

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/76849>

Please be advised that this information was generated on 2020-09-28 and may be subject to change.

De rol van *en* en *s* in samenstellingen

Veranderend Nederlands verandert gedachten over het taalsysteem

Anneke Neijt & Robert Schreuder

Summary

For decades linking elements in compounds in both Dutch and English have been seen predominantly as unrelated to plural and other semantically loaded morphemes, despite their formal similarity. Nevertheless, experimental evidence shows that language users do attach plurality values to certain linking elements in Dutch. This suggests a different view of the relation between form and function, which in turn invalidates much of the popular words-and-rules theory. We propose a multi-tiered, interconnected language system in which elements can have several functions simultaneously. This yields, among other things, a quite natural explanation for the severe limitations on using plurals in English compounds. Furthermore, a tendency for linking elements to disappear from Dutch compounds seems to have come to a halt in recent years. We consider the possibility that this is an effect of changes in spelling rules which, if true, would establish spelling as a more or less fully fledged subsystem of language.

1. Inleiding¹

‘Panta rhei’, liet Plato Heraclitus al zeggen, en zo is het ook vandaag de dag bij de vorming van samenstellingen. Daar zijn ontwikkelingen gaande die voorheen ongetwijfeld juiste uitspraken op losse schroeven zetten. Dat geldt bijvoorbeeld voor Ariane van Santens uitspraken in haar leerboek van vijftientig jaar geleden en in de opvolger daarvan die ze samen met Geert Booij schreef:

Op het ogenblik is er, zoals Van den Toorn laat zien aan de hand van recente nieuwvormingen, een sterke tendens tot het weglaten van een bindfonem. (Van Santen 1984:62-63)

We concluderen dat de bindfonemen /ə/ en /s/ niet geïdentificeerd kunnen worden met de meervoudssuffixen *-en* en *-s*. (Booij & Van Santen 1998:158)

De eerste uitspraak lijkt op de voorspelling van Neijt (2007) dat er minder tussenklanken zullen worden gebruikt in het Nederlands aan het einde van deze eeuw. Het is niet eenvoudig om zulke uitspraken te toetsen. We doen desalniettemin een poging in een corpusstudie

1 Met dank aan de redactie en aan Arina Banga, Esther Hanssen, Frans van der Slik en Rik Smits voor commentaar op een eerdere versie, en aan Marije Boer en Kobie van Krieken voor hun hulp bij het inventariseren van de corpora. Dit artikel is tot stand gekomen mede dankzij steun van de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO), programma Numerosity 360-70-360.

(paragraaf 2). Het blijkt dat er op dit moment van een sterke tendens tot weglaten geen sprake meer is. De tweede uitspraak bespreken we en onderzoeken we in een experimentele studie. We gaan in op de betekenis van tussenklanken en hun theoretische status, met als conclusie dat de bindfonemen wel degelijk geïdentificeerd kunnen worden met de meer-voudssuffixen *-en* en *-s* (paragraaf 3 en 4). Tot slot bespreken we vanuit deze gegevens opnieuw de vraag wat er de oorzaak van kan zijn dat de tendens tot weglaten is verdwenen. Vermijdingsgedrag veroorzaakt door de spelling kan wel eens de oorzaak zijn geweest.

2. Corpusonderzoek van veranderingen

Om veranderingen in kaart te brengen zijn vergelijkbare corpora nodig van verschillende tijdspannes. Die corpora ontbreken op dit moment. We hebben wel enkele steekproeven kunnen nemen met behulp van het geavanceerde zoekstelsel van Google, dat de mogelijkheid biedt om het aantal vindplaatsen van websites die het laatste jaar aangepast zijn te scheiden van alle vindplaatsen van een bepaald woord. Dan vind je bijvoorbeeld dat het *en*-gebruik is toegenomen van 17 naar 29% in het paar *steekproefgemiddelde-steekproefgemiddelde* en van 74 naar 78% in het paar *panne(n)koekhuis-panne(n)koekhuis*. Binnen het paar *frisdrankindustrie-frisdrankenindustrie* is het gebruik van *en* daarentegen drastisch afgenomen: van 84 naar 43%. *Steekproefsgewijs* groeit weer een beetje ten opzichte van *steekproefsgewijs*, van 92 naar 94%, terwijl *werkwoordsvormen* – *werkwoordsvormen* gelijk blijft op 98% *s*-gebruik.

Wat voor bovengenoemde vijf samenstellingen geldt, kan natuurlijk bepaald niet worden geëxtrapoleerd naar een bewering die de schier onmetelijke hoeveelheid samenstellingen van het Nederlands betreft. Aan het aantal waarnemingen kunnen we niet tegemoetkomen in het bestek van dit artikel, maar aan de willekeur wel, door de veranderingen in het gebruik van *en* na te gaan in de veertig samenstellingen die met een geheel ander oogmerk in de experimentele studie van de volgende paragraaf gebruikt worden. Het blijkt dat het gemiddelde aandeel van de vormen met *en* gelijk is gebleven in het laatste jaar (62%). Geen tendens tot afname dus die zo sterk is dat die in een enkel jaar meetbaar is (de metingen vonden plaats in oktober en november 2008). Binnen de groep van veertig woorden die in deze studie gebruikt zijn, overtreft het aantal vormen dat een stijging in het gebruik van *en* laat zien (23) overigens het aantal vormen met een daling (14) ruimschoots. In drie van de veertig vormen is geen verandering aangetroffen. De tendens van weglaten blijkt dus geheel te zijn verdwenen (vergelijk tekentoets; het distributiepatroon laat geen significant verschil zien, noch de ene, noch de andere kant uit). Bijlage 1 bevat de precieze gegevens voor de lijst van veertig woorden.

Hoewel ten aanzien van aan Google ontleende gegevens de nodige reserve op zijn plaats is (zie bijlage 2), blijkt van een sterke tendens tot weglaten in de hier besproken voorbeelden geen sprake te zijn. De stelling van Van Santen (1984) lijkt dus achterhaald, en de uitspraak van Neijt (2007) kan wel eens onjuist blijken te zijn. Aan het eind van ons verhaal komen we erop terug. Eerst bespreken we een experiment dat laat zien dat de conclusie

van Booij en Van Santen (1998) dat *en* en *s* tussen de leden van samenstellingen niet geïdentificeerd kunnen worden met de meervoudssuffixen *-en* en *-s*, niet houdbaar is.

3. Het onderzoek naar de betekenis van *en* in samenstellingen

Eerder gepubliceerd onderzoek waarin een *n* en *s* gemanipuleerd werd, wees al uit dat taalgebruikers *en* en *s* binnen samenstellingen als meervoudsuitgangen interpreteren (Schreuder e.a. 1998 en Neijt e.a. 2006). Hier doen we verslag van nog lopend onderzoek waaruit blijkt dat wat voor de manipulatie van *n* en *s* geldt, ook opgaat voor *en*, een letterpaar dat met een gehele syllabe in de uitspraak correspondeert (Neijt & Schreuder, in voorbereiding).

Draagt *en* in samenstellingen een meervoudsbetekenis? Bij vormparen zoals *celstructuur* en *cellenstructuur*, of *bankwereld* en *bankenwereld* lijkt dat evident, omdat de vormen lichtelijk verschillen van betekenis. Het woord *celstructuur* duidt eerder op de structuur binnen een cel, *cellenstructuur* op een structuur die bestaat uit cellen. *Bankwereld* verwijst naar de wereld van de bank in zijn algemeenheid, maar met *bankenwereld* benadrukt de taalgebruiker dat er veel banken in het spel zijn. Zulke semantisch interessante paren zijn voor ons doel ongeschikt. Wij keken uitsluitend naar vormparen die ogenschijnlijk dezelfde semantische waarde hebben, zoals *naamval(len)systeem*. De vraag is of ook die een meervoudigheidsverschil laten zien, hoewel daar conceptueel geen reden voor is.

Op grond van de resultaten van de experimenten in Schreuder e.a. (1998) zouden we verwachten inderdaad een verschil te vinden. Daar ging het om vormen met of zonder *n*, zoals *bloemenbak* en *bloembak*. Voor veel Nederlanders hebben die dezelfde klankwaarde, zodat het 'slechts' om een orthografisch ingrijpen ging. Het hier beschreven experiment is anders, omdat we de semantische waarde van *en* onderzoeken, een letterpaar dat voor alle taalgebruikers samengaat met een klankverschil. *Naamvallensysteem* en *bloemenbak* hebben nu eenmaal een syllabe meer dan *naamval(s)systeem* en *bloembak*. Omdat alle varianten evenwel gebruikt lijken te worden zonder dat er een opmerkelijk semantisch verschil mee bedoeld wordt, zou ons vermoeden onjuist kunnen zijn.

We legden de vraag *hoe meervoudig is het eerste deel van de samenstelling?* in werkcolleges voor aan 27 beginnende eerstejaarsstudenten Nederlandse taal en cultuur. De studenten waren moedertaalsprekers van het Nederlands, gemiddeld 19 jaar oud. Als materiaal fungeerden veertig samenstellingen met een variabele tussenklank, zoals *adres(en)lijst* en *bloem(en)bak*. De samenstellingen drukken zonder en met *en* hetzelfde concept uit (zie bijlage 1 voor de lijst testwoorden).

We legden het begrip meervoudigheid van het eerste deel van een samenstelling uit aan de hand van voorbeelden zoals *mannenstem*, waarin *mannen* aan een enkelvoud doet denken, tegenover *kleuterklas*, waarin *kleuter* meer aan een meervoud doet denken. Er werd gevraagd om zulke oordelen over enkelvoud en meervoud weer te geven op een zevenpuntsschaal, met links enkelvoud, te gebruiken wanneer het eerste deel alleen maar als een enkelvoud kan worden begrepen en de waarden rechts daarvan voor een oplopende schaal van meer-

voudigheid. Omdat de proefpersonen vormen met en zonder *en* in een willekeurig gevarieerde lijst met samenstellingen te zien kregen, werden ze ervoor gewaarschuwd dat er in de lijst woorden zouden kunnen staan die ze zelf niet zo zouden schrijven. 'Denk daar niet te lang over na', luidde de instructie, want 'Het gaat om de *betekenis*, niet om hoe het woord geschreven is. Laat u vooral leiden door wat het eerst in u opkomt.'

Er waren twee spiegelbeeldig samengestelde lijsten, waarin iedere proefpersoon zowel vormen met als zonder *en* te lezen kreeg, maar van elke samenstelling steeds slechts een van beide varianten. Van elke lijst werden nog eens vier willekeurige volgordes gemaakt, om volgorde-effecten te voorkomen. De instructie bij het experiment is opgenomen in bijlage 3.

De zo bevraagde proefpersonen bleken het eerste lid van de samenstellingen met *en* inderdaad meervoudiger te vinden dan zonder (bijlage 1 bevat per woord de gemiddelde waarden). Het verschil van ruim 1 punt op de zevenpuntsschaal is significant (gepaarde t-toetsen over de proefpersoongemiddelden, $t(26) = 4.08$, $p < .001$ en over de itemgemiddelden, $t(39) = 11.13$, $p < .0001$). Onze conclusie is, dat *en* in samenstellingen als meervouds-suffix geïnterpreteerd wordt.

Gegeven deze uitkomst en de eerdere bevindingen ten aanzien van de *s*, kunnen we ons als taalkundigen slechts neerleggen bij het feit dat taalgebruikers *s* en *en* in samenstellingen als meervoudsaanduiders interpreteren. Dat haalt een van de belangrijkste pijlers onderuit van de words-and-rules-theorie van Pinker (1999), recentelijk nog verdedigd in Berent & Pinker (2007). Opslag en regelgebruik zijn niet strikt gescheiden op grond van taalkundige categorieën maar lopen elegant door elkaar heen al naar gelang het taalgebruik dat vereist. Opslag is geëigend voor wat vaak voorkomt, en regelgebruik voor de minder vaak gebruikte vormen. Voorts kan de term bindfoneem, die al in 1998 door ons op de schop is genomen (Schreuder e.a. 1998), definitief op de schroothoop. Het bindelement, dat interessanter genoeg semantische lading met ritmische conditionering en analogiewerking combineert moet voortaan gewoon een morfeem genoemd worden.

4. De words-and-rules-theorie

In de wereld van de morfologie is de vorming van woorden en samenstellingen op basis van combinatorische regels bekend geworden als de words-and-rules-theorie van Pinker (1999). Deze theorie, die voortbouwt op de level-orderingshypothese van Kiparsky (1982), stelt dat regelmatige flexie de expressie vormt van het regelsysteem, en dat uitsluitend onregelmatige vormen uit opslag betrokken worden. Met Berent zette Pinker recentelijk die visie op de morfologie nog eens grondig uiteen aan de hand van de vorming van samenstellingen (Berent & Pinker 2007). Een niet op grammaticale categorieën, maar op statistische patronen gebaseerde connectionistische benadering zoals die van Rumelhart en McClelland (1986), Haskell, MacDonald en Seidenberg (2003) of Seidenberg, MacDonald en Haskel

(2007) vindt in hun ogen geen genade. Zij prefereren een taalkundige benadering die uitgaat van grammaticale categorieën met productieve combinatorische regels.

Engelse samenstellingen spelen een voorname rol bij het debat over de verdiensten van een scheiding van opslag en regels die precies samenvalt met de scheiding tussen de morfologie (of het lexicon) en de syntaxis, omdat het regelmatige meervoud er slechts zelden in gebruikt wordt. In de oorspronkelijke words-and-rules-optiek volgt dat uit de verdeling van taken tussen het regelsysteem en de opslag van vormen. Het onregelmatige *mice* 'muizen' ligt in het lexicon opgeslagen en kan dus als lexicale bouwsteen fungeren, wat samenstellingen als *mice-infested* 'van muizen vergeven' oplevert. Dat **rats-infested* daarentegen niet voorkomt, zou eraan liggen dat regelmatige meervouden niet beschikbaar zijn op het moment dat samenstellingen gevormd worden.

Dat traditionele model met level-ordering en regelordering als centrale sturende principes wordt in Berent & Pinker (2007) evenwel losgelaten. Liever verklaren ze het geringe voorkomen van regelmatige meervoudsvormen in Engelse samenstellingen op grond van de status van het eerste deel van samenstellingen. Is dat niet een woord, maar een woordgroep, dan komen meervoudsvormen wel degelijk voor, getuige gevallen als *seat-of-the-pants-reasoning* (Lieber 1980) en *red rats eater* (Alegre & Gordon 1996), een vorm die volgens Haskell e.a. (2003:124) aanvaardbaarder is dan *rats eater*. Dat soms toch regelmatig gevormde meervoudige woorden als linkerdeel lijken op te treden, bijvoorbeeld in *publications-catalog*, komt, stellen Berent en Pinker, doordat woordgroepen ook uit een enkel woord kunnen bestaan.

Ter onderbouwing van deze aanpak, die meteen de vraag oproept waarom voor *rats eater* (een niet toegelaten vorm) niet dezelfde weg openligt, verwijzen Berent en Pinker naar Alegre & Gordon (1996), een artikel dat beoogt het belang van analogie in natuurlijke talen te minimaliseren. Daarin wordt gewezen op het onderscheid tussen 'veel individuele ratten' bij *rats-eater* en 'verschillende soorten publicaties' bij *publications-catalog*. Het lijkt ons allemaal veel te ver gezocht. Met enig voorbehoud omdat onze intuïties zouden kunnen verschillen van die van moedertaalsprekers van het Engels: met een Nederlandse pet op gaat een samenstelling zoals *flowers-fantasy* gewoon over bloemen in het meervoud, of het nu een stemmige bos blanke lelies betreft of een woest gemengd boeket.

5. Meerlagigheid en een andere verklaring voor de onwelgevallige meervouden in het Engels

'Dutch disallows the regular -s plural in compounds, but allows the -s suffix as a linking morpheme or diminutive' [sic] beweren Berent en Pinker aan het eind van hun artikel (Berent & Pinker 2007:165). Maar ons onderzoek wijst uit dat voor de taalgebruiker zowel *s* als *en* een meervoudsaanduiding is (voor de *s*, zie Neijt e.a. 2006; voor *en* zie Schreuder e.a. 1998, en het experiment van paragraaf 3). Waarom je dan in bochten wringen om te bewijzen dat het tegendeel waar is? Het is alleszins aannemelijk dat samenstellingen in het

Nederlands op grond van meervouden gevormd mogen worden. Op zijn minst laten onregelmatige meervouden zien dat meervouden binnen samenstellingen net als in het Engels niet principieel uitgesloten zijn. Parallel aan *mice-infested* heeft het Nederlands immers vormen als *goederentrein*, *neerlandicongres* en vele andere.

En er is meer. *Dames* in *damestasje* roept onmiskenbaar een meervoud op. Nu zijn er weliswaar samenstellingen als *zinstype*, waarin de *s* als scheider optreedt (Van Lessen 1928; Krott e.a. 2004), maar dat die *s* ook een meervoudsbetekenis heeft wanneer de fonologische context dat toelaat, hoeft niet te verbazen. Dezelfde *s* die als scheider of als genitief dienst doet, fungeert blijkbaar tevens als meervoud.

Het bindmorfeem *en* heeft eveneens meerdere, heel verschillende taken. Het element kan, zoals hierboven besproken is, ingezet worden om een (soms heel subtiel) verschil in meervoudsbetekenis aan te geven, maar het element kan tevens om ritmische redenen gebruikt worden (Neijt e.a. 2002 en Neijt & Schreuder 2007). Dat effect van ritme is er eveneens wanneer de werking van analogie, een van de belangrijkste sturende principes van woordvorming (Krott e.a. 2001), is uitgesloten. Dat laat het experimenteel onderzochte paar *bloem(en)bak* en *bloem(en)bakontwerp* zien. In *bloem(en)bak* is er een grotere voorkeur voor de vorm met *en* dan in *bloem(en)bakontwerp*, omdat de ritmische structuur van *bloembak* onwelgevormd is. Met alleen een maatslag op *bloem*, *bloembāk*, mist de vorm de noodzakelijke herhaling van maatslagen; met een tweede maatslag op *bak*, *bloembāk*, is er een botsing van maatslagen die liever vermeden wordt. De samenstelling *bloembāköntwerp* daarentegen is ritmisch welgevormd, evenals de variant met *en*, vergelijk *bloembāköntwerp*. In dit geval is er geen voorkeur voor een van de varianten. Dat er zo verschillende en uiteenlopende functies in een enkel morfeem zijn verenigd als meervoud en ritme is misschien bijzonder, maar niet ongeloofwaardig. Zo'n combinatie van taken wordt al in de *Algemene Nederlandse spraakkunst* (Haeseryn e.a. 1997:683) vermeld.

Met dat alles is natuurlijk niet zomaar gezegd dat vormen die op elkaar lijken ongelimiteerd elkaars betekenissen overnemen. Maar wel moeten we concluderen dat er niet zo'n heel sterke een-op-eenrelatie is tussen vorm en betekenis in natuurlijke talen. Sommige vormen hebben meer functies tegelijkertijd. In het Nederlands verenigt *s* de functies van scheider, meervoud en misschien een beetje genitief. Het bindmorfeem *en* is misschien ook een scheider, maar in ieder geval zowel een meervoud als een ritmisch morfeem. Zo beschouwd is het taalsysteem een meerlagige structuur met een soepele relatie tussen vorm en betekenis. Toevallige vormovereenkomsten brengen betekenissen met zich mee. Dat kan in lijn zijn met wat de spreker bedoelt, of het kan hinderlijk zijn. Hindernissen hebben waarschijnlijk invloed op de productie van woorden.

Vergelijken we het Engels en het Nederlands vanuit het meerlagige perspectief, dan ontstaat een heel ander plaatje dan geschetst in Berent & Pinker (2007). Mogelijk verhindert in het Engels het genitieve karakter van de *s* dat woorden zoals *flowers door* 'bloemendeur' gevormd worden. *Flowers door* doet denken aan het conceptueel onwaarschijnlijke *flower's door*

‘deur van de bloem’, en dan kan maar beter *flower door* gezegd worden. Voor het Nederlands geldt die hindernis niet. De tussenklank *s* heeft nauwelijks nog een genitieve kleuring, die functie vinden we eigenlijk alleen nog maar bij eigennamen terug.

De Nederlandse *s* lijkt daarentegen, net als *en*, op een meervoudsuffix. Een onbedoeld meervoudige interpretatie lijkt echter minder hinderlijk dan een onbedoelde interpretatie als genitief. Sterker nog, een meervoudsvorm kan binnen een samenstelling als neutraal opgevat worden, zoals een meervoudsvorm in zinnen ook neutraal kan zijn (Mattens 1970). De stellingen *banken doen zoiets niet* en *een bank doet zoiets niet* hebben ongeveer dezelfde betekenis. Zowel enkelvouden als meervouden worden in bepaalde contexten generiek of anumeriek gebruikt. Zo beschouwd is het helemaal niet vreemd dat er variatie is in paren van het type *adreslijst* en *adressenlijst* met vrijwel dezelfde betekenis. Over deze vormvariatie hoeven we niet moeilijker te doen dan over *steekproefgewijs* naast *steekproefsgewijs*.

6. Tot slot. De sterke tendens tot het weglaten van een bindmorfeem

Dat, in weerwil van wat Booij en Van Santen beweerden, *e(n)* en *s* tussen de leden van een samenstelling door taalgebruikers wel degelijk geïdentificeerd worden met de gelijkkluidende meervoudsuffixen, is onzes inziens voldoende aangetoond. En daarmee vervalt meteen de zin van het spreken van bindfonemen /ə/ en /s/, zoals Booij en Van Santen doen. Beide elementen gedragen zich immers als morfemen. Raadselachtig blijft echter nog de stelling van Van Santen uit 1984 en de voorspelling van Neijt (2007). Waarom zou er destijds sprake zijn geweest van een sterke tendens tot het weglaten van bindmorfemen? En waarom zou die tendens dan, zoals we zagen, wel eens verdwenen kunnen zijn?

De enige verklaring die we kunnen bedenken is de spellingwijziging van het Nederlands in 1995. Voordien stuurde de semantiek de keuze tussen *e* en *en* (schrijf *bessesap* want dat kan het sap van een enkele bes zijn, maar schrijf *bessenjam* want van een enkele bes kun je geen jam maken). Daarna heersten fonologie en morfologie: schrijf alleen een *s* of *en* als je die hoort, en alleen *en* wanneer het om een woord gaat dat buiten de samenstelling ook alleen maar met *en* wordt geschreven. (Deze korte formulering geeft de hoofdregel weer; voor woorden als *huilebalk* en *zikenauto* gelden andere regels, zie Neijt en Kroon (aangeboden).) Hindert de huidige spelling de gebruiker misschien minder bij het gebruik van bindmorfemen? De huidige spelling is in overeenstemming met het morfologische principe van gelijkvormigheid, dat abstraheert van allerlei fonologische details, inclusief het ritme. Ook uitsluitend ritmisch bedoelde bindmorfemen schrijf je volgens de huidige spelling met *en* als het linkerlid een meervoud op *en* heeft. Een onbedoelde meervoudsbetekenis kun je daarvoor niet ontlopen, maar dat nadeel kan wel eens in het niet vallen bij de vormstabiliteit die het morfologische principe oplevert. *Naamvallesysteem*, *steekproefgemiddelde* en *teksteschrijver* zijn misschien minder aanvaardbare vormen, omdat buiten de samenstelling alleen *naamval-*
len, *steekproeven* en *teksten* geschreven mag worden.

Als stelling durven we ze niet te poneren, maar als hypothese gebaseerd op de gedachte dat taal een meerlagig systeem is, vinden we de volgende twee uitspraken interessant:

- (a) De tendens tot vermindering van bindvormen in vroeger tijden is het gevolg geweest van vermijdingsgedrag op grond van een spellingprobleem.
- (b) Het bindmorfeem *s* in het Engels wordt vermeden omdat de genitiefinterpretatie in de weg zit.

Beide hypothesen gaan uit van vermijdingsgedrag als een van de sturende principes in natuurlijke talen. In het Nederlands zitten de vormen elkaar in de weg, in het Engels zitten de betekenissen elkaar in de weg. Het interessante van hypothese (a) is dat zou kunnen blijken dat de spelling als een van de lagen van het taalsysteem moet worden beschouwd. Zou ze juist blijken te zijn, dan heeft interferentie tussen gesproken en geschreven taal immers soortgelijke effecten als interferentie binnen de gesproken taal.

Bibliografie

- Alegre, M. & P. Gordon (1996). 'Red rats eater exposes recursion in children's word formation'. In: *Cognition*, 60, 65-82.
- Berent, I. & S. Pinker (2007). 'The dislike of regular plurals in compounds. Phonological familiarity or morphological constraint?' In: *The Mental Lexicon*, 2, 129-181.
- Booij, G. & A. van Santen (1998). *Morfologie. De woordstructuur van het Nederlands*. 2^e, geh. herz. en uitgebr. dr. Amsterdam: Amsterdam University Press.
- Haeseryn, W. [e.a.] (red.) (1997). *Algemene Nederlandse spraakkunst*. 2^e, geh. herz. dr. Groningen [etc.]: Martinus Nijhoff [etc.]. 2 dln.
- Haskell, T.R., M.C. MacDonald & M.S. Seidenberg (2003). 'Language learning and innateness. Some implications of *compounds research*'. In: *Cognitive Psychology*, 4, 119-163.
- Kiparsky, P. (1982). 'Lexical morphology and phonology'. In: I.S. Yang (ed.). *Linguistics in the morning calm*. Seoul: Hansin, 3-91.
- Krott, A. [e.a.] (2004). 'Probability in the grammar of German and Dutch: interfixation in triconstituent compounds'. In: *Language and Speech*, 47, 83-106.
- Krott, A., R. Schreuder & R.H. Baayen (2001). 'Analogy in morphology: modeling the choice of linking morphemes in Dutch'. In: *Linguistics*, 39, 51-93.
- Lessen, J.H. van (1928). *Samengestelde naamwoorden in het Nederlandsch*. Groningen [etc.]: Wolters. Diss. Groningen.
- Lieber, R. (1980). *The organization of the lexicon*. Cambridge MA: MIT Press.
- Mattens, W.H.M. (1970). *De indifferentialis. Een onderzoek naar het anumerieke gebruik van het substantief in het Algemeen Bruikbaar Nederlands*. Assen: Van Gorcum [etc.].

- Neijt, A. (2007). 'Meer variatie en minder tussenklanken door veranderde spellingregels'. In: Genootschap Onze Taal. *De toekomst van het Nederlands. Onze taal over 75 jaar*. 's-Hertogenbosch: Adr. Heinen, 47-48.
- Neijt, A., H. Baayen & R. Schreuder (2006). 'Formal and semantic constraints on the interpretation of the suffix -s in Dutch nominal compounds'. In: *Written Language and Literacy*, 9, 247-264.
- Neijt, A., L. Krebbers & P. Fikkert (2002). 'Rhythm and semantics in the selection of linking elements'. In: H. Broekhuis & P. Fikkert (eds.). *Linguistics in the Netherlands*. Amsterdam [etc.]: John Benjamins, 117-127. (AVT publications 19).
- Neijt, A. & S. Kroon (aangeboden). 'Onderwijs in de spellingregels voor de tussenklanken'.
- Neijt, A. & R. Schreuder (2007). 'Rhythm versus analogy. Prosodic form variation in Dutch compounds'. In: *Language and Speech*, 50, 533-566.
- Neijt, A. & R. Schreuder (in voorbereiding). 'Rhythm versus numerosity. The balance of prosody and semantics'.
- Neijt, A., R. Schreuder & H.R. Baayen (2004). 'Seven years later. The effect of spelling on interpretation'. In: L. Cornips & J. Doetjes (eds.). *Linguistics in the Netherlands 2004*. Amsterdam [etc.]: John Benjamins, 134-145. (AVT publications 21).
- Pinker, S. (1999). *Words and rules. The ingredients of language*. New York: Basic Books.
- Rumelhart, D.E. & J.L. McClelland (1986). 'On learning the past tense of English verbs. Implicit rules or parallel distributed processing?' In: J.L. McClelland, D.E. Rumelhart & T.P.R. Group (eds.). *Parallel distributed processing. Explorations in the microstructure of cognition. Vol. 2. Psychological and biological models*. Cambridge MA: MIT Press, 210-271. (A Bradford book).
- Santen, A. van (1984). *De morfologie van het Nederlands*. Dordrecht: Foris.
- Schreuder, R. [e.a.] (1998). 'Regular plurals in Dutch compounds: linking graphemes of morphemes?' In: *Language and Cognitive Processes*, 13, 551-573.
- Seidenberg, M.S., M.C. MacDonald & T.R. Haskell (2007). 'Semantics and phonology constrain compound formation'. In: *The Mental Lexicon*, 2, 287-312.
- Toorn, M.C. van den (1984). *Nederlandse grammatica*. 9^e dr. Groningen: Wolters-Noordhoff. [1^e dr. 1973.]

Bijlage 1. De woorden van de corpusstudie en van het experimentele onderzoek

Google biedt de mogelijkheid om de vormvariatie binnen samenstellingen te onderzoeken. We hebben dat middel ingezet in het corpusonderzoek om te bepalen of er een sterke tendens tot weglaten van *e* en *en* is. Voor het experimentele onderzoek gebruikten we de tellingen om na te gaan of de vormen inderdaad variatie lieten zien. De tabel op deze en de volgende bladzijde bevat de verzameling van veertig samenstellingen die in het corpusonderzoek van paragraaf 2 en in het experimentele onderzoek van paragraaf 3 zijn opgenomen.

De tabel is van boven naar beneden geordend op de kolom *verschil*, zodat duidelijk wordt dat in 23 woorden een toename in het gebruik van *e(n)* gevonden is, en in 14 woorden een afname van *e(n)*. De tabel bevat van links naar rechts:

- de samenstelling, waarbinnen schuingedrukt de tussenklank die onderzocht is
- het *aantal* vindplaatsen in Google (de vorm met *e* of *en* opgeteld bij de vorm zonder bindmorfeem)
- het aantal *oude* vindplaatsen (van alle vindplaatsen zijn de *nieuwe* vindplaatsen met als tijdsbeperking het laatste jaar afgetrokken)
- het aantal *nieuwe* vindplaatsen (dat aantal wordt gevonden wanneer de zoekactie wordt beperkt tot websites die het afgelopen jaar zijn bijgewerkt)
- de *percentages* van de samenstellingen met *e* of *en* als bindmorfeem, uitgesplitst voor de *oude* en *nieuwe* vindplaatsen, en het *verschil* tussen oud en nieuw, met een plus bij groei van het percentage bindmorfemen en een min bij afname van dat percentage
- het gemiddelde van de *meervoudsoordelen* die de proefpersonen gaven
- uitgesplitst voor de samenstellingen die geschreven zijn *met* en de samenstellingen die geschreven zijn *zonder* het bindmorfeem.

Het gaat om oordelen op een zevenpuntsschaal, waarbij de uitersten enkelvoud en meervoud met respectievelijk 1 en 7 zijn gecodeerd. De instructie aan de proefpersonen was om vanuit het enkelvoud een oordeel te vormen (bijlage 3). Het eerste rondje inkleuren betekent dat alleen een enkelvoudsinterpretatie mogelijk is. Het gaat dus om een asymmetrisch gebruik van de zevenpuntsschaal, en het is niet zo dat een gemiddeld oordeel van 4 begrepen moet worden als een keuze precies tussen enkelvoud en meervoud in.

samenstelling	aantallen		percentages <i>e(n)</i>			meervoudsoordelen	
	oud	nieuw	oud	nieuw	verschil	met <i>en</i>	zonder <i>en</i>
zwaluwenei	76	2	18%	50%	+32%	2,38	1,86
dierengedrag	14.549	1892	33%	47%	+14%	5,38	3,93
sardienblik	265	43	74%	86%	+12%	4,46	3,36
pannenkoekenmix	1.256	178	64%	76%	+11%	4,62	3,71
handdoekenhanger	478	51	21%	29%	+8%	3,85	2,43
talenstudie	17.979	2631	60%	68%	+8%	2,85	1,79
artiestemaam	77.129	9156	82%	90%	+8%	3,36	1,15
huisdierenvoer	932	84	83%	89%	+6%	4,46	3,08
klassenonderwijzer	286	24	70%	75%	+5%	3,79	2,15
artsendiploma	2.942	388	54%	59%	+5%	2,69	1,29

samenstelling	aantallen		percentages <i>e(n)</i>			meervoudsoordelen	
	oud	nieuw	oud	nieuw	verschil	met <i>en</i>	zonder <i>en</i>
aardbeieveld	528	41	93%	98%	+4%	6,00	4,85
vissenkorn	39.274	4446	81%	85%	+4%	2,85	2,57
cellentekort	5.658	515	92%	95%	+4%	5,69	2,64
pakketdienst	41.598	4452	3%	6%	+3%	6,00	4,38
lampenolie	16.428	1447	53%	55%	+3%	3,62	3,14
verzenbundel	1.188	125	97%	99%	+3%	5,77	4,07
antwoordenblad	9.347	778	32%	34%	+2%	5,57	3,92
schilderijencollectie	2.784	402	97%	99%	+2%	6,15	5,36
manestraal	14.658	2582	98%	99%	+1%	1,50	1,31
bloemenbak	42.009	4200	7%	8%	+1%	3,93	3,00
brillenglas	5.100	575	85%	86%	+1%	4,21	1,85
pannendecksel	7.178	675	57%	58%	+1%	3,00	1,31
olifanten huid	8.748	922	99%	100%	+1%	3,00	1,77
doelpuntmakers	69.837	4926	98%	98%	0%	4,62	4,00
vakkenpakket	143.641	16401	100%	100%	0%	5,43	3,46
druktoetsentelefoon	612	43	7%	7%	0%	3,62	3,57
aardappelenoogst	11.631	2953	1%	0%	-1%	6,50	5,92
boekenlegger	54.929	7601	97%	96%	-1%	2,69	3,00
handtekeningstempel	205	41	4%	2%	-1%	2,93	1,92
perziken huid	2.037	126	7%	6%	-2%	3,79	1,62
werktuigenberging	5.763	764	92%	90%	-2%	5,71	3,46
adressenlijst	886.721	158101	94%	90%	-4%	6,07	4,92
olijvenpit	142	9	26%	22%	-4%	2,08	1,50
papegaaienbek	2.678	204	87%	82%	-5%	2,14	1,46
fietsenrek	170.655	14213	93%	87%	-6%	5,92	4,00
kamelendrijver	1.574	183	65%	57%	-8%	4,69	3,21
brievenpost	37.937	4368	31%	22%	-9%	4,46	3,69
dierenfiguur	3.800	224	49%	26%	-23%	3,08	1,86
gordijnenmarkt	119	6	46%	17%	-30%	6,14	4,46
frisdrankindustrie	11.514	326	84%	43%	-41%	5,86	3,62

Bijlage 2. Onderzoek op het internet, via Google

Gegevens verzameld met behulp van Google moeten met de nodige voorzichtigheid gebruikt worden, zie de waarschuwende opmerkingen van David Beaver en Geoff Nunberg op Language Log van januari 2005. Alleen wanneer het aantal vindplaatsen laag is, lijkt Google echte tellingen te geven, waarbij dan echter de vindplaatsen die erg op elkaar lijken, zijn weggelaten. Deze weggelaten vindplaatsen zijn in onze tellingen ook niet opgenomen, omdat het meestal gaat om het taalgebruik van een enkele bron. Uit voorzorg hebben we overigens de woorden tussen aanhalingstekens ingevoerd, zodat precies op de gewenste woordvorm gezocht zou worden, en de gesuggereerde varianten niet in de tellingen meegenomen zouden worden.

Aanvankelijk wilden we onze zoektocht beperken tot samenstellingen uit het taalkundige jargon. Bij het zoeken naar de taalkundige termen zoals *woordgroep(en)leer* en *woordgroep(s)leer* bleek er echter een storende invloed te zijn van de opbouw van het Googlecorpus in een kleine taalgemeenschap zoals het Nederlands. Omdat in de DNBL, de digitale bibliotheek voor de Nederlandse letteren, bepaalde teksten wel zijn opgenomen en andere niet, kan het zijn dat Google het gebruik van een enkele auteur weergeeft. Dat bleek het geval te zijn bij *woordgroepsleer* (6-2), *woordgroepenleer* (9-0) en *woordgroepleer* (5-2) (oud-nieuw, dat wil zeggen aantallen vormen in alles behalve het laatste jaar, en alleen het laatste jaar). De meeste vindplaatsen van het laatste jaar komen uit de *Nederlandse grammatica* van Van den Toorn (1984), een leerboek dat recent is opgenomen in de DNBL. Voor diachroon morfologisch corpusonderzoek van neerlandistische termen is Google dus zeker niet geschikt.

Bijlage 3. De letterlijke instructie aan de proefpersonen van het experiment in paragraaf 3

Het eerste deel van een samenstelling heeft soms een meervoudsbetekenis, soms een enkelvoudsbetekenis. Bijvoorbeeld: meervoud voor *kleuter* in *kleuterklas*, want het gaat om veel kleuters, en enkelvoud voor *mannen* in *mannenstem*, want het gaat om de stem van één man. Op de achterkant van dit blaadje staat een woordenlijst. In deze lijst geeft u aan hoe meervoudig u het eerste deel van de samenstelling vindt. Bijvoorbeeld:

mannenstem	enkelvoud	●	○	○	○	○	○	○	meervoud
kleuterklas	enkelvoud	○	○	○	○	○	●	○	meervoud
mierenhoop	enkelvoud	○	○	○	○	○	○	●	meervoud
sinaasappelschil	enkelvoud	○	○	●	○	○	○	○	meervoud
appelsap	enkelvoud	○	○	○	●	○	○	○	meervoud
merelnest	enkelvoud	○	●	○	○	○	○	○	meervoud

Wanneer u het eerste deel alleen maar als een enkelvoud kunt interpreteren, dan kleurt u het eerste rondje in. Kan het eerste deel ook een meervoudsbetekenis hebben, dan kiest u een van de rondjes rechts daarvan. Hoe meervoudiger u het eerste deel vindt, hoe meer naar rechts u het rondje inkleurt.

Er kunnen woorden in de lijst staan die u zelf niet zo zou schrijven. Denk daar niet te lang over na. Het gaat om de *betekenis*, niet om hoe het woord geschreven is. Laat u vooral leiden door wat het eerst in u opkomt.