

# Bachelorproject

## Agent-gebaseerd model over de verandering van taal

Pieter de Bie [1513168]

Begeleider: Bart de Boer

Rijksuniversiteit Groningen

6 februari 2007

### Samenvatting

Onderzoek dat zich bezighoudt met het ontstaan van taaldiversiteit maakt gebruik van modellen waarmee het lastig is taalpatronen te vinden die wij in de echte wereld zien, bijvoorbeeld het bestaan van zowel geleidelijke taalovergangen, zoals we die tussen Nederland en Duitsland zien, als het vóórkomen van harde taalbarrières, zoals bijvoorbeeld de Nederlands-Frans-lijn in België.

Geïnspireerd door resultaten in de reactiediffusietheorie, waar met behulp van simpele differentiaalvergelijkingen een variëteit aan ruimtelijk patronen kan ontstaan is er in dit project een spatiaal, agentgebaseerd model ontwikkeld waarin taaldiversiteit ontstaat door het introduceren van taalmutaties. Agenten kunnen deze mutaties overbrengen via communicatie, waarbij gebruik gemaakt is van de Social Impact Theory. Deze stelt dat weinig voorkomende mutaties een relatief grote invloed hebben.

Met dit model is het inderdaad mogelijk om verschillende taalgrenzen tegelijk te laten ontstaan. Met experimenten is bepaald aan welke voorwaarden de parameterinstelling moet voldoen om verschillende taalbarrières te laten ontstaan.

## 1 Inleiding

Binnen de taalkunde wordt er veel onderzoek gedaan naar het ontstaan en de evolutie van taal (Wang & Minett, 2005). Met de opkomst van computers worden er ook computationele modellen gebruikt die helpen linguïstische gebeurtenissen te verklaren. Een van de onderzoeken bestudeert hoe taalveranderingen kunnen ontstaan binnen een wereld waar taal homogeen verdeeld is. Een belangrijk gereedschap hiervoor zijn agentgebaseerde modellen. Het voordeel van deze modellen is dat zij schijnbaar simpele regels volgen, maar toch complex gedrag kunnen vertonen. Op deze manier valt een ingewikkeld verschijnsel dus simpel te verklaren. Hoewel deze modellen verschillende taalgedragingen kunnen laten zien, zoals het behoud of ontstaan van diversiteit, zijn er ook fenomenen die wij in de echte wereld zien, die nog weinig bestudeerd worden.

Een opmerkelijk verschijnsel is dat er vaak geleidelijke overgangen tussen verschillende taalgebieden zijn, waarbij er verschillende dialecten bestaan in het grensgebied. Een voorbeeld hiervan is de overgang tussen Nederlands en Duits in het oosten van Nederland. Naast deze geleidelijke overgangen worden er echter ook harde taalgrenzen gevonden, waarbij een overgangszone ontbreekt en bewoners rond de grenslijn dus slecht kunnen communiceren met een deel van zijn nabije omgeving. Een voorbeeld hiervan is de Frans-Nederlands taalgrens in België (Willemys, 2002). Vaak worden deze taalbarrières echter verklaard door sterke geografische of culturele verschillen (Livingstone, 2002). Sterke taalgrenzen veroorzaakt door geografische barrières zijn al succesvol gemodelleerd, bijvoorbeeld door Patriarca & Leppanen (2004).

Het is echter ook interessant of deze verschillende taalgrenzen ook kunnen ontstaan zonder het introdu-

ceren van geografische of culturele eigenschappen, die een specifiek deel van de populatie beïnvloedt. Deze modellen gaan vaak uit van functionele selectie, of van culturele selectie die niet gebonden is aan een specifieke populatie (Livingstone, 2002, Nettle, 1999a,b). Het slechts introduceren van taalmutaties is niet genoeg voor het ontstaan van taaldiversiteit. Dit staat in contrast met natuurlijke selectie, waar willekeurige mutatie en fitness vanzelf kunnen leiden tot het doorvoeren van de mutatie in een hele populatie. Dit verschil ontstaat omdat, hoewel het heel verleidelijk is taalverandering te vergelijken met genetische mutatie, er een aantal verschillen zijn.

### 1.1 De averaging en thresholding problemen

Een belangrijk verschil tussen biologische en linguïstische mutatie is de manier waarop een mutatie wordt doorgegeven. Bij biologische evolutie wordt een mutatie, zelfs als deze geen of slechts heel weinig invloed heeft, met 50% kans (bij twee ouders) doorgegeven aan kinderen. Bij linguïstische evolutie is deze overdracht echter niet zo vanzelfsprekend. Taalacquisitie is een sociaal proces, waarbij niet alleen de taal van de ouders wordt overgenomen, maar deze ook wordt beïnvloed door de hele sociale omgeving. Een optimale taalverwervingsstrategie is dan om de taaleigenschappen van zoveel mogelijk mensen over te nemen, zodat de gemiddelde fout het laagst is. Deze strategie heeft echter twee problemen.

Het eerste probleem wordt door Nettle het *averaging problem* genoemd (Nettle, 1999a), en houdt in dat kleine, willekeurige mutaties uiteindelijk optellen tot tot nul, waardoor er geen verandering optreedt. Dit kan bijvoorbeeld worden gezien bij de uitspraak van de /e/, waarbij elk persoon een willekeurige variatie in de frequentie van een formant heeft. Als deze variaties gelijk zijn verdeeld, zal het gemiddelde altijd hetzelfde blijven, waardoor er geen taalvariatie doorgevoerd kan worden.

Het tweede probleem bevindt zich bij het introduceren van discrete taaleigenschappen, bijvoorbeeld de woordvolgorde in een zin, en wordt door Nettle het *threshold problem* genoemd. Als er twee varianten van woordvolgorde mogelijk zijn, zal er altijd een zijn die de meerderheid van de sprekers gebruikt. Daarom zal het voor een nieuwe spreker het efficiëntst zijn om deze variant aan te leren, waardoor de mindere variant geen kans krijgt.

### 1.2 Functionele en Sociale selectie

Om deze twee problemen op te lossen moet er een selectieprocedure plaatsvinden, die bepaalde mutaties prefereert boven andere. Traditioneel zijn twee verschillende soorten selectie veelgebruikt voor het invoeren van mutaties. Dit zijn de functionele en de sociale selectie (Nettle, 1999a).

Functionele selectie houdt in dat een agent een bepaalde mutatie prefereert boven een andere omdat dit voordeel oplevert. Dit kan zijn dat een bepaalde klank of woord makkelijker is uit te spreken, makkelijker te verstaan of beter te onthouden is. Door deze voorkeur voor een variatie kan een minderheidsvariant de overhand krijgen.

Sociale selectie is een andere manier van selectie. Hierbij wordt een agent beïnvloed door andere agents in zijn omgeving. Hoe dichterbij een agent is, en hoe meer totale invloed hij heeft, hoe meer invloed hij uitoefent op een specifieke agent. Een agent kan ook de keuze maken om tot een bepaalde groep te behoren en zijn taalgebruik aanpassen aan leden binnen die groep. Deze groepen kunnen dan onafhankelijk van andere taalsprekers muteren, waardoor er een taalverschil kan ontstaan tussen de twee groepen.

## 2 Literatuuroverzicht

Er is al een aantal eerdere onderzoeken gedaan naar de divergentie van taal. Veel van deze onderzoeken maken ook gebruik van computersimulaties om theorieën te testen. Zo heeft Axelrod (1997) een model gebruikt waarin verschillende agenten met elkaar communiceren en daarmee cultuur, waaronder taal, uitwisselen. In zijn model heeft elke agent een lijst met culturele eigenschappen. Elke eigenschap kan een bepaalde discrete waarde hebben, de totale hoeveelheid mogelijke waarden kan worden ingesteld, net als het totaal aantal eigenschappen. De agenten communiceren met elkaar aan de hand van de gelijkheid van deze eigenschappen: de communicatiekans neemt lineair af met het verschil in eigenschappen. Als er communicatie optreedt, kan één van de agenten een eigenschap van een andere agent overnemen. Aangezien communicatie waarschijnlijker is als er meer eigenschappen overeenkomen, komt Axelrod in zijn simulaties ook situaties tegen waarin een hele groep wordt uitgesloten van een andere groep – zij hebben geen manier meer om te communiceren aangezien alle eigenschappen anders zijn. Het hebben van meer eigenschappen zorgt voor meer communicatie, aan-

gezien er meer gronden zijn waarop een agent kan communiceren. Meer mogelijkheden per eigenschap maakt het juist minder waarschijnlijk dat agenten communiceren. De grootte van de ruimte heeft ook invloed: met te weinig individuen zullen er slechts weinig cultureel onafhankelijke gebieden ontstaan, bij te veel individuen ontstaan er ook slechts weinig gebieden, omdat er genoeg ruimte is om deze op te laten lossen. Axelrod's manier van taalrepresentatie is interessant vanwege de simpliciteit en complexiteit die ermee valt te bereiken.

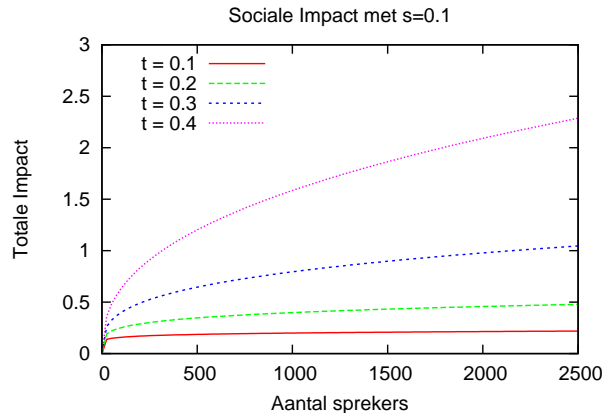
Barr (2004) heeft onderzoek gedaan naar de convergentie van taal, zonder gebruik te maken van een centrale taalrepresentatie. In zijn model hebben agenten alleen onderlinge interactie. Alle individuen beginnen met een willekeurige koppeling van betekenissen en woorden. Door met elkaar te communiceren proberen deze agenten betekenis en woorden aan elkaar te koppelen. Hij laat zien dat het mogelijk is naar een enkele koppeling te convergeren, zonder dat er gebruikgemaakt wordt van een specifieke taalstrategie of globale taalrepresentatie. Een interessant bijverschijnsel van zijn model is dat in sommige situaties er meerdere dialecten kunnen ontstaan, waarbij agenten zich ook ruimtelijk groeperen op basis van hun taal.

Livingstone (2002) heeft met behulp van agenten, gemodelleerd als neurale netwerken, onderzoek gedaan naar het ontstaan van taaldiversiteit. In zijn model, waarbij agenten zich in een rij bevinden, is de communicatiekans afhankelijk van de afstand tussen agenten. Verder heeft hij agentsterfte in zijn model opgenomen, waarbij een agent vervangen wordt na 5 tijdsstappen en agenten alleen kunnen leren in de eerste tijdsstap. In zijn model ontstaat een geleidelijke dialectvorming, waarbij de taalrepresentatie van elke agent dicht bij die van zijn burens ligt, maar de uiteinden van de rijen (die niet met elkaar zijn verbonden) erg verschillende taalrepresentaties hebben. Hij maakt echter gebruik van een hele simpele taalrepresentatie, waarbij niet meer dan 8 verschillende talen mogelijk zijn. Om zijn onderzoek te ondersteunen, heeft hij een reeds bestaand model gemodificeerd en soortgelijke experimenten gedaan. Het gaat hier om een model van de Boer (1997), waarbij agenten verschillende klinkers hebben die zijn gemodelleerd als een eerste en tweede formantfrequentie. In het originele model van de Boer convergeren agenten naar een enkel klinkersysteem. Met de aanpassingen van Livingstone, waarbij de agenten weer in een niet-continue rij worden geplaatst, ontstaan ook dialecten,

waarbij de klinkerlocaties verplaatsen naarmate men verder in de rij van agenten komt, op eenzelfde manier als in het eerdere model. Livingstone laat dus zien dat taaldiversiteit te behouden is zonder expliciete sociale selectie.

Nettle (1999a) heeft ook onderzoek gedaan naar klinkersystemen. Zijn model bestaat uit groepen van agenten die allemaal een eigen klinkersysteem (8 klinkers met elk 2 formanten) hebben. Elk experiment begint met 64 groepen die allemaal hetzelfde klinkersysteem hebben. Binnen deze groepen van 20 individuen sterven individuen en worden er nieuwe geboren, die door willekeurige ruis een klinkersysteem krijgen dat verschilt van de rest van de groep. Elk individu leeft 5 tijdsstappen en kan alleen in de eerste tijdsstap leren van andere individuen in de groep. Door de kleine groeps grootte is het mogelijk verschillende groepen te creëren die significant van elkaar verschillen. Wordt er echter groepscommunicatie toegevoegd aan het model, sterft deze variatie uit en is er geen significant verschil meer. Pas na het introduceren van superindividuen, willekeurige individuen die na geboorte meer invloed krijgen dan de rest, kunnen er weer significante verschillen ontstaan. Het toevoegen van functionele selectie, in de vorm van het "wegduwen" van een formant als deze te dicht bij een andere formant komt, leidt niet tot significante verbeteringen.

Nettle heeft nog een tweede model gemaakt om taaldiversiteit te bestuderen (Nettle, 1999b). In dit model zijn er slechts twee taalvarianten,  $p$  en  $q$ . Nettle introduceert weer leeftijd in zijn model, waardoor alleen jonge agenten taal kunnen overnemen en oude agenten sterven om plaats te maken voor nieuwe. Nettle gebruikt verschillende manieren om zijn model te beïnvloeden. Met een weging waardoor kleinere populaties van een taal relatief zwaarder worden meegeteld, krijgt hij simulaties waarin de twee taaldialecten relatief stabiel naast elkaar kunnen bestaan. Hierbij maakt hij gebruik van de Social Impact Theory. Als hij sociale factoren modelleert, waarbij sommige individuen invloedrijker zijn dan anderen, krijgt hij ook situaties waarin het minderheidsdialect de overhand in de populatie kan krijgen. Naast deze vorm van sociale selectie experimenteert hij ook met functionele selectie, door de ene taalvariant een constante bonus te geven boven de andere. Deze soort bias blijkt echter niet in staat te zijn zijn model realistisch te beïnvloeden. Omdat hij taal echter heel simpel heeft gemodelleerd, met slechts twee taalvarianten, zijn er geen grensovergangen te onderscheiden



**Figuur 1:** Sociale impact afhankelijk van aantal sprekers, voor verschillende waarden van  $t$

in zijn model. Nettle's werk geeft informatie over de effectiviteit van functionele selectie en een gereedschap om sociale steun te modelleren in de vorm van de Social Impact Theory.

## 2.1 De Social Impact Theory

Latané (1981) heeft onderzocht hoe populatiegrootte en *sociale impact* zich relateren. Social impact heeft in dit geval een brede definitie, en beslaat alles waar mensen anderen beïnvloeden. Latané gaat specifiek in op de relatie tussen groeps grootte en total impact. Voor deze relatie heeft hij een formule opgesteld aan de hand van zijn en andere ervaringen, en deze getoetst aan verschillende situaties waarin sociale impact een rol speelt. Zijn formule is:

$$I = sN^t \quad (1)$$

Hierbij staat  $I$  voor de totale sociale indruk, staat  $N$  voor de grootte van de groep die die invloed uitoefent, en zijn  $s$  en  $t$  twee constanten. Deze formule is voor verschillende waarden van  $t$  uitgezet in figuur 1.

Deze formule heeft hij getoetst aan een aantal verschillende situaties, waaronder groepsdruk bij meningsvorming, zenuwachtigheid tegen publieks grootte en waargenomen belangrijkheid van een nieuwsbericht waarin groepen van verschillende grootte het onderwerp waren. Met behulp van de Social Impact Theory is het mogelijk de overdrachtskans van verschillende taaleigenschappen te berekenen.

## 3 Probleemstelling

In dit artikel wordt onderzoek gedaan naar de volgende vraag: Onder welke voorwaarden kunnen verschillende soorten taalgrenzen binnen één taalmodel ontstaan? Deze vraag wordt onderzocht met behulp van een computermodel, waarbij verschillende parameters worden gevarieerd.

## 4 Het Model

Het hier gepresenteerde model is geïnspireerd door verschillende van de eerder besproken modellen. Het doel van het onderzoek is het onderzoeken van het ontstaan van taalgrenzen. Een gevolg hiervan is dat het belangrijk is dat er een goede spatiële representatie is. Er is gekozen voor een rechthoekige 2-dimensionale omgeving, waarbij alle zijden even lang zijn. Het model heeft een totale populatie van  $N$  agenten. De omgeving is opgedeeld in een raster, waarbij elke agent maximaal vier directe burens heeft; agenten langs de grens van het gebied hebben minder burens. Deze ruimte representeert fysieke afstand, en omvat niet de sociale structuur zoals in Nettle's model, aangezien ons model ervan uitgaat dat sociale werkingen tussen agenten geen invloed hebben op het ontstaan van taaldiversiteit. Er is gekozen voor een niet-torodiale ruimte zodat ook patronen die voorkomen door beperkte ruimte onderzocht kunnen worden.

Een ander gevolg van het onderzoeksdoel is dat het belangrijk is dat de taal in het model redelijk complex kan worden, aangezien taalgrenzen alleen kunnen ontstaan in een model met voldoende mogelijkheid tot taaldiversiteit. In Nettle's model, met slechts een enkele discrete variabele, is het voor een agent alleen mogelijk om tot één van twee taalgroepen te behoren. Deze representatie laat weinig ruimte voor bijvoorbeeld geleidelijke taalovergangen. Daarom is in het model hier gekozen voor een taalrepresentatie vergelijkbaar met Axelrod's model: taal is gerepresenteerd als een vector met een van te voren vastgesteld aantal taaleigenschappen. Een taaleigenschap kan twee verschillende varianten hebben. Voor het gemak bij berekeningen krijgen deze de waarden 0 en 1. Over de precieze functie van de taaleigenschappen wordt geen uitspraak gedaan; ze zijn vergelijkbaar met Nettle (1999a)'s *linguistic items*, een stuk structuur dat onafhankelijk kan worden geleerd en daarom kan worden overgedragen van de ene individu op de andere. Woorden zijn voordehandliggende linguisti-

sche items, maar geluiden en zinsstructuren vallen ook onder deze definitie.

Verder interessant aan Nettle’s model is het gebruik van de Social Impact Theory. Deze stelt dat taaleigenschappen die nog weinig sprekers hebben relatief meer invloed per spreker opleveren. Deze druk om een taaleigenschap voort te zetten kan de drijfveer zijn achter het ontstaan van taaldiversiteit. Hoewel Nettle’s model een enkele variabele gebruikt voor sociale motivatie, wordt de SIT in dit model toegepast op elke eigenschap. Hoewel er dus geen sociale bias tussen de sprekers binnen de wereld is, bestaat er wel een soort van sociale druk tussen verschillende taaleigenschappen.

## 4.1 Het agentproces

Elke beurt doorloopt een agent verschillende processen. Deze vallen op te delen in vier processen: agentsterfte, communicatie, eigenschapsoverdracht en mutatie. Communicatie en eigenschapsoverdracht zijn sterk aan elkaar gerelateerd, de andere processen staan wat losser van elkaar. De drie goed te onderscheiden processen zullen dan ook los besproken worden.

## 4.2 Agentsterfte

Een agent heeft ook een leeftijd. De leeftijd is gelijk aan het aantal cycles dat een agent heeft geleefd. Aan het begin van het model heeft elke agent een willekeurige leeftijd. Een agent kan alleen taalmutaties overnemen van een andere agent, als hij jonger is dan een vooraf gestelde grens (`max_learn_age`). Agenten ouder dan die leeftijd zullen dan ook niet proberen te communiceren. Als een agent ouder wordt dan `max_age`, heeft hij elke cycle een 50% kans om te overlijden. Er is hier niet voor een strikte dood gekozen, omdat een agent anders altijd even oud is relatief van zijn burens. Met de hier ingevoerde overlijdenskans ontstaat een dynamische wereld waarin agenten steeds burens met andere leeftijden hebben.

Als een agent overlijdt, komt er een nieuwe agent voor hem in de plaats. Deze neemt de meestvoorkomende taalvector over van zijn burens. Dit betekent dat een kind leert van zijn omgeving, en alleen de meestgesproken taal aanleert. Een alternatieve keus is dat voor elke mutatie de meestvoorkomende vorm wordt gekozen. Dit zou betekenen dat kinderen in grensgevallen een mengelmoesje van beide talen leert, waardoor het met niemand succesvol kan praten, iets wat niet realistisch is.

---

## Algorithm 1 Communicatie tussen twee agents

---

```

for  $\forall a \in N$  do {for all agents}
  for  $\forall b \in \text{Neighbourhood}(a)$  do
    if  $\text{random} \leq I(a,b)$  then
       $i \leftarrow$  random vector index
      if  $a_i \neq b_i$  then
        if  $\text{random} \leq O(i, b_i)$  then
           $a_i \leftarrow b_i$ 
        end if
      end if
    end if
  end for
end for

```

---

## 4.3 Interactie en overdracht

Taalmutatie vindt plaats door interactie van de agent met zijn directe burens. Voor elke buur wordt gekeken of er gecommuniceerd wordt met behulp van de interactiekans. Lukt dit, dan wordt er een willekeurige waarde in de taalvector vergeleken tussen de twee individuen. Is deze verschillend, dan heeft de agent een kans om de waarde van zijn buurman over te nemen aan de hand van een bepaalde kans, de overdrachtskans. Dit hele proces is ook uitgezet in algoritme 1. Hierin is  $I(a, b)$  de interactiekans tussen twee agents  $a$  en  $b$ , is  $O(i, a_i)$  de overdrachtskans van taaleigenschap  $i$  en is  $a_i$  de variant van eigenschap  $i$  zoals gesproken door agent  $a$ .

Elk van deze stappen zullen los uitgebreider besproken worden.

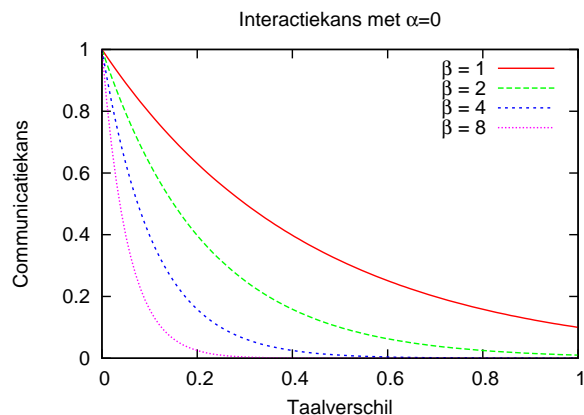
### 4.3.1 Interactie

Elke beurt probeert elke agent met al zijn burens te communiceren, mits hij jonger is dan `max_learn_age`. Om te bepalen of de communicatie slaagt, wordt er een interactiekans uitgerekend. Deze wordt als volgt berekend:

$$I(a, b) = \alpha + (1 - \alpha) \cdot 10^{-d_t(a,b)\beta} \quad (2)$$

Hierbij is  $\alpha$  een minimale interactiekans en  $\beta$  een parameter die de stijlheid van de functie beïnvloedt. Verschillende waarden voor  $\beta$  zijn te zien in figuur 2.  $d_t(a, b)$  is het verschil in taal tussen  $a$  en  $b$ . Deze wordt uitgerekend als de genormaliseerde Hammingafstand:

$$d_t(\mathbf{a}, \mathbf{b}) = \frac{1}{M} \sum_{i=1}^M |a_i - b_i| \quad (3)$$



**Figuur 2:** Verschillende waarden voor  $\beta$  in de interactiekans

Waarbij  $a$  en  $b$  twee taalvectoren zijn van agents, en  $M$  de grootte van de taalvector is. We gaan er hier van uit dat de twee varianten van een taaleigenschap de waarden 0 en 1 hebben. Twee agents met dezelfde taalvector hebben dan een afstand van 0, terwijl twee agents die geen enkele gemeenschappelijke mutatie hebben een afstand van 1 hebben.

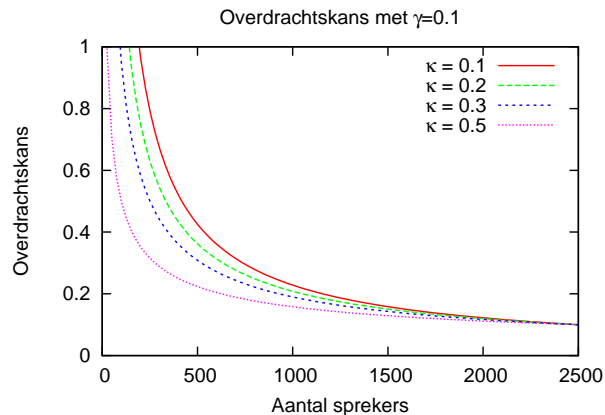
### 4.3.2 Overdracht

Voor het overgeven van een taaleigenschap van de ene agent naar de andere wordt een overdrachtskans berekend. Deze is gebaseerd op de Social Impact Theory zoals beschreven in sectie 2.1. Deze overdrachtskans is dus afhankelijk van de hoeveelheid agents die al een bepaalde taaleigenschap spreken.

Als een agent daadwerkelijk gecommuniceerd heeft, wordt een willekeurige eigenschap uit de taalvector gekozen. Verschilt de variant van deze eigenschap van de agent met die van zijn gesprekspartner, dan kan de communicerende agent de taalvariant van de andere overnemen. Of dit gebeurt wordt bepaald aan de hand van de overdrachtskans. Deze kans wordt als volgt berekend:

$$O(i, v) = \gamma \left( \frac{N_{i,v}}{N} \right)^{\kappa-1} \quad (4)$$

Hierbij is  $\gamma$  een minimale overdrachtskans en is  $\kappa$  een schalingsconstante vergelijkbaar met  $t$  in de originele SIT-formule (zie (1)).  $N_{i,v}$  is de totale hoeveelheid agents die van eigenschap  $i$  variant  $v$  spreken. Omdat we echter niet de totale invloed op een agent willen berekenen, maar de invloed per agent, delen we door de fractie van agents die deze taal reeds



**Figuur 3:** Mutatiekans aan de hand van aantal agents die mutatie spreken

spreekt,  $\frac{N_{i,v}}{N}$ . Dit komt terug in de formule door de  $\kappa - 1$  kwadraatsparameter. Door  $\kappa$  te variëren kan de stijfheid van de curve worden aangepast. Dit is te zien in figuur 3, waar de mutatiefunctie is uitgezet voor verschillende waarden van  $\kappa$ , voor een totale populatiegrootte van 2500.

Belangrijk hierbij is dat de overdrachtskans alleen afhankelijk is van het aantal sprekers van een bepaalde variant. Hoewel er dus wel sociale selectie in het model zit, is deze niet expliciet gemodelleerd tussen twee agents, maar is er sprake van meerdere globale selectieprocessen.

## 4.4 Mutatie

Taalmutaties worden willekeurig ingevoerd in het model. Elke agent heeft elke cycle een kans van `mutation_chance` om een nieuwe taalmutatie aan te nemen. Dit is onafhankelijk van de leeftijd van een agent. Het aannemen van een taalmutatie houdt in dat van een eigenschap in de taalvector, de alternatieve variant wordt gekozen. We kunnen dit “flippen” opschrijven als:

$$a_i \leftarrow 1 - a_i \quad (5)$$

## 5 Implementatie

Het model zoals hier gepresenteerd is ontwikkeld met behulp van de agentmodelleromgeving *Repast*, aan de hand van het overzicht in (Railsback et al., 2006).

## 5.1 Initialisatie

Aan het begin van een run wordt elke agent met een random leeftijd geïnstantieerd. Omdat we beginnen met een homogene taalsituatie, spreken alle agenten dezelfde taal. Dit betekent dat zij allemaal beginnen met dezelfde taalvector.

## 5.2 Communicatie

Voor het communiceren met de burens wordt een vector opgesteld met burens. Deze vector wordt willekeurig ingedeeld voordat communicatie plaatsvindt; dit betekent dat de agenten geen vast patroon gebruiken voor communicatie. Ook het kiezen van een agent die moet communiceren gebeurt willekeurig.

## 5.3 Overig

Voor het genereren van willekeurige getallen is gebruik gemaakt van de Mersenne Twister. De specifieke implementatie is MT19937 (Matsumoto & Nishimura, 1998). Met deze generator worden willekeurige, uniform verdeelde getallen getrokken in het bereik (0, 1).

# 6 Meetmethoden

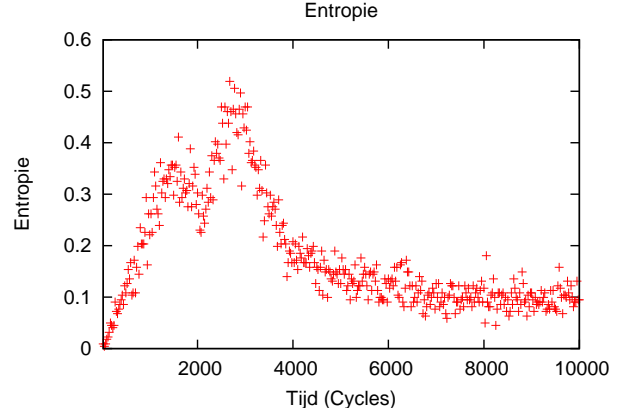
De grote zoekruimte die het systeem moet doorlopen maakt het essentieel betrouwbare meetmethoden ter beschikking te hebben om verzamelde gegevens te analyseren. De data uit het model zijn gebruikt om de entropie, de covariantie en de verdeling van taalgrenzen te berekenen. Verder is er ook een clustering-algoritme geïmplementeerd om agenten te groeperen op basis van hun fysieke en linguïstische afstand.

## 6.1 Entropie

Entropie is een maat die aangeeft hoeveel informatie er in een set gegevens zit (Shannon, 1948). Deze entropie wordt als volgt uitgerekend:

$$H = - \sum_{i=0}^{N_t} p_i \log p_i \quad (6)$$

Om te kijken hoe complex situaties in het model zijn, kan de entropie van alle verschillende talen worden uitgerekend. Hierbij stelt  $N_t$  het totaal aantal talen voor, en  $p_i = \frac{n(i)}{N_t}$ , waarbij  $n(i)$  het aantal sprekers van taal  $i$  is. Entropie is maximaal als alle talen evenveel sprekers hebben. Hoe meer talen er zijn, hoe



*Figuur 4: Verloop van entropie in de tijd*

hoger de entropie. Zijn er slechts enkele talen die veel sprekers hebben, en een groot aantal talen met weinig sprekers, dan is de entropie dus laag.

Met deze maat is te bepalen welke situaties interessant zijn. Ook kunnen we hiermee verschillende situaties vergelijking op basis van complexiteit. Verder kunnen we snel zien hoe een situatie verandert in de tijd: of hij stabiel wordt, of juist wegvalt. Een voorbeeld van het verloop van de entropie, uitgezet tegen de tijd, voor een enkele run is te zien in figuur 4.

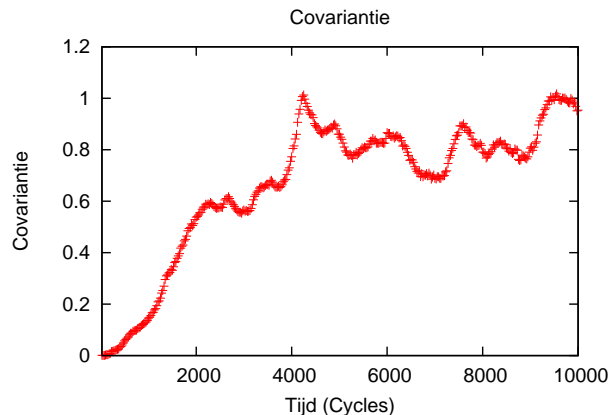
## 6.2 Covariantie

Covariantie is een maat die de correlatie tussen twee variabelen berekent (Kreyszig, 1993). Als twee variabelen geneigd zijn tegelijk te veranderen, is er een positieve covariantie. Als een variabele juist tegen een andere variabele in verandert, is er sprake van een negatieve covariantie. Twee variabelen die compleet onafhankelijk van elkaar zijn hebben een covariantie van 0. Covariantie wordt als volgt berekend:

$$cov(X, Y) = \sum_{i=0}^n \sum_{j=0}^n \frac{(X_i - \mu)(Y_j - v)}{n} \quad (7)$$

waarbij  $\mu$  de gemiddelde waarde voor  $X$  is en  $v$  het gemiddelde van  $Y$ .  $X$  en  $Y$  hebben beiden evenveel ( $n$ ) metingen.

Covariantie is in het model toegepast om de correlatie tussen afstand in de fysieke ruimte en afstand in de taalruimte te bepalen. Deze correlatie is als volgt te berekenen:



**Figuur 5:** Verloop van covariantie in de tijd

$$\sum_{i=0}^N \sum_{j=0}^N \frac{(d_m(a_i, a_j) - \mu)(d_t(a_i, a_j) - v)}{n} \quad (8)$$

Waarbij  $N$  het totaal aantal agenten,  $d_m(a_i, a_j)$  de manhattan distance tussen agenten  $a_i$  en  $a_j$  is,  $d_t(a_i, a_j)$  de Hamming-taalafstand is zoals gegeven in vergelijking (3),  $\mu$  de gemiddelde fysieke (manhattan-)afstand tussen twee agenten en  $v$  de gemiddelde taalafstand tussen twee agenten is. Manhattan distance wordt uitgerekend als de optelling van de absolute verschillen in  $x$  en  $y$ -coördinaten tussen twee agenten:

$$d_m(a, b) = |x_a - x_b| + |y_a - y_b| \quad (9)$$

Hiermee kunnen we de relatie tussen fysieke afstand en verschil in taal uitrekenen; Als agenten een positieve covariantie hebben, verschillen zij veel van taal naarmate ze verder uit elkaar liggen. Agenten met een negatieve covariantie vertonen juist meer overeenkomsten als ze verder uit elkaar liggen. Een positieve covariantie geeft ons informatie over of het systeem een soort van patroon volgt, maar kan ons niet vertellen wat voor patronen we in een situatie kunnen verwachten. Zo zullen zowel geleidelijke grensovergangen als verschillende groepen agenten, met elk een eigen taal, een positieve correlatie opleveren. Een voorbeeld van deze covariantie, uitgezet tegen de tijd, voor dezelfde simulatie als bij de entropie is te vinden in figuur 5.

### 6.3 Clustering

Voor het detecteren van groepen sprekers met dezelfde taal is gebruikgemaakt van een clusteringal-

goritme. Er is gekozen voor een implementatie gebaseerd op het Leader-Follower-clusteringalgoritme (Duda et al., 2001) vanwege de snelheid van dit algoritme en de relatieve simpliciteit.

Agenten worden geclusterd op zowel hun fysieke locatie als hun taal. Omdat deze twee eigenschappen niet dezelfde eenheden hebben, is er gekozen voor een schalingsfactor. Om de afstand tussen twee punten te berekenen, is dan ook de volgende formule gebruikt:

$$d_{lf}(a, b) = \frac{d_m(a, b)}{2\sqrt{N}} + \psi d_t(a, b) \quad (10)$$

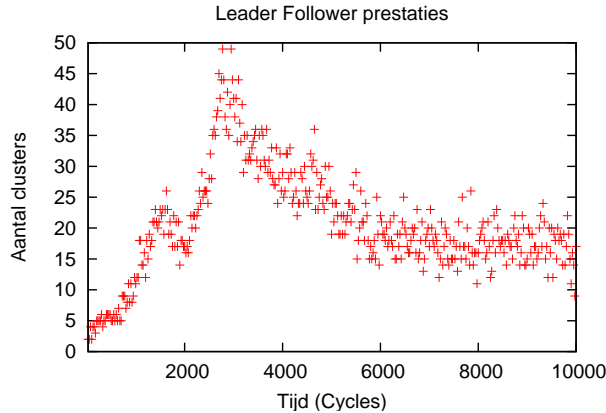
Waarbij  $d_t(a, b)$  het verschil in taalvectoren aangeeft zoals in formule (3) en  $d_m(a, b)$  de manhattan-distance in de fysieke ruimte van twee agenten is, zoals gegeven in formule (9).  $\psi$  is de schalingsfactor waardoor taalafstand anders kan worden gewogen dan fysieke afstand. Met de factor  $\frac{1}{2\sqrt{N}}$  wordt de fysieke afstand genormaliseerd naar een getal in  $[0, 1]$ . Immers, de maximale fysieke manhattanafstand in een raster van  $\sqrt{N} \times \sqrt{N}$  is  $\sqrt{N} + \sqrt{N}$ . Met behulp van de  $\psi$ -parameter kan de afstand in taal belangrijker worden gemaakt bij het clusteren. Een juiste waarde hiervoor kiezen is subjectief, in de huidige implementatie is voor een vaste waarde van 4 gekozen.

Het Leader-Follower clusteringalgoritme werkt niet met een vooraf vastgestelde hoeveelheid clusters, maar creëert een nieuw cluster zodra huidige clusters een te grote afstand hebben van een nieuw punt in de dataset. Deze afstand, de `max_distance`, is de tweede parameter van het algoritme. Doordat het aantal clusters niet van tevoren vaststaat, kunnen wij deze informatie gebruiken als meetmethode. Het verloop van het aantal clusters, uitgezet tegen de tijd, van een run is te zien in figuur 6. Deze meetmethode geeft echter weinig informatie naast de entropie, als te zien is in een vergelijking van figuur 4 ( $r = 0.63$ ).

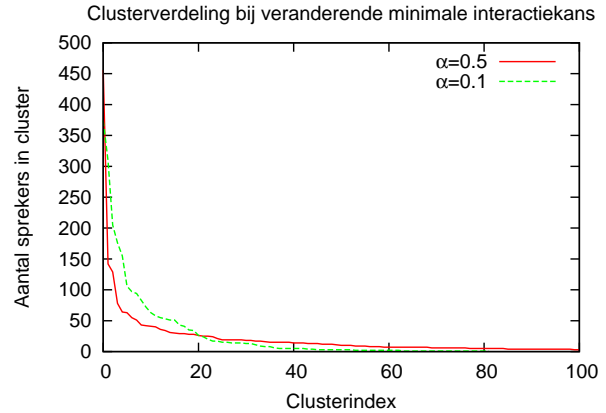
Interessanter aan het clusteringalgoritme is de verdeling van agenten in de clusters. Als er een scheve verdeling is, waarbij een groot aantal agenten zich slechts in een relatief kleine proportie van de clusters bevindt, duidt dit niet alleen aan dat er groepen agenten zijn die dezelfde taal spreken: als er veel clusters zijn met slechts enkele agenten kan dit een vage grensovergangen aanduiden. Een voorbeeld van zo'n scheve verdeling is te zien in figuur 7.

### 6.4 Aantal grenslijnen

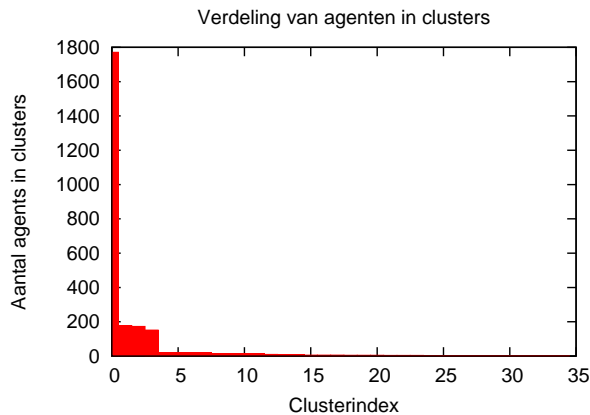
Een laatste telling die ons iets kan zeggen is het totaal aantal grenslijnen van een bepaalde sterkte. Omdat



**Figuur 6:** Verloop van aantal clusters van LF-clusteringalgoritme in de tijd



**Figuur 9:** Verschil in clusterverdeling bij verschillende  $\alpha$



**Figuur 7:** Verdeling van agenten in clusters

er maar een beperkt aantal taaleigenschappen zijn, zijn er slechts evenveel niveau's van taalverschil tussen twee agenten. Door te tellen hoeveel elke agent verschilt in taal van zijn burens, ontstaat er een overzicht van welke grenslijnen aanwezig zijn in een model.

Deze methodiek kan ons vertellen of harde grenslijnen überhaupt wel in het model voorkomen. Het zegt helaas niets over de lengte van zo'n grens. Ook is het geen meetmethode voor vage grenslijnen.

## 7 Experimenten

### 7.1 Parameterruimte

Een overzicht van alle belangrijke parameters is te vinden in tabel 1.

Alle experimenten zijn gedaan met een vaste po-

pulatiegrootte van 2500, wat neerkomt op een ruimte van  $50 \cdot 50$  groot. Ook is de mutatiegrootte, `mutation_size` constant gehouden op 16. 16 verschillende taaleigenschappen komt neer op  $2^{16} = 65536$  mogelijke talen, wat genoeg moet zijn om verschillende taalgrenzen te kunnen onderscheiden. De `max_distance` is vastgezet op 0.4. Agenten kunnen sterven vanaf hun 15e beurt, maar zijn al niet meer in staat om te leren vanaf hun 10e beurt.

### 7.2 Interactiekans

De communicatiekans is te beïnvloeden door twee verschillende parameters in het model.  $\alpha$  bepaalt een minimale interactiekans en  $\beta$  bepaalt de stijlheid van deze kans.

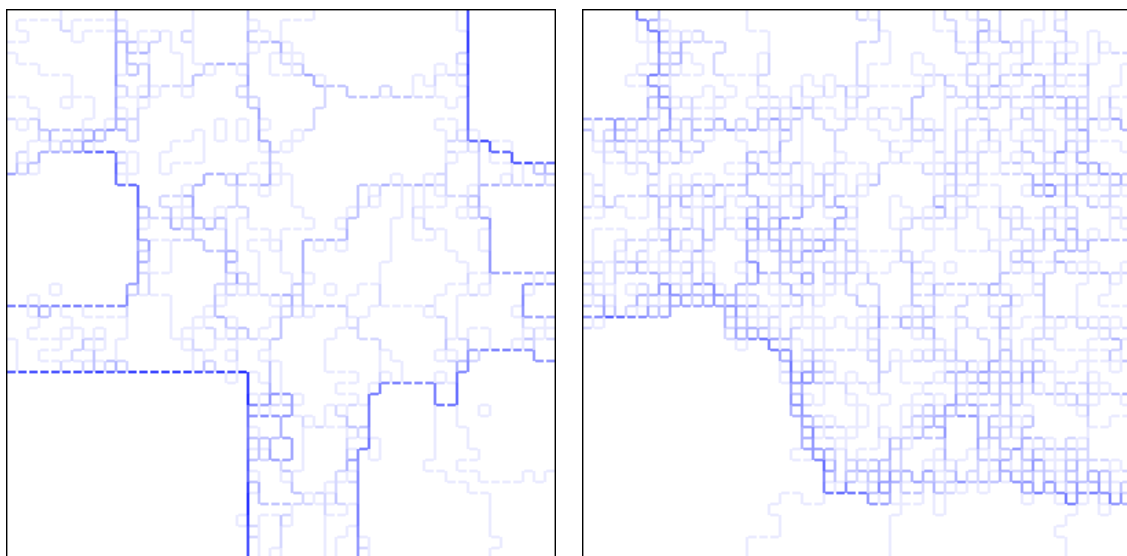
#### 7.2.1 Minimale interactiekans

Het verschil tussen een hoge en een lage minimale interactiekans is te zien in figuur 8. Figuur 8(a) heeft een lage minimale interactiekans (0.1), en figuur 8(b) een hoge (0.5). Deze figuren laten de situatie van een simulatie op een tijdstip zien. Elk wit vlak stelt een agent voor. De blauwe lijnen tussen de agenten stellen communicatiebarrières voor. Een gebied zonder lijnen geeft dus een groep agenten aan die dezelfde taal spreken. Hoe feller groen de lijn is, hoe slechter agenten met elkaar kunnen communiceren.

Het verschil in de twee situaties is duidelijk te zien als we de clusterverdeling van de eerste clusters bekijken. De situatie in figuur 8(a) op pagina 10 heeft in totaal 82 clusters, 8(b) heeft er 256. De situatie met een lage interactiekans heeft meer leden in de eerste clusters zitten dan in de rest, en is dus sche-

Naam	Symbol	Uitleg
num_agents	$N$	Aantal agenten, gelijk aan aantal vakken in omgeving
mutation_size	$M$	Grootte van de taalvector
mutation_introduction_chance	$p_{newmut}$	Kans op introduceren nieuwe mutatie
<b>Communicatie en Mutatie</b>		
com_min_chance	$\alpha$	Minimale communicatiekans
com_mod	$\beta$	Stijlheid van communicatie
takeover_min_chance	$\gamma$	SIT minimale interactiekans
takeover_mod	$\kappa$	SIT paramater
<b>Agentsterfte</b>		
max_age	–	Maximale leeftijd vóór sterfte van agent
max_learn_age	–	Maximale leeftijd waarbij taalmutatie mogelijk is
<b>Clustering</b>		
max_distance	–	LeaderFollower maximale clusterafstand
cluster_scale	$\psi$	Schalingsconstante (zie 6.3)

*Tabel 1: Parameters van het model*



(a) Een lage minimale communicatiekans ( $\alpha = 0.1, \beta = 3$ ) (b) Een hoge minimale communicatiekans ( $\alpha = 0.5, \beta = 3$ )

*Figuur 8: Verschil in minimale interactiekans*

ver. Dit komt naar voren in figuur 8 door de grotere hoeveelheid groepen en minder variëteit buiten deze groepen.

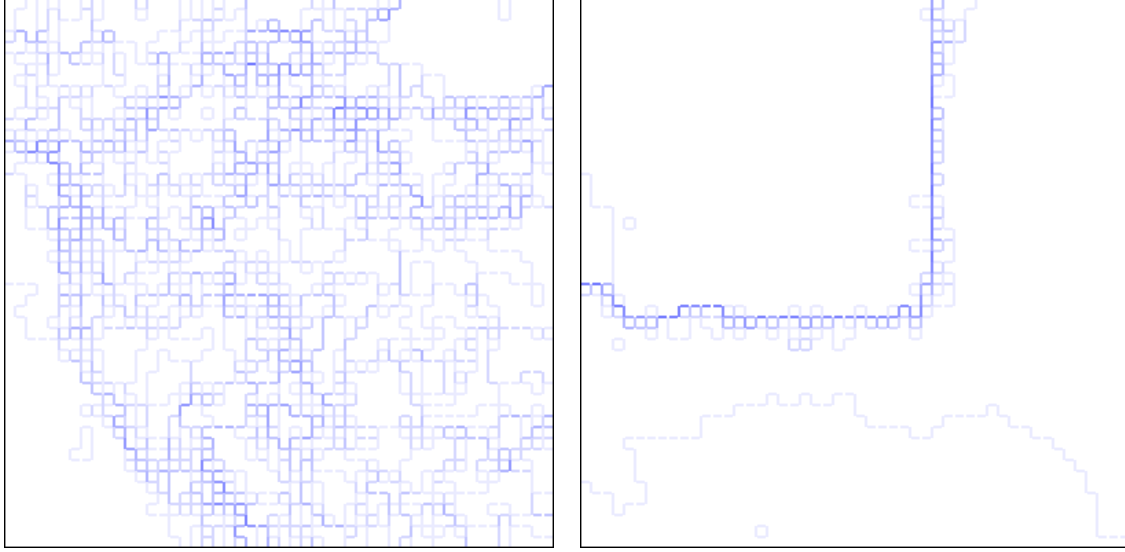
### 7.2.2 Stijlheid van interactiekans

De stijlheid van de interactiekans bepaalt vooral het ontstaan van grenzen en variëteit buiten deze grenzen. In figuur 10 zijn twee situaties uitgebeeld met een verschillende stijlheid. In figuur 10(a) zijn geen scherpe taalgrenzen te onderscheiden. In figuur 10(b) is daarentegen een heel gebied afgescheiden van de

rest. Dit is goed terug te zien in de entropie. Van beide parameterinstellingen zijn 10 runs gedaan. Met een  $\beta$  van 1 is de entropie gemiddeld 2.12 met een standaardafwijking van 0.32, met een  $\beta$  van 11 is de entropie gemiddeld 0.21 met een standaardafwijking van 0.09.

### 7.3 Overdrachtskans

De overdrachtskans is afhankelijk van twee parameters,  $\gamma$  en  $\kappa$ .  $\gamma$  bepaalt een minimale overdrachtskans,



(a) Een vlakke communicatiekans ( $\alpha = 0.4, \beta = 1$ )

(b) Een stijle communicatiekans ( $\alpha = 0.4, \beta = 11$ )

**Figuur 10:** *Verskil in stijlheid van interactiekans*

en  $\kappa$  bepaalt de stijlheid van deze kans.

### 7.3.1 Minimale overdrachtskans

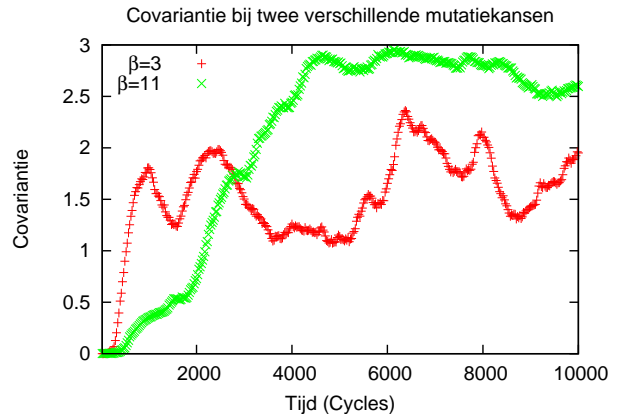
Een hogere minimale overdrachtskans veroorzaakt meer variëteit in de situatie, wat een hogere entropie oplevert. Een lage minimale overdrachtskans kan er voor zorgen dat er helemaal geen taaldifferentiatie ontstaat. Een hoge overdrachtskans is vergelijkbaar met de situatie van een stijle interactiekans.

### 7.3.2 Stijlheid overdrachtskans

De stijlheid van de overdrachtskans blijkt belangrijk te zijn voor de complexiteit van de situatie. Bij een te lage stijlheid is er nauwelijks sprake van dialectvorming. In figuur 11 is het verschil te zien tussen een  $\kappa$  van 0.3 en 0.5. Hoewel er wel sprake is van grensvorming, is deze minder scherp in figuur 11(a). Ook is er minder sprake van grensvervaging.

## 7.4 Mutatiekans

Een hoge mutatiekans leidt tot een complexere situatie. Zelfs met een lage mutatiekans is het echter nog mogelijk grensvorming te krijgen. Figuur 12 laat situaties met hoge en lage mutatiekansen zien. Het introduceren van een hoge mutatiekans leidt ook tot meer variatie binnen het systeem, zoals te zien is in figuur 13. Hier zijn de entropiën van de twee situaties in figuur 12 tegen elkaar uitgezet. Zoals valt te zien is de

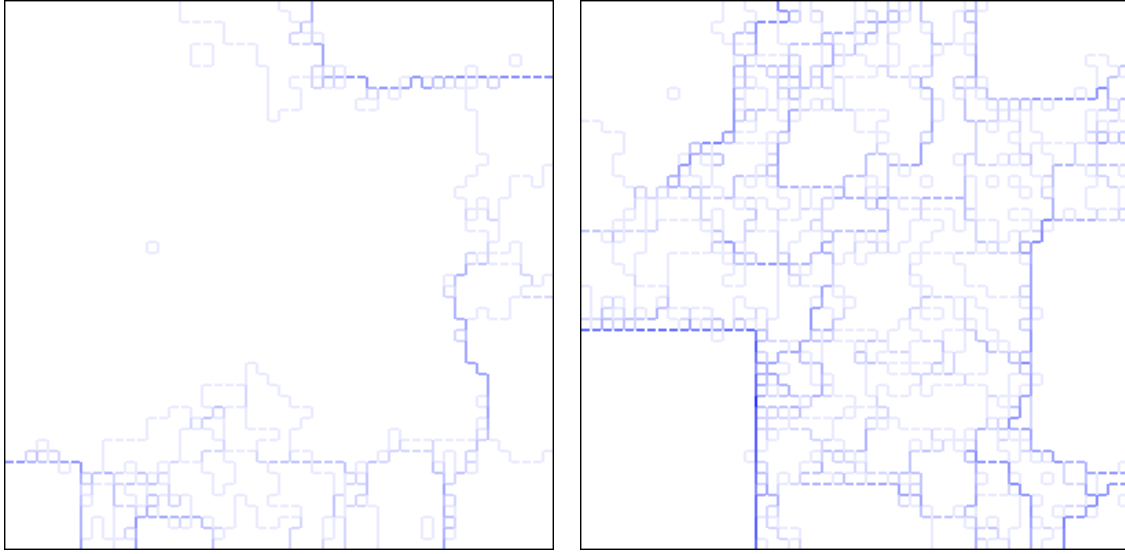


**Figuur 13:** *Verskil in covariantie bij verschillende mutatiekansen*

entropie niet alleen hoger, maar is dit maximum ook sneller bereikt. Dit valt te verklaren doordat verschillende mutaties sneller worden geïntroduceerd. Door deze snellere introductie is de kans dat twee mutaties elkaar tegenkomen ook groter, waardoor er een nieuwe taal ontstaat die de entropie verhoogt.

## 7.5 Agentsterfte

Om de invloed van agentsterfte te bepalen, zijn er een aantal simulaties gedaan waarin agentsterfte is uitgeschakeld. In dit model hebben agenten geen leeftijd, en zijn zij altijd in staat met andere agenten te com-



(a) Een vlakke communicatiekans ( $\gamma = 0.5, \kappa = 0.3$ )

(b) Een stijlvolle communicatiekans ( $\gamma = 0.5, \kappa = 0.5$ )

**Figuur 11:** *Verskil in stijlheid overdrachtskans*

municeren om taaleigenschappen over te nemen. Er worden dus ook geen nieuwe agenten geboren. Twee situaties van het model zonder agentsterfte zijn te zien in figuur 14.

Dit model geeft interessante resultaten. Zonder agentsterfte is er geen sprake van grensvorming. Ook is er geen groepsvorming te zien. Dit is te verklaren doordat agentsterfte in het model een stabiliserende en convergerende werking heeft. Het convergeert doordat nieuwe agenten de meestgesproken taal binnen zijn buurt overneemt. Stabilisatie ontstaat doordat de agent na een bepaalde leeftijd geen nieuwe taaleigenschappen meer kan overnemen.

## 7.6 Het neutrale model (Geen SIT)

Om de invloed van de Social Impact Theory te testen, is de overdrachtskans vervangen door een constante waarde. De kans of een agent een eigenschap overneemt van een andere agent is dan niet meer afhankelijk van hoeveel agenten die specifieke variant spreken, maar is altijd hetzelfde.

Dit is gedaan onder een wijde mogelijkheid van overdrachtskans, variërend tussen de 0 en 1. Onder geen enkele parameterinstelling was het model in staat om een variant door te voeren. Dit is te verklaren doordat er veel meer agenten zijn die de alternatieve taalvariant spreken. Het resultaat is in feite hetzelfde als die van Axelrod (1997). Hij noemt dit gedrag het resultaat van ‘random walk with absor-

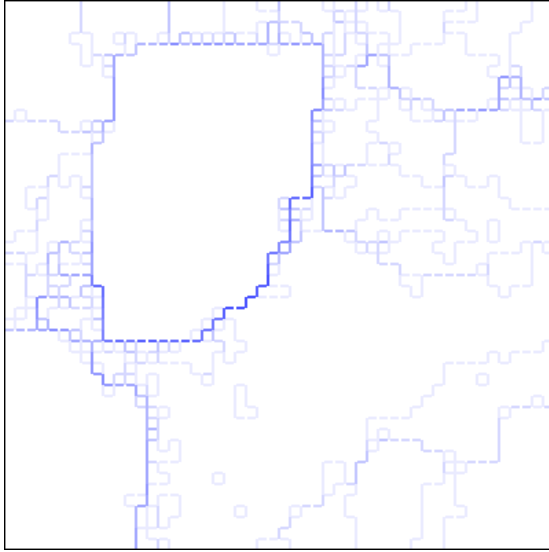
bing barriers’, wat betekent dat in een situatie waarin genoeg ruimte is om te muteren, de kleinste mutatie waarschijnlijk zal worden geabsorbeerd door de grotere.

## 8 Discussie

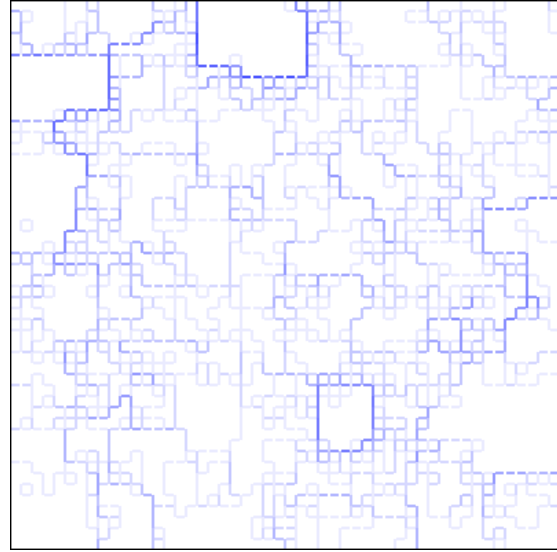
Aan het hier gepresenteerde model kunnen nog verschillende aanpassingen worden gedaan. Uit de literatuur komt een aantal suggesties over het creëren van een model voor taaldiversiteit. Enkele hiervan zijn het waard om nog expliciet genoemd te worden.

### 8.1 Taalverwerving

In het hier gepresenteerde model kan een agent alleen taal overnemen van zijn directe burens. Dit betekent dat hij niet aan de grote hoeveelheid taaldiversiteit om hem heen blootgesteld wordt, waardoor het threshold-probleem minder invloed heeft op een agent. In hoeverre dit uitmaakt moet onderzocht worden in een vervolgonderzoek. Aangezien het threshold-probleem zorgt voor een snelle convergentie, is het interessant te onderzoeken in welke mate een grotere interactiebuurt het missen van agentsterfte kan compenseren. Axelrod (1997) heeft in zijn model ook geëxperimenteerd met verschillende interactiebereiken. In zijn model zorgt een groter bereik voor minder onderscheidbare cultuurregio’s.

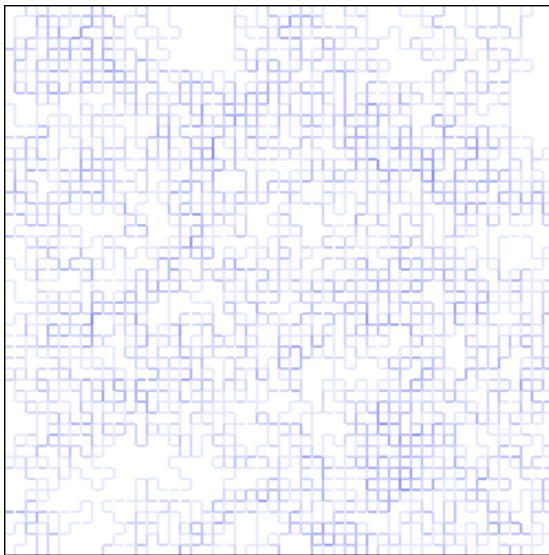


(a) Een lage mutatiekans ( $p_{newmut} = 10^{-5}$ )

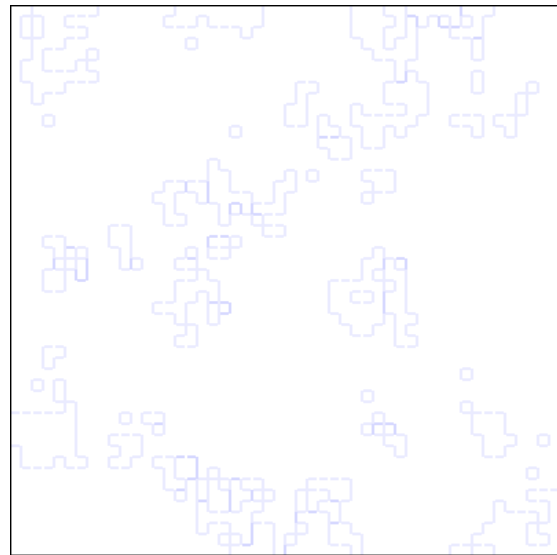


(b) Een hoge mutatiekans ( $p_{newmut} = 10^{-3}$ )

**Figuur 12:** *Vershil in mutatieintroductiekansen*



(a) Een hoge mutatieintroductiekans



(b) Een lage mutatieintroductiekans

**Figuur 14:** *Situaties zonder agentsterfte*

Verder stelt Nettle (1999a) dat taalverwerving een strategie heeft die fouten minimaliseert. De verschillende aspecten van taal moeten dus ook afzonderlijk worden bekeken en kunnen per situatie een andere variatie aannemen. In het hier gepresenteerde model wordt gekeken naar welke taal het meest voorkomt bij de burens van een agent. Deze wordt gebruikt als basistaal voor een agent. Volgens Nettle moet er per mutatie worden gekeken welke variant het meest

voorkomt om zo het taalverschil met alle burens te minimaliseren. Deze vorm van taalverwerving mist echter de stabilisatie die voorkomt in het hier besproken model.

## 8.2 Onafhankelijkheid van taaleigenschappen

In het model zijn taaleigenschappen gemodelleerd als onafhankelijke *linguistic items*. Hoewel taaleigenschappen onafhankelijk van elkaar kunnen worden aangeleerd, is het onwaarschijnlijk dat taaleigenschappen geheel geen invloed op elkaar hebben. Een voorbeeld hiervan zijn klinkersystemen: klinkergeluiden zullen zich op zo'n manier verdelen dat zij een zo groot mogelijke afstand in de fonetische en akoestische ruimte hebben. Welke bijdrage levert dit proces? Nettle (1999a) heeft dit specifieke voorbeeld geïmplementeerd in de vorm van functionele selectie, waarbij hij geen significante verschillen kon vinden. Meer onderzoek kan echter aanwijzen

## 8.3 Taalstabilisatie

In het huidige model wordt taalverwerving stopgezet zodra een agent een bepaalde leeftijd heeft bereikt. Hoe waarschijnlijk is dit? Deze afsnijding kan ook worden gemodelleerd door een functie, waarbij een agent minder snel taaleigenschappen kan overnemen naarmate hij ouder wordt. De taalafsnijding is echter niet compleet willekeurig gekozen: zowel Nettle (1999a,b) als Livingstone (2002) maken gebruik van eenzelfde soort afsnijden. Zij hebben er echter voor gekozen taalverwerving alleen in de eerste levenscyclus van agent plaats te laten vinden. Deze waarden lijken grotendeels arbitrair te zijn gekozen. Ook hebben zij geen experimenten gedaan met het weglaten van agentsterfte, waardoor er niet bekend is welke bijdrage deze processen in hun model leveren. Komt de stabilisatie in Nettle's systeem door de grotere globale interactie, of ook door zijn sterfte-implementatie? Meer onderzoek naar de effecten van agentsterfte en taalverwerving kan waardevolle informatie opleveren over welke processen noodzakelijk zijn voor de stabilisatie van een systeem.

## 9 Conclusie

Uit het model is gebleken dat het inderdaad mogelijk is taaldiversiteit te verkrijgen in een situatie die begint met linguïstische homogeniteit. Met het model zijn ook verschillende taalgrenzen succesvol in eenzelfde situatie gemodelleerd. Taaldiversiteit is echter al eerder aangetoond, onder andere door Nettle (1999b).

Belangrijker dan het daadwerkelijk aantonen van

de taalgrenzen zijn de omstandigheden waaronder dit gebeurt. In het model zijn zowel de sociale druk om bepaalde eigenschappen over te nemen, als de convergentie van de agentsterfte noodzakelijk om tot een stabiele oplossing te komen.

In tegenstelling tot Axelrod's model, waarbij communicatie wordt gezien als een vorm van leren die naar convergentie leidt, zorgt de interactie met behulp van de Social Impact Theory in dit model juist voor divergentie. Deze sociale motivatie is echter niet vastgesteld tussen verschillende agenten, maar is alleen afhankelijk van het totaal aantal sprekers dat een specifieke taalvariant spreekt. Het gevolg van scherper wordende grenzen is dat agenten zich in groepen met dezelfde taal organiseren. Doordat zij allemaal dezelfde taaleigenschappen spreken, worden zij indirect wel afhankelijk van elkaar. Dit effect treedt echter pas op nadat de grenzen al bestaan.

Door het wegvallen van het convergentiemechanisme zoals Axelrod deze heeft gemodelleerd, is er een nieuwe mechaniek nodig om het model stabiel te houden. In het huidige model komt dit naar voren in de vorm van agentsterfte en -geboorte. Agentsterfte zorgt voor convergentie, doordat de meestgesproken taal in de directe omgeving van de agent wordt gekozen als spreektaal. Door het afsnijden van taalverwerving kunnen agenten gestabiliseerd worden, waardoor minderheidsmutaties voor een langere tijd doorgegeven kunnen worden.

Met dit model is het mogelijk om taaldiversiteit te behouden, er is geen sprake van convergentie naar een enkele taal. Dit gedrag is te verkrijgen onder een zeer ruime instelling van de parameters. Voor het behouden van de taaldiversiteit is geen expliciete sociale selectie tussen de agenten nodig, maar slechts een sociale motivatie om een bepaalde variant van een enkele taaleigenschap te spreken, die kan worden verklaard door bijvoorbeeld nieuwigheid of "coolheid" van deze eigenschap. Dit idee heeft dan ook steun van de Social Impact Theory.

## Referenties

- Axelrod, R. (1997). The dissemination of culture: A model with local convergence and global polarization. *The Journal of Conflict Resolution*, **41**(2):203–226.
- Barr, D. J. (2004). Establishing conventional communication systems: Is common knowledge necessary? *Cognitive Science*, **28**(6):937–962.

- de Boer, B. (1997). Generating vowel systems in a population of agents. *Fourth European Conference on Artificial Life, Brighton*.
- Duda, R., Hart, P., en Stork, D. (2001). *Pattern Recognition*, chapter 10.11 ‘On-line Clustering’, pages 559–563. New York: A Wiley-Interscience.
- Kreyszig, E. (1993). *Advanced engineering mathematics*, chapter 22.9 ‘Distributions of several Random Variables’, pages 1098–1099. Wiley New York.
- Latané, B. (1981). The psychology of social impact. *American Psychologist*, **36(4)**:343–356.
- Livingstone, D. (2002). The evolution of dialect diversity. In Cangelosi, A. en Parisi, D., editors, *Simulating the Evolution of Language*, chapter 5, pages 99–118. London, Springer Verlag.
- Matsumoto, M. en Nishimura, T. (1998). Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, **8(1)**:3–30.
- Nettle, D. (1999a). *Linguistic Diversity*. Oxford University Press.
- Nettle, D. (1999b). Using social impact theory to simulate language change. *Lingua*, **108(2-3)**:95–117.
- Patriarca, M. en Leppanen, T. (2004). Modeling language competition. *Physica A: Statistical Mechanics and its Applications*, **338(1-2)**:296–299.
- Railsback, S., Lytinen, S., Jackson, S., en Computing, J. (2006). Agent-based Simulation Platforms: Review and Development Recommendations. *Simulation*.
- Shannon, C. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, **27(3)**:379–423.
- Wang, W. S.-Y. en Minett, J. W. (2005). The invasion of language: emergence, change and death. *Trends in Ecology and Evolution*, **20(5)**:263–269.
- Willemyns, R. (2002). The Dutch-French Language Border in Belgium. *Journal of Multilingual and Multicultural Development*, **23(1-2)**:36–49.