

Klare taal: wat kunnen computermodellen ons leren over taalevolutie?

Tony Belpaeme en Joris Van Looveren

Artificial Intelligence Lab, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussel,

tony@arti.vub.ac.be

joris@arti.vub.ac.be

Abstract

Natuurlijke taal is een bijzonder complex fenomeen en verschillende aspecten zijn nauwelijks of helemaal niet bestudeerd in het traditionele taalonderzoek. In deze bijdrage gaan we in op het dynamische gedrag van taal in een groep taalgebruikers. Een uitgelezen manier om dynamisch gedrag te bestuderen is door gebruik te maken van computersimulaties. In deze simulaties wordt een individu gemodelleerd als een eenvoudig computerprogramma; meerdere van deze computerprogramma's vormen dan een populatie van individuen. Computers laten het probleemloos simuleren van grote groepen van eenvoudige individuen toe, zodat het mogelijk wordt om interacties op grote schaal en over een grote tijdsperiode te volgen.

In deze bijdrage beschrijven we de componenten van dergelijke computermodellen en als concreet voorbeeld beschouwen we enkele stapsgewijs complexere modellen die het aanleren van een lexicon bestuderen, en een model dat nagaat in hoeverre menselijke kleurcategorieën kunnen beïnvloed worden door taal.

Sleutelwoorden: taalevolutie, computersimulatie, dynamische processen, zelforganisatie, kleurcategorieën.

1. Inleiding

Stel je eens een bioloog voor die de evolutie van de mens onderzoekt louter op basis van de nu levende mensen. Stel je voor dat hij geen fossielen ter beschikking heeft die kunnen dienen om na te gaan of bepaalde veronderstellingen of voorspellingen kloppen. Er zijn ook geen levende organismen, zoals mensapen, die een gelijkaardige evolutie hebben doorlopen. En stel je bovendien voor dat mensen volwassen geboren worden, zodat onze bioloog enkel volwassen, levende mensen heeft om de sleutel tot de menselijke evolutie te vinden. Zonder twijfel is dit een situatie die door echte biologen als weinig benijdenswaardig beschouwd zou worden.

Nochtans is dat precies de situatie waarin de onderzoeker geïnteresseerd in de evolutie van taal zich bevindt: er zijn namelijk hoegenaamd geen overblijfselen van taal die dateren van vóór het schrift—en toen het schrift er was, was taal al lang “volwassen”. Van taal vóór het schrift zijn geen tastbare bewijzen overgebleven. Ook zijn er geen levende talen die lager op de evolutionaire ladder staan dan de bestaande talen, alle huidige talen hebben namelijk hetzelfde niveau van complexiteit.

Er is dus een gebrek aan direct bewijsmateriaal. Maar, net zoals bij forensisch onderzoek een *reconstructie* van de misdaad uitsluitend kan geven over wat er gebeurde, proberen ook taalonderzoekers te reconstrueren wat er gebeurd kan zijn in de periode tussen de oorsprong van taal en de uitvinding van het schrift. Deze reconstructie kan zich op allerlei vlakken situeren.

Op het meest abstracte vlak kan men taalevolutie reconstrueren door het opstellen van een opeenvolging van de verschillende stappen die taal doorlopen moet hebben tijdens zijn ontstaan en evolutie. Een aanname “bij verstek” die lang heeft standgehouden binnen de linguïstiek, is dat taal het resultaat was van één enkele genetische mutatie. Deze hypothese werd eigenlijk geboren uit desinteresse voor het probleem: de gedurende lange tijd dominante Chomskiaanse linguïstiek was vooral geïnteresseerd in de formele beschrijving van de syntax en veel minder in het hoe en waarom van het ontstaan ervan. Gevraagd naar het ontstaan was het antwoord eenvoudig en weinig doordacht: een genetische mutatie (Chomsky, 1980).

Pas met het onderzoek naar pidgin- en creooltalen en de prototaal-hypothese van Derek Bickerton (1981) is er enige interesse ontstaan voor de mogelijkheid dat er vroeger eenvoudigere vormen van taal bestonden, en de mogelijkheid dat deze primitievere vormen alsnog de kop opsteken als de “volledige” taal het laat afweten. In pidginsituaties is geen enkele moedertaal dominant en de sprekers moeten terugvallen op een vereenvoudigde omgangstaal. Daarop volgen al snel een creolisatiefase, waarin binnen één generatie een pidgin evolueert tot een volwaardige taal.

Recent is Bickertons tweefasenplan, verder uitgewerkt door Ray Jackendoff (2002), die een volledig schema geeft dat alle cruciale transities in het ontstaan van taal bevat. Deze vorm van reconstructie blijft vrij abstract, maar levert een interessant discours op dat kan gebruikt worden als leidraad bij het bekijken van taal en de complexiteiten die het bevat.

Een concretere vorm van reconstructie is reconstrueren *van de taal zelf*. De kiem van deze aanpak ligt bij de neogrammatici die probeerden om de onderlinge verschillen binnen en tussen verschillende taalgroepen (vooral de Romaanse en de Germaanse) te verklaren. Dit empirische onderzoek leverde een aantal wetten op, die aantonen dat grote veranderingen in taal met een opmerkelijke regelmaat gebeuren.

De technieken die ontwikkeld werden door de neogrammatici in de negentiende eeuw, worden door moderne taalkundigen als Merritt Ruhlen (1996) gerekruteerd om op

grotere schaal (zowel qua aantal talen als qua historische schaal) talen met elkaar te vergelijken en ontbrekende tussenschakels te reconstrueren. Door steeds verder te gaan, proberen deze onderzoekers aan te tonen dat er één oorspronkelijke taal, de prototaal, bestond waaruit alle andere zijn voortgekomen.

Een impliciete aanname die in dit onderzoek gemaakt wordt, is dat de taalcapaciteit van de sprekers altijd dezelfde is. Deze methode is dus per definitie niet bruikbaar om de evolutie van taal te onderzoeken, omdat we net willen onderzoeken waar deze taalcapaciteit vandaan komt. In ieder geval geldt dat hoe langer je met deze methode teruggaat in de tijd, des te onzeker de resultaten worden.

Nog een vorm van reconstructie werkt op het fysiologische vlak: het doel is om, aan de hand van de ontwikkeling van de biologische kenmerken van het menselijke lichaam, te achterhalen op welk moment taal fysiek mogelijk werd. Men merkt op dat bij mensen het strottenhoofd hoger zit dan bij aan mensen gerelateerde dieren, en neemt aan dat dit onderscheidende kenmerk nodig is voor taal. Als men dus kan achterhalen wanneer in de fysiologische evolutie van de mens het strottenhoofd gestegen is, kan men extrapoleren wanneer taal ontstond. Deze invalshoek is uitgewerkt door Philip Lieberman (1998).

Het grote probleem waar deze benadering mee te kampen heeft, is dat de moderne fysiologische kenmerken niet noodzakelijk zijn om een bruikbaar communicatiesysteem op te leveren (zoals bij de vorige methode, die impliciet uitgaat van de noodzaak van een volgroeid taalvermogen). Het is zelfs perfect mogelijk om met slechts enkele klinkers en medeklinkers een moderne taal te vormen: de taal met de reputatie het minste fonemen te hebben (met name 10) is Piraha (Everett 1986). Bovendien is er geen bewijs dat de ontwikkeling van talige vermogens en spraak gelijk lopen; andere modaliteiten voor expressie zoals gebarentaal zijn mogelijk. Bij nader inzien lijkt het er dus op dat ook deze methode geen definitief uitsluitsel geeft over wanneer taal ontstaan is. Bovendien, zelfs al zouden we weten wanneer taal ontstaan is, dan weten we nog steeds niet hoe het gebeurd is.

Als besluit van dit (onvolledige) overzicht kunnen we stellen dat elke methode een aantal premissen moet aannemen die de besluiten potentieel op losse schroeven kunnen zetten. Geen enkele van deze methoden biedt bijgevolg op zichzelf uitsluitsel over wat wanneer en hoe gebeurd is. Het is dus belangrijk om zo veel mogelijk verschillende informatiebronnen te hebben, en de gegevens te integreren om de waarheid zo dicht mogelijk te benaderen¹.

2. Computermodellen

De meest recente aanpak in het onderzoek naar taalevolutie maakt gebruik van computers om virtuele reconstructies te maken. Hierbij worden individuen en hun sociolinguïstische interacties gesimuleerd in een computerprogramma. Een belangrijke overweging die in deze context gemaakt moet worden, is dat een computersimulatie nooit op zichzelf bestaat. Ze vormt altijd de belichaming van een achterliggende theorie. Deze theorie bepaalt de vorm die de individuen en hun interacties krijgen in het computerprogramma, bijvoorbeeld hoe het lexicon van een individu eruit ziet en hoe een woord of betekenis erin opgezocht wordt. In deze zin staan computersimulaties orthogonaal op andere onderzoeksmethoden: ze laten toe om eender welke theorie concreet te maken en uit te testen, en kunnen dan ook in veel contexten gebruikt worden. Dit impliceert ook dat een werkend computermodel op zich niet

¹ Zie (Christiansen & Kirby, 2003) voor een uitgebreider overzicht van verschillende visies op en onderzoeksmethoden naar het ontstaan en de evolutie van taal.

zaligmakend is; als een computermodel gebaseerd is op een wankel theorie, blijft de theorie wankel, ook al werkt het computermodel. Anderzijds kan een computermodel wel aantonen dat een theorie hiaten bevat. Als voorbeeld kunnen we het werk noemen van Sakas & Fodor (2001), waarbij een computermodel gemaakt wordt, gebaseerd op de *principles and parameters*-theorie (Chomsky, 1995). Volgens deze theorie wordt een kind geboren met een taalmodule en moet het enkel een paar schakelaars juist zetten naar gelang de taal waar het tijdens zijn eerste levensjaren mee in contact komt. De aanname is dat de oorspronkelijke complexiteit (één van de 2^n mogelijke grammatica's leren die gedefinieerd worden door n binaire parameters) gereduceerd wordt tot het leren van n parameterwaarden. Na hun experimenten komen Sakas & Fodor echter tot de conclusie dat het probleem niet eenvoudiger wordt: het proces om n binaire parameters juist te zetten komt neer op een zoekproces waarvan de complexiteit 2^n benadert. Dit illustreert hoe een computermodel nooit mag gezien worden als een doel op zich, maar als een verificatie van een theoretisch standpunt.

Veel theorieën over de evolutie van taal zijn in hetzelfde bedje ziek: ze zijn moeilijk of helemaal niet *testbaar* en daardoor niet falsifieerbaar. Dit betekent dat elke theorie voor een groot deel speculatie blijft en louter een beroep doet op het inzicht van de onderzoeker voor validatie. De vooruitgang van wetenschappelijk onderzoek zou echter niet enkel mogen worden bepaald door een dialectiek waarin alleen op basis van retoriek argumenten goed of slecht bevonden worden; hoe meer *voorspellend* een theorie is, en dus hoe meer verifieerbare feiten een theorie genereert, des te meer vertrouwen we kunnen hebben in een theorie. Zoals uit de bovenstaande voorbeelden al blijkt, kunnen computermodellen aan allerlei theorieën een testbaar aspect toevoegen: computermodellen zijn een implementatie van theorie en leveren resultaten op die geverifieerd en gevalideerd kunnen worden.

3. Onze theorie

Een belangrijk aspect van taal dat door de dominantie van op Chomsky's theorieën gebaseerde paradigma's lange tijd uit het oog verloren is, is het sociolinguïstische aspect van taal. Taal bestaat niet in isolatie; het is bij uitstek een fenomeen dat zich manifesteert in een *groep* mensen. Het feit dat taal een groepsfenomeen is, betekent ook dat het onderworpen is aan de dynamiek die bestaat tussen de verschillende leden van de groep, en ook beïnvloed wordt door deze dynamiek.

3.1. Zelforganisatie

Het idee van zelforganisatie kan het best verduidelijkt worden met het klassieke voorbeeld van zelforganiserend mierengedrag (Camazine et al., 2001). Een groep mieren is in staat om het transporteren van voedsel efficiënt te coördineren, ondanks het feit dat een mier toch een erg eenvoudig insect is en dat er nergens een mier is die een coördinerende functie heeft. Hoe doen ze dat? Bij nader inzicht blijkt dat mieren slechts enkele eenvoudige gedragingen nodig hebben, om dit complexe, globale gedrag te veroorzaken. (1) Als een mier rondzwerft op zoek naar voedsel, laat ze een chemisch spoor van feromonen achter, een beetje zoals Klein Duimpje broodkruimels achterliet om zijn weg terug te vinden. (2) Als een mier rondzwerft en ze komt feromonen tegen, wordt ze daardoor aangetrokken, en volgt ze met een zekere waarschijnlijkheid het feromonepad. (3) Hoe meer feromonen een pad bevat, des te sterker wordt de mier erdoor aangetrokken. Deze eenvoudige gedragingen zorgen ervoor dat paden naar voedselvoorraden na verloop van tijd steeds meer feromonen bevatten, zodat ze door meer mieren gevolgd worden. Deze mieren laten op hun beurt weer

feromonen achter, zodat het pad nog meer versterkt wordt. Opmerkelijk is het dat de paden naar de voedselvoorraden meteen ook de meest efficiënte blijken te zijn; het gedrag van elke mier zorgt er uiteindelijk voor dat de weg naar het voedsel het snelst is of het minste aantal hindernissen bevat. Er is dus sprake van een zelfversterkend gedrag dat niet door individuele mieren op zich veroorzaakt wordt, maar wel door de interactie van de gedragingen van de verschillende mieren. Het geheel is dus meer dan de som van de gedragingen van de individuele mieren.

Deze zelfversterkende interactie tussen de gedragingen van individuele leden van de groep wordt *zelforganisatie* genoemd, omdat er geen centrale entiteit is die de gedragingen coördineert. Het zijn louter de eenvoudige, individuele gedragingen die tot het globale, “intelligent aandoende” gedrag leiden.

3.2. Taal en zelforganisatie

Zoals het voorbeeld van de mierenkolonie en zijn fourageergedrag, kunnen groepen mensen en hun communicatiegedrag ook beschouwd worden als een complex dynamisch systeem. Taal is niet terug te brengen tot één enkel individu, maar is een fenomeen dat plaatsvindt op groepsniveau. De vraag blijft dan: welke precieze mechanismen vindt je bij het individu? De onderzoeksmethode die we hier gebruiken, beschouwt taal als een groepsfenomeen. We bouwen computermodellen die enerzijds dienst doen als concretisering van theorieën over de gedragingen van individuen, en anderzijds ook als bekrachtiging van de theorie. Als de modellen een gedrag vertonen dat overeenkomt met het gedrag van natuurlijke taal in gelijkaardige omstandigheden, dan verhoogt dit de waarschijnlijkheid dat de theoretische aannames juist zijn, en dus op één of andere manier ook in groepen mensen aan het werk zijn.

Concreet hebben we al modellen uitgewerkt op verschillende niveaus: de “interne” representatie van taal (bijvoorbeeld het lexicon), de “externe” representatie van taal (zoals fonetiek, zie de Boer, 2001) en de betekenis die aan taal moet vooraf gaan om taal zin te geven (bijvoorbeeld voor perceptuele categorieën, Steels & Belpaeme, 2005).

Op het niveau van “interne taal” vertrekken we van het allereenvoudigste geval, namelijk dat waar individuen namen geven aan objecten in hun omgeving. Dit geval is het eenvoudigste, omdat de namen slechts uit één woord bestaan. Taalgebruikers communiceren dus enkel door middel van enkelvoudige woorden. We beschouwen nog niet het combineren van woorden om samengestelde betekenissen uit te drukken.

Complexere modellen zijn ook mogelijk. Individuen zijn dan in staat om samengestelde zinnen te produceren. Hiervan bestaan drie varianten: een eerste variant waarin expressies nog steeds uit één woord bestaan, maar waar een relatief complexe semantiek achter zit, een model waarin expressies uit meerdere woorden kunnen bestaan, maar waar nog geen grammatica bestaat, en een derde variant die een aanzet tot grammatica bevat. Deze laatste modellen hebben telkens een niet-triviale semantiek.

3.3. In deze bijdrage

Verder in dit artikel worden enkele van deze modellen in detail beschreven. Concreet gaat het om een aantal lexicon-modellen: het basismodel, het “multi-word” model en om Belpaemes kleurenmodel. Van elk model wordt de opbouw besproken, de resultaten die ermee verkregen zijn, de interpretatie van deze resultaten, evenals wat ze (kunnen) betekenen voor het taalonderzoek.

4. Naming games

Het leren van een lexicon is op zich al een complex probleem. Bij het bestuderen ervan zijn we ruwweg in de volgende vragen geïnteresseerd.

1. Hoe komt het dat mensen dezelfde betekenissen gebruiken?
2. Hoe komt het dat mensen dezelfde woorden gebruiken voor dezelfde betekenissen?
3. Hoe komt het dat mensen woorden in zinnen rangschikken volgens dezelfde patronen?

Voor elk van deze problemen kunnen we een computermodel bouwen. In eerste instantie concentreren we ons op het ontstaan van het lexicon zelf (punt 2). Kijkend naar natuurlijke taal, dienen de volgende observaties als leidraad bij het ontwerpen van het model.

1. Kinderen moeten taal vanaf nul leren als ze geboren worden.
2. Taal verandert voortdurend, bijvoorbeeld door bewuste introducties van nieuwe uitdrukkingen, doordat er nieuwe voorwerpen onze leefwereld binnendringen, maar vooral ook onbewust, door klankverschuivingen en door subtiele herinterpretaties van bestaande uitdrukkingen, zie bijvoorbeeld (Lightfoot, 1999) die onder andere beschrijft hoe het suffix *-ed* in het Engels het standaardmorfeem voor het uitdrukken van verleden tijd is geworden.
3. Elke taal bestaat in verschillende varianten: dialecten, maar ook allerlei jargons, waarin bepaalde concepten andere namen krijgen dan in de overkoepelende taal, of concepten benoemd worden die in de overkoepelende taal niet benoemd zijn omdat ze daarin niet relevant zijn.

Punt 1 leidt ertoe aan te nemen dat mensen taal leren door *interacties* met mensen die de taal spreken, kinderen leren hun taal door communicatieve interacties met hun ouders of verzorgers. We moeten dus een manier voorzien waarop individuen kunnen interageren. Punt 2, en dan vooral de manier waarop taal onbewust verandert, en punt 3 leiden ons ertoe aan te nemen dat er geen centrale entiteit is die de taal definieert. Het gevolg hiervan is dat taal als een *gedistribueerd fenomeen* gezien wordt.

Bovendien is het onwaarschijnlijk dat taal vanaf het begin kan geleerd worden zonder ondersteunende communicatie. Een ouder laat bijvoorbeeld vaak op niet-linguïstische wijze aan zijn kind zien of hij hem of haar begrepen heeft of niet, door bijvoorbeeld te knikken, te lachen of het voorwerp aan te reiken waar het kind om vraagt.

We krijgen dus het volgende lijstje eisen waaraan ons model alvast moet voldoen (Steels et al, 2002).

- Er is geen centrale definitie van de taal.
- Individuen moeten met elkaar interageren;
- Individuen mogen bij aanvang geen taal hebben.
- Er moet een (beperkter) parallel communicatiekanaal zijn om een taal bij zijn ontstaan te ondersteunen.
- Individuen moeten een manier hebben om te leren uit het al dan niet slagen van een communicatieve interactie.

Er zijn ook een aantal basiseisen waaraan het model moet voldoen.

- Individuen moeten in staat zijn om te spreken en te luisteren.
- Individuen moeten in staat zijn om een collectie woorden samen met hun betekenis bij te houden;
- Er moeten dingen zijn waarover de individuen met elkaar kunnen spreken.

Deze laatste eisen lijken misschien triviaal, maar de manier waarop ze vorm gegeven worden in een model heeft toch een belangrijke invloed op hoe het zal werken. Neem bijvoorbeeld de eis dat individuen moeten kunnen spreken en luisteren. We kunnen woorden “schriftelijk” tussen de individuen uitwisselen, of een volledige akoestische ruimte simuleren waarin ze interageren. Dit laatste zou ideaal zijn, maar is veel moeilijker en volgens ons ook niet essentieel op dit niveau. We kiezen in onze modellen dus voor de eerste methode, maar niet zonder grondige evaluatie van de mogelijkheden.

4.1. Het basismodel

Aangezien we bestuderen hoe taal evolueert in een groep, hebben we meerdere individuen nodig. Het type simulaties waarbij meerdere *agents* interageren, noemt men *multi-agent* systemen (Jennings & Wooldridge, 1998). Een *multi-agent* systeem bevat een populatie van individuen (*agents*) die zich bevinden in een gemeenschappelijke omgeving, en waarbij ze elk op zich (of collectief) proberen een taak op te lossen. In dit geval bestaat het gemeenschappelijke doel erin om een succesvol communicatiesysteem op te bouwen. De gemeenschappelijke omgeving (ook wel *context* genoemd) van de *agents* bevat een aantal objecten die door de *agents* kunnen waargenomen worden.

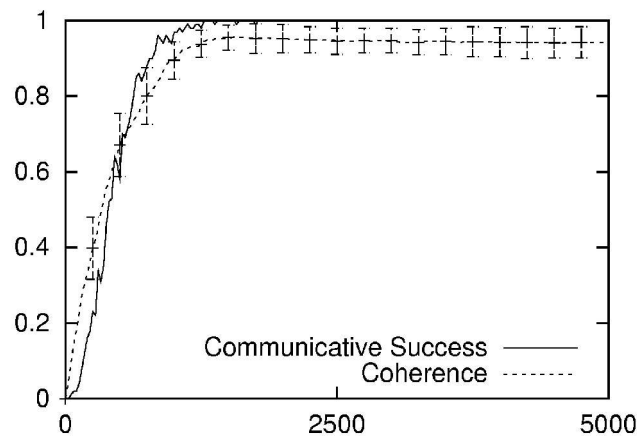
Verder moeten de *agents* nog met elkaar kunnen interageren, wat ze in dit geval doen door te spreken en te luisteren. Het eenvoudigste hier is om ze twee aan twee te laten interageren, waarbij de ene *agent* de spreker is, en de andere de luisteraar. De eenvoudigste interactie is het uitwisselen van een enkel woord. Het ligt dus voor de hand om de *agents* de objecten te laten *benoemen*. Ze moeten het met elkaar eens worden over een lijst namen voor de objecten in hun omgeving. Deze interactie noemen we een *naming game* (Steels, 1996). In het kort wordt het protocol voor een interactie.

1. de spreker kiest een object uit de omgeving;
2. de spreker kiest een woord voor dit object;
3. de spreker spreekt het woord uit;
4. de luisteraar hoort het woord;
5. de luisteraar probeert het woord te interpreteren;
6. als dit lukt, wijst de luisteraar in de omgeving het object aan die volgens hem bij het woord hoort;
7. de spreker bevestigt (niet-linguïstisch) of de luisteraar het juiste object aanwees.

De laatste stap biedt een mogelijkheid tot leren. Als de luisteraar het juiste object aanwijst, is de associatie tussen het gebruikte woord en de gevonden betekenis duidelijk goed, en moet ze versterkt worden. Omgekeerd, als de interpretatie verkeerd is, moet ze verzwakt worden, zodat deze combinatie in de toekomst met minder waarschijnlijkheid gebruikt zal worden.

4.1.1. Evaluatie van het model

De meest directe manier om de werking van het model te evalueren is te kijken hoe goed de communicatie werkt. Dit kan op een eenvoudige manier door per n spelletjes te kijken hoeveel er succesvol waren en hoeveel niet. Figuur 1 geeft de evolutie weer van een serie van 5000 spelletjes tussen 10 *agents*. Het communicatieve succes wordt gemeten per 25 spelletjes, zodat de meting tot op 4% nauwkeurig is. De curve begint bij 0 % (geen enkel gelukt spel), maar bereikt na ongeveer 2000 spellen het maximum van 100%. Dit betekent dat er eerst een “onderhandelingsperiode” is van 2000 spellen, waarin de *agents* verschillende namen geven aan elk object. Naar het einde van deze periode toe bereiken de *agents* een akkoord over de namen voor alle objecten, en raken de lexicons van alle tien *agents* voldoende op elkaar afgestemd om alle communicatie succesvol te laten verlopen. Dit impliceert ook dat na deze inleidende fase er in principe geen niet-linguïstische feedback meer nodig is, omdat de *agents* er voor 100% op kunnen vertrouwen dat hun linguïstische boodschap correct overkomt bij de luisteraar.



Figuur 1. Communicatief succes en coherentie in een populatie van 10 *agents*.

De tweede curve toont de coherentie van de lexicons van de tien *agents*. Hierboven werd al gezegd dat de lexicons voldoende *op elkaar afgestemd* geraken om 100% succes te kunnen garanderen in de communicatie. Dit betekent echter niet dat ze identiek zijn. In het bijzonder is het zo dat de *agents*, in hun eenvoud, geen mechanisme hebben dat ervoor zorgt dat ze weinig- of niet-gebruikte associaties tussen woorden en betekenissen vergeten. Deze associaties blijven met andere woorden passief aanwezig in de lexicons. Het netto-effect hiervan is niet dat zij de communicatie hinderen. In zekere zin geldt zelfs het tegendeel: het effect is dat een bepaalde *agent* meerdere woorden begrijpt voor een bepaald object, zelfs al verkiest hij één specifiek woord als hij zelf de spreker is. *Agents* kunnen elkaar dus begrijpen, zelfs al verkiezen ze verschillende woorden om over hetzelfde onderwerp te spreken.

4.1.2. Wat leert dit model over taal?

Essentieel kunnen we zeggen dat het belangrijkste resultaat is, dat het mogelijk is om in een populatie individuen vanaf nul tot een coherent, bruikbaar lexicon te komen door negotiatie, zonder dat dit lexicon van buitenaf opgelegd wordt. Het mechanisme dat daarvoor nodig is,

is het aanpassen van de associatie tussen woorden en betekenissen. Het bovenstaande mechanisme komt neer op het monitoren van de kennis van de andere *agents* en jezelf daaraan aanpassen om zo goed mogelijk begrepen te worden. Het model illustreert hoe een lexicon niet meer is dan een verzameling conventies die ontstaan binnen een groep.

4.1.3. Wat leert dit model niet?

Het model is zo eenvoudig dat het zeker niet zonder meer werkt voor complexere situaties. Zo blijft in dit model bijvoorbeeld de populatie *agents* stabiel. In werkelijkheid is dit natuurlijk niet zo: mensen worden geboren, sterven of migreren. Dat dit een invloed heeft op de dynamiek van de taal is onontkoombaar, en moet in complexere modellen bestudeerd worden.

Waar dit model ook geen rekening mee houdt is het feit dat echte communicatie niet perfect is. Het is mogelijk, bijvoorbeeld als er veel achtergrondlawaai is, dat wat de spreker zegt niet perfect door de luisteraar ontvangen wordt. Toch zijn mensen in staat om in moeilijke situaties betrouwbaar te communiceren, onder andere doordat communicatie wordt geïnterpreteerd aan de hand van contextuele informatie.

Dat deze tekortkomingen het basisbesluit van het model echter niet invalideren, wordt aangetoond in (Steels et al., 1998) waarin experimenten beschreven worden met modellen die uitgebreid zijn met onzekere transmissie en dynamische, evoluerende populaties. Ook in deze experimenten blijft 100% communicatief succes haalbaar, op voorwaarde dat storende factoren, zoals ruis op de communicatie of een te grote influx van *agents* in populatie, niet boven een bepaalde drempel uit komen. Welke deze drempel is, verschilt per model, en het heeft op dit niveau dan ook weinig zin om te proberen dit te extrapoleren naar natuurlijke taal.

Het belang van het al dan niet vergeten van weinig gebruikte of succesvolle associaties is niet duidelijk. Enerzijds zorgt dit voor hogere coherentie en kleinere lexicons, wat gunstig lijkt als men beschikt over een beperkte mentale capaciteit voor onthouden van een lexicon. Anderzijds kan men argumenteren dat mensen een veel grotere passieve woordenschat hebben dan de woordenschat die ze actief gebruiken. Weinig gebruikte woorden moeten dus wel in zekere mate nog in het lexicon van de *agents* blijven zitten.

4.2. Naming games met semantiek

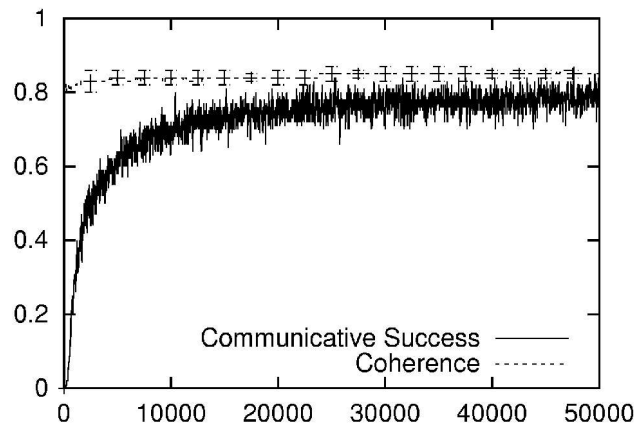
Misschien wel de belangrijkste kritiek op het basismodel is dat er semantiek ontbreekt. De objecten waar de woorden aan gekoppeld worden, zijn abstract. In het bovenstaande model, zijn de objecten rechtstreeks de betekenis. De semiotiek leert echter dat objecten en waarnemingen twee gescheiden dingen zijn. Wanneer men een object waarneemt, resulteert dit in een betekenis².

Dit heeft belangrijke consequenties, omdat de relatie referent-woord in het basismodel rechtstreeks is. Met het gebruik van betekenis vervalt deze rechtstreekse relatie en komt er een extra variabele bij die voor ambiguïteit kan zorgen: vooraleer een woord kan opgezocht worden voor een object, moet eerst een passende betekenis gevonden worden, die op zich (in een andere context) van toepassing kan zijn op andere objecten.

² Dit lijkt op de semiotiek van de Saussure (1916,1974), waarbij een “teken” bestaat uit een tastbaar object en het waargenomen object.

4.2.1. Evaluatie van het model

De grafiek in Figuur 2 toont een experiment met een naming-gamemodel dat semantiek bevat. Het is duidelijk dat de loskoppeling van referenten en betekenissen, die vroeger één en hetzelfde waren, een belangrijke impact heeft op het model: het communicatieve succes ligt in de limiet ongeveer 20% lager dan in het basismodel. De coherentie ligt ongeveer 10% lager, maar aangezien de coherentie berekend wordt over de lexicons van de *agents*, wordt in dit geval de coherentie tussen woorden en betekenissen berekend, en niet meer tussen woorden en referenten. Beide vormen van coherentie zijn bijgevolg niet rechtstreeks te vergelijken.



Figuur 2: Communicatief succes en coherentie met 10 *agents* (met semantiek)

4.2.2. Wat leert dit model over taal?

In elk geval is het zo dat de loskoppeling van referenten en betekenissen de dynamica van het model een stuk complexer maakt. Een object kan nu immers door meerdere betekenissen voorgesteld worden, en één betekenis kan (afhankelijk van de context) naar verschillende objecten verwijzen. Dit betekent dat expressies ambigu kunnen zijn, wat weerspiegeld wordt door het lagere communicatieve succes.

Een aanpassing van de mechanismen kan het succes misschien weer doen stijgen, zoals het gebruiken van een vergeetmechanisme om het aantal mogelijke hypothesen bij interpretatie te beperken, of het gebruik van een mechanisme dat beter om kan met meerdere hypothesen, zoals beschreven in (Steels et al., 1998).

4.2.3. Wat leert dit model niet?

Uiteraard is de betekenismodule geen getrouwe kopie van de “menselijke betekenismodule,” voor zover die al in een gelijkaardige vorm zou bestaan. Waar het hier om gaat is de loskoppeling van referenten en woorden, en het simuleren van een aantal karakteristieken van betekenis, zoals het feit dat één object met meerdere betekenissen kan beschreven worden (groot, blauw) of dat één betekenis van toepassing kan zijn op meerdere objecten. Het model toont hoe talige interacties erin slagen om ambiguë betekenissen te ontrafelen. Dit is analoog met het “gavagai”-voorbeeld van Quine (1980). Quine vertelt het verhaal van een ontdekking-reiziger die een indiaan ontmoet. Bij het zien van een voorbij rennend konijn

roept de inboorling “gavagai”. De ontdekkingsreiziger kan niet weten wat gavagai betekent: het kan “konijn” betekenen, maar evengoed kan het “rennen” betekenen, of “vierdimensionele aanwezigheid van iets konijnachtigs”. Echter, herhaaldelijk het woord gavagai horen in verschillende contexten laat toch toe om de betekenis ervan te deduceren. Op een zelfde manier gaan de *agents* in de simulatie tewerk. Ze gebruiken en horen woorden in verschillende contexten, en komen zo tot een consensus van wat elk woord betekent.

4.3. Meerwoord-naming games

Als we, bovenop de aannames impliciet en expliciet in de voorgaande modellen, aannemen dat de linguïstische geheugen- en verwerkingscapaciteit van een individu niet oneindig is, wordt het duidelijk dat een communicatiesysteem dat enkel één woord toelaat (een holistisch systeem) erg belastend is voor de cognitieve capaciteit van het individu. Een logische “technische” stap zou dus zijn om meerwoords-expressies in onze modellen toe te laten.

Ook taalkundig gezien komt dit overeen met recente voorstellen over hoe taal geëvolueerd is. Jackendoff (2002) stelt een fase genaamd “concatenation of symbols” voor als opvolger van de fase die gemodelleerd wordt door de éénwoord-naming game (“use of symbols in a non-situation-specific fashion”).

Op zich lijkt het een kleine stap te zijn: in feite gaat het gewoon over meerdere éénwoords-games na elkaar, met hetzelfde onderwerp. Zo eenvoudig is het echter niet. In meerwoord-games moet de betekenis, net als in éénwoord-games, uniek zijn voor het onderwerp van de conversatie, en dit laat niet toe om een meerwoord-game te modelleren door verschillende éénwoord-games. Als een spreker slechts één woord kan gebruiken om een betekenis te benoemen, betekent dit dat bij het opzoeken in het lexicon telkens de volledige betekenis moet overeenkomen met de uit te drukken betekenis. Als de spreker meerdere woorden kan gebruiken, moet er ook rekening gehouden worden met woorden die gedeeltelijk overeenkomen. Vervolgens moeten de gevonden woorden zodanig gecombineerd worden dat de combinaties van de betekenissen de volledige oorspronkelijke betekenissen dekken.

Ook de taak van de luisteraar wordt ingewikkelder: hij moet niet enkel één woord in zijn lexicon opzoeken om de betekenis te vinden, maar moet de betekenissen van alle woorden opzoeken en combineren. De extra moeilijkheid hierbij is dat een woord meerdere verschillende betekenissen kan hebben, en dat er dus verschillende combinaties mogelijk kunnen zijn. Stel bijvoorbeeld dat een *agent* een lexicon heeft die de volgende woord-betekenisparen bevat:

wabaku	ROOD
wabaku	LICHT
wogglesplat	GROOT

Tabel 1. Voorbeeldlexicon

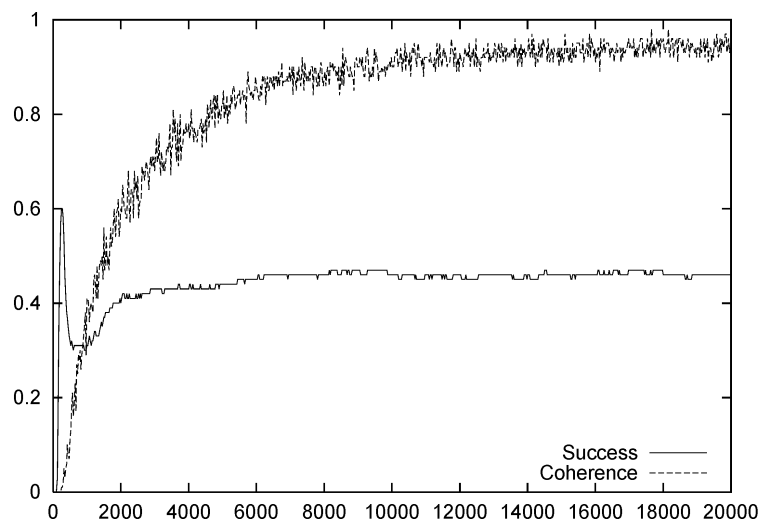
Als de *agent* “wabaku wogglesplat” hoort, zijn er niet één, maar twee mogelijke interpretaties: ROOD-GROOT en LICHT-GROOT. De *agent* moet op basis van andere informatie beslissen welke betekenis hij kiest. Dit kan op basis van een waarschijnlijkheid

zijn: woord-betekenisparen kunnen een sterkte krijgen, die uitdrukt hoe sterk de associatie is, en deze sterktes kunnen gebruikt worden om de totaalsterkte voor een expressie te berekenen. Op een moment dat de *agent* nog niet genoeg vertrouwen heeft in zijn communicatiesysteem, zal hij de betekenis(sen) die hij vindt ook toetsen aan de context om te achterhalen welke de spreker bedoelde. Naarmate zijn vertrouwen in de taal stijgt, wordt ze minder context-afhankelijk, en laat ze ook betrouwbare communicatie toe over abstractere onderwerpen.

In het computermodel is dit een vrij complexe operatie. Het is echter niet duidelijk hoe groot deze stap is voor een biologisch neurale netwerk. Aangezien biologische neurale netwerken goede patroonherkenners zijn, ook voor onvolledige patronen, is de stap hier misschien niet zo groot.

4.3.1. Evaluatie van het model

Figuur 3 laat zien dat het communicatieve succes minstens even hoog is als voor de éénwoord-*naming game*. Het lijkt er zelfs op dat het succes hoger is dan voor de éénwoord-*naming game* met betekenis. De coherentie is laag, maar, net zoals voor de semantische *naming game* ligt dit aan het feit dat er geen rechtstreekse correlatie meer is tussen woorden en referenten; referenten zijn hier volledig extern aan de *agent*, en worden intern voorgesteld door geabstraheerde sensor-informatie, waaruit de betekenissen geconstrueerd worden die voor de communicatie gebruikt worden.

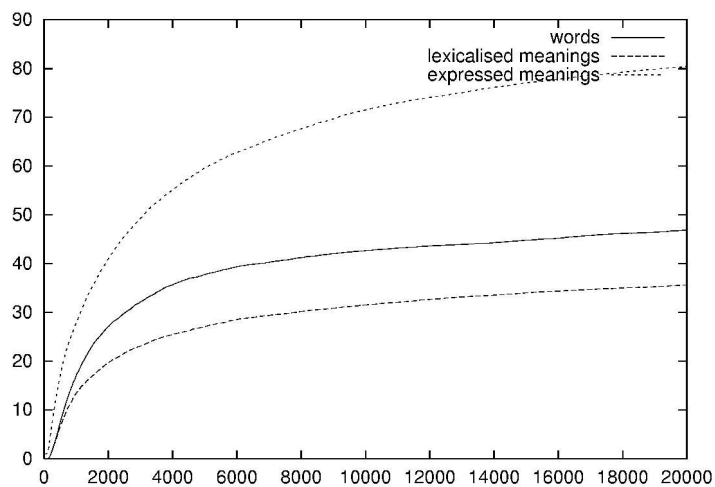
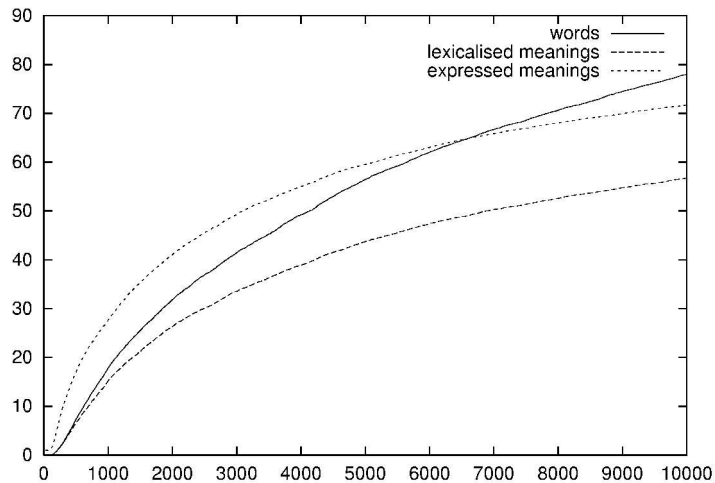


Figuur 3. Succes en coherentie in een meerwoord-*naming game* met 10 *agents* (gemiddelde van 10 reeksen).

Figuur 4 laat de evolutie van het lexicon zien in éénwoord en meerwoord-*agents* die in gelijkaardige omstandigheden taal moeten ontwikkelen. Het is duidelijk dat in het éénwoordmodel het aantal woorden veel sneller stijgt dan in het meerwoordmodel (merk op dat in het meerwoordmodel dubbel zoveel games gedaan zijn als in het éénwoordmodel).

In het éénwoordmodel lopen de curves die het aantal uitdrukkingen en uitgedrukte betekenissen weergeven parallel. Het aantal woorden stijgt sneller dan de beide andere curves in de grafiek, wat betekent dat er meer woorden aangemaakt worden dan er (verschillende) betekenissen uitgedrukt moeten worden. Het komt bijgevolg voor dat er

meerdere woorden per nieuwe betekenis ontstaan. In het meerwoordmodel lopen het aantal woorden en het aantal betekenissen parallel, terwijl het aantal uitgedrukte betekenissen veel groter is. Het is dus duidelijk dat meerwoord-*agents* met veel kleinere lexicons toch in staat zijn om meer betekenissen uit te drukken.



Figuur 4. Aantal betekenissen en woorden in éénwoord- respectievelijk meerwoord *naming games* met 10 *agents* (10.000 respectievelijk 20.000 spellen, gemiddelden over 10 reeksen).

4.3.2. Wat leert dit model over taal?

Een duidelijk verschil tussen de éénwoord-*naming game* en de meerwoord-*naming game* is dat het lexicon kleiner blijft: de *agents* zijn in staat om veel meer betekenissen uit te drukken met een veel kleiner lexicon. Dit lijkt op het eerste zicht triviaal, maar het is toch wel duidelijk dat de mechanismen die nodig zijn om de productie en interpretatie van expressies een stuk complexer zijn voor de meerwoord-*naming game*. Het is niet een kwestie van

verschillende éénwoord-*naming games* na elkaar te spelen met hetzelfde onderwerp; betekenissen moeten samengesteld en uit elkaar gehaald worden aan de hand van de woorden die de expressie van de spreker bevat. Wat de impact van deze complexiteit op biologisch vlak zou kunnen zijn, is echter niet duidelijk.

4.3.3. Wat leert dit model ons niet?

We mogen zeker niet de fout begaan om te denken dat de mechanismen die gebruikt zijn in deze modellen, ook de mechanismen zijn die in het echt het werk doen in de menselijke hersenen. Biologische systemen werken anders dan computers, en zijn goed in patroonherkenning. Dit zou erop kunnen wijzen dat de “technische” sprong die in de simulaties nodig was om van éénwoords naar meerwoord-*naming games* te kunnen gaan, in biologische modellen kleiner zou kunnen zijn. Met andere woorden, wat in het computermodel een grote technische sprong is, zou in de biologische variant een “continue” evolutie kunnen zijn. Echter, zoals al eerder gezegd weten we op dit moment te weinig van de biologische werking (laat staan over de evolutie) van de hersenen om hierover steekhoudende uitspraken te doen.

5. De oorsprong van kleurcategorieën

De hierboven beschreven methodologie leent zich uitstekend tot het onderzoeken van één van de meest controversiële onderwerpen van de cognitieve psychologie: de aard van perceptuele categorieën. De mens deelt zijn belevingswereld op in categorieën: hoewel een regenboog miljoenen verschillende tinten bevat, nemen we slechts een handvol verschillende kleuren waar. Categorieën zijn voor verschillende redenen nuttig. Eén reden is het vereenvoudigen van de wereld om ons heen, wat toelaat om op een zuinige manier om te springen met cognitieve capaciteiten: in plaats van duizenden verschillende stoelen te onthouden die we ooit hebben waargenomen, onthouden we één stoelcategorie die een aantal eigenschappen heeft zoals het hebben van een vlak om op te zitten en het hebben van een rugleuning, die samen uitmaken wat een stoel precies is. Categorieën zijn ook van uitzonderlijk belang voor communicatie. Mensen communiceren met elkaar door het overdragen van woorden, die elk gekoppeld zijn aan één of meerdere categorieën. Zonder categorieën zou het onmogelijk zijn om een menselijke taal te hebben³. Een voorwaarde voor geslaagde communicatie is dat beide partijen min of meer gelijkaardige categorieën hebben. Zoniet is het onmogelijk om met een woord eenzelfde categorie of concept aan te duiden en faalt de communicatie.

Perceptuele categorieën zijn categorieën die onmiddellijk gekoppeld zijn aan onze perceptie. Zo bestaan er olfactorische of gustatorische categorieën afkomstig van geur- en smaakperceptie (Dubois, 2000). De perceptuele categorieën waarmee we misschien het meest mee vertrouwd zijn, zijn kleurcategorieën. Het aantal verschillende kleuren die een persoon met normale kleurperceptie kan onderscheiden wordt geschat op 10 miljoen, maar het aantal kleurcategorieën is veel beperkter. Om kleurcategorieën te onderzoeken, wordt meestal een omweg gemaakt via kleurtermen. Aan de hand van geheugentesten en

³ Niet alle communicatiemiddelen hebben nood aan categorieën. Verschillende diersoorten communiceren met een communicatiesignaal dat niet op te splitsen valt in eenheden die gekoppeld zijn aan categorieën. Sommige vogels, bijvoorbeeld, zingen een langgerekt lied, maar bakenen hiermee enkel hun territorium af of imponeren wijfjes en mogelijke concurrenten (Catchpole & Slater, 1995).

benoemingstesten (Lenneberg & Roberts, 1956) kan nagegaan worden welke kleurcategorieën een proefpersoon heeft. Algemeen wordt aanvaard dat het aantal kleurtermen, en dus ook het aantal kleurcategorieën, in verschillende culturen varieert van twee tot een twintigtal.

Lange tijd werd vermoed dat verschillende culturen het continue kleurenspectrum willekeurig opdeelden in kleurcategorieën. Gleason (1961) geciteerd in (Berlin & Kay, 1969, p. 159) schrijft: “There is a continuous gradation of color from one end of the spectrum to the other. Yet an American describing it will list the hues as red, orange, yellow, green, blue, purple, or something of the kind. There is nothing inherent either in the spectrum or the human perception of it which would compel its division in this way.”

De these dat kleurcategorieën uniek waren voor elke cultuur werd grondig herzien na de publicatie van Berlin en Kays monografie in 1969. Hierin worden resultaten beschreven van experimenten waarbij proefpersonen werden gevraagd om op een kleurenkaart de kleurtermen van hun taal aan te duiden. Het bleek dat focussen van kleurtermen uit verschillende talen, van Libanees Arabisch tot Amerikaans Engels, op zo goed als dezelfde plaatsen vielen op de kleurenkaart. Kleurcategorieën leken dus toch niet relatief te zijn, maar eerder universeel. Het universele karakter van kleurcategorieën werd in meerdere studies bevestigd (onder andere Rosch-Heider, 1972; Rosch, 1973; Bornstein, 1985; Kay & Regier, 2003), maar werd ook af en toe in twijfel getrokken. Saunders en van Brakel (1997) bekritiseerden de methodologie van de experimenten en drongen aan op een meer genuanceerde definitie van kleurencategorisering. Ook kwam er kritiek uit de hoek van experimentele psychologen die er niet in slaagden de resultaten van Rosch-Heider (1972) te repliceren (Davidoff et al., 1999).

Toch blijft er een groeiende consensus bestaan over het feit dat kleurcategorieën een, in mindere of meerdere mate, universeel karakter hebben. De vraag die zich meteen opdringt is: als kleurcategorieën een universeel karakter hebben, hoe komt dat dan? In de literatuur kunnen we drie verschillende verklaringen onderscheiden, die hieronder worden beschreven.

De eerste verklaring beroept zich op het aangeboren zijn van mechanismen die leiden tot universele kleurcategorieën: deze strekking noemen we *nativisme*. Een sterke versie van het nativisme beweert dat kleurcategorieën rechtstreeks genetisch bepaald zijn. Een zwakkere versie tracht de aard van kleurcategorieën te verklaren vanuit het feit dat alle mensen dezelfde perceptuele and cognitieve functies delen. De neurofysiologie van menselijke kleurenperceptie stelt een aantal grenzen, waarbinnen de ontwikkeling van kleurcategorieën gedwongen wordt. Onze trichromatische kleurenperceptie, en het opponent karakter van niet alleen de subjectieve kleurenwaarneming, maar ook van de neurale paden naar de visuele cortex (De Valois et al., 1966), zorgen ervoor dat er zich een aantal voorspelbare kleurcategorieën ontwikkelen. Het nativisme, zowel de sterke als de zwakke versie, sluit elke invloed van taal of cultuur op het vormen van kleurcategorieën uit (Bornstein, 1985).

Een tweede verklaring voor de aard van kleurcategorieën kan gezocht worden in de omgeving. Onze omgeving, of ecologie, bevat een bepaalde chromatische distributie: sommige kleuren komen meer voor dan andere. Uit deze statistische eigenschap kunnen kleurcategorieën gededuceerd worden. Deze strekking noemen we *empirisme*. Het empirisme is aanlokkelijk om de aard van perceptuele categorieën te verklaren: onze zintuiglijke waarneming monstert de wereld en onttrekt uit de waargenomen structuur categorieën. Shepard (1992; 1994) stelt dat verschillende ecologische constanten, zoals zwaartekracht, aardrotatie en ook chromatische structuur van de wereld, geïnternaliseerd

zijn. Shepard gelooft dat chromatische ecologie in de loop van onze evolutie geabsorbeerd werd in onze genen, maar het kan evengoed zo zijn dat kleurcategorieën ontogenetisch geleerd worden, zoals gesuggereerd door Van Wijk (1959) en Yendrikhovskij (2001). Opnieuw wordt hier elke invloed van taal of cultuur uitgesloten.

Een derde verklaring stelt dat kleurcategorieën cultureel worden geleerd, de strekking noemen we *culturalisme*. Kleurcategorieën zijn dus specifiek voor een cultuur, wat echter niet uitsluit dat twee verschillende culturen geen gelijkaardige kleurcategorieën kunnen hebben. De belangrijkste invloed op het verwerven van kleurcategorieën is de taal. Het communiceren over kleur bepaalt hoe onze kleurcategorieën zich vormen en vice versa: kleurcategorieën bepalen de manier waarop er over kleur gecommuniceerd wordt. Merk op dat bij empirisme, waar categorieën ook geleerd worden, taal geen enkele invloed heeft op het leerproces.

Het idee dat taal en cognitie met elkaar verbonden zijn staat bekend als *linguïstische relativiteit* (Sapir, 1921; Whorf, 1956). Opponenten van de linguïstische relativiteits-hypothese (LRH) ridiculiseren vaak een sterke vorm van de hypothese, hierin wordt gesteld dat concepten waarvoor geen woord bestaat niet kunnen gedacht worden (zie bijvoorbeeld Pinker, 1994). Niemand echter verdedigt de sterke variant, waardoor de aanhoudende kritiek bijna vermakelijk lijkt. Maar de zogenaamde “zwakke” LRH, waarin men een causaal verband mogelijk acht tussen taal en cognitie, kent toenemende aanhang, ook onder kleurenonderzoekers (Lucy & Shweder, 1979; Kay & Kempton, 1984; Gellatly, 1995; Davies, 1998; Roberson, 2005).

Aanhangers van de drie verklaringen gebruiken argumenten uit verschillende disciplines om hun stellingen kracht bij te zetten. Ontwikkelingspsychologie, waar de aanwezigheid (Bornstein, 1985; Davies & Franklin, 2002) of de afwezigheid (Gerhardstein *et al.*, 1996) van kleurcategorieën bij pasgeborenen wordt nagegaan. Antropologisch onderzoek naar het al dan niet aanwezig zijn van cognitieve kleurcategorieën⁴ (Rosch-Heider, 1972; Roberson *et al.*, 2000) of het aanwezig zijn van pan-culturele gelijkmatigheden in het benoemen van kleuren (Kay & Regier, 2003; Kay *et al.*, 2003). Psychologen proberen aan de hand van subjectieve kleurenperceptie de oorsprong van kleurcategorieën te verklaren (Boynton & Olson, 1987; Jameson & D’Andrade, 1997). Neurofysiologie, waarbij men de oorsprong van kleurcategorieën tracht te vinden in de structuur van neurale paden (De Valois & De Valois, 1975).

De discussie over de oorsprong van kleurcategorieën leent zich perfect tot een studie met een simulatiemodel. Aangezien na meer dan honderd jaar academisch over en weer vuren van argumenten het conflict nog steeds laait, maken we ons helemaal niet de illusie dat de hieronder beschreven experimenten en resultaten het pleit beslechten. Ze kunnen hoogstens een andere zienswijze bieden en eventueel nieuwe experimenten suggereren aan de cognitieve wetenschap.

5.1. Simuleren van *culturalisme*

Het *culturalisme*, of linguïstisch relativisme, heeft een bijzondere aantrekkingskracht. Enerzijds is het een proces dat zich op de tijdschaal van een generatie afspeelt, in tegenstelling tot *nativisme* wat een proces is wat zich spreidt over honderden, mogelijk

⁴ Om verwarring te vermijden wordt vaak het onderscheid gemaakt tussen *cognitieve categorieën*, categorieën die aanwezig zijn maar niet geassocieerd zijn met een woord, en *talige categorieën*, dit zijn categorieën die identiek zijn aan cognitieve categorieën met uitzondering van het feit dat ze gelexicaliseerd zijn.

duizenden, generaties. Een culturalistisch leerproces is bijgevolg in staat om snel veranderende ecologische en cultu-rele omstandigheden bij te benen. Genetische evolutie baseert zich op het voortplanten van willekeurige mutaties en recombinaties van genetisch materiaal en is daardoor noodgedwongen een traag proces. Daarnaast is het feit dat taal de perceptuele ervaring opdeelt een erg aantrekkelijke verklaring voor het vormen van kleurcategorieën.

Maar, nog nooit werd beschreven hoe linguïstische relativiteit precies zou kunnen tewerk gaan. Om een inzicht te krijgen in dit culturalistisch proces wordt een “minimaal” simulatiemodel gebouwd. Dit houdt in dat we enkel modelleren wat nodig geacht wordt om de verklaringen te bestuderen, en dat alle balast buiten het model gehouden wordt. Zo wordt vermeden dat onnodige elementen in het model het gedrag van het model vertekenen. Als men bijvoorbeeld wenst te bestuderen hoe files ontstaan aan de hand van een simulatiemodel, dan moeten er wegen en wagens geïntegreerd worden in het model, maar de kleur van elke wagen hoeft niet gemodelleerd te worden, aangezien je redelijkerwijs mag veronderstellen dat dit geen invloed heeft op het ontstaan van files.

De simulatie gebruikt opnieuw *agents*, gesimuleerde individuen die elk in staat zijn om kleuren waar te nemen en kleuren te categoriseren. Elke *agent* heeft dezelfde manier van kleurenperceptie en kleurencategorisering. In de volgende paragrafen gaan we dieper in op de perceptie en de categorisatie.

5.1.1. Kleurenperceptie

Kleurenstimuli worden aangeboden in de vorm van tristimuluswaarden in de CIE $L^*a^*b^*$ -ruimte. CIE $L^*a^*b^*$ is een kleurenmodel waar elke kleur voorgesteld wordt door drie waarden. De L^* -waarde geeft aan hoe licht een kleur is, de a^* -waarde is het rood-groen kanaal en het b^* -waarde is het geel-blauw kanaal. CIE $L^*a^*b^*$ stelt een psychologisch equidistante kleurenruimte voor, dit betekent dat de ruimte de menselijke subjectieve kleurenperceptie weergeeft en dat kleuren met een eenvoudige afstandsmaat kunnen vergeleken worden. Hoewel CIE $L^*a^*b^*$ ontworpen werd voor technische applicaties, leent ze zich uitstekend tot het modelleren van menselijke kleurenperceptie (o.a. Lammens, 1994; Sturges & Whitfield, 1995; Yendrikhovskij, 2001). Merk op dat hoewel CIE $L^*a^*b^*$ een model is voor de menselijke kleurenperceptie, dit niet betekent dat het ook menselijke kleurencategorieën modelleert; CIE $L^*a^*b^*$ geeft enkel weer hoe gelijk we twee verschillende kleuren ervaren.

5.1.2. Kleurencategorisering

Elke *agent* kan kleurenstimuli categoriseren. Hierbij gaan we ervan uit dat *agents* kleuren categoriseren als prototypes (Rosch, 1978). Prototypecategorieën hebben een typische “piek”, die het prototype voorstelt. Alles wat niet met het prototype overeenkomt, stemt gradueel minder overeen met die categorie. Bijvoorbeeld, de kleur van een rijpe tomaat is prototypisch voor de categorie ROOD; een roze tint behoort tot op zekere hoogte nog bij ROOD, maar is niet meer prototypisch. Kleurcategorieën hebben daarnaast nog de eigenschap dat één enkele kleurcategorie steeds een aaneengesloten gebied in de psychologische kleurenruimte bestrijkt.

Om dit te simuleren bestaan er verschillende mogelijkheden, in deze simulaties gebruiken we *adaptieve netwerken*. Een categorie wordt voorgesteld door precies één adaptief netwerk (voor details zie Belpaeme, 2001; Belpaeme, 2002; Steels & Belpaeme, 2005). De adaptieve netwerken implementeren daarenboven categorische perceptie. Categorische perceptie (Harnad, 1987; Bornstein, 1987) is het effect waarbij stimuli behorende tot eenzelfde

categorie als minder verschillend worden ervaren, dan stimuli die even ver van elkaar afliggen, maar tot verschillende categorieën behoren. Een categorie oefent als het ware aantrekkingskracht uit op percepten, vandaar dat men soms ook spreekt van het “magneeteffect”.

5.1.3. Kleurlexicalisering

Om in staat te zijn te communiceren over kleuren, moet elke *agent* zijn kleurcategorieën kunnen verbinden met kleurwoorden. Een verbinding tussen een woord en categorie heeft een bepaalde sterkte. Wanneer de sterkte een waarde 1 heeft is de verbinding het sterkst; hoe kleiner de sterkte hoe zwakker de band tussen het woord en de categorie.

5.1.4. Simulatie

Elke *agent* heeft de mogelijkheid om kleuren waar te nemen, te categoriseren en te verbinden met een woord. Maar wat nog niet duidelijk is, is hoe een *agent* zijn categorieën en lexicon leert, en hoe beide processen elkaar kunnen beïnvloeden.

Om categorieën te leren, gebruikt een *agent* een eenvoudige interactie, een discriminatiespel (*discrimination game*) genoemd (Steels, 1997; Belpaeme et al., 1998). Een discriminatiespel wordt gespeeld door slechts één *agent*, en dient om een repertoire van categorieën op te bouwen waarmee de *agent* onderscheid kan maken in de verschillende kleuren die hij waarneemt. Een discriminatiespel wordt gevolgd door een *naming game*, zoals beschreven in 4.1.

Tijdens een discriminatiespel worden een aantal kleuren getoond aan de *agent*, dit noemen we de context, uit die kleuren wordt één kleur gekozen als het onderwerp. Wanneer het onderwerp behoort tot een categorie waartoe geen enkele andere kleur uit de context behoort, dan slaagt het discriminatiespel. Maar, wanneer de *agent* er niet in slaagt een categorie te vinden die uniek bij het onderwerp past, dan faalt het spel. Bij falen wordt ofwel een categorie toegevoegd aan de *agent* ofwel worden de bestaande categorieën van de *agent* zo aangepast dat de kans op slagen in toekomstige discriminatiespellen verhoogt. Zo worden met elk discriminatiespel kleine wijzigingen aangebracht in de kleurcategorieën van een *agent*, net zolang tot de *agent* foutloos kleuren kan onderscheiden.

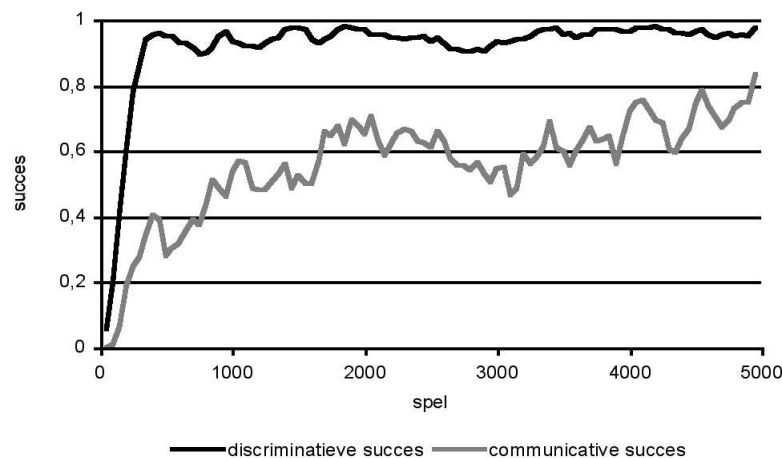
Twee *agents* kunnen samen een *naming game* spelen. Hierbij vervult één *agent* de rol van spreker en de andere *agent* de rol van luisteraar. Beide wordt een aantal kleuren getoond, waarvan enkel de spreker weet wat het onderwerp van de communicatie is. De spreker speelt eerst een discriminatiespel: wanneer een categorie gevonden wordt die uniek past bij het onderwerp, zal de spreker het woord dat het sterkst geassocieerd is met deze categorie uitspreken. In het geval de categorie nog geen bijhorend woord heeft, verzint de spreker een nieuw woord. De luisteraar probeert het woord te interpreteren door na te gaan welke van zijn categorieën het sterkst geassocieerd is met het woord. Daarna wijst de luisteraar de kleur aan die het best overeenstemt met deze categorie. De spreker laat de luisteraar weten of dit inderdaad het onderwerp was.

Een *naming game* kan op een aantal punten mislukken: de spreker vindt geen discriminerende categorie voor het onderwerp, of de luisteraar kent het woord niet wat de spreker heeft gesproken, of de luisteraar slaagt er niet in het onderwerp aan te wijzen. Maar, als een *naming game* mislukt, dan biedt dat een mogelijkheid om de categorieën en het lexicon van elke *agent* aan te passen, zodat de kans op slagen in volgende *naming games* toeneemt. Hierbij zijn twee mechanismen van belang. Het eerste mechanisme past de sterkte aan van de associatie tussen een woord en een categorie. Als een *naming game* slaagt, wordt

de sterkte tussen het gebruikte woord en de categorie verhoogd, en tegelijkertijd worden de sterktes van concurrerende categorieën verlaagd. Deze techniek noemen we *laterale inhibitie* en dient om een eenduidige associatie te leren tussen woorden en categorieën, zo wordt synonymie vermeden. Het tweede cruciale mechanisme maakt kleine aanpassingen aan de categorieën van de spreker en de luisteraar. Bij een geslaagd spel wordt de categorie van de luisteraar sterker gemaakt; bij een mislukt spel, wordt de categorie van de luisteraar iets in de richting van het onderwerp opgeschoven (meer details in Belpaeme, 2002; Steels & Belpaeme, 2005).

5.2. Resultaten

Wanneer een populatie van *agents* een reeks *naming games* speelt, dan bouwen ze elk een repertoire van categorieën en daarmee geassocieerde woorden op. Figuur 5 toont een simulatie waar 10 *agents* een reeks van 5000 *naming games* spelen. De grafiek toont enerzijds het discriminatieve succes (DS), hoe goed kunnen de *agents* kleuren discrimineren, en anderzijds het communicatieve succes (CS), hoe goed kunnen de *agents* kleuren communiceren. Het communicatieve succes is altijd lager of gelijk aan het discriminatieve succes, omdat het communiceren steunt op het kunnen onderscheiden van kleuren. Beide stijgen al snel, het DS benadert 100%, het CS stijgt tot ongeveer 80%. De *agents* hebben dus geen perfecte kleuren communicatie, wat niet helemaal verwonderlijk is. Ook mensen slagen er vaak niet in om overeenstemming te bereiken over het benoemen van kleuren, vooral als het gaat om kleuren die tussen twee categorieën in liggen, zoals geelachtig groen.



Figuur 5: Discriminatief en communicatief succes voor *naming games* over kleur.

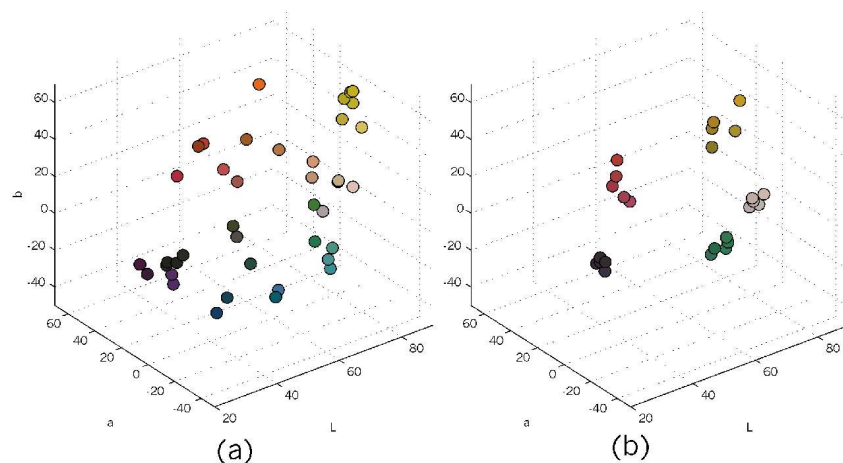
Wat echt interessant is aan dit model is hoe de semantiek gekoppeld aan de woorden verandert onder invloed van de talige interacties. Telkens als twee *agents* een *naming game* spelen heeft het al of niet slagen van het spel een invloed op de kleurcategorieën van de *agents*. Zo worden tijdens het spelen van taalspellen de kleurcategorieën van de *agents* gedwongen op elkaar te lijken. Mochten de categorieën van de *agents* niet op elkaar lijken, dan zouden de *agents* er niet in slagen om met elkaar te communiceren over de waargenomen kleuren. Figuur 6 illustreert dit, (a) toont de kleurcategorieën van vijf *agents* die enkel discriminatiespellen spelen. Elke *agent* heeft een verzameling van kleurcategorieën waarmee perfect kleuren te discrimineren zijn, maar de kleurcategorieën liggen verspreid in de kleurenruimte, er is geen sprake van coherentie tussen de kleurcategorieën van de *agents*.

Figuur 6(b) toont de kleurcategorieën van vijf *agents* na het spelen van een reeks *naming games*. Elke *agent* heeft een verzameling kleurcategorieën die de *agent* in staat stelt om kleuren te onderscheiden van elkaar, ook is elke kleurcategorie nu geassocieerd met een woord. De talige interacties hebben echter een invloed uitgeoefend op de kleurcategorieën van *agents* om op elkaar te lijken.

5.3. Wat leert dit over kleurcategorieën en taal

Inzicht in de aard van perceptuele categorieën geeft inzicht in een cruciaal aspect van menselijke cognitie. Hoe perceptuele categorieën ontstaan en hoe ze gevormd worden, kan een antwoord geven op hoe andere categorieën en concepten zich vormen en wat de invloed is van taal op dit alles. Over het onderzoek naar de aard van kleurcategorieën en het verband met taal zegt Deacon (1997, p. 120): "... this may at first appear to be a comparatively trivial example of some minor aspect of language, but the implications for other aspects of language evolution are truly staggering."

Het eenvoudige model wat hier getoond wordt, suggereert hoe taal een invloed kan uitoefenen op het vormen van categorieën en concepten. Meer nog, het toont ook hoe taal de motor kan zijn achter het creëren van coherente semantische representaties tussen verschillende individuen. Er zijn natuurlijk nog andere mogelijke verklaringen voor de coherentie tussen mentale representaties, zoals het hierboven besproken nativisme en empirisme, die hier niet gemodelleerd zijn. Het besproken culturalistisch model kan dus niet dienen om een uitspraak te doen voor of tegen één van de drie verklaringen. Het toont enkel aan dat culturalisme minstens een even sterke kandidaat is als het nativisme en het empirisme om de fenomenologie van kleurcategorieën te verklaren.



Figuur 6: De locatie van de kleurcategorieën van vijf *agents*, (a) zonder talige interactie en (b) met talige interactie. Taal zorgt ervoor dat de categorieën van de *agents* coherenter worden.

6. Besluit

Het onderzoek naar de oorsprong van taal is bijzonder complex en veelomvattend. Het feit dat mensen uit zoveel verschillende disciplines betrokken bij het onderzoek zijn, wijst erop dat de oorsprong en de evolutie van taal fascineert en intrigeert. Ook de vele verwijzingen naar en het vergelijkingen met mensapen –die tenslotte evolutionair het dichtst bij ons staan, maar geen taal hebben– toont dat taal beschouwd wordt als een definiërend kenmerk voor het mens-zijn. Het is bijgevolg ook in bredere maatschappelijke zin belangrijk om een bevredigend antwoord op deze vraag te vinden.

Tot op dit ogenblik is er nog geen consensus over een scenario over het ontstaan en de evolutie van taal. Het simplistische “Big Bang”-scenario (één enkele genetische mutatie) dat lang heeft standgehouden in formeel georiënteerde linguïstische kringen wordt gradueel verfijnd tot een stappenplan dat de individuele cognitieve sprongen beschrijft. Het globale probleem van taalevolutie wordt dus opgesplitst in een aantal subproblemen die elk op zich te onderzoeken zijn.

Computersmodellen zijn een veelbelovende methode om de verschillende voorgestelde stadia te kunnen bestuderen. Ze maken het mogelijk om theorieën over de oorsprong van taal te modelleren in een werkend systeem, en na te gaan of de geponeerde theorie ook effectief verklaart wat zij belooft.

In het artikel zijn drie modellen besproken: het basislexiconmodel, het meerwoord-*naming game* model, en het kleurmodel. De eerste twee tonen in twee fasen aan hoe de organisatie van het lexicon in zijn werk kan gaan in afwezigheid van een overkoepelend mechanisme dat woorden en betekenissen koppelt: een relatief eenvoudig score-mechanisme in het lexicon laat de *agent* toe om statistieken bij te houden over het communicatieve succes van individuele woord-betekenisparen. De *agent* kan die informatie gebruiken om de waarschijnlijkheid dat hij begrepen wordt te maximaliseren. Het meerwoord-*naming game* model laat zien welke de benodigde mechanismen zijn om compositionele expressies te kunnen produceren en decoderen. Het kleurmodel op zijn beurt toont hoe betekenis kan beïnvloed worden door taal. Het geeft een mogelijke verklaring voor hoe perceptuele categorieën en, in het bijzonder hoe kleurcategorieën, gevormd kunnen worden en coherent worden binnen een groep taalgebruikers.

Computersmodellen zijn in staat om een belangrijke bijdrage te leveren aan het onderzoek naar de oorsprong en de evolutie van taal. De voornaamste verdienste van computersmodellen is het concreet maken van theorieën waardoor haar consistentie kan getoetst worden. Op een abstracter niveau, dwingt computersmodelleren dat een theorie ook concreetiseerbaar is. Theorieën die dit niet zijn, zijn moeilijk of helemaal niet verifieerbaar. Het kunnen omzetten van een theorie in een computersmodel vormt het fundament van elke goede theorie.

7. Referenties

Belpaeme, T. 2001 “Simulating the formation of color categories.” In B. Nebel (Ed.), *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI’01)*, 393–398. San Francisco, CA: Morgan Kaufmann.

Belpaeme, T. 2002 *Factors Influencing the Origins of Colour Categories*. PhD thesis, Vrije Universiteit Brussel, Artificial Intelligence Laboratory.

- Belpaeme, T., Steels, L., & Van Looveren, J. 1998 "The construction and acquisition of visual categories." In A. Birk and Y. Demiris (Eds.), *Proceedings of the 6th European Workshop on Learning Robots*, Lecture Notes on Artificial Intelligence. Berlin: Springer.
- Berlin, B. & Kay, P. 1969 *Basic Color Terms: Their Universality and Evolution*. Berkeley, CA: University of California Press.
- Bickerton, D. 1981 *Roots of Language*. Ann Arbor, USA: Karoma.
- Bornstein, M. H. 1985 "On the development of color naming in young children." *Brain and Language*, 26, 72–93.
- Bornstein, M. H. 1987 "Perceptual categories in vision and audition". In S. Harnad (Ed.), *Categorical Perception*, 287–300. Cambridge, MA: Cambridge University Press.
- Boynton, R. M. & Olson, C. X. 1987 "Locating basic colors in the OSA space." *COLOR Research and Application*, 12(2), 94–105.
- Camazine, S., Deneubourg, J., Franks, N., Sneyd, J., Theraulaz, G., & Bonabeau, E. 2001 *Self-Organization in Biological Systems*. Princeton: Princeton University Press.
- Catchpole, C. & Slater, P. 1995. *Bird song: Biological themes and variations*. Cambridge, MA: Cambridge University Press.
- Chomsky, N. 1980 "Rules and representation." *Behavioral and Brain Sciences*, 3, 1-61.
- Chomsky, N. 1995 *The Minimalist Program*. Cambridge, MA: The MIT Press.
- Christiansen, M. H. & Kirby, S. 2003 *Language Evolution*. Oxford, UK: Oxford University Press.
- Davidoff, J., Davies, I., & Roberson, D. 1999 "Colour categories in a stoneage tribe". *Nature*, 398, 203–204.
- Davies, I. and Franklin, A. 2002 "Categorical perception may affect colour pop-out in infants after all." *British Journal of Developmental Psychology*, 20, 185–203.
- Davies, I. R. 1998 "A study of colour grouping in three languages: A test of the linguistic relativity hypothesis." *British Journal of Psychology*, 98, 433–452.
- de Boer, B. 2001 *The Origins of Vowel Systems*. Oxford, UK: Oxford University Press.
- Deacon, T. W. 1997 *The Symbolic Species: the Co-Evolution of Language and the Brain*. New York: W.W. Norton.
- de Saussure, F. 1974. *Course in General Linguistics*. New York: Fontana/Collins. Manuscript uit 1916, vertaald door W. Baskin.
- De Valois, R. L. & De Valois, K. K. 1975 "Neural coding of color." In E.C. Carterette and M.P. Friedman (Eds.), *Handbook of Perception, Volume V: Seeing*, 117–166. New York: Academic Press.
- De Valois, R., Abramov, I., & Jacobs, G. 1966 "Analysis of response patterns of LGN cells." *Journal of the Optical Society of America*, 56(7), 966–977.
- Dubois, D. 2000 "Categories as acts of meaning: the case of categories in olfaction and audition." *Cognitive Science Quarterly*, 1, 33–66.

- Everett, D. 1986. "Piraha," in Derbyshire, D. & Pullum, G. *Handbook of Amazonian languages I*. Mouton De Gruyter, Berlin.
- Gamble, C. 1999 *The Palaeolithic Societies of Europe*. Cambridge, UK: Cambridge University Press.
- Gellatly, A. 1995 "Colourful Whorfian ideas: Linguistic and cultural influences on the perception and cognition of colour, and on the investigation of them." *Mind and Language*, 10(3), 199–225.
- Gerhardstein, P., Renner, P., & Rovee-Collier, C. 1999 "The roles of perceptual and categorical similarity in colour pop-out in infants." *British Journal of Developmental Psychology*, 17, 403–420.
- Gleason, H. 1961 *An Introduction to Descriptive Linguistics*. New York: Holt, Rinehart and Winston.
- Harnad, S. 1987 *Categorical Perception*. Cambridge, MA: Cambridge University Press.
- Jackendoff, R. 2002 *Foundations of Language*. Oxford, UK: Oxford University Press.
- Jameson, K. & D'Andrade, R. 1997 "It's not really red, green, yellow, blue: an inquiry into perceptual color space." In C.L. Hardin and L. Maffi (Eds.), *Color categories in thought and language*, 295–319. Cambridge, MA: Cambridge University Press.
- Jennings, N. & Wooldridge, M. 1998 *Agent Technology: Foundations, Application and Markets*. Berlin, Germany: Springer Verlag.
- Kay, P. & Kempton, W. 1984 "What is the Sapir-Whorf hypothesis?" *American Anthropologist*, 86(1), 65–79.
- Kay, P. & Regier, T. 2003 "Resolving the question of color naming universals." *Proceedings of the National Academy of Sciences*, 100(15), 9085–9089.
- Kay, P., Berlin, B., Maffi, L., & Merrifield, W. R. 2003 *The World Color Survey*. Center for the Study of Language and Information, Stanford.
- Lammens, J. M. 1994 *A Computational Model of Color Perception and Color Naming*. PhD thesis, State University of New York.
- Lenneberg, E. H. & Roberts, J. M. 1956 "The language of experience: A study in methodology." *International Journal of American Linguistics*, memoir 13.
- Lieberman, P. 1998 *Eve Spoke: Human Language and Human Evolution*. New York: W.W. Norton and Company.
- Lightfoot, D. 1999 *The Development of Language: Acquisition, Change and Evolution*. Oxford, UK: Blackwell.
- Lucy, J. A. & Shweder, R. A. 1979 "Whorf and his critics: Linguistic and nonlinguistic influences on color memory." *American Anthropologist*, 81, 581–615.
- Pinker, S. 1994 *The Language Instinct: How the Mind Creates Language*. New York: W. Morrow.
- Quine, W. 1960 *Word and Object*. Cambridge, MA: The MIT Press.

- Roberson, D. 2005 "Color categories are culturally diverse in cognition as well as in language." *Cross-Cultural Research: The Journal of Comparative Social Science*, 39. *Ter perse*.
- Roberson, D., Davies, I., & Davidoff, J. 2000 "Color categories are not universal: replications and new evidence from a stone-age culture." *Journal of Experimental Psychology: General*, 129(3), 369–398.
- Rosch, E. 1973 "Natural categories." *Cognitive Psychology*, 4, 328–350.
- Rosch, E. 1978 "Principles of categorization." In E. Rosch and B. Lloyd (Eds.), *Principles of categorisation, in cognition and categorisation*, 27–48. Hillsdale, NJ: Erlbaum.
- Rosch-Heider, E. 1972 "Universals in color naming and memory." *Journal of Experimental Psychology*, 93, 10–20.
- Ruhlen, M. 1996 *The Origins Of Language: Tracing the Evolution of the Mother Tongue*. Indianapolis, USA: Wiley.
- Sakas, W., Fodor, J. 2001 "The Structural Triggers Learner." In S. Bertolo (Eds.), *Language Acquisition and Learnability*. Cambridge, UK: Cambridge University Press.
- Sapir, E. 1921 *Language: An Introduction to the Study of Speech*. New York: Harcourt, Brace and Co.
- Saunders, B. & van Brakel, J. 1997 "Are there nontrivial constraints on colour categorization?" *Behavioral and Brain Sciences*, 20(2), 167–228.
- Shepard, R. N. 1992 "The perceptual organization of colors: An adaptation to regularities of the terrestrial world?" In J. Barkow, L. Cosmides and J. Tooby (Eds.), *Adapted Mind*, 495–532. Oxford: Oxford University Press.
- Shepard, R. N. 1994 "Perceptual-cognitive universals as reflections of the world." *Psychonomic Bulletin & Review*, 1:2–28. Reprinted in *Behavioral and Brain Sciences*, 24(3).
- Steels, L. 1996 "Self-organizing vocabularies." In C. Langton and T. Shimohara (Eds.), *Proceedings of the Conference on Artificial Life V (Alife V)* (Nara, Japan). Cambridge, MA: The MIT Press.
- Steels, L. 1997 "Construction and sharing perceptual distinctions." In M. van Someren and G. Widmer (Eds.), *Proceedings of the European Conference on Machine Learning*. Berlin: Springer Verlag.
- Steels, L. & Belpaeme, T. 2005 "Coordinating perceptually grounded categories through language. A case study for colour." Verschijnt in *Behavioral and Brain Sciences*.
- Steels, L., Kaplan, F. 1998 "Stochasticity as a source of innovation in language games" In Adami, G. et al, *Proceedings of the Conference on Artificial Life VI (Alife VI)*, Cambridge, USA: MIT Press.
- Steels, L., Kaplan, F., McIntyre, A., Van Looveren, J. 2002 "Crucial Factors in the Origins of Word-Meaning." In A. Wray (Ed.), *The Transition to Language*. Oxford, UK: Oxford University Press.
- Sturges, J. & Whitfield, T. A. 1995 "Locating basic colours in the Munsell space." *COLOR Research and Application*, 20(6), 364–376.

Van Wijk, H. 1959 “A cross-cultural theory of colour and brightness nomenclature.”
Bijdragen tot de taal-, land- en volkenkunde, 115, 113–137.

Whorf, B. L. 1956 *Language, Thought and Reality: Selected Writings of Benjamin Lee Whorf*. Cambridge, MA: The MIT Press. Edited by Carrol, J.B.

Yendrikhovskij, S. N. 2001 “Computing color categories from statistics of natural images.”
Journal of Imaging Science and Technology, 45(5), 409–417.

Affiliatie

Tony Belpaeme is een postdoctoraal onderzoeker van het Fonds voor Wetenschappelijk Onderzoek - Vlaanderen. Joris Van Looveren heeft een specialisatiebeurs van het IWT (Instituut voor de Promotie van Innovatie door Wetenschap en Technologie).