

Beginnen aan een eindverslag

Over impliciete activiteiten bij preposities en nomina

Michiel de Winter
Begeleider: Joost Zwarts
25 juli 2011

Samenvatting

Deze scriptie zal zich richten op de invloed van preposities op de implicatie van activiteiten bij zinnen met het werkwoord *beginnen* gecombineerd met nomina. Hierbij wordt specifiek gekeken naar de preposities *aan*, *met* en *in*. Met behulp van Pustejovsky's qualiastructuur, wordt een hypothese opgesteld die deze preposities en VP's zonder preposities verbindt aan bepaalde qualiarollen. Dit gebeurt op basis van data uit Google Search en het nomen *boek*. Hierna wordt deze hypothese getest op andere nomina: *relatie* en *winkel*.

Nadat blijkt dat deze hypothese wordt niet wordt ondersteund, zal worden gezocht naar een nadere categorisatie van nomina, in de hoop betere ondersteuning voor de hypothese te vinden. Hoewel deze uitblijft, zal blijken dat er andere contrasten te vinden zijn binnen de nomina, zoals de speciale rol van de prepositie *in*, die containerfuncties met verschillende werkingen veroorzaakt. Omdat de hypothese niet alleen niet ondersteund wordt, maar ook niet ontkracht, worden er enkele voorstellen gedaan om dit onderzoek uit te breiden, voor het vinden van resultaten die het wellicht wel de mogelijkheid geven de hypothese te bevestigen danwel te ontkrachten.

Inhoud

1. Inleiding	4
2. Probleemstelling en wetenschappelijk kader	6
2.1. Moeilijkheden met preposities in de Nederlandse taal	6
2.2. Pustejovsky en de qualiastructuur	6
2.3. Vraagstelling	9
3. Onderzoek	11
3.1. Onderzoeksmethode	11
3.1.1. <i>Onderzoek naar het Nederlands</i>	11
3.1.2. <i>Onderzoek op basis van Google Search Engine</i>	11
3.1.3. <i>Zoektermen</i>	12
3.2. Analyse	12
3.2.1. <i>Een boek beginnen</i>	12
3.2.2. <i>Een relatie/winkel beginnen</i>	15
3.2.3. <i>Tussentijdse conclusie</i>	17
3.3. Categorisatie van nomina	17
3.4. Nadere categorisatie	20
3.4.1. <i>Nomina uit de categorie object</i>	20
3.4.2. <i>Containerconstructies met in</i>	21
4. Conclusie	23
4.1. Verband tussen preposities en implicatie van activiteit	23
4.2. Positieve resultaten	23
4.3. Mogelijkheden voor vervolgonderzoek	24
Referenties	25

1. Inleiding

Voor het creëren van een Kunstmatige Intelligentie die in staat is met mensen te communiceren, is het van cruciaal belang dat er begrip is van de door mensen gebezigde taal. Taal is immers de manier waarop mensen onderling communiceren, en waaraan de mens haar grote ontwikkeling ten opzichte van andere diersoorten te danken heeft. Hoewel deze taal voor mensen heel natuurlijk en vanzelfsprekend is, blijft het vooralsnog een zeer groot probleem om het accuraat aan een computer te leren. Oorzaak hiervan is het bestaan van vele uitzonderingen en ingewikkelde constructies die voor een computer moeilijk te volgen zijn. Onderzoek naar dit probleem kan van twee kanten bekeken worden: de taalkundige, waarbij bekeken wordt hoe de taal precies in elkaar zit en uit welke regels deze is opgebouwd die aan de computer gegeven kunnen worden; of de cognitieve, waarbij de processen bij de mens worden bekeken in een poging dit proces na te bootsen, en zo de computer de taal op een 'natuurlijke' manier laten leren.

Een van de beperkingen van computers wanneer zij zich bezighouden met natuurlijke taal, is dat zij slechts kunnen werken met de regels die door de programmeur zijn ingegeven. Natuurlijke taal is echter meer dan regels: mensen interpreteren de woorden van een ander en kunnen gevoelsmatig meer informatie krijgen dan er expliciet is uitgedrukt door de taal.

Voorbeeld 1

- a) Ik begin een boek
- b) Ik begin aan een boek
- c) Ik begin met een boek
- d) Ik begin in een boek

Een voorbeeld van dergelijke impliciete betekenis is te zien in voorbeeld 1. Bij 1a en b is hier sprake van het schrijven van een boek, bij 1d van het lezen ervan en bij 1c zijn beide mogelijkheden aanwezig. Waar een moedertaalspreker van het Nederlands deze betekenis direct uit de zinnen kan halen, zullen het voor een computer onbegrijpelijke zinnen zijn. Het werkwoord *beginnen* impliceert hier immers een bezigheid, maar een boek is een object. Dit past niet bij elkaar, en maakt dat de computer een foutmelding zal geven.

Al sinds de jaren vijftig, toen Noam Chomsky zijn eerste werk over generatieve grammatica's publiceerde, wordt er gewerkt aan regelsystemen waarmee de complexiteit van natuurlijke taal wordt opgevangen. Hoewel dit inmiddels vele successen heeft opgeleverd – denk bijvoorbeeld aan spelling- en grammaticacontrole en chatbots – blijven er nog steeds veel onopgeloste problemen gerelateerd aan uitzonderingen op vaste regels en complexe constructies zoals de hierboven genoemde impliciete betekenissen. Een veel gebruikte methode om deze gevallen op te vangen, is *brute force*. Hierbij worden per geval extra entries gemaakt in de taaldatabase van een programma, om zo de verschillende gebruiken van een woord te ondervangen. Dit resulteert echter vaak in gigantische databases, waar bovendien het risico groeit op dubbelingen, en incorrecte combinaties. Een betere oplossing, hoewel moeilijker te vinden, is het vinden van hogere regels, waarmee dit verschillende gebruik kan worden beperkt tot een aantal instanties van hetzelfde woord.

Ook voor impliciete betekenissen zoals te zien in voorbeeld 1 is er de mogelijkheid om de verschillende opties stuk voor stuk te programmeren. Men zou bijvoorbeeld kunnen denken aan het los invoegen van de combinaties van *beginnen*, *aan* en *boek* en de bijbehorende betekenis *schrijven*. Niet alleen brengt dit een zeer grote hoeveelheid entries met zich mee, is dit echter ook ingewikkeld te programmeren. Hoewel er altijd uitzonderingen zullen blijven bestaan, is het daarom van belang om toch lijnen te vinden hoe dergelijke impliciete betekenis wordt gevormd. Deze scriptie zal pogen dergelijke lijnen te vinden en zo bij te dragen aan het verklaren van de werking van natuurlijke taal.

2. Probleemstelling en wetenschappelijk kader

2.1 Moeilijkheden met preposities in de Nederlandse taal

Bij het modelleren van de Nederlandse taal, vormen de preposities vaak een struikelblok. In het Nederlands worden regelmatig deze kleine woordjes in zinnen geplaatst, zonder dat het altijd duidelijk is waar ze precies voor dienen. We nemen nogmaals de zinnen gegeven in voorbeeld 1.

Voorbeeld 1

- a) Jan begon een boek
- b) Jan begon aan een boek
- c) Jan begon met een boek
- d) Jan begon in een boek

Hoewel de preposities in principe een alternatief zijn voor naamvallen, en ze helder kunnen worden ingedeeld in categorieën als richting (*naar*), bezit (*van*) en dergelijke, zijn de preposities een woordsoort met veel problematiek. In voorbeeld 1 zien we hoe de zinnen 1b-d gebruik maken van een prepositie, maar zin 1a niet. Desondanks lijken 1a en 1b dezelfde betekenis te hebben.¹ Bovendien zou de prepositie *aan* hier ook vervangen kunnen worden door *met*. Waarom dit zo is blijft onduidelijk. Moedertaalsprekers voelen dit vanzelf correct aan, maar een computer kent dit niet en heeft een verklaring nodig via regels. Volgens de theorie van *iconiciteit* (Haspelmath, 2008) zouden zinnen met meer woorden automatisch ook meer informatie moeten uitdrukken. Dit zou dus betekenen dat 1b meer informatie geeft dan 1a. Als dit zo zou zijn – hier komen we later op terug – moet er dus informatie in de prepositie besloten zijn. Zoals eerder genoemd, zien we in voorbeeld 1 hoe er impliciete informatie wordt verstrekt over een activiteit die verder niet genoemd wordt. Uit het verschil tussen 1a,b (*schrijven*), 1d (*lezen*) en 1c (*schrijven* of *lezen*) blijkt dat het de preposities zijn die de betekenis van de zin bepalen. Om hiermee om te gaan kan er in een model natuurlijk worden geprogrammeerd hoe deze combinaties werken, maar wanneer er meer van dit soort gevallen zijn, wordt het een niet te overziene hoeveelheid werk. Zoals eerder gesteld, zou het dus beter zijn te zoeken naar regels die de individuele gevallen overstijgen, om deze gevallen te kunnen verklaren en ermee om te kunnen gaan.

2.2 Pustejovsky en de qualiastructuur

Een mogelijkheid voor het opvangen van ingewikkelde regels en het wegwerken van meervoudige entries in taalbases, wordt geopperd door James Pustejovsky. In zijn artikel *The generative Lexicon* (Pustejovsky, 1991) merkt hij op hoe in het Engels vaak woorden worden gecombineerd die niet met elkaar kloppen wat betreft type. Omdat ze toch gecombineerd moeten kunnen worden, worden er dan nieuwe entries gemaakt, waarbij er een ander type wordt toegekend aan een woord, zodat de betreffende woorden toch in de

¹ Hoewel hier dient te worden opgemerkt dat over *Jan begon aan een boek* enige twijfel bestaat; een enkeling zegt in gesprek hier de lezing *lezen* aan te geven. Later in deze scriptie zal echter blijken dat de meest gangbare vorm toch *schrijven* is.

gewenste combinatie kunnen worden gebruikt. Bovendien kunnen woorden vaak met verschillende betekenissen worden gebruikt, en zijn ze dus ambigue.

Voorbeeld 2

- a) Mary finished her cigarette
- b) John finished his beer

(Pustejovsky, 1991)

In voorbeeld 2 zien we hoe het werkwoord *finished* bij 2a de betekenis *oproken* heeft, maar in 2b *opdrinken* betekent. Bovendien verlangt een werkwoord als *to finish* normaal gesproken een woord dat een bezigheid uitdrukt. *Cigarette* en *beer* echter zijn beide statische objecten. Deze types zouden daarom niet samen moeten kunnen gaan. Desondanks zijn het beide zeer natuurlijke zinsconstructies, waarvan we willen dat deze goed verwerkt worden door de computer.

Om met dergelijke ambiguïteiten en type errors om te gaan, stelt Pustejovsky dat er gebruik wordt gemaakt van *type coercion*. Hierbij zouden woorden in staat zijn om de types van andere woorden dusdanig te veranderen, dat de types op de goede manier op elkaar aansluiten, waar dat normaal gesproken niet zou gebeuren. Op deze manier ontstaat er een coherente zin, zonder dat er een extra entry in het lexicon hoeft te worden geplaatst. Deze coercion zou volgens hem moeten plaatsvinden op basis van de gegevens uit de *qualiastructuur*. Qualia zijn hierin stukken informatie die inherent aan een woord kunnen worden opgeslagen in het lexicon. Hij onderscheidt hierin vier verschillende qualiarollen:

- I. Constitutive role: de relatie tussen een object en zijn onderdelen
(Bijvoorbeeld materiaal, gewicht, constructieve opbouw)
- II. Formal role: waarmee een object zich onderscheidt van een ruimer domein
(Bijvoorbeeld oriëntatie, grootte, vorm, kleur)
- III. Telic role: het doel en functie van het object
- IV. Agentive role: factoren van het ontstaan van het object
(Bijvoorbeeld maker, soort, oorzaak)

Op basis van deze rollen zou er vervolgens voldoende informatie beschikbaar moeten zijn in een entry om de ambiguïteit van een woord te behouden, zonder dat er meerdere entries nodig zijn.

Voorbeeld 3

Ik begin een roman

Het werkwoord *beginnen* vraagt een complement, en wel van het type *transitie*. Hiermee wordt bedoeld dat er een bezigheid plaatsvindt, zoals schrijven, die kan beginnen. In voorbeeld 3 is het complement *een roman*. Een roman is echter een statisch object, en voldoet dus niet aan de vraag naar een transitie. Coercie kijkt daarom naar de qualia van *roman*, en zoekt hierin naar een transitief gebruik.

Voorbeeld 4

$\lambda x(\text{roman}(x) \ \&$

$\text{Constitutive}(x) = \text{narratief}'(x) \ \&$

$\text{Formal}(x) = \text{boek}'(x) \ \&$

$\text{Telic}(x) = \lambda y, e^T [\text{lezen}'(x)(y)(e^T)] \ \&$

$\text{Agentive}(x) = \lambda y, e^T [\text{schrijven}'(x)(y)(e^T)]$

(Pustejovsky, 1991)

In voorbeeld 4 zien we de vier qualiarollen van *roman* genoteerd op basis van de lambda-calculus. Te zien is hoe de constitutive en formal rollen nieuwe nomina opleveren, maar hoe de telic en agentive rollen beide een werkwoord met een transitiefunctie geven. Omdat de telic en agentive rollen een transitie geven, kan er op basis van deze informatie een keus gemaakt worden tussen deze twee gebruiken, wat resulteert in de twee mogelijkheden van voorbeeld 5. Zoals gezegd wordt hierbij de keuze gemaakt voor de tweede optie. Waarom dit zo is, kom ik later op terug.

Voorbeeld 5

a) Ik begin een roman te lezen

b) Ik begin een roman te schrijven

(Pustejovsky, 1991)

De theorie van Pustejovsky leverde veel reactie op, waarvan een behoorlijk deel met een kritische noot. Met name Markus Egg (2003) gaat er kritisch op in, en bemerkt verschillende probleempunten. Zo stelt hij dat de telic en agentive qualia teveel wereldkennis vereisen om te modelleren. Volgens Egg hoort deze kennis niet thuis in een taalkundig model, omdat het daar slechts om de taal gaat. Een ander probleem dat hij benoemt is het feit dat de context van een zin de betekenis veel meer kan veranderen dan Pustejovsky voorstelde. Hierdoor zou de genoemde coercie veel minder goed werken dan was aangenomen.

Voorbeeld 6

a) John began a novel

b) John began to read a novel

c) John began to lecture on a novel

(Egg, 2003)

In voorbeeld 6a vinden we dezelfde zin die Pustejovsky in zijn voorbeeld gebruikte. Waar deze op basis van zijn qualiastructuur echter zin 6b als logische keuze achtte, argumenteert Egg dat, gegeven een context waarin

John een docent literatuurgeschiedenis is, zin 3c veel sterker staat.² Hiermee wil hij laten zien dat de ambiguïteiten van zinnen nog verder gaan dan Pustejovsky aannam, waarmee diens qualiatheorie ontoereikend blijft om met alle vormen om te gaan. Tot slot merkt Egg op dat Pustejovsky's theorie niet goed zou kunnen werken met quantifiers, zeker wanneer een context de scope van de quantifier beperkt.

Voorbeeld 7

- a) At first, only half of the windows was open.
- b) But then, Max opened all the windows.

Volgens Egg zou uit de qualiatheorie blijken dat in 7b Max van alle ramen de oorzaak is dat ze open zijn (ofwel dat hij ze allemaal open doet). Wanneer echter 7a erbij genomen wordt als context, blijkt dat echter dat hij alleen gezorgd heeft dat ze allemaal open zijn in plaats van de helft. Een subtiliteit die aan de qualia voorbij gaat.

Naast de kritiek van Egg zijn er ook voorbeelden van andere onderzoekers te vinden die de theorie van Pustejovsky ondergraven. Voorbeelden hiervan zijn Maarten Janssen (1997) en Fodor&Lepore (1998). Hoewel een deel van deze kritiek zeker de moeite waard is om te overwegen – inclusief de reactie daarop van Pustejovsky (1998) – en op meerdere punten ook reden bevat om terughoudend om te gaan met de hierboven genoemde qualiastructuur, wil ik toch argumenteren dat het nuttig kan zijn uit te gaan van de theorie zoals hierboven beschreven. Ten eerste is zowel de theorie als alle kritiek gebaseerd op voorbeelden uit het Engels. In deze taal zijn er meerdere voorbeelden gevonden die de theorie bevestigen, maar ook gevallen die (delen) onhoudbaar maken. Op het Nederlands is de theorie echter – voor zover mij bekend – niet getoetst. Zolang dat niet gebeurd is, staat ook niet vast of de theorie op het Nederlands toegepast al dan niet zal werken. Het zou immers kunnen zijn dat de Nederlandse taal zich sterker volgens de voorgestelde lijnen gedraagt. Ten tweede biedt Pustejovsky's theorie, of hij nu precies klopt of niet, een heldere structuur op basis waarvan eenvoudig hypothesen gesteld kunnen worden, en eventueel nieuwe theorieën kunnen worden opgezet.

2.3. Vraagstelling

Zoals hierboven beschreven is het onverstandig om bij het modelleren van taal losse entries te gaan maken voor bijzondere gevallen. Impliciete informatie die moet blijken uit het combineren moet daarom te voorspellen zijn aan de hand van onderdelen van de zin die wel bekend zijn:

Is het mogelijk op basis van preposities en nomina en Pustejovsky's qualiastructuur te voorspellen welke activiteit wordt geïmpliceerd?

Ik zal dit testen door te kijken of preposities steeds naar eenzelfde qualiarol verwijzen en daarmee dus een bepaald type activiteit impliceren. Om de omvang te beperken richt ik mij op de preposities uit voorbeeld 1

² De voorbeelden zijn hier in het Engels; mijn begrip van de nuances in het Engels is te laag om dit echt te kunnen beoordelen, maar naar mijn gevoel is het een wankel voorbeeld. In het Nederlands gaat het mijns inziens in ieder geval niet op. Dit blijft echter een gevoelskwestie, waarbij er geen eenduidige waarheid is.

(aan, met en in) en de zinnen zonder prepositie. Met deze laatste wil ik ook testen of preposities altijd een meerwaarde hebben, of dat juist het weglaten ervan informatie biedt.

3. Onderzoek

3.1. Onderzoeksmethode

3.1.1. Onderzoek naar het Nederlands

Zoals gezegd is op dit gebied bijna uitsluitend onderzoek gedaan naar het Engels. Dit is op zich niet verwonderlijk, aangezien dit de meest voorkomende taal is, maar het beperkt wel de visie. Doordat de theorie in zijn eerste vorm – Pustejovsky heeft zijn theorie na verschillende kritieken nog enkele malen bijgesteld – op basis van het Engels veel is afgekeurd, krijgt hij niet meer de kans zijn potentie te laten zien in andere talen. Waar het Engels zich weliswaar niet geheel houdt aan de door Pustejovsky bedachte structuur, zou het zo kunnen zijn dat het Nederlands dusdanig anders functioneert dat de qualiastructuur hier wel meer inzicht kan geven. Met dit onderzoek zal ik daarom toch nogmaals de voordelige kanten van de theorie gebruiken om te testen of deze toegepast op het Nederlands wel nieuwe inzichten kunnen verschaffen.

3.1.2. Onderzoek op basis van Google Search engine

In haar PhD proposal³ schrijft Josefiën Sweep hoe op het gebied van deze type coercion voornamelijk onderzoek is gedaan op basis van bijpassende, gemaakte voorbeelden, maar relatief weinig op basis van werkelijk gebruikte taal (Sweep, 2006). Hoewel er wel is gewerkt met corpora (Verspoor 1997, Pustejovsky 1993b), stelt Sweep dat ook hierbij vaak wordt teruggegrepen naar bekende voorbeelden, die vervolgens binnen het corpus worden opgezocht. Ze merkt hierbij op dat in de ‘echte’ taal nog veel meer uitzonderingen voorkomen dan in uitgezochte voorbeelden kunnen worden verkregen.

Het gros van de onderzoeken op basis van deze ‘echte’ taal wordt gedaan met behulp van zorgvuldig samengestelde corpora. Voor dit onderzoek heb ik echter besloten om gebruik te maken van de Google Search Engine. Mijns inziens heeft dit verschillende voordelen, waarvan de belangrijkste de ongekende omvang is van het internet dat hiermee bereikt kan worden. Hoewel er verschillende zeer uitgebreide corpora zijn uitgebracht, vallen deze altijd in het niet bij de hoeveelheid data die door mensen online wordt geplaatst. Omdat ik in het onderstaande ook naar minder gebruikelijke constructies zal zoeken, is het risico dat hiervan in een vaststaand corpus geen voorkomens te vinden zijn, terwijl het wel wordt gebruikt. Met Google Search wordt dit probleem tot een minimum beperkt. Daarnaast zijn de data op internet niet geredigeerd, en kan er ook taalgebruik voorkomen wat misschien niet heel gebruikelijk is. Ook dit is echter een onderdeel van natuurlijke taal, aangezien dit de veranderingen zijn waardoor talen na verloop van tijd worden gevormd. Tot slot is het gebruik van een zoekmachine een zeer snelle manier om direct een grote hoeveelheid data te verzamelen, die bovendien direct gebruikt en geanalyseerd kan worden.

Uiteraard heeft dit systeem ook belangrijke nadelen. Zo is het bijvoorbeeld niet mogelijk om met reguliere expressies preciezere output te genereren. Zoals je in een getagged corpus zoektermen kunt gebruiken die

³ Dit onderzoek loopt helaas nog, en heeft daarom nog geen artikel opgeleverd. Hoewel ik het werkelijke onderzoek niet heb kunnen zien, sluit het onderwerp dusdanig aan dat ook het voorstel alleen informatie inspiratie bevatte.

specifiek een vp en een np naast elkaar willen hebben, of verschillende uitgangen van hetzelfde woord, is het bij Google niet mogelijk verder te gaan dan een vast aantal woorden in een specifieke volgorde. Gevolg hiervan is dat een deel van de input bestaat uit zinnen die weliswaar dezelfde woorden in de goede volgorde hebben staan, maar toch een heel andere constructie vormen. Een veelzeggend voorbeeld hiervan is hoe de zoekterm *begon met een boek* – bedoeld om te zoeken naar constructies als *hij/zij begon met een boek (te schrijven)* – zinnen oplevert met de constructie *het begon met een boek dat*. Dergelijke zinnen moeten met de hand weggefilterd worden, om zo de correcte data te selecteren. Opgemerkt dient te worden dat dit probleem zich ook voordoet bij corpora zonder POS-tagging. Bij deze niet-getagde corpora blijft daardoor slechts het geredigeerde aspect over; datgene waarvan ik hierboven heb geargumenteed dat het geen voordeel is. Ondanks de grote mogelijkheden die de POS-tagging met zich meebrengt, wegen mijns inziens de voordelen van een snelle, niet geredigeerde en dus puurdere bron op tegen de nadelen van veel ruis en minder intelligente zoektermen.

3.1.3. Zoektermen

Bij het verwerven van data ben ik steeds uitgegaan van de drie genoemde preposities *aan*, *met* en *in* en de zinnen zonder prepositie. Deze heb ik vervolgens toegepast op de zinnen met “beginnen” en een nomen. De zinnen heb ik verdeeld in 1^e of 3^e persoon enkelvoud verleden tijd en 1^e persoon enkelvoud tegenwoordige tijd. Deze scheiding en beperking heb ik aangehouden om de hoeveelheid data overzichtelijk te houden. Bovendien verwacht ik niet dat een uitbreiding van deze set meer informatie zal geven dan hieruit te verkrijgen is. Bij het verwerven van data ben ik in principe uitgegaan van de eerste vijf resultaten. Hoewel dit niet erg veel is, blijkt dit met een aantal combinaties van prepositie en nomen al te veel te zijn, en met andere combinaties genoeg om te zien wat er zal gebeuren. In de meeste gevallen waren de resultaten per combinatie eenduidig, en leverde het nemen van meer resultaten geen extra variatie op.

3.2. Analyse

3.2.1. Een boek beginnen

Bij dit onderzoek ben ik uitgegaan van het eerstgenoemde voorbeeld, *Een boek beginnen* en diens varianten. Wanneer deze VP zonder preposities wordt gebruikt, wordt er geïmpliceerd dat het onderwerp van de zin gaat beginnen met het schrijven van een boek.⁴ Wanneer er een prepositie aan wordt toegevoegd, kan er behalve *schrijven* ook *lezen* worden geïmpliceerd. Op basis van deze zoektermen geeft de Google Search Engine strings bestaande uit de zoekterm met daaromheen een deel van de context uit de bron. Van deze strings heb ik, zoals hierboven beschreven, in principe de eerste vijf resultaten genomen, tenzij het een andere constructie betrof met toevallig dezelfde woordvolgorde. In tabel 1 zijn de resultaten te zien van de zoektermen gebaseerd op het nomen *boek* en de bovengenoemde termopbouw. In de derde kolom staat genoteerd welke activiteit er geïmpliceerd wordt. Deze labeling is vervolgens handmatig uitgevoerd en is gebaseerd op de zin zelf. Hierna is deze gecontroleerd met de context buiten de zin.

⁴ Sommige mensen zijn van mening dat deze VP ongrammaticaal is. Dit is wederom een voorbeeld van gevoelsmatigheid van taal.

<u>Zoekterm</u>	<u>Zin</u>	<u>Impliciete activiteit</u>
Begon een boek	Ik begon een boek over Mohammed, niet als heilige, maar als persoon, als leider.	Schrijven
	Maar Mozes begon een boek, het boek van de Wet en het bevat vijf van zijn Geschriften.	Schrijven
	<i>Verder geen geldige resultaten</i>	
Begon aan een boek	Hij begon aan een boek over zijn ervaringen in Tahiti, 'Noa Noa', maar het boek kwam helaas niet af.	Schrijven
	Ik begon aan een boek over een spiegel naar een andere werkelijkheid. ik werd aangemoedigd om zeker door te gaan.	Schrijven
	Dat veranderde toen ik samen met Bas Krajenbrink en Hemmo Battjes begon aan een boek.	Schrijven
	De jonge filosoof raakte zo in de ban van de 'geest van de vrijheid' die hij daar ontwaarde dat hij begon aan een boek.	Schrijven
	Ik begon aan een boek dat de verontwaardigde titel: "Ik word al negen!" zou dragen.	Schrijven
Begon met een boek	Ik begon met een boek "basiskennis van de Islam". Ik begon s' avonds te lezen en zat twee uren later nog in volle concentratie.	Lezen
	Elisabeth Mollema begon met een boek voor volwassenen. Maar haar uitgever wilde ook wel eens een kinderboek uitgeven.	Schrijven
	Ik begon met een boek, dan een volgende, tot ik ze allemaal uithad.	Lezen
	Hij begon met een boek over een Elfje in de sneeuw (Zwijsen). Later schreef hij de Drakenrilogieën.	Schrijven
	Ze begon met een boek voor volwassenen maar al gauw begon ze boeken voor jongere kinderen en jeugdromans te schrijven.	Schrijven
begon in een boek	ik had mijn eigen boeken allemaal al uit en begon in een boek van mijn moeder.	Lezen
	Jits begon in een boek (Haar naam was Sara)	Lezen
	Ik begon in een boek, maakte een wandeling, zat op de bank slechts te zitten en ontcijferde op eigen houtje een zelfhulpformule	Lezen
	Welke onderzoeker heeft nooit meegemaakt dat hij hoopvol begon in een boek of archief en daar niet aantrof wat hij dacht te vinden.	Lezen
Begin een boek	Begin een boek met een spannende gebeurtenis. Zo zet je meteen de toon.	Schrijven
	meid begin een boek, je hebt een goeie fantasie	Schrijven
	Begin een boek en ga naar een uitgever (saai)	Schrijven
	Ik begin een boek ook altijd met een zin die op zichzelf kan staan.	Schrijven
	Ik begin een boek met de hoofdpersonen, die zijn het allerbelangrijkst.	Schrijven
Begin aan een boek	Voordat ik begin aan een boek denk ik net zo lang over een idee na tot ik weet wat het allerbelangrijkste is dat ik wil vertellen.	Schrijven
	Als ik begin aan een boek, weet ik precies wat er gaat gebeuren, maar meestal ga ik al bij bladzijde twee een heel andere kant op'	Schrijven
	Begin aan een boek/verhaaltje en schrijf er iets in als je je verveelt	Schrijven
	Ik begin aan een boek door nota's op te schrijven, met de pen, het liefst in een rood atoma-schrift.	Schrijven
	Als ik begin aan een boek dan rauzer ik ze er ook doorheen, een boek van 500 bladzijden doe ik dan in 2 dagen wel.	Lezen ⁵

⁵ Dit geval is een uitzondering op de betekenis. Zie hiervoor voorbeeld 8 en uitleg.

Begin met een boek	als ik bijvoorbeeld begin met een Boek : Het jaar was 1500, en een man liep op straat... Dán heb je al “waarheden” getypt.	Schrijven
	Als ik begin met een boek, moet het uit.	Lezen
	Ik lees vaak zo, begin met een boek en als het dan boeit, wordt het een 'thread'.	Lezen
	Jeroen Theunissen: Voor ik begin met een boek, reis ik meestal.	Schrijven
	Ik begin met een boek van John Welwood	Lezen
Begin in een boek	Als ik eenmaal begin in een boek kan ik het niet NIET uitlezen.	Lezen
	Als ik eenmaal begin in een boek dat mij aanspreekt ben ik als het ware verslaafd en moet ik doorlezen tot ik het boek uit heb.	Lezen
	... als ik weer eens een boek uit hebt of begin in een boek wat me niet meer los laat.	Lezen
	Ik heb het soms ook dat ik wel begin in een boek maar dat ik het halverwege wegleg en niet meer inkijk.	Lezen
	Ik begin in een boek, vervolgens in een ander, dan weer een etc.	Lezen

Tabel 1: resultaten met *begon een boek* en varianten

Voorbeeld 8

Als ik begin aan een boek dan rauzer ik ze er ook doorheen, een boek van 500 bladzijden doe ik dan in 2 dagen wel.

Op basis van deze output lijken we te kunnen stellen dat de preposities inderdaad invloed hebben op welke activiteit wordt geïmpliceerd door de VP. Kort gezegd lijken we de volgende lijn te kunnen trekken:

- geen prepositie → schrijven
- *aan* → schrijven
- *met* → schrijven of lezen
- *in* → lezen

In voorbeeld 4 zagen we hoe *schrijven* en *lezen* respectievelijk de agentieve en telische qualiarollen zijn van het nomen *boek*. Wanneer we bovenstaande lijn vertalen naar qualiarollen krijgen we dus het volgende schema:

- geen prepositie → agentief
- *aan* → agentief
- *met* → agentief of telisch
- *in* → telisch

Bij de data hierboven vallen naast deze lijn twee bijzonderheden op. Ten eerste levert de verleden-tijdsvorm bij de zoekterm zonder prepositie slechts twee resultaten op. Hoewel deze combinatie veel voorkomt op internet, wordt deze bijna altijd gecombineerd met *te* en een infinitief. In tegenstelling tot de tegenwoordige tijdsvorm, is het hier blijkbaar gebruikelijker de activiteit expliciet te melden in plaats van via coercie impliciet te geven. Een tweede opmerkelijkheid is het laatste resultaat van *Begin aan een boek*, te zien in voorbeeld 8, waarbij de implicatie *lezen* is. Dit is naar mijn idee een voorbeeld van hoe mensen de taal na verloop van tijd anders gaan gebruiken, waardoor dus ook impliciete betekenissen langzaam kunnen veranderen. Bij het bekijken van meer resultaten, blijkt ook dat deze lezing slechts zelden voorkomt. Om deze reden is deze lezing bij bovenstaand schema niet toegevoegd aan de implicatie van *aan*.

3.2.2. Een relatie/winkel beginnen

Op basis van het schema hierboven kunnen we de volgende hypothese opstellen:

- VP's zonder prepositie of met de prepositie *aan* impliceren een agentieve qualiarol
- VP's met de prepositie *in* impliceren een telische qualiarol
- VP's met de prepositie *met* impliceren een agentieve of telische qualiarol

Deze hypothese kan nu getest worden op andere nomina. Hiervoor zullen we gebruik maken van de nomina *relatie* en *winkel*. Deze nomina zijn willekeurig gekozen, maar met bij beide de overweging dat het om een geheel ander soort nomen gaat; bij beide is sprake van een woord dat niets met *boek* te maken heeft. Als gevolg hiervan wordt de hypothese getest op heel verschillende gebieden en zal blijken of type coercie in zijn simpele vorm, zoals in eerste instantie door Pustejovsky (1991) gegeven is, werkt met Nederlandse preposities en nomina. De resultaten zijn te zien in respectievelijk tabel 2 en 3.

Zoekterm	Zin	Impliciete activiteit
Begon een relatie	Ze begon een relatie met Alex Kelly	Aangaan
	Geertje begon een relatie met Gerhard	Aangaan
	Leonardus begon een relatie met Antonia Maria Brokling.	Aangaan
	Ik begon een relatie met een Japanse vrouw toen ik nog gehuwd was.	Aangaan
	In 1972 scheidden ze en Julie begon een relatie met Doug Williams	Aangaan
Begon aan een relatie	Hij wist dat hij begon aan een relatie met iemand die nét een huwelijk achter de rug had, maar toch.	Aangaan
	dat ik niet begon aan een relatie waarvan ik weet dat die zal mislopen,	Aangaan
	Als je nou nooit begon aan een relatie, had je nu ook geen problemen.	Aangaan
	Maurice begon aan een relatie op 1 januari 2007 met Josiane Poublon.	Aangaan
	Vriendin van mij was ook 17 toen ze begon aan een relatie met een jongen van 25..	Aangaan
Begon met een relatie	<i>Geen geldige resultaten</i>	n.v.t.
Begon in een relatie	<i>Geen geldige resultaten</i>	n.v.t.
Begin een relatie	Dag 13: Begin een relatie met God.	Aangaan
	Ik begin een relatie met iemand wanneer ik die erg leuk vind	Aangaan
	Ik begin een relatie als ik mij goed voel. Ik bestempel het niet als verliefdheid, dat is iets sterkers.	Aangaan
	begin een relatie en maak een kind ;).	Aangaan
	Begin een relatie met je eigen pluizige puppy of kitten, en kijk hoe ze naar je toe komen in volledig 3D.	Aangaan
Begin aan een relatie	Ik denk niet dat ik begin aan een relatie met een gehandicapt iemand	Aangaan
	Als ik begin aan een relatie ga ik er niet 100 maar 1000% voor.	Aangaan
	ik moet eerst nog wat meer mezelf worden voor ik eventueel begin aan een relatie	Aangaan
Begin met een relatie	Begin met een relatie die leidt tot een liaisonfunctie om bedrijven aan elkaar te koppelen.	Zoeken
Begin in een relatie	<i>Geen geldige resultaten</i>	Aangaan

Tabel 2: Resultaten met *begon een relatie* en varianten

Voorbeeld 9

Begin met een relatie die leidt tot een liaisonfunctie om bedrijven aan elkaar te koppelen

Wanneer we kijken naar de resultaten van de zoektermen met *relatie*, zien we dat de hypothese zoals hierboven gesteld direct problemen oplevert. Ten eerste blijkt het moeilijk te zijn een heldere activiteit te benoemen die door coercie wordt geïmpliceerd. In dit geval heb ik gekozen voor het woord *aangaan*, omdat dit mijns inziens het best de lading dekt van wat er gaande is: het begin van een relatie is het aangaan ervan. Ten tweede blijkt dat de zinnen met daarin de preposities *met* en *in* geen resultaten oplevert, behalve de zin gegeven in voorbeeld 8. Van deze zin is het echter de vraag of hier werkelijk sprake is van coercie. In plaats daarvan lijkt dit een constructie te zijn met een opsomming, waarvan de hier genoemde relatie het eerste element is. Dit in tegenstelling tot de overige zinnen in tabel 2, waar het specifiek de relatie is die wordt begonnen en waar het aangaan ervan dus de activiteit is. Dit is daarom duidelijk de agentieve qualiarol. Deze twee opmerkelijkheden kunnen veroorzaakt worden door een onduidelijk bestaan van de telische qualiarol van *relatie*. Het is immers niet erg duidelijk welke activiteit het doel van een relatie aanduidt.

Zoekterm	Zin	Impliciete activiteit
Begon een winkel	Mijn over-overgrootvader begon een winkel waar alles werd verkocht wat een dorp als Maasbree (Noord-Limburg) nodig had.	Opstarten
	wijlen Feijer sr. die al veel handel dreef in kaas en eieren begon een winkel "Het Zuivelhuis".	Opstarten
	Cor, mijn opa, begon een winkel in de Dijkstraat in Honselersdijk.	Opstarten
	Zij begon een winkel en wijdde zich aan de bescherming van priesters en religieuzen.	Opstarten
	De oudste zoon Jan begon een winkel in Deventer aan de Smedenstraat.	Opstarten
Begon aan een winkel	De kruidenier Adriaan van Well begon aan een winkel onder de naam De Spar	Opstarten
Begon met een winkel	Zijn vrouw Doortje begon met een winkel in een kippenhok.	Opstarten
	Het bedrijf begon met een winkel in de PC Hoofdstraat te Amsterdam.	Opstarten
	Onze start begon met een winkel in 1981 als jonge knaap in de Lange Nieuwstraat te Antwerpen.	Opstarten
	Dinsdag 2 maart is het tien jaar geleden dat Evelien van Driesten begon met een winkel welke gespecialiseerd is in linnengoed met o.a. keukendoeken,	Opstarten
	Hij begon met een winkel in witgoed aan de Irisstraat te Ridderkerk. Daarnaast verrichtte hij alle voorkomende electrawerkzaamheden bij zowel particulieren	Opstarten
Begon in een winkel	Vader Smit begon in een winkel in witgoed en onderhoud.	Verkopen
Begin een winkel	Begin een winkel, lok klanten, geef je klanten de service die ze nodig hebben en verdien sterren	Opstarten
	Begin een winkel met eigengemaakte spulletjes	Opstarten
	ik begin een winkel in Poolse delicatessen.	Opstarten
	Ik begin een winkel in hondenkleding	Opstarten
	Begin een winkel met eigengemaakte spulletjes, beetje de combi van hobby en kunst.	Opstarten
Begin aan een winkel	<i>Geen geldige resultaten</i>	n.v.t.
Begin met een winkel	Begin met een winkel op Internet, kosten zijn zeer laag.	Verkopen
Begin in een winkel	<i>Geen geldige resultaten</i>	n.v.t.

Tabel 3: Resultaten met *begon een winkel* en varianten

Wanneer we in Tabel 3 naar de resultaten van de combinaties met "winkel" kijken, blijkt de bovenstaande hypothese wederom niet erg sterk te staan. Opnieuw zijn niet alle combinaties met de gebruikte preposities mogelijk; dit keer zijn het "aan" en "in" die niet werken, op bij beide één hit in de verleden tijd na. Bij *aan* lijkt

er sprake te zijn van een uitzondering zoals we zagen bij de zin in voorbeeld 8, waarbij er een voorbeeld zichtbaar lijkt van de levendigheid van natuurlijke taal. De uitzondering met *in* lijkt echter eenzelfde geval te zijn als de zin in voorbeeld 8. Daarnaast is er een subtiel verschil op te merken tussen de resultaten zonder prepositie en die met *met*: waar bij de eerste wederom duidelijk het opstarten van het gebruikte nomen is – de agentieve rol – is bij *met* de scope omgedraaid: hier is het het onderwerp van de zin dat begint, waarbij de winkel het middel is.⁶ Dit is echter een ander soort onderscheid dan we bij de combinaties met *boek* aantreffen. *Met* zou dan namelijk algemeen moeten zijn, waar hij hier juist niet hetzelfde kan doen als de zinnen zonder prepositie.

3.2.3. Tussentijdse conclusie

Uit de resultaten van *relatie* en *winkel* wordt al gauw helder dat de eerste hypothese erg zwak is. Hoewel hij nog niet direct krachtig wordt tegengesproken – er zijn geen voorbeelden waar een prepositie verwees naar een qualiarol die volgens de hypothese niet zou kunnen – zijn er ook geen volledig bevestigende resultaten gekomen. Dit kan verschillende oorzaken hebben. Om te beginnen is niet zeker of elk nomen wel alle vier de qualiarollen, en in dit geval in het bijzonder de telische en agentieve rollen, bezit. Het zou immers best kunnen zijn dat een nomen enkel de constitutive, formal en agentive roles heeft. Een andere optie is dat de nomina op hun beurt nog weer zijn op te delen in categorieën, waarbij de hypothese door een of meerdere subcategorieën wél wordt bevestigd. Het zou immers kunnen zijn dat *boek* in een andere categorie kan worden geplaatst dan *relatie* en *winkel*. Omdat het buiten het bereik van dit onderzoek ligt om in kaart te brengen welke qualiarollen bij elk nomen te vinden zijn, zullen we ons in het onderstaande richten op de tweede optie en pogen een onderscheid te maken tussen verschillende nomina.

3.3. Categorië van nomina

Om een categorisatie te kunnen maken van verschillende soorten nomina, is het van belang om een nieuwe lijst te verwerven. Hiertoe gebruiken we nogmaals dezelfde zoektermen, maar laten we nu steeds het nomen weg. Door het gebruik van *een* aan het eind van de zoekterm kunnen we, zij het enigszins provisorisch vergeleken met reguliere expressies, afdwingen dat er een nomen volgt op de gevraagde woordengroep. Daarnaast is een onderwerp aan de zoekterm toegevoegd – *hij* bij verleden tijd en *ik* bij tegenwoordige tijd – om de hoeveelheid ruis in de output te beperken.⁷ De resultaten van deze zoektermen zijn te vinden in tabel 4.

Zoekterm	Zin	Categorie nomen
Hij begon een	Hij begon een laboratorium met eenvoudige middelen.	Object
	Hij begon een productiebedrijf	Object
	Hij begon een relatie met Elisabeth Levitsky	Object
	Maar hij begon een zoektocht naar zijn eigen stijl in een bijzonder gunstig	Object

⁶ Eventueel zou men kunnen argumenteren dat het hier om een telische qualiarol gaat, maar dit gaat wat mij betreft te ver: het doel van een winkel lijkt mij in eerste instantie om producten te verkopen, niet om een keten of persoon te laten groeien of beginnen aan een carrière.

⁷ Bij eerdere zoektermen waren bij enkele combinaties al weinig resultaten; die zouden zijn weggevallen door het toevoegen van deze specifieke onderwerpen. Bij deze zoektermen zijn echter voldoende resultaten om dit middel hier wel te gebruiken. De specifieke keuze voor deze persoonlijk voornaamwoorden is dat ik verwacht dat deze het vaakst zullen voorkomen.

	klimaat	
	hij begon een gesprek met de bewakers	Event
Hij begon aan een	hij begon aan een mooie loopbaan.	Periode
	Hij begon aan een werk voorstellend een dame in rood fluweel,	Object
	hij begon aan een zwerftocht die zich uitte in de ene verhuizing na de andere	Periode
	Hij begon aan een intensieve, gedisciplineerde leesstudie van 9 uur per dag,	Periode
	Hij begon aan een huttencollege, maar woorden waren niet goed genoeg.	Periode
Hij begon met een	Hij begon met een team van vijftig dat later teruggebracht werd naar vijftien en uiteindelijk slechts tien man	Object
	Hij begon met een studie naar het effect van wind en golven op offshore windturbines	Object
	hij begon met een opsomming van alle voordelen die een ERP pakket te bieden heeft	Object
	Hij begon met een grote tournee, en na het succes van "I walk the line"	Object
	Hij begon met een 'afvalrace', door zich obsessief te storten op wielrennen, afvallen en laxeren.	Periode
Hij begon in een	Hij begon in een Sittards café en eindigde als een van de grootste entertainers die ons land heeft gekend: Toon Hermans.	Locatie
	Hij begon in een enorm startveld, baande zich een weg naar het geld en wist uiteindelijk de top 500 te bereiken	Locatie
	Hij begon in een plaatselijke operaproductie, maar later kwam hij onder de hoede van dirigent Serge Koussevitzky,	Locatie
	Hij begon in een verbouwde school met twee leslokalen en een afzonderlijke onderwijzers- woning.	Locatie
	Hij begon in een smartlappen-koor en vertolkte daarnaast op menig feestje het Nederlandse repertoire.	Locatie
ik begin een	ik begin een horecabedrijf	Object
	Ja, ik begin een RNI revival station op de middengolf.	Object
	Ik begin een jeneverfabriek.	Object
	Ik Begin Een Nieuw Leven	Object
	Ik begin een koffiezaak.	Object
Ik begin aan een	'Ik begin aan een wereldreis	Periode/Object
	"IK begin aan een tweede carrière"	Periode
	"Ik begin aan een nieuwe etappe"	Periode
	Ik begin aan een nieuw project.	Periode
	"Ik begin aan een vervolgopleiding.	Periode
Ik begin met een	Ik begin met een nieuwe baan.	Object
	ik begin met een mooi frans nummer	Object
	Ik Begin Met Een Fotoboek	Object
	Ik begin met een puzzel. 7 februari 2011	Object
	Dames en heren, ik begin met een anecdote	Object
Ik begin in een	Als ik begin in een taal	Locatie
	maar ik begin in een nieuwe hoofdstuk	Locatie
	ik begin in een normaal leuk goedkoop huis van rond de 16.000	Locatie
	ik begin in een heel lage divisie	Locatie
	Ik begin in een hoekje van het papier en wil steeds meer.	Locatie

Tabel 4: Resultaten met *begon een* en varianten

Wanneer we naar deze nieuwe output kijken, kunnen we vier verschillende categorieën onderscheiden:

- *Object*: objecten zijn die nomina die puur op zichzelf staan. Deze nomina zijn statisch, en het is het onderwerp van de zin dat de (impliciete) activiteit van de zin uitvoert.

Voorbeeld 10

- Hij begon een relatie met Elisabeth Levitsky
- Hij begon met een opsomming van alle voordelen die een ERP pakket te bieden heeft

- c) Ik begin een horecabedrijf
- d) Ik begin met een nieuwe baan

Bij voorbeeld 10 is te zien hoe deze objecten steeds puur de functie van lijdend voorwerp hebben.

- *Event*: events zijn gebeurtenissen, waarbij het nomen al direct de activiteit aangeeft. Voorbeeld 11 geeft dit geval weer met het nomen *gesprek*. Dit nomen is een object wat een bezigheid op zichzelf is, en daarmee geen twijfel overlaat van de activiteit.

Voorbeeld 11

Hij begon een gesprek

- *Periode*: periodes zijn objecten of events waar duidelijk een tijdsduur aan verbonden is. Anders dan gewone objecten of events, geven ze dus al informatie dat het om een activiteit van beperkte duur gaat.

Voorbeeld 12

- a) Hij begon aan een mooie loopbaan
- b) Ik begin aan een nieuw project
- c) hij begon aan een zwerftocht die zich uitte in de ene verhuizing na de andere
- d) Ik begin aan een vervolgopleiding

- *Locatie*: locaties geven informatie over waar een activiteit plaatsvindt. Dit kan zowel concreet als abstract zijn, zoals te zien in respectievelijk voorbeeld 13a,b en 13c.

Voorbeeld 13

- a) Hij begon in een verbouwde school met twee leslokalen
- b) Ik begin in een normaal leuk goedkoop huis van rond de 16.000
- c) Maar ik begin in een nieuw hoofdstuk

Van deze vier categorieën lijkt *object* de belangrijkste te zijn in dit onderzoek. Met deze nomina is er namelijk sprake van statische dingen die zelf niet veranderen, en ook wordt er geen activiteit aangegeven. Het is hier dus de coercie die moet zorgen voor een activiteit die begonnen wordt. Dit in tegenstelling tot *events*, waar het nomen juist draait om een bepaalde activiteit. Doordat hier al sprake is van een event, hoeft er geen gebruik gemaakt te worden van coercie voor begrip van de activiteit. Wel is coercie nodig om het type van het nomen om te zetten in een transitie, zoals gevraagd door *beginnen*. De categorie *periode* ligt hier vervolgens een beetje tussen in. Zoals gezegd zijn dit objecten of events, waarbij er extra informatie is gegeven dat het activiteiten betreft van beperkte duur. In voorbeeld 12a-b is sprake van objecten, aangezien de nomina in deze zinnen zelf statisch zijn en geen informatie geven over de precieze activiteit. Bij 12c-d is er echter geen twijfel mogelijk van wat het onderwerp van de zin gaat doen. Het onderscheid van deze gevallen moet echter handmatig gemaakt worden.

Tot slot blijft de categorie *locatie* over. Deze categorie lijkt echter een vreemde eend in de bijt te zijn. De nomina worden hier telkens ingeleid door de prepositie *in* en duiden telkens op een containerconstructie. Het onderwerp van de zin voert een actie uit, en doet dit in de plaats waar het nomen naar verwijst. Het is echter maar de vraag of hier sprake is van coercie, aangezien de zinnen stuk voor stuk kunnen worden uitgebreid naar een zin waar *beginnen* wel degelijk een woord van het gewenste type krijgt, zoals te zien is in voorbeeld 14. Omdat dit bij de prepositie *in* steeds het geval lijkt te zijn, moeten we ons afvragen of zinnen met deze prepositie wel een gelijke behandeling mogen krijgen als die met de preposities *aan* en *met* en diegene zonder prepositie.

Voorbeeld 14

- a) Hij begon in een Sittards café
- b) Hij begon te zingen in een Sittards café

Er lijken zich nu dus eigenlijk twee speciale gevallen af te tekenen. Aan de ene kant zijn er de objecten, waarbij coercie plaatsvindt. Hier moet er een activiteit worden geïmpliceerd die van zichzelf nog niet aanwezig is. Met deze implicatie wordt tevens voldaan aan de vraag van het werkwoord *beginnen* om een woord van het type transitie, wat bij de activiteiten het geval is. Aan de andere kant hebben we de locaties, die met behulp van de prepositie *in* een containerconstructie aankondigen. Hier is geen sprake van coercie die het nomen een activiteit laat impliceren, maar van informatie van waar de plaatshebbende activiteit zich afspeelt.

3.4. Nadere categorisatie

3.4.1. Nomina uit de categorie *object*

Ondanks het verwerven van nieuwe data en het opschonen ervan op basis van bovengenoemde categorisatie, is het nog niet mogelijk om de hypothese uit sectie 3.2.1 nogmaals te testen, of aan te wijzen waarom deze een zodanig wisselend effect heeft als we hebben gezien. De nomina in de resultaten uit tabellen 1 tot en met 3 zijn immers alle drie objecten, maar geven toch steeds een andere uitwerking. Een mogelijkheid om hiermee om te gaan is verder te gaan met subcategorieën binnen *object*. Wanneer bij deze subcategorieën groepen te vinden zijn waarbij de hypothese wordt bevestigd, kan daarmee verklaard worden waarom de preposities met deze nomina functioneren zoals ze dat doen, en wellicht ook waarom dit bij andere niet zo is. Een voorstel voor deze subcategorisatie zou kunnen zijn om onderscheid te maken op concreet versus abstract of grootte. Het is echter de vraag of dit veel effect zal hebben.

Voorbeeld 15

- a) Aan een boek beginnen
- b) *Aan een melkpak beginnen
- c) Aan een relatie beginnen
- d) *Aan een droom beginnen
- e) Aan een winkel beginnen

f) ?Aan een kantoor beginnen

In voorbeeld 15 zijn drie maal twee zinnen gegeven. In 15a,c,e zijn de nomina uit de eerste zoektermen en tabellen 1 tot en met 3 gebruikt. In 15b,d,f zijn dezelfde zinnen gemaakt, maar nu zijn nomina ingevuld die volgens een subcategorisatie op basis van de criteria concreet vs. abstract en grootte in dezelfde klasse zouden moeten vallen. Het wordt echter al direct duidelijk dat deze nomina problemen opleveren wanneer we ze combineren met één enkele prepositie. Blijkbaar kan de prepositie *aan* bij de nomina *melkpak*, *droom* en *kantoor* niet de agentieve rol aanroepen met behulp van coercie, terwijl dit bij *boek*, *relatie* en *winkel* wel zo is. Hoewel hiermee nog niet zeker is of dit ook geldt voor combinaties met de andere hierboven gebruikte preposities, is wel duidelijk dat deze nomina niet overeen kunnen komen in subcategorie; immers, wanneer dit wel het geval zou zijn, zouden ze ook met dezelfde preposities gecombineerd moeten kunnen worden.

Het voorstel van de criteria abstract vs. concreet en grootte is weliswaar simpel, maar is wel het meest voor de hand liggend. Hoewel de genoemde voorbeelden geen gesloten bewijs vormen dat deze uitleg ontoereikend is, valt niet te verwachten dat er andere criteria te vinden zijn waarmee de objecten dusdanig te ordenen zijn dat de verschillende preposities wel op dezelfde manier effect hebben. Het zou te verwachten zijn dat een dergelijke verdeling zeer intuïtief zou moeten zijn, om geworteld te raken in het dagelijks taalgebruik. Wanneer een intuïtieve categorisatie als hierboven niet voldoende blijkt en geen van de in voorbeeld 15 gekoppelde nomina als gelijk kan behandelen, acht ik de kans klein dat een verdere zoektocht meer resultaat zal bieden. Hoewel onbevredigend, lijkt dan ook de conclusie gerechtvaardigd dat er geen categorisatie mogelijk is waarmee preposities standaard naar eenzelfde qualiarol verwijzen.

3.4.2. Containerconstructie met *in*

Tot slot blijft er nog de containerconstructie over die gepaard ging met de prepositie *in*. In voorbeelden 13 en 14 was al te zien hoe in deze zinnen de prepositie niet coercie veroorzaakt, maar in plaats hiervan een containerfunctie impliceert. Om dit nader te bekijken nemen we nogmaals de zoektermen met *in* zonder nomen als input voor Google Search.

Voorbeeld 16

- a) Hij begon in een smartlappen-koor
- b) Don heeft ook al meerdere bands gehad, hij begon in een punkrock bandje
- c) Ik begin in een heel lage divisie
- d) Ik begin in één van mijn meegebrachte boeken
- e) Ik begin in een gedeeltelijke crunch-positie
- f) Ik begin in een heel rustige draf

Voorbeeld 17

- a) Hij begon in een lichtbeveiligde afdeling
- b) Hij begon in een paar oude arbeiderswoningen aan de Boelekade

c) Ik begin in een appartement

In de zinnen in voorbeelden 16 en 17 is steeds te zien hoe er sprake is van een containerfunctie. Een verschil tussen beide voorbeelden is echter dat bij de eerste ook informatie wordt gegeven over de activiteit, terwijl dit bij de tweede onduidelijk blijft. Een contrast dat hierbij dient te worden opgemerkt is dat de nomina in voorbeeld 17 concrete locaties zijn, in tegenstelling tot die in voorbeeld 16, waar sprake is van abstracte locaties. Bij de eerste moet informatie over de activiteit geheel uit de context worden gehaald. Wanneer de locatie abstract is, blijkt er echter ook al duidelijkheid te bestaan over de activiteit. Blijkbaar is er in het geval van een abstracte containerfunctie toch sprake van coercie, waardoor een activiteit wordt geïmpliceerd.

4. Conclusie

4.1. Verband tussen preposities en implicatie van activiteit

Hoewel de gegevens vanuit de zinnen gebaseerd op het nomen *boek* veelbelovend waren, blijkt er niet in de verwachte mate een structureel verband te zijn tussen het gebruik van bepaalde preposities en de implicatie van activiteiten. We hebben gezien dat het niet mogelijk is te voorspellen wanneer welke preposities kunnen worden gebruikt in combinatie met een nomen. Met betrekking tot de qualiastructuur van Pustejovsky blijkt dat er geen sterke link te maken is tussen preposities en bijbehorende qualiarollen. De hypothese dat zinnen zonder prepositie en zinnen met de prepositie *aan* verwezen naar de agentieve rol, zinnen met *in* naar de telische rol en zinnen met *met* naar beide rollen, werd weliswaar niet ontkracht, maar vond ook te weinig bevestiging van data om bewezen te worden geacht.

De oorzaak kan liggen in het feit dat wellicht niet alle nomina alle vier de qualiarollen hebben. Dit zou nader onderzocht kunnen worden in verder onderzoek door per qualiarol na te gaan wat de invulling ervan is bij een nomen. Hierbij zou dan ook geconcludeerd kunnen worden dat een nomen deze rol niet vervult. Een andere optie is dat nomina nog verder dienen te worden gecategoriseerd, waarna er per groep heldere lijnen te onderscheiden zijn. Een poging een categorisatie te maken op basis van concreet vs. abstract en grootte heeft echter geen resultaat opgeleverd: nomina die hierdoor in dezelfde categorie vielen, bleken niet inwisselbaar. Aangezien de geteste verdeling een zeer intuïtieve betrof, maar direct werd ontkracht, lijkt ook het zoeken naar andere criteria weinig hoopvol. Er is echter niet uitgesloten dat er wel criteria te vinden zouden zijn voor een correcte subcategorisatie waarbij de bovengenoemde theorie wel wordt bevestigd.

Al met al lijkt de informatie van de activiteit meer uit het nomen te komen dan uit de prepositie, en wordt in specifieke gevallen door de prepositie bepaald welke activiteit het dan betreft. De resultaten van *boek*, waarmee de uiteenzetting begonnen is en waar de prepositie een veel grotere rol lijkt te spelen, lijkt dan ook voornamelijk een bijzonder geval te zijn.

4.2. Positieve resultaten

Ondanks de tegenvallende resultaten op het gebied van de link tussen preposities en hun implicaties, zijn er wel andere interessante bevindingen. De voornaamste daarvan is de afwijkende werking van de prepositie *in*, ten opzichte van *aan* en *met*, die telkens een containerfunctie veroorzaakt. Dit kan voorkomen als concrete locatie, waarbij de activiteit volledig moet blijken uit de context, of als abstracte locatie, waarbij wel degelijk informatie wordt gegeven over de activiteit die plaatsvindt. In deze abstractere gevallen van de containerfunctie is blijkbaar wel een vorm van coercie aanwezig, waardoor de activiteit geïmpliceerd wordt.

Verder blijkt de toevoeging van preposities zeker niet in elk geval extra informatie te verschaffen. In meerdere gevallen blijken zinnen zonder prepositie en zinnen met de prepositie *aan* exact dezelfde betekenis te hebben. Ook blijft de betekenis tussen verschillende preposities onderling gelijk. Waar ik in de inleiding aangaf dat de theorie van iconiciteit stelt dat meer woorden ook een complexere betekenis bewerkstelligde – meer woorden

betekent meer betekenis – blijkt hier uit de resultaten van het onderzoek dat dit niet per se het geval hoeft te zijn. Daarentegen blijkt juist dat bij zinnen als in voorbeeld 18 dat bepaalde preposities niet mogelijk zijn, terwijl een zin zonder prepositie wel degelijk een activiteit impliceert. Hiermee is er een sterk argument tegen de iconiciteitstheorie gevonden.

Voorbeeld 18

- a) Hij begon een relatie
- b) Hij begon aan een relatie
- c) *Hij begon in een relatie
- d) ?Hij begon met een relatie

4.3. Mogelijkheden voor vervolgonderzoek

Zoals aangegeven in sectie 3.2 bestaat er nog de mogelijkheid dat de qualiastructuur simpelweg niet altijd compleet hoeft te zijn voor elk nomen. Zo zou het kunnen zijn dat sommige nomina bijvoorbeeld geen telische rol kennen, waardoor daar ook niet naar verwezen kan worden door preposities. Bovendien kan het zijn dat er juist ook meer qualiarollen te formuleren zijn, waardoor preposities specifiek te werk kunnen gaan. Op die manier zouden preposities naar verschillende qualiarollen kunnen verwijzen. Hierbij zou dan standaard de mogelijkheid kunnen worden meegenomen dat niet elke qualiarol vervuld hoeft te worden. Wanneer dan van een nomen bekend is dat bepaalde qualiarollen niet aanwezig zijn, kan de combinatie met de daarnaar verwijzende preposities worden uitgesloten als grammaticale zin.⁸ Binnen de resultaten zoals die hierboven gevonden zijn, wordt dit niet ontkracht. Om dit te testen moet echter per nomen worden gekeken welke qualiarollen er kunnen worden opgesteld, en welke qualiarollen voldoende zouden zijn voor een adequaat systeem. Wanneer dit gedaan zou zijn, kan er opnieuw getest worden of hierboven genoemde hypothese na de beschreven uitbreiding alsnog bevestigd zou kunnen worden.

Bij mijn onderzoek heb ik mij beperkt tot de preposities *aan*, *met* en *in*. Er zijn echter uiteraard veel meer preposities. Het meenemen van meer preposities zou echter een vele malen groter werk tot gevolg hebben gehad. Voor een completer beeld van de kracht van de qualiastructuur zouden deze andere preposities alsnog in acht moeten worden genomen, om te zien of er daarmee ook meer informatie kan worden voorspeld, zonder dat deze specifiek staat vermeld in de zin. Samen met de bovengenoemde uitbreidingen zou dit kunnen leiden tot een definitief oordeel over de houdbaarheid van Pustejovsky's theorie, toegepast op het Nederlands.

⁸ Mits de volgorde niet in een andere constructie kan ontstaan. Wanneer dit wel gebeurt, zou dit echter via andere regels van het systeem deze zinnen moeten worden goedgekeurd.

Referenties

- Egg, M.: *Beginning Novels and Finishing Hamburgers: Remarks on the Semantics of 'to begin'*. Journal of Semantics 20: 163-191, ©Oxford University Press 2003
- Fodor, J. en Lepore, E.: *The emptiness of the lexicon: Reflections on James Pustejovsky's 'The Generative Lexicon'*. Linguistic Inquiry 29: 269-288, 1998
- Haspelmath, M.: *Iconicity versus frequency in explaining grammatical asymmetries*. Cognitive Linguistics 19.1:1-33, 2008
- Janssen, M.: *Een kritische beschouwing van de Generative Lexicon Theory*, 1997 (<http://maarten.janssenweb.net/Papers/link.pdf>)
- McIntyre, A.: *Functional interpretations: borderline idiosyncrasy in prepositional phrases and other expressions*. University of Leipzig, 2001
- Pustejovsky, J.: *The generative lexicon*. Computational Linguistics Volume 17, Number 4. © 1991 Association for Computational Linguistics
- Pustejovsky, J.: *The Syntax of Event Structure*. Cognition 41 (1991): 47-81 © 1991 Elsevier
- Pustejovsky, J.: *Type coercion and lexical selection*. Semantics and the Lexicon: 73-94 © 1993 Klüwer Academic Publishers
- Pustejovsky, J. et al.: *Lexical Semantic Techniques for Corpus Analysis*. Computational Linguistics Volume 19, Number 2, © 1991 Association for Computational Linguistics
- Pustejovsky, J.: *Events and the Semantics of Opposition*. Events as Grammatical objects, © 2000 CSLI Publications
- Sweep, J.: *Logische metonymie in het directe object*. PhD proposal, Universiteit van Amsterdam, 2006
- Verspoor, C.M.: *Contextually-Dependent Lexical Semantics*. PhD-Thesis, University of Edinburgh, 1997
- Amsterdam Center for Language and Communication (AACL): *the metonymy of logical metonymy, analyzing real language data*. 1991. (Slides)