

DE UNIFICATIE VAN MENSELIJKE COGNITIE

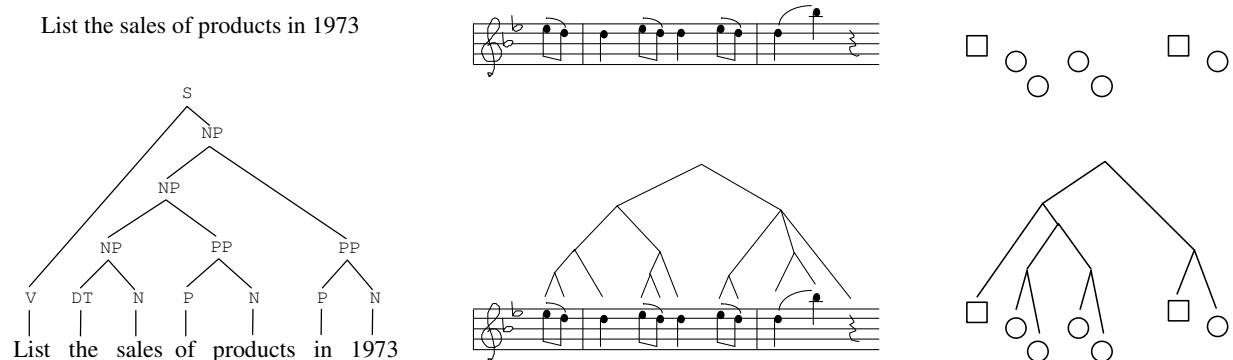
Rens Bod

Institute for Logic, Language and Computation
University of Amsterdam
Plantage Muidergracht 24
1018 TV Amsterdam
rens@science.uva.nl

Hoe leren mensen taal, en wat heeft dit gemeenschappelijk met het luisteren naar muziek, het kijken naar een afbeelding of het oplossen van een probleem? Deze vragen vormen de kern van het Vici-project *Integrating Cognition* dat poogt een algemeen computationeel model te ontwikkelen voor het leren van taal, muziek en redeneren. Zo'n algemeen model zou een klein deel van de zogeheten *Grand Challenge* (Newell 1990) kunnen oplossen: het ontwerpen van een unificerend model voor menselijke cognitie. In dit artikel zal ik ingaan op de doelstellingen en wijze van aanpak van dit project, en op enkele relevante toepassingen ervan.

1. Het probleem: hiërarchische representaties en ambiguïteit

Er zijn slechts weinig robuuste resultaten in de cognitiewetenschappen, maar één ervan is deze: mensen nemen sensorische stimuli niet waar als lineaire sequenties van symbolen maar als hiërarchische groepenstructuren die kunnen worden weergegeven in de vorm van 'boomstructuren'. Boomstructuren zijn gebruikt voor het beschrijven van linguïstische perceptie (Wundt 1901; Chomsky 1965), muzikale perceptie (Longuet-Higgins 1976; Lerdahl and Jackendoff 1983) en visuele perceptie (Marr 1982; Grenander 1996). Figuur 1 toont drie eenvoudige voorbeelden van (transcripties van) talige, muzikale en visuele stimuli met daaronder hun corresponderende boomstructuren.

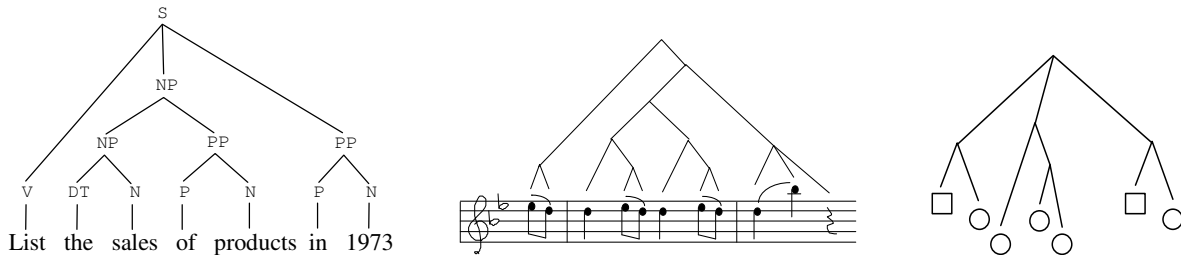


Figuur 1. Voorbeelden van talige, muzikale en visuele stimuli met hun boomstructuren

Een boomstructuur geeft aan hoe perceptuele informatie wordt gesegmenteerd in delen en hoe deze delen samenhangen in een groter geheel. De linguïstische boomstructuur is bovendien verrijkt met syntactische categorieën zoals V (*verb*), DT (*determiner*), N (*noun*), P (*preposition*), NP (*noun phrase*), PP (*prepositional phrase*) etc. Op deze manier kunnen syntactische beperkingen worden opgelegd op de manier waarop bepaalde woorden worden gecombineerd tot grotere eenheden ('constituenten'), terwijl er in muziek en visuele perceptie geen harde

beperkingen zijn: in principe kan elke noot of visueel element worden gecombineerd met elke andere noot of element.

Behalve dit verschil tussen perceptuele modaliteiten, is er ook een belangrijke overeenkomst: de perceptuele invoer ondergaat een proces van hiërarchische decompositie die niet intrinsiek aanwezig is in de invoer. Een belangrijk probleem in de cognitiewetenschappen is hoe we de door mensen waargenomen boomstructuur voor een bepaalde invoer kunnen voorspellen. Dat dit probleem niet triviaal is kan worden geïllustreerd door het verschijnsel dat aan de perceptuele stimuli in figuur 1 ook alternatieve boomstructuren kunnen worden toegekend, zoals in figuur 2.



Figuur 2. Alternatieve boomstructuren voor de stimuli in figuur 1

De linguïstische boomstructuur in figuur 2 heeft een andere syntactische structuur (en bijbehorende semantische interpretatie) dan die in figuur 1. De twee muzikale boomstructuren geven twee verschillende analyses weer bestaande uit verschillende motieven en muzikale frasen. En de twee visuele structuren vormen twee verschillende visuele *Gestalts*. Maar hoewel de alternatieve boomstructuren in principe kunnen worden waargenomen, komen ze niet overeen met de structuren die daadwerkelijk door mensen worden waargenomen. Het verschijnsel dat aan eenzelfde perceptuele invoer verschillende structuren kunnen worden toegekend, staat bekend als het *ambiguïteitsprobleem*. Dit is een van grootste problemen uit de cognitiewetenschappen. Zelfs voor taal, waar een formele grammatica specificeert welke woorden kunnen worden gecombineerd tot constituenten, is het ambiguïteitsprobleem enorm: Charniak (1997) toont aan dat voor gemiddelde zinnen uit de *Wall Street Journal* er meer dan een miljoen verschillende syntactische ontledingen en bijbehorende semantische interpretaties bestaan.

Het ambiguïteitsprobleem voor muzikale en visuele invoer is zo mogelijk nog groter aangezien er vrijwel geen beperkingen zijn hoe perceptuele elementen kunnen worden gecombineerd tot grotere eenheden. Longet-Higgins en Lee (1987) merken op dat "Any given sequence of note values is in principle infinitely ambiguous, but this ambiguity is seldom apparent to the listener".

2. Twee principes: waarschijnlijkheid en eenvoud

Hoe kunnen we uit alle mogelijke boomstructuren de boom selecteren die wordt waargenomen door mensen? Er zijn sinds jaar en dag twee concurrerende principes in omloop. De eerste, voorgesteld door Helmholtz (1910), behelst het waarschijnlijkheidsprincipe: sensorische informatie wordt georganiseerd in de vorm van de meest waarschijnlijke structuur die consistent is met de invoer. De tweede, voorgesteld door Wertheimer (1923) omvat het eenvoudsprincipe: het perceptuele systeem zoekt naar de eenvoudigste structuur die consistent is met de invoer (zie Chater 1999). Ik zal beide principes kort behandelen voor taal-, muziek- en beeldverwerking.

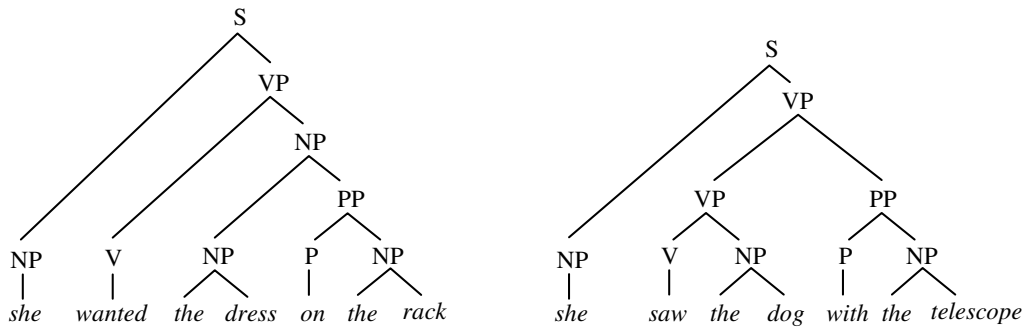
Het *waarschijnlijkheidsprincipe* is bijzonder invloedrijk in de natuurlijke taalverwerking en computerlinguïstiek (zie Manning en Schütze 1999). De waarschijnlijkheid van een boomstructuur wordt berekend aan de hand van de waarschijnlijkheden van haar delen, die op hun beurt worden geschat aan de hand van een grote verzameling taaldata, een *corpus*. De beste ‘automatische ontleders’ behalen op deze manier een nauwkeurigheid van rond de 91% correct voorspelde boomstructuren op een internationaal erkende *benchmark* (zie Bod 2003; Charniak en Johnson 2005). Het waarschijnlijkheidsprincipe is tevens gebruikt in muziekperceptie (zie Bod 2002a,b; Temperley 2007). Een bekende benchmark op dit gebied is de *Essen Folksong Collection*, waarmee rond de 87% nauwkeurigheid wordt bereikt. Ook op het gebied van visuele perceptie is er het laatste decennium een opleving van waarschijnlijkheidsmodellen. Mumford (1999) spreekt zelfs van de *Dawning of Stochasticity*.

Het *eenvoudsprincipe* kent een lange traditie in de perceptiepsychologie, vooral in de Structurele Informatie Theorie (Leeuwenberg 1971; Simon 1972). De eenvoudigste structuur van een visueel patroon wordt berekend aan de hand van zijn kortste codering volgens een codeertaal die lijnsegmenten en hoeken als basiselementen gebruikt. Deze notie van eenvoud is ook met succes toegepast in het modelleren van muziekperceptie (Collard et al. 1981). Bekender is echter de theorie van Lerdahl en Jackendoff (1983), die gebaseerd is op voorkeursregels die Gestaltperceptie beschrijven zoals voorgesteld door Wertheimer (1923). Noties van eenvoud bestaan ook in natuurlijke taalverwerking in de vorm van de kortste afleiding (Frazier 1978). Hoewel zo’n kortste afleiding (bestaande uit de kortste combinatie van constructies uit een groot taalcorpus) een minder hoge nauwkeurigheid behaalt op *benchmarks* dan de meest waarschijnlijke boomstructuur, lijken de resultaten complementair (Bod 2002a).

3. Een geïntegreerde benadering

Een van de basisideeën van dit project is dat beide principes een rol spelen in perceptuele organisatie, maar op een verschillende manier: het eenvoudsprincipe als een algemene voorkeur voor ‘economie’, en het waarschijnlijkheidsprincipe als een frequentie-gebaseerde tendens ten gevolge van eerdere ervaringen. Onze werkhypothese is dat het menselijke perceptuele systeem streeft naar de eenvoudigste structuur, maar dat dit streven wordt beïnvloed door de frequentie van voorkomen van eerdere waarnemingen. Om deze werkhypothese te instantiëren dienen we eerst de verzameling *mogelijke* structuren te definiëren. In dit project zullen we dit doen aan de hand van een model dat bekend staat als DOP (Data-Oriented Parsing) en waarvan bekend is dat het generaliseert over vrijwel alle andere bestaande computationele taalmodellen die gebruik maken van boomstructuren (zie Carroll en Weir 2000). Het basisidee van het DOP model is dat nieuwe invoer wordt geanalyseerd middels het combineren van fragmenten van eerder verwerkte invoer opgeslagen in een *corpus*. Fragmenten kunnen van willekeurige grootte zijn: zowel de kleinst mogelijke deelbomen als volledige boomstructuren worden in acht genomen. Door restricties aan te brengen op de grootte van de fragmenten kan een zeer groot aantal grammaticale formalismen worden gesimuleerd (zie Bod 1998; Bod, Hay and Jannedy, 2003). Het voordeel van het werken met fragmenten van willekeurige grootte is tevens dat hiermee vaste linguïstische constructies - of muzikale riedels - kunnen worden bijgehouden.

We zullen het DOP model illustreren aan de hand van een eenvoudig taalkundig voorbeeld. Stel dat we een heel klein corpus hebben van slechts twee zinnen met hun boomstructuren, zoals weergegeven in figuur 3 (realistische corpora bestaan al gauw uit meerdere miljoenen zinnen).



Figuur 3. Een zeer klein corpus bestaande uit slechts twee boomstructuren

Hoe kunnen we nu aan de hand van dit corpus een nieuwe zin - die niet in het corpus voorkomt - analyseren, zoals *She saw the dress with the telescope*? Welnu, de betreffende zin kan worden geanalyseerd door (ondermeer) het combineren van twee deelbomen uit bovenstaand corpus gebruikmakende van een substitutie-operatie die we aanduiden met ‘ \circ ’, zoals weergegeven in figuur 4.

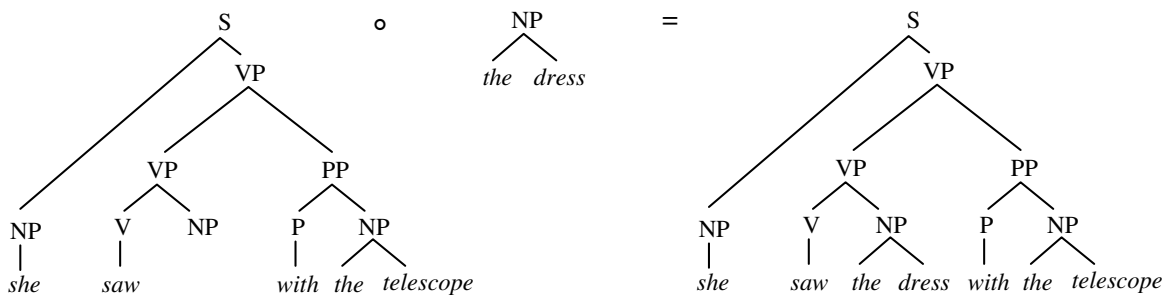


Figure 4. Analyse van een nieuwe zin middels het combineren van deelbomen uit figuur 3

We kunnen aan de hand van het DOP model ook een alternatieve boomstructuur produceren voor de betreffende invoerzin, namelijk door het combineren van drie deelbomen uit het corpus in plaats van twee, zoals blijkt uit figuur 5.

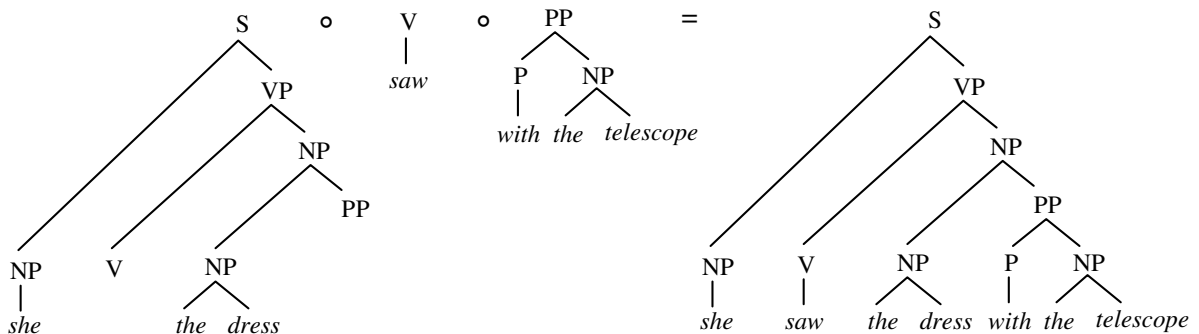


Figure 5. Een alternatieve analyse voor de zin *She saw the dress with the telescope*

Welke van de twee resulterende boomstructuren is nu ‘correct’? Volgens het eenvoudsprincipe zou de boomstructuur in figuur 4 de geprefereerde zijn, aangezien deze kan worden opgebouwd uit het kleinste aantal fragmenten (twee deelbomen). Echter, het is eenvoudig aan te tonen dat volgens het waarschijnlijkheidsprincipe de boomstructuur in figuur 5 de geprefereerde is -- als we alle frequenties van deelbomen meenemen (in dit sterk-geïdealiseerde corpus). De twee principes zijn dus met elkaar in contrast, en een belangrijk deel van dit project gaat over het vaststellen van een optimale combinatie van deze principes. We zullen ons hierbij aanvankelijk laten leiden door een hypothese die veel aanhang heeft verworven in de cognitiewetenschappen (zie bv. Newell 1990; Tomasello 2003): het menselijke perceptuele systeem streeft naar de meest economische structuur (de kortste, en daarmee eenvoudigste afleiding), maar in dit streven hebben mensen alleen toegang tot de top van de distributie van meest waarschijnlijke structuren. Deze hypothese kan in onze benadering worden geformaliseerd door de eenvoudigste structuur te selecteren uit de n meest waarschijnlijke boomstructuren, waarbij n een vrije parameter is die empirisch dient te worden vastgesteld. We willen ons echter niet vastpinnen op deze ene werkhypothese en zullen ondermeer ook de ‘omgekeerde’ hypothese testen, namelijk het selecteren van de meest waarschijnlijke boomstructuur uit de n kortste afleidingen. Bovendien willen we deze optimale mix van eenvoud en waarschijnlijkheid niet alleen voor taal proberen te bepalen maar ook voor andere vormen van perceptie. Het hier geschetste DOP model is namelijk domeinonafhankelijk: het kan even goed worden gebruikt voor het analyseren van muziek, beeld of spraak - zolang we maar een corpus met boomstructuren hebben voor het betreffende domein (zie Bod 2002a,b).

Ook zal DOP worden ingezet voor het modelleren van ‘hogere’ vormen van cognitie, zoals redeneren en probleem-oplossen. Wat voor talige en muzikale analyse geldt, geldt *grosso modo* ook voor redeneren en probleem-oplossen: een stelling of probleem kan vele afleidingen of oplossingen hebben. In dit Vici-project zullen we ons concentreren op afleidingen van problemen in de wiskunde en (klassieke) natuurkunde. Zulke afleidingen kunnen veelal worden weergegeven in de vorm van een *bewijsboom*, hetgeen weer een boomstructuur is (met iets andere eigenschappen dan de hierboven weergegeven boomstructuren). In Bod (2007) is een DOP model ontwikkeld dat nauwkeurig het proces van probleem-oplossen door natuurkundestudenten kan simuleren. Dat DOP model is voornamelijk gebaseerd op het eenvoudsprincipe: studenten blijken een nieuw probleem op te lossen door het zo veel mogelijk te matchen op eerdere opgeloste problemen zoals behandeld in leerboeken. Het is verrassend te zien hoe goed de kortste derivatie (die bestaat uit de grootste deel-derivaties van voorbeeldproblemen) overeenkomt met de oplossingen gegenereerd door derdejaars studenten. Natuurlijk speelt frequentie van voorkomen ook een rol in probleem-oplossen, maar veel minder dan bij het structureren van taal en muziek.

4. Het leren van boomstructuren

Hoewel het DOP model een zekere mate van succes heeft geboekt in het simuleren van talige, muzikale, visuele en andere vormen van cognitie, waaronder redeneren en probleem-oplossen, zegt het model niets over hoe de eerste boomstructuren worden geleerd. Waar komen boomstructuren vandaan als er nog geen structuren in het corpus zijn? In Bod (2006) wordt een uitbreiding voorgesteld van (de talige variant van) het DOP model, die nu bekend staat onder de naam *Unsupervised DOP* ofwel *U-DOP*. Dit model gaat uit van het volgende: als we niet weten welke boomstructuren moeten worden toegekend aan eerste taaluitingen, dienen we aanvankelijk

alle boomstructuren toe te laten, om vervolgens aan de hand van het DOP model te bepalen welke boomstructuren het meest geschikt zijn om deze en nieuwe uitingen te analyseren. Een voorbeeld kan dit verduidelijken. Stel dat een taalverwerper de volgende zinnnetjes hoort: *watch the dog* en *the dog barks*. Hoe zouden we dan de structuren kunnen vinden die bij deze zinnnetjes horen?

Volgens het U-DOP model zou aanvankelijk elk mogelijk fragment een constituent kunnen vormen waarna aan de hand van het DOP model de beste structuur wordt bepaald, middels een nog vast te stellen combinatie van het eenvouds- en waarschijnlijkheidsprincipe. Aangezien er nog geen syntactische categorieën zijn geleerd, nemen we aan dat de (talige) boomstructuren ongelabeld zijn. Dit betekent dat elke categorie feitelijk identiek is aan elke andere categorie, hetgeen we zullen weergeven met een *X*. De mogelijke boomstructuren voor de zinnen *watch the dog* en *the dog barks* kunnen dan worden weergegeven zoals in figuur 6 (we beperken ons in dit voorbeeld tot zogenaamde binaire boomstructuren)

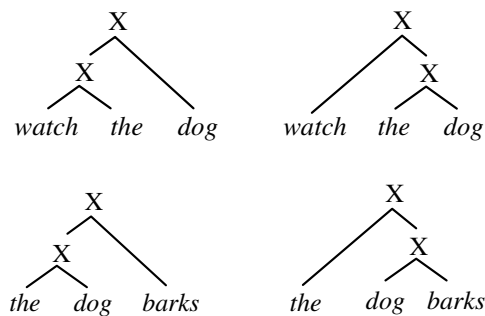
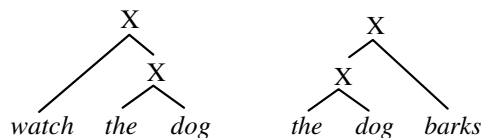


Figure 6. De verzameling binaire boomstructuren voor *watch the dog* en *the dog barks*

Hoewel het aantal mogelijke binaire boomstructuren exponentieel groeit met de zinslengte, kunnen deze efficiënt worden berekend en opgeslagen met behulp van een zogeheten *tree-forest* - maar dit terzijde.

Uitgaande van de boomstructuren in figuur 6 kunnen we aan de hand van het oorspronkelijke DOP model zowel de meest waarschijnlijke boomstructuren als de kortste derivaties uitrekenen voor deze zinnen. Zo valt misschien al op dat er precies één deelboom in figuur 6 is die twee keer voorkomt, terwijl alle andere deelbomen slechts één keer voorkomen. Dit is de deelboom bestaande uit het deelzinnetje *the dog* met de bovenliggende *X*. Als we uit zouden gaan van het waarschijnlijkheidsprincipe zou dit betekenen dat *watch the dog* en *the dog barks* de structuren in figuur 7 als geprefereerde boomstructuren krijgen omdat deze zijn opgebouwd uit (ondermeer) de meest frequente deelboom (voor *the dog*) terwijl de frequenties van de andere deelbomen constant is.



Figuur 7. De geleerde boomstructuren volgens het waarschijnlijkheidsprincipe

Het eenvoudsprincipe daarentegen zou in dit voorbeeld geen onderscheid maken tussen de verschillende structuren voor *watch the dog* en *the dog barks* in figuur 6, aangezien ze allemaal zijn te produceren door een kortste derivatie: namelijk de boomstructuur zelf. Het lijkt dus dat bij het *leren* van boomstructuren de notie van waarschijnlijkheid belangrijker is dan de notie van eenvoud. Dit komt overeen met een recente, invloedrijke stroming in de taalkunde: de zogeheten *usage-based* of *constructionistische* theorie van taalverwerving (Goldberg 2006; Bybee 2006). Volgens deze stroming geldt, net als in U-DOP, dat frequentie de belangrijkste factor is in het leren van taal: hoe vaker twee of meer woorden naast elkaar voorkomen, hoe sterker ze een groep of constituent vormen. Dit uitgangspunt staat in contrast met de zogeheten *nativistische* theorie van taalverwerving volgens welke er een aangeboren grammatica nodig is om de structuur van taal te leren (Chomsky 1965; Crain 1991).

Tot nu toe hebben we ons beperkt tot het leren van de syntactische structuur van taal. Het voert te ver om in dit verband uit te leggen hoe U-DOP kan worden uitgebreid tot het leren van (logische) semantiek, ofwel de betekenis van zinnen (zie Bod 2006). Bovendien hebben we ons hier beperkt tot een talig voorbeeld. Een van de hoofddoelen van dit Vici-project is juist om U-DOP te generaliseren tot een leermodel voor andere vormen van cognitie: van muziek tot redeneren.

5. Relevantie en toepassingen

De belangrijkste innovatie van dit Vici-project is het ontwerpen van een algemeen, ongesuperviseerd leermodel voor verschillende cognitieve domeinen. Zo zijn er deelprojecten op het gebied van (1) taal, (2), muziek, (3) visuele analyse, (4) redeneren en (5) probleem-oplossen, waarvoor drie postdocs en twee aio's worden aangesteld. Echter, het ontwerp van een algemeen leermodel voor cognitie is niet het einddoel. We willen het te ontwerpen leermodel tevens uittesten in concrete toepassingen. Het vinden van de juiste boomstructuur is namelijk niet slechts een academische exercitie, maar is van groot nut in toepassingen zoals automatisch vertalen, automatische spraakherkenning, automatische muzikale begeleiding, en automatische stellingsbewijzers. Het U-DOP model zal als deelmodule in zulke toepassingen worden geëvalueerd waarbij het dan niet gaat om hoe nauwkeurig we een bepaalde boomstructuur kunnen voorspellen maar om de mate waarin zo'n boomstructuur kan bijdragen tot het verbeteren van de nauwkeurigheid van de betreffende toepassing. Op deze manier heeft het project zowel een wetenschappelijk doel (het oplossen van een deel van de *Grand Challenge*: de unificatie van cognitie), als een meer toegepast doel (het verbeteren van bestaande applicaties).

Literatuur

- Bod, R. (1998) *Beyond Grammar: An Experience-Based Theory of Language*. Stanford: CSLI Publications, Cambridge: Cambridge University Press.
- Bod, R. (2002a) A unified model of structural organization in language and music, *Journal of Artificial Intelligence Research* **17**, pp. 289-308.
- Bod, R. (2002b) Memory-based models of melodic analysis: challenging the Gestalt principles, *Journal of New Music Research* **31**, pp. 27-36.
- Bod, R., J. Hay, en S. Jannedy (red.) (2003) *Probabilistic Linguistics*. Cambridge, MA: MIT Press.
- Bod, R. (2003) Do all fragments count?, *Journal of Natural Language Engineering* **9**, pp. 307-323.
- Bod, R. (2006) Exemplar-based syntax: how to get productivity from examples, *The Linguistic Review* **23**, pp. 289-318.

- Bod, R. (2007) Getting rid of derivational redundancy or how to solve Kuhn's problem, *Minds and Machines* **17**, pp. 47-66.
- Bybee, J. (2006) From usage to grammar: the mind's response to repetition, *Language* **82**, pp. 711-733.
- Carroll, J., en D. Weir (2000). Encoding frequency information in stochastic parsing models, in: H. Bunt en A. Nijholt (red.). *Advances in Probabilistic and Other Parsing Technologies*. Boston: Kluwer, pp. 13-18.
- Charniak, E. (1997) Statistical techniques for natural language parsing, *AI Magazine*, Winter 1997, pp. 32-43.
- Charniak, E., en M. Johnson (2005) Coarse-to-fine n-best parsing and MaxEnt discriminative reranking, in: *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*. Morristown: Association for Computational Linguistics, pp. 173-180.
- Chater, N. (1999) The search for simplicity: a fundamental cognitive principle?, *The Quarterly Journal of Experimental Psychology*, **52A**, pp. 273-302.
- Chomsky, N. (1965) *Aspects of the Theory of Syntax*. Cambridge MA: MIT Press.
- Collard, R., P. Vos, en E. Leeuwenberg (1981) What melody tells about metre in music, *Zeitschrift für Psychologie* **189**, pp. 25-33.
- Crain, S. (1991) Language acquisition in the absence of experience, *Behavioral and Brain Sciences* **14**, pp. 597-612.
- Frazier, L. (1978) *On Comprehending Sentences: Syntactic Parsing Strategies*. PhD Thesis, University of Connecticut.
- Goldberg, A. (2006) *Constructions at Work: the nature of generalization in language*. Oxford: Oxford University Press.
- Grenander, U. (1996) *Elements of Pattern Theory*. Baltimore: Johns Hopkins University Press.
- Helmholtz von, H. (1910) *Treatise on Physiological Optics* (Vol. 3). Dover NY: New York.
- Leeuwenberg, E. (1971) A perceptual coding language for perceptual and auditory patterns, *American Journal of Psychology* **84**, pp. 307-349.
- Lerdahl, F. en R. Jackendoff (1983) *A Generative Theory of Tonal Music*. Cambridge MA: MIT Press.
- Longuet-Higgins, H. (1976) Perception of melodies, *Nature* **263**, pp. 646-653.
- Longuet-Higgins, H., en C. Lee (1987) The rhythmic interpretation of monophonic music, in: *Mental Processes: Studies in Cognitive Science*. Cambridge MA: MIT Press.
- Manning, C., en H. Schütze (1999) *Foundations of Statistical Natural Language Processing*. Cambridge MA: MIT Press.
- Marr, D. (1982) *Vision*. San Francisco: Freeman.
- Mumford, D. (1999) *The dawning of the age of stochasticity*. Invited Lecture at the Accademia Nazionale dei Lincei.
- Newell, A. (1990) *Unified Theories of Cognition*. Harvard: Harvard University Press.
- Simon, H. (1972) Complexity and the representation of patterned sequences of symbols, *Psychological Review* **79**, pp. 369-382.
- Temperley, D. (2007) *Music and Probability*. Cambridge MA: MIT Press.
- Tomasello, M. (2003) *Constructing a Language*. Harvard: Harvard University Press.
- Wertheimer, M. (1923) Untersuchungen zur Lehre von der Gestalt, *Psychologische Forschung* **4**, pp. 301-350.
- Wundt, W. (1901) *Sprachgeschichte und Sprachpsychologie*. Leipzig: Engelmann.