

Over het voorkomen van adjacente identieke elementen

Eric Hoekstra en Henk Wolf, Fryske Akademy, Leeuwarden

1. Inleiding

De menselijke geest bezit het vermogen om taal te produceren. In die zin denkt men in de generatieve grammatica in al haar varianten aan de grammatica als een productiesysteem. De nadruk wordt heel sterk gelegd op het creatieve aspect van het taalvermogen. De menselijke geest heeft echter ook het vermogen om taal te interpreteren. Het deel van het taalvermogen dat gebruikt wordt voor interpretatie wordt doorgaans aangeduid met de term "parser". De vragen die men nu zou kunnen formuleren zijn:

- (A) Wat is de relatie tussen grammar en parser?
- (B) Wat zijn de formele eigenschappen van de parser, in het bijzonder die formele eigenschappen die de grammar ontbeert? En daarmee samenhangend:
- (C) Vertonen taalfeiten eigenschappen die niet aan de grammar maar aan de parser zijn toe te schrijven?

De praktijk van het meeste grammatica-onderzoek is dat men de vraag naar de rol van de parser buiten beschouwing laat.

Stel nu dat het volgende het geval is. De grammar en de parser leggen beide hun eigen beperkingen aan natuurlijke taal op. Natuurlijke taal zou dan het produkt zijn van een interactie tussen grammar en parser. De huidige praktijk is dat men taalfeiten exclusief onder de noemer van de grammar probeert te verantwoorden. Die huidige praktijk zou, onder dit scenario, tot mislukken gedoemd zijn, omdat zij impliciet ten onrechte ervan uitgaat dat de rol van de parser veronachtzaamd mag worden. De vraag is dan: zijn er aanwijzingen dat bovengenoemd scenario een kern van waarheid bevat? Zijn er vreemde verschijnselen in taal die we wanhopig met behulp van de grammar proberen te verklaren maar waar we onvoldoende in slagen?

In deze lezing gaan we een lijstje van verschijnselen bespreken die ons inziens een aanwijzing zouden kunnen vormen dat er inderdaad zoiets is als een parser die principieel verschilt van de grammar. Dat zou ons dwingen tot de positie dat het mechanisme om taal voort te brengen principieel anders is, ten dele, van het mechanisme om taal te interpreteren.

De verschijnselen die we de revue laten passeren vormen geen homogene groep. Het zijn echter wel allemaal verschijnselen die problematisch zijn voor het gewone grammatica-onderzoek, en die daarom, terecht vanuit dat perspectief, naar de periferie van het onderzoek worden geschoven. Binnen deze groep verschijnselen onderscheiden we twee soorten.

Eenzijds zijn er verschijnselen die te maken hebben met een verbod op het tweemaal voorkomen van hetzelfde element. We geven het achterliggende principe weer als hieronder:

(1) Het principe van distinctiviteit: vermijd alfa alfa

Dit principe is gerelateerd aan principes zoals de Stuttering Prohibition van Ross of Postal, de Unlike Category Condition van T. Hoekstra (?). Recentelijk hebben Ackema en Neeleman in NLLT nog een principe voorgesteld dat van twee identieke kenmerken één deleert. Op deze plaats gaan we niet op de historische kant van de zaak in. We bespreken een aantal verschijnselen die instantiaties van Vermijd Alfa Alfa lijken te zijn. De verschijnselen zelf gebruiken we om de condities op Vermijd alfa alfa scherper te stellen.

2. Een merkwaardige restrictie op sententiële inbedding: * Dat (COMP) Dat (Comp)

Er is een merkwaardige restrictie op sententiële inbedding in subjectpositie, die reeds bekend is sinds de jaren zeventig (Ref ?). Het is namelijk zo dat een DAT-zin wel in onderwerpspositie kan staan, maar niet ingebed mag worden:

- (1a) [Dat zij een 10 heeft gehaald] bevalt mij.
 (1b) * Ik zei dat [dat zij een 10 heeft gehaald] mij beviel.
 (1c) Ik zei dat het mij beviel dat zij een 10 heeft gehaald.

Het is een uitermate vreemd verschijnsel. Het is niet in zijn algemeenheid zo dat linkerrecursie is uitgesloten in natuurlijke taal. Wat is de generalisatie? Heeft het iets maken met de twee DAT'en die naast elkaar staan? Die hypothese valt te testen door met OF en DAT te gaan werken. (2) laat eerst zien dat OF zich net zo gedraagt als DAT.

- (2a) Ik weet niet [of zij een 10 heeft gehaald]
 (2b) * Ik weet niet [of [of zij een 10 had gehaald] haar verteld werd]]
 (2c) Ik weet niet of haar verteld werd of zij een 10 had gehaald

Nu gaan we OF en DAT afwisselen:

- (3a) ? Ik weet niet [of [dat zij een 10 had gehaald] haar verteld is]]
 (3b) Ik weet niet of haar verteld is dat zij een 10 had gehaald

Zoals uit de vergelijking van (3a) met (2b) en (1b) blijkt, is er sprake van een significante verbetering van de zin. We concluderen daarom dat de onacceptabiliteit van (1b) en (2b) ten dele samenhangt met het optreden van twee opeenvolgende voorkomens van hetzelfde voegwoord, namelijk DAT in (1b) en OF in (2b). We formuleren deze observaties hieronder:

(4) Het voegwoord DAT mag niet onmiddellijk volgen op het voegwoord DAT

Met een nevenschikkend en een onderschikkend voegwoord OF lijkt dezelfde restrictie te gelden:

- (5a) * Ik weet niet of ze komt of of ze gaat.
 (5b) Ik weet niet of ze komt of dat ze gaat.

Het tweede OF in moet worden vervangen door het substituutvoegwoord DAT om een grammaticale zin te krijgen (Hoekstra 1992).

De generalisatie is echter niet dat er geen twee datten op elkaar mogen volgen:

(6) Ik denk dat dat korrekt is.

Het gaat om twee datten met dezelfde syntactische categorie. Een voorlopige conclusie kan zijn dat twee identieke voegwoorden niet adjacent mogen zijn. Die restrictie betreft de fonologische vorm van het voegwoord. Dat is vanuit streng syntactisch perspectief onverwacht. Waarom zouden in de grammatica syntactische principes een dergelijke gevoeligheid voor fonetische vormen vertonen? Binnen een aanpak die de parser veronachtzaamt, zou dit verschijnsel op PF verantwoord worden. Maar het probleem doorschuiven naar PF is eveneens onbevredigend aangezien het niet het soort verschijnsel betreft dat in de fonologie behandeld wordt. Vanuit de fonologie geredeneerd is de verwijzing naar de syntactische categorie "voegwoord" weer bevreemdend. We hebben te maken met een type restrictie dat enerzijds naar domme oppervlaktevormen verwijst, anderzijds een referentie aan syntactische categorieën vereist.

3. AL en WOL in het Fries: * AL AL

Het woord AL kan in het Fries zowel "al" als "wel" betekenen. Zie de volgende zinnen:

(7a) Komt er al?

"Komt hij al?"

(7b) Komt er al as komt er net?

"Komt hij wel of komt hij niet."

'al' en 'wol' in de betekenis 'wel degelijk', 'wel in tegenstelling tot iets anders' zijn normaliter uitwisselbaar, zoals uit blijkt uit de zinnen hieronder. Beide woorden kunnen zonder problemen met een bijwoord van tijd voorkomen, zoals blijkt uit de zinnen hieronder. We zouden nu verwachten dat ook 'al' ('wel') in combinatie met 'al' ('al') mogelijk is. Zin (8c) laat zien dat dat niet zo is.

(8a) Fan 'e moarn hied er de krante noch net sjoen, mar no wol / al.

vanmorgen had hij de krant nog niet gezien, maar nu wel

(8b) Fan 'e moarn hied er de krante noch net sjoen, mar no al wol / wol al

vanmorgen had hij de krant nog niet gezien, maar nu al wel / wel al

(8c) * Fan 'e moarn hied er de krante noch net sjoen, mar no al al

vanmorgen had hij de krant nog niet gezien, maar nu al wel.

We moeten dus concluderen dat er een mechanisme bestaat dat het voorkomen van de reeks 'al al' uitsluit. We zouden kunnen uitgaan van een oppervlakkige, fonologische regel, die hier uitsluit dat de fonologisch identieke woordjes 'al' adjacent voor de heren komen. Dat zou niet correct zijn, want in de volgende zin komen wel twee allen voor:

(9) Hy tocht dat er no al al it kwea yn himsels oerwûn hie.

hij dacht dat hij nu wel al het kwaad in zichzelf overwonnen had

Hier gaat het echter om twee allen die syntactisch verschillen: de eerste al is een bijwoord bij VP en de tweede al een pronominaal element. In het ongrammaticale (f)-voorbeeld zijn beide allen bijwoorden bij VP. De dubbele datten en de dubbele allen zijn beide verboden als ze dezelfde syntactische kenmerken hebben.

4. *(ER ER) in het Nederlands

De bovenstaande verschijnselen doen denken aan observaties van Van Riemsdijk (1978), die constateert dat een locatief en een pronominaal 'er' in het Nederlands niet adjacent mogen zijn, zoals in de zin hieronder wordt getoond

(10) * Zij heeft er er vaak over gesproken.

Bennis (1986) concludeert op grond de zinnen hieronder dat er geen sprake kan zijn van een syntactische restrictie op het tegelijkertijd voorkomen van locatieve adverbia en R-pronomen:

(11a) Zij heeft eri daarjvaak [tj over] gesproken.

(11b) Zij heeft erjdaari vaak [tj over] gesproken.

Blijkbaar speelt ook hier de vorm van de woorden een rol: de fonologisch identieke woordjes 'er' mogen niet naast elkaar staan in een zin.

Dat *(ER ER)-filter functioneert vaker, zoals blijkt wanneer we de hoofdzin (18a) veranderen in de bijzin (18b), waardoor de twee woordjes 'er' naast elkaar zouden komen te staan. Dan wordt de zin ongrammaticaal. Wanneer een van beide voorkomens van 'er' wordt gewist, zoals in (18c), dan is de zin wel weer grammaticaal.

(12a) Er zijn er twee (van).

(12b) * Dat er er twee (van) zijn.

(12c) Dat er twee (van) zijn.

Vervangen we het pronominaal of expletieve 'er' door 'daar', dan kan er zonder probleem een bijzin worden gemaakt met 'er' erin:

(13a) Daar zijn er twee (van).

(13b) Dat er daar twee (van) zijn.

Verskillende woordjes 'er' moeten in veel gevallen samenvallen. Bennis (1986) geeft daarvan een hele lijst voorbeelden. Enkele daarvan:

(14a) dat er (*er) twee jongens op zaten. (expletief en R-pronomen vallen samen)

(14b) dat er (*er) niemand meer woonde. (expletief en locatief vallen samen)

(14c) dat Jan er (*er) twee over gelezen heeft (R-pronomen en partitief vallen samen).

(14d) dat er (*er/*er er) niemand drie over geschreven heeft (expletief, partitief en R-pronomen vallen samen)

Bennis (1986) verklaart de ongrammaticaliteit van de ongrammaticale zinnen met behulp een PF-filter dat de opeenvolging van bepaalde typen lettergrepen met een sjwa uitsluit. Vervolgens worden er strategieën toegepast zoals het laten 'samensmelten' van verschillende woordjes 'er' tot een enkel voorkomen daarvan. Dus terwille van een fonologische restrictie wordt er over woordgrenzen heen syntactische informatie gewist. Deze feiten kunnen geassimileerd worden aan de voorgaande, maar dan moeten we aannemen dat het bij al deze erren om dezelfde syntactische soort ER gaat.

5. Syntactische identiteit en coordinatie

Het is niet altijd zo dat het gaat om dezelfde woorden met dezelfde syntactische categorie. In het Nederlands verliest het werkwoord in de tweede persoon enkelvoud zijn uitgang -T in inversie. Volgens Ackema en Neeleman (2003:687) is dat het gevolg van een taalspecifieke regel die van een sequentie van twee identieke kenmerken er eentje wist: KOMT JE wordt KOM JE door eliminatie van het kenmerk <2SG> op V onder invloed van de <2SG> op het pronomen.

In de voorgaande gevallen was er sprake van lexicale en syntactische identiteit. In dit geval is er echter alleen sprake van syntactische identiteit. Dat vormt volgens Ackema & Neeleman een potentieel aangrijpingspunt voor taalspecifieke regels. Voor ons is van belang dat zowel de inversie-feiten als de DAT-DAT feiten en de AL-AL feiten suggereren dat syntactische identiteit van twee adjacente elementen een kontekst is (i) waar lexicale identiteit is uitgesloten (ii) waar lexicale herschrijffregels lexicale inhoud kunnen veranderen (T-verlies in het Nederlands).

De visie van Ackema & Neeleman is gerelateerd aan de onze. Wij hebben net laten zien dat identiteit van syntactische kenmerken en van leksikale inhoud überhaupt uitgesloten is. Ackema & Neeleman laten zien dat als er identiteit van syntactische kenmerken is zonder leksikale identiteit, dat dan syntactische taalspecifieke regels daarop aangrijpen kunnen.

Nu speelt syntactische identiteit ook een rol in coordinatie. Conjuncten hebben immers vaak dezelfde structuur en kenmerkinhoud. Als er een soort distinctiviteit een rol speelt bij taalspecifieke processen, zoals Ackema & Neeleman veronderstellen, dan zou dat principe er ook toe kunnen leiden dat twee conjuncten de neiging hebben om zich verschillend te gaan gedragen. Dat lijkt inderdaad zo te zijn.

Op het tweede lid en eventuele latere leden van coördinatieconstructies worden allerlei vereenvoudigingsstructuren toegepast. De zinnen hieronder laten daar een aantal voorbeelden van zien.

(15a) Ik wilde even bij de bakker langsgaan en een broodje kopen.

(15b) Op het tweede lid en eventuele latere leden.

(15c) Als je nu vroeg naar bed gaat en je slaapt morgen lekker uit, ben je weer helemaal het heertje.

(15d) Gehen Sie und nehmen den kleinen grünen Asteroiden auf.

Zin (a) toont een vrij eenvoudig voorbeeld van deletie van subject, persoonsvorm en modaliserend bijwoord in het tweede lid. De zinsnede na 'en' wordt geïnterpreteerd als ving hij aan met het zelfde 'ik wilde even' dat ook in het eerste lid staat. Iets soortgelijks ziet men bij (b), waar het tweede lid wordt geïnterpreteerd als bevatte het het in het eerste lid genoemde voorzetsel 'op'. Zoals genoeglijk bekend is dit type deletie uitgesloten in het eerste lid of in niet-gecoördineerde zinsneden.

Zin (c) laat zien hoe het tweede lid van een in het regeerbereik van een voegwoord vallende coördinatie niet de verwachte geïnverteerde woordvolgorde heeft, maar inplaats daarvan, werkwoordvooropplaatsing zoals men die men normaal in een niet-conditionele hoofdzin aantreft. Het moge duidelijk zijn dat toepassing van deze woordvolgorde in het eerste lid van de coördinatie of buiten een coördinatieconstructie tot een onwelgevormde zin zou leiden. Tenslotte (d): hier is sprake van subjectdeletie in de tweede van twee gecoördineerde Duitse imperatiefzinnen. Ook hier geldt weer dat we te maken hebben met een verschijnsel dat niet kan optreden in het eerste lid of in een niet-gecoördineerde zin.

Hoekstra (1987) verklaart zinnen als (15c) hierboven in de syntaxis op een zelfde soort manier als Ackema & Neeleman de morfologische feiten op PF verklaren, namelijk door bij naast elkaar voorkomende gelijke features de tweede op rij te schrappen. Hoekstra poneert de stelling dat ondergeschikte zinnen een nominaal karakter hebben ([+N]), terwijl niet-ondergeschikte zinnen verbaal ([+V]) van aard zouden zijn. Een verdere aanname is dat er [+V] de ongemarkeerde optie is. Wanneer nu in een coördinatieconstructie twee zinsneden met een nominaal feature worden nevenschikt, vervalt het feature van het tweede lid. Nu verstoken van een morfologische status zou het het ongemarkeerde verbale feature toebedeeld krijgen en daarmee de vorm van een hoofdzin aannemen.

In hetzelfde theoretisch kader is de Friese nevenschikte IPI-constructie te verantwoorden. Een paar voorbeelden vindt men hieronder.

(16a) De plysje soe by him komme en helje him op.

(16b) Ik woe ferve keapje en stryk doch myn stek net oer.

In deze zinnen neemt het tweede lid van de in het bereik van het hulpwerkwoord (soe, woe) staande coördinatie een nogal merkwaardige vorm aan. Het werkwoord staat op de eerste plaats in de zin en het heeft de vorm van de gebiedende wijs. Door vorm en volgorde heeft de hele zin rechts van 'en' de vorm van een gebiedendewijszin.

Hoekstra's idee is dat het tweede lid van de beide nevenschikte nominale (want ondergeschikte) zinnen in het domein van het hulpwerkwoord zijn feature verliest en het ongemarkeerde verbale feature krijgt toebedeeld. Dat zou de vooropplaatsing van het werkwoord triggeren. Dat het werkwoord daarbij de vorm van een gebiedende wijs aanneemt, is een vooralsnog niet te verantwoorden verschijnsel. Dat het hier slechts een oppervlakkig vormkenmerk betreft blijkt aan het feit dat de IPI-constructie in vroegere taalfasen van het Fries op de plaats van de huidige gebiedende wijs een onbepaalde wijs had staan, dus bijvoorbeeld:

(17) Ik woe ferve keapje en strike doch myn stek net oer.

Dat geldt overigens ook voor diverse Nederduitse en Hoogduitse dialecten die zinnen als die in (23) kennen.

Hoekstra heeft het vraagstuk van de IPI-constructie in de syntaxis proberen op te lossen. Wij vragen ons af of dat vol te houden is. In ieder geval in zijn ontstaan kan de IPI ook goed een paratactische constructie zijn die wordt ingezet wanneer de afstand tussen het regerend element en het geregeerde zinsdeel te groot wordt. De gelaagde herschrijfstructuur van de zin wordt te complex voor het lineaire spraakproductieapparaat en moet het afleggen tegen een eenvoudiger lineaire structuur die semantisch ondergeschikte zinsneden realiseert als waren ze niet ondergeschikt.

De waarneming dat in oudere taalfasen van het Fries het werkwoord in de IPI-positie diverse vormen kon aannemen: gebiedende wijs, onbepaalde wijs, diverse persoonsvormen in de tegenwoordige en verleden tijd, duidt erop dat de zinnen zodanig complex waren dat er bij de sprekers grote onzekerheid optrad wat de juiste vorm van het werkwoord betrof. De uiteindelijke keuze voor toch maar een vorm was niet altijd dezelfde. Dat vormt natuurlijk een extra aanwijzing voor een verklaring in de paratactische sfeer.

6. Uitleiding

We hebben verscheidene gevallen gezien van taalverschijnselen waarbij sprake is van restricties die geformuleerd moeten worden met behulp van syntactische begrippen enerzijds en fonologische vormen anderzijds. Bij de restricties op fonologische vormen gaat het in een aantal gevallen om een opeenvolging van fonologische vormen, zoals bij de restrictie op twee DAT'en. Dat doet denken aan de Stuttering Prohibition (fan wa ek mar wer???)

Al deze verschijnselen zouden mogelijkerwijs op het conto van de parser kunnen worden geschreven. Een kenmerk van de parser is dat er restricties op oppervlaktevormen gelden gekoppeld aan syntactische begrippen. De formalisering van de rol van de parser zou ons in staat stellen een abstracte visie op de grammatica, die zijn kracht door de jaren heen ook heeft bewezen, vast te houden. Zonder een parsercomponent zouden we dergelijke feiten in de grammatica gaan proberen te verantwoorden. Veelal wordt dan uitgeweken naar PF, als 'restcategorie' binnen de grammatica. Zulke oplossingen lijken ons ongewenst, aangezien ze strijdig zijn met het abstracte en universele karakter van de grammatica.

Wat we hier gepresenteerd hebben zijn een verzameling puzzels die wat gemeen lijken te hebben. Naar aanleiding daarvan hebben we gesuggereerd dat die puzzels te maken hebben met de aanwezigheid van een parser, die eveneens zijn stempel op de taalfeiten drukt.

Adres van de auteurs:
Eric Hoekstra (ehoekstra@fa.knaw.nl), Henk Wolf (hwolf@fa.knaw.nl)
Fryske Akademy
Postbus 54
NL 8900 AB Leeuwarden

Referenties

- Bennis, H. (1986), Gaps and dummies, Dordrecht: Foris.
Hoekstra (1992), 'Of en dat nader bekeken', *De nieuwe taalgids* 85, p. 441-445.
Hoekstra (1994), Agreement and the nature of specifiers, *GAGL* 37, p. 159-168.
Riemsdijk, H. van (1978), 'On the diagnostics of WH-movement', in: S.J. Keyser (ed.), *Recent transformational studies in European languages*, MIT-press, p. 189-206.
Wolf, H. (1996), 'IPP en morfologische markering', *TABU* 26, p. 33-40.