meten. Op die manier hebben we de meervoudige bevindingen uit de oorspronkelijke studie gerepliceerd, wat verschillende nieuwe inzichten heeft opgeleverd. Figuur 2 geeft een overzicht van de resultaten voor de gewogen uniformiteitswaarden op basis van elf kledingconcepten[15]. In wat volgt vatten we eerst de voornaamste conclusies samen in verband met de verhouding *tussen* de nationale variëteiten en vervolgens bespreken we de evoluties *binnen* die nationale variëteiten in het licht van het destandaardiseringsdebat.

4.1 De verhouding *tussen* de nationale variëteiten: het Belgisch en Nederlands Nederlands groeien niet langer naar elkaar toe

Wanneer we over de periode 1950-2012 kijken naar de verhouding tussen het Belgisch Nederlands en het Nederlands Nederlands, kunnen we niet langer spreken van convergentie tussen het Belgisch Nederlands en het Nederlands Nederlands, maar observeren we een divergerende trend voor de periode 1990-2012. De langetermijntrend naar diachrone convergentie, zoals vastgesteld in Geeraerts, Grondelaers en Speelman (1999), lijkt dus omgebogen te zijn. Deze vaststelling bevestigt nu ook aan de hand van kwantitatieve metingen (zij het op een bescheiden steekproef) wat reeds elders in de literatuur geopperd werd. Waar de Belgisch-Nederlandse geschreven pers, de modebladen in deze studie, zich in de tweede helft van de twintigste eeuw qua woordgebruik steeds dichter naar de Nederlandse "norm" toebewoog, zien we nu, in het begin van de eenentwintigste eeuw, voor het eerst een lichte verwijdering optreden. Op de oorzaken van deze convergentiestop gaan we hier niet verder in, maar we verwijzen naar de in de literatuur (zie, o.a., Grondelaers en Van Hout 2011a, p. 228) reeds aangehaalde factoren zoals het grotere culturele en talige zelfbewustzijn in Nederlandstalig België, een verminderde aandacht voor taalzuivering en belgicismenvermijding, en het verminderde contact van de Vlaming met de Nederlandse media: allemaal factoren die de gerichtheid van het Belgisch Nederlands op de exogene Nederlands-Nederlandse norm deden afnemen. Interessant is nog dat de toenemende invloed van het Engels, in zekere zin een nieuwe exogene "norm", deze keer inwerkend op beide nationale variëteiten, niet sterk genoeg is om een tegengewicht te bieden voor divergentie, zelfs niet in een trendgevoelig lexicaal veld als mode.

4.2 Evoluties *binnen* de nationale variëteiten: standaardisering én (in)formalisering?

De tweede hoofdhypothese uit de oorspronkelijke studie betrof de grotere afstand tussen het bovengewestelijke formele en het regionaal gekleurde informele niveau in het Belgisch Nederlands in vergelijking met het Nederlands Nederlands. De uniformiteitsmaten uit 2012 bevestigen dat dit nog steeds het geval is. We vinden zelfs dat de intranationale uniformiteit, d.i. de overlapping tussen het supraregionale en het regionale register, in Nederland groter is dan de supranationale overeenkomst tussen het Nederlands Nederlands en het Belgisch Nederlands. De data duiden bovendien op een veranderende taalinterne situatie in beide nationale variëteiten. In het Nederlands Nederlands observeren we verschuivingen in het lagere register, terwijl in Nederlandstalig België net het omgekeerde het geval is.



Figuur 2 Gewogen uniformiteitsverhoudingen binnen het kledingmateriaal (elf concepten)

Tot slot vermelden we dat we in het Nederlands Nederlands een grotere terughoudendheid observeren ten opzichte van het Engels dan in het Belgisch Nederlands, alsook dat er een afname van de Franse invloed is op beide taalniveaus in de twee variëteiten.

Hoewel het de hoofdopzet van dit artikel was om de bevindingen van deze replicatiestudie (in opvolging van Geeraerts, Grondelaers en Speelman 1999) te beschrijven en de vergelijking te maken met de oorspronkelijke studie, bespreken we ook kort hoe de vaststellingen in deze studie in

verband gebracht kunnen worden met het lopende onderzoek naar destandaardisering en demotisering van het Nederlands en het tussentaaldebat (zie o.a. Auer en Spiekermann 2011 voor een algemene beschouwing; Grondelaers en Van Hout 2011a; Absillis, Jaspers en Van Hoof 2012; Cajot 2012; Van Hoof en Jaspers 2012). Hierbij is het van belang om termen als (de) standaardisering te vertalen naar de verschillende processen die we bestuderen vanuit een profielgebaseerd perspectief. We maken een conceptueel onderscheid tussen drie dimensies en volgen hierin Geeraerts en Speelman (2014). De eerste dimensie gaat kijken of de afstand tussen twee taalregisters toe- of afneemt. We spreken dan van **convergentie of divergentie**[16] en meten dit aan de hand van de uniformiteitsmaat. Als de afstand tussen twee taalregisters verandert, kan dat in twee richtingen. Wanneer het hogere taalregister zich naar het lagere beweegt, spreken we van **informalisering**. Het omgekeerde gebeurt bij **formalisering**. Tot slot, de variabiliteit binnen een bepaald register (voornamelijk, het hogere taalniveau) kan ook af- of toenemen. We spreken dan van **(de)homogenisering**. Een register wordt homogener als de variabiliteit afneemt en dus de interne uniformiteit groter wordt. Standaardisering, zoals opgevat door Auer en Spiekermann (2011), kan dan vanuit een profielgebaseerd perspectief in verband gebracht worden met convergentie, formalisering en homogenisering. Echter, deze drie dimensies zijn onafhankelijk van elkaar en kunnen dus in verschillende combinaties voorkomen.

Met de data waarover we nu beschikken, vinden we dat er zowel binnen het Belgisch Nederlands als binnen het Nederlands Nederlands convergentie (de afstand tussen de registers vernauwt, want de uniformiteit stijgt) en homogenisering (de interne uniformiteit stijgt in het hoogste register) optreedt. Deze tendensen voor twee van de drie dimensies wijzen dus op een klassiek standaardiseringproces. Bij de derde, (in)formaliserings-dimensie, zien we in de twee nationale variëteiten echter tegenovergestelde processen. In het Belgisch Nederlands beweegt het hogere register zich naar het lagere[17]: de traditionele exonormatieve oriëntatie maakt plaats voor een endogene, informelere norm. Deze verschuiving kan als demotisering in de zin van Kristiansen (2009) geïnterpreteerd worden. Meteen wordt in de Belgisch-Nederlandse situatie dan ook duidelijk dat aspecten van standaardisering en demotisering elkaar niet noodzakelijk uitsluiten. Bij het Nederlands in Nederland observeren we echter een verschuiving van het lagere taalniveau richting het hogere. In de Nederlandse situatie wijzen de drie dimensies voor de periode 1990-2012 dus op een klassiek standaardiseringsproces, en niet op aan de gang zijnde destandaardisering. Een mogelijke verklaring voor dit verrassende resultaat

is dat in Nederland woordenschat historisch minder een *markeerder* is geweest van een lagere status dan in Nederlandstalig België, waar het net in het middelpunt stond in de taalverzorgingsliteratuur. Dit kan ook de tegenstelling met de bevindingen van Grondelaers en Van Hout (2011a) verklaren, doordat uitspraak en grammatica in Nederland misschien meer expliciete taalattitudes teweegbrengen. Hoe dan ook zijn verder onderzoek en een kritische beschouwing van termen als (de)standaardisering en (in)formalisering vanuit een profielgebaseerd perspectief (zoals in Geeraerts en Speelman (2014)) nodig om conclusies te trekken over de standaardiseringsstatus van het Nederlands.

4.3 Verder onderzoek: grotere corpora, meer registers en andere (lexicale) velden

Met de bevindingen in deze studie is nog niet alles gezegd over de taalsituatie in de Lage Landen. Verscheidene uitbreidingen zijn nodig om de vaststellingen uit dit onderzoek, hoe robuust ook, door te trekken met enige mate van zekerheid naar conclusies over de status en de structuur van het Nederlands. In eerste instantie is de uitbreiding van het corpus een cruciale stap. Zelfs voor slechts veertien kledingconcepten is het een tijdrovend proces om manueel data te verzamelen in tijdschriften en winkeletalages. De automatisering van het tellen van observaties in een elektronisch corpus zoals in Ruette, Speelman en Geeraerts (2011) en Heylen en Ruette (2013), is hiervoor deels een oplossing en kan een significante schaalvergroting opleveren, ook al leidt deze automatisering tot het verlies van de referentiële informatie uit afbeeldingen die de oorspronkelijke verzamelingswijze oplevert. Toch is het een voor de hand liggende keuze om voor de verzameling van concepten uit andere lexicale velden of van grammaticale constructies die lijn te vervolgen. Een corpusgebaseerde studie naar variatie zoals in dit onderzoek voorgesteld is, kan immers bijdragen aan het standaardtaalonderzoek. Daarnaast is de stratificationele opbouw van dit artikel slechts een beperkte weerspiegeling van de talige realiteit van het Nederlands. De toevoeging van andere registers zou toelaten om de complexiteit van het brede spectrum tussen de standaardtaal en de dialecten, in zowel geschreven als gesproken vorm, beter te beschrijven. Tot dan biedt deze studie enkel een gedeeltelijke inkijk in de lectale structuur van het Nederlands.

Noten

1. Graag bedanken we masterstudente Tine De Cnodder (academiejaar 2012-2013) voor het verzamelen van de data uit zowel de tijdschriften en de winkeletalages. We danken ook de anonieme reviewers voor hun suggesties bij een eerdere versie van dit artikel. Het onderzoek waarover dit artikel rapporteert wordt gesteund door het OT-project 3H110243 van de KU Leuven.
2. In dit artikel geldt, alsook het geval is in Geeraerts, Grondelaers en Speelman (1999, pp. 6-7), dat we zullen spreken van *Belgisch Nederlands* en *Nederlands Nederlands* om te verwijzen naar de twee geografisch gescheiden variëteiten van het Nederlands. In deze context vermijden we dus de term *Vlaams,* die ook verwijst naar een verzameling dialecten, evenals de spelling met koppelteken *Belgisch-Nederlands*, welke vooronderstelt dat het een aparte taalvariëteit betreft. Wanneer deze termen bijvoeglijk gebruikt worden, is een koppelteken onvermijdelijk..
3. De *City-block distance* (CBD) uit Speelman, Grondelaers en Geeraerts (2003; 2006) meet niet de overlapping of uniformiteit (U) tussen twee taalvariëteiten, maar het verschil ertussen. Een simpele rekensom laat de overgang van de ene maat naar de andere toe, namelijk CBD = 1 - U, of anders genoteerd: U = 1 - CBD.
4. Appendix 1 bevat een overzicht van de veertien kledingconcepten en hun frequentie in aantal woorden in de verschillende deelverzamelingen. Appendix 3 somt de tijdschriftbronnen voor het materiaal van 2012 op.
5. De absolute frequenties van de kledingtermen in deze twee profielen zijn te vinden in Tabel 4 in Appendix 1.
6. Door te kwadrateren wordt meer gewicht gegeven aan de dominantie van een één bepaalde term, wat niet het geval is door louter het gemiddelde te nemen. Een kleine vergelijking illustreert dit goed. In bron A wordt een bepaald concept aangeduid met vier synoniemen, in de verhouding 50%, 25%, 15% en 10%. Datzelfde concept kent in bron B een andere distributie, namelijk 85%, 5%, 5% en 5%. Door de interne uniformiteit te berekenen op basis van de som van de kwadraten, betekent dit dat in bron A, met 34,5%, de uniformiteit veel lager ligt dan in bron B, met 73%. Het gemiddelde in beide gevallen is evenwel 25%, en dus nietszeggend met betrekking tot de homogeniteit van de taalsituatie in bron A en die in bron B.
7. Deze wegingsfactor wordt momenteel manueel bepaald door de onderzoeker, maar een andere mogelijkheid is het experimenteel meten van de herkenbaarheid van gallicismen en anglicismen. Voor de manuele toekenning onderscheiden we op verschillende gronden de waarden 0 (geen aandeel), 1 (volledig aandeel) en enkele tussenliggende gevallen. Kledingtermen waarvan zowel de schrijfwijze als de uitspraak nog dezelfde is als in de brontaal krijgen waarde 1, zoals *caleçon* en *leggings* voor LEGGING_V voor respectievelijk de Franse en Engelse aandeelmaat. Vernederlandste termen qua uitspraak en/of schrijfwijze die niet onmiddellijk als termen van vreemde herkomst herkenbaar zijn, ontvangen de score 0,5. Dit is bijvoorbeeld het geval bij *vest* voor VEST_MV. Aan samenstellingen met een Nederlands grondwoord en een grondwoord van Franse of Engels herkomst wordt waarde 0,5 toegekend, zoals bij *jeansbroek* voor JEANS_MV. Aan *windbloes* voor JACK_MV kennen we slechts 0,25 toe, op grond van het vernederlandste grondwoord *bloes*. Voor het aandeel van de taalzuiveringswerken geldt in eerste instantie een binaire toekenning, 0 (niet opgenomen in het taalzuiveringswerk) en 1 (afgekeurd). Vervolgens wordt het gemiddelde genomen van de waarden gebaseerd op de twee geraadpleegde werken.

8. De uniformiteitswaarden in Figuur 1 zijn gebaseerd op de berekening voor elf kledingconcepten en kunnen dus voor het hedendaagse materiaal verschillen van de waarden die we opgeven bij de vergelijkingen (14-19; 24-31), omdat we daar uitgaan van alle veertien kledingconcepten.
9. Net zoals voor de oorspronkelijke materiaalverzameling, zijn er vooral kleinere winkels (en niet grotere winkelketens) onderzocht (Geeraerts, Grondelaers en Speelman, 1999, p. 34). Het is echter aannemelijk dat een groot aantal van de boetieks in het straatbeeld plaats hebben gemaakt voor net die grotere winkelketens, wat verklaart waarom het aantal observaties voor het etalagemateriaal in 2012 ligt lager dan in 1990.
10. Niet alle subhypotheses uit Geeraerts, Grondelaers en Speelman (1999) worden hier herhaald en getest met de nieuwe data uit 2012, maar de vergelijkingen zijn repliceerbaar met de tabellen uit Appendix 2.
11. Voor het uitvoeren van de *log likelihood ratio test* gebruiken we de implementatie van Pete Hurd in R, genaamd de 'g-test', waarvoor de code te vinden is op: http://www.psych.ualberta.ca/~phurd/cruft/g.test.r.
12. De significantiegrens van 5% die we hier vermelden verwijst naar de *log likelihood ratio test* en is dus niet hetzelfde als de foutenmarge gehanteerd in Geeraerts, Grondelaers en Speelman (1999) voor het vergelijken van twee uniformiteitswaarden. Bij het vergelijken van berekeningen die steunen op de interne uniformiteitsmaat en de aandeelmaat gebruiken we wel de statistische foutenmarge van 5%.
13. De uniformiteitsindices voor het materiaal van 1950, 1970 en 1990 verschillen van die in Geeraerts, Grondelaers en Speelman (1999) omdat het concept MANTELPAKJE (VROUW) omwille van dataschaarste niet mee opgenomen is 2012. De originele uniformiteitswaarden werden daarom herrekend voor 14 concepten.
14. Ook wanneer we de tijdsblokken opsplitsen, zien we een geleidelijke toename van de uniformiteit wat de plausibiliteit verhoogt van de interpretatie dat de ontwikkelingen vooral aan Belgische kant voorvallen. In dit opzicht, net als in Geeraerts, Grondelaers en Speelman (1999, p. 74), zijn er ook geen retardatie-effecten te observeren. We verwijzen hier slechts beknopt naar deze berekeningen omdat ze qua interpretatie niet verschillen van de oorspronkelijke studie en geen nieuwe inzichten opleveren. Appendix 2 geeft wel de relevante berekeningen weer om alle analyses uit de oorspronkelijke studie na te gaan.
15. Doordat Figuur 2 een globaal overzicht geeft waarin zowel de basisbevindingen voor de diachrone convergentiehypothese als voor de stratificatiehypothese af te lezen zijn, geven we de uniformiteitswaarden weer voor elf kledingconcepten. Hierdoor verschillen de waarden voor onder andere het regionale materiaal met die van de vergelijkingen in sectie 3 van dit artikel, omdat we daar rekenen met veertien kledingconcepten.
16. De termen convergentie en divergentie worden door de focus van Geeraerts, Grondelaers en Speelman (1999) voornamelijk geassocieerd met respectievelijk een vernauwing en een verwijding van de afstand tussen *geografische* variëteiten (cf. de diachrone convergentie tussen het Belgisch en Nederlands Nederlands), maar dit hoeft niet het geval te zijn. Convergentie betekent meer algemeen de vernauwing van de afstand tussen twee taalvariëteiten, zijnde geografisch, stratificationeel… In een discussie over standaardisering, wordt de term convergentie gebruikt om de afstand tussen twee verschillende registers te bestuderen. Merk hierbij op dat dit verschilt van de terminologie in Geeraerts en Speelman (2014), waar de dimensie 'convergentie-divergentie' aangeduid wordt met '(de)standaardisering'.
17. In een profielgebaseerde aanpak operationaliseren we het concept 'informalisering' aan de hand van de volgende vergelijking: $U(L_i,H_{i+1}) > U(H_i,L_{i+1})$. Hierbij staat 'L' voor het

lagere register, en 'H' voor hogere taalniveau, voor de twee tijdstippen 'i' en 'i+1'. Voor de data uit deze studie wordt dat:

(52)	U'(LeuKor90,B12) 56,22	>	U'(B90,LeuKor12) 52,29	*(14C)*
(53)	U'(LeiMaa90,N12) 59,23	<	U'(N90,LeiMaa12) 85,54	*(14C)*

Bibliografie

Absillis, Kevin, Jürgen Jaspers en Sarah Van Hoof (red.), *De manke usurpator. Over Verkavelingsvlaams*, Gent: Academia Press, 2012.

Auer, Peter en Helmut Spiekermann, 'Demotisation of the standard variety or de-standardisation? The changing status of German in late modernity (with special reference to southwestern Germany),' In: Tore Kristiansen en Nikolas Coupland (red.), *Standard Languages and Language Standards in a Changing Europe*, Oslo: Novus Press, 2011, pp. 161-176.

Cajot, José, 'Waarom het Verkavelingsvlaams onvermijdelijk was. De ontwikkeling van een informele omgangstaal in Vlaanderen', In: Kevin Absillis, Jürgen Jaspers en Sarah Van Hoof (red.), *De manke usurpator. Over Verkavelingsvlaams*, Gent: Academia Press, 2012, pp. 39-66.

De Cnodder, Tine, 'Convergentie en divergentie in de Nederlandse woordenschat anno 2012. Een onderzoek naar kledingtermen', Meesterproef onder begeleiding van Dirk Geeraerts, Leuven: KU Leuven, 2012-2013.

Geeraerts, Dirk en Dirk Speelman, 'A lectometric definition of demotisation and destandardisation', Presentatie op *Taal & Tongval 2014* (28 november 2014), 2014.

Geeraerts, Dirk en Stefan Grondelaers, 'Purism and fashion. French influence on Belgian and Netherlandic Dutch', *Belgian Journal of Linguistics,* 13 (1999), pp. 53-68. doi:10.1075/bjl.13.04gee.

Geeraerts, Dirk, Stefan Grondelaers en Dirk Speelman, *Convergentie en divergentie in de Nederlandse woordenschat: een onderzoek naar kleding- en voetbaltermen*. Amsterdam: P.J. Meertens-Instituut, 1999.

Goossens, Jan, 'De toekomst van het Nederlands in Vlaanderen', *Ons Erfdeel* 43 (2000), pp. 3-13.

Grondelaers, Stefan, Dirk Geeraerts, Dirk Speelman en José Tummers, 'Lexical standardisation in internet conversations. Comparing Belgium and The Netherlands', In: Josep M. Fontana, Louise McNally, M. Teresa Turell en Enric Vallduví (red.), *Proceedings of the First International Conference on Language Variation in Europe*, Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada, Unitat de Investigació de Variació Lingüística, 2001a, pp. 90-100.

Grondelaers, Stefan, Hilde Van Aken, Dirk Speelman en Dirk Geeraerts, 'Inhoudswoorden en preposities als standaardiseringsindicatoren. De diachrone en synchrone status van het Belgische Nederlands', *Nederlandse Taalkunde,* 6-3 (2001b), pp. 179-202.

Grondelaers, Stefan en Roeland van Hout, 'The Standard Language Situation in the Low Countries: Top-Down and Bottom-Up Variations on a Diaglossic Theme', *Journal of Germanic Linguistics* 23-3 (2011a), pp. 199-243.

Grondelaers, Stefan en Roeland van Hout, 'The standard language situation in The Netherlands',

In: Tore Kristiansen en Nikolas Coupland (red.), *Standard Languages and Language Standards in a Changing Europe*, Oslo: Novus press, 2011b, pp. 113-118.

Hendrickx, Els, 'Het effect van lexicale taaladvisering op het Belgisch-Nederlandse taalgebruik. Een diachroon corpusonderzoek naar factoren van invloed', Dissertatie, Leuven: KU Leuven, 2013.

Heylen, Kris en Tom Ruette, 'Degrees of semantic control in measuring aggregated lexical distances'. In: Lars Borin en Anju Saxena (red.), *Approaches to Measuring Linguistic Differences*, (Trends in Linguistics. Studies and Monographs 265), Berlin: Mouton de Gruyter, 2013, pp. 353-374. doi:10.1515/9783110305258.353.

Kristiansen, Tore, 'The nature and role of language standardisation and standard languages in late modernity', Projectvoorstel, 2009. http://dgcss.hum.ku.dk/exploratoryworkshops/proposal/ (18 mei 2015).

Koops, Bert-Jaap, Pim Slop, Paul Uljé, Kees Vermeij en Dick van Zijderveld, '2400 x liever Nederlands. Woordenlijst onnodig Engels', Stichting Nederlands, 2005. http://snn.vvb.org/woordenlijst/a.htm (22 augustus 2013).

Ruette, Tom, Dirk Speelman en Dirk Geeraerts, 'Measuring the lexical distance between registers in national varieties of Dutch', In: Augusto Soares da Silva, Amadeu Torres en Miguel Gonçalves (red.), *Línguas Pluricêntricas. Variação Linguística e Dimensões Sociocognitivas*, Braga: Publicações da Faculdade de Filosofia, Universidade Católica Portuguesa, 2011, pp. 541-554.

Speelman, Dirk, Stefan Grondelaers en Dirk Geeraerts, 'A profile-based calculation of region and register variation: the synchronic and diachronic status of the two main national varieties of Dutch'. In: Andrew Wilson, Dawn Archer en Paul Rayson (red.), *Corpus Linguistics around the World*, Amsterdam: Rodopi, 2006, pp. 181-194.

Speelman, Dirk, Stefan Grondelaers en Dirk Geeraerts, 'Profile-based linguistic uniformity as a generic method for comparing language varieties', *Computers and the Humanities,* 37-3 (2003), pp. 317-337. doi:10.1023/A:1025019216574.

Theissen, Siegfried en Peter Debrabandere. *Woordenboek voor correct taalgebruik*. Deurne: Wolters Plantyn, 2004.

Van Hoof, Sarah en Jürgen Jaspers. 'Hyperstandaardisering', *Tijdschrift voor Nederlandse Taal- en Letterkunde,* 128-2 (2012), pp. 97-125.

Zenner, Eline, Dirk Speelman en Dirk Geeraerts, 'Cognitive Sociolinguistics meets loanword research: Measuring variation in the success of anglicisms in Dutch', *Cognitive Linguistics* 23-4 (2012), pp. 749-792. doi:10.1515/cog-2012-0023.

Over de auteurs

Jocelyne Daems is doctoraatsbursaal aan de KU Leuven in de onderzoeksgroep *Quantitative Lexicology and Variational Linguistics* (*QLVL*) en is de corresponderende auteur: Jocelyne.
E-mail: Daems@kuleuven.be

Kris Heylen is postdoctoraal onderzoeker aan de KU Leuven in de onderzoeksgroep *Quantitative Lexicology and Variational Linguistics* (*QLVL*).

Dirk Geeraerts is gewoon hoogleraar Algemene Taalkunde aan de KU Leuven en is aldaar verbonden aan de onderzoeksgroep *Quantitative Lexicology and Variational Linguistics* (*QLVL*).

Appendices

Appendix 1 Overzicht observaties in alle subcorpora in aantal woorden

Tabel 4 Kwantitatief overzicht voor de veertien kledingconcepten in het standaardtaalmateriaal in aantal woorden

concept	variant	B12	N12	B90	N90	B70	N70	B50	N50
overhemd_M	*hemd*	27	29	11	9	33	21	40	5
overhemd_M	*overhemd*	0	19	24	25	38	62	23	23
overhemd_M	*shirt*	2	107	0	20	1	11	2	1
bloes_V	*bloes(je)*	188	146	46	51	51	174	149	56
bloes_V	*bloeze/blouse/blouze*	112	324	176	473	172	67	291	322
T-shirt_MV	*T-shirt*	139	136	57	27	1	4	0	0
T-shirt_MV	*shirt*	3	80	1	18	0	0	0	0
trui_MV	*pull(etje)*	2	0	0	5	105	3	0	0
trui_MV	*pullover*	2	7	2	9	29	24	49	44
trui_MV	*trui(tje)*	202	251	377	341	82	57	13	9
trui_MV	*sweater*	28	32	0	0	0	0	0	0
vest_MV	*cardigan*	130	3	90	2	13	0	0	0
vest_MV	*gil(l)et*	7	1	41	24	1	0	3	0
vest_MV	*jasje*	100	140	142	176	11	6	14	8
vest_MV	*vest(je)*	72	328	166	246	62	66	59	22
vest_MV	*golf*	0	0	0	0	0	0	0	0
broek_MV	*broek*	304	602	243	388	390	232	59	291
broek_MV	*pantalon*	35	74	12	39	154	162	11	124
broek_MV	*pant*	0	0	0	0	0	0	0	0
broek_MV	*pants*	4	1	0	0	0	0	0	0
jeans_MV	*jeans*	121	185	64	81	8	2	0	0
jeans_MV	*spijkerbroek*	0	67	2	34	1	3	0	0
jeans_MV	*jeansbroek*	4	2	0	0	0	0	0	0
legging_V	*caleçon*	0	0	40	0	0	0	0	0
legging_V	*legging*	28	62	26	113	0	0	0	0
legging_V	*leggings*	0	0	39	25	0	0	0	0
legging_V	*leggingbroek*	1	0	0	0	0	0	0	0
rok_V	*rok(je)*	441	423	328	392	429	405	373	642
rok_V	*skirt*	8	1	0	0	0	0	0	0
jurk_V	*japon(netje)*	0	0	0	2	40	351	131	792
jurk_V	*jurk(je)*	621	891	167	184	1493	1015	104	305
jurk_V	*kleed(je)*	0	0	0	0	43	0	376	1
jurk_V	*dress*	3	16	0	0	0	0	0	0

concept	variant	B12	N12	B90	N90	B70	N70	B50	N50
colbert_M	*blazer*	6	12	4	9	1	7	0	0
colbert_M	*k/colbert(je)*	0	6	3	109	1	21	30	275
colbert_M	*colbertjasje*	0	0	1	0	0	0	0	3
colbert_M	*jas*	0	0	1	0	12	2	4	42
colbert_M	*jasje*	0	49	19	10	29	18	2	7
colbert_M	*vest(je)*	1	0	2	0	5	1	10	0
colbert_M	*veston*	0	0	0	0	0	0	1	0
colbert_V	*blazer*	185	81	62	129	5	19	0	0
colbert_V	*colbert(je)*	0	91	7	43	0	0	0	0
colbert_V	*colbertjas(je)*	0	4	0	0	0	0	1	0
colbert_V	*jas*	0	0	2	1	1	0	0	0
colbert_V	*jasje*	19	98	117	77	49	11	20	23
colbert_V	*mantel(tje)*	0	0	0	1	0	0	10	10
colbert_V	*tailleurmantel(tje)*	0	0	0	0	0	0	0	8
colbert_V	*vest(je)*	4	4	14	0	5	0	39	0
colbert_V	*kostuumvest*	2	0	0	0	0	0	0	0
jack_MV	*bloes*	0	0	0	0	0	0	4	0
jack_MV	*blouson(netje)*	13	5	13	31	31	0	0	0
jack_MV	*jack(je)/jek*	11	70	15	141	18	36	1	0
jack_MV	*jacket*	12	0	0	1	0	0	0	3
jack_MV	*jak(je)*	0	0	0	0	0	0	0	3
jack_MV	*windbloes(je)*	0	0	0	0	0	0	7	1
jack_MV	*windjack*	1	1	1	0	3	1	8	26
jack_MV	*windjacket*	0	0	0	0	0	0	2	0
kostuum_M	*colbertcostuum*	0	0	0	0	0	0	8	19
kostuum_M	*k/costuum*	2	1	7	119	63	39	14	49
kostuum_M	*pak*	4	14	25	57	9	4	60	0
	TOTAAL	2844	4363	2347	3412	3389	2824	1918	3114

Tabel 5 Kwantitatief overzicht voor de veertien kledingconcepten in het substandaard-taalmateriaal in aantal woorden

concept	variant	Leu12	Kor12	Lei12	Ma-a12	Leu90	Kor90	Lei90	Ma-a90
overhemd_M	*hemd*	91	46	0	46	367	364	7	126
overhemd_M	*overhemd*	0	0	0	0	0	0	20	1
overhemd_M	*shirt*	0	0	37	14	0	0	112	79
bloes_V	*bloes(je)*	13	20	0	2	245	285	1	6
bloes_V	*bloeze/blouse/blouze*	2	1	10	31	83	12	114	155
T-shirt_MV	*T-shirt*	24	24	6	17	49	33	21	20
T-shirt_MV	*shirt*	4	1	11	25	0	0	3	0
trui_MV	*pull(etje)*	61	45	0	4	783	754	19	113
trui_MV	*pullover*	0	5	10	3	0	0	70	36
trui_MV	*trui(tje)*	7	19	17	57	3	25	96	174
trui_MV	*sweater*	3	7	1	0	0	0	0	0
vest_MV	*cardigan*	37	13	0	0	62	19	0	0
vest_MV	*gil(l)et*	13	30	0	0	135	132	0	0
vest_MV	*jasje*	0	1	12	15	3	0	0	0
vest_MV	*vest(je)*	16	4	44	58	28	31	48	84
vest_MV	*golf*	1	0	0	0	0	0	0	0
broek_MV	*broek*	86	74	30	80	296	356	27	34
broek_MV	*pantalon*	1	6	29	18	31	28	109	159
broek_MV	*pant*	0	0	3	0	0	0	0	0
broek_MV	*pants*	0	0	0	1	0	0	0	0
jeans_MV	*jeans*	31	25	30	23	112	152	74	71
jeans_MV	*spijkerbroek*	0	0	0	0	0	0	16	2
jeans_MV	*jeansbroek*	0	0	0	0	0	0	0	0
legging_V	*caleçon*	0	0	0	0	132	106	0	0
legging_V	*legging*	1	2	1	8	7	17	40	28
legging_V	*leggings*	0	0	0	0	0	0	0	0
legging_V	*leggingbroek*	0	0	0	0	0	0	0	0
rok_V	*rok(je)*	18	16	13	32	464	437	69	162
rok_V	*skirt*	1	0	0	1	0	0	0	0
jurk_V	*japon(netje)*	0	0	1	3	0	0	27	23
jurk_V	*jurk(je)*	7	9	10	28	5	7	7	22
jurk_V	*kleed(je)*	48	39	0	0	108	101	0	0
jurk_V	*dress*	0	0	0	1	0	0	0	0
colbert_M	*blazer*	9	3	3	5	15	57	5	6
colbert_M	*k/colbert(je)*	0	0	23	6	0	0	66	95
colbert_M	*colbertjasje*	0	0	0	0	0	0	1	0
colbert_M	*jas*	0	0	0	0	29	2	0	0

concept	variant	Leu12	Kor12	Lei12	Ma-a12	Leu90	Kor90	Lei90	Ma-a90
colbert_M	*jasje*	0	0	0	0	0	0	0	0
colbert_M	*vest(je)*	7	4	0	0	56	51	0	0
colbert_M	*veston*	0	0	0	0	0	0	0	0
colbert_V	*blazer*	4	5	5	33	25	70	37	78
colbert_V	*colbert(je)*	0	0	2	3	0	0	2	5
colbert_V	*colbertjas(je)*	0	0	0	0	0	0	0	0
colbert_V	*jas*	0	0	0	0	0	0	0	5
colbert_V	*jasje*	0	0	2	3	0	0	6	7
colbert_V	*mantel(tje)*	0	0	0	0	0	0	0	0
colbert_V	*tailleurmantel (tje)*	0	0	0	0	0	0	0	0
colbert_V	*vest(je)*	10	2	0	0	142	103	0	0
colbert_V	*kostuumvest*	0	0	0	0	0	0	0	0
jack_MV	*bloes*	0	0	0	0	0	0	0	0
jack_MV	*blouson(netje)*	2	3	0	0	19	13	0	0
jack_MV	*jack(je)/jek*	0	0	32	12	1	6	80	82
jack_MV	*jacket*	0	0	0	3	0	4	0	0
jack_MV	*jak(je)*	0	0	0	0	0	0	0	0
jack_MV	*windbloes(je)*	0	0	0	0	0	0	0	0
jack_MV	*windjack*	0	0	0	0	0	0	0	0
jack_MV	*windjacket*	0	0	0	0	0	0	0	0
kostuum_M	*colbertcostuum*	0	0	0	0	0	0	0	0
kostuum_M	*k/costuum*	7	9	23	5	38	45	64	79
kostuum_M	*pak*	0	0	0	0	2	0	0	0
	TOTAAL	504	413	355	537	3240	3210	1141	1652

Appendix 2 Uniformiteitswaarden, interne uniformiteitswaarden en aandeelmaten

Tabel 6 Uniformiteitswaarden voor elf en veertien kledingconcepten met en zonder de *log likelihood ratio* (LLR) significatietest

	11C		11C_LLR		14C		14C_LLR	
	U'	U	U'	U	U'	U	U'	U
B12_N12	80.79	64.61	81.50	67.40	79.83	67.57	80.50	70.01
B12_LeuKor12	58.72	58.17	59.99	67.21	62.64	66.14	63.95	73.72
B12_Leu12	56.07	56.38	57.97	65.90	59.95	64.38	61.82	72.34
B12_Kor12	58.96	58.93	63.07	72.29	63.07	67.12	67.02	78.23
N12_LeiMaa12	80.63	69.31	82.29	71.08	80.20	72.24	81.65	73.62
N12_Lei12	77.25	67.13	80.47	71.10	77.05	70.26	79.87	73.37
N12_Maa12	81.61	66.59	85.06	70.30	81.16	70.20	84.18	73.12
LeuKor12_LeiMaa12	41.67	41.55	41.70	41.61	45.99	50.39	46.02	50.44
N12_Leu12	41.78	32.81	44.00	41.76	46.13	43.64	48.08	50.67
N12_Kor12	44.40	34.46	45.07	34.81	47.94	44.21	48.53	44.48
Lei12_Maa12	73.73	78.09	79.24	86.85	76.95	82.41	82.06	89.66
Leu12_Kor12	84.47	85.97	88.67	93.46	85.77	88.24	90.05	94.87
B90_B12	81.61	69.42	81.78	70.50	81.39	70.34	81.57	71.22
B50_B12	52.52	47.96	53.35	49.94	-	-	-	-
N90_N12	86.33	76.61	87.09	77.90	87.07	79.89	88.00	81.34
N50_N12	57.43	45.02	57.75	45.39	-	-	-	-
B12_N90	75.18	63.03	76.74	66.45	74.79	64.84	76.20	67.53
LeuKor90_LeuKor12	87.31	87.41	90.19	93.93	85.11	82.94	87.71	88.07
LeiMaa90_LeiMaa12	80.04	82.10	82.18	85.05	80.02	81.27	81.93	83.59
B90_N90	84.33	75.73	86.15	78.99	81.83	72.22	83.49	74.78
B90_LeuKor90	43.39	41.70	43.83	42.03	46.44	50.09	46.86	50.47
B90_Leu90	43.24	40.26	43.79	40.69	46.00	48.66	46.63	49.34
B90_Kor90	42.84	41.12	43.47	44.00	46.24	49.96	46.96	52.57
N90_LeiMaa90	66.53	66.98	66.53	66.98	67.72	69.07	67.72	69.07
N90_Lei90	65.70	65.40	66.09	66.12	67.45	68.71	68.33	70.12
N90_Maa90	67.08	67.48	67.52	68.29	67.70	68.38	68.10	69.02
Leu90_Kor90	92.19	86.16	92.89	90.50	92.57	88.50	93.56	92.54
Lei90_Maa90	84.93	88.66	86.92	92.47	85.34	89.13	87.33	93.01
N90_Leu90	39.98	35.86	40.03	35.90	39.85	37.86	39.89	37.88
N90_Kor90	40.69	38.38	40.94	38.55	41.06	40.46	41.29	40.59
B50_N50	67.58	66.58	69.21	68.98	-	-	-	-
B70_N70	77.44	68.36	77.50	68.65	-	-	-	-
B50_B90	59.25	54.49	59.53	55.37	-	-	-	-
N50_N90	65.86	59.64	67.28	61.31	-	-	-	-
B70_N50	59.44	59.77	59.67	59.92	-	-	-	-

	11C		11C_LLR		14C		14C_LLR	
	U'	U	U'	U	U'	U	U'	U
B90_N70	69.80	64.33	69.80	64.33	-	-	-	-
B50_B70	54.54	54.80	54.99	56.38	-	-	-	-
N50_N70	65.63	59.93	66.22	62.81	-	-	-	-
B70_B90	81.47	74.93	82.21	78.12	-	-	-	-
N70_N90	71.50	71.71	72.35	73.34	-	-	-	-

Tabel 7 Interne uniformiteitswaarden voor elf en veertien kledingconcepten

	11C		14C	
	I'	I	I'	I
B12	77.56	69.68	79.35	74.96
B90	72.61	67.62	72.10	69.21
B70	77.68	63.22	-	-
B50	61.06	57.07	-	-
N12	75.34	71.04	73.71	71.06
N90	74.72	70.74	73.69	68.48
N70	65.86	66.15	-	-
N50	71.05	66.23	-	-
LeuKor12	74.11	76.15	76.28	80.04
LeuKor90	85.56	77.05	86.25	80.78
Leu12	79.48	78.85	80.56	81.63
Leu90	85.39	79.76	86.33	83.41
Kor12	71.49	76.96	74.61	81.35
Kor90	87.29	78.09	87.59	81.08
LeiMaa12	68.43	74.91	69.56	76.89
LeiMaa90	71.77	78.12	73.20	80.50
Lei12	74.21	78.67	75.51	79.98
Lei90	76.34	81.58	76.77	81.88
Maa12	73.50	75.60	73.33	77.39
Maa90	72.85	78.71	74.61	82.89

Tabel 8 Aandeelmaten voor de kenmerken 'Frans', 'Engels' en 'afkeuring in taalzuiveringswerken' voor veertien kledingconcepten

	Frans				Engels				taalzuivering	
	11C		14C		11C		14C		14C	
	A'	A	A'	A	A'	A	A'	A	A'	A
B12	11.83	15.40	10.60	12.10	15.18	26.13	23.92	41.72	10.20	21.89
N12	19.39	16.98	17.03	13.34	7.71	16.24	17.40	32.28	11.98	19.10
B90	17.63	19.20	17.62	17.81	7.83	8.34	15.03	25.05	-	-
N90	31.71	29.59	28.94	23.25	7.80	12.77	14.86	29.35	-	-
B70	14.90	28.50	14.86	22.39	3.29	7.57	3.55	19.44	-	-
N70	20.78	27.12	20.72	21.31	2.87	15.36	3.07	22.07	-	-
B50	29.00	27.39	29.00	21.52	2.87	9.12	2.87	7.16	-	-
N50	39.49	35.48	39.49	27.88	1.75	10.48	1.75	8.23	-	-
LeuKor12	13.91	32.45	12.21	25.49	17.39	17.06	27.48	34.83	23.83	29.13
LeiMaa12	26.46	32.71	22.87	25.70	18.68	20.99	29.71	37.92	22.76	26.82
LeuKor90	17.23	33.80	19.29	33.04	17.53	13.50	21.61	25.55	-	-
LeiMaa90	37.45	41.84	33.76	32.87	22.64	20.70	29.61	36.91	-	-

Appendix 3 Overzicht van de bronnen voor het materiaal uit 2012

Onderstaande bronnen werden geraadpleegd door Tine De Cnodder in de stedelijke bibliotheken van Herentals, Leuven en Eindhoven, alsook in de Koninklijke Bibliotheken van België (Brussel) en Nederland (Den Haag). De eerste twee maanden werden weggelaten omwille van de onbeschikbaarheid van de magazines Flair en Feeling uit die periode in de Koninklijke Bibliotheek van België.

1 Kledingnamenbronnen voor het vergelijkingspunt Nederland 2012 (N12)
   a. Burda
   b. KNIP
   c. Margriet
   d. Esquire
   e. Cosmopolitan
   f. Libelle (NL)
2 Kledingnamenbronnen voor het vergelijkingspunt België 2012 (B12)
   a. Flair
   b. Feeling
   c. Libelle (BE)

Ethnolect speakers and Dutch partitive adjectival inflection

A corpus analysis

Dirk Pijpops & Freek Van de Velde

TET 67 (2): 343–371
DOI: 10.5117/TET2015.2.PIJP

Abstract

This study applies the methodology described by Gries & Deshors (2014) within the framework of the Contrastive Interlanguage Analysis (Granger, 1996) to the partitive genitive inflection in post-quantifier adjectives in the Moroccan Dutch ethnolect. This implies fitting a logistic regression model on data from the complementary ConDiv and Moroccorp corpora to investigate the differences between the L1 variety and the (early L2 /2L1) ethnolect variety. It was found that the Moroccan Dutch language users do not differ from 'ordinary' Dutch language users in the realisation of the partitive genitive *-s* suffix, neither through an outspoken preference for one of the inflectional variants, nor in the factors determining the alternation. This is considered a rather surprising result, as such differences do exist for a number of other grammatical phenomena (Cornips and Rooij, 2003; Van de Velde and Weerman, 2014). This finding can tell us something about the inflectional status of the partitive genitive. It appears that it is less non-transparent than other quirks in adjectival inflection.

Keywords: partitive, genitive, logistic regression, Contrastive Interlanguage Analysis, Moroccan Dutch, adjectival inflection, ethnolect

1 Introduction[1]

1.1 The partitive genitive *-s* in Dutch

Over time, Dutch attributive adjectival inflection has been drastically reduced, as part of an overall deflexion tendency (van der Horst, 2008, p.143;

van der Horst, 2013), which has especially targeted the nominal domain (Schönfeld, 1970, p.117; Harbert, 2007, p.90). It seems the only vestige of the once rich adjectival case inflection system is an alternation between a schwa and -∅ ending. The inflectional schwa is, however, not the only ending a Dutch adjective may receive. Hiding in an inconspicuous corner of Dutch grammar, an *-s* ending has also survived the turmoils of deflexion. This *-s* suffix can be attached to adjectives when they are postmodifying an indefinite pronoun or numeral (Haeseryn et al., 1997, p.412), as in (1)-(3).[2]

(1) *iets* *bijzonder-s*
something special-GEN
'something special'

(2) *wat* *zinnig-s*
something sensible-GEN
'something sensible'

(3) *veel* *goed-s*
much good-GEN
'a lot of good things'

The genitive case has proven to be quite resilient, surviving well into the twentieth and even twenty-first century (Weerman and De Wit, 1999; Hinrichs and Szmrecsanyi, 2007; Scott, 2011, 2014). Perhaps its most well-known remnant is the prenominal genitive *-s*, which is used to mark possession, as in (4). However, in present-day Dutch, the possessive genitive can only be used on proper names and common names used as terms of address and can only take the form of an *-s* suffix, indiscriminately applied to all genders (Haeseryn et al. 1997: 163). Even so, it faces competition, in particular by the so-called *z'n*-construction, as in (5) (also known as prenominal periphrastic possessive, resumptive possessive pronoun or possessor doubling construction, see (Weerman and De Wit, 1999; van der Horst and van der Horst, 1999, pp. 164-165; Harbert, 2007, pp. 158-161; Allen, 2008, pp. 186-222; Hendriks, 2012).

(4) *Dirk-s* *boek*
Dirk-GEN book
'Dirk's book'

(5) *Dirk z'n boek*
Dirk his book
'Dirk's book'

Like its more famous possessive sibling, the partitive genitive, which will be the subject of this article, has become more limited in use and form throughout its history, as well as subject to competition threatening its very existence. As its name implies, it could express a much wider range of partitive meanings, and could appear on nouns following any kind of quantifier, as exemplified in (6)-(8). In present-day Dutch, however, the partitive suffix can only be used if the quantifier is an indefinite pronoun or numeral followed by an adjectival phrase. Meanwhile, formally, only the *-s* survives, with all other genitive endings disappearing from the language's history. Lastly, an alternative exists in the form of a construction without *-s* suffix (9), which is most popular in – but not limited to – informal language use in the South of the Dutch language area (Pijpops and Van de Velde, 2014).

(6) *veel goed-er ghedachten*
many good-GEN thoughts
'many good thoughts'
(Middle Dutch, van der Horst, 2008, p.575)

(7) *een pont speck-s*
a pound bacon-GEN
'a pound of bacon'
(Middle Dutch, van der Horst, 2008, p.575)

(8) *een corse broot-s*
a crust bread-GEN
'a crust of bread'
(Early Modern Dutch 16th century, van der Horst, 2008, p.1033)

(9) *iets bijzonder*
something special
'something special'

All this leaves the partitive genitive *-s* contrastively, synchronically and diachronically in a peculiar place. Contrastively and synchronically, there seems to be no need to have the suffix, as is demonstrated by English, a

sister language of Dutch, and by Dutch itself, both of which make use of the alternative in (9). Also note that the *-s* in German *etwas Gutes* is not analogous to the Dutch partitive genitive *-s*, but is rather part of the regular and productive general German adjectival inflection. Here, it signifies the neutral singular nominative or accusative case of the entire phrase, as opposed to the Dutch *-s*, which historically signified the genitive case of only the post-modifier, as can be seen when comparing (10) to (11).

(10) German: *etwas* Gut-***(e)s***
[something good]-NOM/ACC-NTR-SG
Dutch: *iets* *goed-s*
[something good-GEN]
'something good'

(11) German: *zu* *etwas* *Gut-**em***
to [something good]-DAT-NTR-SG
Dutch: *tot* *iets* *goed-s*
to [something good-GEN]
'to something good'

Another peculiarity of the partitive genitive is that the adjective follows the quantifier/numeral it is modifying instead of preceding it, as is usual in Dutch. This tendency for a modifier-head sequence is actually becoming ever stricter (van der Horst, 2008, pp. 1946-1961; Van de Velde, 2009, Ch. 3).[3] Lastly, the use of genitival inflection on the adjective is quite uncommon, especially in light of the observation that nouns lost their genitive morphology completely in partitive constructions. The present-day counterparts of examples (6)-(8) all lack genitival inflection.

This peculiar situation in Dutch grammar has attracted a number of diverse theoretical analyses (Schultink, 1962, p. 62; Abney, 1987; Kester, 1996; van Marle, 1996; Broekhuis and Strang, 1996; Haeseryn et al., 1997, p. 356, 432; Hoeksema, 1998a; Booij, 2010a, pp. 223-228; Broekhuis, 2013, pp. 419-461). It is not the aim of this article to delve into this discussion. A large part of it revolves around the question which element is the head, and what bracketing structure should be assumed for this binominal construction, and this is not our central concern here. What is of importance to us is how 'transparent' the function of the *-s* suffix is.

1.2 Non-transparent morphology and the partitive genitive

Some affixes transparently map onto a certain lexical meaning or grammatical function, whereas others have a meaning or function that is not clearly delineated or are non-transparently constrained in their use. Such non-transparent affixes are often obsolescent. An example are the thematic vowels of Indo-European nouns. They are assumed to originally have had derivational meaning, but the meaning is no longer reconstructable. One option for language users is to tolerate non-transparent morphology, and use it as superfluous or irregular inflection. But there is another option as well: quite frequently, language users can be seen to refunctionalise obsolescent morphology in a process of 'exaptation' (Lass, 1990; Van de Velde and Norde, forthc.).[4]

The aversion to quirky morphology is arguably bigger in L2 speakers. The later they acquire the language, the more difficulty they have with morphology, especially with morphology that is constrained in a grammatically complex way (Kortmann and Szmrecsanyi, 2012). This is the reason that languages with a high proportion of L2 speakers tend to be morphologically less complex (Kusters, 2003; Lupyan and Dale, 2010; Trudgill, 2011; Bentz and Winter, 2013). Simplification of complex morphology can either be achieved by discarding the morpheme, as has been done in English plural verbs morphology, or by simplifying the constraints on the morpheme, as has been done in Dutch plural verbs, where the *-en* suffix extended to second person, which used to be expressed by *-t*, to give a simple example. This makes the *-en* more transparent: it marks PLURAL, rather than 1/3.PLURAL.[5]

(12)	Middle Dutch (Loey, 1980, p. 55)	Present-day Dutch
	wi nem-en (we take-1/3PL)	*wij nem-en* (we take-PL)
	gi neem-t (you:PL take-2PL)	*jullie nem-en* (you:PL take-PL)
	si nem-en (*they take* 1/3PL)	*zij nem-en* (they take-PL)

In the domain of Dutch adjectival inflection, the notion of transparency is demonstrably at play: Dutch adjectives alternate between an inflected form with *-ə* and a bare form. In their most basic form, ignoring numerous semantically, grammatically or phonologically conditioned exceptions, the rules can be outlined as in (13): the bare form is associated with predicative use (13a) and the inflected form with attributive use (13b), except in the condition where the adjective is part of a singular indefinite neuter NP (13c).[6]

(13) a. Predicative use: ADJ-Ø

het	*boek*	*is*	*moeilijk*
the	book	is	difficult:BARE

'the book is difficult'

b. Attributive use: ADJ-ə

het	*moeilijk-e*	*boek*
the	difficult-INFLECTED	book

'the difficult book

c. Except: [+sg -def +neutr] NPs: ADJ-Ø

een	*moeilijk*	*boek*
a	difficult:BARE	book

'a difficult book'

A radical simplification of the situation as outlined in (13a-c) would be to get rid of the *-ə* inflection. This is the option that English took in the course of its history. Another, less radical simplification would be to keep the *-ə*, but just ignore the constraint in (13c). That would amount to a refunctionalisation of the *-ə* suffix, which would then transparently mark attributive use, as opposed to predicative use. Van de Velde and Weerman (2014) show that this is indeed what is happening in Dutch, where all kinds of exceptional patterns are made to conform to a simpler system (13a-b), and that the ongoing changes are more outspoken in the Moroccan-Dutch ethnolect, underscoring the role of (early) L2 acquisition.[7] We will return to the adjectival *-ə* below.

The partitive genitive *-s* in Dutch is also part of the adjectival inflection domain, and is, at first sight, rather quirky and non-transparent. The reason is that there is variation in the expression of the *-s* in the partitive construction of a quantifier + adjective. The *-s* ending can be dropped. The variation in the *-s* drop is mentioned in passing by Booij (2010a, p. 244) and Broekhuis (2013, p. 426), but it is dismissed as a regional feature of southern varieties. Now, it is true that large-scale *-s* omission is typical of the southern varieties, but as van der Horst (2008, p. 1624-1625) points out, occasional omission is also attested in northern varieties, and such omissions are possibly even on the increase. Pijpops and Van de Velde (2014a) show that the expression of *-s* is multifactorially determined, as summarised in (14).

(14) More *-s* omission in:
- a. colour adjectives
- b. adjectives *beter*, *fout*, *goed* en *verkeerd*

c. informal registers
d. low-frequency adjectives
e. patterns with quantifiers *iets* and *niets*, as opposed to *wat* en *weinig* (quantifiers *veel* en *zoveel* give equivocal results) in Flanders. In the Netherlands, no difference can be detected between the quantifiers.

This looks like a strange assortment of constraints on the partitive *-s* suffix, but there is a good explanation for most of them.

Constraints (14a) and (14b) can be explained by the influence of superficially resembling constructions: colour adjectives are morphologically indistinguishable from colour nouns, so that *zoveel oranje* 'so many orange things' in (15), as opposed to *zoveel interessant* 'so many interesting things', is structurally ambiguous between Q + ADJ and Q + N (Van de Velde 2001: 150-151). This structural ambiguity is then carried over to cases like *iets oranje* 'something orange', which can only be interpreted as Q + ADJ, as *iets* does not premodify nouns.

Similarly, *iets goed doen* as in (16) is structurally ambiguous between a reading where *goed* is in the partitive Q + ADJ construction: [$_{VP}$ do [$_{object}$ something good]] and a reading where *goed* is an adverb: [$_{VP}$ do [$_{Object}$ something] in a good way].[8] Pijpops and Van de Velde (accept.) were able to demonstrate that *-s* omission occurred even when the adverbial reading was semantically infelicitous, suggesting a similar 'contamination effect' as with the colour adjectives. To put it another way, in both cases, the *-s*-less variant is 'primed' by another, etymologically unrelated, but superficially analogous construction. This effect ties in with the idea that speakers often act on fortuitous similarities, creating 'local generalisations' (see Joseph, 1992; Enger, 2013; Van de Velde and van der Horst, 2013; Van de Velde and Weerman, 2014).

(15) *Hey heeft nl vandaag ook gevoetbalt ofzo ?? <TiredV> ik zie*
Hey has nl today also played_football or_so <TiredV> I see
zoveel oranje [Moroccorp]
so_much orange
'Hey, have the Netherlands also played football today? <TiredV> I see so many orange things.' **[Q + ADJ]**
'Hey, did the Netherlands also play football today? <TiredV> I see so much orange' **[Q + N]**

(16) *<katertje> ik heb toch nog iets goed gedaan vandaag*
<katertje> I have still yet something good done today [ConDiv]

'<katertje> at least, I've done something good today.'
[$_{VP}$ **do** [$_{object}$ **something good**]]
'<katertje> at least, I've done something correctly today.'
[$_{VP}$ **do** [$_{Object}$ **something**] **in a good way**]

The constraint in (14c) is a more general effect that standard norms are more strictly adhered to in formal written registers and (14d) shows the well-known Conserving Effect (Bybee, 2006; Bybee and Beckner, 2010) that applies cross-linguistically and states that more frequent forms are better shielded against morphological change. If Van der Horst (2008: 1624-1625) is right in assuming that the *-s* may be losing ground in the north, the constraints in (14c) and (14d) point to an ongoing change 'from below'.[9] Constraint (14e) is a so-called interaction effect: there is a difference between the quantifiers, but it only plays in Flanders.

1.3 L2 speakers, non-transparent morphology, and the Dutch partitive genitive

The question we address in this paper is what early L2 or 2L1 speakers of Dutch do with the morphological quirk of the Dutch partitive genitive.[10] From earlier studies, we know that early L2 or 2L1 speakers speed up refunctionalisation in adjectival inflection, as argued above (see also Van de Velde and Weerman 2014 and references cited there).[11] Early L2 speakers increase the transparency of adjectival *-ə* by loosening the unmotivated constraint in (13c). But what happens with adjectival *-s*? Do early L2 speakers similarly experience difficulties with the constraints on the partitive construction, or is the suffix transparent enough to let early L2 speakers pick up on this morphological signal?

We will investigate to what extent early L2 speakers differ from L1 speakers with regard to the partitive genitive *-s* by applying the method described in Gries and Deshors (2014). To this end, we make use of two complementary corpora: the internet chat relay part of the ConDiv Corpus (Grondelaers et al., 2000) and Moroccorp, a corpus of the Moroccan ethnolect variety of Dutch, as used by adolescent Moroccan early L2 and 2L1 speakers of Dutch (Ruette and Van de Velde, 2013).

This study has four foci, two specifically related to Dutch and two more general ones. On a more specific level, this study first fits in the research on the language use of multilingual youngsters, more specifically street language and ethnolect and its (potential) influence on the future of Dutch, as discussed in Cornips and Rooij (2003). Second, it is an extension of a quantitative study of the partitive genitive in Dutch commenced in Pijpops

and Van de Velde (2014a), and will further flesh out the description of the partitive genitive in present-day Dutch grammar. On a more general level, this article can be seen as a concrete study of deflexion and the role of second language learners or language contact in this process, tying in with recent findings on the relation between demography and language change (Lupyan and Dale, 2010; Bentz and Winter, 2013). Lastly, in comparing Moroccorp to a 'standard' in the form of a subcorpus of ConDiv, the corpus study reported here aspires to contribute to the methodology of Granger's (1996) Contrastive Interlanguage Analysis (CIA). This framework aims to compare a native language to an 'interlanguage', by making use of quantitative analyses of comparable corpora. Its goal is to uncover "factors of foreign-soundingness" (Granger, 1996, p. 43) by investigating in what way the interlanguage corpus differs from the native language corpus, mostly with respect to a specific phenomenon, e.g. the use of adverbial and adnominal participles by learners of English with a French and Dutch background (Cosme, 2008; Gries and Deshors, 2014, p. 110). More specifically, this paper hopes to answer Gries and Deshors' (2014) call for the use of more sophisticated analytical approaches in this research tradition, by applying the regression-based methodology described in their article. In fact, the first step in the methodological procedure proposed by Gries and Deshors (2014) has already been completed in Pijpops and Van de Velde (2014a). This step is needed to assess which factors determine a particular linguistic choice by native speakers (Gries and Deshors, 2014, p. 111). As explained below, the Moroccorp and ConDiv corpora are ideally suited for such contrastive analyses; while the present investigation zooms in on the partitive genitive, these corpora can be used to apply this methodology to any type of linguistic alternation.

Our concrete research question is the following:

> Do the language users in Moroccorp differ from the language users in the ConDiv chat corpus in their realisation of the partitive genitive alternation, either in absolute numbers (e.g. less partitives or more *-s* omission in Moroccorp) and/or in the number and/or choice of factors determining the alternation?

We have four competing scenarios:

a. Moroccorp language users consistently use the [+ s] variant, simplifying the rule by generalizing the most frequent variant ('the hyper-Netherlandic *overgeneralisation* option'). This would mean that adjectival *-ə* and *-s* inflection are treated similarly. Like the adjectival *- ə* ending, the

-s would be an integral part of the grammar of Moroccorp language users and would arguably have experienced an increase in transparency.

b. Moroccorp language users consistently use the [-∅] variant, simplifying the rule by deleting the superfluous morpheme in an unproductive case system ('the English *deflexion* option'). This would suggest that the partitive *-s* is a vulnerable victim in an ongoing deflexion trend, and does not carry a transparent function that can be easily picked up by early L2 language users of Dutch.

c. Moroccorp language users do not differ from their ConDiv L1 peers ('the constraint-sensitive *no-difference* option'). Firstly, this can be taken to indicate that Moroccorp language users are quite capable of picking up the exact construction in which to apply the *-s* ending. Secondly, it would suggest that unlike the factors determining the adjectival *-∅/-ə* alternation in L1 Dutch, those determining *-∅/-s* alternation operate at a level which is not readily bypassed by Moroccorp language users.

d. Moroccorp users differ from their ConDiv L1 peers, in using more or other constraints ('the hyper-sensitive *aemulatio* option'). They might for instance use a wider range of constraints on the realisation of the *-s*, as is done by southern L1 speakers, for whom there is a differentiation between the quantifiers (see 13e, above).

A difference in the number and/or choice of factors determining the alternation (i.e., scenario d) would be of particular interest to the CIA analysis of Moroccan Dutch. In which cases do these speakers have more trouble with the *-s* suffix and why? Is the multifactorial grammar behind the partitive genitive different in Moroccan Dutch? To answer this question, we will have to look into the interactions in the regression model (Gries and Deshors, 2014, p. 120-126).

This article is structured as follows. Section 2 introduces the corpora used in this study (Moroccorp and ConDiv), explains why they are suited for the type of research envisaged here, and gives information on the extraction and analysis of the data. In Section 3, we carry out a regression analysis on these data and discuss the results in the perspective of the research question posed above. Section 4 summarizes the conclusions.

2 Corpora and data

2.1 Corpora

The data used in this corpus study have been extracted from Moroccorp and the subcorpus of ConDiv containing the Netherlandic chat material.[12] ConDiv is a lectally stratified corpus of Netherlandic and Flemish Dutch (Grondelaers et al., 2000). The subcorpus we have used here amounts to roughly 7 million tokens. Moroccorp (see Ruette and Van de Velde, 2013) is a 10 million token corpus compiled to reflect Dutch as is used by speakers of second or third generation Moroccan immigrants, who use an ethnolectal variety of Dutch. Their language use exhibits a number of characteristics typically associated with early L2 or 2L1 varieties of Dutch, like the well-known *Ausgleich* of Dutch attributive adjectival inflection in indefinite singular neuter NPs (see 12c, above). The advantage of these corpora are that they are comparable in size and register,[13] which is one of the most important requirements of the methodological approach followed here (Gries and Deshors, 2014, p. 110), but also that the results we get for the adjectival partitive *-s* can be compared to the findings on adjectival *-ə* in Van de Velde and Weerman (2014), who used the same two corpora.

The ConDiv corpus is regionally stratified, and contains both a Netherlandic chat component and a Flemish chat component. We only used the Netherlandic component, as Moroccorp only contains Netherlandic ethnolect speakers. Including the Flemish ConDiv chat component might therefore compromise the comparability of both corpora. Indeed, it has already been established that there exist important differences between the partitive genitive alternation in Flanders and in the Netherlands (Pijpops and Van de Velde, 2014).

As our dataset will thus be entirely comprised of chat material, this particular form of language use requires some further specifications. To use the words of Koch and Österreicher (2007, p. 359):

> *Der Chat ist sogar eines der schönsten Beispiele dafür, dass im graphischen Medium eine relative, natürlich immer limitierte Annäherung an dialogische, spontane Nähesprachlichkeit möglich ist.*

While on the medial dichotomy, chat is of course part of *graphic*, i.e. written, language use, its place on the conceptual continuum from written to spoken is more on the side of spoken language (Koch and Österreicher, 2007; Söll, 1974; Grondelaers et al., 2000, p. 358). Still, Koch and Österreicher (2007, p. 359), as well as Ágel and Hennig (2007, pp. 202, 206-214) point

to a possible inhibiting influence of the graphic medium on language use. As for manners, however, its written and physically distant nature may also generate a disinhibiting effect (Suler 2004). That is, as can be inferred from a quick glance at the comments section of a newspaper's website, people may feel freer to speak their mind or to use ruder language from behind a computer screen – especially if anonymity is guaranteed – than would be considered acceptable in face-to-face conversations.

Not all types of chat conversations can be seen as part of the same discourse tradition, or as positioned on the same place in the conceptual continuum from spoken to written language. A chat conversation between a software technician and a customer experiencing a programme malfunction will be of a different nature than one between brother and sister. The chatlogs gathered in our corpora contain language from youngsters who typically only know one another from the chat box. To further characterise the conversational setting, we make use of Koch and Österreicher (2007, p. 351) communicative parameters:

- **Privatheit/Öffentlichkeit:** Anyone can join, store and even publish the chat conversations – as has been done in the compilation of the corpora – and the chatters are sometimes explicitly aware of this (Ruette and Van de Velde, 2013, p. 461, 464). Still, the chatters mostly assume they are amongst like-minded peers – they can see who is 'listening in' and, after all, their identity is protected by anonymous nicknames. As such, their conversations are often of a private nature.
- **Vertrautheit/Fremdheit der Kommunikationspartner:** While the chatters mostly do not know one another in real life, this is not to say that they are necessarily strangers. Some are regular visitors of the chat box, and recognise each other by their nicknames.
- **Starke/geringe emotionale Beteiligung:** personal problems and (strong) personal opinions are often discussed. Emotional participation is rather strong (cf. Suler's 2004 Online Disinhibition Effect).
- **Situations- und Handlungseinbindung/-entbindung:** a brief look at the corpus material shows that the discussed subjects are mostly situational and act-oriented matters, rather than abstract ideas.
- **Referenzielle Nähe/Distanz:** Referential distance.
- **Raum-zeitliche Nähe/Distanz:** Spatio-temporal distance.
- **Kommunikative Kooperation/keine komm. Koop.:** Communicative cooperation.
- **Dialogizität/Monologizität:** Multiple dialogs are often held simultaneously.
- S**pontaneität/Reflektiertheit**: Spontaneity.

– **Freie Themenentwicklung/Themenfixierung**: Free subject development.

These parameters accumulate to a conversational setting of considerable communicative proximity (Koch and Österreicher, 2007, p. 351). As for language material that can easily be gathered in digital corpora on a large scale, this kind of data approximates 'natural' spoken language use fairly well. This is important, as normative works explicitly repudiate the variant without *-s* ending (Taaladvies.net, Taaltelefoon.be, Taalnet) and we know from earlier work that this norm affects formal language use – although even there, it is not strictly adhered to (Pijpops and Van de Velde, 2014).

2.2 Data

The data from the Netherlandic chat corpus could be reused from an earlier study of the partitive genitive in Pijpops and Van de Velde (2014a). The Moroccorp data, however, had to be extracted from scratch, and because Moroccorp is not syntactically annotated they had to be manually checked in order to ensure that the occurrences gathered were in fact genuine partitive genitives.[14]

All instances in which one of the following quantifiers preceded one of the following adjectives, with or without -s suffix were extracted from Moroccorp and manually checked. The extraction was done using the freely available *AntConc* software (Anthony, 2011).

– Quantifiers: *iets* ('something'), *niets* ('nothing'), *wat* ('something'), *veel* ('a lot'), *zoveel* ('so much')
– Adjectives: *aardig* ('nice'), *apart* ('apart'), *belangrijk* ('important'), *beter* ('better'), *bijzonder* ('particular'), *blauw* ('blue'), *concreet* ('concrete'), *deftig* ('decent'), *dergelijk* ('similar'), *erg* ('awful'), *geel* ('yellow'), *gek* ('crazy'), *goed* ('good'), *groen* ('green'), *interessant* ('interesting'), *klein* ('small'), *lekker* ('tasty'), *leuk* ('fun'), *mooi* ('beautiful'), *nieuw* ('new'), *nuttig* ('useful'), *oranje* ('orange'), *positief* ('positive'), *purper* ('purple'), *raar* ('weird'), *rood* ('red'), *spannend* ('exciting'), *speciaal* ('special'), *verkeerd* ('wrong'), *verschrikkelijk* ('horrible'), *vreemd* ('weird'), *warm* ('warm'), *wit* ('white'), *zinnig* ('sensible'), *zwart* ('black')

For the selection of the adjectives, we used three criteria:
– The adjective had to occur at least seven times in post-quantifier position in the Corpus of Spoken Dutch (CGN) (van Eerten, 2007). The reason why we resorted to another corpus for this selection criterion

is that we needed a PoS-tagged corpus to ensure we included all the relevant attestations. To the resulting set, we added the major colour adjectives, if they did not violate the next criterion, because we have a special interest in them (see 13a).

– Adjectives ending in a (post-)alveolar fricative were excluded, because the presence of a partitive *-s* is phonologically indiscernible. This is the reason why the list above includes *purper* ('purple'), but not the more common colour term *paars* ('purple'), for instance.
– Adjectives that are homonymous with plurals of nouns were excluded as well. Cases in point are *ouders* ('elder-GEN' or 'parents') and *extra's* ('extra-GEN' or 'bonuses')

The selection of the quantifiers is based on the lists in Haeseryn et al. (1997, p. 356, 432), applying a threshold of at least one occurrence in both corpora of the combination of the quantifier and one of the selected adjectives. Note that our dataset does not contain the quantifier *weinig* ('few'). The hits of *weinig* were originally extracted from Moroccorp, but it was later found that the entire Moroccorp and ConDiv dataset contained no occurrences of *weinig* without *-s* suffix. This leads to problems with the estimates in the regression model. A total of 11 hits of *weinig* were removed from the dataset, leaving us with a dataset of 2378 partitive genitives occurrences in total.

These datapoints are used to carry out a logistic regression analysis. This statistical technique measures the effect of one or more explanatory variables, or 'predictors', on a binary response variable (see Baayen, 2008; Gries, 2013; Speelman, 2014). In the case at hand, we want to predict under what circumstances the partitive *-s* suffix is realised or not. Instead of testing each of the relevant predictors (to be introduced below) bivariately, by means of an association test like a Chi-Square or Fisher's Exact test, a multiple logistic regression allows one to get an idea of the impact of a predictor while controlling for all other predictors. Such an approach is statistically superior, as separate bivariate testing can yield problematic results, for instance by ignoring 'interaction effects', when the effect of one explanatory variable differs depending on the levels of another explanatory variable.

Technically, what regression does is minimizing the distances between observed values in a Cartesian n-dimensional space and a so-called 'hyperplane'. The geometric properties of this hyperplane are 'fitted' to the data. A crucial step in this fitting process is to decide on the number of predictors to be included in the regression model. The more predictors we add to

the model, the more accurately we can predict the values of our observations, but regressing means finding a balance between having enough predictors to make reasonable predictions on the one hand, and 'overfitting' the model, with a concomitant lack in the summarising power. To find the optimal balance, we take two measures.

The first measure is to make use of a so-called 'mixed-effect' logical regression model. Such models make a distinction between 'fixed effects' – the predictors you are actually interested in – and 'random effects' – predictors that can be assumed to have an actual effect, but are not of interest, and the levels of which vary randomly if you were to redo the analysis on a different dataset. Technically, random effects assume different slopes and/or intercepts of each of the 1-dimensional regression lines that together make up the hyperplane. In our study, we use random effects with different intercepts.

The second measure for finding the optimal balance is using a bidirectional step-wise procedure for variable selection. This procedure involves introducing and dropping variables in the model and seeing how this affects the Akaike Information Criterion (AIC), a measure allowing for quality comparison between models. The lower the AIC, the better the model fits the data.

The analyses have been carried out with the aid of the open-source software R (R Core Team, 2013).[15] The variables included in this procedure are listed below.

Response variable:

– *-s*: with, without

Explanatory variables:

– Corpus:	*ConDiv, Moroccorp*
– Quantifier:	*iets* ('something'), *niets* ('nothing'), *wat* ('something'), *veel* ('a lot'), *zoveel* ('so much')
– Type-Adjective:	*other, deviant, colour*
– Length-Adjective:	*1, 2, 3, 4*
– Number-of-words-AP:	*1, 2*
– Frequency:	log-transformed frequency of the phrase
– Phrase:	*iets leuk*(*s*) ('something fun'), *niets zinnig*(*s*) ('nothing sensible'), *weinig concreet*(*s*) ('few concrete things'),…

The set-up of the present study requires a slightly different set of predictor variables than the one in Pijpops and Van de Velde (2014a). Four adjustments were necessary. First, the variable *Corpus* has of course been added, which distinguishes between the ConDiv subcorpus of Netherlandic chat language and Moroccorp material. In the present study, this explanatory variable carries major theoretical weight. Second, the variable *Variety*, distinguishing between the Netherlands and Flanders was removed, for reasons discussed above. Third, the variable *Register* was dropped. This variable exploited the register stratification of ConDiv, but could obviously not be retained in a model that exclusively looks at chat data. In Moroccorp we have no counterpart for the more formal registers in ConDiv. Fourth, we have used base 10 for the logarithmic transformation of frequency, as it is more easily interpretable than the base we used in Pijpops and Van de Velde (2014a). This does not make a real difference for the model as such, as different bases only differ by multiplication by a constant (Fox and Weisberg, 2011, p. 127).

The rest of the variables are copied from the Pijpops and Van de Velde (2014a) study:

– *Type-Adjective* makes a distinction between (i) colour adjectives, (ii) the adjectives *beter* ('better'), *goed* ('good'), *fout* ('incorrect') and *verkeerd* ('wrong'), which are here called 'deviant adjectives' for terminological convenience, and (iii) all the other adjectives. The reason is that (i) and (ii) were assumed to display higher rates of *-s* drop, for reasons stated above, see (14a-b).
– The variable *Length-Adjective* indicates the phonological weight of the adjective, expressed as the number of syllables. Note that the adjective *interessant* ('interesting') is counted as a three-syllable word, as the first schwa is often syncopated.
– *Number-of-words-AP* distinguishes between single-word adjectives and adjectives that are premodified by a degree adverb, e.g. (17) and (18).
– *Frequency* is the logarithmically transformed frequency of the phrase type – phrase type being the unique combination of a Q + ADJ pair. The uniqueness disregards the presence of the *-s* suffix, so that *iets moois* ('something beautiful-GEN) and *iets mooi* ('something beautiful-Ø') belong to the same phrase type. It does however take into account whether or not the adjective is premodified by a degree adverb, so that *iets beter*(*s*) ('something better(-GEN)') and *iets veel beter*(*s*) ('something much better(-GEN)') do belong to different phrase types.
– The variable *Phrase* has all these distinct phrase types as individual levels.

(17) <*engeltje*> *dag:* *liever* ***iets*** ***lekker*** ***warm:***) <*dag*> *ok,*
<engeltje> dag preferably something tasty warm <dag> ok
er *is* *melk,* *koffie,* *nesquick*… **[Condiv]**
there is milk coffee nesquick
'<engeltje> dag: I'd prefer something nice and hot. <dag> ok, there is milk, coffee, hot chocolate milk …'

(18) *ik* *heb* *die* *van Utrecht laten lopen, maar geloof me ik heb*
I have that_one of Utrecht let run but believe me I have
iets ***veel*** ***beter-s*** *nu* **[Moroccorp]**
something much better-GEN now
'I let the one from Utrecht go, but believe me, I've got something much better now.'

The variable *Phrase* will be implemented as a random effect in the regression models, whereas all other explanatory variables will be entered as fixed effects into the variable selection procedure. *Corpus, Quantifier* and *Type-Adjective* are categorical variables, *Length-Adjective, Number-of-words-AP* and *Frequency* numeric ones.

3 Results and discussion

3.1 Results

As said in the previous section, we analysed the dataset by means of mixed-effects (multiple) logistic regression modelling, more specifically, by using the procedure proposed by Gries and Deshors (2014, p.122-136). Our method does, however, differ from their example in one aspect. Whereas Gries and Deshors (2014) only entered interactions between the fixed effect *Corpus* and the other fixed effects in the variable selection procedure, we included all possible two-way interactions between any of the fixed effects under scrutiny.

The model that we arrived at was subjected to a number of additional diagnostics (see Speelman 2014). First, all predictors which did not significantly improve the model were dropped. Second, we tested whether the residual deviance was not much larger than the degrees of freedom. Large residual deviance signals overdispersion, suggesting that the data behave too heterogeneously to have confidence in the model. Third, we carried out a Hosmer-Lemeshew-Cessie goodness-of-fit test. If the test result is significant, the model does not fit well, for instance because there are important

predictors missing from it (Speelman, 2014). Fifth, we checked whether all Variance Inflation Factors were below 4, to make sure the model does not run into the problem of multicollinearity, meaning that several variables measure the same thing – obviously something one wants to avoid in a maximally parsimonious regression model. None of the diagnostics yielded problematic results, so, in a final step, we added the random effect *Phrase* and once again removed all predictors which no longer made a significant contribution to the model's quality.

This left us with the model presented in Table 1. As can be appreciated, the number of parameters is well below the number of observations of the response variable's least frequent level divided by twenty, which is another rule-of-thumb in logistic regression modelling. The C-index gives an idea of the overall quality of the model. Values above 0.80 are considered as satisfactory. With 0.8420, we have a powerful regression model. The predictors are ordered from most to least important.

Note that the categorical variables were implemented using dummy coding. This means that one level is taken as the reference level, and the others are used as separate regressors. The numeric variables could be implemented directly. All estimates and confidence intervals of the estimates are rounded off to 2 decimals, the p-values are rounded off to 4 decimals.

The estimates give an indication of the effect size. Because the variant without *-s* is the success level of the response variable, a positive value for the estimate of a numeric predictor means the probability of *-s* drop rises as the value of the variable increases. Meanwhile, a positive value for the estimate of one of the levels of a categorical variable means that there is more *-s* drop than in the reference level. Conversely, a negative sign means that the level in question favours *-s* retention. The higher the absolute value, the more severe the impact is. A value of 4.34 for colour adjectives, for instance, means that the logit of the *-s* drop is 1.28 (4.34 plus the intercept -3.06).

The logit is a (double-)transformed measure of the probability, and the corresponding probability can be computed by taking inverse logit function of the estimates. For the adjectives, this means that colour adjectives have a 'partial' estimated probability of 0.78.

A more human-friendly visualisation of the results is given in the 'effect plots' (Fox 2003) in Figure 1, which directly give the estimated probability of *-s* drop and a corresponding confidence interval.[16] The 'bar code' on the x-axis on the right side is a so-called 'rug', indicating the marginal distribution of the Frequency observations. Each bar stands for one phrase type.

— AIC: 1240
— C-index: 0.842
— Number of phrases: 96
— Total number of hits: 2378
— Hits with *-s*: 2143
— Hits without *-s*: 235

Table 1 Regression model predicting *-s* drop

Predictors	Levels	Estimates	Confidence intervals		P-values
			2,5%	97,5%	
	intercept	-3.06	-4.32	-1.80	< 0.0001
Type-Adjective	*other*	Reference level			
	deviant	1.78	1.07	2.48	< 0.0001
	colour	4.34	2.71	5.97	< 0.0001
Frequency		0.14	-0.61	0.89	0.7136
Corpus	*ConDiv*	Reference level			
	Moroccorp	0.98	-0.33	2.28	0.1416
Interaction Frequency - Corpus	*ConDiv*	Reference level			
	Moroccorp	-0.87	-1.59	-0.15	0.0175

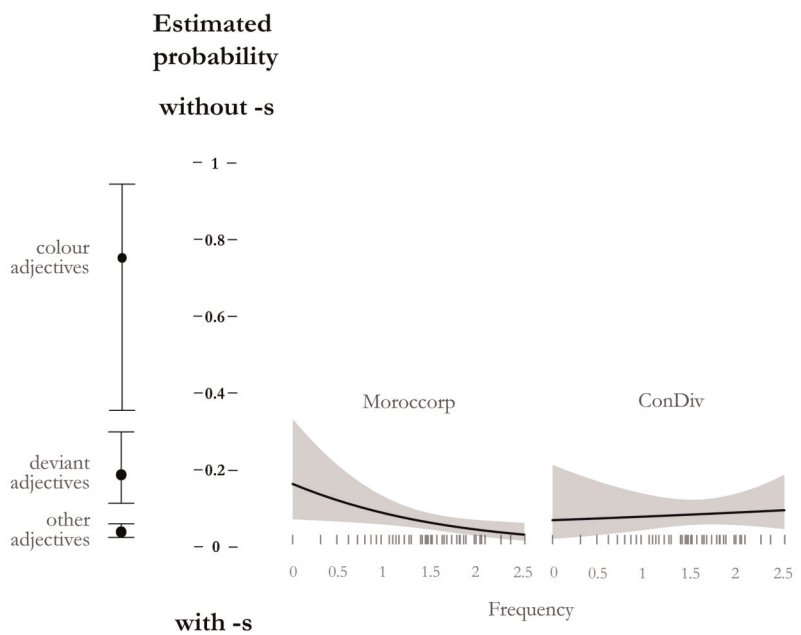


Figure 1 Probabilities, estimated by the regression model. The influence of Type-Adjective is strong and stable across both corpora. The influence of Frequency seems to differ, at first sight.

Our most important predictor is *Type-Adjective*, followed by *Frequency* and *Corpus*. The model also contains an interaction between *Frequency* and *Corpus*. This interaction seems to indicate that the Moroccorp chatters tend more towards *-s* omission in the low frequency phrases than the Netherlandic chatters of ConDiv. However, it would be inadvisable to make sweeping conclusions on the basis of this effect. First, exactly because the infrequent phrases are infrequent, there is a lot of uncertainty about their behaviour, as can be seen in the large confidence intervals in the low frequencies in Figure 1. Second, as can be seen in Figure 2, over two thirds of our dataset is made up of Moroccorp material. This means that the calculated frequencies are more strongly influenced by the Moroccorp material than the ConDiv material, which may explain why we only find the frequency effect in the Moroccorp data. Finally, as explained below, neither *Frequency* nor *Corpus* seems vital to the overall model.

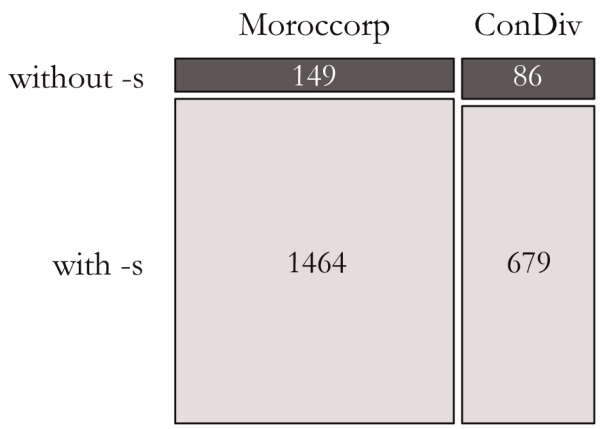


Figure 2 Partitive genitives in the Moroccorp and ConDiv corpora

As can be seen in Figure 1 and Figure 2, there is hardly any difference between the Moroccorp and ConDiv material regarding a general preference for *-s* omission. In fact, *-s* retention is even slightly more frequent in Moroccorp, though this difference is not significant, neither in the regression model (see Table 1) nor in a bivariate chi-squared test ($p = 0.126$). Also, with regard to the factors determining the alternation, Moroccorp and ConDiv are very similar. The regression model contains only a single interaction with *Corpus*, which, as stated above, does not seem to be of crucial importance. If we were to remove not only this interaction from the model, but also the main effects of *Corpus* and *Frequency*, retaining only the random effect *Phrase* and the fixed effect *Type-Adjective*, the model's predic-

tive quality is even slightly raised, to a new C-index of 0.844.[17] The high predictive quality of this last model at the cost of so few predictors is perhaps what is most surprising about our analysis, and another testimony to the power of the *Type-Adjective* predictor (cf. Pijpops and Van de Velde, 2014, p. 18).

It appears then, that Moroccorp users are equally sensitive to the *Type-Adjective* constraint as the ConDiv users. This is confirmed if we measure how well a ConDiv-only model can predict the Moroccorp observations. The reasoning behind this procedure is as follows: let's first build a bivariate logistic regression model with *Type-Adjective* as the only fixed effect based on the ConDiv data only. Then we fit this ConDiv model on the Moroccorp data. If the Moroccorp data are highly comparable to the ConDiv data, the model should yield a good fit, and the differences between the estimated values of the model and the observed values should be small. This is indeed what we find. The mean estimated difference is only off by 0.04.[18]

In theory, there is a possibility that the similarity between ConDiv and Moroccorp is overestimated: Moroccorp users may avoid the partitive genitive construction when in doubt, and only use it in straightforward cases.[19] We tested this hypothesis in two ways. First, we checked whether Moroccorp has fewer instances of the partitive genitive per 10,000 tokens. This is emphatically not the case: ConDiv has 1.10 instances per 10,000 tokens, while Moroccorp has 1.57 instances per 10,000 tokens. So if anything, Moroccorp chatters use the construction *more* than ConDiv chatters. Second, we looked at whether there is more variation in the phrase types used in either of the corpora. Moroccorp has more different phrase types than ConDiv (86 vs. 67), but of course, we have to correct for the fact that Moroccorp is 47% larger than Condiv. Correcting is not so easy in this case. It does not make sense to just calculate the number of phrase types per 10,000 tokens, because type accumulation peters out as the corpus size grows, due to the Zipfian distribution of phrase types (see Bentz et al. 2014). To correct for this, we divided the number of phrase types by the logarithm of the corpus size. Again, Moroccorp users outperform the ConDiv chatters in using the construction with more different phrase types.

3.2 Discussion

We have found there to be hardly any difference between the Moroccorp and ConDiv corpora concerning the partitive genitive alternation. Also, both corpora exhibit a variance-suffocating dominance of nearly 90% of the *-s* retaining variant. Because of this, we believe there is little reason to

continue along the path sketched out by Gries and Deshors (2014, p. 126-131) by applying what they call the 'MuPDAR' approach to the partitive genitive alternation in Moroccorp and ConDiv. This MuPDAR analysis ('Multifactorial Prediction and Deviation Analysis with Regressions') is a statistical analysis of the deviation in the choices the two varieties make for a certain linguistic response variable. Instead, we can immediately turn to answering the research question.

The partitive *-s* realisation of the Moroccorp language users does not differ from that of the Netherlandic native speakers of the ConDiv chat corpus, or even from the Dutch written standard language (if we take into account the results of Pijpops and Van de Velde 2014, p. 19-20, 23), neither in absolute numbers (e.g. more *-s* omission in Moroccorp), nor in the number and/or choice of factors determining the alternation. Moreover, Moroccorp chatters do not use the construction less frequently, or in a lexically less varied way. Especially striking is the strong influence of *Type-Adjective* in both ConDiv and Moroccorp. With regard to the four scenarios set out in Section 1.3, our data suggest that the third scenario, the so-called 'constraint-sensitive *no-difference* option', is the correct one.

This entails that the Moroccorp language users do not simplify the adjectival *-∅/-s* alternation, by overgeneralizing the *-s* inflection, i.e. Scenario (a) in Section 1.3., in contrast to what is the case with the *-∅/-ə* alternation (see Van de Velde and Weerman, 2014 for several case studies of adjectival inflection patterns, where Morrocorp and ConDiv do differ). This suggests that the factors governing the *-∅/-s* alternation, most notably *Type-Adjective*, are of a different nature than those governing the *-∅/-ə* alternation. L2 language users appear equally responsive as L1 users to links between superficially resembling, yet structurally unrelated constructions (see the explanation of constraints (13a) and (13b) in Section 1.2).[20]

Furthermore, contrary to what could have been expected on the basis of morphological difficulties in L2 language use, Moroccorp language users do not wash away the partitive *-s* in the waves of ongoing deflexion either, i.e. Scenarion (b) in Section 1.3. If Moroccorp chatters are statistically more likely to jettison non-transparent morphology, then our results could be taken as an indication that the *-s* suffix is not so 'odd and quirky' after all. In our view, the transparency of the partitive *-s* derives from the fact that it has a construction-marking function.

The partitive genitive is best seen as a construction, in the sense of Construction Morphology (see Booij, 2010a,b for an extensive motivation for the need of a constructional approach to morphology), and the *-s* is interpreted as a partitive ending only when it occurs in the construction at

issue. In other words: it derives its function from the construction it occurs in. This is why Booij (2010a, p. 211-236) speaks of 'construction-dependent morphology'. In itself, the partitive genitive *-s* is ill-motivated, as it does not form part of a case paradigm. Dutch has largely lost its case system, and the refunctionalised *-s* of the genitive is found on nouns, not on adjectives. In isolation the inflectional *-s* does not signify partitivity; it needs the constructional template to be interpreted felicitously. Following Booij (2010a, p. 227), we can use the formalisation in (19) to capture the constructional meaning, in which the formal part is specified on the left-hand of the double arrow, and the function part is specified on the right-hand side.

10a

(19) $[NP_i [\ldots [X\text{-}s]_A]_{APj}]_{NPk} \leftrightarrow [Quantity_i \text{ with } Property_j]_k$

In a construction grammar approach, the *-s* is not necessarily quirky. If it surfaces across the board as soon as the constructional template occurs, it is more transparent than the rule in (13), with its intricate grammatical condition in (13c). Admittedly, the regression model shows that there are conditions on (19) as well, but as explained in Section 1.2, these conditions are not grammatical, but semantic and pragmatic in nature, and derive from contamination from superficially resembling constructions.[21]

4 Conclusions

We conclude this article by summarizing what the results presented here mean for the various 'tracks' on which this study is situated. As for the research into the Dutch language use of early L2 ethnolects, it is remarkable how good the ethnolect language users are at adopting what at first sight can only be considered – in comparison to related languages – a weird quirk in Dutch grammar. It may prove to be a fruitful undertaking to further investigate exactly for which linguistic phenomena these language users do diverge from 'mainstream' language use, and for which phenomena they do not. For instance, do they prove to be more creative with contextual-inflectional, inherent-inflectional, derivational or syntactic processes (see Booij, 1996 for the first two terms)?

Focusing on the partitive genitive, we think this study can be taken as evidence that the partitive genitive has good prospects for survival, at least in the Netherlands. Grand-scale *-s* omission seems at present limited to Belgium. In the Netherlands, though it is not supported by a productive

case system, the suffix survives in a specific constructional niche. This has been observed for other remnants of the genitive case as well by Hoeksema (1998b). In fact, it is not unusual for old constructions to revive in specific corners of the grammar (Van de Velde 2015).

Finally, we hold the present study to be of relevance to CIA-research. Although the findings can be considered null results in terms of a CIA-analysis, our study still shows the applicability and feasibility of the method of Gries and Deshors (2014) on morphological alternations, just as they have already shown it for lexico-syntactic alternations. In this respect, the present study contributes to the methodology of the field of Dutch language variation.

We hope that this study will add to the understanding of the fascinating inflectional quirk of the Dutch language that is the partitive genitive. Its future might turn out to be somewhat brighter than is sometimes assumed (van der Horst, 2008, p. 1624-1625), and its quirkiness seems to be less of a problem for Moroccorp chatters than for Dutch linguists.

Notes

1. This article has profited from the comments by two reviewers, as well as by Timothy Colleman, whom we would all like to thank. The research carried out was supported by a BOF research grant from the University of Leuven and a fellowship from the Research Foundation Flanders (FWO).
2. Although historically, the partitive genitive *-s* is clearly part of the Dutch case inflection system, it may be synchronically more cautious to consider it an isolated irregular suffix, as the case inflection system has long collapsed and its debris is scattered around in Dutch grammar. In the glosses, the *-s* has been marked -GEN.
3. This is sometimes seen as an argument to assume that the quantifier is no longer the head of the construction, but the adjective is (Van Marle 1996: 73, 80).
4. To illustrate the notion of exaptation, consider the fate of negation in some dialects of Dutch. Where an erstwhile negation marker, *en* was reanalyzed into a marker of subordination. The clitic *en* lost its transparency when negation *niet* took over in the Dutch Jespersen Cycle. In the gradual decline of *en*, relic attestations were more frequent in subordinate clauses, and this created the conditions for a reanalysis in which *ne* was seen as a subordinator (see Van der Auwera 2012: 413, with reference to earlier work by Overdiep and Neuckermans). In example (i), the *en* historically derives from a negation particle, but it occurs in subordinate clauses with positive polarity. (i)
   Toen we bij de poort en kwamme ...
   when we at the gate SUBORDINATOR came
   'when we arrived at the gate'
5. The sitation is more complex, as the *-en* affix is also used for the infinitive and many varieties of Dutch have apocope of the *n* in speech. The old 2PL -t ending is attested until the 20th century (b.v. *jullie gokt* 'you gamble', WNT s.v. wereld) and is still present

in some dialects. For a more in-depth theory of transparency in morphology, see among others Leufkens (2015).

6. In the southern spoken Dutch variety of Flanders, the condition in (13c) is slightly different: definiteness does not play a role.
7. In actual fact, the account in Van de Velde and Weerman (2014) is more complicated. They show that in premodifying position the *-ə* suffix is used to demarcate the determiner from the adjective. The gist of the account is the same as what is reported here, however: the non-transparent condition in (13c) is replaced by a more transparent one, which amounts to refunctionalisation / exaptation.
8. The bracketed representation is meant as an approximate formulation. No theoretical significance should be attributed to it.
9. The actual history is more complicated. The *-s* was not consistently expressed in earlier stages of Dutch, so it is possible that it was introduced in the north by a change 'from above' first, in a more general tendency of reviving the case system in Early Modern Dutch, by copying the revered Latin language, and then later suffered from change from below, reintroducing the *-s*-less variant.
10. We are looking into informal Dutch produced by language users of Moroccan ethnicity. For convenience sake, we will refer to the variety under study as an '(early L2) ethnolect', though it may be more accurate to see it as a 2L1 variety.
11. Late L2 speakers are more likely to 'fossilise' the construction (see Matras 2009: 75 for this term).
12. To avoid misunderstandings, we will use Dutch to refer to the Dutch language, including the Flemish variety and Netherlandic to refer to Dutch spoken in the Netherlands, as others have done before us (cf. Geeraerts 2010).
13. There is a possible confound to the comparability, as Moroccorp was compiled about 15 years after ConDiv.
14. We are not aware of the existence of reliable taggers for Dutch chat material, which is exceptionally hard to annotate because of its fragmented nature. Here, the biggest advantage of chat data for linguists, i.e. its proximity to natural, spoken language, becomes a practical disadvantage.
15. For the analysis and visualisation, we made use of the MASS (Venables and Ripley, 2002), rms (Harrell, 2013), lme4 (Bates et al., 2013) and effects (Fox, 2003) packages.
16. Note that the fitted probabilities in the effect plots may slightly differ from the values based on the estimates in the model summary in Table 1 because the other variables are left 'as is' in the model, while the values in the effect plots adjust the value of the other variables to the mean (Buis, 2007).
17. Of course the AIC is raised as well, to a value of 1252, which is why this model was not selected by the variable selection procedure.
18. The ConDiv-only model has no random effects, as the attested phrase types are not identical in each of the corpora. This makes the small mean difference even more impressive.
19. We would like to thank one of the anonymous reviewers for pointing out this possibility.
20. Or they are exceedingly good at picking up and reproducing lexical preference patterns of L1 speakers. A vast body of CIA-research, however, shows that this is exactly where L2 language users struggle most (see Granger 2004: 132; Cosme 2008; Gries and Deshors 2014, among others).
21. An alternative explanation for the fact that Moroccorp differs from ConDiv in the expression of the adjectival schwa inflection, but does not differ in the expression of

the partitive -s genitive, could be sought in the 'stylisation' function of these grammatical markers. Overgeneralisation of the adjectival schwa is a known shibboleth of Dutch ethnolects, and can accordingly be used to (semi—)consciously index the ingroup speech, whereas partitive genitive *-s* is not. We think this is unlikely, as the findings in Van de Velde & Weerman (2014) show that there are very subtle effects in the refunctionalisation of the schwa in adnominal elements that go well beyond the (13c) condition on which the shibboleth hinges.

References

Steven Paul Abney, *The English noun phrase in its sentential aspect*, Cambridge: Dissertation Massachusetts Institute of Technology, 1987.

Vilmos Ágel and Mathilde Hennig, 'Überlegungen zur Theorie und Praxis des Nähe- und Distanzsprechens [Considerations on the theory and practice of near and distant speaking]', In: Vilmos Ágel and Mathilde Hennig (red.), *Zugänge zur Grammatik der gesprochenen Sprache*. Tübingen: Niemeyer, 2007, pp. 179-214.

Cynthia Allen, *Genitives in early English: typology and evidence*, Oxford: Oxford University Press, 2008.

Laurence Anthony, *AntConc* (*Computer Software, version 3.3.3*), Tokyo: Waseda University, 2011, http://www.antlab.sci.waseda.ac.jp/.

Johan van der Auwera, 'Wat het Nederlands ons kan leren over de Jespersencyclus [What Dutch can teach us about the Jespersen's cycle]', *Nederlandse taalkunde* 17 (2012), pp. 403-413.

Rolf Harald Baayen, *Analyzing linguistic data: a practical introduction to statistics using R*, Cambridge University Press, 2008.

Douglas Bates, Martin Maechler, Ben Bolker and Steven Walker, *lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0-4*, 2013, http://cran.r-project.org/package=lme4.

Christian Bentz, Douwe Kiela, Feli Hill and Paula Buttery, 'Zipf's law and the grammar of languages: A quantitative study of Old and Modern English parallel texts', *Corpus Linguistics And Linguistic Theory* 10, 2 (2014), pp. 175-211.

Christian Bentz and Bodo Winter, 'Languages with More Second Language Learners Tend to Lose Nominal Case', *Language Dynamics and Change* 3, 1 (2013), pp. 1-27.

Geert Booij, 'Inherent versus contextual inflection and the split morphology hypothesis', In: Geert Booij and Jaap van Marle (red.), *Yearbook of Morphology 1995*. Dordrecht: Kluwer, 1996, pp. 1-16.

Geert Booij, *Construction morphology*, Oxford: Oxford University Press, 2010a.

Geert Booij, 'Construction morphology', *Language and Linguistics Compass* 4, 7 (2010b), pp. 543-555.

Hans Broekhuis, *Syntax of Dutch: Adjectives and Adjective Phrases*, Amsterdam: Amsterdam University Press, 2013.

Hans Broekhuis and Anke Strang, 'De partitieve genitiefconstructie [The partitive genitive construction]', *Nederlandse taalkunde* 1, 3 (1996), pp. 221-238.

Maarten L Buis, 'Predict and adjust with logistic regression', *Stata Journal* 7, 2 (2007), pp. 221-226.

Joan Bybee, 'From Usage to Grammar: The Mind's Response to Repetition', *Language* 82, 4 (2006), pp. 711-733.

Joan Bybee and Clay Beckner, 'Usage-based theory', In: Bernd Heine and Heiko Narrog (red.), *The Oxford Handbook of Linguistic Analysis*. Oxford: Oxford University Press, 2010, pp. 827-855.

Leonie Cornips and Vincent de Rooij, 'Kijk, Levi's is een goeie merk: maar toch hadden ze 'm

gedist van je schoenen doen 'm niet', In: Jan Stroop (red.), *Waar gaat het Nederlands naartoe? Panorama van een taal*. Amsterdam: Bert Bakker, 2003, pp. 131-142.

Christelle Cosme, 'Participle clauses in learner English: the role of transfer', In: Gaëtanelle Gilquin, Szilvia Papp and Mara Belén Díez-Bedmar (red.), *Linking up Contrastive and Learner Corpus Research*. Amsterdam/Atlanta: Rodopi, 2008, pp. 177-200.

Laura van Eerten, 'Over het Corpus Gesproken Nederlands', *Nederlandse Taalkunde* 12, 3 (2007), pp. 194-215.

Hans-Olav Enger, 'Morphological theory and grammaticalisation: the role of meaning and local generalisations', *Language Sciences* 36 (2013), pp. 18.

John Fox, 'Effect displays in R for generalised linear models', *Journal of Statistical Software* 8 (2003), pp. 1-27.

John Fox and Sanford Weisberg, *An R companion to applied regression*, 2nd edn. Los Angeles: Sage, 2011.

Sylviane Granger, 'From CA to CIA and back: an integrated approach to computerized bilingual and learner corpora', In: Karin Aijmer, Bengt Alternberg and Mats Johansson (red.), *Languages in Contrast: Text-based Crosslinguistic Studies*. Lund: Lund University Press, 1996, pp. 37-51.

Sylviane Granger, 'Computer learner corpus research: current status and future prospects', In: Ulla Conner and Thomas A. Upton (red.), *Applied Corpus Linguistics: A Multidimensional Perspective*. Amsterdam: Rodopi, 2004, pp. 123-145.

Stefan Thomas Gries, *Statistics for linguistics with R. A practical introduction*, 2nd edn. Berlin: De Gruyter, 2013.

Stefan Thomas Gries and Sandra Deshors, 'Using regressions to explore deviations between corpus data and a standard/target: two suggestions', *Corpora* 9, 1 (2014), pp. 109-136.

Stefan Grondelaers, Katrien Deygers, Hilde Van Aken, Vicky Van den Heede and Dirk Speelman, 'Het CONDIV-corpus geschreven Nederlands [The CONDIV-corpus of written Dutch]', *Nederlandse Taalkunde* 5, 4 (2000), pp. 356-363.

Walter Haeseryn, Kirsten Romijn, Guido Geerts, Jaap de Rooij and Maarten van den Toorn, *Algemene Nederlandse Spraakkunst*, Groningen: Nijhoff, 1997.

Wayne Harbert, *The Germanic languages*, Cambridge: Cambridge university press, 2007.

Frank Harrell, 'rms: Regression Modeling Strategies. R package version 4.0-0', , 2013, http://cran.r-project.org/package=rms.

Jennifer Hendriks, 'Re-examining the "origins" of the prenominal periphrastic possessive construction Jan z'n boek in Dutch', *Diachronica* 29, 1 (2012), pp. 28-71.

Lars Hinrichs and Benedikt Szmrecsanyi, 'Recent changes in the function and frequency of Standard English genitive constructions: a multivariate analysis of tagged corpora', *English Language and Linguistics* 11, 3 (2007), pp. 437-474.

Jack Hoeksema, 'Adjectivale inflectie op -s: geen geval van transpositie [Adjectival inflection on -s: not a case of transposition]', In: Eric Hoekstra and Caroline Smits (red.), *Morfologiedagen 1996* [*Morphology Days 1996*]. Amsterdam: P. J. Meertens-Instituut, 1998a, pp. 46-72.

Jack Hoeksema, 'Een ondode kategorie: de genitief [An undead category: the genitive]', *Tabu* 28 (1998b), pp. 162-167.

Joop van der Horst, *Geschiedenis van de Nederlandse syntaxis* [*History of Dutch syntax*], Leuven: Universitaire Pers Leuven, 2008.

Joop van der Horst, *Taal op drift. Lange-termijnontwikkelingen in taal en samenleving* [*Language adrift. Long term developments in language and society*], Amsterdam: Meulenhoff, 2013.

Joop van der Horst and Kees van der Horst, *Geschiedenis van het Nederlands in de twintigste eeuw* [*History of Dutch in the twenthieth century*], Den Haag: Sdu, 1999.

Brian Joseph, 'Diachronic explanation: putting the speaker back into the picture', In: Garry W.

Davis and Gregory K. Iverson (red.), *Explanations in historical linguistics*. Amsterdam: John Benjamins, 1992, pp. 123-144.

Ellen-Petra Kester, 'The nature of adjectival inflection', Utrecht: LEd, 1996.

Peter Koch and Wulf Österreicher, 'Schriftlichkeit und Kommunikative Distanz [Writtenness and communicative distance]', *Zeitschrift fur Germanistische Linguistik* 35, 3 (2007), pp. 346-375.

Bernd Kortmann and Benedikt Szmrecsanyi, *Linguistic complexity: second language acquisition, indigenization, contact*, Berlin: De Gruyter, 2012.

Christiaan Wouter Kusters, *Linguistic complexity : the influence of social change on verbal inflection*, Utrecht: LOT, 2003.

Roger Lass, 'How to do things with junk: exaptation in language evolution', *Journal of linguistics* 26, 1 (1990), pp. 79-102.

Sterre Leufkens, *Transparency in language: A typological study.*, Utrecht: LOT, 2015.

Adolphe van Loey, *Middelnederlandse spraakkunst. Deel I. Vormleer*, 9th edn. Groningen: Wolters-Noordhoff, 1980.

Gary Lupyan and Rick Dale, 'Language structure is partly determined by social structure.', *PloS one* 5, 1 (2010), pp. e8559.

Jaap van Marle, 'The unity of Morphology: on the interwovenness of the derivational and inflectional dimension of the word', *Yearbook of Morphology 1995*. Dordrecht: Kluwer, 1996, pp. 67-82.

Yaron Matras, *Language Contact*, Cambridge: Cambridge University Press, 2009.

Muriel Norde and Freek Van de Velde, 'Exaptation: taking stock of a controversial notion in linguistics', In: Muriel Norde and Freek Van de Velde (red.), *Exaptation in language change*. Amsterdam: John Benjamins, 2016, pp. 1-35.

Dirk Pijpops and Freek Van de Velde, 'Constructional contamination. A case study.', *Folia Linguistica* (Forthcoming).

Dirk Pijpops and Freek Van de Velde, 'A multivariate analysis of the partitive genitive in Dutch. Bringing quantitative data into a theoretical discussion', *Corpus Linguistics and Linguistic Theory* (2014), Published online, ahead of print.

R Core Team, *R: A language and environment for statistical computing. R Foundation for Statistical Computing*, Vienna, 2013, http://www.r-project.org/.

Tom Ruette and Freek Van de Velde, 'Moroccorp: tien miljoen woorden uit twee Marokkaans-Nederlandse chatkanalen', *Lexikos* 23 (2013), pp. 475-556.

Moritz Schönfeld, *Historische grammatica van het Nederlands* [*Historical grammar of Dutch*], (Red.) Adolphe van Loey, 8th. edn. Zutphen: W.T. Thieme & Cie, 1970.

Hendrik Schultink, *De morfologische valentie van het ongelede adjectief in modern Nederlands*, Den Haag: Van Goor, 1962.

Alan Scott, 'The position of the genitive in present-day Dutch', *Word Structure* 4, 1 (2011), pp. 104-135.

Alan Scott, *The genitive case in Dutch and German: a study of morphosyntactic change in codified languages*, Leiden: Brill, 2014.

Ludwig Söll, *Gesprochenes und geschriebenes Französisch* [*Spoken and written Dutch*], Berlin: Schmid, 1974.

Dirk Speelman, 'Logistic regression: A confirmatory technique for comparisons in corpus linguistics', In: Dylan Glynn and Justyna A. Robinson (red.), *Corpus Methods for Semantics: Quantitative studies in polysemy and synonymy*. (Human Cognitive Processing [HCP]). Amsterdam: John Benjamins, 2014, pp. 487-533.

John Suler, 'The Online Disinhibition Effect', *CyberPsychology Behavior* 7, 3 (2004), pp. 321-326.

Taaladvies.net, 'Taaladvies.net', *Nederlandse Taalunie*, http://taaladvies.net/ (7 April, 2015).

VRT Taalnet, 'VRT Taalnet', http://www.vrt.be/taal/ (7 May, 2015).

Taaltelefoon.be, 'Taaltelefoon.be', *Vlaamse overheid*, http://www.taaltelefoon.be/ (7 April, 2015).

Peter Trudgill, *Sociolinguistic typology: social determinants of linguistic complexity*, Oxford: Oxford University Press, 2011.

Freek Van de Velde, *Iets taalkundig(s): een functioneel georiënteerde analyse van deflexie en de genitiefontwikkeling in het Nederlands* [*Something linguistic: a functionally oriented analysis of deflexion and the development of the genitive in Dutch*], Leuven: Univerisity of Leuven MA thesis, 2001.

Freek Van de Velde, *De nominale constituent. Structuur en geschiedenis* [*The noun phrase. Structure and history*], Leuven: Leuven University Press, 2009.

Freek Van de Velde and Joop van der Horst, 'Homoplasy in diachronic grammar', *Language Sciences* 36, 1 (2013), pp. 66-77.

Freek Van de Velde and Fred Weerman, 'The resilient nature of adjectival inflection in Dutch', In: Petra Sleeman, Freek Van de Velde and Harry Perridon (red.), *Adjectives in Germanic and Romance*. (Linguistik Aktuell/Linguistics Today). Amsterdam: John Benjamins, 2014, pp. 113-145.

William Venables and Brian Ripley, *Modern applied statistics with S*, 4th edn. New York: Springer, 2002.

Fred Weerman and Petra De Wit, 'The decline of the genitive in Dutch', *Linguistics* 37, 6 (1999).

About the authors

Dirk Pijpops, University of Leuven, RU Quantitative Lexicology and Variational Linguistics, FWO Flanders, Blijde-Inkomststraat 21, 3000 Leuven.
E-mail: dirk.pijpops@kuleuven.be

Freek Van de Velde, University of Leuven, RU Quantitative Lexicology and Variational Linguistics, Blijde-Inkomststraat 21, 3000 Leuven.
E-mail: freek.vandevelde@kuleuven.be

Colofon

taal & tongval
tijdschrift voor taalvariatie – language variation in the Low Countries
Jaargang 67-1, 2015
http://www.taalentongval.eu

Subsidiënten:

Universitaire Stichting, Brussel

taal & tongval verscheen voor het eerst in 1949 met als ondertitel *Tijdschrift voor de studie van de Nederlandse volks- en streektalen*. Het was van het begin af een Nederlands-Vlaamse onderneming. In 1977 werd de ondertitel gewijzigd in *Tijdschrift voor Dialectologie*, hoewel toen reeds het veld dat door het tijdschrift bestreken werd, ruimer was dan de klassieke dialectologie. Met name de sociolinguïstiek kwam er ook regelmatig aan bod naast allerlei aspecten van de historische variatie binnen het Nederlands. Vanaf 2001 heet *taal & tongval* dan ook *tijdschrift voor taalvariatie*. Daarmee wordt aangegeven dat het tijdschrift alle taalvariatieaspecten (ruimtelijke, sociale, pragmatische en historische) binnen het Nederlands taalgebied tot zijn publicatiedomein rekent.

Van *taal & tongval* verschijnen per jaar twee nummers.
Prijs per jaargang: particulier abonnement: 40 Euro; institutioneel abonnement: 60 Euro.

Wilt u op de hoogte blijven van het laatste nieuws, interessante nieuwe artikelen en boeken? Schrijf u dan nu in op de nieuwsbrief van AUP, via www.aup.nl.

Voor advertenties, abonnementen en ruilabonnementen wende men zich tot:

Amsterdam University Press
Herengracht 221
Telefoon: +31 (0)20 420 00 50
Fax: +31 (0)20 420 32 14
e-mail: abonnementen@aup.nl
ING Bank:
IBAN: NL98 INGB 0664 1034 21,
BIC: INGBNL2A

UGent, afd. Nederlandse Taalkunde
Blandijnberg 2
9000 Gent (België)
Telefoon: +32 (0)9 264 40 75
Fax: +32 (0)9 264 41 70
e-mail: valerie.bouckaert@ugent.be
Belfius bank:
IBAN: BE40 0682 1390 8063
BIC: GKCCBEBB