Samenvatting

Het doel van dit proefschrift is *Optimality Theory* (Prince and Smolensky, 1993) zo te implementeren dat we de variatie in spraak kunnen beginnen te verklaren. Het *Simulated Annealing Optimality Theory* algoritme (SA-OT, Fig. 2.8, pagina 64) combineert Optimaliteitstheorie met de zogenaamde *Simulated Annealing*, een wijd verspreid heuristisch optimalisatie-algoritme waarmee we de frequentie van verschillende alternatieven nauwkeurig kunnen modelleren. Verder onderzoek zal in de toekomst moeten uitwijzen of SA-OT op termijn succesvoller is dan de bestaande, stochastische optimaliteitstheoretische modellen.

In Hoofdstuk 1 worden deze concurrerende modellen op een rijtje gezet, en wordt de "filosofische" achtergrond van het proefschrift geschetst. In Hoofdstuk 1 wordt het gebruik van een heuristisch optimalisatiealgoritme – zoals *Simulated Annealing* – verantwoord, en het SA-OT algoritme geïntroduceerd. Tabel 2.1 (pagina 43) formuleert een van de centrale ideeën van dit proefschrift: door het scheiden van het taalkundige model en de implementatie ervan, kunnen taalkundige competentie èn performancefouten verklaard worden. Een adequaat model, zoals een optimaliteitstheoretische grammatica, kan voorspellen welke vormen door de moedertaalspreker als grammaticaal worden beoordeeld. Dit deel van het model komt overeen met de *statische kennis van de taal* in de hersenen van de moedertaalspreker. Daarnaast is er de implementatie van de grammatica, die gezien kan worden als model van het *dynamische taalproductieproces*. Net als de menselijke spraak hoeft de implementatie van de grammatica niet exact (juist) te zijn, maar de door de implementatie gemaakte fouten dienen wel overeen te komen met geobserveerde performancefouten.

SA-OT introduceert een topologie van de verzameling optimaliteitstheoretische kandidaten, namelijk een *nabijheidsstructuur* (Engels: 'neighbourhood structure'). Op deze manier kan het begrip *lokaal optimum* gedefinieerd worden: een lokaal optimum is een kandidaat die een hogere harmonie heeft dan zijn directe omgeving. Het zijn deze lokale optima die, ookal zijn ze niet globale optima, als 'onjuiste' uitvoer geselecteerd kunnen worden, en zo performancefouten modelleren. Het globale optimum komt overeen met de grammaticale vorm.

Hoe gaat dit in zijn werk? SA-OT voert een *random walk* uit op de kandidatenverzameling. Bij elke iteratie kiest de random walker willekeurig vanuit de huidige positie w een naburige kandidaat w' (er is bijvoorbeeld een basisoperatie die kandidaat w in w' verandert), op basis van de *a priori probabilities*. Vervolgens verplaatst de random walker zich naar w' met een waarschijnlijkheid die we de *transition probability* zullen noemen. De *a priori probabilities* blijven constant tijdens de simulatie, maar de *transition probability* neemt af,

als functie van de paramater T ('temperatuur'), tenzij w' een harmonieuzere (of een even zo harmonieuze) kandidaat is dan w, in welk geval de *transition probability* altijd 1 is. Aan het eind van een simulatie staat de random walker altijd in een lokaal optimum dat niet noodzakelijk het globale optimum is. De uitvoer van het algoritme is de eindpositie van de random walker. Vaak – maar, zoals is gebleken, lang niet altijd – is de kans dat het juiste, globale optimum wordt bereikt groter wanneer de temperatuur T langzamer afneemt, en er meer iteraties worden uitgevoerd.

Het combineren van Simulated Annealing en Optimaliteitstheorie is niet triviaal. De *targetfuncties* die geoptimaliseerd worden in de traditionele Simulated Annealing hebben reeële getallen als functiewaarden, maar de Harmoniefunctie uit de Optimaliteitstheorie niet. De oplossing voor dit probleem wordt informeel geïntroduceerd in sectie 2.2, en wiskundig uitgewerkt in hoofdstuk 3. Dit hoofdstuk begint met een formele definitie van Optimaliteitstheorie, gevolgd door een bespreking van twee manieren om de Harmoniefunctie te formuleren: door middel van polynomen, of met ordinale getallen. Beide benaderingen leiden tot dezelfde combinatie van Optimaliteitstheorie met Simulated Annealing.

In het daaropvolgende hoofdstuk speculeer ik over twee taalkundige 'hot topics': het mentale lexicon (de woordenschat), en het leren van een grammatica. Naast vragen die ik in dit hoofdstuk en deze dissertatie verder onbeantwoord zal laten, introduceer ik ook een nieuwe formele definitie van de zogenaamde *Output-Output Correspondence*, of liever *Constutituent-Output Correspondence*, die in het volgende hoofdstuk gebruikt zal worden.

De laatste drie hoofdstukken laten zien hoe een aantal concrete, taalkundige verschijnselen aangepakt kunnen worden met behulp van SA-OT. Hier ligt de nadruk dan ook niet op de fonologische details van de analyses, maar op de methodologie van het toepassen van SA-OT. De besproken modellen laten het belang van de modelparameters zien in het algoritme. Daarnaast worden er enkele technieken en trucs geïntroduceerd en verschillende experimenten die men kan uitvoeren met de modellen, en wordt besproken hoe de uitkomsten van deze experimenten begrepen moeten worden. Sectie 8.1 vat de verschillende methodologische observaties nog eens uitvoerig samen.

Metrische klemtoon in het Nederlands dient in hoofdstuk 5 als voorbeeld van hoe spreeksnelheid met behulp van het algoritme gemodelleerd kan worden. Bij een langzamere simulatie (met meer iteraties, met een geleidelijker *cooling schedule*) veranderen de relatieve frequenties van de door het model geproduceerde vormen op dezelfde wijze als de door een mens geproduceerde vormen bij het veranderen van de spreeksnelheid in een laboratoriumexperiment.

Hoofdstuk 6 gaat over stemassimilatie, en de methodologische kwesties uit dit hoofdstuk gaan terug op enkele problemen uit sectie 2.3. De 'landschappen', dat wil zeggen de topologieën met de Harmoniefunctie, zijn eenvoudiger in deze modellen dan die van de modellen in hoofdstuk 5, waardoor een analytische bespreking van het gedrag van de modellen mogelijk is. Verder bewijzen deze modellen dat – in tegenstelling tot traditionele Simulated Annealing – een hoger aantal iteraties niet altijd leidt tot een grotere kans het globale optimum, de beste kandidaat, te vinden. Als SA-OT daadwerkelijk een adequaat model van de spraakproductie is, dan suggereert dit dat we onregelmatige vormen ook met een eenvoudige grammatica kunnen uitleggen. Namelijk, de onregelmatige vormen zijn dan de 'foutjes' die het dynamische taalproductieprocess onherroeplijk maakt.

In hoofdstuk 7 passeren twee verschijnselen de revue, die met woord- en lettergreepstructuur te maken hebben. Eerst bestudeer ik de clitisering van het Hongaarse lidwoord, afhankelijk van het spreektempo en de gekozen allomorf. Daarna implementeer ik met SA-OT het klassieke paradigma van lettergreepstructuur van Prince en Smolensky: *Basic CV Theory*. De topologie in het model van het Hongaarse lidwoord heeft een algemene structuur die overeenkomt met een van de topologieën die in hoofdstuk 6 gebruikt zijn. De topologie die gebruikt is bij het implementeren van Basic CV Theory is een verdere ontwikkeling hiervan. Samengenomen hebben we in hoofdstukken 6 en 7 verschillende aspecten van modellen met een vergelijkbare topologie bestudeerd. Ik vermoed dat deze 'familie' van modellen in toekomstig werk ook prominent aanwezig zal zijn.

In het achtste en laatste hoofdstuk worden de bevindingen samengevat, en SA-OT vergeleken met andere modellen van variatie in taal. Tot slot wordt SA-OT geplaatst in de context waarin de Optimaliteitstheorie meer dan een decennium geleden ontstond: de cognitiewetenschap.