

Woordvolgordevariatie: theorie versus empirie?

Helen de Hoop

NT 21 (2): 275–284

DOI: 10.5117/NEDTAA2016.2.HOOP

Abstract

Word order variation: theory vs. empiricism?

In their chapter on word order variation in the middle field (*scrambling*), Broekhuis & Corver (2016) present a set of syntactic rules (generalizations) to derive word order patterns in Dutch. I will argue that their theoretical assumptions lead to the rejection of some sentences as being ungrammatical, while in my view they are perfectly grammatical. I conclude that it would be better to base grammaticality judgments on empirical data rather than on theoretical assumptions.

Keywords: scrambling, Dutch, intonation, grammaticality judgments, empirical approach

1 Introductie

Taalwetenschappers bepalen niet wat grammaticaal is en wat niet.¹ Zij schrijven niet aan de taalgemeenschap voor wat acceptabele zinnen zijn en wat onacceptabele. Zoals gebruikelijk binnen een empirische wetenschap, observeren zij slechts wat er is (en wat er gebeurt) en proberen zij wat ze waarnemen zo goed mogelijk te beschrijven. Als ze een stapje verdergaan, kunnen ze proberen het beschrevene ook te verklaren.

Lees de volgende drie dialogen eens hardop voor:

¹ Graag wil ik Sebastian Collin, Ad Foolen, Gijs Mulder, Jimmy van Rijt en Joost Swarts bedanken voor nuttig commentaar op de eerste versie van dit artikel.

- (1) Spreker A: Ik heb het aan Peter verteld.
Spreker B: Dan heb je de verkeerde waarschijnlijk ingelicht.
- (2) Spreker A: Waarom is Jan zo kwaad over dat papiertje?
Spreker B: Omdat waarschijnlijk de jongens het gelezen hebben.
- (3) Spreker A: Is Jan eindelijk uit de kast?
Spreker B: Ja, ze zeggen dat hij het waarschijnlijk zijn moeder heeft verteld.

Moedertaalsprekers van het Nederlands zullen, als we Broekhuis & Corver (2016) mogen geloven, waarschijnlijk moeite hebben met het vertolken van de rol van spreker B in bovenstaande drie dialogen. Spreker B produceert namelijk alleen maar ongrammaticale zinnen, getuige de asterisk die zulke zinnen (respectievelijk (51b'), (58a) en (63b) in hoofdstuk 13) van de auteurs krijgen ten teken van hun ongrammaticaliteit. Wat mij betreft zijn de uitingen van spreker B in bovenstaande dialogen wel degelijk grammaticaal.

Het is natuurlijk interessant dat moedertaalsprekers van het Standaardnederlands (Broekhuis & Corver enerzijds, ik anderzijds) onderling zo kunnen verschillen in wat zij als grammaticaal beoordelen en wat niet. In deze bijdrage aan de discussie zal ik de vermeende ongrammaticaliteit van de uitingen van spreker B in (1)-(3) aan een nadere beschouwing onderwerpen aan de hand van de analyses die deze ongrammaticaliteit volgens Broekhuis & Corver zouden moeten verklaren. Mijn conclusie zal zijn dat met de huidige mogelijkheden van taalwetenschappelijk onderzoek het beschrijven van de grammatica van het Nederlands niet langer louter op basis van intuïties van taalkundig onderlegde moedertaalsprekers zou moeten gebeuren, omdat de totstandkoming van deze intuïties deels gebaseerd lijkt te zijn op onderliggende theoretische analyses.

2 Ongemarkeerde woordvolgorde in het middenveld

Broekhuis & Corver (2016) besteden in hoofdstuk 13 allereerst aandacht aan de ongemarkeerde volgorde in het Nederlandse middenveld. Zin (4) als antwoord op de vraag *Wat is er gisteren gebeurd?* presenteert de ongemarkeerde volgorde van een passieve zin volgens Broekhuis & Corver (p. 1579, zin (17)):

- (4) Gisteren werden door Jan Marie de boeken aangeboden.

Omdat in (4) het indirect object *Marie* voorafgaat aan het subject *de boeken* (als je de zin goed interpreteert, dus niet met *Jan Marie* als één persoon), beargumenteren Broekhuis & Corver dat de ongemarkeerde volgorde van nominale argumenten van een werkwoord in het Nederlands niet gebaseerd is op de grammaticale functies die de argumenten hebben, maar op hun semantische rol. In ongemarkeerde zinnen gaat de Agens vooraf aan de Recipient (Goal) en deze komt voor het Thema. Een andere generalisatie is dat nominale objecten (NP's) voorafgaan aan prepositionele objecten (PP's). Als er een PP gebruikt wordt voor het indirect object in (4), dan zou de ongemarkeerde volgorde die in (5) worden:

- (5) Gisteren werden door Jan de boeken aan Marie aangeboden.

Behalve deze generalisaties over de volgorde van argumenten van het werkwoord, is er ook een belangrijke rol weggelegd voor zinsaccent bij het bepalen van de ongemarkeerde woordvolgorde. Broekhuis & Corver nemen aan dat het zinsaccent syntactisch bepaald wordt: in ongemarkeerde volgordes valt het zinsaccent op de diepst ingebedde constituent in het *lexicale domein* (de vP). Beschouw nu de volgende twee zinnen (zinnen (21a,b), p. 1581):

- (6) Jan heeft waarschijnlijk mijn zuster bezocht.
 (7) Jan heeft mijn zuster waarschijnlijk bezocht.

In zin (6) staat het object in *ongescrambelde* positie, d.w.z. ter rechterzijde van het zinsadverbium *waarschijnlijk* (Broekhuis & Corver willen de term *scrambling* niet gebruiken omdat er verschillende soorten scrambling zijn, maar ik gebruik voor het gemak de term wel). In zin (7) staat het direct object in *gescrambelde* positie, ter linkerzijde van het adverbium. Volgens Broekhuis & Corver valt het zinsaccent in (6) op *zuster* (er moet dan worden aangenomen dat het object dieper is ingebed in de syntactische structuur dan het werkwoord), maar in (7) is de originele objectpositie leeg (het object staat nu hoger in de structuur) en valt het zinsaccent volgens Broekhuis & Corver op het werkwoord *bezocht*. De zinnen (8) en (9) worden als extra bewijs hiervoor gepresenteerd (zinnen (24a,b), p. 1582):

- (8) Jan heeft op zijn vader gewacht.
 (9) Jan heeft op het perron gewacht.

Volgens Broekhuis & Corver valt het zinsaccent in (9) niet op *perron* maar op *gewacht*. De verklaring die ze hiervoor geven is dat *op het perron* geen voorzetselvoorwerp is, maar een bijwoordelijke bepaling. Daarom is *gewacht* de diepst ingebedde constituent en valt het accent dus niet op *perron*. Echter, als ik zin (9) zo neutraal mogelijk uitspreek, valt het accent wel degelijk op *perron* en niet op *gewacht*. Ook valt bij mij het accent in (7) nog steeds op *zuster* en niet op *bezocht*. Beschouw ook de volgende zin:

(10) Jan heeft de hele dag op zijn nieuwe auto gewacht.

Deze zin is ambigu, omdat *op zijn nieuwe auto* een voorzetselvoorwerp kan zijn (zoals in (8)) of een bijwoordelijke bepaling van plaats (zoals in (9)). Volgens Broekhuis & Corver zou deze zin echter onmiddellijk gedesambigueerd worden door hem uit te spreken. In de ene lezing zou het zinsaccent op *auto* vallen en in de andere lezing op *gewacht*. Volgens mij valt in beide gevallen het accent op *auto* en blijft de zin dus ook ambigu als je hem uitspreekt.

Dat intonatie niet bepaald wordt door de syntactische structuur per se, wordt ook duidelijk als we de volgende twee zinnen met elkaar vergelijken (de Hoop et al. 2007):

- (11) De meeste mensen slapen 's nachts.
 (12) De meeste mensen drinken 's nachts.

De meest voor de hand liggende interpretatie van (11) is dat op de vraag wat de meeste mensen 's nachts doen, het antwoord *slapen* is, terwijl de meest voor de hand liggende interpretatie van (12) is dat de meeste mensen die drinken, dat 's nachts doen. In zin (11) zal daarom het accent op *slapen* vallen en in (12) op *'s nachts*. Nu kunnen we natuurlijk aannemen dat er in (11) en (12) twee verschillende onderliggende syntactische structuren zijn, die de twee ongemarkeerde intonatiepatronen bepalen. Het punt is echter dat het verschil in intonatie (en dus ook het eventuele verschil in syntactische structuur) tussen de twee zinnen een gevolg is (en niet de oorzaak) van de (ongemarkeerde) interpretatie die de zinnen krijgen. Oppervlakkig gezien, d.w.z. op basis van de zinsstructuur, is er geen enkele reden om (11) anders uit te spreken dan (12). Dat we (11) toch anders uitspreken dan (12), komt omdat de intonatie een gevolg is van de (ongemarkeerde) interpretaties die we op basis van onze wereldkennis aan de zinnen toekennen. Die ongemarkeerde interpretaties kunnen zelf dus niet het gevolg zijn van de verschillen in intonatie.

3 Dan heb je de verkeerde waarschijnlijk ingelicht

Een laatste factor waaraan een belangrijke rol wordt toegedicht in het al dan niet scramblen van een direct object over een adverbium, is informatiestructuur. Directe objecten in gescrambelde positie zijn *discourse-old*, terwijl directe objecten in ongescrambelde positie *discourse-new* zijn. Dat verklaart volgens Broekhuis & Corver (p. 1592) de ongrammaticaliteit van de uiting van spreker B in de dialoog in (1), hieronder herhaald als (13):

- (13) Spreker A: Ik heb het aan Peter verteld.
Spreker B: Dan heb je de verkeerde waarschijnlijk ingelicht.

Hoewel *de verkeerde* verwijst naar *Peter* is de informatie dat Peter de verkeerde persoon is, nieuw in de discourse. Daarom kan volgens Broekhuis & Corver het object in (13) niet in de positie links van het zinsadverbium *waarschijnlijk* staan (tenzij het contrastieve klemtoon krijgt). Dat definitieve NP's niet per se *discourse-old* hoeven te zijn of contrastieve klemtoon hoeven te krijgen om in gescrambelde positie te verschijnen, is ook te zien in de dialoog in (14). Hier is de gescrambelde definitieve NP een zogenaamde *zwakke definitief* die *discourse-new* is en geen contrastieve klemtoon draagt:

- (14) Spreker A: Marie was ineens weg!
Spreker B: Dan heeft ze de bus waarschijnlijk genomen.

Volgens Broekhuis & Corver valt in de uiting van spreker B het zinsaccent op *genomen* omdat het object in gescrambelde positie staat. Dat kan alleen als het object *discourse-old* is (maar dat is het in bovenstaande dialoog niet) of als het contrastieve klemtoon krijgt (maar ook dat is niet zo in bovenstaande dialoog). Daarom zou deze zin in de analyse van Broekhuis & Corver ook als ongrammaticaal worden bestempeld. De variant waarin het object in ongescrambelde positie staat, zal voor sprekers van het Nederlands ontegenzeggelijk de voorkeur hebben, maar dit maakt de gescrambelde variant nog niet ongrammaticaal (contra Broekhuis & Corver).

4 Omdat waarschijnlijk de jongens het gelezen hebben

De volgorde van nominale argumenten op basis van hun semantische rol (Agens-Recipient-Thema), moet gewaarborgd blijven, ook als een van de

argumenten scrambelt vanwege de informatiestructuur. Volgens Broekhuis & Corver is het handhaven van die ongemarkeerde volgorde sterker dan de informatiestructurele principes die zeggen dat nominale argumenten die discourse-new zijn in het lexicale domein moeten blijven staan, terwijl nominale argumenten die discourse-old zijn moeten scramblen naar een functionele naamvalspositie. Dat betekent dat in een actieve zin een Thema dat scrambelt omdat het discourse-old is, ook de Agens mee omhoog in de syntactische structuur dwingt, zelfs als de Agens zelf niet discourse-old is. Daarom is de uiting van spreker B in een dialoog als in (2), hieronder herhaald als (15) ongrammaticaal volgens Broekhuis & Corver:

- (15) Spreker A: Waarom is Jan zo kwaad over dat papiertje?
 Spreker B: Omdat waarschijnlijk de jongens het gelezen hebben.

Omdat *het* terugverwijst naar het papiertje uit de vraag van spreker A, is het discourse-old. Het moet dus scramblen en het subject dat discourse-new is, moet dus mee om de volgorde Agens-Thema veilig te stellen. Het probleem is dat de uiting van spreker B misschien minder voor de hand ligt dan *Omdat de jongens het waarschijnlijk gelezen hebben*, maar wel mogelijk is (voor mij) en zeker niet ongrammaticaal. Volgens Broekhuis & Corver kan de zin in (15) alleen grammaticaal gemaakt worden door contrastieve klemtoon op *jongens* te leggen: *Jan is kwaad omdat waarschijnlijk de jón-gens het gelezen hebben, niet de meisjes*. Echter, een contrastieve lezing is wat mij betreft niet nodig om zin (15) van spreker B grammaticaal te maken.

In Bouma & De Hoop (2008) wordt een OT-analyse van de interactie tussen intonatie en interpretatie van gescrambelde en ongeschrambelde pronomina gegeven. De puzzel die we daar presenteren (en pogen op te lossen), is dat pronomina objecten in ongeschrambelde positie vaak klemtoon en daardoor een contrastieve lezing krijgen (cf. De Hoop 2004), maar niet noodzakelijk. Een voorbeeld van een onbeklemtoond, niet-contrastief voornaamwoord komt uit *Kees de jongen* (1923) van Theo Thijssen:

- (16) Maar deze mevrouw was een edele dame, en als ze ooit hem nodig had ...

De meest natuurlijke (ongemarkeerde, neutrale) volgorde zou *en als ze hem ooit nodig had* zijn, maar toch is het niet zo dat in deze afwijkende volgorde het voornaamwoord *hem* in (16) klemtoon krijgt, noch een contrastieve lezing. In Bouma & De Hoop (2008) wordt verondersteld dat het

adverbium *ooit* klemtoon draagt in (16) en dat het pronomen onbeklemtoond in ongescrambelde positie blijft staan om een botsing met het zinsaccent op *nodig* te voorkomen (in navolging van Féry 2007). Ook in (15) hoeft er geen contrastieve klemtoon op *het* noch op *jongens* te vallen om de zin grammaticaal te maken (contra Broekhuis & Corver).

5 Ze zeggen dat hij het waarschijnlijk zijn moeder heeft verteld

Het moge duidelijk zijn dat de uiting van spreker B in dialoog (3), hieronder herhaald als (17), ook ongrammaticaal moet zijn volgens de analyse van Broekhuis & Corver (p. 1596):

- (17) Spreker A: Is Jan eindelijk uit de kast?
 Spreker B: Ja, ze zeggen dat hij het waarschijnlijk zijn moeder heeft verteld.

Een direct object, of het nu nominaal of pronominaal is, mag in de theorie van Broekhuis & Corver niet over een indirect object heen scramblen. Omdat het direct object *het* in (17) naar de linkerkant van het adverbium *waarschijnlijk* gescrembeld is, terwijl het indirect object in ongescrambelde positie is achtergebleven, wordt de volgorde-eis die stelt dat een Recipient vooraf moet gaan aan een Thema (ook als dat Thema gescrembeld zou zijn om informatiestructurele redenen), geschonden. In de uiting van spreker B in (17) verwijst *het* terug naar het uit de kast zijn van Jan en *hij* verwijst naar Jan. Het ligt dus voor de hand om zowel het subject als het direct object in de uitspraak te reduceren:

- (18) Ja, ze zeggen dat ie 't waarschijnlijk z'n moeder heeft verteld.

Wat mij betreft kan het Thema *het* in de uiting van spreker B in (17) of 't in (18) zonder problemen ter linkerkant van *waarschijnlijk* staan, terwijl de Recipient *zijn moeder* die discourse-new is, achterblijft ter rechterkant van het adverbium (contra Broekhuis & Corver).

6 Een empirische aanpak

Hierboven heb ik laten zien dat de grammaticaliteitsoordelen van Broekhuis & Corver soms bepaald lijken te worden door de theorie die zij opstellen. Tegenwoordig ligt een empirische aanpak voor de hand om theoretische aannames te toetsen. Als een grammatica verschillende volgordes toelaat, kunnen allerlei factoren de keuze van een spreker voor de ene of de andere volgorde beïnvloeden. Welke volgorde de voorkeur krijgt of welke volgorde uiteindelijk gekozen wordt, kan in een psycholinguïstisch experiment getest worden (zie bijvoorbeeld Lamers & De Hoop 2014) of in een corpus onderzocht (zie Van Bergen 2011 voor verschillende voorbeelden van deze aanpak).

Een mooi voorbeeld van een dergelijke corpusstudie met verrassende resultaten is de studie van Van Bergen & de Swart (2010) naar het al dan niet scramblen van een direct object over een zinsadverbium. In het Nederlands komen beide volgordes van object en adverbium ongeveer even vaak voor: Van Bergen & de Swart (2010) bekijken in totaal 8656 zinnen met een direct object en een zinsadverbium uit het *Corpus Gesproken Nederlands* en constateren dat het object in net iets meer dan de helft daarvan ($n = 4708$) in ongescrambelde positie (ter rechterzijde van het zinsadverbium) staat en in net iets minder dan de helft ($n = 3948$) in gescrambelde positie (ter linkerzijde van het adverbium).

Sprekers van het Nederlands kunnen dus kiezen tussen twee volgordes en blijkbaar kiezen ze ongeveer even vaak voor de ene als voor de andere volgorde (wat op zich al een verklaring kan zijn voor het feit dat beide volgordes nog steeds bestaan in het Nederlands). Van Bergen & de Swart laten zien dat de keuze voor een van beide volgordes hoofdzakelijk afhangt van het type van het object: persoonlijke voornaamwoorden verschijnen bijna altijd in gescrambelde positie, indefiniëte NP's bijna nooit. Hun statistische model dat slechts gebaseerd is op verschillen in type object (waarbij ze zes verschillende typen onderscheiden) doet in 97,4% van de zinnen de juiste voorspelling over de volgorde van object en adverbium. Eigenamen laten in hun data de meeste variatie zien: zij vallen precies tussen persoonlijke voornaamwoorden (die vrijwel altijd scramblen) en *volle* NP's (die vrijwel niet scramblen) in. Van Bergen & de Swart (2010) verklaren hun bevindingen in termen van sprekersgemak: als sprekers zinnen bouwen, dan zullen ze (wanneer de grammatica dat toestaat) eerst de woorden uiten die het gemakkelijkst bij hen naar boven komen en dan pas de wat minder toegankelijke woorden, waarvoor ze wat meer moeite moeten doen. Op basis van een steekproef van 1000 willekeurige zinnen uit het

CGN definiëren Van Bergen & de Swart toegankelijkheid (*predictability*) op basis van *class size* (n), d.w.z. het aantal woorden in een bepaalde woordklasse (typen, geen tokens): $1/n$. Omdat er maar een klein aantal verschillende pronomina is ($n = 41$), is de toegankelijkheid daarvan groot (0,024). Daarentegen is de verzameling verschillende nomina in de 1000 zinnen heel groot ($n = 1270$) en de toegankelijkheid van nomina dus klein (0,0008). Adverbia en eigennamen zitten daartussenin: adverbia ($n = 166$) hebben een toegankelijkheid van 0,006 en eigennamen ($n = 191$) hebben een toegankelijkheid van 0,005. Dit verklaart de voorkeur van pronomina voor de gescrambelde positie, de voorkeur van gewone nominale constituenten voor de ongescrambelde positie en de grote mate van variatie in het geval van eigennamen. Dit is geïllustreerd in Tabel 1 (gebaseerd op Tabel 12 van Van Bergen & de Swart 2010).

Tabel 1 Scrambling op basis van toegankelijkheid

	PRONOMEN	ADVERBIUM	EIGENNAAM	NOMEN
Toegankelijkheid (1/n)	0,024	0,006	0,005	0,0008
Scrambling	+		+/-	-

Als de grammatica beide volgordes toestaat, zullen sprekers meestal een woord met een hoge toegankelijkheid eerst uiten (dat is gemakkelijk) en dan pas het woord met een lagere toegankelijkheid (dat kost relatief meer tijd en moeite). Het kleine verschil in toegankelijkheid tussen een adverbium en een eigennaam verklaart het wispelturige scramblinggedrag van eigennamen. In zo'n geval is er volgens Van Bergen & de Swart ruimte voor andere factoren om het al dan niet scramblen mede te beïnvloeden. Ze laten zien dat voor de subset van eigennamen een model op basis van animacy (de eigennaam verwijst naar een levend wezen of niet), klemtoon (de eigennaam draagt klemtoon in de zin of niet – de fragmenten zijn daartoe beluisterd door de auteurs) en lengte (gemeten in aantal tekens) in bijna 75% van de gevallen de juiste voorspellingen doet met betrekking tot het al dan niet scramblen van de eigennamen. Voor de andere typen NP's doen deze factoren er echter nauwelijks toe, eenvoudigweg omdat ze in het niet vallen bij de factor 'type NP' (Van Bergen & de Swart 2010).

7 Conclusie

In deze bijdrage heb ik gefocust op enkele grammaticaliteitsoordelen van Broekhuis & Corver (2016) waar ik het niet mee eens ben (er staan natuur-

lijk in het betreffende hoofdstuk ook veel oordelen die ik wel deel). Deze oordelen over de grammaticaliteit van bepaalde woordvolgordevarianten in het Nederlands lijken gebaseerd te zijn op theoretische aannames van Broekhuis & Corver (2016), die niet getoetst zijn aan de empirie. Het zou beter zijn om 'echte taal' als uitgangspunt te nemen voor de beschrijving én de verklaring van woordvolgordevariatie in het Nederlands.

Referenties

- Bergen, Geertje van (2011). *Who's first and what's next? Animacy and word order variation in Dutch language production*. Proefschrift Radboud Universiteit Nijmegen.
- Bergen, Geertje van & Peter de Swart (2010). Scrambling in spoken Dutch: Definiteness versus weight as determinants of word order variation. *Corpus Linguistics and Linguistic Theory* 6, 267-295.
- Bouma, Gerlof & Helen de Hoop (2008). Unscrambled pronouns in Dutch. *Linguistic Inquiry* 39, 669-677.
- Broekhuis, Hans & Norbert Corver (2016). *Syntax of Dutch: Verbs and Verb Phrases Volume III*. Amsterdam: Amsterdam University Press.
- Féry, Caroline (2007). The prosody of topicalization. In: Kerstin Schwabe & Susanne Winkler (red.), *On information structure, meaning and form: Generalizations across languages*. Amsterdam: John Benjamins, 69-86.
- Hoop, Helen de (2004). On the interpretation of stressed pronouns. In: Reinhard Blutner & Henk Zeevat (red.), *Optimality Theory and pragmatics*. New York: Palgrave Macmillan, 25-41.
- Hoop, Helen de, Petra Hendriks & Reinhard Blutner (2007). On compositionality and bidirectional optimization. *Journal of Cognitive Science* 8, 137-151.
- Lamers, Monique & Helen de Hoop (2014). Animate object fronting in Dutch: a production study. In: Brian MacWhinney, Andrej Malchukov & Edith Moravcsik (red.), *Competing motivations in grammar and usage*. Oxford: Oxford University Press, 42-53.

Over de auteur

Helen de Hoop, Centre for Language Studies (CLS), Radboud Universiteit Nijmegen

E-mail: h.dehoop@let.ru.nl