



Radboud University Nijmegen

Cladistische methoden in de taalwetenschap en de wet van Dollo

Lukas Reinarz
s4216202

Eerste lezer: Prof. Dr. Helen de Hoop
Tweede lezer : Prof. Dr. Pieter Muysken

Instelling: Radboud Universiteit Nijmegen
Academisch jaar 2014-2015

Inhoudsopgave

1	Inleiding	1
2	Cladogrammen in de taalwetenschap	2
2.1	Taalverandering en evolutie	2
2.2	Cladogrammen in de taalwetenschap	5
3	Cladogrammen in de biologie	7
3.1	Fylogenetische methoden in de biologie	7
3.2	De wet van Dollo	12
4	Taal als organisme	14
4.1	Taalkundige cycli	14
4.2	Talige kenmerken als fenotypen	17
4.3	Terug naar het begin: cladogrammen in de taalwetenschap	20
5	Conclusie	21
	Literatuur	23

1 Inleiding

Taalverandering en taalverwantschap zijn een veel bestudeerd onderwerp in de taalwetenschap. Er wordt geprobeerd om kenmerken te identificeren die talen met elkaar delen om vervolgens te kunnen vaststellen in welke mate talen met elkaar verwant zijn. Sommige talen blijken een heel nauwe verwantschapsrelatie te onderhouden - zoals het Nederlands en het Duits -, andere talen zijn helemaal niet verwant, bijvoorbeeld het Fins en het Quechua.

Om vast te stellen hoe talen met elkaar verwant zijn en welke verwantschapsrelaties ze met elkaar onderhouden, worden er methoden gebruikt die het doel hebben om verwantschapsbomen aan te leggen die de mate van verwantschap tussen talen en de afstamming van talen tonen. Deze bomen worden gemaakt op basis van evolutiebiologische methoden.

Omdat het in de evolutiebiologie er ook over gaat om vast te stellen welke soorten aan elkaar verwant zijn en uit welke oersoorten ze zijn voortgekomen, zijn er methoden bedacht die mensen in staat stellen om op basis van genetische kenmerken van soorten te achterhalen of soorten een gemeenschappelijke voorouder hebben en hoever ze van elkaar af staan.

Deze methoden worden toegepast om ditzelfde doel te bereiken voor talen. Zo presenteren Dunn, Greenhill, Levinson en Gray (2011) cladogrammen, dat wil zeggen stambomen, van vier grote taalfamilies (de Austronesische, Indo-Europese, Uto-Aztekische en de Bantutalen) die gebaseerd zijn op woordvolgorde. Ze zeggen dat

[by d]rawing on the powerful methods developed in evolutionary biology, we can [...] track correlated changes during the historical processes of language evolution as languages split and diversify.
(Dunn et al., 2011, p. 79)

Ze geven dus aan dat biologische methoden geschikt zijn om taalkundige bomen te maken.

Er zijn echter taalkundige fenomenen die we in de evolutie van soorten slechts beperkt zien zoals cycli. Taalveranderingen kunnen op een cyclische manier verlopen. Volgens van Gelderen (2009) wil dat zeggen dat een talig kenmerk zoals een woord of een frase vervangen wordt door een ander talig kenmerk,

maar dat het patroon van de verandering steeds weer terugkeert. Met betrekking tot cyclische verandering is er een wet in de biologie geformuleerd, de *wet van Dollo*, die zegt dat alleen kenmerken die *niet* in de genen verankerd zitten en die beïnvloed kunnen worden door externe factoren cyclisch kunnen veranderen. Voor andere, genetisch bepaalde kenmerken geldt dit niet.

Het probleem dat zich voordoet is dat er in de evolutiebiologie gewerkt wordt met genetische kenmerken, zogeheten *genotypen*, en die kunnen volgens de wet van Dollo niet cyclisch veranderen. In taal daarentegen worden voor het maken van cladogrammen géén genotypische kenmerken gebruikt, maar *fenotypen*. Deze kenmerken kunnen op een cyclische wijze veranderen wat volgens dezelfde wet wijst op door externe omstandigheden beïnvloedbare kenmerken.

In de biologie worden dus genetische kenmerken gebruikt voor het maken van stambomen en in de taalwetenschap niet-genetische kenmerken. Omdat deze echter ook onder de invloed van externe factoren kunnen variëren, kunnen ze niet betrouwbaar zijn als men wil achterhalen welke talen aan elkaar verwant zijn. Biologische methoden worden dus ten onrechte toegepast op taal.

In deze scriptie zal ik deze hypothese uitwerken. Het resultaat zal erop neerkomen dat onder andere de bomen die door Dunn et al. (2011) zijn gemaakt geen betrouwbare afspiegeling kunnen zijn van taalverwantschapsrelaties. In de volgende sectie leg ik uit hoe cladogrammen in de taalwetenschap gemaakt worden. Vervolgens zal ik erop ingaan hoe dat in de biologie gebeurt en ten slotte waarom deze biologische methoden niet toepasbaar zijn op taal.

2 Cladogrammen in de taalwetenschap

2.1 Taalverandering en evolutie

Een van de klassieke doelen van de historische linguïstiek is het reconstrueren van een prototaal die ten grondslag ligt aan andere talen die uit die prototaal zijn voortgekomen. Er wordt in dit geval gebruikt gemaakt van de *Comparative Method* waarmee door middel van taalvergelijking een prototaal gereconstrueerd kan worden. Volgens Campbell (2013) bestaat die methode uit zeven stappen die uit te voeren zijn aan de hand van cognaten, woorden die in verschillende talen voorkomen en etymologische aan elkaar zijn gerelateerd.

Als eerste stap moet materiaal verzameld worden op basis waarvan de analyse uitgevoerd kan worden. Dat materiaal bestaat uit een aantal cognaten die gebruikt worden om een prototaal te reconstrueren. Een voorbeeld van zo'n cognatenset zijn de woorden *tafel*, *Tafel* en *table* in het Nederlands, Duits en Engels. Nadat een cognatenset met cognaten uit verschillende talen is samengesteld, wordt er bijvoorbeeld gekeken naar klankcorrespondenties. Dit wil zeggen dat de klanken van de cognaten in één taal vergeleken worden met de klanken van de cognaten in de andere talen. Voor de genoemde set zouden de correspondenties eruitzien als in Tabel 1.

Tabel 1: Klankcorrespondenties (KK) van de woorden *tafel*, *Tafel* en *table* in het Nederlands, Duits en Engels

	Nederlands	Duits	Engels
	ta:fɫ	ta:fl	tɛɪbɫ
KK 1	t	t	t
KK 2	a:	a:	ɛɪ
KK 3	f	f	b
KK 4	ɫ	l	ɫ

Op basis van deze klankcorrespondenties wordt er als derde stap een protoklank gereconstrueerd. Hierbij moet rekening gehouden worden met een aantal factoren. Ten eerste is er in veel gevallen een cross-linguïstische directionaliteit vast te stellen. Dit betekent dat een klank A vaker in een klank B verandert dan dat klank B klank A wordt. Een voorbeeld hiervan is $s > h$ (Campbell, 2013) terwijl $h > s$ bijna onbekend is. In het voorbeeld van de drie cognaten uit het Nederlands, Duits en Engels moet dan de /b/ als protoklank aangenomen worden omdat de /b/ tussen twee klinkers in vaker in een /f/ verandert dan de andere kant op. Als er geen klank gevonden kan worden op basis van directionaliteit, moet gekeken worden welke klank in de verschillende talen het vaakst voorkomt. De klank die in de meeste talen voorkomt wordt de protoklank. In hetzelfde voorbeeld zou dit de /f/ zijn omdat die in twee van de drie talen voorkomt, maar omdat er al een protoklank op basis van directionaliteit is gevonden, is de keuze van de /f/ als protoklank niet van toepassing.

Het kan gebeuren dat een protoklank zich in een taal in twee (of meer) verschillende klanken splitst. Dat wil zeggen dat één klank uit de prototaal veranderd is in twee of meer klanken in een dochtertaal. Als in een taal twee

klanken voor één protoklank lijken op te treden - wat vaak het geval is bij klankveranderingen die fonologisch geconditioneerd zijn -, moet er achterhaald worden welke protoklank die verschillende klanken representeren. In het voorbeeld is dit niet te illustreren omdat er maar één cognaat in is opgenomen. In een grotere cognatenset is dat soort veranderingen van één protoklank naar meerdere klanken vaak beter te zien.

Als er na al deze stappen een set protoklanken is gereconstrueerd, moet er gecheckt worden of de klanken plausibel zijn binnen de gereconstrueerde foneeminventaris van de prototaal. Dit betekent dat er gezocht moet worden naar gaten in de foneeminventaris of fonologische onregelmatigheden die niet verklaarbaar zijn. Als bijvoorbeeld alle plosieven voorhanden zijn behalve de /d/ en als er een /r/ aanwezig is, dan is het voor de hand liggend dat de /r/ een verkeerde reconstructie is en in de prototaal eigenlijk de /d/ representeert. Dit kan ondersteund worden door het feit dat veranderingen van het type $d > r$ voorkomen in talen (Campbell, 2013).

Als voorlaatste stap wordt de foneeminventaris in een cross-linguïstisch perspectief bekeken. Als er bijvoorbeeld klanken in zitten die taaltypologisch heel zeldzaam zijn, is het aan te raden om die klank nader te bekijken en te achterhalen of er misschien een fout in de analyse zit. Als de foneeminventaris volledig en plausibel is gereconstrueerd, kunnen aparte morfemen worden gemaakt. Dit is dan de stap waarin de woordenschat van de prototaal gemaakt kan worden.

De beschreven methode kan toegepast worden omdat taal verandert. Door taalverandering ontstaat er een enorme variatie die met behulp van de *Comparative Method* kan worden teruggebracht tot een prototaal van een taalfamilie. Nu is het zo dat de processen van taalverandering lijken op die van de evolutie van de soorten (Croft, 2008). Zo wordt bijvoorbeeld gezegd dat exaptatie - het gebruiken van een kenmerk voor een bepaalde functie waarvoor dat kenmerk oorspronkelijk helemaal niet was ontstaan - zowel in de biologie als in taalverandering optreedt. Verder wordt beweerd dat variatie ontstaat doordat kenmerken gerepliceerd worden maar dat die replicatie niet perfect is. Er worden bovendien varianten die niet voordelig zijn geselecteerd en deze kunnen dus doorevolueren.

Een kenmerk dat gekopieerd wordt, wordt *replicator* genoemd (Croft, 2008).

Een speciaal type replicator is een *meme*. Volgens Croft (2008, p. 221) zijn memen

cognitive entities in the minds of humans, instances of human behavior, or artifacts. (...) Hence concepts are replicators and minds or brains of individuals are vehicles.

Wat dit betekent is dat memen cognitieve eenheden zijn (bijvoorbeeld gedachten) die door middel van communicatie overgedragen en gerepliceerd kunnen worden. Ritt (2004) heeft een analyse van taalverandering gemaakt in termen van memen. Volgens hem kunnen met name fonologische entiteiten zoals klanken beschreven worden als memen. Taalverandering vindt dienovereenkomstig plaats via spraak, dat wil zeggen door het produceren van uitingen en de variatie die daardoor natuurlijkerwijs ontstaat.

2.2 Cladogrammen in de taalwetenschap

In de vorige sectie is duidelijk geworden dat taal verandert en dat taalverandering en evolutie op elkaar lijken. Doordat taal veranderingen ondergaat is het mogelijk om vast te stellen welke talen verwant zijn aan elkaar en welke talen dat niet zijn. Een vaak gebruikte manier om dit te doen is via het idee van *fylogeneese*. Volgens Brinkman en Leipe (2001, p. 323) is fylogeneese het bestuderen van evolutionaire relaties. Deze methode is afkomstig uit de biologie en heeft als doel om soorten in te delen in families en te kunnen vaststellen welke soorten in welke mate aan elkaar verwant zijn. Dunn (2014, p. 190) zegt hierover dat fylogeneese “can be applied to any domain which varies according to general evolutionary processes”. Het verschil tussen fylogeneese en de Comparative Method die hierboven is beschreven, is dat fylogenetische methoden de nadruk leggen op evolutionaire processen die met name in de biologie gebruikt worden. Wat hieraan ten grondslag ligt is de aanname dat evolutie van taal op dezelfde manier werkt als evolutie in de biologie.

Dunn (2014) beschrijft dat de meeste fylogenetische methoden die toegepast worden in het bestuderen van taalverwantschap cognatensets als data gebruiken. Er zijn verscheidene methoden om vervolgens fylogenetische relaties te bepalen. In het geval van de lexicostatistiek wordt er gekeken naar hoeveel

cognaten in twee verschillende talen niet hetzelfde zijn. Hoe hoger dat getal is, hoe minder verwant de twee talen zijn.

Bij de *Levenshtein distance* ligt de nadruk meer op de fonologische woordvormen. Dat wil zeggen dat twee woorden vergeleken worden en elke verandering die nodig is om een woord in een ander woord te laten veranderen bijgehouden wordt. Het aantal veranderingen dat nodig is, is een maat voor de Levenshtein distance. Dunn (2014, p. 195) noemt het volgende voorbeeld. Er wordt gekeken naar de Levenshtein distance tussen *Shakespeare* en de naam *Jacques Pierre*. In Tabel 2 worden de twee namen in fonetische transcriptie weergegeven.

Tabel 2: Veranderingen die plaatsvinden van *Shakespeare* naar *Jacques Pierre*

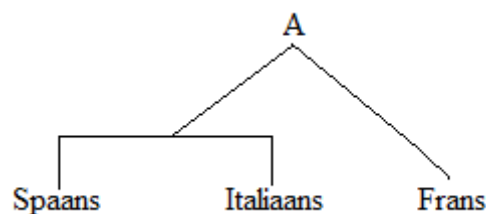
/	f	ɛi	k	s	p	i	r	/	
	substitutie	substitutie	deletie				insertie		
/	ʒ	a	k		p	i	ɛ	r	/

Hier is te zien dat er twee keer een klank vervangen, één keer een klank verwijderd en één keer een klank ingevoegd wordt. De Levenshtein distance is dus vier. Hoe groter die afstand is, hoe minder verwant de talen zijn.

Een directe link tussen biologie en taalwetenschap kan gelegd worden door het identificeren van *homologe kenmerken* in talen (Platnick & Cameron, 1977). Kenmerken worden homoloog genoemd wanneer ze dezelfde oorsprong hebben - in de biologie zijn dat gensequenties die van een gemeenschappelijke voorouder zijn overgenomen (Brinkman & Leipe, 2001). Als je in twee talen kenmerken kunt vinden die homoloog zijn, zijn die talen dus aan elkaar verwant omdat ze blijkbaar kenmerken van een gemeenschappelijke prototaal hebben overgenomen. Op basis van zulke kenmerken kunnen er zogeheten *cladogrammen* gemaakt worden. Dit zijn stambomen die laten zien welke talen of soorten in welke mate verwant zijn en van welke voorouder ze afstammen.

Om zulke homologe kenmerken te vinden moet in eerste instantie gekeken worden naar ogenschijnlijk op elkaar lijkende dingen in twee talen, maar - zoals Platnick en Cameron (1977, p. 383) stellen - het moet uit nader onderzoek blijken of die dingen ‘similarities of the entire system’ zijn. Als voorbeeld

wordt een homologie uit de Romaanse talen genoemd, namelijk de gelijkenis van de telwoorden *un*, *une* in het Frans, *uno*, *una* in het Spaans en *uno*, *una* in het Italiaans. Dat deze getallen een homoloog kenmerk van de drie talen zijn wordt ondersteund door het feit dat de woorden voor *twee* en *drie* ook op elkaar lijken (Fr. *deux*, *trois*, Sp. *dos*, *tres*, It. *due*, *tre*). In Figuur 1 wordt geïllustreerd hoe het Frans, het Spaans en het Italiaans op basis van de eerste drie telwoorden geclusterd zouden kunnen worden. De *A* staat voor de gemeenschappelijke voorouder van die drie talen.



Figuur 1: Een cladogram met het Frans, het Italiaans en het Spaans als inputtalen. Dit is een mogelijk cladogram dat gemaakt kan worden op basis van de eerste drie telwoorden in de drie talen. De letter *A* representeert de prototaal waaruit de drie talen zijn voortgekomen.

Het zoeken naar homologe kenmerken is de methode die het dichtst staat bij de biologie. De vraag is echter of het gerechtvaardigd is om een methode uit de evolutiebiologie waarin duidelijk onderscheid gemaakt kan worden tussen homologe en niet-homologe kenmerken op basis van de genetica, toe te passen op talen waarin dat onderscheid niet altijd even helder is. Platnick en Cameron (1977) zeggen dat cladistische methoden algemeen aanvaard zijn omdat ze vaak toegepast worden, maar ze geven geen harde argumenten waarom deze zouden werken voor talen. In de volgende secties zal ik laten zien hoe cladistische methoden in de biologie werken en in hoeverre ze gegeneraliseerd mogen worden naar taal.

3 Cladogrammen in de biologie

3.1 Fylogenetische methoden in de biologie

Het oorspronkelijke idee van cladogrammen komt uit de biologie. Met behulp van cladistische methoden proberen evolutionaire biologen om verwantschaps-

relaties tussen soorten in kaart te brengen en te visualiseren. De belangrijkste notie achter cladogrammen is dat soorten die op elkaar lijken niet noodzakelijkerwijs ook een nauwe verwantschapsrelatie onderhouden (Lipscomb, 1998). Er zijn namelijk twee redenen waarom twee soorten op elkaar kunnen lijken.

De eerste reden is dat ze echt aan elkaar verwant zijn en dat de kenmerken die ze gemeenschappelijk hebben genetisch van dezelfde voorouder afkomen. Zulke kenmerken worden dan *homoloog* genoemd. Echter, als twee soorten in (bijna) dezelfde omstandigheden leven en bepaalde kenmerken dus voor beide soorten voordelig zijn, dan gaan ze op elkaar lijken. De kenmerken die beide soorten dan verwerven worden in zo'n soort situatie *analoog* genoemd.

Het feit dat twee soorten homologe kenmerken hebben zorgt er echter nog niet voor dat ze ook noodzakelijkerwijs nauw aan elkaar verwant zijn. Lipscomb (1998) geeft hier het volgende voorbeeld bij. Mensen en krokodillen hebben vijf vingers aan hun handen, paarden hebben er maar één met een hoef. Nu is het evolutionair gezien zo dat mensen nauwer verwant zijn aan paarden dan aan krokodillen. Het feit dat krokodillen in tegenstelling tot paarden vijf vingers hebben zorgt er niet automatisch voor dat ze meer verwant zijn aan mensen dan paarden. Het belangrijke punt hier is dat een vijftal vingers een primitief kenmerk is, dat wil zeggen dat de gemeenschappelijke voorouder van mensen, paarden en krokodillen vijf vingers had. Bij paarden is dit aantal in de loop der tijd gereduceerd tot één vinger. Op basis van het feit dat mensen en krokodillen een homoloog kenmerk delen kan dus niet geconcludeerd worden dat ze nauwer aan elkaar verwant zijn dan mensen en paarden.

Om een cladistische analyse te kunnen uitvoeren moet er bepaald worden of een kenmerk *primitief* of *afgeleid* is (Lipscomb, 1998). Afgeleide kenmerken zijn kenmerken die bij twee (of meer) soorten zichtbaar zijn maar die er niet waren in hun gemeenschappelijke verre voorouder (Brinkman & Leipe, 2001). Als er zo'n kenmerk gevonden wordt, betekent dat dat de groep soorten die dat kenmerk heeft dus verwantschappelijk bij elkaar hoort en niet bij andere soorten met dezelfde voorouder die het betreffende kenmerk niet hebben. Een primitief kenmerk daarentegen is een kenmerk dat niet afgeleid is en dat dus ook de gemeenschappelijke voorouder al had.

Deze manier om primitieve en afgeleide kenmerken te vinden wordt *outgroup comparison* genoemd (Lipscomb, 1998). Hierbij wordt een groep soorten die

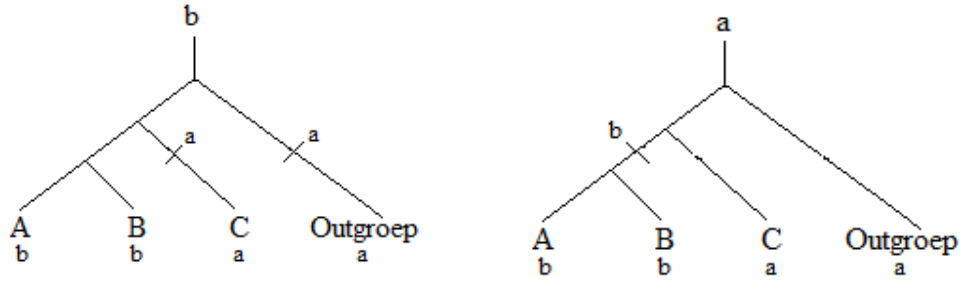
geclassificeerd worden de *ingroep* genoemd, de andere soorten de *outgroep*. Er zijn in principe twee redenen waarom deze methode lijkt te werken.

De eerste reden draait om homologe kenmerken. Als een ingroep en een outgroep een homoloog kenmerk delen, dan moet het zo zijn dat ze het betreffende kenmerk hebben overgenomen van een voorouder die ouder moet zijn dan de voorouder van de ingroep. Het volgende voorbeeld moet dit duidelijk maken. In Figuur 2a is te zien dat het paard, de kangoeroe en de aap binnen de familie van de zoogdieren vallen. Het paard heeft één vinger en de kangoeroe en de aap hebben er vijf (aangegeven tussen haakjes onder de namen van de soorten). Nu is het zo dat de hagedis ook vijf vingers heeft maar duidelijk geen zoogdier is. Hagedis, aap en kangoeroe delen dus een homoloog kenmerk. De enige mogelijkheid om dit te verklaren is dat zoogdieren en hagedis een gemeenschappelijke voorouder moeten hebben die vijf vingers had, zoals te zien is in Figuur 2b.

De tweede reden waarom de methode van *outgroup comparison* lijkt te werken is parsimonie of spaarzaamheid. Als er twee genetisch aan elkaar gerelateerde kenmerken a en b zijn, dan is óf a primitief en b afgeleid óf andersom. Als we nu aannemen dat kenmerk a niet alleen in de ingroep maar ook in een outgroep te vinden is, dan moeten er meer processen plaatsvinden wanneer b het primitieve kenmerk is. In Figuur 3 op pagina 11 wordt dit getoond. In Subfiguur 3a is b het primitieve kenmerk, in Subfiguur 3b is a het primitieve kenmerk. Het is duidelijk dat in 3a er twee veranderingen nodig zijn om tot de kenmerken die als kleine letters onder de hoofdletters staan te komen. Eén keer moet er iets in de ingroep veranderen en één keer iets in de outgroep. In 3b daarentegen hoeft er alleen een verandering plaats te vinden binnen de ingroep. De hypothese dat a het primitieve kenmerk is, is dus economischer en voldoet aan het principe van parsimonie.

Als eenmaal bekend is welke kenmerken primitief en welke afgeleid zijn, kan er een cladogram gemaakt worden. Een methode hiervoor is de *Henning-methode* (Lipscomb, 1998). Deze methode scheidt een aantal soorten die bij de analyse betrokken worden stap voor stap van elkaar. Dit zal nu geïllustreerd worden aan de hand van een voorbeeld.

Laten we aannemen dat er een ingroep is die bestaat uit drie soorten, A , B en C . Er zijn vier kenmerken die gebruikt worden bij de analyse. Elke soort



(a) Als *b* het primitieve kenmerk is, moet er op twee plekken worden aangenomen dat kenmerk *a* afgeleid is uit *b*.

(b) Als *a* het primitieve kenmerk is, moet aangenomen worden dat *b* op maar één plek is afgeleid uit *a*.

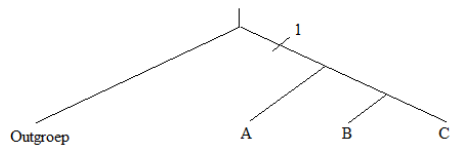
Figuur 3: De twee hypothesen met één keer *b* en één keer *a* als primitief kenmerk. De lettertjes onder de hoofdletters geven aan welk kenmerk de soorten hebben. In de rechter figuur is er maar één plek van verandering, in de linker figuur zijn het er twee.

Tabel 3: De verdelingen van de kenmerken binnen de soorten. De nullen staan voor primitieve kenmerken, de enen voor afgeleide kenmerken.

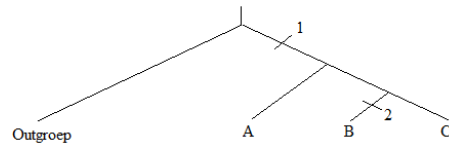
	Kenmerk 1	Kenmerk 2	Kenmerk 3	Kenmerk 4
Outgroep	0	0	0	0
A	1	0	0	0
B	1	1	0	1
C	1	0	1	1

van afgeleide en primitieve kenmerken. Een belangrijk punt hierbij is dat die kenmerken - los ervan of ze afgeleid of primitief zijn - in de genen van de soorten verankerd moeten zitten. De genen spiegelen als het ware de ontwikkelingsgeschiedenis van de soorten weer en daarom zijn alleen de genen geschikt om verwantschapsrelaties tussen soorten te onderzoeken. De kenmerken die genetisch vastgelegd zijn, worden *genotypen* genoemd. Deze worden dus puur door genetische mutaties veranderd. Kenmerken waarop ook door de omstandigheden of het milieu waarin een soort leeft invloed kan worden uitgeoefend, worden *fenotypen* genoemd. Deze kenmerken zitten dus niet in de genen en zijn aan veranderingen van niet-genetische aard onderhevig.

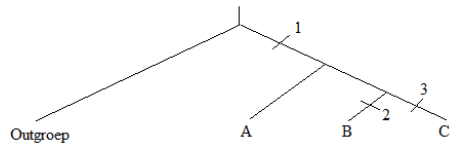
Met betrekking tot het onderscheid tussen genotypen en fenotypen is er door Dollo (1893) een fenomeen beschreven dat tegenwoordig *de wet van Dollo* genoemd wordt. In de volgende sectie zal hierop worden ingegaan omdat die wet een belangrijk inzicht biedt in de toepassing van fylogenetische methoden in de taalwetenschap.



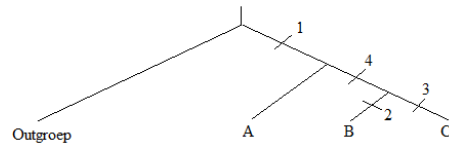
(a) Het eerste kenmerk kan gebruikt worden om de outgroep te scheiden van de drie andere soorten.



(b) Het tweede kenmerk scheidt *B* af van de andere twee soorten.



(c) Kenmerk nummer drie is alleen afgeleid bij soort *C*.



(d) Het laatste kenmerk neemt *B* en *C* bij elkaar en scheidt deze af van *A*.

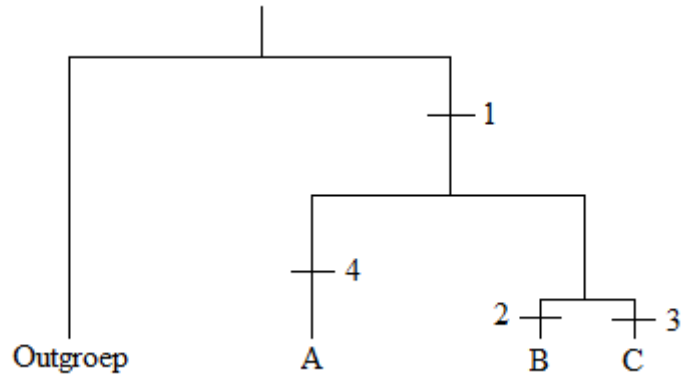
Figuur 4: De stappen die ondernomen moeten worden om een cladogram op basis van Tabel 3 te maken.

3.2 De wet van Dollo

De wet van Dollo zegt dat kenmerken die in de loop van de evolutie zijn verdwenen niet meer kunnen optreden. Evolutie is volgens de wet van Dollo dus *onomkeerbaar*. Dit betekent dat een kenmerk nooit meer terugkeert, ook als de soort waarbij dat kenmerk ooit is opgetreden zich in precies dezelfde omstandigheden bevindt als toen het kenmerk nog aanwezig was. Maar Dollo zegt ook dat de sporen van dat kenmerk er altijd in de genen van de soort zullen zijn.

Het is belangrijk om op te merken dat Dollo zelf zei dat functionele of fysiologische kenmerken wel degelijk kunnen terugkeren, maar structurele en morfologische kenmerken niet (Gould, 1970). Dat functionele eigenschappen wel vaker kunnen optreden ligt waarschijnlijk aan de al genoemde exaptatie van een kenmerk. Dat wil zeggen dat een kenmerk voor een functie wordt gebruikt waarvoor het oorspronkelijk niet was ontstaan. Morfologische en structurele eigenschappen kunnen niet zo gemakkelijk door exaptatie beïnvloed worden omdat ze niet zo'n duidelijke functie kunnen overnemen.

Verder is het zo dat de wet van Dollo alleen van toepassing is op genotypen. Fenotypische kenmerken kunnen steeds weer terugkeren omdat deze niet afhankelijk zijn van de genen van een soort maar van de omstandigheden waarin een soort leeft. Een voorbeeld hiervan is lichaamslengte bij mensen. Als er



Figuur 5: Het cladogram dat op basis van Tabel 3 is gemaakt. Het is equivalent met het cladogram in Figuur 4d.

weinig voedsel beschikbaar is, zijn mensen gemiddeld korter dan wanneer er genoeg voedsel is. Als de situatie verandert, kan dat weer van invloed zijn op de lichaamslengte. Dit betekent echter niet dat de genen door de veranderingen in het milieu veranderen, maar die veranderingen in het milieu zelf zijn de reden waarom de lichaamslengte varieert.

Een voorbeeld van een genotypisch kenmerk zijn tanden bij vogels. Collin en Miglietta (2008) zeggen dat vogels sinds 60 miljoen jaar geen tanden meer hebben en dat er tot nu toe geen soorten vogels bekend zijn die weer tanden hebben gekregen. Echter, kuikenembryo's laten een soort primitieve plaat in hun snavel zien die niet doorontwikkelt. Dit komt overeen met wat Dollo zegt, namelijk dat de sporen van een verloren kenmerk altijd aanwezig zijn. Het is zelfs zo dat een genmutatie bij kuikens, die *talpid*² kuikens worden genoemd (Collin & Miglietta, 2008), ervoor zorgt dat die kuikens tandachtige structuren laten zien die lijken op een vroege vorm van alligatortanden.

Het feit dat genotypische kenmerken die door evolutionaire processen zijn verdwenen niet meer terugkeren is verklaarbaar omdat de kans dat een verdwenen kenmerk weer terugkomt extreem gering is. De genetische veranderingen die door miljoenen jaren heen hebben plaatsgevonden zouden in precies de omgekeerde volgorde weer moeten gebeuren om het oorspronkelijke kenmerk terug te krijgen. De kans dat dit gebeurt is bijna nul. Onomkeerbaarheid is dus een kwestie van kans.

Een ander belangrijk punt is dat de wet van Dollo niet ongeldig is omdat

eigenschappen die een soort ooit heeft gehad te zien zijn in de ontogenie van een soort, dat wil zeggen in de ontwikkeling van de embryo's. Volgens Gould (1970) is de ontogenie een recapitulatie van de fylogenie, dus de ontwikkeling van de soort in het verloop van de evolutie. Kenmerken die ooit aanwezig zijn geweest bij de voorouders van een soort komen dus in de ontwikkeling van een embryo even terug maar functioneel wordt er verder niets mee gedaan.

In de wet van Dollo is dus geformuleerd dat in de genen verankerde structurele en morfologische kenmerken nooit meer optreden bij dezelfde soort, als die kenmerken eenmaal zijn verdwenen. De reden hiervoor is de tot nul naderende kans dat een door de evolutie veranderd kenmerk precies dezelfde mutaties in de omgekeerde volgorde meemaakt om weer bij het originele kenmerk te komen.

In het volgende hoofdstuk zal duidelijk worden waarom de wet van Dollo zo belangrijk is bij het toepassen van fylogenetische methoden in taal. Het zal blijken dat uit die wet volgt dat het gebruik van biologische methoden in een taalwetenschappelijke oftewel taalhistorische context niet gerechtvaardigd is.

4 Taal als organisme

4.1 Taalkundige cycli

Omdat de wet van Dollo iets zegt over de afwezigheid van cyclische veranderingen in de evolutie van soorten is het interessant om na te gaan hoe deze wet zich verhoudt tot taalverandering. Talen veranderen namelijk niet lineair maar veel meer op een cyclische manier. Van Gelderen (2009, p. 2) definieert een taalkundige cyclus op de volgende manier:

The Linguistic Cycle is a name for changes where a phrase or word gradually disappears and is replaced by a new linguistic item.

De naam *cyclus* kan misleidend zijn omdat de specifieke vormen die in het verloop van de cyclische verandering verdwijnen niet terug hoeven te keren. Von der Gabelentz (1901, p. 256) spreekt daarom over een spiraal:

[D]ie Entwicklungslinie krümmt sich zurück nach der Seite der Isolation, nicht in die alte Bahn, sondern in eine annähernd parallele.

Darum vergleiche ich sie der Spirale.¹

Omdat een cyclus echter een makkelijkere conceptuele notie is, zal ik het begrip *cyclus* blijven gebruiken. Het klassieke voorbeeld van een taalkundige cyclus is de *Jespersen cyclus* die voor het eerst beschreven is door Jespersen (1917). Van der Auwera (2009) presenteert deze cyclus voor onder andere het Frans en onderscheidt vijf stappen in het proces.

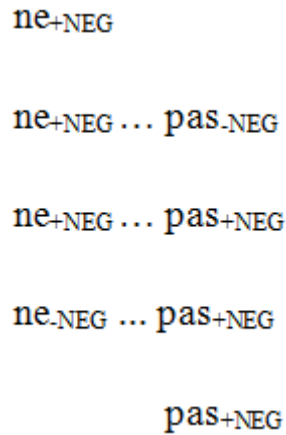
Als eerste stap was er maar één negatiepartikel in het Frans, namelijk *ne* (uitgesproken als [nə]). Met behulp van dit morfeem werd een zin ontkend. Vervolgens werd dit fonetisch zwakke element formeel en semantisch versterkt door er een tweede lexicale element *pas* 'stap/pas' bij te betrekken. Dit nieuwe morfeem was nog niet gegrammaticaliseerd en had een puur lexicale betekenis. Omdat *pas* heel handig bleek omdat de negatie nu weer sterker uitgedrukt kon worden - er was immers een fonetisch sterk element bij gekomen - ging het grammaticaliseren en kreeg uiteindelijk een ontkennende betekenis waardoor het ook niet meer optioneel was. Er was een tweeledige negatie ontstaan die bestond uit de delen *ne* en *pas*.

Nu er twee elementen waren om negatie te coderen en één van die elementen fonetisch sterker was dan het andere, werd het oorspronkelijke element *ne* semantisch zwakker en kon weggelaten worden. Terwijl in een vroeger stadium het nieuwe element *pas* optioneel was, werd nu het oude element optioneel. Door verdere verzwakking van *ne* werd het uiteindelijk helemaal weggelaten, zoals het tegenwoordig in het gesproken Frans het geval is.

Het is dus duidelijk dat er over een cyclus gesproken kan worden omdat negatie in het Frans door de tijd heen eerst met één negatiemorfeem uitgedrukt werd en vervolgens met twee morfemen die weer tot één partikel gereduceerd werden. Hierbij hebben processen zoals semantische en fonetische afzwakking en versterking een rol gespeeld. Een schematisch overzicht is te zien in Figuur 6. Het subscript *+NEG* betekent dat het woord een ontkennende functie in de zin heeft, *-NEG* wil zeggen dat dat niet het geval is.

Een andere cyclus wordt beschreven door van Gelderen (te verschijnen). In haar artikel presenteert ze de naamvalscyclus. In het verloop van de geschiedenis van een taal die naamvallen heeft, wordt vaak geobserveerd dat de naam-

¹De verandering gaat terug richting isolatie, niet op de originele maar op een parallelle manier. Daarom vergelijk ik haar met een spiraal (eigen vertaling).



Figuur 6: Een schematisch overzicht van de Jespersen cyclus. Een woord met *+NEG* als subscript heeft een ontkennende functie, een woord met *-NEG* heeft er geen.

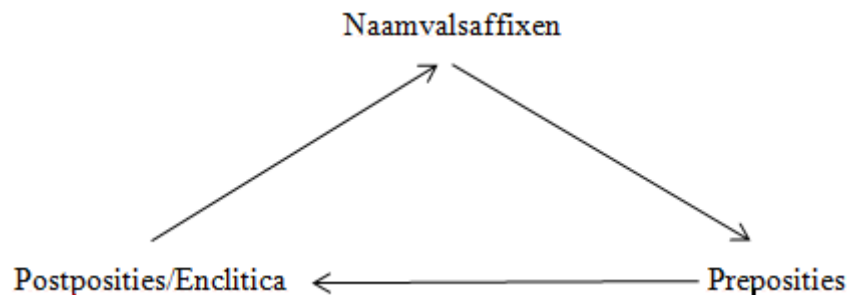
valsaffixen verdwijnen. De codering van de functies door die affixen valt dus weg waardoor er een nieuw middel nodig is om de functies te markeren. Dit geschiedt meestal door adposities en dus worden de functies syntactisch in plaats van morfologisch gemarkeerd.

Een voorbeeld hiervan is de genitief in het Nederlands. Tegenwoordig zijn er nog archaïsche resten van de genitief te vinden, zoals te zien is in (1) waarin het bezit door middel van het suffix *-(e)s* wordt uitgedrukt. In het moderne Nederlands wordt zo'n relatie echter gecodeerd met de prepositie *van*, geïllustreerd in (2).

- (1) De heer des huizes
- (2) De heer **van** het huis

Het verlies van naamval resulteert dus in het gebruik van adposities om de functies te coderen die door de niet meer voorhanden naamvallen uitgedrukt werden. Na het vervangen van naamvalsaffixen door adposities grammaticaliseren deze steeds meer en kunnen als postposities achter het nomen worden gezet. Dit is de tweede stap in de naamvalscyclus. Als laatste stap worden de postposities steeds enclitischer waardoor ze uiteindelijk niet meer los kunnen staan van hun nomen. Dit is het moment waarop er weer nieuwe naamvalsaffixen zijn ontstaan en de cyclus begint opnieuw. Een voorbeeld van dit laatste

proces is, aldus van Gelderen (te verschijnen), het Turks. Deze taal heeft de postpositie *ile* 'met' die langzamerhand een naamvalsuffix aan het worden is. Zo kun je bijvoorbeeld zeggen *Mehmet-le* wat dan 'met Mehmet' betekent. Figuur 7 illustreert de naamvalscyclus op een visuele manier.



Figuur 7: Een schematisch overzicht van de naamvalscyclus. Als de naamvalsaffixen verdwijnen, worden deze vervangen door preposities die vervolgens als postposities achter het nomen worden geplaatst. De postposities grammaticaliseren en worden weer naamvalsuffixen.

Taalkundige cycli zijn dus verschijnselen in taalverandering waarin een bepaald patroon zich steeds herhaalt en dus steeds weer terugkeert. Dat dit soort verandering in taal bestaat impliceert een belangrijke consequentie voor cladistische methoden die in de taalwetenschap toegepast worden. Die consequentie zal in de volgende twee secties besproken worden.

4.2 Talige kenmerken als fenotypen

Zoals in Sectie 3.2 besproken is, heeft de wet van Dollo alleen betrekking op genotypen. Dat wil zeggen dat in de loop van de evolutie verdwenen genotypische kenmerken bij een soort niet meer zullen terugkeren, maar dat dat bij fenotypische eigenschappen wel het geval kan zijn.

Dit laatste feit is belangrijk als het gaat over taalverandering. Omdat taal - zoals in de vorige sectie besproken is - cyclisch verandert en evolueert, kunnen talige kenmerken die veranderen niet genotypisch zijn; ze zijn noodzakelijkerwijs fenotypisch - dit is de consequentie die volgt uit de wet van Dollo. Bij nader inzien is dit ook vrij voor de hand liggend omdat er bepaalde fenomenen bestaan die taalverandering beïnvloeden maar die in biologische evolutie onmogelijk zijn.

Een van die verschijnselen is taalcontact. Taalcontact wordt veelal genoemd als een van de belangrijkste motoren voor taalverandering (Thomason, 2001; Lee, 1987; Silva-Corvalán, 1994; Cabral, 2003; Heine & Kuteva, 2005; Miestamo, Sinnemäki & Karlsson, 2008). Aikhenvald (2002) beschrijft in haar boek talrijke voorbeelden van taalveranderingen die door taalcontact hebben plaatsgevonden. Een van die voorbeelden betreft de taal Retuarã die onderdeel uitmaakt van de Yucuna-taalfamilie. Het Retuarã heeft een aantal oblique naamvallen zoals instrumentalis of locatieve naamval, wat niet typisch is voor de Yucuna-talen. Volgens Aikhenvald (2002) is deze verandering tot stand gekomen door contact met de Arawak-talen die dat soort naamvallen wel regelmatig laten zien.

Een ander voorbeeld dat tevens naamvallen betreft zijn locatieve naamvallen in het Tariana dat hoort bij de Arawak-taalfamilie. Deze taal heeft de volgorde van locatieve naamvalssuffixen geleend uit het Oost-Tucano terwijl de locatieve morfemen van Arawakse oorsprong zijn. Aikhenvald (2002, p. 107) bespreken het voorbeeld uit Tariana dat getoond wordt in (3). In (4) is een woord uit het Baniwa (Arawak-taalfamilie) te zien dat toont dat de morfemen *haiku*, *haiku*, *se*, *hre* en *naku*, *naku* cognaten zijn van elkaar. De twee talen behoren immers tot dezelfde taalfamilie.

- (3) haiku-se-naku
boom-LOCATIE-OP
"op deze boom hier"(Tariana)
- (4) haiku-naku-hre
boom-OP-NAARTOE
"naar het oppervlak van de boom"(Baniwa)

Wat opvalt is dat de twee locatieve morfemen in het Baniwa en het Tariana in precies de omgekeerde volgorde optreden. Terwijl het Baniwa de volgorde vertoont die typisch is voor de Arawak-talen, is de volgorde van de morfemen in het Tariana typerend voor het Oost-Tucano. Het Tariana heeft de volgorde van de locatieve naamvalssuffixen maar niet die suffixen zelf ontleend aan een andere taal en heeft dus een taalverandering door taalcontact ondergaan.

Een andere factor die taalverandering kan beïnvloeden is cultuur. In maatschappijen met schoolplicht en relatief grote verschillen in opleiding van de bevolking is het effect van de prescriptieve grammatica meestal heel groot en

worden taalveranderingen gauw stigmatiserend (Drake, 1977). Hubers en de Hoop (2013) rapporteren een effect van de prescriptive regel dat alleen *dan* in comparatieve constructies gebruikt mag worden. In (5a) is een voorbeeld te zien van een comparatief met *dan* als partikel, in (5b) dezelfde zin met *als*. De eerste zin is volgens de prescriptieve regels 'goed', de andere niet.

- (5) a. De vrouw is groter **dan** de man.
b. De vrouw is groter **als** de man.

Hubers en de Hoop (2013) vonden dat *dan* in comparatieven vaker voorkomt dan *als* hoewel *als* al in de zestiende eeuw werd gebruikt en tegenwoordig in vrijwel alle dialecten van het Nederlands voorkomt. Het hoge percentage van *dan* in comparatieven weerspiegelt dus de aanwezigheid van de prescriptieve regel. De onderzoekers vonden dat de keuze tussen de twee partikels in hoge mate samenhangt met het opleidingsniveau van de sprekers. Hoe hoger de mensen zijn opgeleid, hoe meer ze zich aan de prescriptieve regel houden. Dit leidt ertoe dat ze de prescriptief 'foute' vorm *als* in comparatieven onderdrukken. Het prescriptivisme - een cultureel verschijnsel - zorgt er dus voor dat er een taalverandering geblokkeerd wordt.

Een ander voorbeeld van invloed van prescriptivisme op taalverandering wordt gegeven door van Gelderen (2006). Zij rapporteert dat Amerikaanse studenten het voegwoord *that* maar heel zelden gebruiken wanneer er ook *who* of *which* kan staan. Het vervangen van deze twee woorden door *that* in relatieve bijzinnen is heel gebruikelijk in het Engels, maar wordt studenten afgeleerd waardoor *that* steeds zeldzamer wordt.

Al deze voorbeelden tonen aan dat taalverandering in grote mate onderhevig is aan externe factoren zoals cultuur en taalcontact. Vanuit een biologisch oogpunt kan dit alleen als de dingen die er veranderen fenotypisch van aard zijn en niet genotypisch. Ook de wet van Dollo wijst erop dat talige kenmerken fenotypen zijn. Hieruit volgt dat er geen genotypen bestaan in taal en dat taalkundige kenmerken fenotypisch van aard zijn. De consequentie hiervan zal besproken worden in de volgende sectie.

4.3 Terug naar het begin: cladogrammen in de taalwetenschap

Zoals in de vorige secties is uitgelegd worden cladogrammen in de taalwetenschap gebruikt om taalverwantschapsrelaties te vinden. Hierbij worden de cladistische methoden uit de biologie gebruikt die gebaseerd zijn op genotypen van soorten. Hierboven heb ik beargumenteerd dat de talige kenmerken die bij taalvergelijking worden gebruikt en die in de tijd veranderen, fenotypisch zijn. Dit volgt direct uit het feit dat taal cyclisch verandert - wat volgens de wet van Dollo alleen van toepassing is op fenotypische kenmerken - en uit het feit dat taalverandering sterk van buitenaf beïnvloed wordt. De eigenschappen van talen die veranderbaar zijn en die men gebruikt om cladogrammen te maken zijn dus eenduidig fenotypisch en kunnen niet gebruikt worden in op genotypen gebaseerde methoden.

Desalniettemin worden er toch cladogrammen van talen gemaakt die in de biologie alleen op basis van genotypische kenmerken gecreëerd kunnen worden. De conclusie is dus dat er geen betrouwbare cladogrammen van talen gemaakt kunnen worden omdat de data die gebruikt wordt voor zo'n cladogram niet betrouwbaar zijn. De precieze vorm van een cladistische boom hangt af van de keuze van de data hoewel dit eigenlijk niet zou moeten. Het toepassen van cladistische methoden uit de biologie op talen is gebaseerd op een verkeerde aanname, namelijk dat de gebruikte kenmerken genotypisch zijn. Beweringen over taalverwantschapsrelaties die gestoeld zijn op cladogrammen zijn dus niet betrouwbaar.

Voor de bomen die door Dunn et al. (2011) zijn gemaakt, geldt dit dus ook. De auteurs laten zelf zien dat de aan veranderingen onderhevige talige kenmerken - concreet de volgorde van adpositie-nomen en werkwoord-object - alle kanten op kunnen veranderen en dat deze dus fenotypisch van aard zijn. Volgens hen is de kans groot dat bijvoorbeeld de volgorde werkwoord-object in een taal met postposities verandert naar object-werkwoord. Ze laten zien dat een object-werkwoordtaal met postposities ook in een taal met werkwoord-objectvolgorde kan veranderen, hoewel de kans dat dit gebeurt wellicht kleiner is. Voor alle mogelijke combinaties (postposities en werkwoord-object, postposities en object-werkwoord, preposities en werkwoord-object, preposities en

object-werkwoord) tonen ze aan dat er tenminste een kleine kans is dat het één in het ander verandert. Dit geeft dus aan dat het over een cyclische verandering gaat waaruit volgt dat ook woordvolgordekenmerken fenotypische kenmerken zijn.

Voor hun cladogram gebruiken ze echter op woordvolgorde gebaseerde data. Dit is dus niet gerechtvaardigd en zet de betrouwbaarheid van de uitkomsten van hun analyse op losse schroeven. Het feit dat ze bestaande bomen gebruiken voor hun analyses staat deze conclusie niet in de weg. Immers, voor de methoden die toegepast zijn om de bestaande bomen te verkrijgen zijn tevens taalkundige data gebruikt die fenotypisch van aard zijn. Bij de Comparatieve Methode bijvoorbeeld, die in Sectie 2.1 is besproken, worden lexicale elementen gebruikt en met elkaar vergeleken waaruit taalverwantschappen afgeleid worden. Het is echter voor de hand liggend dat woorden in sterke mate onderhevig zijn aan culturele en maatschappelijke aspecten. Zo kan taalcontact ervoor zorgen dat woorden veranderen, verdwijnen of nieuw ontstaan. Ook kunnen bepaalde woorden uit een vreemde taal een bepaald prestige hebben dat ervoor zorgt dat die woorden in een andere taal worden opgenomen of juist worden vermeden. Hierdoor ontstaan mengelingen die alleen kunnen optreden bij fenotypische kenmerken. Dat Dunn et al. (2011) bestaande bomen gebruiken, is dus geen argument voor de juistheid van hun uitkomsten.

5 Conclusie

We hebben gezien dat het toepassen van cladistische methoden die gebaseerd zijn op genotypische kenmerken niet gerechtvaardigd is voor het modelleren van taalverwantschappen. De reden hiervoor is dat de kenmerken die betrokken worden bij een analyse van taalverwantschapsrelaties beïnvloed worden door de omstandigheden waarin een taal gesproken wordt, zoals bijvoorbeeld cultuur en taalcontact. Bovendien is het zo dat die kenmerken op een cyclische manier in de tijd veranderen en de wet van Dollo zegt dat dit alleen mogelijk is voor kenmerken die fenotypisch zijn. Dit alles wijst erop dat de data die gebruikt worden bij het maken van cladogrammen niet genotypisch kunnen zijn en dus niet geschikt zijn om er cladogrammen van talen mee te maken.

Cladogrammen zoals in Figuur 1 van het Frans, het Spaans en het Italiaans

zijn dus geen betrouwbare weergaven van taalverwantschapsrelaties. De data die als basis gebruikt zijn voor Figuur 1 - de woorden voor *één* en *twee* in de drie talen - hebben hun specifieke vorm alleen op basis van de omstandigheden waarin die drie talen zijn veranderd. Het is niet zo dat er iets in de 'genen' van de talen is veranderd dat het telwoord er zo deed uitzien.

Hoewel cladistische methoden in de taalwetenschap vaak toegepast worden, is gebleken dat de aanname die hiervoor nodig is niet verantwoord is. Om erachter te komen hoe betrouwbare representaties van taalverwantschapsrelaties gemaakt kunnen worden, moet er onderzoek gedaan worden dat er rekening mee houdt dat de veranderbare kenmerken die in taalvergelijking gebruikt worden niet genotypisch, maar fenotypisch zijn en dus hun vorm hebben op basis van externe invloeden.

Literatuur

- Aikhenvald, A. Y. (2002). *Language contact in Amazonia*. Oxford: Oxford University Press.
- Auwers, J. van der (2009). The Jespersen cycles. In Elly van Gelderen (red.), *Cyclical change* (pp. 35–71). Amsterdam: John Benjamin Publishing Company.
- Brinkman, F. S. & Leipe, D. D. (2001). Phylogenetic analysis. *Bioinformatics: a practical guide to the analysis of genes and proteins*, 323–358.
- Cabral, A. S. A. C. (2003). *Contact-induced language change in the western Amazon: The non-genetic origin of the Kokama language* (Academisch proefschrift). UMI, Ann Arbor.
- Campbell, L. (2013). *Historical linguistics: An introduction*. Edinburgh: Edinburgh University Press.
- Collin, R. & Miglietta, M. P. (2008). Reversing opinions on Doll's law. *Trends in Ecology & Evolution*, 23(11), 602–609.
- Croft, W. (2008). Evolutionary linguistics. *Annual Review of Anthropology*, 37(1), 219–234.
- Dollo, L. (1893). The laws of evolution. *Bulletin de la Societe Belge de Geologie, de Paleontologie et d'Hydrologie*, 7, 164–166.
- Drake, G. F. (1977). *The Role of Prescriptivism in American Linguistics 1820-1970* (Dl. 13). Amsterdam: John Benjamins Publishing.
- Dunn, M. (2014). Language phylogenies. In C. Bower & B. Evans (red.), *The Routledge handbook of historical linguistics* (pp. 190–211). New York: Routledge.
- Dunn, M., Greenhill, S. J., Levinson, S. C. & Gray, R. D. (2011). Evolved structure of language shows lineage-specific trends in word-order universals. *Nature*, 473(7345), 79–82.
- Gabelentz, G. von der (1901). *Die Sprachwissenschaft: ihre Aufgaben, Methoden und bisherigen Ergebnisse*. CH Tauchnitz.
- Gelderen, E. van (2006). Economy against Prescriptivism: Internal and External Factors of Language Change. *South West Journal of Linguistics*, 25, 1–17.
- Gelderen, E. van (2009). *Cyclical change*. John Benjamins Publishing.

- Gelderen, E. van (te verschijnen). The dependent marking cycles: Case. In Elly van Gelderen (red.), *Linguistic cycles*.
- Gould, S. J. (1970). Dollo on Dollo's law: irreversibility and the status of evolutionary laws. *Journal of the History of Biology*, 3(2), 189–212.
- Heine, B. & Kuteva, T. (2005). *Language contact and grammatical change*. Cambridge: Cambridge University Press.
- Hubers, F. & de Hoop, H. (2013). The effect of prescriptivism on comparative markers in spoken Dutch. *Linguistics in the Netherlands*, 30(1), 89–101.
- Jespersen, O. (1917). Negation in English and other languages. *Høst & Søn*.
- Lee, J. R. (1987). *Tiwi today: a study of language change in a contact situation*. Canberra: Dept. of Linguistics, Research School of Pacific Studies, The Australian National University.
- Lipscomb, D. (1998). Basics of cladistic analysis. *George Washington University*.
- Miestamo, M., Sinnemäki, K. & Karlsson, F. (2008). *Language complexity: Typology, contact, change* (Dl. 94). Amsterdam: John Benjamins Publishing.
- Platnick, N. I. & Cameron, H. D. (1977). Cladistic methods in textual, linguistic, and phylogenetic analysis. *Systematic Biology*, 26(4), 380–385.
- Ritt, N. (2004). *Selfish sounds and linguistic evolution: a darwinian approach to language change*. Cambridge: Cambridge University Press.
- Silva-Corvalán, C. (1994). *Language Contact and Change: Spanish in Los Angeles*. Oxford: Oxford University Press.
- Thomason, S. G. (2001). *Language contact*. Edinburgh: Edinburgh University Press.