



Universiteit  
Leiden  
The Netherlands

## Building a Phonological Inventory

Veer, B.M. van 't

### Citation

Veer, B. M. van 't. (2014, January 7). *Building a Phonological Inventory*. *LOT dissertation series*. Retrieved from <https://hdl.handle.net/1887/30242>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/30242>

**Note:** To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/30242> holds various files of this Leiden University dissertation

**Author:** Veer, B.M. van 't

**Title:** Building a phonological inventory : feature co-occurrence constraints in acquisition

**Issue Date:** 2015-01-07

---

## Samenvatting in het Nederlands

---

De centrale stelling van dit proefschrift is dat de verwerving van de consonantinventaris, een probleem dat reeds meer dan een eeuw de gemoederen bezighoudt, het best gezien kan worden als de gelijktijdige ontwikkeling van een inventaris van distinctieve kenmerken en een verzameling constraints die het aantal mogelijke combinaties van die kenmerken inperkt. Hierin verschilt het van eerdere aanpakken van de verwerving van de klankinventaris (Jakobson, 1941/1968; Beers, 1995, bijvoorbeeld), die voornamelijk naar de representatieve ontwikkeling keken, en daarmee niet goed in staat bleken een verantwoording te geven van de ontwikkeling van de inventaris inclusief de gaten die daarin onherroepelijk ontstaan.

De voorgestelde theorie is simpel en restrictief, en maakt gebruik van alleen monovalente kenmerken en van slechts twee typen constraints:

- (76) a. \*FG  
assign a violation mark for every segment  $\Sigma$  iff [F] is in  $\Sigma$  and [G] is in  $\Sigma$  (*c-constraint*)
- b. F→G  
assign a violation mark for every segment  $\Sigma$  iff [F] is in  $\Sigma$  and [G] is not in  $\Sigma$  (*i-constraint*)

Het eerste type noemen we ‘c-constraint’, waarbij de ‘c’ staat voor ‘co-occurrence’. Dit type wordt geschonden wanneer de twee kenmerken waarnaar het verwijst ([F] en [G] in voorbeeld 76) samen in één segment voorkomen. Het tweede type, de ‘i-constraint’ (waar ‘i’ staat voor ‘implicational’) wordt geschonden wanneer [F] in een segment voorkomt *zonder* [G]. Een ongeschonden i-constraint over [F] en [G] betekent dus dat de aanwezigheid van [F] de aanwezigheid van [G] impliceert.

Het proefschrift beslaat dus eigenlijk twee thema’s: enerzijds de ontwikkeling van een fonologische theorie over Feature Co-occurrence Constraints (FCC’s), en anderzijds de verwerving van de segmentinventaris. **Hoofdstuk 3** houdt zich bezig met het eerste thema, terwijl de verwervingskwestie aan

bod komt in **hoofdstuk 4**. De andere hoofdstukken behandelen elk aspecten van beide thema's.

## De inventaris

Het idee dat er restricties moeten bestaan op het aantal mogelijke combinaties van kenmerken is niet nieuw; een groot deel van **hoofdstuk 1** behandelt eerdere voorstellen die van FCCs gebruikmaken. Het beeld dat hieruit naar voren komt is dat dit type constraint bijzonder veelzijdig ingezet kan worden: het hoofdstuk geeft voorbeelden uit de literatuur over onderspecificatie (Archangeli, 1984, 1988), specificatievariëteit (Yip, 2011) en licentie (Itô et al., 1995), klinkerharmonie (Van der Hulst & Smith, 1986; Van der Hulst, 1988; Piggott & van der Hulst, 1997; Van der Hulst, 2012; Padgett, 2002; Kiparsky & Pajusalu, 2003; Linke en van Oostendorp, in voorbereiding), en de derivatie van de segmentinventaris (Hayes, 1999; Van der Hulst, 2012).

Dat brengt ons bij de segmentinventaris *an sich*. In **hoofdstuk 2** wordt voorgesteld dat eigenschappen van de segmentinventaris te groeperen aan de hand van twee aspecten van de inventaris: de *vorm* en de *structuur*. De vorm heeft betrekking op de manier waarop de inventaris gebruik maakt van de beschikbare fonetische ruimte, hoe de individuele leden zich tot elkaar verhouden in het licht van akoestische of perceptuele dimensies. Dispersion Theory (Liljencrants & Lindblom, 1972; Flemming, 2004), bijvoorbeeld, gaat over de vorm van de inventaris, en, zo stelt **hoofdstuk 2**, niet over de structuur. Voorbeelden van structurele eigenschappen van de inventaris zijn bijvoorbeeld 'gaten': segmenten die geen deel uitmaken van de inventaris terwijl dat, gegeven de kenmerken waar de taal gebruik van maakt, wel zou moeten kunnen. Denk bijvoorbeeld aan de klank /g/ in het Nederlands, die wij alleen kennen in leenwoorden ('goal', 'golf'). Daarnaast zijn er juist ook regelmatigheden waaraan inventarissen zich lijken te conformeren. Zo is er een tendens om symmetrisch te zijn (ondanks de gaten), om economisch om te gaan met kenmerken (zo veel mogelijk segmenten met zo min mogelijk kenmerken), en zijn de individuele leden van de inventaris vaak gelijkmatig verdeeld over de fonetische ruimte.

Nu we onderscheid hebben aangebracht tussen vorm-gerelateerde en structuurgerelateerde eigenschappen van de inventaris doet zich de vraag voor in hoeverre deze eigenschappen daadwerkelijk eigenschappen van de inventaris zijn, of, duidelijker gesteld, in hoeverre de inventaris zelf 'bestaat' als fonologische entiteit. Om een voorbeeld te noemen, dispersie-effecten zijn typisch aspecten van de inventaris als geheel: het is onmogelijk om de dispersie van een segment te bepalen. De vraag is echter of het bepalen (berekenen) van de fonetische dispersie behoort tot het domein van de mentale, fonologische grammatica. Hetzelfde geldt voor eigenschappen zoals economie en symmetrie. Een strikte en minimale visie op fonologie, zo wordt voorgesteld, houdt in dat fonologische computatie zich alleen bezighoudt met het coderen en decoderen van oppervlakte- en onderliggende vormen. Dat wil zeggen, met de derivatie van individuele woorden of morfemen, en dus niet met het berekenen van zaken die

zich op een systeemniveau afspelen. De inventaris als systeem, of als geheel, zou niet onderwerp moeten kunnen zijn van de fonologische grammatica, maar juist het gevolg (of: de output) daarvan. Voorstellen die daar niet aan voldoen worden ‘holistisch’ genoemd (de inventaris wordt ‘in z’n geheel’ geëvalueerd door de fonologische grammatica). Het ontwikkelen van een niet-holistische theorie over de segmentinventaris is een belangrijk doel van deze dissertatie.

Als we terugkijken naar het voorbeeld van Dispersion Theory, zien we dat die niet noodzakelijkerwijs holistisch is. Dispersie-effecten zijn daadwerkelijk aangetoond, maar daarmee is niet gezegd dat ze het gevolg zijn van de synchrone fonologische grammatica. Dat is ook direct het belangrijkste punt van kritiek op de OT-versie van Dispersion Theory die wordt voorgesteld in, onder andere, Flemming (2004). Na de behandeling van Dispersion Theory worden nog twee theorieën besproken. Deze zijn interessant omdat zij niet alleen gaan over de inventaris, maar ook een belangrijke leercomponent bevatten. De Parallel Bidirectional Phonetics and Phonology (Boersma & Hamann, 2008) is een omvangrijk model van zowel de productie als de perceptie van taal, waarbij alle niveaus van de derivatie worden voorgesteld als representaties, tezamen met de bijbehorende verzamelingen van constraints. Traditionele gemarkeerdheidsconstraints hebben betrekking op de derivatie van de onderliggende vorm naar de oppervlaktevorm, terwijl een andere set constraints betrekking heeft op de derivatie van oppervlaktevorm naar ‘Auditory form’: de zogenoemde cue constraints. Boersma en Hamann (2008) laten onder andere zien dat bepaalde dispersie-effecten gemodelleerd kunnen worden door cue constraints en gemarkeerdheids-constraints. Het voorbeeld van de verdeling van sibilanten op het continuüm van spectrale ruis komt weer terug in **hoofdstuk 5**, waar ook besproken wordt tot in hoeverre het model van Boersma en Hamann (2008) verenigbaar is met de Feature Co-occurrence Constraint-theorie die in dit proefschrift voorgesteld wordt. Het (voor ons) belangrijke punt van Boersma en Hamann (2008) is dat dispersie ontstaat door het leerproces; de auteurs maken hierbij gebruik van een specifiek leeralgoritme.

Datzelfde kan gezegd worden van de Modified Contrastive Hierarchy (Dresher, 2009; Hall, 2007). Deze theorie over contrast en inventaris is kort samen te vatten in wat Hall (2007) de ‘Contrastivist Hypothesis’ noemt: alleen contrastieve kenmerken zijn actief in de fonologie van de betreffende taal. Om te bepalen wat de contrastieve kenmerken zijn wordt een leeralgoritme voorgesteld, het ‘Successive Division Algorithm’. De (hypothetische) leerder begint met de hypothese dat alle klanken allofoon zijn; er is geen contrast. Wanneer de leerder ontdekt dat er een contrast is, selecteert deze een kenmerk (F) dat dit contrast het beste tot uitdrukking brengt en verdeelt de verzameling allofonen in tweeën: de ene onderverzameling wordt [-F], de andere wordt [+F]. deze stappen worden herhaald totdat er geen onderverzamelingen groter dan één zijn. Wanneer een onderverzameling nog maar één lid heeft, dan worden de volgende kenmerken niet meer toegekend aan dat lid. Op deze manier worden alle segmenten contrastief gespecificeerd. In **hoofdstuk 2** lezen we een uitge-

breide beschrijving van het algoritme, en hoe de Modified Contrastive Hierarchy een niet-holistische theorie van de inventaris voorstelt.

In **hoofdstuk 5** komen we terug op bovengenoemde theorieën, en kijken we in hoeverre deze verenigbaar of complementair zijn met het voorgestelde model. Dispersion Theory en de Modified Contrastive Hierarchy blijken niet goed te combineren met de Feature Co-occurrence Constraint theorie, waarbij bij die laatste het probleem komt dat het niet geheel duidelijk is wat de relatie is tussen de hypothetische leerder die het algoritme toepast en een ‘echte’ leerder (zie het laatste deel van **hoofdstuk 2** voor een verhandeling over het verschil tussen die twee in algemene zin). Een andere theorie die we al eerder zijn tegengekomen (in **hoofdstuk 1** en **hoofdstuk 4**), inductive grounding, maakt ook gebruik van Feature Co-occurrence Constraints, zij het in een iets andere zin, en levert een manier waarop deze constraints geleerd kunnen worden uit functioneel gemotiveerde fonetische kennis.

De constraints in voorbeeld 76 worden verder uitgewerkt in **hoofdstuk 3**. Enerzijds moeten theorieën over ‘volwassen’ grammatica een leerbaarheidstoets doorstaan, maar anderzijds moet een verwervingstheorie ook een beeld geven van de uiteindelijke staat, de grammatica van de volwassen taalgebruiker. In ons geval is de finale staat de (consonant)inventaris van het Nederlands, en deze verzameling segmenten wordt dan ook gebruikt als illustratie van de werking van de Feature Co-occurrence Constraint-theorie. **Paragraaf 3.2** geeft een overzicht van het Nederlands en de kenmerkverzameling die we gebruiken om de Nederlandse consonanten te specificeren, en laat zien hoe de Nederlandse consonantinventaris volgt uit een kleine groep kenmerken en een zeer beperkte verzameling Feature Co-occurrence Constraints. In het vervolg van dit hoofdstuk gaan we verder in op de logische en formele aspecten van de constraints (ze verwijzen naar exact twee kenmerken, er zijn slechts twee typen nodig, en wat telt als een kandidaat voor evaluatie door FCCs) en de kenmerken: monovalente kenmerken zijn restrictiever, sommige aspecten van klanken worden niet fonologisch gerepresenteerd (coronaliteit en plosiviteit), en een belangrijke implicatie voor kenmerktheorie is dat zogenoemde major class features zoals [Sonorant] te algemeen zijn om te kunnen functioneren in een FCC-systeem. Dit type kenmerken wordt dan ook niet gebruikt in de voorgestelde theorie, evenmin als een kenmerkgeometrie. Om niet belangrijke inzichten uit de kenmerkgeometrieliteratuur verloren te laten gaan wordt de theorie van ‘Feature Classes’ (Padgett, 2002; Yip, 2011) als alternatief voorgesteld.

## De verwerving

Gewapend met de inzichten over constraints en kenmerken die uit hoofdstuk 3, richten we ons in **hoofdstuk 4** op de verwerving van de Nederlandse consonantinventaris. Hierbij maken we gebruik van de data van een aantal kinderen in de CLPF-database (Fikkert, 1994; Levelt, 1994). Deze database bevat longitudinale spontane spraak en is dus uitermate geschikt voor onderzoek naar de ontwikkeling van fonologische kennis en kunde. In **paragraaf 4.2** lezen we over

de methoden die gebruikt zijn, de keuzes die gemaakt zijn om de verschillende stadia van ontwikkeling te onderscheiden, en de manier waarop de constraints bepaald zijn. Feitelijk is dat een iteratieve toepassing van het algoritme dat in **paragraaf 3.2.3** wordt voorgesteld. Deze methode blijkt keer op keer de juiste inventaris te genereren.

Het verschil in de toepassing van het algoritme uit **paragraaf 3.2.3** tussen **hoofdstuk 3** en **hoofdstuk 4** is dat in het laatste geval tijd een belangrijke rol speelt (per definitie, omdat we naar longitudinale ontwikkeling kijken). We nemen aan dat een constraint over de kenmerken [F] en [G] maar op één moment geactiveerd kan worden, te weten het moment dat de leerder het laatste van de twee kenmerken verwerft. Deze aanname blijkt correct, en aangezien het algoritme iteratief wordt toegepast en geen ‘geheugen’ heeft (dat wil zeggen: het weet niet wat de constraintverzameling in het vorige stadium was) is dat geen methodologisch artefact. Dat betekent ook dat het algoritme, wanneer het eenmaal een bepaald gat in de inventaris heeft beschreven aan de hand van een bepaalde constraint, dat niet later door een andere constraint doet.

Het is daarentegen wel zo dat het algoritme soms ‘redundante’ constraints introduceert, wanneer een gat op verschillende manieren kan worden verklaard. Hoewel dat op het eerste gezicht wellicht oneconomisch lijkt, is het daadwerkelijk een belangrijke bijdrage aan de flexibiliteit van het systeem. Noch het algoritme, noch de leerder kan in de toekomst kijken, dus noch het algoritme, noch de leerder kan weten of een constraint in een later stadium herroepen moet worden. Het kan zo zijn dat wanneer een bepaald gat in de inventaris door verschillende constraints verklaard wordt, een van die constraints later moet worden herroepen door de verwerving van een nieuw segment, terwijl het gat wel blijft bestaan. De andere constraint blijft dan actief. Een voorbeeld van zulke redundantie zien we in de uitgebreide beschrijving van het verwerkingstraject van Noortje (een van de CLPF-kinderen) in **paragraaf 4.3.1**.

Hoewel het algoritme dus over het algemeen zeer goed in staat blijkt om de juiste constraints te genereren, werkt het niet volledig foutloos. In een beperkt aantal gevallen wordt een verzameling constraints gegenereerd die meer segmenten toestaat dan er daadwerkelijk in de data aangetroffen worden (**paragraaf 4.5**). Een gedetailleerde blik op de CLPF-data leert dat de meeste van die gevallen het gevolg zijn van de inclusiecriteria; de betreffende segmenten worden wel degelijk geproduceerd, maar niet vaak genoeg, of in voldoende verschillende lexicale vormen, om in de uiteindelijke dataset terecht te komen. Dat illustreert een probleem dat inherent is aan dit type onderzoek: het in acht nemen van alle ruwe data maakt de analyse zeer kwetsbaar voor toevalligheden in de productie, terwijl het toepassen van inclusiecriteria altijd het risico met zich mee brengt dat een klein deel van de data onterecht niet meegenomen wordt. Het is aan de analist (en het werkveld) om hier een evenwicht in te brengen. Daarnaast dienen zulke onverwachte zaken als overpredicties als zogenoemde *red flags*: waarschuwingssignalen voor de analist om de ruwe data nogmaals gericht te bestuderen.

Niet alle gevallen van overpredicties blijken van genoemde aard te zijn. Een aantal gevallen is het gevolg van het feit dat we een ‘leeg’ segment hebben: eerder werd al genoemd dat coronaliteit en plosiviteit niet door kenmerken gespecificeerd worden. Dat betekent dat /t/ geen kenmerken heeft. Het is dus de fonetische realisatie van het lege segment. Wat geen kenmerken heeft, daar kan geen kenmerkconstraint tegen optreden. Een direct gevolg van die aanname over kenmerkspecificatie (of het gebrek daaraan) is dus de voorspelling dat /t/ vanaf het eerste moment aanwezig is in de inventaris van de taalleerder. Dit is het geval in verreweg de meeste bestudeerde gevallen. Uiteindelijk blijkt deze vorm van overpredictie alleen in Noortjes eerste stadium voor te komen. De andere gevallen van overpredictie zijn labiale fricatieven in de ontwikkeling van Noortje. Deze segmenten (/f,v/) vormen wat kenmerksamenstelling betreft een onderverzameling van een ander segment: /v/. De rest van paragraaf 4.5 behandelt de formele omstandigheden waarin overpredicties kunnen ontstaan, gegeven het constraintgenererende algoritme dat we hanteren.

Een belangrijk deel van dit proefschrift besteedt aandacht aan de vraag of de elementen waarvan gebruik wordt gemaakt – kenmerken en constraints – aangeboren of aangeleerd zijn. Deze vraag is bepaald niet nieuw, noch definitief beantwoord, maar in **paragraaf 2.3** behandelen we een aanzienlijke hoeveelheid literatuur over de aangeborenheid van kenmerken, en in **paragraaf 4.6** kijken we naar constraints. Anders dan we vaak zien bij discussies over dit probleem, vergelijken we hier niet verschillende voorspellingen die beide polen (nativisme versus maturationalisme) over concrete data doen, maar nemen we een definitie van aangeborenheid (Spelke, 2012), en toetsen we een grote hoeveelheid literatuur hieraan.

In **paragraaf 2.3**, waar we ons richten op kenmerken, maken we onderscheid tussen drie verschillende ‘rollen’ die kenmerken spelen in een grammatica. Allereerst worden kenmerken gebruikt om categorieën van klanken, wanneer die eenmaal onderscheiden worden, te labelen. Daarnaast zijn kenmerken de elementen die het lexicon bevolken: de representatie van lemma’s is, zo wordt meestal aangenomen (zie Lahiri en Reetz (2002) voor een expliciet voorstel in deze richting) opgebouwd uit ‘horizontaal’ en ‘verticaal’ (of: paradigmatisch en syntagmatisch) geordende kenmerken. Als laatste zijn kenmerken ook die zaken (naast prosodische categorieën) die door regels, constraints, of andere derivatieve of generaliserende mechanismen gemanipuleerd worden. Vanuit elk van deze drie perspectieven komt steeds het beeld naar voren dat aangeboren kenmerken een reële mogelijkheid zijn.

Met dezelfde definitie kijken we in **paragraaf 4.6** naar constraints. De vraag of constraints substantief aangeboren zijn blijkt moeilijker te beantwoorden, en er is niet veel experimenteel onderzoek beschikbaar. In de (vroeg) OT literatuur (Prince & Smolensky, 1993/2004) wordt aangenomen dat alle constraints universeel en aangeboren zijn, maar een theorie als Inductive Grounding (Hayes, 1999) laat zien dat ten minste sommige (kenmerk)constraints aangeleerd kunnen zijn, en tot in zekere mate universeel omdat ze door de leerder



geëvalueerd worden op basis van een (universele) fonetische moeilijkheid.

Onderzoek naar de vroege activiteit van constraints dat over prosodische categorieën gaat (bijvoorbeeld Inkelas en Rose, 2008) laat meer evidentie zien van aangeborenheid dan experimenteel onderzoek naar kenmerkconstraints (Jusczyk et al., 2002). Vandaar dat we, in tegenstelling tot het geval was bij kenmerken, ervan uitgaan dat Feature Co-occurrence Constraints niet volledig substantief (dat wil zeggen, gevuld met kenmerken) aangeboren zijn.

Dit proefschrift schaart zich in een traditie van kindertaalonderzoek waarin getracht wordt om zowel formeel als ontwikkelingsgericht een verklaring te zoeken voor de ontwikkeling van het menselijk taalvermogen. In **paragraaf 1.3** kijken we naar twee verschillende tradities van kindertaalonderzoek. Het onderscheid dat we hanteren, tussen het formele en het ontwikkelingsperspectief, wordt ontleend aan Ingram (1989).

De ontwikkelingstraditie richt zich voornamelijk op het beschrijven en verklaren van de taalontwikkeling van kinderen, en doet dat vaak vanuit een cognitief-psychologisch of psycholinguïstisch perspectief. Veel experimenteel perceptieonderzoek valt in deze categorie (dit wordt uitgebreid behandeld in **paragraaf 2.3**), maar ook productieonderzoek als dat van Ferguson en Farwell (1975) en van Ingram zelf (Ingram, 1981, 1988, 1989). De andere traditie, die van de formele aanpak, kenmerkt zich door een focus op het logische probleem van leerbaarheid, of het ontwikkelen van abstracte leeralgoritmes die niet in de eerste plaats bedoeld zijn om de daadwerkelijke ontwikkeling van het taallerende kind te modelleren (een duidelijk voorbeeld hiervan is het eerdergenoemde Successive Division Algorithm).

Zoals gezegd bestaat er ook een synthetische traditie waarin getracht wordt beide problemen (het ontwikkelingsprobleem en het leerbaarheidsprobleem) aan te pakken. Belangrijke voorbeelden van deze traditie zijn Smith (1973); Fikkert (1994); Levelt (1994); Beers (1995); Pater (1997) en Rose (2000).

Een meer recent voorstel in de synthetische traditie is Levelt en van Oostendorp (2007). Deze studie is de eerste aanzet tot het gebruiken van Feature Co-occurrence Constraints om fonologische verwerving te verklaren, en vormt daarmee de prelude naar het onderzoek waarvan dit proefschrift verslag doet.

De Feature Co-occurrence Constraint-theorie die in dit proefschrift wordt voorgesteld is verre van compleet, maar levert een belangrijke bijdrage aan de bestaande inzichten over zowel de klankinventaris als de verwerving daarvan.

