

Taaldistributie: een interdisciplinaire benadering

Bacheloronderzoek

Yuri Villa Rikkers

Dr. Ria van der Lecq

Dr. Herman Hendriks

Prof. dr. Sjef Barbiers

Interdisciplinair begeleider

Interdisciplinair begeleider

Taalkundig adviseur

Taaldistributie: een interdisciplinaire benadering

LAS 301: het interdisciplinaire sluitstuk van de bacheloropleiding Liberal Arts & Sciences, Universiteit Utrecht, juli 2010.

Student:

Yuri Villa Ridders (3136108)

Hoofdrichting: Taalkunde van het Nederlands

Begeleiders:

Dr. Ria van der Lecq: Interdisciplinair begeleider

Dr. Herman Hendriks: Interdisciplinair begeleider, taalkundig adviseur

Prof. Dr. Sjef Barbiers: Hoofdadviseur (generatieve) taalkunde

Afbeelding voorpagina: De ontwikkeling van het Hebreeuwse alfabet, opgeschreven door Wilhelm Gesenius begin 19^e eeuw, en later vertaald door A.E. Cowley in 1910.

Abstract

In deze scriptie wordt ingegaan op de factoren die een rol spelen bij taaldistributie, in termen van taalspreiding en taalevolutie. De eerste helft bestaat uit een zuiver taalkundige analyse van taaldistributie, waarbij zowel theoretische taalkunde als historisch-vergelijkende taalkunde aan bod komt. Vervolgens worden er parallellen getrokken tussen taalkundige en biologische begrippen, en zal blijken dat de populatiegenetica belangrijke inzichten verschaft in de werking van taaldistributie. Hierna wordt taaldistributie op microniveau bekeken vanuit de antropologie. Dit is van belang om de concrete werking van taalspreiding te kunnen begrijpen. In het kort worden dan nog enkele inzichten uit minder relevante disciplines besproken. Deze verzameling van multidisciplinaire inzichten worden tot slot geïntegreerd in een overkoepelend interdisciplinair model van taaldistributie.

Voor Ria, en haar aanstekelijke enthousiasme

Voor Herman, omdat het zo moeilijk is hem iets nieuws te vertellen

Voor Sjef, omdat het kritisch denken pas echt bij hem begon

Voor mijn ouders, die nu hopelijk eindelijk begrijpen wat ik studeer

Voor alle kopjes koffie, en de mensen die ze met me hebben gedeeld

Voor mij, omdat ik nu pas begrijp wat interdisciplinariteit betekent

Inhoud

0 Inleiding	9
0.1 Probleemstelling	9
0.2 Afbakening onderzoeksvraag	10
0.3 Verantwoording	10
0.4 Doelstelling	11
1 Taalkundige benadering	12
1.1 Microniveau	13
1.1.1 Fonologie	13
1.1.2 Morfologie	14
1.1.3 Semantiek	14
1.1.4 Syntaxis	15
1.1.5 Parameters	15
1.1.6 Het Y-model	17
1.1.7 Het Kind als centraal element	18
1.1.8 Optimaliteitstheorie	19
1.1.9 Optimaliteit in syntaxis	20
1.1.10 Samenvatting	22
1.2 Macroniveau	22
1.2.1 De Talen van de wereld	22
1.2.2 De comparatieve methode	24
1.2.3 Tekortkomingen van de comparatieve methode	25
1.2.4 Typologische prototalen	25
1.2.5 Isolatietalen	26
1.2.6 Samenvatting	28

1.3 Samenvatting	28
2 Biologische benadering	29
2.1 Prototaal	29
2.1.1 Biologische prototalen	30
2.1.2 Het continuïteitsprobleem	31
2.1.3 Eén of meerdere prototalen?	32
2.1.4 Homologie vs. homoplasie	33
2.1.5 Samenvatting	38
2.2 Populatiegenetica	38
2.2.1 Algoritmische methoden	38
2.2.2 The maximum parsimony method	39
2.2.3 The bayesian maximum likelihood method	40
2.2.4 Een boomstructuur reconstrueren	40
2.2.5 Migratie	41
2.2.6 Samenvatting	42
2.3 Samenvatting	42
3 Linguïstisch antropologische benadering	43
3.1 Taalcontact	43
3.1.1 Taalcontact en contacttalen	43
3.1.2 Multilinguïsme	44
3.1.3 Taalverschuiving	44
3.1.4 Diglossie	45
3.1.5 Samenvatting	45

3.2 Contacttalen	45
3.2.1 Mengtalen	46
3.2.2 Pidgintalen en creolentalen	46
3.2.3 Samenvatting	47
3.3 Microlinguïstische factoren	48
3.4 Samenvatting	49
4 Samenvatting	50
5 Integratie	51
5.1 Conflicterende begrippen	51
5.1.1 Taalvermogen	52
5.1.2 Taalevolutie	53
5.1.3 Taaldiversiteit	57
5.1.4 Taal distributie	59
5.1.5 Samenvatting	60
5.2 Factoren en inzichten	60
5.2.1 Generatieve taalkunde	60
5.2.2 Historische taalkunde	61
5.2.3 Evolutiebiologie	62
5.2.4 Populatiegenetica	62
5.2.5 Antropologie	63

5.3 Het model	64
5.3.1 Overzicht	64
5.3.2 Een overkoepelend model	67
5.4 Voor toekomstig onderzoek	68
Referenties	70
Artikelen	70
Boeken	72
Websites	73

0 Inleiding

0.1 Probleemschets

Volgens populaire definities telt de wereld momenteel ruim 6000 natuurlijke mensentalen (Unesco, 2010). De definities van taal lopen echter uiteen, en daarmee dus ook de schattingen van het aantal talen. In de Van Dale is taal bijvoorbeeld een “middel om gedachten en gevoelens uit te drukken”. Deze definitie is echter niet specifiek genoeg, omdat hij pretendeert dat taal een willekeurig element is uit de verzameling {middelen om gedachten en gevoelens mee uit te drukken}, waaronder ook het maken van een schilderij zou kunnen vallen. Binnen de taalkunde wordt vaak gesproken over taal als een “dialect met een leger en een marine” (oorspronkelijke bron en jaartal onbekend). Hoewel deze metafoor er wellicht goed in slaagt om bepaalde intuïties over taal te belichamen, brengt het ons eigenlijk geen meter verder. We zitten nu namelijk met het nog ongedefinieerde begrip *dialect*, waarvan het onmogelijk lijkt het te kunnen definiëren zonder te verwijzen naar het begrip *taal*. Daarnaast geven de termen “leger” en “marine” je de indruk dat een standaarddialect zodanig beschermd wordt dat het niet verandert, terwijl een standaardtaal net zo veranderlijk is als ieder ander dialect. Steeds vaker wordt er door taalkundigen liever gesproken over *idiolecten* (individuele dialecten) in plaats van talen, wat betekent dat er zoveel talen worden gesproken als er mensen zijn (ongeveer 6,79 miljard). Los van de definitie van taal is er ook te weinig onderzoek gedaan in de taalgebieden van de wereld. Met name van zuid Azië, Afrika en Papua Nieuw Guinea zijn er onvoldoende pogingen gedaan hun talen structureel in kaart te brengen (Asher, 1995). In elk geval is de taaldiversiteit op aarde erg groot: niemand hoeft ver te reizen om zich onverstaaanbaar te willen maken. Het ligt niet voor de hand in hoeverre taalvariatie mogelijk is en waarom. Sinds het begin van de jaren zestig, toen de linguïst Noam Chomsky felle kritiek leverde op het behaviorisme (Ashcraft, 2006), wordt veelal aangenomen dat taal niet iets kan zijn dat volledig cultureel bepaald is (Chomsky, 1957). Eerder ligt er aan taal volgens hem een cognitief taalvermogen ten grondslag dat de mens als soort uniek maakt. De enorme typologische¹ verscheidenheid aan talen

¹ Met typologische kenmerken worden vanaf nu bedoeld: kenmerken van geschreven taal die direct waarneembaar zijn. Er valt onderscheid te maken tussen lexicale typologische kenmerken, ofwel

probeert Chomsky vervolgens te verklaren met parameters: abstracte schakels in het taalvermogen die taalvariatie aan regels verbinden. Een echte verklaring voor taaldiversiteit is dit nog niet. Wat ontbreekt is bijvoorbeeld een reconstructie van de manier waarop taal door de tijd heen is verspreid. Voor dergelijke reconstructies is vervolgens kennis nodig van statistische modellen, en diepe verklaringen voor motivaties van mensen om andere talen te gaan spreken. Om deze motivaties in kaart te brengen blijkt het relevant te zijn het probleem vanuit meerdere invalshoeken te benaderen. Dit onderzoek is gericht op het in kaart brengen van relevante invalshoeken en de inzichten in taaldistributie die deze opleveren. De centrale onderzoeksvraag is daarbij:

Welke factoren spelen een rol in taaldistributie?

0.2 Afbakening onderzoeksvraag

Omdat er in het kader van de beperkte omvang van deze scriptie geen ruimte is om een verklaring te bieden voor de volledige distributie van talen in de wereld, wordt er slechts ingegaan op de belangrijkste factoren die een rol spelen in taaldistributie. Ten eerste zijn enkele factoren nog te weinig onderzocht om er iets relevants over te kunnen zeggen. Ten tweede zijn sommige factoren van zeer beperkte invloed op taaldistributie. Deze factoren worden dan ook beperkt behandeld. Tot slot is het een onmogelijke opgave om voor iedere taal in de wereld te achterhalen waarom deze daar gesproken wordt. Bovendien is deze opgave niet erg interessant voor deze scriptie, omdat hij niet tot inzicht leidt in de algemene werking van taaldistributie.

0.3 Verantwoording

Er wordt op dit moment een breed scala aan multidisciplinaire thema's onderzocht waar taal een rol in speelt. Een goed voorbeeld van een dergelijk thema is het *taalkundig relativisme*. Vanuit deze hypothese wordt onderzoek gedaan naar de invloed van taal op perceptie en de vorming van cultuur (Saeed, 2008). Ook taaldistributie wordt multidisciplinair onderzocht: Onderzoeken naar correlaties tussen politiek en taaldiversiteit (Currie & Mace, 2008), geografische kenmerken en taaldiversiteit (Silva & De Oliveira, 2008), en demografische groei en taalomvang (Zanette, 2008) groeien in populariteit door de ontwikkeling van nieuwe statistische modellen. Deze onderzoeken

eigenschappen van woordvormen, en structurele (syntactische) typologische kenmerken, ofwel eigenschappen van de manier waarop woorden in frasen en zinnen geordend zijn.

zijn niet volledig interdisciplinair van aard, omdat de inzichten niet worden geïntegreerd. Theorieën blijven hier binnen het bereik van verschijnselen uit de oorspronkelijke discipline, en er wordt niet getracht verschijnselen uit verschillende vakgebieden te integreren in een overkoepelend model. Om volledig te begrijpen hoe een willekeurige taal haar betreffende typologische karakter heeft gekregen en op haar plek is terechtgekomen, is zo'n overkoepelend model wel nodig. Hier ontleent dit onderzoek zijn relevantie aan. De beslissing om prioriteit te geven aan respectievelijk de taalkunde, de populatiegenetica en de antropologische linguïstiek komt niet uit de lucht vallen. De taalkunde is allereerst belangrijk voor het benadrukken van de complexiteit van taaldistributie, en daarmee het onderstrepen van mijn onderzoeksvraag. Taalkundig onderzoek heeft een overzicht opgeleverd van de huidige mondiale verdeling van taaltypen. Een verklaring voor de geanalyseerde situatie wordt door de taalkunde vervolgens niet geboden. Omdat de werking van taaldistributie grotendeels te maken heeft met de werking van menselijke interacties, is het noodzakelijk deze laatste bloot te leggen. De biologie en de antropologie blijken zich hier uitstekend voor te lenen. Door (metaforische) parallellen te trekken tussen genetische verwantschap en taaltypologische verwantschap is het mogelijk om statistische modellen uit de populatiegenetica te gebruiken voor historisch vergelijkend taalonderzoek. Tot slot komt de antropologie om de hoek kijken omdat de verspreiding van taal direct afhangt van het gedrag van groepen mensen. Mogelijk correleren er verschijnselen uit veel meer disciplines met het gedrag van talen, maar onderzoek moet nog uitwijzen in hoeverre deze verschijnselen van cruciale invloed zijn op taaldistributie. In elk geval worden de taalkunde, de biologie en de antropologie hier als de belangrijkste disciplines beschouwd om de onderzoeksvraag te beantwoorden.

0.4. Doelstelling

In interdisciplinaire werken is het doorgaans prettig om zo min mogelijk tussentijds te integreren. Hoewel hier meerdere redenen voor zijn, is de belangrijkste dat het vaak speculatieve karakter van integratieprocessen niet past bij de objectieve manier waarop disciplinaire inzichten worden beschreven. Inzichten die in de disciplinaire gedeeltes naar boven komen zullen dus pas aan het einde van deze scriptie met elkaar verbonden worden. Het doel is om dit te doen in de vorm van een interdisciplinair model, waarin alle inzichten verwerkt zijn. In het ideale geval vormen deze inzichten de input van het model, met als output een verklaring voor de manier waarop de talen van de wereld

momenteel gedistribueerd zijn. Het model is in dat geval zelfs in staat voorspellingen te doen over de manier waarop talen zich in de toekomst zullen gedragen. In werkelijkheid zal het model dit ideaal slechts benaderen. Vanzelfsprekend zal het bijvoorbeeld eenvoudiger zijn om met behulp van het model kwantitatieve uitspraken te doen dan kwalitatieve. Zo zal het aantal talen dat er over 1000 jaar gesproken wordt gemakkelijker te voorspellen zijn dan het uiterlijk van deze talen, of de plaats waar ze worden gesproken. Dit komt doordat er voor het voorspellen van de toekomst hetzelfde geldt als voor het reconstrueren van de geschiedenis: het is eenvoudiger te achterhalen wát er is gebeurd dan hóe het is gebeurd. Wellicht is dit onderscheid zelfs terug te zien in algemene verschillen tussen bètawetenschappen en andere wetenschappen; in dit geval tussen bijvoorbeeld de biologie en de antropologie. Waar de populatiegenetica zich bezig houdt met exacte reconstructies en voorspellingen van populatiespreiding, richt de antropologie zich meer op de mechanismen die deze spreiding aansturen. De voorspelling is daarom dat de antropologie de meeste vragen over taaldistributie openlaat. Eerder zal het een opsomming van mogelijk betrokken ingrediënten opleveren, dan een kookboek waarin deze ingrediënten gecombineerd worden volgens een afgemeten recept. Dit neemt niet weg dat het mogelijk moet zijn het model van een architectuur te voorzien die plaats biedt voor inzichten die de toekomst mogelijk oplevert. Het bieden van deze ruimte is de laatste eis die aan het model gesteld wordt.

1 Taalkundige benadering

De opkomst en ondergang van talen is een voortdurend dynamisch proces. Op het moment zijn er “officieel” (Unesco, 2009) ruim 6000 talen bekend in de wereld. Met ongeveer 6 miljard mensen in 195 landen komt dat neer op een gemiddelde van 30 talen per land, en één miljoen sprekers per taal. Veel taalkundigen ontkennen echter dat een taal gevangen kan worden binnen geografische grenzen of typologische definitives. (Asher, 1995) Ondersteuning voor dit kritische standpunt is te vinden in bijvoorbeeld de dialectvariatie binnen het Nederlands. Alleen al op syntactisch niveau zijn er meer dan 250 Nederlandse dialecten bekend (Barbiers, 2006). Taalkundigen spreken daarom vaak liever van idiolecten in plaats van talen. Dit betekent dat ieder individu op aarde een unieke taal spreekt. Technisch gesproken is er dan dus tussen alle mensen een

taalafstand. In het ideale geval betekent dit dat gesproken taal in de wereld iedere kilometer langzaam en gradueel typologisch van karakter verandert. Je verwacht dan bijvoorbeeld dat Nederlandse dialecten richting het oosten steeds meer op het standaard Duits gaan lijken. In dit geval blijkt dat inderdaad te kloppen, maar op mondiaal niveau gaat de vlieger niet meer op. Door grootschalige volksmigraties zijn talen soms abrupt in een vreemd taalgebied terechtgekomen. De geleidelijkheid van taalafstand is in deze gebieden ver te zoeken. Hoewel het zou kunnen dat talen zich op grensgebieden aan elkaar aanpassen gaat hier de nodige tijd overheen. Vanaf nu wordt er voor het gemak uitgegaan van 6000 talen.

1.1 Microniveau

Hoewel er in deze scriptie uiteindelijk naar een inzicht wordt gestreefd in de mechanismen die ten grondslag liggen aan taalvariatie, is het noodzakelijk eerst te bekijken in welke opzichten taal überhaupt kan variëren. Dit wordt onderzocht in het kader van de generatieve ofwel theoretische taalkunde. Deze discipline onderscheidt vier autonoom veronderstelde hoofddomeinen van het cognitieve taalvermogen: de *fonologie*, de *syntaxis*, de *morfologie* en de *semantiek*. Deze domeinen corresponderen met de representatieniveaus van een taaluiting, en worden hier behandeld omdat ze stuk voor stuk beperkingen opleggen aan taalvariatie. Ook is er in deze paragraaf aandacht voor *Optimality Theory*, waarin nog andere cognitieve bronnen voor taalvariatie een rol spelen. Bewijzen voor het bestaan van, de werking van, en de onderlinge relaties tussen de domeinen zijn altijd van empirische aard.

1.1.1 Fonologie

De fonologie is het systeem dat beperkingen oplegt aan klankvormen die je in een taal kunt realiseren. Het fonologische systeem van het Nederlands staat bijvoorbeeld de klankvorm *stzek* niet toe. In het Tzotzil, een taal gesproken in Mexico, is deze klankvorm wel toegestaan, en betekent zelfs iets (*rok*) (Baker, 2001). Bewijs voor een fonologisch systeem wordt geleverd door sterke intuïties van mensen over klanken. *Stzek* zou nooit een plek in het Nederlandse lexicon kunnen innemen, maar *zoeg* zou dat wel kunnen, hoewel dat nu nog niets betekent. Het fonologische regelsysteem is dus de eerste beperking bij het genereren van een taal. Zodra deze set regels is opgenomen in een taal kunnen betekenissen niet meer aan willekeurige klankvormen gerelateerd worden. Aangezien het regelsysteem niet geschreven maar gesproken taal betreft, is het

fonetisch alfabet bedacht waar klankvormen mee worden beschreven. Het Engelse woord *dogs* wordt in dit alfabet bijvoorbeeld weergegeven als /dogz/.

1.1.2 Morfologie

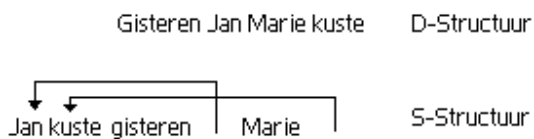
De morfologie, ook wel de wordsyntaxis genoemd, legt beperkingen op aan de interne geleiding van woorden. Ten eerste stelt deze component eisen aan de manier waarop woorden gecombineerd kunnen worden tot samenstellingen. Zo kunnen *rood* en *huid* gecombineerd worden tot *roodhuid*, maar *rood* en *arm* niet tot **roodarm* (de asterisk geeft de onmogelijkheid van deze samenstelling weer). Ten tweede heeft het morfologisch systeem betrekking op het toegestane gedrag van ongebonden morfemen. Deze hebben zelf geen betekenis, maar voegen betekenis toe aan een woord waarmee ze gecombineerd worden. Een voorbeeld is de mogelijkheid van *paard* – *paarden* tegenover de onmogelijkheid van *panda* - **pandaen*, waarbij *-en* dan het ongebonden (meervouds)morfeem is. Steeds vaker wordt het morfologische domein verder opgesplitst in twee subdomeinen, omdat wordt aangenomen dat het om verschillende processen gaat. Bewijzen hiervoor kunnen onder andere bij kinderen worden gezocht, die twee typen morfemen in verschillende fasen verwerven (Don et al., 1994). De subdomeinen heten *derivatie* en *flexie*. Derivatie heeft onder andere te maken met woordsoortveranderende morfemen. Zo zorgt het morfeem *-en* achter het naamwoord *golf* ervoor dat het verandert in een werkwoord (*golven*). Onder flexie valt met name werkwoordsvervoeging. Naast de fonologie beperkt de morfologie het lexicon nog iets verder in de mogelijkheden voor taaluitingen, en net als in de fonologie gelden er voor ieder taaltype andere regels.

1.1.3 Semantiek

De semantiek is het domein waarbinnen een taaluiting betekenis krijgt. Binnen deze discipline proberen taalkundigen de betekenis achter taaluitingen formeel te representeren, en relaties tussen vorm en betekenis en zinsvolgorde en betekenis bloot te leggen. Of de semantiek beperkingen kan opleggen aan taalvariatie staat sterk ter discussie (Semantics, 2008). Deze discussie wordt gedomineerd door de vraag of mensen uit verschillende culturen ook verschillende mentale concepten vormen die mogelijk van invloed zijn op het gedrag van taal of andersom. Hier wordt nu verder niet op ingegaan.

1.1.4 Syntaxis

De syntaxis geeft een set regels die de woordvolgorde in een zin bepalen. Naar dit onderdeel van de generatieve taalkunde is het meeste onderzoek gedaan. Het is het enige generatieve domein waarvan de verschillende mogelijke regels zijn vastgelegd in een universele theorie, ofwel een die op alle talen toepasbaar zou moeten zijn. De grondlegger van deze theorie is Noam Chomsky. Hij verklaart de in de wereld waargenomen syntactische variatie aan de hand van zogeheten *parameters* (Chomsky, 1995). Ook neemt Chomsky een zekere verzameling van universele syntactische eigenschappen waar, die in elke taal voorkomen. Die noemt hij *principes*. Verder spreekt Chomsky van twee interpretatieniveaus in de syntaxis: de *Deep Structure* en de *Surface Structure*. Op D-structuur-niveau hebben alle woorden in een zin een vaste volgorde. Om de juiste uitgespelde zin te kunnen krijgen moeten *verplaatsingen* worden toegepast. Hierna vindt *spellout* plaats, ofwel het uitspellen van de uiting. In figuur 1.1 staat een voorbeeld.



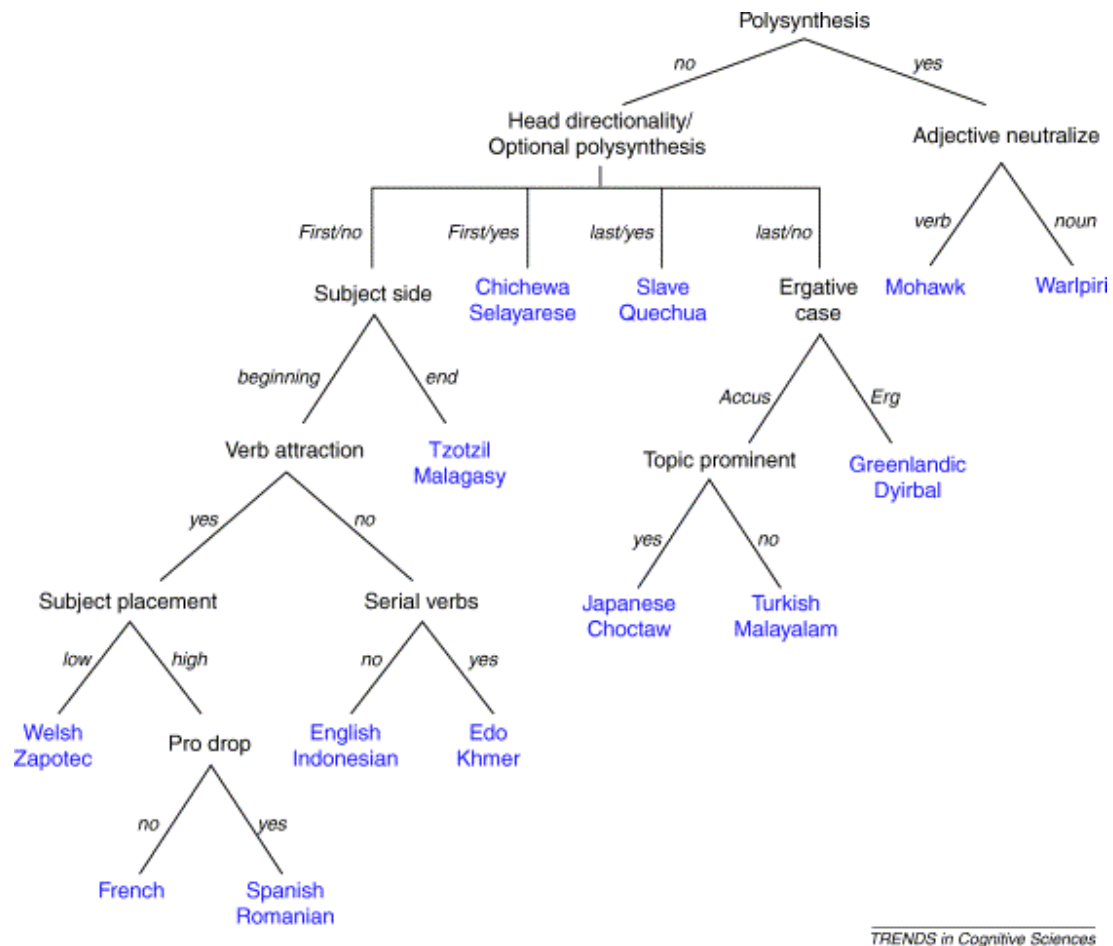
Figuur 1.1 Verplaatsingen in een zin.

Bewijzen voor woordverplaatsingen komen uit empirisch crosslinguïstisch onderzoek. Alle potentiële posities waarin woorden kunnen staan blijken een dergelijk woord slechts onder bepaalde omstandigheden te kunnen herbergen. Het is de taak van syntactici om de verschillen in eigenschappen tussen deze posities expliciet te maken in modellen. De precieze werking van verplaatsing is voor deze scriptie irrelevant, en wordt niet besproken. Op parameters wordt nu wel wat dieper ingegaan.

1.1.5. Parameters

Parameters worden door Chomsky voorgesteld als een set binaire schakelaars. Een taal heeft elk van deze schakelaars dus “aan” of “uit” staan. Deze parameters komen tot uiting in de zinsvolgordes van de betreffende taal. Op deze manier probeert Chomsky te verklaren waarom bepaalde volgordes wel, en andere volgordes niet voorkomen in de talen van de wereld. Een diepere analyse van de werking van parameters komt van

Mark Baker. Hij stelt zich een zekere hiërarchie voor waarin de parameters een vaste plaats hebben.



TRENDS in Cognitive Sciences

Figuur 1.2 The Parameter Hierarchy (Baker, 2003)

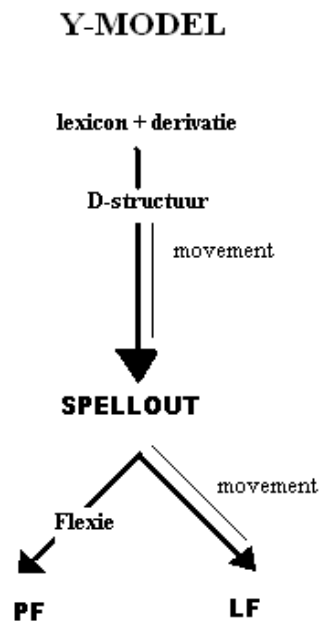
Deze hiërarchie laat bijvoorbeeld zien dat in een polysynthetische taal niet verschillende naamvalsvarianten kunnen voorkomen en andersom². Dit model verklaart nog concreter waarom bepaalde volgordes wel en andere niet mogelijk zijn. Een opmerkelijk feit is overigens dat alle bovenstaande syntactische types op meerdere, historisch-taalkundig

² Bovenaan de hiërarchie in figuur 1.2. staat de parameter *polysynthesis*. Talen waarin polysynthese voorkomt hebben deze parameter op *yes* staan. Daaronder staat een nieuw filter: de Adjective Neutralize Parameter. Dit is de enige parameter die binnen polysynthetische talen voor syntactische variatie kan zorgen. Verschillen in bijvoorbeeld naamval (in de figuur aangegeven met *Erg* vs. *Accus*) zijn hier dus niet mogelijk.

niet direct verwante plekken in de wereld voorkomen. Vergelijk in figuur 1.2 bijvoorbeeld het Engels en het Indonesisch. Hoewel deze talen exact dezelfde syntactische eigenschappen hebben zijn ze historisch niet verwant (Serva & Petroni, 2007). Hieruit blijkt dus dat je historische verwantschap van talen niet kunt achterhalen door alleen naar syntactische eigenschappen te kijken.

1.1.6 Het Y-model

Hoewel de vier domeinen allemaal unieke eigenschappen hebben en als autonoom beschouwd kunnen worden, lijkt er ook een bepaalde onderlinge afhankelijkheid te bestaan. Zo wordt gedacht dat de syntaxis gedreven wordt door de morfologie. Het model dat deze hiërarchie weergeeft heet het Y-model (maar wordt ook wel T-model genoemd). In figuur 1.3 is een versie van dit model afgebeeld.



Figuur 1.3 Een voorbeeld van het Y-model, gebaseerd op Carnie (2002).

De Phonological Form (PF) stelt de fonologische representatie voor en de Logical Form (LF) de semantische. Met het lexicon wordt in dit model bedoeld de input van woordstammen die nog geen derivatie of flexie hebben ondergaan. Met *movement* worden de syntactische verplaatsingen bedoeld, van D-structuur naar S-structuur. Om uit te leggen waar al deze ideeën vandaan komen, is het nuttig eerst wat dieper op de werking van het model in te gaan. Het misleidende van het Y-model en de manier

waarop het model vaak wordt gepresenteerd, is dat het niet bedoeld is als een diachroon model maar als een synchroon model. Het is dus niet mogelijk er een tijdas bij te plaatsen, ook al insinueren termen als "verplaatsing" dit wel. Wat het model wel weergeeft, is de hiërarchie van representatieniveaus van de autonome domeinen. Concreet betekent dat voor figuur 1.3 dat verplaatsingen in het syntactische domein bijvoorbeeld worden gestuurd door (ondergeschikt zijn aan) processen uit het derivatieve domein. Deze ideeën zijn op hun beurt afgeleid van empirisch onderzoek, waaruit bijvoorbeeld letterlijk blijkt dat de woordsoort, welke vastgelegd wordt met derivatieprocessen, bepalend is voor de distributie van het woord. Ideeën over het bestaan van, de werking van, en het uiterlijk van het Y-model variëren echter enorm. Zo is er voor iedere aanname wel een tegenvoorbeeld te bedenken aan de hand van andere empirische gegevens. Voor dit onderzoek is het vooral belangrijk dat ieder verondersteld domein voor ieder type taal een eigen set regels genereert, en daardoor als filter optreedt van de mogelijkheden die in een taal zijn toegestaan.

1.1.7 Het kind als centraal element

Het is onwaarschijnlijk dat kinderen alle zinnen die in hun moedertaal mogelijk zijn uit het hoofd leren. Zo zijn kinderen bijvoorbeeld in staat zinnen te produceren die ze niet eerder gehoord hebben. Het is zelfs niet aannemelijk dat kinderen de hierboven besproken regels uit het hoofd leren als zij een taal beginnen te spreken. Ten eerste zou dit leerproces een enorme taak zijn. Ten tweede zou dan het leerproces afhankelijk moeten zijn van de intentie van het kind. Omdat kinderen hun moedertaal uiteindelijk allemaal even vloeiend leren spreken, is het moeilijk voorstelbaar dat intentie een rol speelt bij het ontwikkelen van taal (Kerstens, 1997). Taal lijkt eerder iets te zijn dat op een bepaalde leeftijd tot uiting komt, vergelijkbaar met bijvoorbeeld de uiting van secundaire geslachtskenmerken op latere leeftijd. Chomsky realiseerde zich dit, en zag kindertaalverwerving daarom als een proces waarbij parameters worden *gefixeerd* in het taalvermogen. Het kind zou zeer weinig talige input nodig hebben om deze afstellingen te ontdekken en toe te passen in de eigen spraak. Een voor deze scriptie belangrijk inzicht van Eric Lenneberg (1964), is het bestaan van een bepaalde kritische leeftijdsgrens waarna kinderen niet meer in staat zijn deze parameters af te stellen. Dit biedt een mogelijke verklaring voor het feit dat volwassenen een vreemde taal alleen maar kunnen leren met heel veel oefening, en dan nog zelden helemaal vloeiend. Lenneberg stelde de kritische leeftijd vast op 12 jaar. Later werd er gesproken van een

kritische "periode", die werd vastgesteld tussen 5 en 16 jaar. Echte bewijzen zijn er echter nooit gevonden. In elk geval kan wel worden aangetoond dat leeftijd een rol speelt bij het gemak waarmee een eerste of een tweede taal aangeleerd kan worden. Hierdoor zijn kinderen waarschijnlijk de enige taalverwervers die een verandering in een taal optimaal tot stand kunnen brengen. Dit maakt dat kinderen een centraal element vormen in het proces van taalspreiding.

1.1.8 Optimaliteitstheorie

In het fonologische domein binnen de taalkunde houdt men zich onder andere bezig met de vraag of er een regelsysteem bestaat achter klankstructuren. Zo kent het Nederlands twee meervoudsuitgangen (-s , -en), en de redenen om voor sommige woordstammen –s en voor andere –en te kiezen zijn van fonologische aard. Het blijkt echter niet eenvoudig een sluitend regelsysteem achter deze variatie te vinden. Om dit probleem op te lossen is het idee van optimaliteit in het leven geroepen (Kooij & Van Oostendorp, 2003). De op dit begrip gebaseerde theorie (optimaliteitstheorie) gaat ervan uit dat het taalvermogen een set fonologische eigenschappen zoveel mogelijk probeert te vermijden. Op het moment dat, in het gebruikte voorbeeld met meervoudsmorfemen, een woordstam met suffix –s regel x schendt, maar met suffix –en regel y schendt, moet er een keuze gemaakt worden tussen twee kwaden³. Door zulke verschijnselen veelvoudig en crosslinguïstisch te analyseren komt er een set restricties tevoorschijn die voor alle talen gelijk is. Verschillen in fonologische regels worden dan verklaard door aan te nemen dat deze restricties per taal hiërarchisch verschillend gestructureerd zijn. Restrictie x is in taal A dan bijvoorbeeld prominenter dan in taal B. Ter illustratie staat in figuur 1.4 de hiërarchie van restricties voor het Engels beschreven: *SS<Agree<Max<Dep<Ident. (Wikipedia, 2010)⁴ Deze restricties worden opgesteld in een zogeheten *tableau*. Aan de linkerkant van het tableau worden meerdere potentiële, willekeurig gekozen vormen neergezet van het woord waarnaar je op zoek bent. In dit

³ Ongebonden morfemen worden ook wel *affixen* genoemd, waarbij *suffixen* achter een woord worden geplakt en *prefixen* ervoor. Bijzondere affixen zijn *infixen* en *circumfixen*. Een infix komt midden in een woord te staan (*opbellen* -> *opgebeld*). Een circumfix bestaat uit een prefix en een suffix die zonder elkaar niet kunnen bestaan (*berg* -> *gebergte*)

⁴ *SS staat voor de restrictie die wordt geschonden wanneer twee sibilanten elkaar in een syllabe opvolgen. Sibilanten zijn klanken die in de fonologie aan de hand van hun gedrag gegroepeerd zijn in een natuurlijke klasse. Agree wordt geschonden wanneer een stemloze en een stemhebbende klank elkaar opvolgen. Max wordt geschonden als de toegevoegde klank (bijvoorbeeld een meervoudsmorfeem) niet wordt uitgesproken. Dep wordt geschonden als er klanken worden toegevoegd voor de uitspraak, wat bijvoorbeeld gebeurt bij *koe* + *en* = *koeien*. Ident tot slot, wordt geschonden zodra klanken anders gaan klinken dan ze contextvrij zouden klinken.

voorbeeld gaat het om de optimale klankvorm van het meervoud van /cat/. Deze worden vervolgens getoetst aan de eisen, beginnend aan de linkerkant. De *SS conditie wordt in dit geval door geen enkele vorm geschonden, dus moet er naar de volgende conditie gekeken worden. Klankvorm /catz/ valt hier af omdat deze de conditie *Agree* schendt, aangegeven met * voor “schending” en ! voor het feit dat de schending fataal is voor de klankvorm, omdat er klankvormen te bedenken zijn die condities schenden die lager in de hiërarchie staan dan deze conditie. Uiteindelijk moet gekozen worden voor /cats/, waarbij de minst belangrijke conditie wordt geschonden.

cat + z > cats

cat + z	*SS	Agree	Max	Dep	Ident
catiz				*!	
catis				*!	*
catz		*!			
cat			*!		
☞ cats					*

Figuur 1.4 Een optimaliteitstableau waar het meervoud van /cat/ mee gezocht kan worden (Wikipedia, 2010)

Optimaliteit is heel vruchtbaar gebleken om de variërende distributie van suffixen in talen mee te verklaren.

1.1.9 Optimaliteit in syntaxis

In de dialectologie is vrij recent ontdekt dat er ook binnen het syntactische domein sprake zou kunnen zijn van optimaliteitsprincipes (Ackema & De Hoop, 2006). Zo kan een dialect een verplaatsing van een woord uit D-structuur in context A wel toestaan, maar in context B niet. Een nabijgelegen ander dialect kan dezelfde verschijnselen hebben, maar dan precies andersom: De verplaatsing wordt niet toegestaan in context A, maar wel in context B. De hiërarchie van eisen lijkt dus anders te zijn in de twee talen, zonder dat er een verschil in parameterafstellingen bestaat. Blijkbaar is de syntactische variatie waarin taal kan voorkomen niet volledig te vangen in parameters. Tot voor kort werden zulke problemen in de syntaxis nooit op een onomstreden manier opgelost. Drie gebruikte strategieën om de problemen op te lossen waren de volgende:

- Aannemen dat niet gevonden verschijnselen wel bestaan, maar niet zichtbaar zijn. Zo zouden bepaalde verplaatsingen wel, en andere niet zichtbaar zijn.

- Aannemen dat er een abstractieniveau bestaat waarop de geschonden restrictie niet wordt geschonden.
- De restrictie zodanig modificeren dat deze de nieuwe gevonden verschijnselen kan verklaren.

De eerste twee strategieën zijn omstreden omdat de betreffende assumpties niet bewezen kunnen worden. De bekende abstractieniveaus zoals D-structuur kunnen empirisch ondersteund worden, maar blijken vaak slechts ad hoc te werken. De laatste strategie is omstreden omdat er altijd wel nieuwe tegenvoorbeelden gevonden kunnen worden, waardoor de restrictie oneindig vaak moet worden bijgesteld. In Optimality Theory ga je ervan uit dat er inderdaad universele restricties bestaan, maar dat deze net als in het fonologische domein in iedere taal anders gerangschikt zijn. Hoewel we dan weer opgescheept zitten met een abstractieniveau dat niet bewezen kan worden, biedt het idee een aantal voordelen ten opzichte van bijvoorbeeld D-structuur. Ten eerste kan verklaard worden waarom een verschijnsel x in context A wel, maar in context B niet kan voorkomen. Ten tweede hoeft er geen onderscheid meer gemaakt te worden tussen principes en parameters. Alle restricties zijn nu namelijk universeel. Een probleem hierbij is dat sommige syntactische varianten heel veel, en andere syntactische varianten heel weinig voorkomen op de wereld, zonder dat verwantschap hierbij een grote rol speelt. Om dit op te lossen is voorgesteld dat sommige hiërarchieën van restricties logischer zijn dan andere, omdat sommige restricties van nature dominanter zijn dan andere. De reden dat deze dominante restricties soms alsnog laag in de hiërarchie komen te staan levert dan weer opnieuw vragen op, en zo kun je door blijven gaan. Inderdaad zijn er theorieën waarin meerdere strata (abstractieniveaus) worden voorgesteld die verantwoordelijk zijn voor de uiteindelijk waarneembare hiërarchie van restricties in talen (Bermúdez-Otero, 2005). Een belangrijke vraag is of de voorkeur voor bepaalde syntactische regels willekeurig is, of cognitief wordt aangestuurd. Er zijn cognitiewetenschappers die beweren dat er voor de productie van sommige zinnen meer mentaal werkgeheugen nodig is dan voor andere (Ackema et al., 2006). In dit geval zou het kunnen zijn dat een taal er altijd naar streeft zo eenvoudig mogelijk te worden, ofwel zo min mogelijk proceskosten van het brein te eisen. Op lexicaal niveau zijn hier aanwijzingen voor te vinden. Zo worden sterke (onregelmatige) werkwoorden vaak verkeerd gebruikt, doordat ze op een zwakke (regelmatige) manier worden vervoegd. Met name bij kinderen worden dergelijke overgeneralisaties aangetroffen,

zoals **loopte* en **valde*. Ook de verklaring voor het feit dat Nederlands sprekenden vaak **groter als* zeggen in plaats van *groter dan* zou zijn oorsprong kunnen vinden in cognitieve efficiëntie. Deze gevallen zijn echter niet van syntactische aard, en zeggen daarom niet zoveel.

1.1.10 Samenvatting

Het cognitieve taalvermogen kent slechts beperkte mogelijkheden voor het genereren en aanleren van een taal. Hoewel het spraakorgaan een grote verzameling klanken kan produceren, bestaan er allerlei restricties op de manier waarop die klanken gecombineerd kunnen worden. De ruimte waarbinnen taal kan variëren is mogelijk dankzij parameters, en hiërarchisch gestructureerde principes. Voor het aanleren van een tweede taal naast de moedertaal is het essentieel dat dit op vroege leeftijd wordt gedaan. Daarom zijn kinderen noodzakelijk voor het verspreiden van taal.

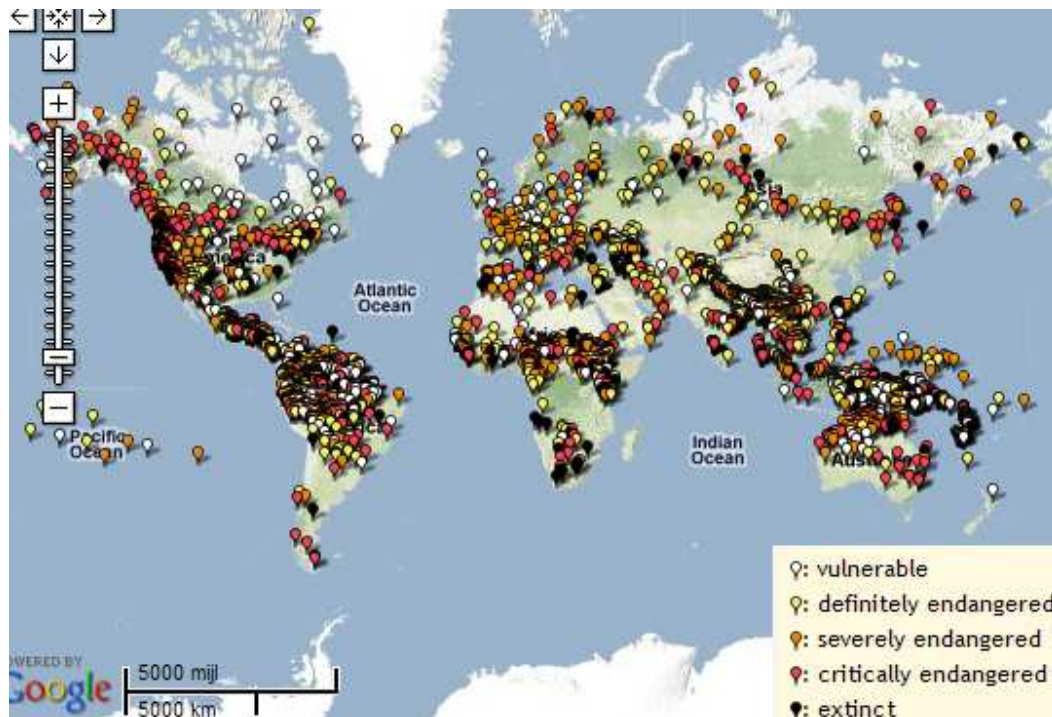
1.2 Macroniveau

Nu is besproken binnen welke cognitieve grenzen taalvariatie mogelijk is, kan er naar macrolinguïstische verschijnselen worden gekeken. Eerst wordt een analyse gegeven van de manier waarop de talen van de wereld momenteel typologisch verdeeld zijn. Tot slot wordt gekeken hoe deze talen door de geschiedenis heen op hun plek zijn gekomen, en welke methoden gebruikt kunnen worden om dit te achterhalen. De kwesties die hier worden onderzocht bevinden zich voornamelijk op het terrein van historische taalkundigen.

1.2.1 De Talen van de wereld

De 6000 talen van de wereld zijn alles behalve homogeen verspreid. Zo worden er in Papua Nieuw Guinea meer dan 800 talen gesproken onder nog geen 7 miljoen mensen (Pawley & Ross, 1993). Dit is een onvoorstelbaar aantal vergeleken met bijvoorbeeld Nederland, waar twee talen (Nederlands en Fries) worden gesproken onder bijna 17 miljoen mensen. Op basis van historisch vergelijkend onderzoek wordt verondersteld dat er ongeveer 17 taalfamilies zijn (Cavalli-Sforza et al., 1988). Of deze families in een ver verleden een gedeelde oorsprong hebben kan moeilijk gezegd worden, zoals verderop in deze scriptie zal blijken. De meest bestudeerde taalfamilie is het Indo-Europees, waaronder ook het Nederlands valt. Meer dan een derde van de talen in de wereld wordt

op dit moment met uitsterven bedreigd. Unesco heeft gegevens hierover recentelijk in kaart gebracht:



Figuur 1.5 De bedreigde talen van de wereld (Unesco, 2009)

Experts verwachten dat de helft van alle talen deze eeuw nog zal uitsterven, en dat de wereld uiteindelijk zelfs overblijft met slechts zo'n 300 talen (Asher, 1995). Hoewel taalsterfte een interessant verschijnsel is, staat in deze scriptie eerst de vraag centraal hoe de bedreigde talen in de eerste plaats op hun plek terechtgekomen zijn, evenals de andere circa 4000 talen van de wereld. Om deze vraag te beantwoorden worden zowel *romantische* als *regressieve* benaderingen gebruikt (Walker, 1980). Bij een romantische aanpak begin je bij de oorsprong, en probeert van daaruit de ontwikkeling in de richting van het heden te verklaren. Theorieën over *biologische prototalen* volgen dit idee. Biologische prototalen worden in hoofdstuk 2 besproken. Bij een regressieve aanpak bekijk je de huidige situatie, en probeer je terug te redeneren naar de oorsprong. *Typologische prototalen* zijn de uitkomsten van de toepassing van regressieve methoden. Typologische prototalen zijn hypothetische oertalen waar alle tegenwoordig

gesproken talen hun eigenschappen aan danken. De klassieke methode om deze talen mee te reconstrueren is de *comparatieve methode*.

1.2.2 De comparatieve methode

Historische taalkundigen maken gebruik van twee vergelijkende methoden om de achtergrond van talen te achterhalen (Pawley & Ross, 1993). De oudste is de *typologische comparatieve methode* (TCM). Hierbij wordt gekeken naar structurele, ofwel syntactische overeenkomsten en verschillen tussen talen. Zo zou een gelijke parameterafstelling kunnen wijzen op een gemeenschappelijke achtergrond. Een voordeel van deze methode is het feit dat structurele kenmerken minder gevoelig zijn voor verandering dan bijvoorbeeld lexicale kenmerken (Dunn et al., 2008). Een nadeel is dat er maar weinig parameters bestaan, die ook nog eens binair van aard zijn. Hierdoor is de kans groot dat twee onverwante talen een gegeven parameter (of zelfs alle parameters) hetzelfde hebben afgesteld. Het blijkt dat alle mogelijke woordvolgorden inderdaad voorkomen in alle delen van de wereld (Baker, 2001). Structurele overeenkomsten wijzen dus niet per definitie op een gelijke afkomst van talen. Dit maakt deze methode enigszins onbetrouwbaar. Een alternatief is de *genetische comparatieve methode* (GCM). Deze gaat uit van het feit dat de relatie tussen klankvorm en betekenis arbitrair is. Zo is er bijvoorbeeld geen logische reden om een stoel een *stoel* te noemen (vergelijk *chair* in het Engels). Omdat mensen een enorm scala aan klanken kunnen produceren is het erg onwaarschijnlijk dat er tijdens het ontstaan van een taal bij toeval dezelfde klankvorm werd toegekend aan een concept als in een andere, onverwante taal.⁵ De genetische comparatieve methode maakt gebruik van deze inzichten door op zoek te gaan naar morfemen in verschillende talen die eenzelfde klank-betekenisrelatie hebben. Een toepassing van de GCM is de *Swadesh lijst*. Dit is een lijst met ruim 200 ongebonden morfemen die naar alle talen in een onderzoeksgebied vertaald worden om ze te kunnen vergelijken. Op deze manier komen *cognaten*, ofwel woorden met een gemeenschappelijke etymologische oorsprong aan het licht (Dunn et al., 2008).

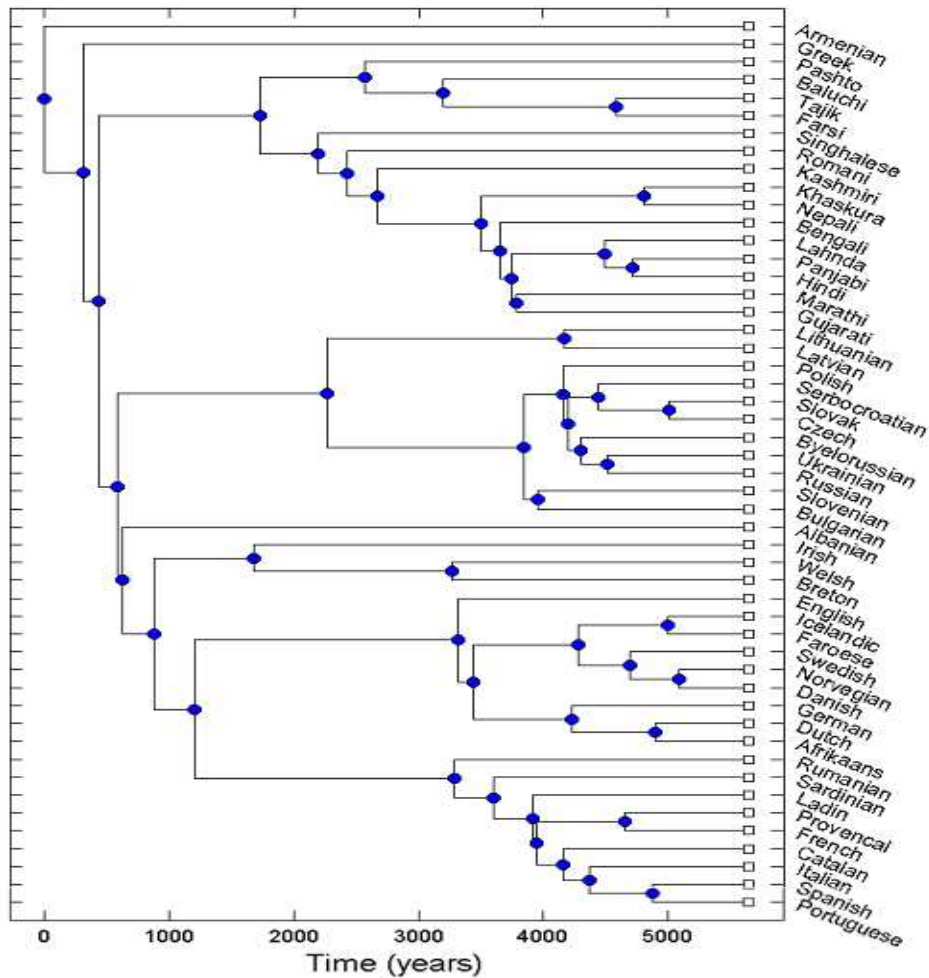
⁵ *Onomatopeeën* vormen hierin een uitzondering. Dit zijn woorden waarvan de klank enigszins verband houdt met de representatie in de werkelijkheid. Een voorbeeld is “koekoek”, de naam van een vogel die een geluid maakt dat op deze naam lijkt. Ook zogeheten *nursery forms* zoals *papa* en *mama* lijken geen arbitraire klanken te zijn, en eerder universeel van karakter.

1.2.3 Tekortkomingen van de comparatieve methode

Hoewel de GCM de meest gebruikte onderzoeksmethode in de historische taalkunde is, blijkt hij niet altijd betrouwbaar. Zelfs in de grootste taalfamilies, of in taalfamilies waarvan de oudste documentatie gevonden is, zijn lexicale relaties tot niet verder dan 10.000 jaar geleden terug te traceren (Dunn et al., 2008). Hoewel de grootste volksmigraties rond die tijd al hadden plaatsgevonden (Dunn et al., 2008), schiet de genetische comparatieve methode tekort. Soms zijn fonologische vormen van woorden zozeer veranderd dat hun oorspronkelijke vorm en daarmee hun afkomst niet meer zichtbaar is (Reesink et al., 2009). Deze veranderingen zijn vaak het gevolg van langdurige isolatie. Op het eiland Melanesië hebben talen bijvoorbeeld mogelijk meer dan 8000 jaar in isolatie verkeer. Dit zou de reden kunnen zijn voor het feit dat ze oppervlakkig gezien onverwant lijken te zijn (Hunley et al., 2008). Hun structurele eigenschappen komen voor een veel groter deel overeen, waardoor een gemeenschappelijke afkomst vooralsnog niet valt uit te sluiten. De belangrijkste tekortkoming van de GCM is het feit dat hij onderzoekers niet in staat stelt het verschil te herkennen tussen cognaten en *leenwoorden*. Leenwoorden zijn letterlijk woorden die geleend worden uit een andere taal via *taalcontact*. Over dit probleem wordt meer uitgelegd in de hoofdstukken 2 en 3. Verder werkt de comparatieve methode met standaardtalen zoals het standaard Nederlands (in de volksmond ook wel Algemeen Beschaafd Nederlands genoemd), en negeert hij dialectvariatie. Dit is een behoorlijk probleem als je bedenkt dat alleen het Nederlands al meer dan 250 dialecten telt (Barbiers, 2006).

1.2.4 Typologische prototalen

Met grootschalige vergelijkingen van taalcorpora en het gebruik van statistische modellen zijn pogingen gedaan de geschiedenis van taalspreiding te reconstrueren. Een recent voorbeeld hiervan is in figuur 1.6 weergegeven.



Figuur 1.6 Een fylogenetische reconstructie van het Indo-Europees (Serva & Petroni, 2007)

Links bovenaan bevindt zich de knoop waaruit alle weergegeven taalfamilies, en uiteindelijk dus alle weergegeven talen zouden zijn vertakt. In de taalkunde wordt deze knoop geacht een volwaardige taal te zijn geweest: het *Proto Indo-Europees*. Andere veronderstelde prototalen zijn het *Proto-Dravidisch* en het *Proto-Oeraals*, waaruit weer andere talen zijn vertakt. Hoeveel prototalen uiteindelijk verantwoordelijk zijn geweest voor het totale aantal talen in de wereld is onduidelijk (Cavalli-Sforza et al., 1988). In hoofdstuk 2 zal bovendien blijken dat er meerdere definities van prototaal bestaan.

1.2.5 Isolatietalen

Een isolatietaal is een taal die typologisch in alle opzichten afwijkt van de talen waardoor zij wordt omgeven. In figuur 1.7 staan in het rood de talen weergegeven die niet

afstammen van het Indo-Europees. Het meest linkse rode vlekje is het taalgebied waar Baskisch wordt gesproken.



Figuur 1.7 De distributie van taalfamilies in Europa. Rood staat voor non-Indo Europees, groen voor Romaans (Wikipedia 2008)

Het Baskische taalgebied is volledig omsloten door Romaanse talen (aangegeven met groen) en andere Indo-Europese talen (aangegeven met grijs). Voor historisch taalkundigen is een verschijnsel als dit een probleem. Het is tot nog toe onduidelijk hoe de taal daar terechtgekomen is (Trask, 1996). Veelal wordt gedacht dat het Baskisch veel ouder is dan de Indo-Europese talen in Europa (Trask, 1996). Het idee hierbij is dat de Basken behoren tot een oervolk dat al in Baskenland zat voor de Europeanen er massaal heen migreerden. Het is in dit geval opmerkelijk dat alleen het Baskisch weerstand heeft weten te bieden tegen de invloeden van de Europeaanse talen. Hierbij moet wel gezegd worden dat het aantal Baskisch sprekers drastisch afneemt, wat waarschijnlijk zal resulteren in het uitsterven van de taal (Unesco, 2009). In figuur 1.8 zijn een aantal woorden uit Romaanse talen vergeleken met hun Baskische equivalenten. Het valt direct op dat lexicale overeenkomsten ver te zoeken zijn.

Nederlands	Frans	Italiaans	Catalaans	Portugees	Roemeens	Spaans	Baskisch
één	un	uno	u	um	unu	uno	bat
twee	deux	due	dos	dois	doi	dos	bi
drie	trois	tre	tres	três	trei	tres	hiru
maandag	lundi	lunedì	dilluns	segunda- feira	luni	lunes	Astlehena
dinsdag	mardi	martedì	dimarts	terça-feira	marti	martes	Asteartea
woensdag	mecredi	mercoledì	dimecres	quarta-feira	miercuri	miércoles	Asteazkena
januari	janvier	gennaio	gener	janeiro	ianuarie	enero	Urtarrila
februari	février	febbraio	febrer	fevreiro	februarie	febrero	Otsaila
maart	mars	marzo	març	março	martie	marzo	Martxoa

Figuur 1.8 Een vergelijking van het Baskisch met de omliggende Romaanse talen (Klein & Stegmann, 2000).

1.2.6 Samenvatting

In deze paragraaf is eerst besproken hoe de talen op dit moment verdeeld zijn over de wereld. Vervolgens werd de vraag gesteld hoe ze op hun plaats terechtgekomen zijn, en welke methoden voornamelijk worden gebruikt om dit te achterhalen. Het bleek dat de meest gebruikte methode, de comparatieve methode, vaak behoorlijk tekort schiet.

1.3 Samenvatting

In dit hoofdstuk is besproken welke rol taalkundigen spelen bij het onderzoeken van taaldistributie. Theoretische ofwel generatieve taalkundigen houden zich bezig met de cognitieve mogelijkheden en restricties die mensen in staat stellen taal te genereren. Zij zijn van mening dat taal niet op een willekeurige manier is opgebouwd, maar zich houdt aan bindende eisen opgelegd door het cognitieve taalvermogen. Binnen deze beperkingen is er echter nog genoeg speling in het taalvermogen om een zeer groot aantal talen te genereren. Dit idee wordt ondersteund door de enorme taalvariatie op aarde. Taaltypologen zijn meer geïnteresseerd in de output van het taalvermogen dan de werking van het taalvermogen zelf. Ze analyseren geschreven en gesproken tekst, en proberen achter de historische verwantschap van talen te komen. De hiervoor gebruikte comparatieve methode kan echter geen rekening houden met verschijnselen als taalcontact, met als gevolg dat hypothetische reconstructies van taalfamilies onbetrouwbaar worden. Taalkundigen houden zich in dit opzicht teveel bezig met de vraag *hoe* talen kunnen bestaan en *hoe* die talen op hun bestemming belanden. Het zal blijken dat de vraag *waarom* talen zich in eerste instantie verspreiden eerst zal moeten

worden beantwoord om taaldistributie zo volledig mogelijk te kunnen doorgronden. Hiervoor is aandacht in hoofdstuk 3. Nu zal eerst worden gekeken in hoeverre ideeën en methoden uit de biologie taalkundigen kunnen helpen dichter bij een realistische historische analyse van taaldistributie te komen.

2 Biologische benadering

Deze scriptie gaat over de factoren die een rol spelen bij taaldistributie. In hoofdstuk 1 is uitgelegd dat alle talen zich aan regels houden. Op basis van deze regels kan uitgezocht worden welke eigenschappen van talen wel, en welke eigenschappen niet op toevalligheid berusten. Vervolgens kan deze kennis gebruikt worden om de herkomst van talen te reconstrueren, en om verwantschap van talen in kaart te brengen. Dit kan op twee manieren. De eerste manier is het bestuderen van lexicale overeenkomsten tussen woorden met behulp van de genetische comparatieve methode. Het nadeel hiervan is dat het meestal onmogelijk is het verschil te zien tussen leenwoorden en cognaten, waardoor het goed mogelijk is dat er valse verwantschap wordt ontdekt, of correcte verwantschap wordt genegeerd. De tweede manier is het bestuderen van structurele ofwel syntactische eigenschappen. Het nadeel is hier dat structurele eigenschappen niet per se op conventionele wijze zijn ontstaan, en daarom niet hoeven te wijzen op verwantschap. Om deze reden is de typologische comparatieve methode in hoofdstuk 1 verworpen. Toch bieden structurele eigenschappen van taal perspectief, omdat ze stug van karakter zijn en daarom in staat verwantschap aan het licht te brengen. Door te kijken naar methoden uit de biologie is men op het idee gekomen om parallellen te trekken tussen structurele eigenschappen en genen, en historisch taalkundig onderzoek te baseren op onderzoek uit de populatiegenetica. Voordat deze technieken aan de orde komen, worden er eerst theorieën over biologische prototaal behandeld met het doel daarmee de zoektocht naar een "oertaal" voort te zetten.

2.1 Prototaal

In hoofdstuk 1 werden prototalen beschouwd als hypothetische volwaardige vooroudertalen. De nadruk lag daar op de *prestaties* van het taalvermogen, en op de vraag hoe taaluitingen door de tijd heen zijn geëvolueerd. Een prototaal is dan de eerste

volwaardige taal, waar alle hedendaags gesproken talen hun eigenschappen aan ontlenu. In de evolutiebiologie heeft de term prototaal verwarrend genoeg een andere betekenis.

2.1.1 *Biologische prototalen*

Met een prototaal wordt in de biologie bedoeld: een eerste, nog onvolwaardig communicatiesysteem waar intentioneel door mensen gebruik van is gemaakt (Tallerman, 2005). Het is erg lastig om dit een plek te geven in theorieën over de ontwikkeling van mensen en hun taal, omdat er allerlei discussies bestaan over de manier waarop de mens zich heeft afgesplitst van haar voorouder. In recente literatuur over dit onderwerp heerst het idee dat de allereerste mens een volwaardige taal beheerste (Aitchison, 1996). Dit vanuit de gedachte dat een diersoort zoals *Homo sapiens sapiens* nu eenmaal wordt gedefinieerd aan de hand van gedeelde genetische eigenschappen, en daarmee gedeelde fysiologische expressies zoals de organen en hersenstructuren die samen het taalvermogen vormen. Verder is het idee dat ons taalvermogen niet van de ene op de andere dag is ontstaan in het brein, maar dat evolutieprocessen dit vermogen langzaam hebben ontwikkeld. De mens wordt geacht zich pas te hebben afgesplitst van haar voorouder nadat het taalvermogen volledig ontwikkeld was. Dit impliceert dat de eerste mensen dus ook een eerste volwaardige taal (typologische prototaal) moeten hebben gesproken. Biologische prototaal betekent in dit geval dus: het communicatiesysteem dat gebruikt werd door een groep humanoïden die zich na voltooiing van het taalvermogen heeft afgesplitst van de soort. Een groot probleem bij dit verhaal is de verborgen aanname dat de populatie zich tijdens het ontwikkelen van het taalvermogen niet heeft verspreid, maar samengepakt is gebleven tot de mens uitgeëvolueerd was om vervolgens – na nog eens een slordige 100.000 jaar⁶ - pas te migreren. Van opgesplitste populaties is immers bekend dat deze zich beide anders gaan ontwikkelen in een andere omgeving, terwijl het taalvermogen bij alle mensen op aarde gelijk is gebleven. Het lijkt echter nogal sterk om aan te nemen dat de populatie humanoïden zich tijdens duizenden jaren nooit heeft verspreid. De volgende vraag die biologen zich stellen is hoe het taalvermogen zich fysiologisch dan precies heeft ontwikkeld, en hoe deze ontwikkelingen zich hebben geuit in communicatie.

⁶ Dit wordt beter uitgelegd in paragraaf 2.1.3

2.1.2 Het continuïteitsprobleem

Er bestaan in de biologie meerdere hypothesen over prototalen, onder te verdelen in *synthetische theorieën* en *holistische (analytische) theorieën* (Tallerman, 2005). Synthetische theorieën gaan uit van het idee dat mensen eerst semantische concepten ontwikkelden, die later gekoppeld werden aan klankvormen. Syntactische principes zouden pas later hun intrede hebben gedaan. Holistische theorieën gaan uit van het idee dat mensen al lang voor het ontstaan van prototaal klanken gebruikten om te communiceren. Deze klanken zouden pas later bewust gekoppeld zijn aan concepten. Verder zouden de klanken aanvankelijk *monomorfologisch* zijn geweest, wat wil zeggen dat ze nog geen interne geleiding hadden. Het opbreken van klanken in kleinere betekenisdragende elementen (morfemen) zou pas later, bij het ontstaan van prototaal, hebben plaatsgevonden. Bewijzen voor deze gedachten worden gezocht in kennis over de ontwikkeling van het strottenhoofd en andere fysiologische structuren, maar ook in kennis over de ontwikkeling van het menselijk brein. In beide typen theorieën is er sprake van een *continuïteitsprobleem* tijdens twee stadia in de evolutie van het taalvermogen. Het eerste stadium begint bij de allereerste overgang van diertaal naar een complexer systeem, en eindigt op het punt waar er voor het eerst sprake is van een intentioneel communicatiesysteem (biologische prototaal). Het probleem is hier dat het onduidelijk is welke processen hiervoor verantwoordelijk zijn geweest. Uit onderzoek blijkt bijvoorbeeld dat dierengeluiden door heel andere hersengebieden worden veroorzaakt dan mensentaal (Tallerman, 2005), en er is niets bekend over de werking van eventuele verschuivingen van functies naar een ander hersengebied. Ook van het daarop volgende stadium, waarin biologische prototaal zich heeft ontwikkeld tot een eerste volwaardige taal, is het – zoals hierboven besproken - niet duidelijk welke stadia het taalvermogen precies heeft doorlopen en welke selectieprocessen hieraan ten grondslag hebben gelegen. Ook hier zit dus een gat van duizenden jaren waarin weinig bekend is over de evolutie van de mens. Hoewel taalkundigen (en dan in het bijzonder typologen) zich in de regel niet bezig houden met deze problemen, heeft Chomsky wel pogingen gedaan dit probleem op te lossen. Hij spreekt van *saltatie*, ofwel een abrupte genetische mutatie die ten grondslag ligt aan een reeks ingrijpende veranderingen in het fenotype. Gezien saltatie een omstrede oplossing is voor dit probleem (Pinker, 1995), zijn er pogingen gedaan het vraagstuk op een andere manier op te lossen. Een populair alternatief is het idee van *rijping*. Rijping staat normaal gesproken voor de ontwikkeling van het taalvermogen bij een kind. In termen van evolutie kan het begrip ook gebruikt

worden om de ontwikkeling van het taalvermogen in de soort mee te verklaren. Het idee hierachter is dat het taalvermogen an sich niet veel complexer is geworden sinds er sprake was van biologische prototaal, maar dat het vermogen zich in iedere generatie mensen complexer heeft geuit. De reden voor deze vertraagde uiting kan het gebrek aan talige input zijn geweest: in tegenstelling tot nu hadden de eerste mensen met een volwaardig taalvermogen geen ouders die een volwaardige taal spraken. Van rijping is bekend dat dit niet tot nauwelijks wordt geactiveerd zonder talige input (Pinker, 1995).

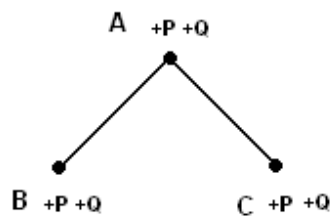
2.1.3 Eén of meerdere prototalen?

200.000 tot 150.000 jaar geleden splitste onze soort (*Homo sapiens sapiens*) zich af van haar voorouder. Er wordt gedacht dat ons taalvermogen zich sindsdien nauwelijks meer heeft ontwikkeld, wat betekent dat de allereerste mensen net als wij een volwaardige taal spraken (Aitchison, 1996). Vermoedelijk ongeveer 60.000 jaar geleden verspreidde de mens zich langzaam vanuit Afrika. Hoe taal zich heeft ontwikkeld en verspreid binnen de populatie in de circa 100.000 jaar ervoor is tot op heden onduidelijk. Als er in die periode helemaal geen populatiespreiding heeft plaatsgevonden, verwacht je dat alle mensen in de populatie dezelfde taal hebben gesproken (typologische prototaal). Hoewel dit een elegant antwoord zou zijn in de zoektocht naar een oertaal, zijn er helemaal geen aanwijzingen voor. Zowel fylogenetische als taaltypologische reconstructies – vervaardigd aan de hand van grootschalig genetisch onderzoek - ontberen namelijk een stam⁷: er zijn geen duidelijke sporen die in de richting wijzen van een enkele moederpopulatie of oertaal (Singh et al., 2004). Bovendien zou het op zijn minst merkwaardig zijn als de populatie mensen in 100.000 jaar tijd onverspreid is gebleven. Er zijn dus redenen om aan te nemen dat er meerdere populaties tegelijk een verschillende volwaardige taal ontwikkelden, ook al is het lastig hierbij een voorstelling te maken. De keuze voor de aanname van één of meerdere prototalen heeft grote gevolgen voor theorieën over taaldistributie. Het is daarom voor deze scriptie relevant hier nog wat dieper op in te gaan.

⁷ De begrippen (*fylogenetische/etymologische*) boom, boomstructuur en reconstructie staan allemaal voor de genetische of taaltypologische grafische weergaven van historische relaties tussen talen of populaties. Hierbij moet gedacht worden aan een boom die op zijn kop staat, waarbij de oorsprong de stam genoemd wordt. Deze staat voor de moedertaal of moederpopulatie. De tijd loopt in de figuren dus naar beneden, waarlangs vertakkingen verschijnen die staan voor opsplitsingen van talen of populaties. Merk op dat er in dit specifieke geval, waarbij de boom een stam ontbeert, technisch gesproken dus helemaal geen sprake is van een boom.

2.1.4 Homologie vs. homoplasie

Het bovengenoemde probleem is erg complex. Samenvattend zijn er vier scenario's mogelijk: Het taalvermogen ontwikkelde zich bij alle aanstaande mensen tot één of meerdere biologische prototalen, afhankelijk van de spreiding in de populatie, en later tot één of meerdere typologische prototalen per biologische prototaal. Het lijkt onmogelijk om één van deze scenario's te bewijzen vanwege tegenstrijdige assumpties over onder andere evolutietheorie, populatiespreiding en archeologische vondsten. Om inzicht te krijgen in de mogelijke gevolgen van alle scenario's is het verhelderend twee termen in het leven te roepen: *homologie* en *homoplasie*. Homologie is het proces waarbij een populatie haar eigenschappen overdraagt op de volgende generatie. Een schematische weergave staat in figuur 2.1.

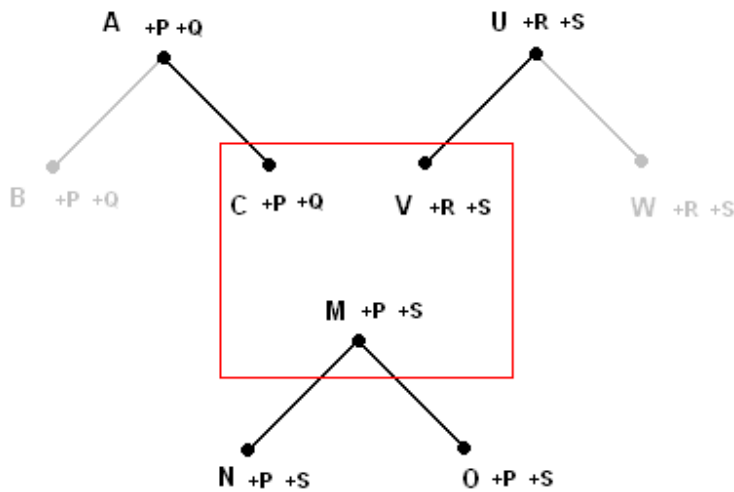


Figuur 2.1 Een voorbeeld van homologie, waarbij een gegeven populatie A haar eigenschappen +p en +q overdraagt op dochterpopulaties B en C.

Het zou hier bijvoorbeeld kunnen gaan om een populatie die zich opsplijt in twee nieuwe populaties. Als er in de moederpopulatie een (proto)taal met kenmerken +p en +q⁸ gesproken wordt, dan gebeurt dit ook in beide dochterpopulaties. Verbasteringen die na enige tijd in de talen kunnen ontstaan zorgen er mogelijk voor dat de talen steeds meer van elkaar gaan verschillen. Dit proces zou je *taalevolutie* kunnen noemen: klankverschuivingen en syntactische veranderingen die het gevolg zijn van langdurige isolatie. De (mensen)soort zelf kan tijdens deze isolatie ook genetisch veranderen door biologische evolutieprocessen. In het gemakkelijke geval zou taalevolutie met alleen homologische processen te verklaren zijn. Hoewel typologen inderdaad met deze aanname werken, zijn zij zich er van bewust dat er meer aan de hand is. Zo zouden homologische processen verstoord kunnen worden door taalcontact. Taalcontact is een

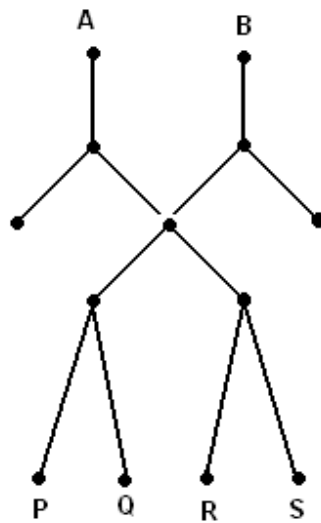
⁸ Het is in de taalkunde gebruikelijk een + te gebruiken wanneer een gegeven typologisch kenmerk aanwezig is in de taal.

vorm van homoplasie, ook wel *convergente evolutie* genoemd (PloSbiology, 2010). Homoplasie is het proces waarbij twee verschillende moederpopulaties voor gemengde dochterpopulaties zorgen. Het is mogelijk dat de mengvorm (i) van typologische aard is, (ii) van genetische aard is of (iii) een combinatie van beide. Met een typologische mengvorm wordt bedoeld dat de talen gesproken binnen de moederpopulaties mengen, en zo een nieuwe taal opleveren. Deze vorm van homoplasie wordt behandeld in hoofdstuk 3. In het geval van genetische mengvormen produceren mensen uit verschillende populaties samen nageslacht. In figuur 2.2 staat homoplasie schematisch weergegeven.



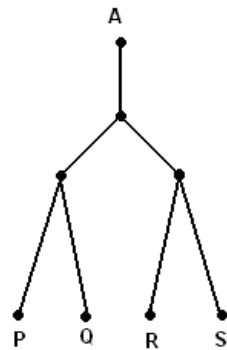
Figuur 2.2 Een voorbeeld van homoplasie, waarbij eigenschappen uit meerdere moederpopulaties worden verenigd in een nieuwe populatie.

Het feit dat zowel homologie als homoplasie altijd kan plaatsvinden heeft gevolgen voor de reconstructie van prototalen. Een simplistische weergave van een mogelijke combinatie van homologie en homoplasie staat in figuur 2.3.

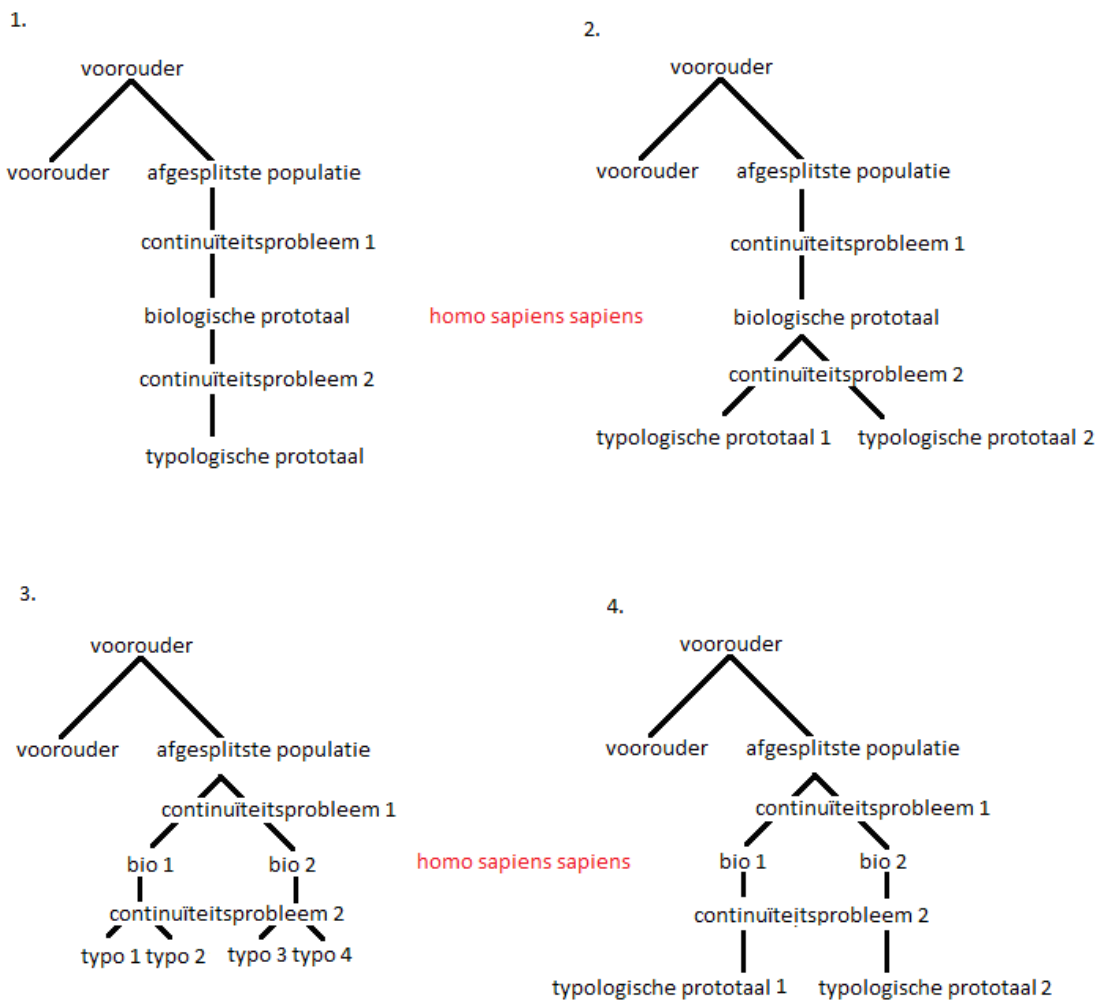


Figuur 2.3 Een mogelijke combinatie van homologie- en homoplasiegestuurde processen.

De letters P, Q, R en S stellen talen voor die ons bekend zijn, en A en B staan voor hypothetische prototalen. Hiervoor geldt altijd dat we slechts in beperkte mate weten in hoeverre deze talen historisch verwant zijn, omdat schrift nu eenmaal een stuk minder oud is (vermoedelijk minimaal 22.000 jaar jonger) dan gesproken taal (Onze Taal, 2004). Historische verwantschap van talen in de periode vóór het schrift is nu eenmaal moeilijk te bewijzen. Figuur 2.3 laat zien dat het zelfs moeilijk is ook maar te speculeren over diepe relaties tussen talen. Onderzoek naar de talen P, Q, R en S zou in eerste instantie namelijk op een enkele prototaal A wijzen, omdat er geen aanwijzingen zijn voor een tweede prototaal. De volgens deze foutieve gedachtegang gereconstrueerde boom is afgebeeld in figuur 2.4.



Figuur 2.4 Een mogelijke foutieve interpretatie van de herkomst van talen P, Q, R en S uit figuur 2.3.



Figuur 2.5 Hypothetische reconstructies van het ontstaan van prototaal. De tijd loopt van boven naar beneden, met in het rood het moment aangeven waarop de mens zich definitief heeft afgezonderd van haar voorouder.

Zelfs in dit eenvoudige model zijn de problemen niet te overzien. In werkelijkheid is het probleem nog veel groter, omdat door homologie- en homoplasie gestuurde processen elkaar mogelijk al duizenden jaren afwisselen. Het nadelige gevolg van deze complexiteit is dat er door alleen naar geschreven taal te kijken geen uitspraken gedaan kunnen worden over de exacte herkomst van talen. In paragraaf 2 worden hiervoor alternatieve methoden besproken. De geïntroduceerde biologische terminologie is nu alvast nuttig om de entree van taal in de menselijke soort mee te beschrijven. De vier scenario's die zojuist in woorden zijn uitgelegd, staan schematisch weergegeven in figuur 2.5. In dit model is de verdere verspreiding van talen achterwege gelaten. Scenario 1 lijkt zeer onwaarschijnlijk. Het model suggereert dat er geen populatiespreiding is geweest sinds het afsplitsen van de voorouder. Dit zou vreemd zijn omdat het hier gaat om een zeer lange periode. Een voordeel is wel dat hierdoor het continuïteitsprobleem minder complex wordt: onverklaarbare evolutionaire veranderingen worden nog onverklaarbaarder als ze op meerdere plekken tegelijk plaatsvonden. Het grootste probleem van scenario 1 is dat het slechts één typologische prototaal impliceert, terwijl hier geen aanwijzingen voor zijn (Singh, et al., 2004). Scenario 3 en 4 vallen eveneens af, omdat zij impliceren dat *Homo sapiens sapiens* zich op twee plaatsen tegelijk definitief afsplitste. Dit is evolutionair gezien moeilijk te verklaren. Scenario 2 blijft dan over als meest plausibele kandidaat. Het grootste minpunt is hier het eerste continuïteitsprobleem: hoe is er in een relatief korte tijd een complex systeem zoals het taalvermogen geëvolueerd? Scenario's 1, 3 en 4 kampen echter met dezelfde vraag. Continuïteitsprobleem 2 wordt eventueel opgelost met rijping. De grootste voordelen van scenario 2 zijn de tegemoetkomingen van ideeën over evolutie en het aantal typologische prototalen. In het geval dat dit model de werkelijkheid inderdaad het beste benadert, komt er een belangrijk inzicht aan het licht: de manier waarop een biologische prototaal verder wordt ontwikkeld is in zekere zin willekeurig. Rijping kan zodoende resulteren in verschillende typologische prototalen. Zelfs is het mogelijk dat rijping niet alleen een willekeurig resultaat oplevert, maar ook in een willekeurige volgorde plaatsvindt. Wellicht verlicht dit inzicht de strijd tussen aanhangers van holistische en synthetische theorieën over prototaal. Nota bene: hoewel het model uitgaat van precies twee typologische prototalen, kunnen het er in werkelijkheid meerdere zijn geweest zonder dat het scenario zijn kracht verliest.

2.1.5 Samenvatting

In deze paragraaf is het verschil uitgelegd tussen biologische prototaal en typologische prototaal. Dit onderscheid bleek vervolgens bruikbaar te zijn om taaltypologische reconstructies te laten rijmen met ideeën over evolutieprocessen. Hoewel er een beeld is geschetst van de manier waarop typologische prototalen mogelijk zijn ontstaan, heeft de introductie van de biologische termen homologie en homoplasie een deur opengedaan naar de diepere werking van taalevolutie, en daarmee de problemen die dit oplevert. In hoofdstuk 3 zullen de oorzaken van homoplasie in taalspreiding en taalevolutie worden besproken, maar het zal niet mogelijk blijken de gevolgen van homoplasie in kaart te brengen bij het reconstrueren van taaltypologische bomen. Een alternatieve oplossing kan gevonden worden in de populatiegenetica.

2.2 Populatiegenetica

Populatiegenetica is de tak van de biologie waarin veranderingen in genetisch materiaal van populaties worden bestudeerd. Kenmerkend voor onderzoek binnen deze discipline is het kwantitatieve en statistische karakter. Om achter genetische verwantschap van diersoorten te komen worden grote hoeveelheden genetisch materiaal vergeleken. Omdat diersoorten onderling geen vruchtbare nakomelingen kunnen produceren, is het effect van homoplasie hier verwaarloosbaar klein.⁹ Dit in tegenstelling tot talen, die continu invloed ondervinden van taalcontact. Hoewel homoplasie door de populatiegenetica dus doorgaans genegeerd wordt bij het reconstrueren van de evolutie van soorten, kan het vruchtbaar zijn modellen uit deze discipline te gebruiken bij het reconstrueren van taalevolutie. Hierover gaat deze paragraaf.

2.2.1 Algoritmische methoden

Modellen uit de populatiegenetica zijn succesvol gebleken om de spreiding van populaties door de geschiedenis heen te beschrijven (Lansing et al., 2007). Zo vertellen gemeenschappelijke allelen van mensen op een directe manier iets over hun vooroudergeschiedenis. Talen hebben deze eigenschappen niet. Typologische

⁹ Homoplasie is niettemin een belangrijk verschijnsel in de genetica, maar heeft alleen grote gevolgen binnen een diersoort zelf. Zo wordt continu genetisch materiaal uitgewisseld wanneer ouders nageslacht produceren. Het zijn voornamelijk uitwisselingen tussen verschillende rassen van dezelfde soort die grote genetische veranderingen opleveren.

overeenkomsten tussen talen hoeven niet te wijzen op een gemeenschappelijke afkomst. De ontwikkeling van taalspreiding kan daarom alleen aan het oppervlak bestudeerd worden. Dit zou geen probleem zijn als er van alle ooit gesproken talen geschriften bestonden. Maar hoewel recent ontdekte geschriften mogelijk 8.000 jaar oud kunnen zijn (Onze Taal, 2004), zijn er duidelijke archeologische aanwijzingen voor dat gesproken taal veel ouder is dan geschreven taal (30.000 tot 200.000 jaar oud, afhankelijk van de bron (Onze Taal, 2004; Aitchison, 1996)). Dit zorgt er samen met de door homoplasie veroorzaakte verstoringen voor dat het moeilijk is hard bewijs te leveren voor de allereerste verspreidingen van talen. In plaats daarvan is ervoor gekozen de geschiedenis van talen zo realistisch mogelijk te benaderen. Populatiegenetici gebruiken voor hun eigen onderzoek twee methoden om historische relaties bloot te leggen: de *maximum parsimony method (MPM)* en de *bayesian maximum likelihood method (MLM)* (Dunn et al., 2008). Het voordeel is dat met deze methoden op zoek gegaan kan worden naar optimale reconstructies, waarbij de effecten van homoplasie zo min mogelijk verstorend zijn.

2.2.2 The Maximal parsimony method

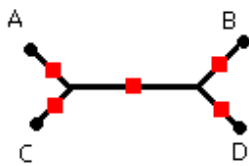
Als aangenomen wordt dat er in totaal twaalf talen zijn ontsproten aan een prototaal als het Proto Indo-Europees, zijn er theoretisch gesproken 13 miljoen manieren te bedenken waarop dit gebeurd kan zijn (Dunn et al., 2008). De volledige reconstructie van 6.000 talen levert een ronduit buitensporig aantal mogelijkheden op. Om vervolgens uit alle potentiële boomstructuren handmatig de juiste te kiezen is een onmogelijke opgave. Daarom zijn er algoritmen ontwikkeld die de boomstructuren aan een aantal eisen onderwerpen. Op deze manier komen de meest voor de hand liggende reconstructies aan het licht. Bij de *maximal parsimony method (MPM)* wordt een algoritme gebruikt dat op zoek gaat naar de boomstructuren met een *maximale parsimonie*. Hiermee wordt bedoeld: de boom met de minste evolutionaire veranderingen. Het nadeel hiervan is dat er mogelijk relaties worden ontdekt die feitelijk niet bestaan, als gevolg van *long branch attraction*, waarbij genetische (of typologische) kenmerken op basis van toevallige overeenkomsten onder een niet bestaande voorouder(taal) worden geschaard. De boom kan op die manier te simplistisch worden.

2.2.3 The bayesian maximum likelihood method

Een alternatieve methode is de maximum likelihood method (MLM). Hiervoor wordt een complexer algoritme gebruikt, dat de boomstructuren aan veel meer beperkingen onderwerpt. Bovendien zijn deze beperkingen niet wiskundig van aard, zoals bij maximale parsimonie, maar (taal)evolutionair gemotiveerd. Op deze manier komen de meest logische boomstructuren aan het licht. Er zijn drie grote nadelen aan deze methode. Ten eerste is het maar de vraag van welke factoren de *likelihood* afhangt. Hiervoor is kennis nodig van antropologische processen en historische gegevens die veelal ontbreken. Ten tweede zijn zelfs de nieuwste supercomputers niet in staat dergelijke berekeningen uit te voeren voor het totale aantal talen in de wereld. Tot slot wordt, zoals eerder genoemd, homoplasie achterwege gelaten, wat hier betekent dat convergerende¹⁰ takken niet mogelijk zijn omdat het aantal theoretisch mogelijke bomen dan niet meer te overzien is. In de praktijk worden de MPM en de MLM door elkaar gebruikt, en hun uitkomsten onderling vergeleken. Uiteindelijk is er voor een enigszins betrouwbare interpretatie dan nog het gezonde verstand van de onderzoeker nodig.

2.2.4 Een boomstructuur reconstrueren

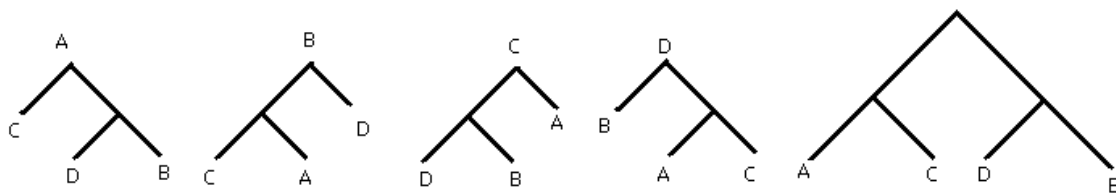
Er is soms sprake van nog een ander probleem bij het reconstrueren van taalfamilies. Om dit uit te leggen volgt nu een voorbeeld van een eenvoudige reconstructie. In figuur 2.6 zijn vier gegeven talen A, B, C en D aangeduid met stippen, en met lijnen zijn de gevonden interne relaties tussen de talen aangegeven.



Figuur 2.6 Een gegeven relatie tussen de vier talen A, B, C en D.

¹⁰ Wanneer alleen homologische vertakkingen worden meegeteld is er sprake van divergerende, ofwel uiteenlopende takken. Door in de reconstructie taalcontact, ofwel homoplasie toe te staan, kunnen takken elkaar oneindig vaak tegenkomen (convergeren) en later weer opnieuw divergeren. Het probleem ligt dan niet zozeer bij de beperkte rekenkracht van computers, maar bij het feit dat er wiskundig gezien een oneindig aantal mogelijke bomen mogelijk is.

Wanneer de figuur als een mobiel voorgesteld wordt, kan deze opgeraapt worden op elk van de vijf punten die zijn aangegeven met rode vierkantjes. In de praktijk is het vaak niet duidelijk welke mogelijkheid de juiste is, omdat de gevonden verbanden weinig verraden over de volgorde waarin opsplitsingen hebben plaatsgevonden. Er is immers te weinig bekend van de leeftijden van talen. In dit simplistische geval zijn er al vijf verschillende reconstructies mogelijk, weergegeven in figuur 2.7.



Figuur 2.7 Vijf mogelijke voorstellingen van verwantschap tussen de talen uit figuur 2.6.

In de empirische werkelijkheid is het aantal geanalyseerde talen veel groter, en daarmee ook de moeilijkheid hiervoor een prototaal te vinden. Wanneer men alle nadelen negeert, beide methoden gebruikt om de meest waarschijnlijke boom te vinden en op rationele wijze naar een vooroudertaal zoekt, komt er een reconstructie tevoorschijn die grotendeels overeenkomt met de veel betrouwbaardere populatiebomen uit de genetica (Seuren, 2009). Dit is een interessant resultaat, maar zegt nog niets over de mogelijke ruis die homoplasie veroorzaakt.

2.2.5 Migratie

Het idee dat homoplasie niet reconstrueerbaar is scheidt, zelfs met toepassing van complexe algoritmen, een somber toekomstbeeld voor de historische taalkunde. Er zijn echter aanwijzingen dat taalcontact in werkelijkheid niet zoveel invloed heeft op taaldiversiteit. Populatiegenetici werken doorgaans met de *isolation by distance hypothesis* (IBD), wat neerkomt op het idee dat populaties meer van elkaar gaan verschillen naarmate de geografische afstand toeneemt als gevolg van migratie. Aanvankelijk werd gedacht dat deze theorie ook zou opgaan voor talen: hoe verder talen met een gedeelde vooroudertaal van elkaar verwijderd raken, hoe meer ze gaan verschillen. De gedachte hierachter is dat talen zelfs zonder contact met andere talen verbasteren van generatie op generatie, en dat daarbij de kans op contact met andere taalgemeenschappen toeneemt naarmate de geografische afstand dit doet. Uit een

grootschalig onderzoek is echter gebleken dat talen zeer sterk aan elkaar verwant blijven, ongeacht het contact met andere taalfamilies of geografische afstand (Holman et al., 2006). Convergerende processen lijken alleen kans van slagen te hebben bij talen die aanvankelijk al verwant waren. Mogelijk zijn cognaten dus een stuk betrouwbaarder dan vaak wordt beweerd!

2.2.6 Samenvatting

In deze paragraaf is uitgelegd welke factoren het reconstrueren van taaltypologische bomen bemoeilijken. Wanneer een computer alle mogelijke bomen modelleert, moet hieruit een selectie gemaakt worden van bomen die de werkelijkheid vermoedelijk het beste benaderen. Hiervoor bestaan twee veelgebruikte algoritmen, die op een verschillende manier de onwaarschijnlijke bomen eruit filteren. Een nadeel hiervan is het feit dat de algoritmen waarschijnlijk gebaseerd zijn op onjuiste aannamen, omdat er te weinig kennis is over de processen die taalevolutie sturen. In het geval dat de juiste voorwaarden die een algoritme hoort te stellen wel bekend zijn, is er vervolgens het probleem dat computers het aantal mogelijke bomen dan mogelijk niet meer kunnen berekenen. Daarnaast wijst verwantschap op zichzelf nog niet op de chronologische volgorde waarin de talen zijn geëvolueerd.

2.3 Samenvatting

Door te kijken naar methoden en aannames die in de biologie gebruikt worden, is duidelijk geworden hoe de tekortkomingen van de comparatieve methode gedeeltelijk opgevangen kunnen worden. Door een onderscheid te introduceren tussen biologische en typologische prototaal komen er meer mogelijkheden aan het licht voor het combineren van tegenstrijdige theorieën over het ontstaan van taalgebruik in de menselijke soort. Hoe de eerste volwaardige taal en de daaraan ontsproten dochtertalen er hebben uitgezien blijft onduidelijk, zelfs na gebruik van geavanceerde statistische technieken. Uiteindelijk is het noodzakelijk meer te weten te komen over het *waarom* van taalevolutie en verspreiding. Het volgende hoofdstuk behandelt daarom de antropologie als aanvullend gereedschap. Hoewel hiermee geen uitputtende verklaring voor de werking van taaldistributie kan worden bereikt, schept deze discipline in elk geval een beeld van de processen die überhaupt van invloed zijn op het gedrag van talen.

3 Linguïstisch antropologische benadering

In de voorgaande hoofdstukken zijn belangrijke factoren besproken die inzicht geven in de werking van taaldistributie. Voor een vollediger overzicht is het nodig om menselijk gedrag in een antropologisch kader te plaatsen. Ten eerste nemen mensen voor een deel actief en intentioneel deel aan taalspreiding. Hun beweegredenen kunnen doorslaggevend zijn voor het ontstaan, veranderen of verdwijnen van een taal. Ten tweede is menselijk gedrag verantwoordelijk voor *taalcontact*, een verschijnsel dat verstoringen veroorzaakt bij het construeren van fylogenetische bomen en daarom de werking van taaldistributie daarom compliceert. De discipline die op deze gebieden de meest bruikbare inzichten kan verschaffen is de linguïstische antropologie. Deze wetenschap bestudeert onder andere taalgemeenschappen, taalcontact en taalvariatie. In dit hoofdstuk wordt bekeken hoe contact tussen taalgemeenschappen kan zorgen voor taalvariatie.

3.1 Frequente uitkomsten van taalcontact

Taalcontact is een verzamelterm voor alle processen die op macroniveau direct verantwoordelijk zijn voor onder andere *taalverschuiving* en *taalverandering*. Met taalverschuiving wordt de volledige vervanging van een taal door een andere bedoeld. Taalverandering is het proces waarbij een taal van uiterlijk veranderd door invloeden van andere talen, opdat er een *contacttaal* ontstaat. In werkelijkheid zijn taalverschuiving en taalverandering zeldzame uitkomsten van taalcontact. In deze paragraaf worden de meest voorkomende gevolgen van taalcontact besproken.

3.1.1 *Barrières voor contacttalen*

Allereerst hangt de mogelijkheid van een contacttaal af van de intentie van de betrokken taalgemeenschappen. Hiermee wordt bedoeld dat zij zich minstens moeten openstellen voor verandering. Ten tweede is de periode van taalcontact van invloed. Hoe langer twee gemeenschappen sociaal contact onderhouden, hoe groter de kans is dat er taalverandering optreedt. Ten derde is de machtsverhouding tussen de taalgemeenschappen van belang. Een dominante gemeenschap wil haar taal misschien letterlijk opleggen. De ondergeschikte gemeenschap kan zich hier vervolgens al dan niet tegen verzetten. In een egalitair systeem zijn ook meerdere uitkomsten mogelijk. Het

streven naar gelijkheid kan zorgen voor respect voor elkaars gesproken talen, waardoor deze mogelijk in originele staat behouden blijven. Tot slot zijn de motivaties voor taalverandering van belang. Er kan in een land bijvoorbeeld de behoefte bestaan aan een universele eerste taal ter bevordering van de communicatie tussen gemeenschappen. Aan de andere kant is het mogelijk dat gemeenschappen zich juist graag identificeren met hun taal. Al met al is een contacttaal slechts een uitzonderlijke uitkomst van taalcontact (Duranti, 2004). Er zijn teveel variabelen van invloed op de kans van slagen. Er zal eerder sprake zijn van verschijnselen als *multilinguïsme*, *taalverschuiving* en *diglossie*, die nu zullen worden besproken.

3.1.2 Multilinguïsme

De meest voorkomende resultaten van taalcontact zijn bilinguïsme en multilinguïsme. In deze situaties worden er in een taalgemeenschap twee of meerdere talen gesproken. Extreme voorbeelden hiervan worden gevonden in het Amazonegebied. Hier bestaan gemeenschappen waarin niet minder dan 20 talen gesproken worden (Duranti, A., 2004). Omwille van de identiteit van de bewoners worden de unieke eigenschappen van de talen streng bewaakt. Om toch onderling te kunnen communiceren wordt een *lingua franca* gebruikt, ofwel één taal die iedereen spreekt.

3.1.3 Taalverschuiving

Tegenover multilinguïsme, waarbij behoud van de taal een grote rol speelt, staat taalverschuiving, waarbij juist niets wordt behouden van de oorspronkelijke taal. Een dergelijke verschuiving naar monolinguisme kan erg abrupt zijn. Taalverschuiving komt steeds vaker voor. Macroantropologische factoren die hierbij een rol spelen zijn industrialisatie, urbanisatie, migratie en globalisering (Duranti, 2004). Het meest van invloed op taalverschuiving is hier globalisering, omdat dit verschijnsel zorgt voor een toename van behoefte aan internationale communicatie, en de verdwijning van geïsoleerde (taal)gemeenschappen. De verwachting dat er op korte termijn nog slechts 5% van de hedendaags gesproken talen zullen bestaan is hierop gebaseerd (Asher, 1995). Hoewel sommige culturele veranderingen dus succesvol voor een taalverschuiving zorgen, is dit lang niet altijd het geval. Waar het in Indonesië is gelukt een taal die slechts door een minderheid gesproken werd succesvol te introduceren als nationale taal, faalt een dergelijk beleid structureel in bijvoorbeeld Singapore, waar Engels al decennialang de voertaal op school is, terwijl men daar thuis meestal

Mandarijn spreekt (Asher, 1995). Kennelijk is de toekomst van een taal onvoorspelbaar en complex. Om het verschijnsel echt te doorgronden zouden microantropologische factoren dus moeten worden bekeken. Zo is het interessant te onderzoeken welke redenen gemeenschappen precies hebben om al dan niet afstand te doen van hun moedertaal. Autoritaire druk of een collectieve behoefte aan modernisering zijn voorbeelden van dergelijke factoren.

3.1.4 Diglossie

Een diglossie is een gemeenschap waar een “hoge” en een “lage” variant van een taal wordt gesproken. De hoge variant wordt gebruikt voor formele gesprekken, en de lage voor informele. In Arabische landen is dit een gebruikelijke situatie, evenals in enkele gekoloniseerde gebieden, waar de Europese talen als hoog beschouwd worden, en de zogenaamde creolentalen, welke verderop nauwkeuriger worden besproken, als laag (Duranti, 2004). Omdat de keuze voor het spreken van een hoge of een lage variant van een taal afhangt van de sociale context is dit verschijnsel van sociolinguïstische aard in plaats van linguïstisch antropologische aard.

3.1.5 Samenvatting

Tot nu toe zijn de meest voorkomende gevolgen van taalcontact in kaart gebracht. De invloed van diglossie en multilinguale gemeenschappen op taalspreiding lijkt niet zo groot. Eigenschappen van talen blijven hier juist goed behouden. Interessanter is taalverschuiving, waarbij een taal volledig verdwijnt. Andere ingrijpende producten van taalcontact zijn contacttalen, die in de volgende paragraaf besproken worden.

3.2 Contacttalen

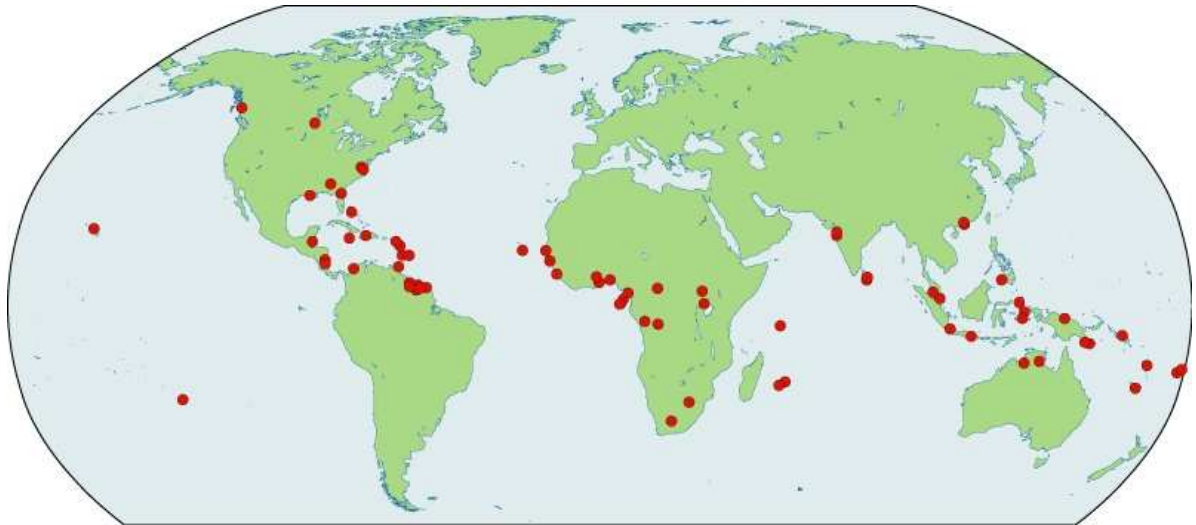
Alleen in uitzonderlijke gevallen resulteert taalcontact in een contacttaal. Omdat contacttalen interessant zijn voor deze scriptie wordt hier toch uitgebreid op ingegaan. Bekende vormen van contacttalen zijn *pidgintalen* en *creolentalen*. Er is echter onenigheid over de betekenis van deze termen. Daarom zijn er pogingen gedaan om meerdere typen talen te onderscheiden die deze termen dekken. Voorbeelden hiervan zijn *semicreolen*, *creoloïden*, *mengtalen*, *vervlochten talen*, *bilinguale mengvormen* en *aangeboren variëteiten*. Op slechts enkele termen wordt hier verder ingegaan.

3.2.1 Mengtalen

Mensen uit bilinguale taalgemeenschappen spreken twee talen vloeiend. Zij zijn in staat deze twee talen te combineren voor zover de genoemde sociale factoren dit toelaten. De taal die dan ontstaat heet een mengtaal of een vervlochten taal. In meer of mindere mate vindt dit op taalkundig niveau altijd op dezelfde manier plaats: het flexiesysteem en het syntactische systeem van één taal gaan samen met het lexicon van de andere taal. Een volledige mix ontstaat echter zelden. Het Media Quechua, een taal gesproken in Ecuador, is hier een uitzonderlijk voorbeeld van. Deze taal heeft het flexiesysteem en het syntactische systeem uit het Quechua (de daar oorspronkelijke gesproken taal) volledig behouden, maar bevat alleen maar Spaanse woordstammen (Duranti, A., 2004). Het ontstaan van mengtalen komt voornamelijk voor in gekoloniseerde gebieden. De kolonisten leren in dit geval de inheemse taal volledig, en de inheemse bevolking leert de taal van de kolonisten. Het mengen gebeurt pas nadat beide gemeenschappen de vreemde taal volledig zijn gaan beheersen.

3.2.2 Pidgintalen en creolentalen

Een pidgintaal – we laten de exacte definitie achterwege - is eveneens een mogelijk gevolg van kolonisatie. In dit geval komen twee of meerdere voor elkaar vreemde gemeenschappen elkaar tegen, die vervolgens proberen te communiceren om bijvoorbeeld de handel te bevorderen. Door de wederzijdse afhankelijkheid van de betrokkenen ontwikkelen pidgintalen zich snel. Kinderen die in een dergelijke “linguïstische chaos” opgroeien zijn in staat de pidgintaal, in feite een handen-en-voeten-taal, verder te ontwikkelen tot een volwaardige taal. Zo'n taal wordt een creolentaal genoemd. Het idee is dat kinderen de pidgintaal alleen maar hoeven “op te vullen” met aangeboren kennis die ze van taal hebben. Na de eerder besproken *kritische periode* blijken kinderen niet meer in staat een beroep te doen op deze kennis (Duranti, 2004). In figuur 3.1 staat een overzicht van creolentalen in de wereld.



Figuur 3.1 De creolentalen van de wereld (APICS, 2009)

Op de kaart is goed te zien dat de creolentalen vooral voorkomen in gekoloniseerde gebieden. Op deze plekken kwamen soms slaven uit verschillende taalgebieden naast elkaar te werken onder de kolonisten. Zoals we zagen is een dergelijke linguïstische chaos een bron voor creolentalen. Hoewel de meeste creolentalen zijn uitgestorven is er nog een tiental over. Creolisering (het ontstaan van creolentalen) kan er samenvattend voor zorgen dat talen min of meer vanuit het niets ontstaan. Hiaten in het rijpingsproces van kinderen resulteren mogelijk in talen die typologische weinig meer te maken hebben met de oorspronkelijk betrokken talen. In zekere zin kunnen creolentalen dus een eigen taalfamilie creëren. Een concreet voorbeeld stamt uit de jaren tachtig van de 20^e eeuw, waar er in Nicaragua vanuit niets een volwaardige gebarentaal is ontstaan waarin alle grammaticale domeinen (zelfs het fonologische domein!) vertegenwoordigd zijn (Senghas et al., 2004). Hoewel de divergerende werking van taalcontact misschien nauwelijks zorgt voor mengtalen, kan het dus wel een nieuwe stamtaal produceren. Omdat het onduidelijk is of er oudere taalfamilies bestaan die ontsproten zijn aan een creolentaal, stelt taalcontact taalkundigen mogelijk toch voor een probleem.

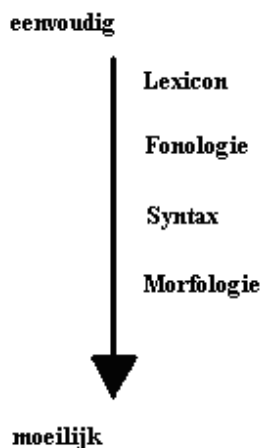
3.2.3 Samenvatting

Mengtalen, creolentalen en isolatietalen zijn de bekendste vormen van contacttalen. Bij het ontstaan van deze talen spelen migratie en kolonisatie een grote rol. De behandeling van creolentalen heeft belangrijke inzichten opgeleverd in taalevolutie, omdat de rol van

taalcontact in typologische reconstructies een nieuwe betekenis heeft gekregen. De daadwerkelijke transitie van een taal naar een andere is op taalkundig niveau nog niet besproken. Daarover gaat de volgende paragraaf.

3.3 Microlinguïstische factoren

Als je wilt begrijpen hoe taalcontact werkt op microniveau spelen vooral taalkundige processen een rol. Linguïst Sarah Grey Thomason heeft hiervoor “the borrowing scale” bedacht (Thomason 2001).



Figuur 3.2 The Borrowing Scale (Thomason 2001)

Hieruit is af te lezen dat het lexicon het gevoeligst is voor verandering, en morfologie (hiermee worden voornamelijk flexiemorfemen bedoeld) het minst gevoelig. Een proces waarin deze hiërarchie niet lijkt op te gaan is *taalconvergentie*. Hierbij blijven de lexicons van de betrokken talen behouden, terwijl de structuren assimileren. Taalconvergentie is bijvoorbeeld verantwoordelijk voor het ontstaan van *Sprachbünde*: taalgebieden waar talen historisch niet direct verwant zijn, maar structureel-typologisch wel (Duranti, 2004). Een voorbeeld hiervan is de Balkan Sprachbund. In figuur 3.3 staan voorbeeldzinnen in enkele talen uit dit taalgebied. Het valt direct op dat er amper lexicale, maar wel structurele overeenkomsten zijn. Ter vergelijking is de Duitse vertaling gegeven, waarin de gebruikte volgorde een onwelgevoemd resultaat geeft.

<i>Taal</i>	<i>Datief</i>	<i>Genitief</i>
<i>Duits</i>	* Ich habe das Buch gegeben an Marie	Das Buch ist Mary's
Albanees	Ja dhashë librin Marisë .	Është libri i Marisë .
Bulgaars	Dadoh knigata na marija	Knigata e na Marija
Roemeens	I-am dat cartea Mariei	Este cartea Mariei
Macedonisch	Ì ja dadov knigata na Marija	Knigata e na Marija
Grieks	Édōsa to biblío stīn maría	Eínai to biblío tīs marías

Figuur 3.3 Voorbeeldzinnen uit de Balkan Sprachbund (Wikipedia, 2010)

Het behoud van het lexicon bij de assimilatie van talen is vaak het resultaat van de behoefte aan een eigen identiteit van de sprekers. De assimilatie van structurele eigenschappen ontstaat door de behoefte aan een betere leerbaarheid van een tweede taal. Zo zou het voor Nederlanders bijvoorbeeld gunstig zijn hun moedertaal syntactisch aan te passen aan het Engels om de leerbaarheid van deze taal te vergroten. Deze sociale factoren zijn in dit geval dus dominantanter dan de linguïstische motivatie om de structuur te behouden.

3.4 Samenvatting

Waar in hoofdstuk 1 en 2 voornamelijk is gesproken over het *hoe* van taalgedrag, ging het in hoofdstuk 3 meer over het *waarom*, ofwel over de oorzaken van taalevolutie en taalspreiding. Hiervoor was het belangrijk een onderscheid te maken tussen taalverandering en taalverschuiving. Er is ingezoomd op het moment waarin de ene taal verandert in de andere, en welke contacttalen hierdoor kunnen ontstaan. Tot slot is behandeld in welke opzichten talen stug zijn en in welke opzichten juist flexibel. Helaas zijn de problemen die door homoplasie worden veroorzaakt nog steeds niet opgelost, en het is de vraag of hier ooit verandering in komt.

4 Samenvatting

Het doel van deze scriptie was om een zo volledig mogelijk beeld te schetsen van de werking van taaldistributie. Het bleek dat de taalkunde, de wetenschap die de werking van taal bestudeert, tekortschiet om dit doel te bereiken. Vanuit taalkundig perspectief kunnen theoretische grenzen worden gedefinieerd waarbinnen taal kan variëren, en kunnen er classificaties van de waargenomen taaltypen worden gemaakt. Bij het classificeren komt in zekere mate historische verwantschap tussen talen aan het licht, en daarmee tekenen hypothetische taalfamilies zich af. Hoe en waarom mensen vervolgens taal verspreiden is hiermee nog niet duidelijk. Daarom is vervolgens een beroep gedaan op de biologie. Vanuit theorieën over prototaal en ideeën over populatiegenetica is een theoretisch kader geschetst waarbinnen taal zich kan verspreiden, en dienden zich alternatieve oplossingen aan om verwantschap tussen talen nauwkeurig te reconstrueren in plaats van de standaard comparatieve methoden uit de taaltypologie. Omdat het gedrag van talen uiteindelijk afhangt van het gedrag van mensen, bleek het nuttig de distributie van talen ook nog door het oog van de antropoloog te bekijken. Hier kwamen menselijke motieven aan het licht die verantwoordelijk zijn voor taalverschuiving, taalverandering, en andere mogelijke gevolgen van taalcontact. Hoewel er dus vanuit allerlei invalshoeken naar de werking van taaldistributie is gekeken, is er nog geen poging gedaan deze inzichten te integreren in een overkoepelend model. Dit zal worden ondernomen in het laatste hoofdstuk. Hoewel deze scriptie nooit uitputtend kan zijn, is het belangrijk nogmaals te stipuleren dat er mogelijk andere factoren een rol spelen bij het gedrag van talen. Zo kan de kans die een taal heeft om te blijven bestaan afhankelijk zijn van demografische groei en de mate van politieke complexiteit (Currie et al., 2008), en kan taaldiversiteit samenhangen met biodiversiteit en geografische of klimatologische diversiteit. Deze relaties zijn belangrijk om te onderzoeken omdat ze dieper ingaan op de verklaring van taaldistributie, ter aanvulling van de meer beschrijvende kracht van de behandelde invalshoeken.

5 Integratie

Binnen de behandelde disciplines zijn bepaalde theorieën en concepten naar voren gekomen die nodig waren om taal te analyseren vanuit de betreffende invalshoek. Dit heeft per discipline de nodige inzichten opgeleverd in het gedrag van talen. Om meer uit al deze inzichten te halen dan de som der delen, moeten zij gesmeed worden tot één overkoepelend inzicht. De eerste stap in dit integratieproces bestaat uit het identificeren van tegenstrijdigheden in theorieën en hun basisassumpties, en de inzichten die ze opleveren. Onderzoekers die vanuit verschillende wetenschappen een onderwerp benaderen lopen tegen verschillende problemen aan. Het is zowel mogelijk dat ze menen over verschillende onderwerpen te praten maar hetzelfde bedoelen als andersom. Eveneens kan het zijn dat er in feite appels met peren vergeleken worden, doordat de basisassumpties van beide onderzoekers verschillend zijn. Het is daarom noodzakelijk om dergelijke conflicten aan het licht te brengen en op te lossen. Vervolgens moet er naar overeenkomsten gezocht worden tussen inzichten die aanvankelijk verschillend leken. Waar twee meestal meer weten dan één, probeert de interdisciplinaire onderzoeker meer te weten te komen dan twee. De inzichten die uiteindelijk hergedefinieerd en wel overblijven, moeten worden geïntegreerd in een overkoepelend model.

5.1 Conflicterende Begrippen

In deze paragraaf komen een aantal begrippen aan bod die vanuit meerdere disciplines beschreven worden. Om overzichtelijk te maken wat deze begrippen per discipline¹¹ inhouden en vanuit welke theorie ze benaderd worden, zijn ze hier samengevat in tabellen. Dit is noodzakelijk omdat het hier begrippen betreft die vaak voor meerdere zaken gebruikt kunnen worden. Er wordt steeds geprobeerd de benaderingen van het begrip samen te vatten in een breder concept, of om het begrip juist op te splitsen in

¹¹ Er zal vanaf nu gewerkt worden vanuit de vijf disciplines generatieve taalkunde, historische taalkunde, evolutiebiologie, populatiegenetica en linguïstische antropologie. Deze opsplitsing is nog niet eerder expliciet gemaakt omdat er vaak slechts metaforische parallellen bestaan tussen de betreffende discipline en taaldistributie, waarbij er theorieën uit de discipline worden geleend voor nieuwe doeleinden. Een hoofdstuk *evolutiebiologie* zou vermoedelijk verkeerde verwachtingen scheppen.

meerdere begrippen om de conflicten tussen begrippen op te lossen¹². Deze stappen zijn ter bevordering van het integratieproces.

5.1.1 Taalvermogen

In figuur 5.1 wordt het *taalvermogen* bekeken. De generatieve taalkunde gebruikt dit begrip om het abstracte systeem aan te duiden dat onze taaluitingen stuurt. Dit wordt benaderd vanuit het *Y-model*, waarin de veronderstelde domeinen van het systeem worden gerepresenteerd. Deze benadering van het begrip is relevant om universele beperkingen op taalvariatie in kaart te brengen.

Taalvermogen	Benadering van het concept	Theorie
Generatieve taalkunde	Nadruk op <u>capaciteit</u> taalvermogen	Y-model
Evolutiebiologie	Nadruk op <u>anatomie</u> taalvermogen	Prototaal
Antropologie	Nadruk op <u>prioriteiten</u> taalvermogen	Borrowing scale

Figuur 5.1 Verschillende benaderingen van het taalvermogen

Ook de evolutiebiologie maakt gebruik van het begrip, maar legt de focus meer op de concrete fysiologische kenmerken van het taalvermogen en de evolutie ervan. Een belangrijke aanname in theorieën over deze evolutie is het bestaan van *prototaal*. De benadering van het taalvermogen vanuit prototaal is relevant voor dit onderzoek, omdat ze inzichten verschaft in de ontstaansmogelijkheden van taal, en daarmee in eventuele valkuilen van taalreconstructies. Tot slot wordt het taalvermogen behandeld in de antropologie. In de zoektocht naar motivaties van mensen om veranderingen in taal teweeg te brengen is de *borrowing scale* aan het licht gekomen. Het taalvermogen wordt binnen deze theorie net als in de generatieve taalkunde benaderd als een abstract systeem, alleen ligt de focus hier op de eis van het vermogen om bepaalde domeinen eerder intact te houden dan andere. Deze benadering dankt haar relevantie aan de inzichten die het oplevert in de kansen die een taal heeft om te overleven, sterven of veranderen, en daarmee in een gedeelte van taaldistributie. Het ligt voor de hand om het begrip *taalvermogen* hier op te splitsen in twee begrippen: *abstract taalvermogen* en *biologisch taalvermogen*. Het abstracte taalvermogen is dan niet meer dan een model dat de werking van taal in het brein weergeeft in termen van competenties, zonder dat

¹² Merk op dat het opsplitsen van een begrip al is gedaan in hoofdstuk 1 en 2, waar noodzakelijk onderscheid is gemaakt tussen biologische en typologische prototaal ten behoeve van de leesbaarheid.

hier fysiologische structuren en hersenprocessen bij betrokken worden. Zowel de generatieve taalkunde als de antropologie kan met deze definitie uit de voeten, zonder dat het begrip zijn verklarende werking verliest. Het biologische taalvermogen aan de andere kant, stelt de verzameling lichamelijke structuren voor die de taaluiting mogelijk maken. Hieronder vallen zowel spraakorganen zoals het strottenhoofd en de stembanden, articulatoren zoals de mondholte en de tong, en cognitieve processen die verantwoordelijk zijn voor het vormen van concepten en het aansturen van de spraakorganen¹³. Om een plaats te krijgen in een interdisciplinair model, moeten de nieuwe begrippen nog gedefinieerd worden in termen van taaldistributie. Dit is vrij eenvoudig. Het abstracte taalvermogen is nodig om te verklaren waarom er universele restricties lijken te bestaan op variatie in talen, en waarom het dus mogelijk is dat talen soms op elkaar lijken zonder dat ze aan elkaar verwant zijn. Kennis van het biologische taalvermogen is belangrijk om theorieën over prototaal te ondersteunen, en daarmee de ontwikkeling van taal te onderzoeken. Dit levert namelijk inzichten over verspreidingsmogelijkheden van taal op. Het zal in de volgende paragraaf blijken dat het ook nuttig kan zijn het biologische en het abstracte taalvermogen samen te nemen om sommige processen te beschrijven. Voor deze gevallen definiëren we het algemene taalvermogen nu als volgt: de set cognitieve en fysieke processen die samen verantwoordelijk zijn voor het totale scala aan betekenisdragende klanken.¹⁴

Begrippen:

- Biologisch taalvermogen
- Abstract taalvermogen

5.1.2 Taalevolutie

Een ander complex begrip is *taalevolutie*. In paragraaf 2.1 is deze term kort geïntroduceerd als het proces waarbij talen door de tijd heen veranderen, en dan specifiek het langzaam uit elkaar groeien van verwante talen. Veranderingen in taal zijn in de voorgaande hoofdstukken echter voortdurend behandeld op steeds andere

¹³ Op de werking van deze organen wordt verder niet ingegaan, omdat dat niet relevant is om de evolutie van het taalvermogen te begrijpen.

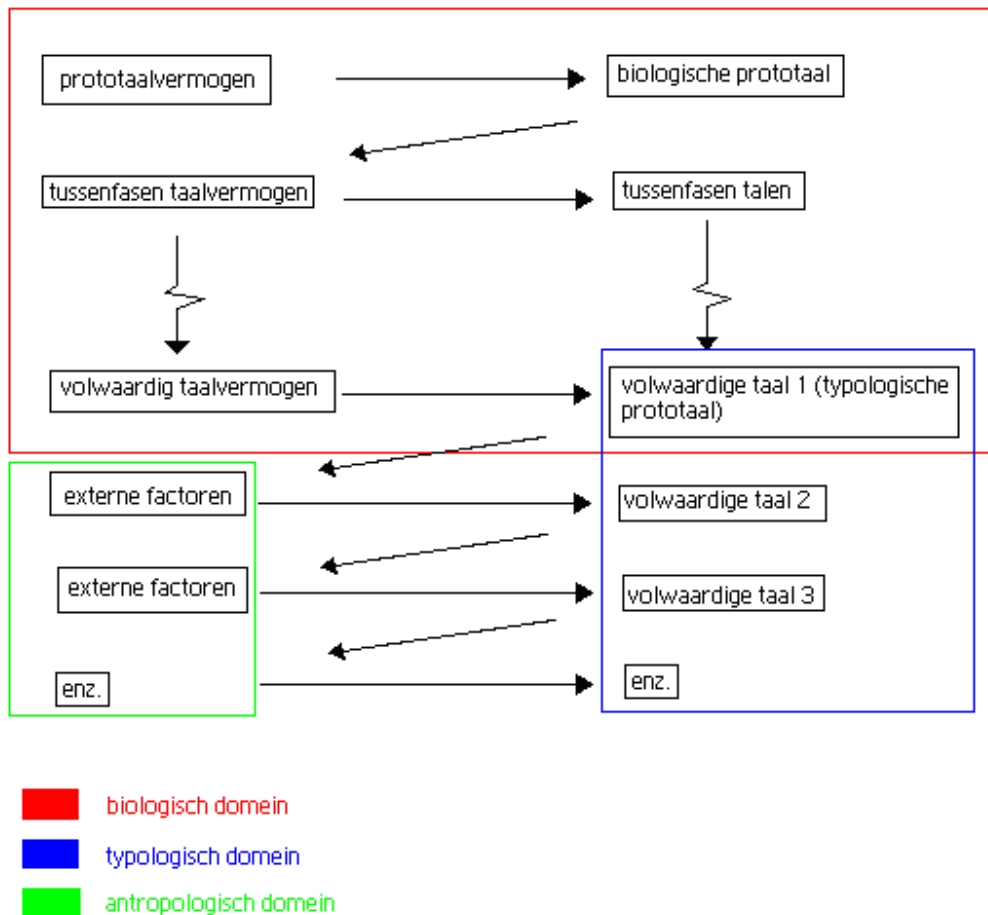
¹⁴ Non-verbale taaluitingen zoals handgebaren vallen ook onder “klanken”. Gebarentalen zijn net als gesproken talen volledig uitgerust met een syntactisch, morfologisch, semantisch en fonologisch systeem.

manieren, zoals te zien is in figuur 5.2. Het zal nuttig blijken om meerdere termen te scharen onder taalevolutie, en andere juist duidelijk buiten taalevolutie om te definiëren.

Taalevolutie	Benadering van het concept	Theorie
<i>Generatieve taalkunde</i>	Nadruk op ontwikkeling taaluitingen	Rijping
<i>Evolutiebiologie</i>	Nadruk op ontwikkeling fysiologie taalvermogen	! Prototaal
<i>Historische taalkunde</i>	Nadruk op verspreiding talen	! Prototaal TCM GCM
<i>Antropologie</i>	Nadruk op principes verantwoordelijk voor taalverandering	Culturele motivaties

Figuur 5.2 Verschillende benaderingen van taalevolutie

Vanuit de generatieve taalkunde wordt er gesproken van *rijping* als de ontwikkeling van het taalvermogen *binnen* het individu. Dit idee biedt een potentiële oplossing voor een van de continuïteitsproblemen, en levert tevens inzichten op over het spontaan ontstaan van talen. Dit is vervolgens weer belangrijk om problemen zichtbaar te maken bij het reconstrueren van taalfamilies. De evolutiebiologie houdt zich ook bezig met de ontwikkeling van het taalvermogen, maar richt zich juist voornamelijk op de structurele veranderingen in de soort: hoe verandert het taalvermogen van generatie op generatie? Dit is noodzakelijk om de beginselen van het taalvermogen te doorgronden, en daarmee theorieën over prototaal te vormen. Prototaal wordt weer gebruikt om aansluiting te zoeken bij theorieën over de verdere ontwikkeling van talen, welke gevormd worden vanuit de historische taalkunde. Ongelukkig genoeg wordt er in deze laatste discipline ook over prototaal gesproken, terwijl er iets anders mee wordt bedoeld. Dit is aangegeven met uitroeptekens in figuur 5.2. De historische taalkunde houdt zich namelijk niet bezig met de ontwikkeling van het taalvermogen, maar met de ontwikkeling van taal van generatie op generatie. Prototaal wordt hier gebruikt om de eerste volwaardige taal aan te duiden die ooit gesproken is. Hoewel het onderscheid tussen de concepten al is gemaakt in hoofdstuk 1 en 2 met de introductie van biologische prototaal en typologische prototaal, is in figuur 5.3 nog wat duidelijker weergegeven hoe deze begrippen zich tot elkaar verhouden.



Figuur 5.3 Het verschil tussen biologische prototaal en typologische prototaal

Aan de linkerkant van de figuur staan factoren die van invloed zijn (geweest) op het uiterlijk van talen. Allereerst zijn klankvormen volledig afhankelijk van het taalvermogen dat ze produceert. In het begin was dit taalvermogen – in de figuur het *prototaalvermogen* genoemd - nog onontwikkeld, wat betekende dat de bijbehorende taaluitingen ook onontwikkeld waren. De ontwikkeling van de taal zelf is steeds aan de rechterkant van de figuur afgebeeld. Uiteindelijk, met een onbekend aantal tussenfasen, ontwikkelde het taalvermogen zich tot het volwaardige systeem dat wij nu kennen. Op dit punt stopt de interesse van evolutiebiologen. Dit is in de figuur aangegeven met het rode kader. De interesse van historische taalkundigen is hier zojuist begonnen, omdat het volwaardige taalvermogen een of meerdere volwaardige talen als gevolg heeft gehad. Dit is aangegeven met blauw. Omdat de ontwikkeling van het taalvermogen hier is gestopt, kunnen alleen externe factoren nog van invloed zijn op het uiterlijk van talen.

Aangezien het menselijk gedrag de grootste factor is, zijn het vooral antropologen die deze externe factoren onderzoeken. Dit is aangegeven in het groen. Het reconstrueren van oertalen en hun dochtertalen gebeurt vervolgens met de GCM en de TCM. In het kader van taaldistributie zijn typologische prototaal, de GCM en de TCM essentieel om de ontwikkeling van taal, en daarmee de verspreiding van taal te onderzoeken. De invloed van mensen op de evolutie van taal wordt door antropologen tot slot onderzocht aan de hand van culturele motivaties. Het gaat hier vaak om de ontwikkeling van taal *binnen één generatie*, waarbij er wordt ingezoomd om het daadwerkelijke moment van verandering. Om een model van taaldistributie enige voorspellende waarde te geven zijn antropologische theorieën over taalverandering het meest essentieel. We hebben samenvattend te maken met vier soorten taalevolutie. Deze zijn in een schema gezet, waarin valt af te lezen wat er per discipline onder taalevolutie wordt verstaan.

Verandering:	<i>Microniveau</i>	<i>Macroniveau</i>
<i>Taal</i>	Antropologie	Historische taalkunde
<i>Taalvermogen</i>	Generatieve taalkunde	Evolutiebiologie

Figuur 5.4 Verschillende benaderingen van verandering.

De meest werkbare oplossing voor deze definitiekwestie is het maken van een onderscheid op basis van onderzoeksgebied in plaats van het niveau waarop dit onderzoek plaatsvindt. Dit onderscheid is tevens conform figuur 5.3, waar de ontwikkeling van taal en de ontwikkeling van het taalvermogen als twee gescheiden trajecten zijn weergegeven. In hoofdstuk 1 en 2 is eerder het onderscheid gemaakt tussen biologische en typologische prototaal. Deze terminologie kan nu weer gebruikt worden om het onderscheid tussen de twee soorten taalevolutie te maken. We definiëren *typologische taalevolutie* nu als volgt: het proces waaronder zowel de daadwerkelijke momenten van transitie in typologische kenmerken van taal vallen, als de grote waarneembare tendens waarmee taal zich langs de tijdas heeft ontwikkeld. Dit is precies het traject dat onderzocht wordt door de antropologie en de historische taalkunde. *Biologische taalevolutie* is dan: het proces waaronder alle vormen van ontwikkeling in groei en expressie van het taalvermogen vallen. Dit dekt zowel de betekenis die het taalvermogen heeft voor de generatieve taalkunde als voor de evolutiebiologie.

Begrippen:

- Typologische taalevolutie
- Biologische taalevolutie

5.1.3 Taaldiversiteit

Het laatste begrip waar veel verwarring over kan ontstaan is *taaldiversiteit*. We hebben namelijk gezien dat talen in meerdere opzichten van elkaar kunnen verschillen. Dit is een probleem omdat het verschil tussen talen een van de weinige variabelen is waarmee je kunt meten hoe talen gedistribueerd zijn over de wereld. In figuur 5.5 staat kort weergegeven waar de disciplines het accent op leggen als het gaat om taaldiversiteit.

Taaldiversiteit	Benadering van het concept	Theorie
<i>Historische taalkunde</i>	Nadruk op aantal talen en taalfamilies	TCM GCM
<i>Generatieve taalkunde</i>	Nadruk op cognitieve beperkingen taaldiversiteit	Y-model
<i>Populatiegenetica</i>	Nadruk op wiskundige beperkingen van taaldiversiteit	MPM MLM
<i>Antropologie</i>	Nadruk op producten die taalevolutie oplevert	Culturele voorwaarden

Figuur 5.5 Taaldiversiteit vanuit de betrokken disciplines

De betekenis van taaldiversiteit lijkt hetzelfde voor zowel de generatieve taalkunde, de historische taalkunde en de populatiegenetica: de mate waarin talen in lexicaal en structureel opzicht van elkaar verschillen. Toch is het in werkelijkheid genuanceerder. De historische taalkunde kijkt om precies te zijn naar de verdeling van *taaltypen* in de wereld, in de zoektocht naar reconstructies van talen, taalfamilies en oertalen. Om die taaltypen te kunnen onderscheiden gebruiken ze meetinstrumenten zoals de TCM en de GCM, die zijn ontwikkeld op basis van kennis uit de generatieve taalkunde. Deze laatste discipline onderzoekt namelijk hoe talen wel en niet van elkaar kunnen verschillen, en zorgt er zo voor dat historische taalkundigen weten waar ze naar moeten zoeken. Taaldiversiteit betekent voor generatieve taalkundigen dus: de (mogelijke) verscheidenheid aan taaltypen in de wereld. Voor historische taalkundigen wordt dit dan: de mate waarin de taaltypen daadwerkelijk voorkomen, en de manier waarop deze taaltypen zijn gedistribueerd over de wereld. De populatiegenetica zit hier ergens

tussenin. Aan de ene kant levert deze discipline net als de generatieve taalkunde kennis over restricties op taaldiversiteit aan de historische taalkunde, op basis van wiskundige modellen. De definitie van taaldiversiteit lijkt echter meer op die van historische taalkundigen, omdat de nadruk ligt op de daadwerkelijke distributie van taaltypen. De drie behandelde benaderingen leveren geen groot conflict op over wat taaldiversiteit *is*, maar meer over de manier hoe je taaldiversiteit een plek geeft in theorieën over taaldistributie. Voor de antropologie is het een ander verhaal. Hoewel ook antropologen verschillen in talen meten op grond van typologische kenmerken, is dit niet de manier waarop zij talen uiteindelijk classificeren. Antropologen houden zich typisch bezig met de manier waarop talen op microniveau veranderen, en welk type mengvorm, taalgeboorte of taalsterfte dit oplevert. Dit betekent dus ook dat *taaltype* een heel andere betekenis krijgt, namelijk: het resultaat van het betreffende type taalevolutie, waarbij dit type afhankelijk is van de processen die het evolutieproces in werking hebben doen treden, of de manier waarop het evolutieproces op microniveau is verlopen. Waar historische taalkundigen spreken van taaltypen zoals *agglutinerende talen* en *polisynthetische talen* (talen met een rijk flexiesysteem), spreken antropologen van bijvoorbeeld *mengtalen* en *vervlochten talen* op basis van de manier waarop ze zijn veranderd, en van *pidgintalen* en *creolentalen* op basis van de manier waarop het veranderingsproces is geïnitieerd. In feite kan taaldiversiteit binnen de antropologie dus twee dingen betekenen: (i) de variatie in processen die taalevolutie in gang kunnen zetten en (ii) de variatie in eindproducten van deze processen. De term *distributie* wordt hier met opzet niet gebruikt, omdat antropologen hier in de eerste plaats niet in geïnteresseerd zijn. De relevantie van benadering (i) en (ii) in het kader van taaldistributie zit hem in de inzichten die ze kunnen verschaffen voor onderzoekers die wel in taaldistributie geïnteresseerd zijn. Samenvattend zitten we nu met meerdere definities van taaldiversiteit, en daar moet wat aan gedaan worden. Het woord *diversiteit* heeft gevoelsmatig een meer kwantitatieve betekenis, tegenover de meer kwalitatieve betekenis van bijvoorbeeld *variatie*. Zo lijkt diversiteit te gaan over de daadwerkelijk aangetroffen verscheidenheid, en variatie over de aard van die verscheidenheid. Dit inzicht kunnen we gebruiken bij het herdefiniëren van de termen. *Taaldiversiteit* heeft in dit geval betrekking op de historische taalkunde en de evolutiebiologie, omdat zij de daadwerkelijk waargenomen verscheidenheid proberen te beschrijven. *Taalvariatie* hoort dan bij de generatieve taalkunde, die zich meer bezighoudt met de kwalitatieve eigenschappen van verscheidenheid. De twee antropologische benaderingen zijn beide van zowel kwalitatieve als kwantitatieve aard,

en worden daarom nu geschaard onder de overkoepelende term *verscheidenheid*. Om toch onderscheid te kunnen maken tussen de twee benaderingen wordt de term nu opgesplitst in (i) *linguïstisch antropologische verscheidenheid*, waarbij het gaat om de verscheidenheid aan antropologische processen, en (ii) *antropologisch linguïstische verscheidenheid* om het scala aan typen talen te beschrijven op basis van de antropologische processen. De nieuwe terminologie is in figuur 5.6 in tabelvorm gegoten:

Discipline	Term
<i>Historische taalkunde</i>	taaldiversiteit
<i>Generatieve taalkunde</i>	taalvariatie
<i>Evolutiebiologie</i>	taaldiversiteit
<i>Antropologie</i>	(i) linguïstisch antropologische verscheidenheid (ii) antropologisch linguïstische verscheidenheid

Figuur 5.6 Diverse benaderingen van de verscheidenheid aan taaltypen

Er is in hoofdstuk 3 ook nog kort sprake geweest van een sociolinguïstische classificatie van talen, op basis van de sociale context waarin talen worden gebruikt. Voorbeelden hiervan waren *hoge* en *lage talen*. Dit perspectief op taaldiversiteit laten we verder echter achterwege.

Begrippen:

- Taalvariatie
- Taaldiversiteit
- Linguïstisch antropologische verscheidenheid
- Antropologisch linguïstische verscheidenheid

5.1.4 Taal distributie

Distributie heeft gewoonlijk te maken met de uitdeling van zekere elementen. In de vorige paragraaf bleek dat het niet eenvoudig is om deze elementen eenduidig te classificeren als het gaat om het brede begrip *taal*. Hoewel de onderzoeksvraag zich richt op *taal distributie*, ligt het dus niet voor de hand wat taal distributie eigenlijk is. Dit

doet af aan de kwaliteit van de onderzoeksvraag en de doelstelling. De vergaarde begrippen stellen ons nu echter in staat om taaldistributie concreet te definiëren:

De waargenomen taaldiversiteit, taalvariatie en antropologisch linguïstische verscheidenheid, ofwel de manier waarop taal zowel kwalitatief als kwantitatief varieert op de wereld, waarbij de kwalitatieve classificatie opgesplitst kan worden in een typologische en een antropologische classificatie.

5.1.5 Samenvatting

In deze paragraaf is een poging gedaan een einde te maken aan het rommelige gebruik van de betreffende terminologie. Dit heeft een helder overzicht van inzichten opgeleverd waar deze scriptie kennelijk rijk aan is, en vormt daarom een belangrijke eerste stap in het integratieproces.

5.2 Factoren en inzichten

In dit onderzoek naar factoren die van invloed zijn op taaldistributie is een groot aantal inzichten naar boven gekomen. In de vorige paragraaf is er waar nodig dieper ingegaan op deze inzichten om ze te vormen tot bruikbare elementen. Er volgt in deze paragraaf per discipline een overzicht van alle verworven hapklare stukken informatie die straks in het interdisciplinaire model verwerkt kunnen worden. Hierbij zullen de factoren de losse bouwstenen van het model vormen. De losse inzichten worden gebruikt om de architectuur van het model, ofwel de manier waarop de bouwstenen worden samengevoegd, vast te stellen.

5.2.1 Generatieve taalkunde

Figuur 5.7 is eenvoudig af te lezen. De *factoren* behoren tot de elementen die in het uiteindelijke model opgenomen moeten worden. Met *invloed* wordt de manier bedoeld waarop de factor gevolgen heeft voor taaldistributie. De *losse inzichten* kunnen belangrijk zijn om structuur aan te brengen in de uiteindelijk lange lijst multidisciplinaire factoren. Binnen de generatieve taalkunde is het belangrijkste inzicht dat taalvariatie niet mogelijk zou zijn zonder taalverwervende kinderen, omdat zij als enige in staat zijn optimaal gebruik te maken van de speelruimte die het taalvermogen biedt. Deze speelruimte wordt bepaald door principes en parameters.

Generatieve taalkunde	<i>Factoren</i>	<i>Invloed</i>
<i>Losse inzichten</i>	Generatieve principes: morphologische fonologische semantische syntactische	Ze beperken taalvariatie in theoretische mogelijkheden
Het kind is het belangrijkste medium	Generatieve parameters: morphologische fonologische syntactische	Ze scheppen theoretische mogelijkheden voor taalvariatie
	Optimaliteit	
	Rijping	Verantwoordelijk voor biologische taalevolutie op microniveau

Figuur 5.7 generatief taalkundige invloeden op taaldistributie

5.2.2 Historische taalkunde

De historische taalkunde leert dat zowel lexicale als syntactische gelijkenissen tussen talen verraderlijk kunnen zijn, vanwege de mogelijke willekeur van syntactische parameterafstellingen en het voorkomen van leenwoorden als gevolg van taalcontact. Daarbij komt nog eens dat historisch vergelijkend onderzoek afhankelijk is van geschreven taal, welke veel jonger is dan gesproken taal. De daadwerkelijke verstoring die taalcontact oplevert kun je niet meten zonder kennis van antropologische processen.

Historische taalkunde	<i>Factoren</i>	<i>Invloed</i>
<i>Losse Inzichten</i>	Stugheid van syntactische eigenschappen	Talen neigen syntactisch amper te evolueren
Geschreven taal is te jong voor volledige reconstructies	Willekeur syntactische eigenschappen	Syntactische overeenkomsten kunnen op toeval duiden
De bijdrage van taalcontact kun je alleen onderzoeken vanuit antropologische theorieën	Taalcontact	Leenwoorden zijn in staat niet bestaande taalfamilies te voorspellen

Figuur 5.8 historisch taalkundige invloeden op taaldistributie

5.2.3 Evolutiebiologie

Omdat evolutie zich op alle niveaus afspeelt levert de discipline ook meerdere soorten inzichten op. Op het niveau van het individu speelt rijping een grote rol. Dit vermogen stelt het individu in staat om met geringe input een taal te ontwikkelen, en biedt daarom mogelijke verklaringen voor de continuïteitsproblemen. Toch is er over rijping te weinig bekend, waardoor theorieën over biologische prototaal elkaar blijven tegenspreken. Op het niveau van de soort zijn er ook onduidelijkheden, omdat de verspreiding van taal zowel convergent als divergent kan verlopen.

Evolutiebiologie	<i>Factoren</i>	<i>Invloed</i>
<i>Losse Inzichten</i>	Rijping	Taal is in staat spontaan te ontstaan
De eigenschappen van rijping zijn mogelijk onvoldoende om het continuïteitsprobleem op te lossen	Homologie	Lexicale overeenkomsten wijzen mogelijk op verwantschap
Theorieën over biologische prototaal in de zoektocht naar een oertaal zijn tegenstrijdig	Convergente taalevolutie (homoplasie)	Talen zijn in staat hun oorsprong onzichtbaar te maken

Figuur 5.9 evolutiebiologische invloeden op taaldistributie

5.2.4 Populatiegenetica

Door talen te benaderen als genenpakketten is het mogelijk gebleken om vrij nauwkeurige reconstructies te maken van taalfamilies die verder teruggaan dan geschreven taal. Hoewel hier aan het licht gekomen is dat onverwante talen praktisch immuun zijn voor taalcontact kan de invloed van volksmigraties op taal niet genegeerd worden. Bij gebrek aan kennis van het aantal historische volksmigraties is het dus moeilijk te bepalen welke van de mogelijke algoritmen de juiste is voor een betrouwbare reconstructie.

Populatiegenetica	<i>Factoren</i>	<i>Invloed</i>
<i>Losse Inzichten</i>	Onverwante talen zijn praktisch immuun voor taalcontact	Het effect van homoplasie is slechts beperkt verstrend
Kennis over migratie is onvoldoende om de verspreiding van talen mee te achterhalen	Computers hebben een beperkte rekenkracht.	Het is moeilijk om betrouwbare reconstructies te maken
	Algoritmen hebben te weinig relevante input	

Figuur 5.10 populatiegenetische parallellen met taaldistributie

5.2.5 Antropologie

De belangrijkste mechanismen die de ontwikkeling van taal sturen zijn van antropologische aard. Tegelijk zijn deze factoren het minst begrepen vanwege hun complexe aard. Zo is het onduidelijk in welke context een gemeenschap zal neigen naar het behoud van de eigen taal of het overstappen op een andere taal. De behoefte aan economische progressie kan zorgen voor de behoefte aan een lingua franca, terwijl de behoefte aan identiteit juist voor een toename in talen kan zorgen. Toch is er een mondiale afname in het aantal talen waargenomen, die mogelijk grotendeels te maken heeft met globalisering. Het ontstaan van nieuwe talen is echter nog steeds mogelijk, en gebeurt dan via creolisering. Dit proces heeft een redelijk voorspellende waarde vanwege de betrouwbaarheid van de borrowing scale.

Antropologie	<i>Factoren</i>	<i>Invloed</i>
<i>Losse inzichten</i>		
de complexiteit van cultuur zorgt voor een lage voorspellende waarde	behoefte aan eigen identiteit	behoud diversiteit
	behoefte aan soepele communicatie	verdwijning diversiteit
	globalisering	verdwijning diversiteit
	linguïstische chaos	creolisering
	borrowing scale	sommige eigenschappen stugger dan andere

Figuur 5.11 antropologische invloeden op taaldistributie

5.3 Het model

Er zijn meerdere manieren om aan de hand van de gevonden factoren een overkoepelend model van taaldistributie te construeren. In het ideale geval is het model in staat op ieder gewenst moment in het verleden of de toekomst een beeld te schetsen van de manier waarop talen verdeeld zijn over de wereld. Het intelligente algoritme dat hiervoor nodig is moet vol zitten met volledige informatie over het gedrag van mensen, de omgevingsfactoren op ieder moment, en de interactie tussen deze variabelen. Kort gezegd is dit een onmogelijke opgave. Een betere manier is het in kaart brengen van zo veel mogelijk factoren die, onafhankelijk van de tijd, een rol zouden kunnen spelen bij de verspreiding van taal. Dit sluit daarbij ook veel beter aan op de onderzoeksvraag. Het mag inmiddels duidelijk zijn dat het gedrag van talen zeer complex is, en dat er geen enkele wetenschappelijke discipline bestaat van waaruit er zelfstandig uitspraken over dit gedrag gedaan kunnen worden. Laat het model een handig stuk gereedschap zijn dat de onderzoeker dwingt het gedrag van talen altijd in een interdisciplinair kader te plaatsen. Unesco heeft reeds getracht een voorspelling te doen over de ontwikkeling van taal in de nabije toekomst, en deed daarbij de voorspelling dat een derde van alle talen deze eeuw nog zal verdwijnen. De besproken omvang van de set factoren die invloed hebben op het aantal talen geeft ons echter reden een kritische houding aan te nemen tegenover deze voorspelling.

5.3.1 Overzicht

Figuur 5.12 biedt een overzicht van de factoren die tot nu toe voorbij zijn gekomen, gesorteerd op discipline waaruit het inzicht is geput en het niveau waarop de factor van toepassing is. Zo wordt de factor *immunititeit voor taalcontact* slechts mogelijk gemaakt door culturele processen die het individu alleen niet kan initiëren. Verder is er ook een onderscheid gemaakt tussen *constanten* en *variabelen*. Constanten zijn hier de factoren die continu dezelfde invloed uitoefenen op taaldistributie. Variabelen kunnen al dan niet aanwezig zijn. In de rechterkolommen staat per factor aangegeven wat voor invloed deze heeft op het aantal talen in de wereld (taaldiversiteit) en op het kader waarbinnen typologische variatie (taalvariatie) kan plaatsvinden. In het grijs staan factoren waarvan de invloed op taaldistributie nog niet voldoende is onderzocht. Zij vormen vanzelfsprekend slechts een kleine indicatie van andere spectra waarlangs taaldistributie onderzocht kan worden.

	Factoren			Invloed	
	Onderzoeksniveau	Discipline	Inzichten	Aantal talen	Typologische variatie
Constanten	<i>Individu</i>	<i>Generatieve taalkunde</i>	principes	neutraal	beperkend
		<i>Historische Taalkunde</i>	willekeur syntactische eigenschappen stugheid syntactische eigenschappen	neutraal neutraal	verrijkend beperkend
	<i>Evolutiebiologie</i>	rijping	neutraal	verrijkend	
	<i>Antropologie</i>	Borrowing scale	neutraal	beperkend	
	<i>Groep</i>	<i>Historische taalkunde</i>	taalcontact	neemt toe/af/blijft neutraal	verrijkend/beperkend, neutraal
<i>Evolutiebiologie</i>		homologie	neutraal/neemt toe	verrijkend/neutraal	
<i>Populatiegenetica</i>		homoplasie immunititeit voor taalcontact	neutraal/neemt toe neutraal	beperkend/neutraal beperkend	
	<i>Externe factoren</i>	<i>Antropologie</i>	Globalisering	neemt af	neutraal
Variabelen	<i>Individu</i>	<i>Generatieve taalkunde</i>	parameters optimaliteit	neutraal neutraal	verrijkend verrijkend
		<i>Antropologie</i>	Behoeftte aan communicatie	neemt af	neutraal
	<i>Groep</i>	<i>Antropologie</i>	Behoeftte aan identiteit	neemt toe	neutraal
			Behoeftte aan communicatie	neemt af	neutraal
	<i>Externe factoren</i>	<i>Antropologie</i>	Linguïstische chaos	neutraal	verrijkend
complex politiek klimaat			neemt af	neutraal	
eenvoudig politiek klimaat			neemt toe	neutraal	
<i>Overige disciplines</i>		isolatie taalgemeenschap	neutraal	verrijkend	
		dicht bij evenaar	neemt toe	neutraal	
		ver van evenaar af	neemt af	neutraal	
		vlak landschap	neemt af	neutraal	
		gebergte	neemt toe	neutraal	

Figuur 5.12 Een overzicht van inzichten en hun invloed op taaldistributie, gesorteerd op discipline en onderzoeksniveau

In figuur 5.13 zijn alle factoren opnieuw gesorteerd, deze keer op hun invloed op taaldistributie. Deze nieuwe indeling is bedoeld als laatste stap voordat de inzichten in een interdisciplinair model kunnen worden opgenomen. We weten nu welke factoren er welke invloed hebben, en of deze invloed constant of variabel is.

Aantal talen	Invloed	Factoren
	Neemt toe	Behoeftte aan identiteit taalcontact (homologie) eenvoudig politiek klimaat dicht bij evenaar gebergte
	Neemt af	taalcontact (homoplasie) Behoeftte aan communicatie Globalisering complex politiek klimaat vlak landschap ver van evenaar af
Typologische variatie	Wordt beperkt	Generatieve principes stugheid syntactische eigenschappen Borrowing scale immuniteit voor taalcontact taalcontact
	Wordt verrijkt	willekeur syntactische eigenschappen rijping parameters optimaliteit Linguïstische chaos taalcontact (homologie/homoplasie) isolatie taalgemeenschap

Figuur 5.13 Alle factoren gesorteerd op hun invloed op taaldistributie

De belangrijkste informatie uit figuur 5.12 en 5.13 is samengevat in figuur 5.14.

	Diversiteit neemt toe	Diversiteit neemt af	Variatie wordt beperkt	Variatie wordt verrijkt
Constant	Taalcontact (homologie)	Taalcontact (homoplasie) Globalisering	Generatieve principes stugheid syntactische eigenschappen Borrowing scale immunititeit voor taalcontact taalcontact	willekeur syntactische eigenschappen rijping taalcontact
Variabel	Behoeftte aan identiteit eenvoudig politiek klimaat dicht bij evenaar gebergte	Behoeftte aan communicatie complex politiek klimaat vlak landschap ver van evenaar af		parameters optimaliteit Linguïstische chaos isolatie taalgemeenschap

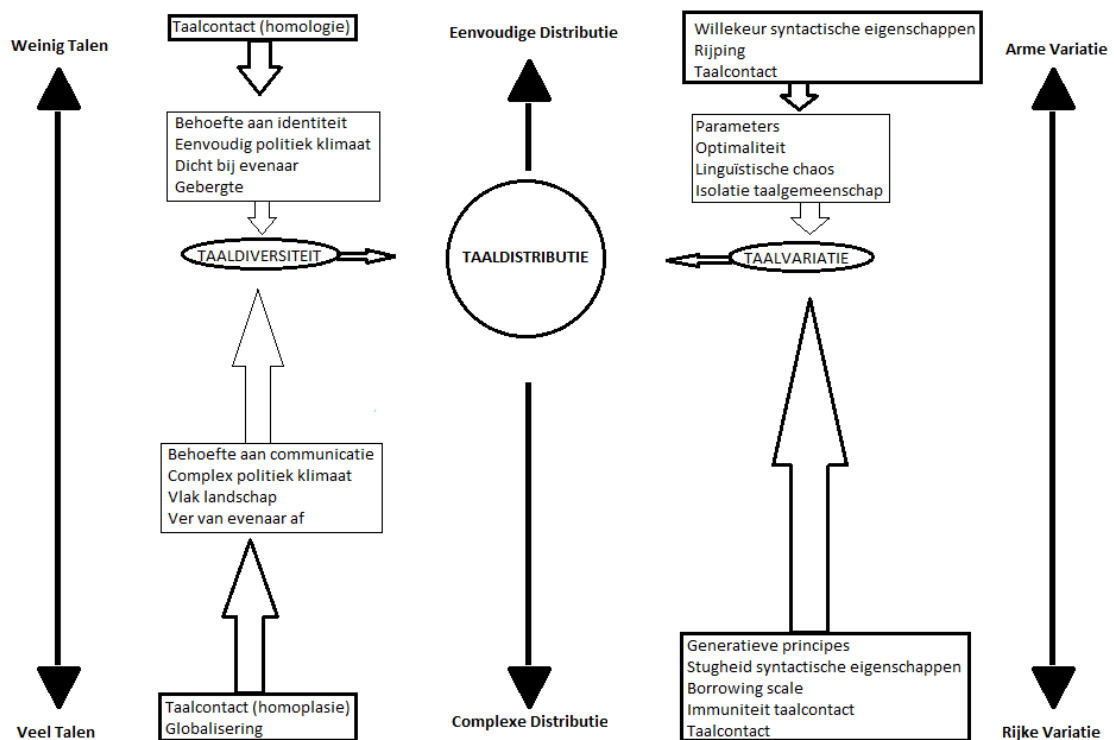
Figuur 5.14 Een totaaloverzicht van de factoren

Er is uit dit onderzoek niet gebleken hoe de factoren zich kwantitatief bezien tot elkaar verhouden. In het definitieve model kan er dus slechts onderscheid gemaakt worden tussen de factoren die constant invloed uitoefenen en de factoren die contextafhankelijk zijn. Verder is er figuur 5.14 geen onderscheid gemaakt tussen taaltypologische- en antropologische taalvariatie, terwijl de gegeven definitie van taaldistributie wel op dit onderscheid doelt. De reden hiervoor is dat taal weliswaar in zowel taaltypologisch als in antropologisch opzicht kan variëren, maar dat deze variatie zich op verschillende niveaus afspeelt die elkaar niet beïnvloeden. Zo worden taaltypologische mogelijkheden an sich niet uitgebreid door de manier waarop een contacttaal kan ontstaan en andersom.

5.3.2 Een overkoepelend model

Nu er een compleet overzicht is van de invloed die alle genoemde factoren hebben op taaldistributie kan deze informatie verwerkt worden in een model. Het model moet in staat zijn in één oogopslag een totaalbeeld te schetsen van de verkregen inzichten, en daarmee in de werking van taalvariatie en taaldiversiteit. Een voorbeeld van een dergelijk model staat in figuur 5.15, waar taaldistributie wordt uitgebeeld als het aantal talen en de waargenomen cross-linguïstische variatie waaraan de wereld rijk is. Een rijke taalvariatie staat dan voor een groot aantal ver verwante taalfamilies, en een arme variatie voor een klein aantal nauw verwante taalfamilies. De complexiteit van de taaldistributie hangt af van de optelsom van het aantal talen en de mate van taalvariatie. In het statisch ogende model lijkt het alsof het aantal talen per definitie gelijk is aan de mate van taalvariatie, maar dit hoeft niet zo te zijn. Zo kan er in theorie één taalfamilie bestaan met duizenden nauw verwante talen, of honderden taalfamilies met ieder

slechts één taal. Voor het gemak hebben de variabelen in dit voorbeeld echter dezelfde waarde gekregen. De reden dat ze beide boven de helft zijn afgebeeld is dat er op dit moment sprake lijkt te zijn van een stijgende vereenvoudiging van taaldistributie, in het model te zien aan de lengte van de pijlen. De dikte van de pijlen en de kaders waaraan ze ontspruiten varieert tussen dik en dun, wat staat voor respectievelijk continue invloed en variabele (contextafhankelijke) invloed.



Figuur 5.15 Het definitieve interdisciplinaire model van mondiale taaldistributie

5.4 Voor toekomstig onderzoek

Er is reden om te zeggen dat de nadruk in dit hele werkstuk heeft gelegen op toekomstig onderzoek. Er is toegewerkt naar een model met een theoretisch onderbouwde architectuur die zich leent voor onderzoek langs het hele wetenschappelijke spectrum. Om het model alvast een inhoudelijke waarde te geven is het ingevuld met informatie uit de belangrijkste onderzoeksgebieden als het gaat om taaldistributie. De grootste

bottleneck ligt misschien wel bij de ruis die het conflict tussen leenwoorden en cognaten oplevert als gevolg van onbegrip van antropologische processen. Inzichten in de wisselwerking tussen taal en cultuur bieden daarom mogelijk de vruchtbaarste toevoegingen aan het model. Een concreet voorbeeld hiervan is globalisering: een recente trend waarvan de oorzaken en gevolgen nog nauwelijks bekend zijn. In een later stadium is het van belang om andere onderzoeksgebieden bij het probleem te betrekken zoals klimatologie, geografie en economie. Het mag duidelijk zijn dat de complexiteit van taaldistributie veruit het belangrijkste inzicht is van deze scriptie. De lijst met invalshoeken van waaruit het probleem benadert kan worden is praktisch oneindig, en dit maakt het dan ook tot een blijvend interessant interdisciplinair onderwerp. Laat deze scriptie bijdragen aan de groeiende populariteit van taalonderzoek, en mensen bewuster maken van de indrukwekkende wereld achter taal en taalgebruik.

Referenties

Artikelen

Ackema, P., Hoop, H. de, *Current Issues in Optimality Theoretic Syntax* (2006), Linguistics.

Aikhenvald, A., *Language Contact in Amazonia* (2007), Language.

Asher, R.E., *The Rise and Fall of Language* (1995), Geographical 67, 18.

Bermúdez-Otero, R., *Stratal Optimality Theory* (2005), Oxford University Press.

Blust, R., *The Position of the Languages of Eastern Indonesia: A Reply to Donohue and Grimes* (2009), Oceanic Linguistics.

Cavalli-Sforza, L.L. et al., *Reconstruction of human evolution: Bringing together genetic, archaeological, and linguistic data* (1988), Department of Genetics, Stanford University.

Currie, T.E. et al., *Political complexity predicts the spread of ethnolinguistic groups* (2008), PNAS

Dunn, M. et al., *Structural Phylogeny in Historical Linguistics: Methodological Explorations Applied in Island Melanesia* (2008), Max Planck Institute.

Dutt, Ashok K. et al., *Spatial Pattern of Languages in India: A Culture-Historical Analysis* (1985), Reidel Publishing Company.

Evans, N., *Australian Languages Reconsidered: A Review of Dixon* (2005), Oceanic Linguistics.

Fitch, W.T., *The evolution of language: a comparative review* (2005), School of Psychology, University of St. Andrews.

Holman, E.W. et al., *On the relation between structural diversity and geographical distance among languages: observations and computer simulations* (2006), Department of Psychology, University of California.

- Hunley, K. et al., *Genetic and Linguistic Coevolution in Northern Island Melanesia* (2008), University of Chicago.
- Kennedy, C., *Comparative deletion and optimality in syntax* (2002), Kluwer Academic Publishers
- Lansing, J.S. et al., *Coevolution of languages and genes on the island of Sumba, eastern Indonesia* (2007), Princeton University
- Lehmann, W.P., *A Definition of Proto-Germanic: A Study in the Chronological Delimitation of Languages* (2010), Linguistic Society of America.
- Lenneberg, E., *New Directions in the Study of Language* (1964), M.I.T.Press.
- Manni, F. et al., *Genetic Structures and Linguistic Boundaries in Italy: A Microregional Approach* (2001), Human Biology.
- Marris, E., *The Language Barrier* (2008), Nature Publishing Group.
- Mortarino, C., *An improved statistical test for historical linguistics* (2007), *Statistic Methods & Applications* 18-2.
- Pawley, A. et al., *Austronesian Historical Linguistics and Culture History* (1993), *Annual Reviews Anthropology*.
- Reesink, G. et al., *Explaining the Linguistic Diversity of Sahul Using Population Models* (2009), Centre for Language Studies, Radboud University.
- Schulze, C. et al., *Monte Carlo simulation of the rise and the fall of languages* (2004), Institute for theoretical Physics, Cologne University.
- Senghas, A. et al., *Children Creating Core Properties of Language: Evidence from an Emerging Sign Language in Nicaragua* (2004), *Science* 305, 1779-1782.
- Serva, M., *Indo-European languages tree by Levenshtein Distance* (2007), *Europhysics Letters*
- Silva, E.J.S. et al., *Evolution of the linguistic diversity on correlated landscapes* (2008), Departamento de Estatística e informática, Universidade Federal Rural de Pernambuco.
- Speelman, D. et al., *Profile-Based Linguistic Uniformity as a Generic Method for Comparing Language Varieties* (2003), Kluwer Academic Publishers
- Tallerman, M., *Did our ancestors speak a holistic protolanguage?* (2005), School of English Literature, Language and Linguistics, University of Newcastle upon Tyne.

Walker, L.D., *A Note on Historical Linguistics and Marc Bloch's Comparative Method* (1980), Backwell Publishing for Wesleyan University.

Zanette, D.H., *Demographic Growth and the Distribution of Language Sizes* (2008), International Journal of Modern Physics.

Boeken

Aitchison, J., *The Seeds of Speech* (1996), Cambridge University Press.

Ashcraft, M.H., *Cognition, Fourth Edition* (2006), Pearson Education International.

Baker, M.C., *The Atoms of Language* (2001), Basic Books

Barbiers, S., *Syntactische Atlas van de Nederlandse Dialecten* (2006), Amsterdam University Press.

Boeckx, C., *Language in Cognition* (2010), Wiley-Blackwell.

Carnie, A., *Syntax* (2002), Blackwell Publishers.

Chomsky, N., *Syntactic Structures* (1957), Mouton & Co.

Chomsky, N., *The Minimalist Program* (1995), MIT Press.

Don, J., et al., *Inleiding in de Generatieve Morfologie* (1994), Coutinho.

Duranti, A., *A Companion to Linguistic Anthropology* (2004), Blackwell Publishing.

Halliburton, R., *Introduction to Population Genetics* (2004), Pearson Education.

Kerstens, J., *Plato's Probleem* (1997), Coutinho.

Klein, H.G., Stegmann, T.D., *EuroComRom: Die Sieben Siebe* (2000), Amsterdam University Press.

Kooij, J., Oostendorp, M. van, *Fonologie* (2003), Amsterdam University Press.

Pinker, S., *The Language Instinct* (1995), Penguin.

Saeed, J.I., *Semantics* (2008), John Wiley and Sons.

Seuren, P.A.M., *Language in Cognition* (2009), Oxford University Press.

Singh, R.S. et al., *The Evolution of Population Biology* (2004), Cambridge University Press.

Trask, R.L., *History of Basque* (1996), Routledge.

Zlatev, J. et al., *Studies in Language and Cognition* (2009), Cambridge Scholars Publishing.

Websites

APICS, *The Atlas of Pidgin and Creole Language Structures*, 2009.

Baker, M.C., *Linguistic differences and language design* (2003), Elsevier
http://www.sciencedirect.com.proxy.library.uu.nl/science?_ob=ArticleURL&_udi=B6VH9-49209CW-1&_user=3021449&_coverDate=08%2F31%2F2003&_rdoc=1&_fmt=high&_orig=search&_sort=d&_docanchor=&_view=c&_acct=C000021878&_version=1&_urlVersion=0&_userid=3021449&_md5=7462108be574bf5cb8e314114da74250

Onze Taal, <http://www.onzetaal.nl/kalender/records/r0504.php>, 2004

Oostendorp, M. van., *Basisbegrippen van de taalwetenschap: Taalverandering en taalcontact* (2004), Universiteit Leiden.
www.meertens.knaw.nl/medewerkers/.../taalverandering.pdf

Public Library of Science Biology (2010)
www.PloSbiology.org

UNESCO, *Interactive Atlas of the World's Languages in Danger* (2009)
<http://www.unesco.org/culture/ich/index.php?pg=00206>

Wikipedia, *Verspreiding van niet-Indo-Europese talen in Europa*, 2008