



Universiteit Utrecht

FACULTEIT GEESTESWETENSCHAPPEN

KUNSTMATIGE INTELLIGENTIE

Complexiteitsbepaling van incrementeel opgebouwde bewijsnetten met kruisende axiomaverbindingen

Auteur:
Elynn WEIJLAND
6418651

Begeleider:
Gijs WIJNHOLDS

Tweede lezer:
Frans ADRIAANS

BACHELORSRIPTIE, 7.5 ECTS

2 juli 2021

Abstract

Dit onderzoek bestudeert bewijsnetten met kruisende axiomaverbindingen en geeft hierbij een complexiteitsbepaling. Morrill (2000) beschrijft een manier waarop bewijsnetten incrementeel opgebouwd kunnen worden. In deze bewijsnetten wordt gebruik gemaakt van de Lambek Calculus (Lambek, 1958). Met behulp van deze bewijsnetten kan een complexiteitscurve gevormd worden die het aantal ongebonden axioma's weergeeft voor elke stap bij het vormen van het bewijsnet. Het uitgangspunt van deze complexiteitsbepaling is dat de bewijsnetten geen kruisende axiomaverbindingen bevatten in de visuele weergave. Bewijsnetten met kruisende axiomaverbindingen bestaan echter wel in bepaalde Nederlandse zinsdelen. In deze studie zijn de ambigue zinsdelen "Dieren die mensen eten" en "Een dier dat een mens eet" geanalyseerd. De typen van de woorden in deze zinsdelen zijn verkregen uit het ÆTHEL corpus (Kogkalidis, Moortgat & Moot, 2020a) en hebben een richting toegekend gekregen. Vervolgens zijn de bewijsnetten gevormd die behoren bij de twee lezingen van de bestudeerde zinsdelen. Twee van deze bewijsnetten bevatten kruisende axiomaverbindingen. Om de complexiteit te kunnen bepalen van de bewijsnetten in dit onderzoek, is de complexiteitsbepaling van Morrill (2000) aangepast. Daarbij is per stap het aantal ongebonden axioma's geteld, waar het aantal kruisingen van axiomaverbindingen bij is opgeteld. Uit de complexiteitscurves is af te leiden dat de lezing met "dieren" en "een dier" als subject van het zinsdeel voorkeur heeft over de lezing met "mensen" of "een mens" als subject. Dit resultaat is vergeleken met het ÆTHEL corpus, waaruit blijkt dat men deze lezing daadwerkelijk preferereert. Er mag voorzichtig geconcludeerd worden dat de geïntroduceerde complexiteitsbepaling in de toekomst toegepast kan worden op andere bewijsnetten met kruisende axiomaverbindingen.

Keywords: *ambiguïteit, intuïnzinnen, incrementeel parseren, Lambek Calculus, bewijsnetten, ÆTHEL, complexiteit.*

Inhoudsopgave

1	Introductie	3
2	Achtergrond	5
2.1	Lambek Calculus	5
2.2	Incrementele bewijsnetten	6
2.3	Complexiteit	9
2.4	Æthel	11
3	Methode	14
4	Resultaten	20
5	Discussie en Conclusie	21
	Bibliografie	24

Hoofdstuk 1

Introductie

In een gesprek tussen twee gesprekspartners ontvangt de luisteraar informatie op een incrementele manier. Dit houdt in dat de informatie in een zin woord voor woord verkregen wordt. Voor een luisteraar is, nadat het eerste woord is ontvangen, het verloop van de zin niet bekend. Echter, de luisteraar kan in een groot deel van de gevallen dit zinsverloop voorspellen (Harley, 2017). In de taalwetenschap, een belangrijk onderzoeksgebied binnen de kunstmatige intelligentie, wordt breed aangenomen dat hier sprake is van incrementeel parsen. De luisteraar maakt een grammaticale structuur aan, gebaseerd op de informatie die hij of zij heeft ontvangen, en past deze structuur aan naarmate er meer informatie wordt verkregen. De grammaticale structuur wordt aangemaakt volgens minimal attachment (Harley, 2017). Dit houdt in dat de structuur altijd zo economisch mogelijk is, oftewel dat de ontledingboom zo min mogelijk knopen bevat. Nadat het laatste woord van de zin is ontvangen door de luisteraar is de grammaticale structuur compleet. Deze syntactische informatie gebruikt de luisteraar vervolgens, onder andere, om de betekenis van de zin te achterhalen.

Echter, ambiguïteit gooit roet in het eten bij deze methode van incrementeel parsen. Globaal ambigue zinnen, zoals “Jan zag de man met de verrekijker”, hebben twee syntactische structuren. De luisteraar maakt een van deze twee structuren aan, maar deze structuur komt niet altijd overeen met de lezing die de spreker voor ogen heeft. Hierdoor kan er een probleem ontstaan in de communicatie. Intuïnzinnen zorgen op dezelfde manier voor verwarring. Deze zinnen zijn lokaal ambigu (Harley, 2017), wat inhoudt dat er aan het begin van de zin sprake is van ambiguïteit, maar dat deze wordt opgelost naarmate de zin vordert. De meest bekende intuïnzin is de volgende: “The horse raced past the barn fell.” (Bever, 1970). Volgens minimal attachment wordt er verwacht dat de zin eindigt na het woord “barn”. Na het horen van het woord “fell” moet de syntactische structuur die de lezer of luisteraar heeft aangemaakt aangepast worden. Dit zorgt voor een verwarringseffect. Ambigue zinnen, zinsdelen en intuïnzinnen bevatten dus verschillende grammaticale structuren, en hierdoor zijn verschillende lezingen mogelijk.

Onderzoeker Morrill (2000) heeft in zijn artikel deze ambiguïteiten geprobeerd te vangen met bewijsnetten. Hij heeft een manier bedacht waarop deze bewijsnetten incrementeel opgebouwd kunnen worden, op dezelfde manier als een luisteraar informatie ontvangt. De Lambek Calculus (Lambek, 1958) wordt gebruikt om deze bewijsnetten te kunnen vormen. Hierbij heeft elk woord een type, dat gelijk staat aan een formule binnen de Lambek Calculus. Met behulp van formuleontvouwingen kunnen de typen opgesplitst worden in axioma's. Deze worden incrementeel gecombineerd, waardoor een bewijsnet ontstaat. Voor intuïnzinnen is er één bewijsnet mogelijk, omdat aan het eind van de zin maar één lezing mogelijk is. Morrill (2000) vergelijkt in deze gevallen het bewijsnet van de intuïnzin met het bewijsnet van de zin met de aangenomen structuur. Bij “The horse raced past the barn fell” is dit “The horse raced past the barn.” Voor globaal ambigue zinnen en zinsdelen zijn twee bewijsnetten mogelijk, een voor elke lezing. Op basis van de gevormde

bewijsnetten creëert Morrill (2000) een complexiteitscurve, waarbij (x,y) staat voor complexiteit y in stap x . Volgens zijn theorie heeft de lezing met een lagere complexiteit, dat wil zeggen de som van de per stap opgetelde complexiteit, de voorkeur ten opzichte van de lezing met een hogere complexiteit. Deze theorie blijkt voor de door hem bestudeerde gevallen juist. Morrill (2000) bespreekt verschillende Engelse ambigue zinnen en bepaalt hiervan de complexiteit. Uitgangspunt is dat voor al deze gevallen geldt dat de bewijsnetten van alle lezingen geen kruisende axiomaverbindingen bevatten in de visuele weergave van het bewijsnet. Echter, voor bepaalde Nederlandse zinsdelen bevatten de bewijsnetten wél kruisende axiomaverbindingen.

In dit onderzoek staat de volgende vraag centraal: “Hoe kan de complexiteit van een zin(sdeel) bepaald worden als blijkt dat een of meer bewijsnetten van deze zin(sdelen) een of meer kruisende axiomaverbindingen bevatten?”

Er worden in dit onderzoek twee ambigue zinsdelen bestudeerd, waarbij een van de lezingen kruisende axiomaverbindingen bevat in het bewijsnet. In de achtergrondsectie zal in detail beschreven worden hoe Morrill (2000) zijn bewijsnetten incrementeel opbouwt en hoe hij hier vervolgens de complexiteit van bepaalt. Ook wordt hier het ÆTHEL corpus (Kogkalidis et al., 2020a) geïntroduceerd. Met behulp van de python code van Kogkalidis (2021) kunnen uit dit corpus de woordtypen in de geanalyseerde zinnen gevonden worden. Deze typen bevatten echter geen richting. De typen zullen zo worden aangepast, dat zij wel een richting bezitten. Hierdoor wordt woordvolgorde gemodelleerd en benadert de methode van dit onderzoek de methode van Morrill (2000) zo dicht mogelijk. Vervolgens zal in de methodesectie een manier beschreven worden waarmee de complexiteit van de gevormde bewijsnetten, inclusief kruisende axiomaverbindingen, bepaald kan worden. Hieruit zal geconcludeerd worden welk van de lezingen van de bestudeerde zinsdelen de voorkeur heeft. Vervolgens zal deze conclusie vergeleken worden met de conclusie die getrokken kan worden uit het ÆTHEL corpus, om te bepalen of men daadwerkelijk de voorkeur heeft voor deze lezing. Op basis van de geïntroduceerde complexiteitsbepaling zal de voorkeur van een lezing van een zin bepaald kunnen worden op basis van het bewijsnet, ook als dit bewijsnet wel kruisende axiomaverbindingen bevat.

Hoofdstuk 2

Achtergrond

2.1 Lambek Calculus

In de introductie is aangegeven dat het vormen van bewijsnetten mogelijk gemaakt wordt door de regels van de Lambek Calculus (Lambek, 1958). Door middel van deze calculus kan er bepaald worden of een zin grammaticaal correct is. Elk woord heeft een type, dat gelijk staat aan een formule binnen de Lambek Calculus. De formules worden met elkaar gecombineerd en hierop worden de regels van de calculus toegepast. De formules binnen de Lambek Calculus zijn atomair, of bevatten de binaire connectieven / ('over'), ('onder') of \otimes ('product'). De formules die gecombineerd worden door deze connectieven, kunnen op zichzelf ook weer connectieven bevatten. Bij de categoriale basislogica staan de volgende zes regels centraal (Lambek, 1958):

$$\frac{}{A \rightarrow A} 1_A \quad \frac{A \rightarrow B \quad B \rightarrow C}{A \rightarrow C} \circ \quad (2.1)$$

$$\frac{A \otimes B \rightarrow C}{A \rightarrow C/B} \triangleright \quad \frac{A \rightarrow C/B}{A \otimes B \rightarrow C} \triangleright^{-1} \quad \frac{A \otimes B \rightarrow C}{B \rightarrow A \setminus C} \triangleleft \quad \frac{B \rightarrow A \setminus C}{A \otimes B \rightarrow C} \triangleleft^{-1} \quad (2.2)$$

De eerste formule in (2.1) staat voor reflexiviteit, de tweede formule voor transitiviteit. De vier formules in (2.2) zijn residuatierregels. Het combineren van deze regels definieert alle mogelijke afleidingen van een zin. Er ontstaat een afleiding, oftewel de zin is grammaticaal juist, als elke formule opgebroken kan worden in axioma's waarop de reflexiviteitsregel kan worden toegepast. Op deze manier kunnen ook de applicatie- en coapplicatierregels worden gevormd. De regels zijn weergegeven in (2.3) (Lambek, 1958), hun afleiding in (2.4).

$$\frac{}{B \otimes B \setminus A \rightarrow A} \text{app} \setminus \quad \frac{}{A/B \otimes B \rightarrow A} \text{app} / \quad (2.3)$$

$$\frac{}{A \rightarrow B \setminus (B \otimes A)} \text{coapp} \setminus \quad \frac{}{A \rightarrow (A \otimes B)/B} \text{coapp} /$$

$$\frac{}{B \setminus A \rightarrow B \setminus A} 1_{B \setminus A} \quad \frac{}{A/B \rightarrow A/B} 1_{A/B}$$

$$\frac{}{B \otimes (B \setminus A) \rightarrow A} \triangleleft^{-1} \quad \frac{}{(A/B) \otimes B \rightarrow A} \triangleright^{-1} \quad (2.4)$$

$$\frac{}{B \otimes A \rightarrow B \otimes A} 1_{B \otimes A} \quad \frac{}{A \otimes B \rightarrow A \otimes B} 1_{A \otimes B}$$

$$\frac{}{A \rightarrow B \setminus (B \otimes A)} \triangleleft \quad \frac{}{A \rightarrow (A \otimes B)/B} \triangleright$$

De Lambek Calculus kan verder uitgebreid worden met de operaties \diamond en \square (Gherke, Nagahashi & Venema, 1958). Deze operaties geven in de context van dit onderzoek een dependentierol aan voor een bepaald type. Dependenterollen geven informatie over onder andere de rol van het type binnen de zin. Bijvoorbeeld, $\diamond_{sub} \text{NP}$ geeft aan

dat het type NP het onderwerp, het subject, is van de zin. Er geldt voor de operaties dat zij monotoon zijn en dat de compositie $\diamond\Box$ contraherend is, terwijl de compositie $\Box\diamond$ juist expanderend is. De composities zijn weergegeven in (2.5), de monotoniteit in (2.6) (Gherke et al., 1958).

$$\diamond\Box A \rightarrow A \quad A \rightarrow \Box\diamond A \quad (2.5)$$

$$\frac{A \rightarrow B}{\diamond A \rightarrow \diamond B} \quad \frac{A \rightarrow B}{\Box A \rightarrow \Box B} \quad (2.6)$$

2.2 Incrementele bewijsnetten

Om de theorie van Morrill (2000) te kunnen gebruiken, en de complexiteitsbepaling aan te passen, is het van belang om te begrijpen hoe deze manier van bewijsnetten vormen in elkaar steekt. Morrill (2000) maakt gebruik van de volgende *logical links*, die formules uit de Lambek Calculus (Lambek, 1958) ontvouwen, weergegeven in (2.7).

$$\frac{A^\circ \quad B^\bullet}{A \setminus B^\bullet} \text{ ii} \quad \frac{B^\circ \quad A^\bullet}{A \setminus B^\circ} \text{ i} \quad (2.7a)$$

$$\frac{B^\bullet \quad A^\circ}{B / A^\circ} \text{ ii} \quad \frac{A^\bullet \quad B^\circ}{B / A^\circ} \text{ i} \quad (2.7b)$$

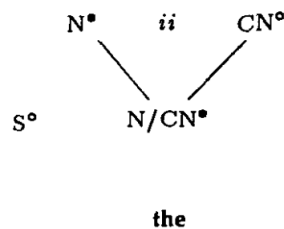
$$\frac{A^\bullet \quad B^\bullet}{A \cdot B^\bullet} \text{ i} \quad \frac{B^\circ \quad A^\circ}{A \cdot B^\circ} \text{ ii} \quad (2.7c)$$

Bovenstaande links bestaan uit formules. De formule onder de streep wordt door de link opgebroken in twee blaadjes. Deze blaadjes zelf zijn ook weer formules. Als deze niet atomair zijn, kan hier ook weer een link op worden toegepast. Alle formules in (2.7) bevatten een polariteit, ze worden gemarkeerd met een input (\bullet) of output (\circ) label. Afhankelijk van de formule onder de streep, worden de blaadjes van deze formule vervolgens gemarkeerd met een input of output label. Morrill (2000) maakt verder gebruik van de definities van bewijsframe, axiomaverbinding en bewijsstructuur. Een bewijsframe is een eindige reeks van formulebomen. Een formuleboom ontstaat door de regels in (2.7) toe te passen op een formule totdat er axioma's overblijven. De reeks van formulebomen moet precies één boom bevatten met een *root*, een formule die geen blaadje is, voorzien van een output label. Een axiomaverbinding is een verzameling van axioma's, waarbij de axioma's met elkaar worden verbonden tot paren. De axioma's die verbonden worden hebben verschillende polariteiten. Hier geldt dat deze axiomaverbindingen elkaar niet mogen kruisen in de visuele weergave van het bewijsnet. Een bewijsstructuur combineert een bewijsframe en zijn axiomaverbindingen; de axioma's van de reeks van formulebomen worden gecombineerd door middel van axiomaverbindingen. De gevormde bewijsstructuur is een bewijsnet als een cykel in deze bewijsstructuur een i-link bevat. De drie i-links zijn weergegeven in (2.7). Nu bekend is wat een bewijsnet is, en wat er nodig is om deze te vormen, kan een bewijsnet incrementeel opgebouwd worden.

Elk woord in een taal heeft een bepaald type. Afhankelijk van of de semantiek of de syntax wordt bestudeerd, krijgt een woord een verschillend type toegewezen. Aan een eigennaam wordt in de syntax het type NP, noun phrase, toegekend. Hierop gebaseerd heeft het woord in de semantiek het type e, voor entiteit (Winter, 2016). Het type NP geeft aan in welke context het woord voor kan komen, het type e geeft aan wat het woord kan betekenen. In dit onderzoek worden bewijsnetten gebouwd, waarin de syntactische typen gebruikt worden. Aan de hand van de bekende intuïzie van Bever (1970) zal hier beschreven worden hoe een bewijsnet incrementeel opgebouwd wordt. Morrill (2000) neemt de volgende lexicale toekenning aan:

$$\begin{aligned}
 \textit{barn} &:= \textit{CN} \\
 \textit{horse} &:= \textit{CN} \\
 \textit{past} &:= ((\textit{N}\backslash\textit{St})\backslash(\textit{N}\backslash\textit{St}))/\textit{N} \\
 \textit{raced} &:= \textit{N}\backslash\textit{S}+ \\
 \textit{the} &:= \textit{N}/\textit{CN}
 \end{aligned}
 \tag{2.8}$$

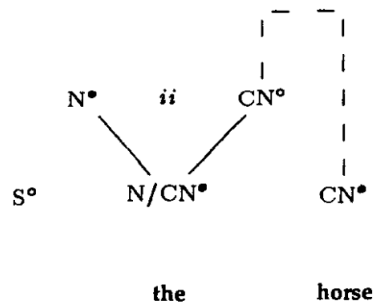
Het symbool + staat hier voor een tijdsmarkering, de verleden tijd van het woord “race”. Tijdens het vormen van bewijsnetten is de tijdsmarkering niet van belang, deze wordt weggelaten. Morrill (2000) doet in zijn artikel de aanname dat de luisteraar zal verwachten dat de spreker een zin zal uiten. Het bewijsnet van een zin begint hierdoor altijd met een type S, voor “sentence”, met een output label. Alle woorden die een luisteraar waarneemt, krijgen het input label. Het eerste woord dat wordt waargenomen, “the”, heeft het type N/CN toegekend gekregen. Vervolgens wordt met de ii-link in (2.7b) het type van N/CN[•] opgebroken tot de axioma’s N[•] en CN[°]. Aangezien alle axioma’s verschillende typen hebben, kunnen er nog geen axiomaverbindingen gevormd worden. De situatie die ontstaat is weergegeven in Figuur 2.1 (Morrill, 2000).



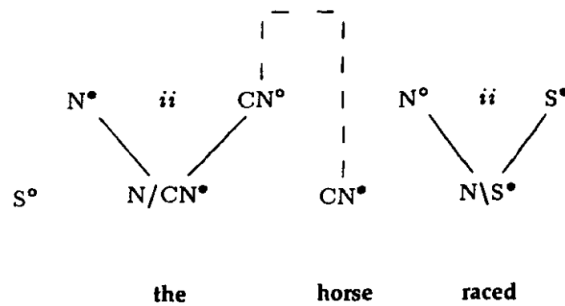
FIGUUR 2.1: Incrementeel bewijsnet: verwachting en “the” (Morrill, 2000).

Hierna neem de luisteraar het woord “horse” waar. Dit heeft type CN, gemarkeerd met een input label. Dit type is op zichzelf al atomair. Er is nu een axiomaverbinding mogelijk tussen CN[•] van het type van “horse”, en CN[°] van het type van “the”. Deze axiomaverbinding is weergegeven in Figuur 2.2.

Daaropvolgend hoort de luisteraar het woord “raced”, met type N\ S. Na het toepassen van de ii-link in (2.7a) komen de axioma’s N[°] en S[•] tot stand. Er zijn nu twee mogelijkheden. De axioma’s N en S kunnen met elkaar verbonden worden. Echter, in deze situatie zijn alle axioma’s met elkaar verbonden, zijn er geen kruisende axiomaverbindingen en bevat dit bewijsnet geen cyclen. Er is in deze situatie een bewijsnet gevormd van de zin “The horse raced.” Aangezien deze zin hier niet van belang is, zullen deze axiomaverbindingen niet ontstaan. Dit te zien is in Figuur 2.3.



FIGUUR 2.2: Incrementeel bewijsnet: verwachting en "the horse" (Morrill, 2000).

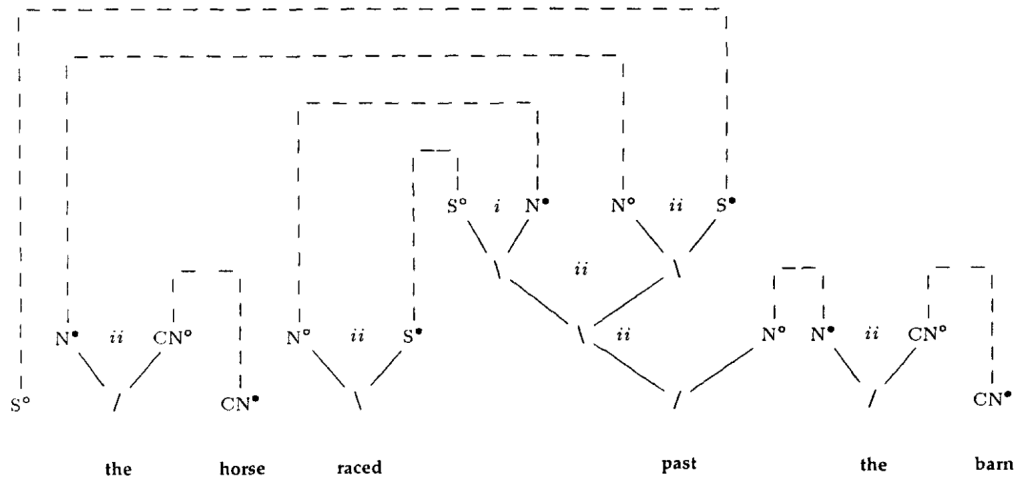


FIGUUR 2.3: Incrementeel bewijsnet: verwachting en "the horse raced" (Morrill, 2000).

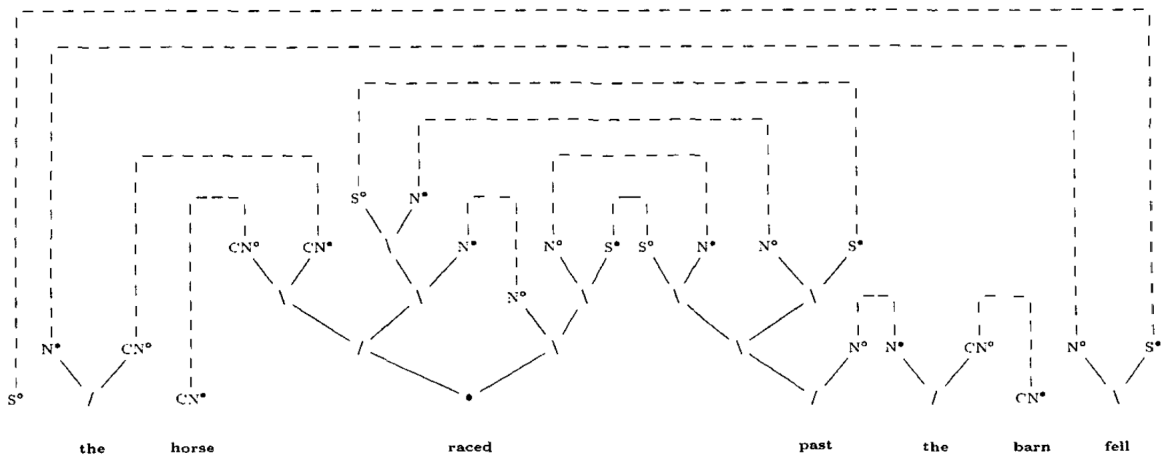
In Figuur 2.4 is de analyse van de gehele zin weergegeven. Het is een bewijsframe, want er is precies één formuleboom met een root voorzien van een output label, namelijk de verwachting S° van de luisteraar. Er zijn axiomaverbindingen gevormd tussen alle axioma's, en deze verbindingen kruisen elkaar niet. Verder is er een cykel aanwezig in de analyse, tussen de axioma's N° en S^\bullet van het type van "raced" en de axioma's N^\bullet en S° van het type van "past". Deze laatste twee axioma's zijn voortgekomen uit de formule $N \setminus S^\circ$. De *i*-link in (2.7a) is hier toegepast. De cykel in de analyse bevat dus een *i*-link. Hieruit volgt dat de analyse in Figuur 2.4 een bewijsnet is (Morrill, 2000). Op dezelfde manier kan het bewijsnet van de garden path sentence, "The horse raced past the barn fell." (Bever, 1970), gevormd worden. De lexicale toekenning is hier als volgt:

$$\begin{aligned}
 \textit{barn} &:= CN \\
 \textit{horse} &:= CN \\
 \textit{past} &:= ((N \setminus St) \setminus (N \setminus St)) / N \\
 \textit{raced} &:= ((CN \setminus CN) / (N \setminus (N \setminus S^-))) \cdot (N \setminus (N \setminus S^-)) \\
 \textit{the} &:= N / CN \\
 \textit{fell} &:= N \setminus S^+
 \end{aligned} \tag{2.9}$$

Er geldt weer dat de symbolen + en - staan voor een tijdsbepaling, deze kunnen weggelaten worden in het bewijsnet. Het uiteindelijke bewijsnet van deze zin is weergegeven in Figuur 2.5.



FIGUUR 2.4: Analyse van "The horse raced past the barn" (Morrill, 2000).



FIGUUR 2.5: Analyse van "The horse raced past the barn fell" (Morrill, 2000).

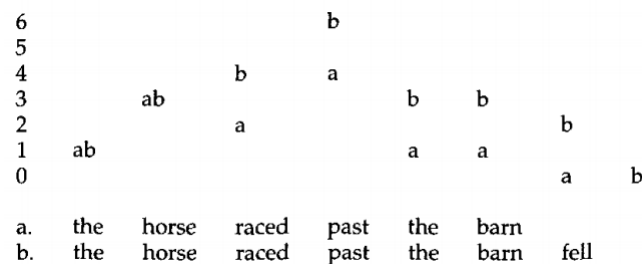
Ook in dit geval is er sprake van een bewijsnet, want er is aan alle condities voldaan: er is één formuleboom met een output label, de axiomaverbindingen kruisen elkaar niet en de twee cyclen in de analyse bevatten een i-link.

2.3 Complexiteit

Nu de bewijsnetten van de zinnen "The horse raced past the barn" en "The horse raced past the barn fell" (Bever, 1970) zijn gevormd, kan de complexiteit van deze zinnen bepaald worden. De complexiteitsbepaling van Morrill (2000) is gebaseerd op de aanname dat de complexiteit van de zin voor de luisteraar afhangt van de mate waarin het geheugen wordt belast. Hieruit volgt het *Principle of Acceptability* (Morrill, 2000): "Acceptability is inversely proportional to the sum in time of the memory load of unresolved valencies." Er geldt dat voor elk punt in de tijd het aantal ongebonden axioma's onthouden moet worden. Als deze "unresolved valencies" voor elk punt worden opgeteld, is de totale complexiteit gevonden. De analyse, ofwel de lezing, met de laagste totale complexiteit heeft de voorkeur ten opzichte van een lezing met een hogere complexiteit. Volgens Morrill (2000) geldt, als alle andere

factoren constant zijn, dat zijn systeem van complexiteitsbepaling een goede voorspelling levert voor de geprefereerde lezing van de luisteraar. Bij synonymie, zoals de bekende intuïzin “The horse raced past the barn” van Bever (1970), is de semantiek constant. Dit geldt ook bij ambigue zinnen en zinsdelen. Op basis van deze constante factor kan voor ambigue zinnen en zinsdelen de lezing met de hoogste voorkeur worden bepaald.

Morrill (2000) stelt in zijn artikel bij verschillende bewijsnetten een complexiteitscurve op. Dit houdt in dat bij elke stap van het bouwen van het bewijsnet het aantal ongebonden axioma’s wordt geteld. Het gevormde bewijsnet van de zinnen “The horse raced past the barn” en “The horse raced past the barn fell” beginnen beide met de verwachting S° . Er is in deze stap dus één axioma dat niet verbonden is. Vervolgens wordt het woord “the” waargenomen. Als het type van dit woord ontleed is, zijn er twee axioma’s ontstaan. Er zijn nu in totaal drie ongebonden axioma’s. Hierna hoort de luisteraar “horse”. In de zin “The horse raced past the barn” verbindt dit atomaire type met het axioma CN° van het type van “the”. Er zijn nu nog twee axioma’s die niet verbonden zijn. Bij “The horse raced past the barn fell”, zal het axioma CN^\bullet van “horse” niet verbinden met het axioma CN° . In deze zin zijn bij deze stap dus vier ongebonden axioma’s. Voor elke stap, oftewel na elk woord dat ontvangen wordt, wordt het aantal ongebonden axioma’s geteld voor beide zinnen. Deze informatie is vervolgens weergegeven in de complexity curve, te zien in Figuur 2.6.



FIGUUR 2.6: Complexiteitsbepaling van “The horse raced past the barn” en “The horse raced past the barn fell” (Morrill, 2000).

Voor beide zinnen, zin a en zin b, is de complexiteit per stap berekend. Er is te zien dat voor elke stap zin b een gelijke of hogere complexiteit heeft dan zin a. Als het principle of acceptability gevolgd wordt, heeft zin a een geprefereerde lezing ten opzichte van zin b. De voorspelling van Morrill (2000) is dat “raced” liever gezien wordt als hoofdwerkwoord, dan als passief werkwoord in een relatieve bijzin. Vergeleken met de achtergrondliteratuur blijkt de voorspelling van Morrill (2000) inderdaad te kloppen.

Morrill (2000) maakt gebruik van de definitie van axiomaverbinding. Er geldt hier dat de verbindingen tussen de axioma’s van verschillende typen van woorden elkaar niet mogen kruisen. Als dit wel het geval is, voldoet de gevonden analyse niet aan de voorwaarden van een bewijsnet en kan de complexiteit van dit zin(sdeel) niet worden bepaald. Deze voorwaarden van bewijsnetten zijn gebaseerd op de bewijsnetten van Engelse zinnen. In de bewijsnetten van verscheidene Nederlandse zinnen en zinsdelen komen er echter wel kruisende axiomaverbindingen voor. In

dit onderzoek zullen twee voorbeelden geanalyseerd worden en hier wordt vervolgens de complexiteit van bepaald.

2.4 Æthel

Er wordt in deze studie gebruik gemaakt van de ÆTHEL database (Kogkalidis et al., 2020a). Deze database bevat een lexicon van ongeveer 900.000 woorden met hun *supertags*, oftewel hun lineaire typen. De typen van deze woorden zijn gevonden door gebruik te maken van het Lassy Klein-corpus (Van Noord et al., 2013), ook wel “the gold standard corpus of written Dutch” (Kogkalidis et al., 2020a). Het corpus bevat ongeveer een miljoen Nederlandse woorden, inclusief syntactische structuren van verscheidene zinnen. Met een algoritme (Kogkalidis et al., 2020a) kunnen de typen van woorden door deze structuren automatisch verkregen worden. Er is immers bekend in welke contexten een woord voor kan komen. Een woord kan verschillende typen toegekend krijgen, maar sommige woorden hebben ook maar één type. De typen van deze woorden kunnen verder dependentierollen bevatten, aangegeven door de operaties \diamond en \square , maar dit is ook niet altijd het geval. De box wordt gebruikt bij het markeren van een adjunct, een non-head functie. Dit is te zien in (2.10). Er is in de twee bomen aan de rechterzijde al functie applicatie toegepast, B/A of $A \setminus B$, voordat de box operatie is toegepast. Een diamond markeert het hoofd van een functie, en hierna wordt pas de functie applicatie uitgevoerd. In (2.10) is weergegeven dat de diamond operatie alleen is toegepast op A , niet op B/A of $A \setminus B$ in zijn geheel. Hier staat d voor dependentie en h voor hoofd.

$$\begin{array}{cccc}
 \begin{array}{c} B \\ \triangleleft \\ h \quad d \\ \swarrow \quad \searrow \\ B/\diamond_d A \quad A \end{array} &
 \begin{array}{c} B \\ \triangleleft \\ d \quad h \\ \swarrow \quad \searrow \\ A \quad \diamond_d A \setminus B \end{array} &
 \begin{array}{c} B \\ \triangleleft \\ h \quad d \\ \swarrow \quad \searrow \\ A \quad \square_d (A \setminus B) \end{array} &
 \begin{array}{c} B \\ \triangleleft \\ d \quad h \\ \swarrow \quad \searrow \\ \square_d (B/A) \quad A \end{array} &
 (2.10)
 \end{array}$$

Een voorbeeld van een type zonder dependentierollen is het type van het woord “mand”. Dit woord heeft het type N, wat staat voor een noun, een zelfstandig naamwoord. Het type van het woord “de” is in de meeste gevallen $[N \rightarrow NP]$ det (Kogkalidis et al., 2020a). Dit wordt ook wel geschreven als $\square_{det}(N \multimap NP)$ (Moot, 2002). Er geldt hier dat de box, met zijn rol “det”, de dependentierol aangeeft van het gedeelte tussen de haakjes. De diamond operatie werkt op vergelijkbare manier. Het type van het woord “sneuvelen” is $\langle NP \rangle su \rightarrow SMAIN$ (Kogkalidis et al., 2020a). Dit wordt ook wel geschreven als $\diamond_{sub}NP \multimap S$ (Moot, 2002). De NP binnen dit type heeft de dependentierol “sub” toegekend gekregen. Het symbool \multimap geeft in beide gevallen aan dat het type aan de linkerzijde wordt verwacht in de context, vervolgens ontstaat het type aan de rechterzijde. In het eerste geval wordt een type N verwacht in de context, bijvoorbeeld “mand”, en samen met het woord “de” vormt dit dan een noun-phrase determiner, “de mand” (Kogkalidis et al., 2020a; Moot, 2002). Bij het tweede voorbeeld wordt er een NP verwacht in de context, die het subject is van de zin, waarna een zin wordt gevormd. Dit is bijvoorbeeld de NP “Piet”. Door het gebruik van het symbool \multimap is er niet bekend of het type $\square_{det}(N \setminus NP)$ of $\square_{det}(NP/N)$ is. Hetzelfde geldt voor $\diamond_{sub}NP \setminus S$ of $S/\diamond_{sub}NP$. De typen in het lexicon van ÆTHEL bevatten dus geen richting. Dit houdt in dat het type beschrijft in welke context

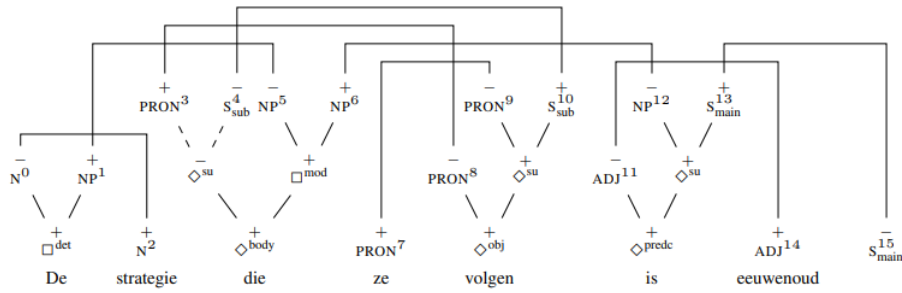
dit woord voor kan komen, maar niet of deze context zich rechts of links van het woord bevindt. Het type van “de” beschrijft dat er een woord met type N in de context voor moet komen en het type van “sneuvelen” doet hetzelfde voor een woord met type NP, maar zij beschrijven niet of het woord in de context links of rechts moet staan. Om de database ÆTHEL te verkleinen is gekozen voor het weergeven van richtingloze typen. Het toevoegen van richting voor alle typen in het corpus zorgt namelijk voor een zeer omvangrijke database, aangezien de typen van woorden vaak op meerdere manieren een richting kunnen bezitten.

ÆTHEL bevat niet alleen een lexicon met woorden en hun typen, de database bevat ook 72.192 analyses van zinnen. Deze zinnen zijn op vier equivalente manieren geanalyseerd: natuurlijke deductie, sequent-style bewijzen, lambdatermen en bewijsnetten. In dit onderzoek is alleen deze laatste manier van belang. Bij bewijsnetten kan namelijk op een incrementele manier een analyse worden gevormd. Zoals Kogkalidis, Moortgat en Moot (2020b) zelf aangeven in hun artikel: natuurlijke deductie heeft, gecombineerd met lambdatermen, in de conclusie vaak meer axioma’s dan in de blaadjes door het aantal hypothesen dat gedaan moet worden. In de sequentcalculi is dit probleem niet te vinden, maar hier zijn oneindig veel bewijzen te vinden voor dezelfde formule, door het herhaaldelijk introduceren en elimineren van connectieven. Bewijsnetten combineren “the best of both worlds”, er is één bewijsnet per lezing te vinden zonder het introduceren van hypothesen. De bewijsnetten in ÆTHEL zijn vergelijkbaar met de bewijsnetten die Morrill (2000) in zijn artikel bespreekt. Het verschil tussen deze twee manieren van bewijsnetten vormen is dat Morrill (2000) het bewijsnet incrementeel opbouwt, terwijl in ÆTHEL de hele zin bekend is voordat het bewijsnet wordt gevormd (Kogkalidis et al., 2020b). Het bewijsnet gebruikt de typen van woorden uit het lexicon van ÆTHEL, en verkrijgt vervolgens de axioma’s door de regels toe te passen die in (2.11) zijn weergegeven.

$$\begin{array}{ccc}
 \begin{array}{c} \bar{A} \quad + \\ \diagdown \quad / \\ \hline A \quad + \quad \circ \quad B \end{array} & \begin{array}{c} \frown \\ \hline + \quad A \quad \bar{A} \quad - \end{array} & \begin{array}{c} + \quad \bar{B} \\ \diagdown \quad / \\ \hline A \quad - \quad \circ \quad B \end{array} & (2.11)
 \end{array}$$

Een positieve polariteit in dit figuur is hetzelfde als een input label dat Morrill (2000) geeft aan zijn formules. Vanzelfsprekend staat negatieve polariteit dan voor een output. De middelste regel staat voor het combineren van axioma’s met twee verschillende polariteiten. Links staat de regel vergelijkbaar met de ii-links in (2.7), rechts staat de regel vergelijkbaar met de i-links. In Figuur 2.7 is het bewijsnet van de zin “De strategie die ze volgen is eeuwenoud” weergegeven.

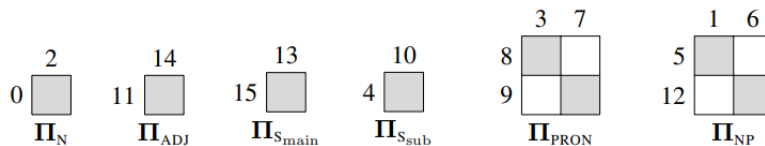
Kogkalidis et al., (2020b) houden hier andere voorwaarden aan voor het vormen van een bewijsnet, er geldt alleen dat er voldaan moet worden aan de correctheids-criteria. Dit zijn dezelfde voorwaarden als die van Morrill (2000), alleen mogen hier wel kruisende axiomaverbindingen voorkomen. De analyse in Figuur 2.7 voldoet hieraan, en is dus een bewijsnet. Er zijn hier duidelijk veel kruisende axiomaverbindingen aanwezig, negen om precies te zijn. Deze verbindingen kruisen elkaar omdat de typen van de woorden geen richting toegekend hebben gekregen. Er moet



FIGUUR 2.7: Analyse “De strategie die ze volgen is eeuwenoud” met ongerichte typen (Kogkalidis, Moortgat & Moot, 2020b).

bij Figuur 2.7 echter wel opgemerkt worden dat de \diamond_{su} in het type van “die” niet correct is, dit hoort een \diamond_{obj} te zijn. In bovenstaand figuur is nu PRON^3 met een subjectrol verbonden met PRON^8 met een objectrol. Dit hoort niet mogelijk te zijn in een juist bewijsnet.

Op basis van het gevonden bewijsnet kunnen er vervolgens meerdere permutatiematrices Π_A (Kogkalidis et al., 2020b) gevormd worden, met een matrix voor elk type van axioma, A . Elk axioma krijgt een cijfer toegewezen. Verticaal worden de axioma’s met negatieve polariteiten uitgezet en horizontaal de axioma’s met positieve polariteiten. Vervolgens wordt een vlak in de matrix grijs gemarkeerd als de twee bijbehorende axioma’s een verbinding vormen in het bewijsnet. Een wit vlak in de matrix geeft aan dat de twee axioma’s in het bewijsnet niet met elkaar verbonden zijn. De permutatiematrices die horen bij het bewijsnet van Figuur 2.7 zijn in Figuur 2.8 weergegeven.



FIGUUR 2.8: Permutatiematrices behorende bij Figuur 2.7 (Kogkalidis, Moortgat & Moot, 2020b).

Door Kogkalidis et al., (2020a), de onderzoekers die de ÆTHEL database hebben gevormd, is een python code ontwikkeld (Kogkalidis, 2021). Met behulp van deze code kan er voor elke mogelijke zin een bewijsnet gevormd worden. Hierbij kunnen de bewijsnetten en hun permutatiematrices weergegeven worden.

Hoofdstuk 3

Methode

In de achtergrondsectie is zojuist beschreven hoe Morrill (2000) bewijsnetten incrementeel opbouwt. Er kan een complexiteitscurve gevormd worden, behorende bij zo'n bewijsnet. De complexiteitscurves van verschillende lezingen van een zin of zinsdeel kunnen vervolgens vergeleken worden. Hieruit kan worden bepaald welke lezing de voorkeur heeft. Het uitgangspunt van Morrill (2000) zijn onderzoek is dat de visuele weergave van dit bewijsnet geen kruisende axiomaverbindingen bevat. Als er gebruik wordt gemaakt van de ongerichte typen uit ÆTHEL (Kogkalidis et al., 2020a) zijn deze kruisende axiomaverbindingen echter wel aanwezig. Zelfs na het geven van richting aan de typen, bevat de visuele weergave van bewijsnetten van bepaalde Nederlandse zinnen toch kruisende axiomaverbindingen. De complexiteitsbepaling van Morrill (2000) kan hier niet op worden toegepast.

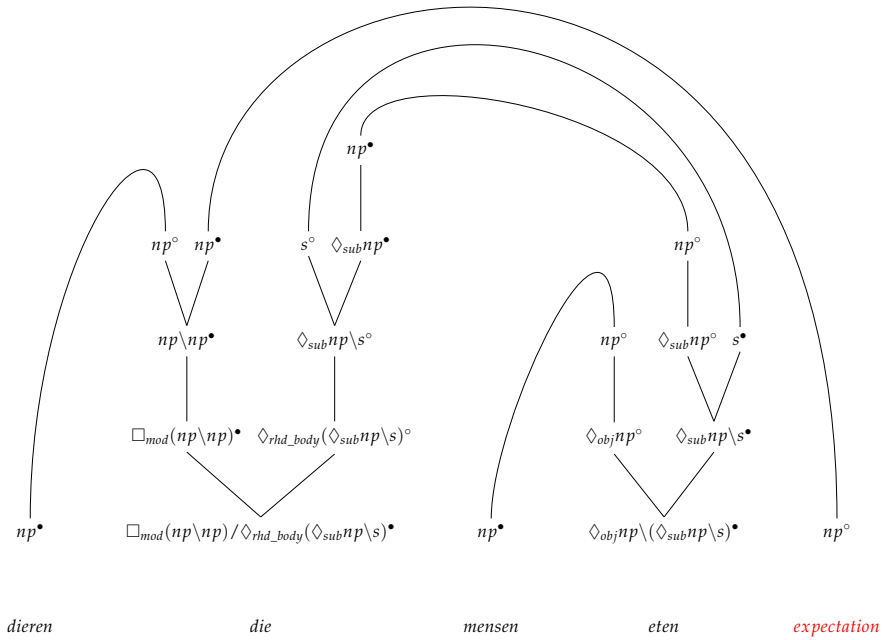
Om de complexiteit bij bewijsnetten met kruisende axiomaverbindingen te kunnen bepalen, zijn er in dit onderzoek twee ambigue zinsdelen geanalyseerd. Er is gezocht naar ambigue zinsdelen, waarbij alle woorden in beide lezingen hetzelfde type hebben. De verwachting bij deze zinsdelen is dat het bewijsnet van één lezing kruisende axiomaverbindingen bevat, terwijl het bewijsnet van de andere lezing deze niet zal hebben. De zinsdelen die zijn geanalyseerd in dit onderzoek zijn "Dieren die mensen eten" en "Een dier dat een mens eet". In het eerste geval kan óf "dieren" óf "mensen" gezien worden als subject, in het tweede geval geldt dit voor "een dier" en "een mens" op dezelfde manier. De gebruikte typen voor de woorden in deze zinsdelen zijn verkregen uit het ÆTHEL corpus. Met behulp van de python code van Kogkalidis (2021) zijn de juiste typen gevonden uit dit corpus. De functie `deco_lexicon()`, aangeropen als bijvoorbeeld `deco_lexicon('dieren')`, uit deze python code geeft voor een woord zijn typen terug uit ÆTHEL, in volgorde van frequentie. Voor alle woorden in de twee bestudeerde zinsdelen is op deze manier het meest voorkomende type in de database gevonden. Vervolgens is gebruik gemaakt van de functie `get_proofs()` (Kogkalidis, 2021) op de volgende manier: `get_proofs(proofs, containing_word_as('dieren', AtomicType('NP')))`. Elk woord wordt samen met het meest voorkomende type in deze functie aangeroepen, waardoor alle bewijsnetten voor alle zinnen in het corpus met het woord "dieren" als NP worden gegeven. Deze functie wordt aangeroepen om te controleren of het gevonden type ook daadwerkelijk de juiste is. Door de gegeven bewijsnetten is te zien of het gegeven woord in de juiste context voorkomt. De lexicale toekenning van de woorden in de bestudeerde zinnen in dit onderzoek is als volgt (Kogkalidis et al., 2020a; Moot, 2002):

$$\begin{aligned}
dieren &:= NP \\
mensen &:= NP \\
eten &:= \diamond_{obj}NP \multimap \diamond_{sub}NP \multimap S \\
die_1 &:= \diamond_{rhd_body}(\diamond_{sub}NP \multimap S) \multimap \square_{mod}(NP \multimap NP) \\
die_2 &:= \diamond_{rhd_body}(\diamond_{obj}NP \multimap S) \multimap \square_{mod}(NP \multimap NP) \\
dier &:= N \\
mens &:= N \\
een &:= \square_{det}(N \multimap NP) \\
dat_1 &:= \diamond_{rhd_body}(\diamond_{sub}NP \multimap S) \multimap \square_{mod}(NP \multimap NP) \\
dat_2 &:= \diamond_{rhd_body}(\diamond_{obj}NP \multimap S) \multimap \square_{mod}(NP \multimap NP)
\end{aligned} \tag{3.1}$$

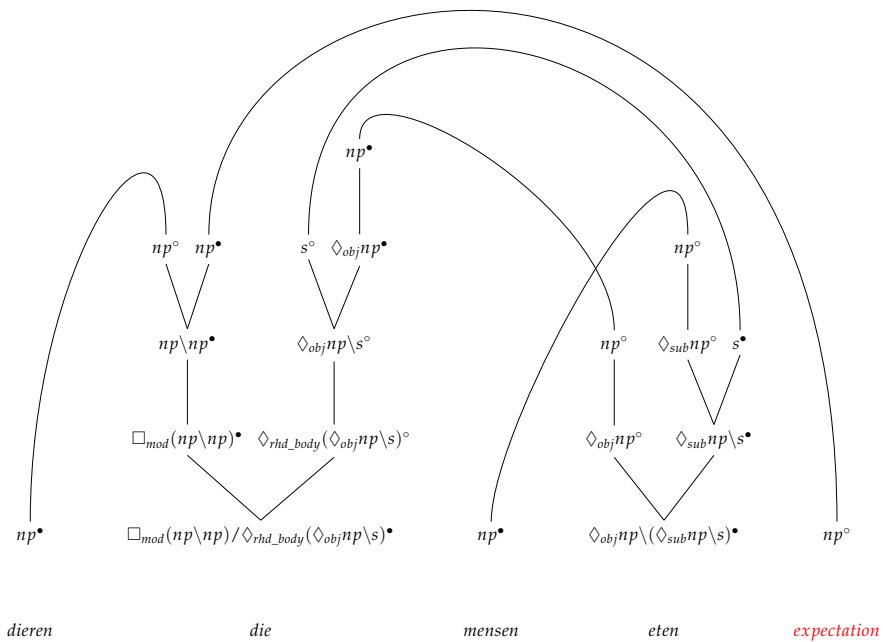
“Die” en “dat” hebben in beide lezingen hetzelfde type, maar verschillende dependentierollen voor het type NP dat zich binnen het type van “die” en “dat” bevindt. Dit is weergegeven in (3.1) met die_1 , die_2 , dat_1 en dat_2 . De gegeven lexicale toekenning bestaat verder uit ongerichte typen, zoals in de achtergrondsectie benoemd is. Er is niet bekend of het symbool \multimap staat voor een / (‘over’), of een \ (‘onder’). Om een bewijsnet te kunnen vormen, moeten deze typen een richting toegekend krijgen. Het geven van richting aan deze typen zorgt ervoor dat de methode van dit onderzoek de methode van Morrill (2000) zo dicht mogelijk benadert. De formuleontvouwingen in (2.7) kunnen namelijk niet op ongerichte typen worden toegepast. Er geldt hier dat de typen een richting moeten krijgen, die in het uiteindelijke bewijsnet zorgt voor zo min mogelijk kruisende axiomaverbindingen. Door middel van trial-and-error is een richting voor alle typen gevonden. Deze typen zijn weergegeven in (3.2).

$$\begin{aligned}
dieren &:= NP \\
mensen &:= NP \\
eten &:= \diamond_{obj}NP \backslash (\diamond_{sub}NP \backslash S) \\
die_1 &:= \square_{mod}(NP \backslash NP) / \diamond_{rhd_body}(\diamond_{sub}NP \backslash S) \\
die_2 &:= \square_{mod}(NP \backslash NP) / \diamond_{rhd_body}(\diamond_{obj}NP \backslash S) \\
dier &:= N \\
mens &:= N \\
een &:= \square_{det}(NP / N) \\
dat_1 &:= \square_{mod}(NP \backslash NP) / \diamond_{rhd_body}(\diamond_{sub}NP \backslash S) \\
dat_2 &:= \square_{mod}(NP \backslash NP) / \diamond_{rhd_body}(\diamond_{obj}NP \backslash S)
\end{aligned} \tag{3.2}$$

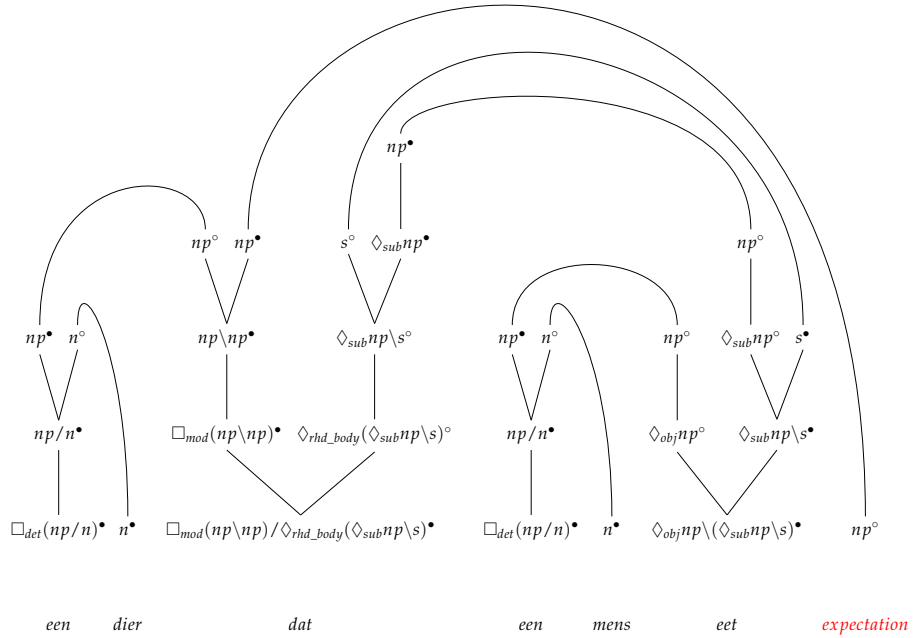
Deze lexicale toekenning is gebruikt om bewijsnetten te vormen voor de twee lezingen van de zinsdelen “Dieren die mensen eten” en “Een dier dat een mens eet”. Op basis van de links in (2.7) zijn de twee zinsdelen geanalyseerd.



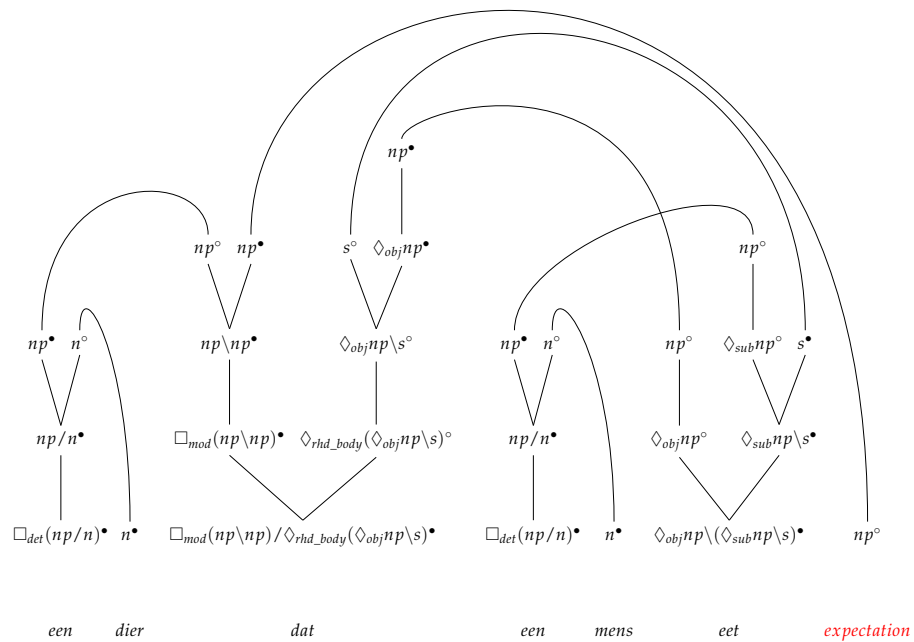
FIGUUR 3.1: Analyse van “Dieren die mensen eten”, “dieren” is subject.



FIGUUR 3.2: Analyse van “Dieren die mensen eten”, “mensen” is subject.

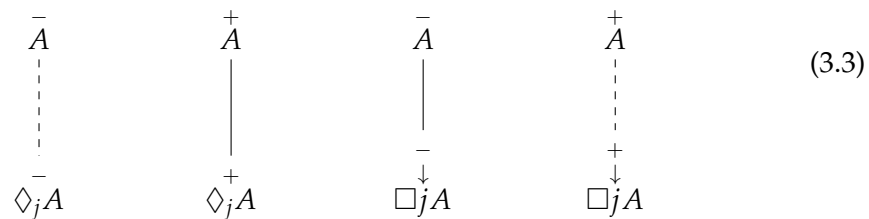


FIGUUR 3.3: Analyse van “Een dier dat een mens eet”, “een dier” is subject.

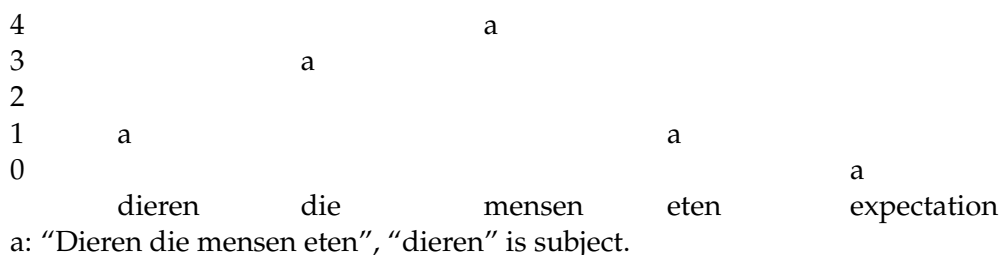


FIGUUR 3.4: Analyse van “Een dier dat een mens eet”, “een mens” is subject.

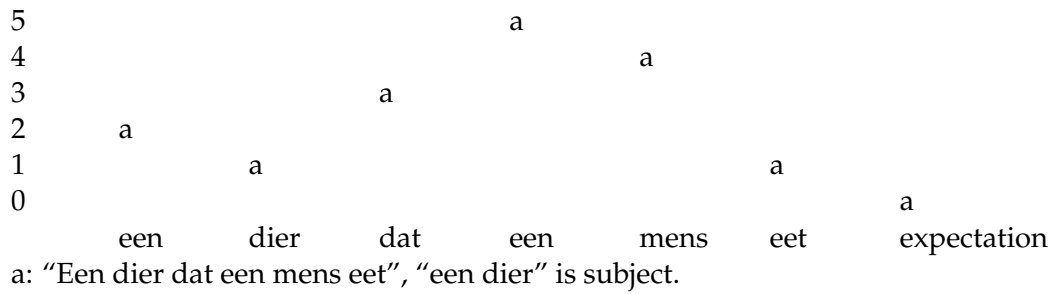
Het verwijderen van dependentierollen binnen deze bewijsnetten gebeurt volgens de regels van Moot (2002). Deze regels zijn weergegeven in (3.3). Net zoals bij Kogkalidis et al., (2020b) staat een negatieve polariteit hier voor een output label, een positieve polariteit voor een input label. De stippellijn staat hier verder voor een ii-link, een getrokken lijn geeft een i-link weer. Er geldt verder in bovenstaande bewijsnetten dat de verwachting NP aan de rechterzijde van het bewijsnet is weergegeven. De keuze hiervoor is gebaseerd op de analyses van Kogkalidis et al., (2020b), in hun bewijsnet staat de verwachting ook aan de rechterzijde. Morrill (2000) geeft deze verwachting aan de linkerzijde van het bewijsnet weer.



In alle vier de gevallen geldt dat voldaan is aan de correctheidscriteria van Kogkalidis et al., (2020b). De vier bewijsnetten bevatten een cykel, maar deze cycli bevatten zelf een i-link. Verder geldt voor de reeksen van formulebomen dat er precies één formuleboom is met een input label. Om te voorkomen dat een NP met de dependentierol als subject een verbinding maakt met een NP met een objectrol, wordt in de bovenstaande gevallen een extra voorwaarde opgenomen in de correctheidscriteria: als er een dependentierol wordt verwijderd door middel van een i-link, kan het resulterende axioma alleen combineren met een axioma dat is ontstaan door dezelfde dependentierol te verwijderen met een ii-link. In de bewijsnetten van Figuur 3.1 tot 3.4 geldt dat het verwijderen van de dependentierol behorende bij de NP in het type van “die” of “dat” mogelijk gemaakt wordt door een i-link. Deze combineert vervolgens met een NP waarbij dezelfde dependentierol is verwijderd door gebruik te maken van een ii-link. De vier gevormde analyses zijn dus bewijsnetten. Op basis van twee van deze bewijsnetten kan de complexiteitscurve gevormd worden, volgens de methode die Morrill (2000) in zijn artikel beschrijft. Dit zijn de lezingen van de twee zinsdelen waarbij “dieren” en “een dier” het subject zijn. Deze bewijsnetten bevatten immers geen kruisende axiomaverbindingen. De complexiteitscurves zijn weergegeven in Figuur 3.5 en 3.6.



FIGUUR 3.5: Complexiteitscurve “Dieren die mensen eten”, “dieren” is subject.



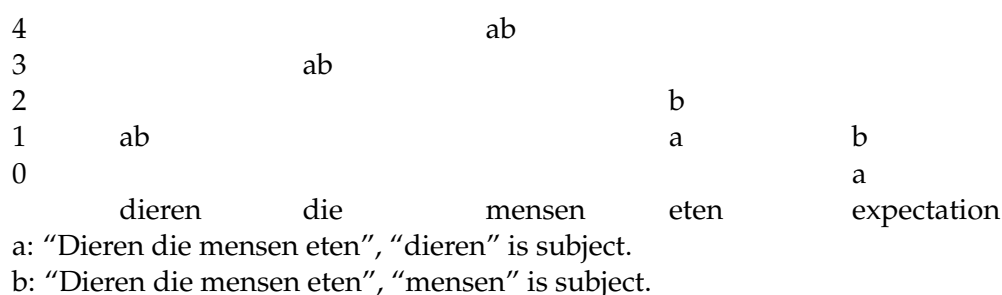
FIGUUR 3.6: Complexiteitscurve "Een dier dat een mens eet", "een dier" is subject.

De complexiteitsbepaling van Morrill (2000) kan niet toegepast worden bij lezingen waarbij het bewijsnet wel kruisende axiomaverbindingen bevat. Het niet hebben van kruisende axiomaverbindingen in het bewijsnet is namelijk een correctheidscriterium dat Morrill (2000) in zijn onderzoek hanteert. Verder geeft Morrill (2000) in zijn onderzoek niet aan hoe zijn complexiteitsbepaling uitgebreid kan worden naar bewijsnetten met kruisende axiomaverbindingen. Om de complexiteit van dit soort bewijsnetten vast te kunnen stellen, is het noodzakelijk om de complexiteitsbepaling van Morrill (2000) aan te passen. Volgens deze bepaling wordt in elke stap het aantal ongebonden axioma's geteld, wat resulteert in de complexiteitscurve. Bij de bewijsnetten met kruisende axiomaverbindingen wordt niet alleen het aantal ongebonden axioma's per stap bijgehouden, maar ook het aantal kruisingen van axiomaverbindingen. Per stap wordt nu de complexiteit berekend door het aantal kruisingen van axiomaverbindingen op te tellen bij het aantal ongebonden axioma's. Het resultaat is wederom een complexiteitscurve, met de complexiteit y in stap x voor punt (x,y) .

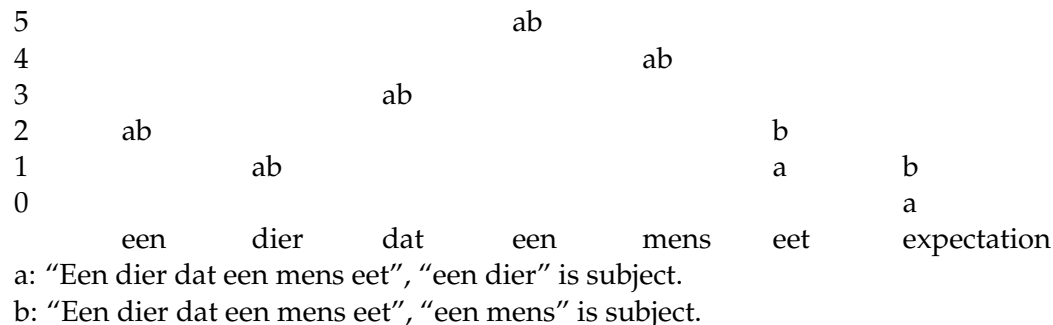
Hoofdstuk 4

Resultaten

De resultaten van dit onderzoek zijn verworven door de zojuist beschreven complexiteitsbepaling toe te passen op de vier bewijsnetten. De complexiteit van de lezingen met bewijsnetten zonder kruisingen blijft exact hetzelfde als eerder is weergegeven. In Figuur 4.1 en 4.2 is de complexiteitscurve van beide lezingen van beide zinsdelen weergegeven.



FIGUUR 4.1: Complexiteitscurve "Dieren die mensen eten" voor beide lezingen.



FIGUUR 4.2: Complexiteitscurve "Een dier dat een mens eet" voor beide lezingen.

Volgens het principle of acceptability (Morrill, 2000) wordt de totale complexiteit van een lezing berekend door de complexiteit van elke stap bij elkaar op te tellen. Voor lezing a in Figuur 4.1 is deze lager dan lezing b, aangezien $9 < 11$. Hetzelfde geldt voor lezing a en b in Figuur 4.2, de complexiteit van lezing a (16) is lager dan de complexiteit van lezing b (18). Uitgaande van het principle of acceptability, heeft de lezing met "dieren" of "een dier" als subject dus de voorkeur over de lezing met "mensen" of "een mens" als subject.

Hoofdstuk 5

Discussie en Conclusie

In deze studie staat centraal hoe de complexiteit van zinsdelen met bewijsnetten met kruisende axiomaverbindingen bepaald kan worden. Van de woorden die zich bevinden in de ambigue zinsdelen “Dieren die mensen eten” en “Een dier dat een mens eet” zijn de bijbehorende typen verkregen uit het ÆTHEL corpus (Kogkalidis et al., 2020a). Deze typen hebben vervolgens een richting toegekend gekregen door middel van trial-and-error, waarvoor de resulterende bewijsnetten een zo klein mogelijk aantal kruisende axiomaverbindingen bevatten. De bewijsnetten van de twee lezingen van de twee zinsdelen zijn incrementeel opgebouwd. Hierbij is de complexiteit bepaald. Voor de bewijsnetten zonder kruisende axiomaverbindingen is de complexiteitsbepaling van Morrill (2000) gebruikt, bij de bewijsnetten met deze kruisingen is voor elke stap het aantal kruisingen opgeteld. Uit de complexiteitscurves is vervolgens te concluderen dat de lezingen met “dieren” en “een dier” als subject de voorkeur hebben over de andere lezing, aangezien de complexiteit van deze eerste lezingen volgens het principle of acceptability (Morrill, 2000) lager is.

Om er zeker van te zijn dat men daadwerkelijk voorkeur geeft aan de lezingen met “dieren” of “een dier” als subject van het zinsdeel, is de ÆTHEL database gebruikt. Zoals eerder al genoemd in de methodesectie, bevat de python code van Kogkalidis (2021) de functie `deco_lexicon()`. Deze functie geeft voor een bepaald woord de typen uit ÆTHEL terug in volgorde van frequentie. Deze functie kan nu ingezet worden om te bepalen welke lezingen van de onderzochte zinsdelen daadwerkelijk de voorkeur heeft. Bij het aanroepen van `deco_lexicon('die')`, wordt gegeven dat het type $\langle\langle\text{VNW}\rangle \text{su} \rightarrow \text{SSUB}\rangle \text{rhd_body} \rightarrow [\text{NP} \rightarrow \text{NP}]$ mod 4613 keer voorkomt in het ÆTHEL corpus. Het type $\langle\langle\text{VNW}\rangle \text{obj1} \rightarrow \text{SSUB}\rangle \text{rhd_body} \rightarrow [\text{NP} \rightarrow \text{NP}]$ mod komt 469 keer voor in deze database. Aangezien ÆTHEL is gebaseerd op het corpus Lassy Klein (Van Noord et al., 2013), kan er aangenomen worden dat men het tweede type voor “die” minder vaak gebruikt dan het eerste type. Er wordt dus voorkeur gegeven aan het gebruik van “die” met het eerste type, waar het voornaamwoord (NP) gemarkeerd is als subject. In de bewijsnetten, weergegeven in de methodesectie, is te zien dat dit type hoort bij de lezing waar “dieren” het subject is van het zinsdeel. Uit de complexiteitscurve van Figuur 4.1 is te af te leiden dat de conclusie die getrokken wordt uit ÆTHEL overeenkomt met de conclusie die getrokken wordt uit de complexiteitsbepaling die is geïntroduceerd in dit onderzoek. Het aanroepen van `deco_lexicon('dat')`, levert op dat het type $\langle\langle\text{VNW}\rangle \text{su} \rightarrow \text{SSUB}\rangle \text{rhd_body} \rightarrow [\text{NP} \rightarrow \text{NP}]$ mod 982 keer voorkomt in het ÆTHEL corpus, terwijl het type $\langle\langle\text{VNW}\rangle \text{obj1} \rightarrow \text{SSUB}\rangle \text{rhd_body} \rightarrow [\text{NP} \rightarrow \text{NP}]$ mod maar 116 keer te vinden is. Ook hier geldt dat men het type van “dat” liever gebruikt als het voornaamwoord de dependentierol van het subject heeft. Dit type komt overeen met de lezing waar “een dier” het subject is van het zinsdeel. Wederom kan uit de complexiteitscurve van Figuur 4.2 en de ÆTHEL database hetzelfde worden geconcludeerd; men preferert de lezing met “een dier” als subject over de lezing met “een mens” als subject. Uit het ÆTHEL corpus volgt dat de geïntroduceerde complexiteitsbepaling het correcte resultaat levert voor de zinsdelen die in dit onderzoek zijn geanalyseerd.

Er zijn meerdere beperkingen te vinden in het uitgevoerde onderzoek. Allereerst wordt er in de analyse van de bestudeerde zinsdelen gebruik gemaakt van een aangenomen verwachting. In de bewijsnetten is deze verwachting aan de rechterzijde weergegeven. Aangenomen wordt dat er bij het luisteren naar een zin of zinsdeel een verwachting is van het type hiervan. De verwachting van de luisteraar is volgens deze aanname altijd juist. Immers, als dit niet het geval is, dan is het bewijsnet van deze zin of dit zinsdeel incorrect. Echter, deze aanname kan worden betwijfeld. Zoals Morrill (2000) al aangeeft zal men over het algemeen een zin verwachten bij het luisteren naar gesproken taal, hoewel dit niet altijd het geval hoeft te zijn. In dit onderzoek worden echter zinsdelen geanalyseerd met het resulterende type NP. De aanname dat de luisteraar altijd verwacht dat dit zinsdeel het type NP heeft, is niet erg realistisch. De aangenomen verwachting kan verder aan de linker- of rechterzijde van een bewijsnet worden weergegeven. In de context van dit onderzoek is dit beide mogelijk, de complexiteitsbepaling geeft nog steeds hetzelfde resultaat als ÆTHEL. Bepaalde gevallen die Morrill (2000) heeft onderzocht zijn hier echter wel gevoelig voor. Het plaatsen van de verwachting aan de rechterzijde van het bewijsnet zorgt in die gevallen voor een complexiteitscurve die de voorkeur geeft aan de verkeerde lezing. Vervolgonderzoek kan bestuderen hoe de complexiteit gelijk kan blijven, als de aangenomen verwachting aan de linker- of rechterzijde in het bewijsnet wordt geplaatst.

Een tweede limitatie betreft de aanname van een nieuw correctheidscriterium; het axioma dat ontstaat uit een unaire i-link mag alleen combineren met een axioma dat ontstaat door een ii-link uit een formule met dezelfde dependentierol. Deze aanname wordt gedaan om geen verbindingen te kunnen vormen tussen axioma's met een objectrol en een subjectrol. Echter, de aanname is niet universeel voor bewijsnetten. Er is geen correctheidscriterium te vinden voor de vorm van bewijsnetten die in dit onderzoek wordt gebruikt, die diamonds en boxen bevatten. De axiomaverbindingen tussen axioma's met verschillende dependentierollen leveren een onjuist bewijsnet op. Zonder het aangenomen correctheidscriterium zouden deze onjuiste bewijsnetten wel aan alle condities van een bewijsnet voldoen. In vervolgonderzoek kan bestudeerd worden of er een correctheidscriterium gevonden kan worden voor dit soort bewijsnetten met diamonds en boxen, waarbij correcte bewijsnetten wel worden toegestaan maar incorrecte bewijsnetten niet.

Daarnaast wordt in dit onderzoek geen rekening gehouden met semantiek. De bestudeerde zinsdelen bevatten twee lezingen en de geprefereerde lezing wordt alleen bepaald door de bijbehorende complexiteitscurve. Echter, Taraban en McClelland (1988) betogen dat semantische informatie het effect van lokale of globale ambiguïteit uitschakelt. Uit alleen de semantische informatie is volgens hen de lezing met voorkeur te bepalen. Bij de zin "Mensen die dieren eten", dat exact hetzelfde bewijsnet heeft als "Dieren die mensen eten", heeft de lezing met "mensen" als subject de voorkeur. Echter, als aan deze zin twee woorden worden toegevoegd, "Mensen die dieren eten zijn vegetarisch", heeft de lezing met "dieren" als subject de voorkeur. Dit komt omdat de andere lezing een tegenstrijdig resultaat oplevert: mensen eten dieren, maar zijn tegelijkertijd vegetarisch. De semantische informatie heeft dus invloed op welke lezing de voorkeur heeft. Deze invloed van semantiek kan in vervolgonderzoek gecombineerd worden met de geïntroduceerde complexiteitsbepaling.

Een laatste beperking die kan worden genoemd is dat de complexiteitsbepaling in dit onderzoek niet erg intuïtief is. Het verschil in complexiteit tussen de twee lezingen van beide zinsdelen is pas bekend na het horen van het woord "eten" of

“eet”. Dit houdt in dat, volgens de gegeven complexiteitscurves, de luisteraar hiervoor beide lezingen als even aannemelijk voorhoudt. Echter, het is onwaarschijnlijk dat de luisteraar na het horen van het laatste woord in het zinsdeel pas een keuze maakt tussen de twee lezingen. Er kan onderzocht worden wanneer deze keuze tussen lezingen precies ontstaat en hoe de complexiteitsbepaling hier vervolgens op kan worden aangepast.

De complexiteitsbepaling die is geïntroduceerd in deze studie komt overeen met de conclusie die getrokken wordt uit ÆTHEL. Het bijhouden van ongebonden axioma's en kruisingen van axiomaverbindingen tijdens het vormen van het bewijsnet levert een juiste complexiteitsbepaling op voor de bestudeerde zinsdelen. De lezingen met “dieren” en “een dier” als subject hebben de voorkeur over de lezingen met “mensen” of “een mens” als subject. Door middel van vervolgonderzoek kan de geïntroduceerde manier van complexiteit bepalen worden aangepast, waarmee de limitaties aan dit onderzoek worden weggenomen. Ondanks de genoemde beperkingen aan dit onderzoek mag niettemin voorzichtig geconcludeerd worden dat de geïntroduceerde complexiteitsbepaling ook juist is voor andere zinsdelen en zinnen. Op basis van deze complexiteitsbepaling kan in de toekomst de complexiteit bepaald worden van andere zinnen en zinsdelen, waarvan de bewijsnetten ook kruisende axiomaverbindingen bevatten.

Bibliografie

- Bever, T. (1970). The Cognitive Basis for Linguistics Structures. In J. Hayes (Red.), *Cognition and the Growth of Language* (pp. 279–352). New York, Amerika. Wiley.
- Gherke, M., Nagahashi, H. & Venema, Y. (1958). A Sahlqvist theorem for distributive modal logic. *Annals of Pure and Applied Logic*, 131(1-3), 65–102.
- Harley, T. (2017). *Talking the Talk* (2de ed.). Psychology Press.
- Kogkalidis, K. (2021). LSD 2021. Verkregen van <https://github.com/konstantinosKokos/presentations/blob/master/LSD2021/LSD%202021.ipynb>
- Kogkalidis, K., Moortgat, M. & Moot, R. (2020a). ÆTHEL: Automatically Extracted Typological Derivations for Dutch. *Proceedings of The 12th Language Resources and Evaluation Conference*, 5259–5268.
- Kogkalidis, K., Moortgat, M. & Moot, R. (2020b). Neural Proof Nets. *Proceedings of The 24th Conference on Computational Natural Language Learning*, 26–40.
- Lambek, J. (1958). The Mathematics of Sentence Structure. *The American Mathematical Monthly*, 65(3), 154–170.
- Moot, R. (2002). *Proof Nets for Linguistic Analysis* (Proefschrift, Universiteit Utrecht, Utrecht). Verkregen van https://www.narcis.nl/search/coll/publication/Language/NL/repositorygroup_id/uu
- Morrill, G. (2000). Incremental Processing and Acceptability. *Association for Computational Linguistics*, 26(3), 319–338.
- Taraban, R. & McClelland, J. (1988). Constituent attachment and thematic role assignment in sentence processing: Influences of content-based expectations. *Journal of Memory and Language*, 27, 597–632.
- Van Noord, G., Bouma, G., Van Eynde, F., de Kok, D., van der Linde, J., Schuurman, I., ... Vandeghinste, V. (2013). Large Scale Syntactic Annotation of Written Dutch: Lassy. In P. Spynes & J. Odijk (Red.), *Essential Speech and Language Technology for Dutch* (pp. 147–164). Den Haag, Nederland: Springer.
- Winter, Y. (2016). *Elements of Formal Semantics* (1ste ed.). Edinburgh University Press.