

Prof.dr. Sjef Barbiers

De zonnige toekomst van de Nederlandse taalkunde



Universiteit
Leiden

Bij ons leer je de wereld kennen

De zonnige toekomst van de Nederlandse taalkunde

Oratie uitgesproken door

Prof.dr. Sjef Barbiers

bij de aanvaarding van het ambt van hoogleraar in de

Nederlandse Taalkunde

aan de Universiteit Leiden

op vrijdag 8 december 2017



**Universiteit
Leiden**

Mijnheer de Rector Magnificus, zeer gewaardeerde toehoorders,

Taal en hiërarchie

De verlossing, een roman uit 1921 van de beroemde Vlaamse schrijver Willem Elsschot, gaat over de jarenlange machtsstrijd tussen het hoofdpersonage Pol van Domburg en de pastoor van het Vlaamse dorp Groendal. Deze Pol is een tot het socialisme geneigde vrijdenker die zijn vrouw slaat en die de enige winkel van het dorp bezit, waar hij zijn klanten ontvangt met wat borrels om ze daarna op te lichten.

De eerste kennismaking van de lezer met het dorp Groendal gaat als volgt:

Toen hij een paar jaar getrouwd was, verliet hij de stad en kwam zich te Groendal vestigen, waar hij 'Pol' gedoopt werd, zonder meer, omdat hij nummer één was van een nieuw geslacht. Hij heette wel Leopold van Domburg, maar familienamen waren te Groendal in onbruik geraakt en het oudste lid van ieder gezin werd daar alleen bij zijn verkorte voornaam genoemd. Heeft Jan een zoon die Jozef heet, dan staat die bekend als Jef van Jan. En wordt Jef vader vóór Jan dood is, dan heet het kleinkind Louis van Jef van Jan. De dochter van Louis wordt Marie van Louis van Jef van Jan. Gaat de stamvader dood, dan wordt hij niet alleen ter aarde besteld, doch tevens worden de namen van al zijn nakomelingen van de zijne gezuiverd. Jef van Jan wordt gewoon Jef, Louis van Jef van Jan wordt Louis van Jef, en Marie van Louis van Jef van Jan wordt Marie van Louis van Jef. Soms blijft de naam van een afgestorven boer toch nog een tijd lang in zwang, en dan wel met het toevoegsel 'zaliger', wanneer er op 't dorp twee mensen met dezelfde voornamen rondlopen. Jan heet dan Jan van Pietzaliger, in tegenstelling met Jan van Warzaliger. Een dode wordt anders ook dáár spoedig vergeten, net zo goed als in de stad, en na één geslacht herinneren zich alleen nog een paar ouden van dagen dat Pietzaliger een goed boogschutter was, of dat hij veel kon drinken.

Deze passage roept naast weemoed interessante vragen op. Waarom zou in een dorpsgemeenschap zo'n ingewikkeld en

omslachtig systeem van naamgeving ontstaan, in plaats van dat iedereen een unieke naam krijgt waarmee hij of zij van de rest van het dorp kan worden onderscheiden? Blijkbaar heeft elke dorpeling van elke andere dorpeling een stuk stamboom in z'n hoofd. Dat komt in conflictsituaties zoals tussen Pol en de pastoor natuurlijk goed van pas, want dan moet je weten in wie zijn kamp iemand zit. En als je op zoek bent naar een partner waar je kinderen mee wilt krijgen is het natuurlijk ook handig als je weet tot welke familie iemand behoort. Voor overleven en voortplanting is het kennen van de familierelaties kortom cruciaal, en dit wordt uitgedrukt in de taal. De structuur van de taal weerspiegelt in dit geval de structuur van de stamboom. Het steeds dieper inbedden van relaties van het type vader-zoon of -dochter in de bestaande stamboom komt overeen met het steeds weer toevoegen van een relatie met het woordje *van*: Marie is van Louis, Louis is van Jef, en Jef is van Jan. En daarvoor zijn Marie en Louis ook van Jan.

Zo is er dus ook in de taal een hiërarchische structuur waarin hetzelfde patroon telkens herhaald wordt. Dat herhalen van een patroon in éénzelfde patroon noemen we recursie, een verschijnsel dat je overal in de natuur en cultuur kunt waarnemen. Denk aan het eindeloze herhaling die je van jezelf ziet als je tussen twee spiegels staat: een beeld van jezelf in een spiegel waarin een beeld zit van jezelf in een spiegel, waarin enzovoort. Of denk aan de matroesjka-poppetjes, met steeds hetzelfde poppetje in hetzelfde poppetje. Of de foto's in de map verjaardagen in de map 2016 in de map Familie in de map etc. op uw computer.

Deze twee eigenschappen, hiërarchische structuur en recursie, zijn kenmerkend voor menselijke taal in het algemeen, niet alleen voor het uitdrukken van familierelaties, zo blijkt uit taalwetenschappelijk onderzoek. Je kunt bijvoorbeeld eindeloos bijzinnen in bijzinnen stoppen, zoals in de zin: *Ik denk dat Elsa hoopt dat Piet zegt dat Karel weet dat Jan toegeeft dat ... etc.* Zo nu en dan staat er een taalkundige op die zegt een niet-recursieve taal te hebben ontdekt, of die meent te kunnen aantonen

dat taal niet hiërarchisch gestructureerd is maar uitsluitend lineair, maar erg overtuigend zijn de argumenten daarvoor tot op heden niet.

Menselijke taal lijkt in dit opzicht uniek. Dieren hebben ook communicatiesystemen, maar er is nog geen dierlijk communicatiesysteem gevonden dat hiërarchisch en recursief in elkaar zit. Ook jarenlange trainingsprogramma's om slimme dieren zoals apen en papegaaien menselijke taal te leren of een ander communicatiesysteem met dezelfde eigenschappen hebben tot nu toe hopeloos gefaald. Dieren kunnen wel de betekenis van een woord, of, als ze geen geschikte spraakorganen hebben, de betekenis van een gebaar of een ander symbool leren, maar ze kunnen die symbolen niet met elkaar combineren tot hiërarchisch gestructureerde zinnen. Misschien de bultrugwalvissen, zoals sommigen beweren, we zullen zien, meer onderzoek is nodig.

4 Het ziet er voorlopig naar uit dat wij mensen alleen staan in de natuur. Deze eenzaamheid wordt gevangen in de titel van een recent boek van de grondlegger van de moderne taalkunde, de Amerikaanse taalkundige Noam Chomsky. Die titel luidt *Why only us?* Chomsky en zijn coauteur Berwick vragen zich in dit boek af hoe het kan dat alleen mensen zo'n communicatiesysteem hebben, een ongebruikelijke toestand in evolutionair perspectief omdat de evolutie doorgaans geen grote sprongen maakt en een hoge mate van continuïteit en overeenkomst tussen de soorten laat zien. Chomsky concludeert daaruit dat het verschil tussen mensen- en diertaal minimaal moet zijn, en precies daaruit bestaat dat mensen uit symbolen, bijvoorbeeld woorden zoals *Jan*, *Louis* en *van*, een oneindig aantal hiërarchisch gestructureerde zinnen kunnen bouwen, en dieren niet.

We hebben zojuist gezien dat menselijke familierelaties uitgedrukt kunnen worden met de hiërarchische structuur van de taal. Als die hiërarchische structuur ontbreekt in het communicatiesysteem van onze naaste verwanten de apen, betekent dat dan ook dat die apen hun eigen familierelaties niet kennen? Dat zou tamelijk verrassend zijn, want apen hebben er net als

wij voor overleving en voortplanting belang bij te weten welke aap tot hun familie behoort en welke niet. En het blijkt ook niet zo te zijn, apen kennen wel degelijk hun familierelaties. De primatologen Cheney en Seyfarth observeerden in een kolonie vervets in Kenia typische wraakacties die alleen kunnen worden begrepen als deze apen net als mensen hun familierelaties kennen. In één observatie delft een jonge aap in een vechtpartij het onderspit, en twintig minuten later bijt de zus van de verliezer zonder enige aanleiding in de staart van de zus van overwinnaar. Dezelfde primatologen hingen in een experiment achter een struik een luidspreker die het gekrijs van een tweejarige aap weergaf. De vrouwtjes in de groep reageerden hierop door de blik te richten op de moeder van de tweejarige aap. Apen kennen dus hun familierelaties, maar ze kunnen er niet op dezelfde wijze over communiceren als wij. Denken is blijkbaar niet hetzelfde als taal en denken over familierelaties is niet uniek voor mensen, in tegenstelling tot taal.

Moderne syntaxis in de schaduw

Als we begrijpen hoe menselijke taal in elkaar zit, als we de structuur van de menselijke taal kunnen doorgronden, begrijpen we een unieke cognitieve eigenschap van de mens, een eigenschap waar een groot deel van de menselijke cultuur, het recht en de politiek op gebouwd is. Dit is één van de hoofdoelen van de moderne taalwetenschap, gesteld in de jaren vijftig van de vorige eeuw door de eerdergenoemde Chomsky. Ik beperk mij hier tot mijn eigen expertise, de zinsbouw of syntaxis, maar ik denk dat veel van wat ik hierna ga zeggen ook geldt voor andere onderdelen van de taalwetenschap, zoals de betekenis- en klankleer. Door diepgaande analyse van zinspatronen in één taal en vergelijking van die patronen in andere talen gaan we ontdekken welke patronen alle talen met elkaar gemeenschappelijk hebben. Dit moet leiden tot een model van menselijke taal dat kan verklaren hoe het mogelijk is dat elk kind elke moedertaal spelenderwijs kan leren, in al zijn complexiteit, veel ervan nog voor het vier jaar is en voor het eerst naar school gaat. Het model moet ook duidelijk maken hoe dat leerproces verloopt. Anders dan wel eens gedacht wordt,

leren we immers onze moedertaal, het woord zegt het al, niet op school maar van onze moeder en van andere sleutelfiguren in onze vroege kindertijd. Op school leren we vooral lezen en schrijven en moeilijke woorden, maar dat kan pas als we onze moedertaal al geleerd hebben.

Syntaxis is, op deze manier beoefend, een onderdeel van de cognitiewetenschap, het onderzoekt de onbewuste kennis die ten grondslag ligt aan ons taalgebruik, het produceren en begrijpen van zinnen. Het gaat hierbij niet, zoals wel eens gedacht wordt, over de vraag hoe wij zinnen bouwen in ons brein, maar over hoe die zinnen in elkaar zitten. In een vergelijking met het bouwen van een huis: Het gaat over de architectuur van het huis, niet over het daadwerkelijk bouwen ervan met de aanvoer van de materialen, het hijsen, metselen, timmeren etc. Er zijn andere onderdelen van de taalkunde die zich met het daadwerkelijk bouwen van zinnen bezighouden, met andere onderzoeksmethoden, de psycho- en neurolinguïstiek. Die disciplines kijken onder andere naar wat er waar gebeurt in ons brein als wij taal produceren of percipiëren, en zij hebben de resultaten van onder andere syntactisch onderzoek nodig om te weten wat voor soort onbewuste kennis daarbij een rol speelt.

Een andere inspiratiebron van de moderne syntaxis in de jaren vijftig, naast de cognitiewetenschappen, waren de opkomende computerwetenschappen. De hypothese was dat het mogelijk moest zijn om menselijke taal te modelleren als een computerprogramma. Je hebt een database van woorden, een mentaal woordenboek of lexicon, en je hebt een beperkte verzameling samenhangende regels, een algoritme, waarmee je die woorden kunt combineren tot zinnen. Een deel van die regels, zo luidde de hypothese, zou voor alle talen hetzelfde zijn, en een deel zou taalspecifiek zijn. De zinsbouwregels zijn immers niet voor elke taal precies hetzelfde. In het Nederlands zeggen we *Adam heeft een appel gegeten*, met het lijdend voorwerp *een appel* voor het werkwoord *gegeten*, maar die volgorde is fout in het Engels, daar moet het zijn *Adam has eaten an apple*, met het lijdend voorwerp na het werkwoord. De verwachting was dat de syn-

taxis van elke taal en dus ook de syntactische verschillen tussen talen, te beschrijven zouden zijn als een verzameling regels die op een paar A4-tjes zouden passen.

Deze twee invalshoeken maakten de moderne syntaxis geweldig sexy en populair en hun beoefenaren tot beroemdheden en felbegeerde huwelijkspartners. De cognitiewetenschappen en de computerwetenschappen waren jong en veelbelovend. Neurocognitief onderzoek stond nog in de kinderschoenen, technieken om processen in levende breinen te onderzoeken waren nog niet zo geavanceerd als tegenwoordig. Ook de gebruikte retoriek droeg bij aan het aanzien en de populariteit van de moderne syntaxis. Zo dook de term dieptestructuur, een technische term die verwijst naar een bepaald niveau van zinsanalyse, plotseling op in andere wetenschappen in een geheel andere en tamelijk loze betekenis. Een term waar diep voor staat is immers goed in de wetenschap, welke wetenschapper wil er nu niet diep zijn? Ik ben daarom benieuwd hoe diep *deep learning* uiteindelijk blijkt te zijn, al moet ik toegeven dat de recente praktische resultaten indrukwekkend zijn.

En nu, ruim zestig jaar later, is het volgens velen gedaan met het sexappeal van de moderne syntaxis. Hier kom ik met mijn betoog in de schaduw terecht, wat onvermijdelijk is als ik het, uiteindelijk, zoals de titel zegt, over de zonnige toekomst van de Nederlandse taalkunde wil gaan hebben.

Hier volgen enkele kritische uitspraken die de crisis verwoorden. Ik begin er met twee waar ik maar kort op in wil gaan. De eerste luidt: “Als je taal als cognitief verschijnsel wilt begrijpen kun je beter direct onderzoeken hoe taal werkt in het brein, in plaats van via de omweg van onderzoek naar de structuur van taal.” En de tweede: “Alle succesvolle taaltechnologie, zoals automatisch vertalen, information retrieval, tekst-naar-spraak en spraak-naar-tekst-systemen, vraag-antwoordsystemen, spraak in navigatiesystemen, werkt op basis van statistiek, niet op basis van taalkundige regels en algoritmes.”

De eerste uitspraak, de gedachte dat je menselijke taal als cognitief verschijnsel kunt begrijpen door taalkundige analyse over te slaan en rechtstreeks onderzoek te doen naar taal in het brein berust op een denkfout, één die ik hiervoor eigenlijk al besproken heb. Neurocognitief onderzoek heeft de resultaten van onder andere syntactisch onderzoek nodig om te weten wat het moet onderzoeken. Het huidige neurocognitieve onderzoek maakt nog te weinig gebruik van de resultaten van 60 jaar syntactisch onderzoek en de onderzoeksvragen die daaruit voortvloeien. Eén van de oorzaken daarvan is dat deze resultaten en vragen soms subtieler zijn dan wat op dit moment met neurocognitieve methoden kan worden gemeten.

Ook op de tweede kritische uitspraak, er is alleen statistiek en geen taalkunde nodig in goedwerkende taaltechnologie, valt wat af te dingen. Het klopt dat succesvolle taaltechnologie hoofdzakelijk gebaseerd is op statistische analyse van grote hoeveelheden data. Maar taaltechnologie kan nog steeds essentiële dingen niet die mensen wel kunnen. Taaltechnologie begrijpt nog steeds niets van taal. Neem Google Translate. Elke keer als er weer een bericht in de media verschijnt dat Google Translate nu echt perfecte vertalingen aflevert toets ik een simpel zinnetje in om te kijken hoever ze zijn. Dat zinnetje luidt: *Waar denk je dat hij de auto mee repareert?* Google Translate vertaalt dit zinnetje als volgt naar het Engels: *Where do you think he will repair the car?* een zin die iets heel anders betekent. De Nederlandse zin kan bijvoorbeeld als antwoord krijgen *Met een schroevendraaier*, terwijl de Engelse zin beantwoord kan worden met *In de garage*. Google Translate ziet niet dat het woord *waar* in de ingetoetste zin bij het woord *mee* hoort. Dat komt doordat die twee woorden in de testzin niet naast elkaar staan. Het programma negeert vervolgens de aanwezigheid van *mee* en slaat er een slag naar. Taaltechnologie werkt kortom vooralsnog heel anders dan menselijke taal. Er zal ongetwijfeld een moment komen dat taaltechnologie bijna perfect werkt. Maar wetenschappelijk onderzoek naar menselijk taal is er in eerste instantie op gericht te begrijpen hoe menselijk taal in

elkaar zit en werkt, en het is niet zo waarschijnlijk dat hedendaagse taaltechnologie daarvoor een goed model is.

Een derde punt van kritiek sluit hier direct op aan. Hier is de gedachte dat ook voor het onderzoek naar menselijke taal zelf alleen maar statistische analyse van grote hoeveelheden taaldata nodig is, uit te voeren door de computer. De rol van de onderzoeker zou hier zo goed als uitgespeeld zijn en een hypothetisch-deductieve benadering, waarin je de structuur van taal onderzoekt door een theorie over taal te toetsen aan steeds nieuwe constructies en talen, overbodig. Een kind dat zijn moedertaal leert zou ook maar gewoon een kleine statisticus zijn die generalisaties afleidt uit de taal in zijn of haar omgeving. De computer en het kind zouden volgens dit scenario ontdekken dat taal bestaat uit een grote hoeveelheid patronen die allemaal opgeslagen worden in een netwerk in het menselijke geheugen. En dat zou betekenen dat een taal niet kan worden beschreven als een algoritme, waarmee de oorspronkelijke doelstelling van de syntaxis zou komen te vervallen.

Dat een kind niet alleen maar een kleine na-aper is blijkt uit taalverwervingsonderzoek waar Caitlin Meyer, Fred Weerman en ik dit jaar over publiceren. We onderzochten in welke volgorde kleuters de rangtelwoorden *eerste*, *tweede*, *derde*, *vierde* etc. verwerven. Als die volgorde van verwerving zou corresponderen met de frequentie van elk van deze rangtelwoorden in het taalaanbod dat zouden we verwachten dat *eerste* eerst wordt verworven, dan *tweede*, dan *derde* en dan *vierde*, want hoe hoger het rangtelwoord hoe lager de gebruiksfrequentie. We vinden echter dat het onregelmatige *derde* significant later wordt verworven dan *eerste*, *tweede* en *vierde*. En in het Engels wordt het nog onregelmatigere *second*, het op één na frequentste rangtelwoord, later verworven dan bijvoorbeeld *fourth*. Nederlandse kinderen gebruiken in het stadium waarin ze *derde* niet begrijpen zelf *driede*. Dit alles suggereert dat kinderen de rangtelwoorden volgens een regel verwerven, niet per woord, en dat het ze in het begin niet duidelijk is hoe *derde* en *second* onder die regel kunnen vallen.

Een ander voorbeeld dat taalverwerving meer is dan na-aperij betreft onderzoek waar Marjo van Koppen en ik enkele jaren geleden een artikel over hebben geschreven. Dit onderzoek begon voor mij met de observatie dat mijn dochters Annika en Jitske in de peuterfase gedurende korte tijd af en toe zinnnetjes produceerden zoals *Dan noem-ik-te jou Sinterklaas*, waarbij de volgorde van de verledentijdsuitgang en het onderwerp van de zin, *ik*, abusievelijk zijn omgedraaid. In volwassenen Nederlands zeggen we: *Dan noem-de ik jou Sinterklaas*, met eerst het werkwoord *noem*, dan de verledentijdsuitgang - *de* en dan het onderwerp van de zin *ik*. *Noem* en de uitgang - *de* horen dus naast elkaar te staan, maar mijn dochters zetten daar het onderwerp *ik* tussen. Dat ze niet zomaar iets deden maar bij hun verwisseling het klanksysteem van het Nederlands onbewust toepasten kan je zien aan het feit dat ze de uitgang - *de* na *ik* als - *te* uitspraken. De slotklank /k/ van *ik*, die direct voor de uitgang staat, zit immers in 't Kofschip en de beginklank van de verledentijdsuitgang is daar gevoelig voor.

Op het eerste gezicht leek deze verwisseling van verledentijdsuitgang en onderwerp een grappige taalfout, een verspreking of misschien wel het symptoom van een taalstoornis. Maar het verschijnsel bleek bekend uit de kindertaalliteratuur, kinderen doen dit in een fase waarin ze ook andere woorden uit elkaar zetten die eigenlijk naast elkaar moeten staan. In het proefschrift van Jacqueline van Kampen zijn daar tal van voorbeelden van te vinden. In plaats van *Mag ik proeven hoe heet het is* zegt de vierjarige Laura bijvoorbeeld *Mag ik proeven hoe het heet is*, met *hoe* en *heet* gesplitst in plaats van direct naast elkaar. Zulke woordvolgordes horen kinderen natuurlijk niet van hun opvoeders, ze maken ze zelf. Kinderen apen niet alleen maar de syntactische patronen na die ze in hun taalomgeving horen, en we moeten verklaren hoe dat kan.

Het patroon van de *Dan noem-ik-te jou Sinterklaas* zin bleek vervolgens ook voor te komen in 6 van de 267 dialecten van het Nederlands die wij in de jaren 2000 aan het Meertens Instituut aan het onderzoeken waren. Bij statistische analyse zonder

theorie zouden deze 6 dialecten ongetwijfeld als ruis uit de data zijn gefilterd, los van het feit dat er nog geen dialect-tekst-corpora bestaan waarin een computer dit soort verschijnselen kan opsporen. De meeste dialecten zijn namelijk alleen gesproken talen en er is nog geen spraak-naar-tekst software die alle dialectvariatie aankan. Deze verschijnselen kunnen alleen gevonden worden met experimenteel onderzoek, door veldwerk.

We zouden nu kunnen volstaan met te constateren dat kinderen en sommige dialecten blijkbaar een van het volwassen Standaardnederlands afwijkende constructie in hun zinsbouwrepertoire hebben, een aardig feit, en dan overgaan tot de orde van de dag. Het verschijnsel wordt echter pas interessant in het licht van de Universal Base Hypothesis, de hypothese dat in alle talen van de wereld de zinsbouw op een abstract niveau hetzelfde is. De *Dan noem-ik-te jou Sinterklaas* zin ondersteunt deze hypothese op verrassende wijze.

Dat zit zo. Een goede kandidaat voor een zinsbouwprincipe dat alle talen en dialecten delen betreft de posities in de zin waar een werkwoord kan staan. In het Nederlands en zijn dialecten zijn dat er op het eerste gezicht twee, één positie vooraan in de zin en één positie achteraan. In de zin *Zij wilde vanmorgen een lied zingen* staat de persoonsvorm *wilde*, de vorm van het werkwoord die vervoegd is voor de verleden tijd, op de tweede plaats in de zin, na het onderwerp *Zij*. Het hele werkwoord *zingen* staat achteraan in de zin. Persoonsvorm en hele werkwoord kunnen niet van plaats wisselen, je kunt niet zeggen *Zij zingen vanmorgen een lied wilde*. Je kunt ook niet van *willen* het hele werkwoord maken en van *zingen* de persoonsvorm, dan krijg je de foute zin *Zij willen vanmorgen een lied zong*. Als je het hulpwerkwoord *willen* uit de zin verwijdert dan moet *zingen* de persoonsvorm worden en op de tweede plaats staan, dan krijgen we *Zij zong vanmorgen een lied*. Er is, met andere woorden, een relatie tussen de vorm van het werkwoord en de plaats waar het kan staan in de zin. Alleen de persoonsvorm kan op de tweede plaats in de zin, en het hele werkwoord moet altijd achteraan in de zin.

Taalvergelijkend onderzoek heeft laten zien dat deze relatie tussen de vorm van het werkwoord en mogelijke positie in de zin in veel talen bestaat. Een theorie van de syntaxis van menselijke taal moet begrijpelijk maken waarom dit zo is. Datzelfde taalvergelijkende onderzoek heeft echter ook aangetoond dat veel talen minimaal drie posities voor werkwoorden in de zin hebben, in plaats van de twee die we tot nu toe in het Nederlands hebben gevonden. Op grond van de Universal Base Hypothesis verwachten we dan dat het mogelijk is om aan te tonen dat er ook in het Nederlands drie posities voor het werkwoord zijn. Jan-Wouter Zwart heeft in zijn boek met een aantal argumenten laten zien dat dit inderdaad het geval is. Het eenvoudigste daarvan is dat we naast *Zij wilde vanmorgen een lied zingen* ook nog de vraagzinvolgorde *Wilde zij vanmorgen een lied zingen* hebben, met de persoonsvorm *wilde* voor het onderwerp *zij* in plaats van *erna*. We kunnen hieruit concluderen dat ook het Nederlands drie zinsposities heeft voor werkwoorden: voor het onderwerp, na het onderwerp en achteraan in de zin. De twee posities voor en na het onderwerp zijn alleen toegankelijk voor persoonsvormen.

Kijken we nu wat beter naar onze *Dan-noem-ik-te-jou-Sinterklaas* zin dan zien we hierin de twee posities van de persoonsvorm direct terug. Voor het onderwerp *ik* staat de stam van de persoonsvorm, *noem*, en na het onderwerp staat de uitgang van de persoonsvorm, *-te*. In de *Dan-noem-ik-te-jou-Sinterklaas* zin kunnen we dus de twee mogelijke posities van de persoonsvorm vooraan in de zin tegelijk gevuld zien, iets wat nooit zichtbaar is in het volwassen Standaardnederlands, waar altijd maar één van de twee posities zichtbaar is. Mijn dochters deden niet zomaar wat, ze maakten gebruik van universeel beschikbare zinsposities. En sommige dialecten van het Nederlands doen dat ook.

Kinderen die hun moedertaal leren, leren dus niet alleen door imitatie maar passen ook algemene principes toe. Dit blijkt uit het feit dat ze constructies maken die niet in het taalaanbod zitten maar die wel consistent zijn met die principes. We kunnen die principes ontdekken door zoveel mogelijk talen en dialecten te vergelijken.

Misschien kan de computer ons daarbij helpen en dat zou mooi zijn, want er zijn zoveel talen, dialecten en constructies en die zijn vaak zo complex, dat het zonder hulp van de computer nog heel lang kan duren voor we begrijpen hoe de syntaxis van natuurlijk taal in elkaar zit. In het promotieproject dat Martin Kroon hier in Leiden uitvoert onderzoeken we daarom in samenwerking met de centers voor digital humanities en voor data science of de computer automatisch syntactische verschillen kan detecteren door analyse van parallelle corpora zoals Europarl, de online beschikbare in 21 talen vertaalde verslagen van het Europese parlement.

De al of niet met behulp van de computer ontdekte algemene zinsbouwprincipes fungeren vervolgens als hypothese bij het bestuderen van een nieuwe taal of constructie. Deze hypothesen fungeren op hun beurt als een zoeklicht, zonder deze hypothesen zien we de interessantste verschijnselen, zoals de verwisseling van onderwerp en verledentijdsuitgang, gemakkelijk over het hoofd en als we ze al zien hebben ze geen betekenis in het grotere geheel. Deze hypothesen voorspellen ook dat sommige taalstructuren, hoewel denkbaar, nooit zullen voorkomen in natuurlijke taal. We hebben bijvoorbeeld nog geen talen of dialecten gevonden waarin de verledentijdsuitgang op een positie stond waar de persoonsvorm niet kan staan.

Aan het begin van een zonnige toekomst

Mogen we nu uit het voorafgaande concluderen dat het moderne syntactische onderzoek al sinds de jaren vijftig van de vorige eeuw op de juiste weg is en vernieuwing niet nodig is? Het eerste wel, het tweede niet. De afgelopen jaren is er een uitstekende uitgangssituatie gecreëerd om Nederland en het Nederlands, inclusief zijn dialecten, internationaal een centrale rol te laten spelen in het onderzoek naar natuurlijke taal. Het Nederlands is waarschijnlijk de best beschreven taal ter wereld en de Nederlandse taalkunde beschikt over een internationaal onvergelijkbaar geavanceerde digitale onderzoeks-infrastructuur.

Zo is recent de *Syntax of Dutch* beschikbaar gekomen in het online Taalportaal. De *Syntax of Dutch* is een monumentale, meer dan 4000 pagina's tellende beschrijving van syntactische eigenschappen van het algemeen Nederlands, grotendeels geschreven door Hans Broekhuis op basis van het syntactisch onderzoek van de afgelopen 60 jaar. Ik benadruk graag nog eens dat die 4000 pagina's de onbewuste grammaticakennis beschrijven van een moedertaalspreker van het Nederlands en dat veel van die kennis al bij een vierjarige kleuter aanwezig is, dus niet op school geleerd. Deze 4000 pagina's leveren een interessante paradox op. U herinnert zich dat het oorspronkelijke doel van modern syntactisch onderzoek was om elke taal te beschrijven als een algoritme van een paar A4-tjes. Als er 4000 pagina's nodig zijn om nog maar een deel van de syntactische eigenschappen van het Nederlands te beschrijven, dan lijkt deze onderneming mislukt.

Toch is dat niet het geval. Wie de *Syntax of Dutch* raadpleegt zal zien dat er hele hoofdstukken zijn die net zo goed over een andere taal hadden kunnen gaan, omdat die taal in dat domein identiek is aan het Nederlands. Dat geldt bijvoorbeeld voor de relatieve plaatsing van verschillende typen bijwoordelijke bepalingen, zoals de zinsbepaling *waarschijnlijk* en de predicaatsbepaling *zuiver* in de zin *Zij heeft het lied waarschijnlijk zuiver gezongen*. De onderlinge volgorde, of beter hiërarchie van verschillende typen bijwoordelijke bepalingen ligt vast, de zin *Zij heeft het lied zuiver waarschijnlijk gezongen* is geen goede zin van het Nederlands. Er zijn meer dan dertig verschillende typen bijwoordelijke bepalingen en in taal-na-taal-na-taal vertonen ze dezelfde onderlinge volgorde-restricties, zo heeft de Italiaanse syntacticus Guglielmo Cinque laten zien. Om dit te kunnen ontdekken zijn uitgebreide beschrijvingen nodig zoals de *Syntax of Dutch*, die daarmee het doel van een algoritme van een paar A4-tjes juist dichterbij brengen.

In het Taalportaal zijn of worden parallelle beschrijvingen van de nauw met het Nederlands verwante talen Afrikaans en Fries beschikbaar gemaakt, hetgeen gedetailleerde vergelijking

mogelijk maakt en beantwoording van de vraag of een taal een verzameling toevallige syntactische eigenschappen is of een systematisch samenhangende verzameling regels waarvan de kern gedeeld wordt door alle talen in de wereld. Diezelfde vraag kan onderzocht worden met behulp van de gegevens van 267 dialecten van het Nederlands die we met een groot Nederlands-Belgisch team verzameld hebben in de jaren 2000 in het project Syntactische Atlas van de Nederlandse dialecten, beter bekend als de SAND. Ook deze data-verzameling is internationaal zonder precedent. De data zijn online beschikbaar, doorzoekbaar en karteerbaar in applicaties als DynaSAND en de CLARIN-tool MIMORE en hebben de afgelopen jaren geleid tot een golf van onderzoek naar syntactische microvariatie, veelal in de vorm van spraakmakende proefschriften zoals die van Jeroen van Craenenbroeck, Marjo van Koppen, Hedde Zeijlstra, Gunther de Vogelaer, Marco René Spruit en Eefje Boef.

Dit type onderzoek is volstrekt nieuw, zeker waar het zich concentreert op de relatie tussen de geografische distributie van syntactische eigenschappen en taal als een cognitief systeem, twee tot voor kort gescheiden onderzoeksdisciplines. In een vorig jaar verschenen artikel laten Marjo van Koppen, Norbert Corver, Hans Bennis en ik aan de hand van syntactische verdubbelingverschijnselen zoals *de* en *die* in *Ik heb d'n dieën gezien* en *-de* en *gij* in *Hedde gij da gedaan*, zien dat Vlaams, Brabants en Nederlands elk hun eigen grammaticasysteem hebben, en hoe die systemen in elkaar overgaan als bepaalde eigenschappen veranderen. Dit type onderzoek is ook nieuw omdat het methodologie van de sociolinguïstiek combineert met die van de theoretische taalkunde. Een andere innovatie en sprong voorwaarts is een combinatie van kwantitatieve en theoretische analyse van syntactische variatie, zoals in het recente werk van Jeroen van Craenenbroeck.

Dit soort onderzoek maakt het mogelijk een potentiële bedreiging voor syntactisch onderzoek, een bedreiging waar ik het nog niet over gehad heb, om te vormen tot een fascinerend onderzoeksprogramma. Deze keer gaat het om een bedreiging

van binnenuit. U herinnert zich dat in het boek van Berwick en Chomsky beargumenteerd wordt dat er maar een heel klein verschil kan zijn tussen menselijke taal en dierlijke communicatie. De combinatorische operatie die dieren missen noemen zij Merge, het samenvoegen van twee woorden of woordgroepen tot een groter geheel. Merge zou volgens hen het enige zinsbouwprincipe zijn dat nodig is voor de beschrijving van de structuur van menselijke taal. Geen 4000 pagina's beschrijving, ook geen algoritme van een paar A4-tjes, maar slechts één simpele combinatorische regel. Dit levert een volgende interessante paradox op.

Als alle talen slechts één en dezelfde zinsbouwregel hebben dan kan er strikt genomen geen syntactische variatie bestaan. Maar we hebben hiervoor gezien dat er juist veel syntactische verschillen zijn, verschillen in woordvolgorde, verdubbeling, splitsing, niet alleen tussen talen maar zelfs tussen dialecten van dezelfde taal. Dan moet het zo zijn dat syntactische verschillen tussen talen het resultaat zijn van de wisselwerking tussen die ene regel en andere modules van de menselijke grammatica, zoals het mentale lexicon, de in ons geheugen opgeslagen woordenschat, en de module waarin abstracte syntactische structuren vertaald worden naar klank- of gebarenreeksen. Dat is precies wat we in het eerder aangehaalde artikel over het Vlaams, Brabants en Nederlands laten zien.

Het onderzoeken van deze interacties brengt het doel, het begrijpen van menselijke taal als cognitief systeem, dichterbij, omdat hierbij gekeken moet worden hoe verschillende typen informatie, lexicale informatie, syntactische informatie, betekenisinformatie en klankinformatie op elkaar afgebeeld worden en hoe daarbij schijnbare syntactische variatie ontstaat. De SAND-data en software-tools vormen daarvoor een ideaal laboratorium, waarin de gevolgen van een klein verschil tussen twee dialecten voor de gehele grammatica's van deze dialecten kunnen worden onderzocht. Een groot voordeel van dit onderzoeksprogramma is dat syntactische verschillen niet altijd syntactisch hoeven te worden verklaard, wat tot op zekere

hoogte een beperking was van het oude onderzoeksprogramma waarin nog verondersteld werd dat syntactische variatie uitgedrukt werd in specifieke syntactische regels. Het betekent ook dat syntactische overeenkomsten tussen talen, zoals de hiërarchische ordening van verschillende typen adverbia of de drie posities voor werkwoorden, niet alleen maar kunnen worden begrepen in termen van de syntactische module van de grammatica.

Te verwachten valt daarom dat een deel van het onderzoek naar syntactische variatie zal verschuiven naar variatie in het lexicon. De integratie van alle belangrijke digitale woordenboeken van het Nederlands van het Instituut voor de Nederlandse Taal in de nationale geesteswetenschappelijke onderzoeksinfrastructuur CLARIAH zal dit onderzoek geweldig faciliteren. Ook het NWO-groot project Nederlab zal een plaats krijgen in deze infrastructuur. Als Nederlab binnenkort klaar is wordt het mogelijk syntactische variatie niet alleen in de geografische ruimte maar ook in de tijd, en dus syntactische verandering, systematisch te onderzoeken, op basis van de grote verzameling historische teksten en met behulp van de software-tools die Nederlab beschikbaar maakt.

De situatie voor het Nederlandse taalgebied is bij mijn weten uniek in de wereld. Nergens in de wereld bestaan van een taal en zijn dialecten zulke omvangrijke, gevarieerde, goed toegankelijke diachrone en synchrone digitale dataverzamelingen, systematische beschrijvingen en taalatlassen, ingebed in een digitale infrastructuur voor de geesteswetenschappen. En ook de wijze waarop de Nederlandse taalkunde georganiseerd is in de Landelijke Onderzoeksschool Taalkunde is internationaal volstrekt uniek. Alle voorwaarden zijn kortom aanwezig om de Nederlandse taalkunde en de Nederlandse taal een centrale rol te laten spelen bij het verder ontsluiten van het taalsysteem.

Hier zijn de studenten, de promovendi en de postdocs aan zet, jullie moeten profiteren van deze mogelijkheden en voor doorbraken gaan zorgen. Ik zou tegen jullie het volgende willen

zeggen. Een studie of baan aan de universiteit hoort geestverruimend te zijn, zeker aan onze faculteit. Ik bedoel hier niet jullie buitenschoolse en -werkse activiteiten, wat ik bedoel is dat je door veel lezen, hard werken, diep nadenken, slim onderzoek, heftig debatteren en de juiste coaching werelden kunt ontdekken die voor anderen gesloten blijven, zoals de cognitieve wereld achter de Nederlandse taal. Dat is waar universiteiten in de eerste plaats voor zijn, niet voor het opdoen van zulke alledaagse dingen als 21st century en transferable skills.

Ik kom tot een afronding. Moderne wetenschap is teamwork, veelal interdisciplinair. Het eenzame genie is een romantische beeld uit vervlogen tijden. Ik ben aan veel mensen dank verschuldigd. Zonder alle mensen waarmee ik de afgelopen 25 jaar op allerlei wijzen heb mogen samenwerken had ik hier nu niet gestaan. Ik kan onmogelijk iedereen noemen, maar wil voor een aantal toch een uitzondering maken. In chronologische volgorde. Mijn leermeesters Hans Bennis en wijlen Teun Hoekstra, mijn eigen gidsen naar de wereld die schuil gaat achter alledaagse taal. Zij hebben het belangrijkste gedaan wat een leermeester kan doen, de verwondering wekken die nodig is om levenslang onderzoek te doen. De taalkundigen en technici van het Meertens Instituut, tijdelijk en vast. De afgelopen 17 jaar hebben we daar met z'n allen in een zeldzaam plezierig onderzoeksklimaat de wat ik eerder genoemd heb totaal-taalkunde uitgevonden en tot ontwikkeling gebracht, bijvoorbeeld in het SAND-project, en een belangrijke bijdrage geleverd aan de totstandkoming van een digitale onderzoeksinfrastructuur voor de taalkunde. Veel dank ook aan Norbert Corver en Marjo van Koppen voor 10 jaar lang vriendschappelijke en inspirerende samenwerking bij de opleiding Nederlands van de Universiteit Utrecht.

Mijn terugkeer in Leiden vorig jaar voelde als thuiskomen. Ik wil alle taalkundigen en medewerkers van LUCL en alle collega's bij de opleiding Nederlands bedanken voor de hartelijke ontvangst en de constructieve en aangename werksfeer die ik het afgelopen jaar heb mogen ervaren. Ik verwacht veel van

onze verdere samenwerking en verheug me erop. Ik wil ook graag het College van Bestuur bedanken, het bestuur van de faculteit, directie van LUCL en alle anderen die hebben bijgedragen aan de totstandkoming van deze benoeming.

De belangrijkste personen in mijn leven bedank ik als laatste. Anneke, Daan, Koen, Annika, Jitske, jullie zijn mijn grote inspiratiebron. Ten slotte bedank ik mijn ouders, voor alles. Mijn moeder is er helaas niet meer bij, maar ze heeft nog wel mijn benoeming meegemaakt en zei bij die gelegenheid: Goh jongen, zijn die taalproblemen nou nog niet opgelost? We doen ons best, moeder, en ik denk dat het de goede kant opgaat.

Ik heb gezegd.

Bibliografie

- Barbiers, Sief, Marjo van Koppen, Hans Bennis en Norbert Corver, 2016. Microcomparative Morphosyntactic Research (MIMORE): Mapping partial grammars of Flemish, Brabantish and Dutch. *Lingua* 178, 5-31.
- Barbiers, Sief en Marjo van Koppen, 2006. Een plaats voor tijd in het middenveld van het Nederlands. *Taal & Tongval*, (Themanummer 19), 24-39.
- Barbiers, Sief, 2013. Microsyntactic Variation. In Marcel den Dikken (ed.) *The Cambridge Handbook of Generative Syntax*. Cambridge Handbooks in Language and Linguistics. Cambridge: Cambridge University Press.
- Barbiers, S., J. van der Auwera, H. Bennis, E. Boef, G. De Vogelaer, M. van der Ham, 2008. *Syntactic Atlas of the Dutch Dialects Volume II*. Amsterdam: Amsterdam University Press.
- 12 Barbiers, S., H. Bennis, G. de Vogelaer, M. Devos and M. van der Ham, 2005. *Syntactic Atlas of the Dutch Dialects Volume I*. Amsterdam: Amsterdam University Press.
- Berwick, Robert and Noam Chomsky, 2015. *Why only us? Language and Evolution*. Cambridge: MIT Press.
- Boef, Eefje, 2013. Doubling in Relative Clauses. Aspects of morphosyntactic microvariation in Dutch. Diss. Utrecht. LOT Dissertations 317.
- Broekhuis, Hans, et al., 2012-2016. *Syntax of Dutch*. 7 volumes. Comprehensive Grammar Resources. Amsterdam: Amsterdam University Press.
- Cheney, Dorothy and Robert Seyfarth, 1992. The representation of social relations by monkeys. *Cognition* 37, 167-196.
- Cinque, Guglielmo, 1999. *Adverbs and Functional Heads. A Cross-Linguistic Perspective*. Oxford Studies in Comparative Syntax. New York/Oxford: Oxford University Press.
- De Vogelaer, Gunther, 2008. *De Nederlandse en Friese subjectmarkerders: geografie, typologie en diachronie*. Diss. Gent. Gent: Koninklijke Academie voor Nederlandse Taal- en Letterkunde.
- Elsschot, Willem, 1921. *De Verlossing*. In W. Elsschot, *Verzameld werk*. Amsterdam, Querido, 1986.
- Kampen, Jacqueline van, 1997. *First Steps in Wh-movement*. Dissertation Utrecht University.
- Koppen, Marjo van, 2005. *One Probe-Two Goals: Aspects of agreement in Dutch dialects*. Diss. Leiden. LOT Dissertations 105.
- Meyer, Caitlin, Sief Barbiers and Fred Weerman, 2017. Ordinals are not as easy as one, two, three. The acquisition of cardinals and ordinals in Dutch. *Language Acquisition*. <http://dx.doi.org/10.1080/10489223.2017.1391266>
- Spruit, Marco René, 2008. *Quantitative perspectives on syntactic variation in Dutch dialects*. Diss. UvAmsterdam. LOT Dissertations 174.
- Van Craenenbroeck, Jeroen, 2004. *Ellipsis in Dutch dialects*. Diss. Leiden. LOT Dissertations 96.
- Van Craenenbroeck, Jeroen, 2017. A quantitative analysis of syntactic microvariation. Word order in Dutch verb clusters. Ms. KU Leuven.
- Zeijlstra, Hedde, 2004. *Sentential Negation and Negative Concord*. Diss. UvAmsterdam. LOT Dissertations 101.
- Zwart, Jan Wouter, 1997. *Morphosyntax of Verb Movement. A Minimalist Approach to the Syntax of Dutch*. Studies in Natural Language and Linguistic Theory 39. Dordrecht: Kluwer Academic Publishers.

PROF.DR. SJEF BARBIERS (AMSTERDAM, 1959)



- 2016- Hoogleraar Nederlandse Taalkunde, Leiden University Centre for Linguistics, Universiteit Leiden
- 2005-2015 Bijzonder hoogleraar Variatielinguïstiek van het Nederlands, Universiteit Utrecht
- 2000-2016 Senior-onderzoeker Variatielinguïstiek, Meertens Instituut (KNAW), Amsterdam
- 1997-2000 Postdoc en Universitair Docent, Opleiding Dutch Studies, Universiteit Leiden
- 1996-1997 Postdoc MIT, Cambridge, USA (Niels Stensen Stipendium)
- 1995 Promotie Universiteit Leiden, cum laude (The Syntax of Interpretation)
- 1991-1995 Assistent in Opleiding, Universiteit Leiden
- 1990 Doctoraal Nederlandse Taal- en Letterkunde, cum laude, Universiteit Leiden
- 1986-1990 Studie Nederlandse Taal- en Letterkunde, Universiteit Leiden

Syntaxis is een subdiscipline van de taalwetenschap die onderzoekt volgens welke principes zinnen in natuurlijke taal worden opgebouwd uit morfemen, woorden en woordgroepen. Dit onderzoek geschiedt vanuit een taalvergelijkend perspectief en tegen de achtergrond van de hypothese dat alle talen een aantal abstracte zinsbouwprincipes met elkaar gemeen hebben. Als we begrijpen hoe syntaxis in elkaar zit, begrijpen we een centraal kenmerk van de menselijke cognitie. Een belangrijke recente ontwikkeling in dit domein is de vergelijking van de zinsbouw van grote hoeveelheden nauw verwante talen, zoals de dialecten van het Nederlands. Deze methode maakt het mogelijk syntactische variatie, zoals woordvolgordevariatie, verdubbelingen en weglatingen te herleiden tot de wisselwerking tussen lexicale eigenschappen, zinsbouwprincipes, het betekenisstelsel en het klanksysteem.



Universiteit
Leiden