Building a Phonological Inventory

Feature Co-occurrence Constraints in Acquisition

Published by

LOT
Trans 10
3512 JK Utrecht
The Netherlands

phone: +31 30 253 6111
e-mail: lot@uu.nl
http://www.lotschool.nl

ISBN: 978-94-6093-163-5
NUR: 616

Building a Phonological Inventory

Feature Co-occurrence Constraints in Acquisition

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof. mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op woensdag 7 januari 2015
klokke 15.00 uur

door

Burght Marinus van 't Veer

geboren 15 augustus 1982
te Amsterdam, Nederland

Promotores:	Prof. dr. C.C. Levelt & Prof. dr. M. van Oostendorp
Promotiecommissie:	Prof. dr. R.W.J. Kager Dr. B. Köhnlein Dr. Y. Rose

The research reported here was conducted within the context of the NWO-funded project *Kindersprache, Aphasie und Algemeine Lautgesetze* Revisited (NWO 360-75-000).

for my children
for my parents.

Contents

Acknowledgements

First of all, I want to thank my advisors, Marc van Oostendorp and Claartje Levelt, to whom I owe so much. Somehow they had trust in me at times when I didn't. I have truly learned a lot form them. Crit Cremers has always been a mentor to me, and his profound thinking about language and other matters have educated and challenged me.

I also want to thank Bert Botma, Ricardo Bermúdez-Otero, Edoardo Cavirani, Colin Ewen, Daniel Currie Hall, Ben Hermans, Frans Hinskens, Peter Jurgec, Björn Köhlein, Jos Pacilly, Patrycja Strycharczuk, Shanti Ulfsbjorninn and many others for the influence they exerted and keep exerting on my phonological development. Special mention must be for Yvan Rose, for at least two reasons: first he gave the fields of phonology and language acquisition Phon, an impressive and important research tool without which I couldn't have done much of the work for this thesis. Also, Yvan has more then once provided incredibly detailed and well-motivated feedback. The amount of time and effort he invested is humbling.

PhD life at LUCL has been great, and I owe so much to my friends and colleagues: Enrico Boone, Martine Bruil, Lisa Cheng, Camelia Constantinescu, Elly Dutton, Andreea Geambasu, Margarita Gulian, Sita ter Haar, Annemieke Hammer, Pepijn Hendriks (thanks for the editorial help!), Caroline Junge, Anne van der Kant, OIga Kepinska, Björn Köhnlein, Martin Kohlberger, Elisabeth Koier, Maarten van Leeuwen, Bora Lushaj, Sara Lusini, Marieke Meelen, Beata Moskal, Khalid Mourigh, Victoria Nyst, Stanly Oomen, Bobby Ruijgrok, Niels Schiller, Mulugeta Tsegaye, Rinus Verdonschot, Stefan Verschoor, Jenneke van der Wal, Josh Wilbur, Jurriaan Witteman, Sima Zolfaghari and the 2009-2010 LEO board: Ron Pat-El, Benjamin Bilski, Olaf van Vliet, Judith Sandriman. Thanks to Kathrin Linke for putting up with me, my loud typing, my restlessness. For introducing me to and helping me with LaTeX, and for being side-by-side for so long. Kathrin, Sita ter Haar and me were given the opportunity to attend the 2009 LSA Summer Institute in Berkeley (for which I acknowledge a grant from the Leiden University Fund). I have learned a lot, and they were great companions.

I was very lucky to be able to spend a month at University College London with John Harris and Andrew Nevins in 2011. Ever since I met John in Berkeley, he has been a source of tremendous inspiration. His way of thinking about language and phonology profoundly influenced me. Andrews support and enthusiasm were highly infectious, I put a little present for him in this sentence. I'm grateful for having had the chance to meet Phillip Backley, Shanti Ulfsbjorninn, Sam Green, and the other folks at UCL. I'm also indebted to Arta and Bob, who opened their house for me, and for Lara for getting me in touch with them. Bob, you were an amazing person, so full of life, wisdom and amazement. The world misses your wonderful smile.

It's a special thing when colleagues become close friends. Rebecca Voll, Allison Kirk & Jessie Nixon have changed me through their friendship, taught me so much about the world and about myself. Discovery is what makes life interesting; it is the reward of science, art, and friendship. I can't thank you enough.

Thanks to Pieter, who repeatedly inquired how the book was going, and who selflessly hosted many Thursday night sessions. And to Titus, who did so before. And to Stefan, for keeping me sailing. We go a long way back, more then most other people mentioned here, and I am immensely grateful for that, as I am for the travels we made and hopefully will make. And simply for all the good times. Thanks to Ruben Bres, David Heijkamp, Corné Knape, Jasper Visser, and many others for all the fun, for your friendship and for keeping me (somewhat) aligned with the 'real world'. I am indebted to Irene Cécile, Paul Romijn, Christiaan van Tol, Hadi Abdi, to Abel Ponger, Roy Pesik and the others in the 2012-14 Fotovakschool class, and most of all to Roel Stevens, for introducing me to and supporting me in discovering an entirely different (or not so different?) world.

My mother has been a great support in so many ways. I can't begin to thank her enough for all she is, and all she's done, not only for me but also for Carl, Huyg, Lise and Erica. This will never go away.

Thanks also to my father and Nelleke for their perpetual support. And to my father for introducing me to the wonderful world of photography so many many years ago. Special thanks also to all grandparents for helping us out so much with babysitting: Opa Luc, Oma Anneke, Oma Liesbeth, Opa Mars and Oma Nelleke. I also want to thank my sisters and their spouses, Toos & Hille, Anna & Stephanie, for their support and all the fun we have.

Thanks to Erica. There for me always, and gave me the greatest gift of all, our children Carl, Huyg and Lise. Always supported me, no matter how selfish or idiotic my plans. I could not have done this without her.

Carl, Huyg and Lise, this book is dedicated to you. In moments of despair, you unknowingly light up the world and in moments of happiness, you make it even better. Your lust for life is wholesomely relativizing and incredibly contagious. Don't ever lose it – the world is your playground!

CHAPTER 1

Introduction

Every language has an inventory of phonological segments. Every child acquires a language. Hence, every child must acquire the inventory of phonological segments of her language. This, in a nutshell, is the topic of the current thesis: to arrive at a theory that describes the phonological segment inventory, and describes how it is acquired.[1] This theory is dubbed Feature Co-occurrence Constraint Theory. We shall develop a theory with minimal ingredients: distinctive features, a generator mechanism which (limitlessly) combines these features, and two types of constraints on the co-occurrence of features. One of the main innovations of the theory is that it views acquisition as the tandem development of representational and computational aspects of phonology, rather than focussing on one aspect in particular, as previous approaches tended to do. Each of the ingredients is in and of itself uncontroversial; combining them is uncontroversial, as well. This thesis attempts to be as explicit about assumptions and implications as is possible.

The Feature Co-occurrence Constraint Theory does not stand in isolation; it has a number of predecessors (see section 1.2). One of these, Hayes (1999), provides a phonetically motivated alternative to the innate constraint hypothesis (see also Hayes & Wilson, 2008; Adriaans & Kager, 2010). Furthermore, Fikkert and Levelt (2008) propose a lexicon-driven constraint generation mechanism (see section 4.6). In our theory, constraint templates are innately present, and these templates are populated (with features) and activated as soon as the conditions for population (i.e., the acquisition of two features) are present. In this sense, the current proposal is firmly couched within the tradition of gen-

[1] In this thesis, we consider only spoken language.

erative approaches to phonology.

The natural habitat for generators and constraints in contemporary phonology is Optimality Theory (Prince & Smolensky, 1993/2004, henceforth: OT). Most certainly, there is no objection to implement the theory of Feature Co-occurrence Constraints in that framework; in fact, we borrow many formalisms and insights from the rich Optimality Theory literature. First, there is a generator that supplies the full permutation of feature combinations, where feature combinations (in the simultaneous sense) stand for possible segments. This generator, to which we shall not devote a great deal of attention, has an obvious analogue in Optimality Theory: GEN. Next, Feature Co-occurrence Constraints are invoked and revoked; the revocation of constraints can be modeled in OT by constraint demotion. The invocation of constraints is somewhat less straightforward; in 'classical' OT, all constraints are thought to be universal and innate (Prince & Smolensky, 1993/2004), and while it has been proposed that in the initial state, markedness constraints collectively outrank faithfulness constraints (Gnanadesikan, 1995; Boersma, 1998; Tesar & Smolensky, 2000), some have argued for the opposite (Hale & Reiss, 1998). In their computational model, Boersma and Hayes (2001) show that (under some assumptions at least) an initial state in which constraints are unranked (or: ranked in the same stratum) can lead to a successful final state grammar. With respect to our current proposal, where constraints are invoked, we can assume that it makes no great difference how constraints are ranked in the initial state. We assume that constraints can only be invoked at one point in the development of the grammar: namely, the point at which the latter of the two features the constraint refers to is acquired (we will argue for innate features in chapter 2). Even if the constraints are innate and every possible Feature Co-occurrence Constraint is present in the initial state, we will only see its effect if its structural description (that is, both features) is present. Hence, while we assume that constraints are invoked (see section 4.6), even if they are in fact innate, their ranking prior to the invocation point is irrelevant.

Whereas OT provides a ranking scale for constraints (be it absolute or gradual (Boersma, 1998)), in other frameworks constraints are either active and not violated, or inactive. A theory of constraints that aspires to be compatible with both Optimality Theory and non-constraint-ranking theories must forego the possibilities offered by a full-fledged infinite constraint hierarchy, and be limited to constraints divided over no more than two strata: one for violated constraints, one for satisfied constraints. Constraints in the latter strata are those that are active in non-OT frameworks. We will see, in chapter 3, that a two-strata OT grammar results from generating the Feature Co-occcurrence Constraints necessary for any inventory, if no faithfulness constraints are considered. Hence, the theory is transportable between frameworks. In chapter 5 we will return to this issue.

1.1 Feature Co-occurrence Constraints: the proposal

As mentioned above, the Feature Co-occurrence Constraint Theory rests on features and constraints. The central idea is quite simple: the acquisition of the segment inventory can best be described in terms of the tandem development of both the representational and the output components of the phonological system (see also Levelt & van Oostendorp, 2007). The representational component is taken to be the lexicon, which is populated by lexical items consisting of structured bundles of monovalent features. Monovalency is not a choice on which the theory depends; it is, however, well motivated (section 3.4.1) and has repercussions for the system of constraints that we propose (chapter 3).

The output system is represented by a system of highly simple Feature Co-occurrence Constraints. Only two types of constraints are necessary; these are represented in the form of constraint templates in (1) below.

(1) a. *FG
assign a violation mark for every segment Σ iff [F] is in Σ and [G] is in Σ (*c-constraint*)

b. F→G
assign a violation mark for every segment Σ iff [F] is in Σ and [G] is not in Σ (*i-constraint*)

We call these constraints 'c-constraints' (where 'c' stands for 'co-occurrence'), and i-constraints (where 'i' stands for 'implicational'). With these two constraint templates, we will show that the inventory of adult language can be described (chapter 3), and likewise the acquisition of that inventory. The c-constraints are straightforward: they are violated by the combination in one segment of the features that are listed in the constraint's structural description. The addition of i-constraints is closely tied to the adoption of monovalent, rather than binary features (section 3.4.1). In monovalent feature theory, the complement of the presence of a feature is its absence, whereas in binary feature theory the complement of [+F] can be expressed by [-F]. Hence, if a feature [F] cannot co-occur without feature [G], this can be formulated in binary feature theory by *[+F, -G], which fits the ordinary c-constraint template. The expression [-G] is unavailable in monovalent feature theory, but the logical formula $\neg(F \wedge \neg G)$, which is the logical representation of the c-constraint mentioned above, is equivalent to the statement F→G, which is the logical representation of i-constraints (this argument will be explored further in chapter 3).

A crucial assumption about Feature Co-occurrence Constraints is that they can be invoked, but also revoked (or demoted). Many individual constraints remain active throughout the final state (adult grammar), but some are transient and only active for some period during acquisition. These constraints are invoked when a feature has been acquired, but is still limited in its combinational options. Revocation implies that the feature becomes more free in its association with other features.

Features are assumed to be strictly monovalent, and innate. In section 2.3 we will review a large body of literature that in some way pertains to the acquisition of features, and we will see that the evidence in favor of innateness is fairly strong. Monovalency will be defended more thoroughly in chapter 3. Furthermore, some properties are not represented in featural terms. There is no feature [coronal], and neither is there a feature [stop]. This implies that the empty segment (which cannot be excluded by the types of FCCs developed here) is interpreted by the phonetics as /t/.

Feature Co-occurrence Constraints act as a filter to separate legal from illegal segments, or grammatical from ungrammatical segments. Even illegal segments in one language, however, are still possible segments, and may be grammatical in another language. We define a 'possible segment' as any combination of distinctive features. This implies that there is no principled difference between universal gaps and language specific gaps: both arise through acquisition, and hence the motivation for universal gaps must be sought elsewhere.

Feature geometry and major class features are difficult to combine with the Feature Co-occurrence Constraint Theory as it is proposed in the current thesis, but at the same time, it remains important to be able to group features into classes, and refer to these classes. This issue is resolved by adopting a version of Feature Class theory (Padgett (2002), see also chapter 3).

Finally, the Feature Co-occurrence Constraint Theory is a theory not so much about the inventory, but about the gaps in the inventory that can occur. Every constraint describes a gap; a fully symmetrical inventory can be described with no FCCs at all. A drive towards as little active (or high ranked) FCCs as possible is also the drive towards more symmetry. A fully symmetrical system is also fully maximally economic, and every gap represents a decrease in economy.

We will see that the number of Feature Co-occurrence Constraints that are activated is relatively small. Only 19 are needed to accurately describe the inventory of Dutch, and the number for each stage in acquisition is of a similar magnitude. Compared to the impressively large inventory of possible constraints, this is number is perhaps in itself an argument against the innateness of substantive Feature Co-occurrence Constraints, instead of the innateness of FCC templates (as we assume here). Before turning to the main matter of the dissertation, we shall briefly consider the earlier uses of Feature Co-occurrence Constraints in adult phonology, and previous approaches to the acquisition of the segment inventory.

1.2 Earlier uses of Feature Co-occurrence Constraints in adult phonology

An early predecessor of Feature Co-occurrence Constraints is the system of Marking Conventions of Chomsky and Halle (1968, henceforth: SPE)'s chapter nine. These state universal constraints on feature co-occurrences, in both the paradigmatic and syntagmatic sense. Interestingly, the Marking Conventions are both violable (they indicate markedness preferences rather then hard restrictions), and universal. Many other conceptions of constraints in the phonological literature take them to be inviolable, language-specific generalisations over the (underlying) structure of morphemes (Stanley, 1967) or surface phonotactics (Shibatani, 1973). It is not until Singh (1987) and Paradis (1988) that the we once more encounter the idea of the universal phonological constraint.

1.2.1 Feature combinations in rule-based phonology

Admittedly not technically constraints by format, the Marking Conventions serve a similar role as our Feature Co-occurrence Constraints: to limit the generative power of a grammar. This is not *entirely* correct, however, because in SPE the marking conventions are "...not part of the grammar but rather conventions for the interpretation of a grammar"(Chomsky & Halle, 1968, p. 403) whereas in the current proposal, the Feature Co-occurrence Constraints *are* the grammar. In SPE, it is a system of highly formal ("overly formal" in the words of Chomsky & Halle, 1968) rules, constituting the phonological grammar, that are capable of generating much more than is attested; in our case, there is a generator function, much like GEN in Optimality Theory, that is equally blind to the "intrinsic content" of features.

Although the Marking Conventions in SPE all take the shape of implicational statements (like our i-constraints), the reliance on binary feature values presents us with an interesting reversal of the situation outlined in section 1.1 above. Whereas we adopt i-constraints to accommodate co-occurrence restrictions that could be expressed as c-constraint only if using binary features, some of the Marking Conventions in SPE utilise the implicational format (or rewrite format) to express what we would express in c-constraints. Take, for example, the "universal constraint on feature combination" that states that it is impossible to have segments that are both [+low] and [+high] (Chomsky & Halle, 1968, p. 404). Both features being positively valued, the constraint can be rephrased in FCC terms as *[low, high]. Chomsky and Halle (1968, p. 405), however, formalise the constraint by two Marking Conventions:[2]

(2) Marking Conventions
  a. Marking Convention 6VII
     [+low] $\rightarrow$ [-high]

[2]See also the paragraph on Kiparsky and Pajusalu (2003) below.

b. Marking Convention 6IX
[+high] → [-low]

In the theory proposed in the current thesis, the segment [low, high] is a possible segment and will be generated; it will also be ungrammatical by virtue of the c-constraint *[low, high], which, we predict, will be activated and remain active in the grammar of learners of every language where no mid vowels exist.[3]

Close in spirit and format to the universal Marking Conventions of SPE are the redundancy and predictability rules in the various flavours of Underspecification Theory (Archangeli, 1984, 1988, etc.). These too serve to limit the generative power of the derivational mechanism, and they do so in a similar way: by limiting the amount of information specified in the input of phonological rules, the range of possible output structures is likewise limited. Among the differences between these types of rules and the SPE Marking Conventions is that, typically, the redundancy and predictability rules of Underspecification Theory are part of the set of operations performed by the grammar. This includes insertion *during* (or as part of) derivation.

As in SPE, redundancy and predictability rules are not constraints but rather implicational statements. In (3) we see an example of fill-in rules in Radical Underspecification (example from Archangeli, 1988). A five-member vowel inventory is specified as in (3a), and the missing values are inserted by the set of rules in (3b).

(3) Fill-in rules in Radical Underspecification

a. Inventory and specifications

	i	e	ɑ	o	u
[high]		-		-	
[low]			+		
[back]				+	+
[voice]					

b. Fill-in rules

i. [+low] → [-high]
ii. [+low] → [+back]
iii. [ ] → [-low]
iv. [ ] → [+high]
v. [ ] → [-back]
vi. [ ] → [+voice]

The main difference between the rules in (3b) and the Feature Co-occurrence Constraints proposed here, apart from the fact that the former are not technically constraints, is that FCCs act as a filter; they come in *after* the fact

[3] It should be noted that mid vowels are represented as the presence of both height features in the current privative approach, contrary to the standard practice in binary feature theory, where mid vowels are [-high, -low].

whereas fill-in rules are applied in the course of derivation. Fill-in rules do not ban what is already generated; they postpone certain types of generation to the point that it no longer applies. We will see that in the current proposal, some traits are unspecified (as in Underspecification Theory), but unspecification is permanent, and not undone by the FCCs. Feature Co-occurrence Constraints do not fill in, they filter out. Another way of stating the difference is to say that FCCs do not limit the generative power of the grammar *per se*, but rather undo some of it.

1.2.2 Constraints on feature combinations in harmony

Proper constraints on the co-occurrence of distinctive features in both OT and non-Optimality Theory phonology we encounter in the treatment of (vowel) harmony. In a government-based account of vowel harmony, Van der Hulst and Smith (1986) and Van der Hulst (1988) employ feature co-occurrence constraints in combination with strict locality to model opacity (see also Piggott & van der Hulst, 1997). In this analysis, the incompatibility of, for example, [low] and [ATR] in the surface vowel inventory blocks [ATR] spreading when a [low] vowel is encountered; strict locality prohibits it from looking further. It should not go unmentioned that these authors employ co-occurrence constraints in a way that is remarkably similar to the current proposal. Van der Hulst employs elements rather than features, and discusses vowel inventories rather than consonant inventories (as the subject of his study is vowel harmony), and does not explicitly limit the types of constraint as we shall do in chapter 3, but otherwise his system is very close indeed to the current proposal.[4] The spirit of Van der Hulst (2012) is very similar to the theory explored in Van der Hulst and Smith (1986) and Van der Hulst (1988), namely that the behaviour of non-alternating vowels in vowel harmony languages is predictable: inert vowels that are compatible with the harmonising feature are transparent, whereas those that are incompatible with the harmonising feature are opaque. In Van der Hulst (2012), incompatibility is explicitly modeled in terms of FCCs.

Let us move on to an Optimality Theoretic account of vowel harmony in which constraints on feature co-occurrences play an important role. Kiparsky and Pajusalu (2003) aim to derive a typology of neutral vowels in [back] harmony. They do so from the Factorial Typology of a harmony driving constraint (Agree[F]), positional faithfulness constraints and constraints on feature co-occurrence. The latter are anti-harmonic: they act against the surfacing of marked segments:

(4) Featural markedness constraints in Kiparsky and Pajusalu (2003)

a. [–low, –round] → [–back]
   If a vowel is non-low and ungrounded, it must be front

[4] It should be noted that Van der Hulst (2012) also uses syntagmatic FCCs, see the discussion on Hayes (1999).

b. [–back] → [–low, –round]
If a vowel is front, it must be non-low and unrounded

Although similar to our Feature Co-occurrence Constraints, the featural markedness constraints in Kiparsky and Pajusalu (2003) are different in several respects: first, they refer to three features rather than to two (see section 3.3.1 for a discussion). Second, they only act against marked segments, whereas our Feature Co-occurrence Constraints are blind to markedness: every combination of two features is a possible candidate for an FCC.

Feature Co-occurrence Constraints are employed in the studies of Turkish vowel harmony by Padgett (2002) and Linke and van Oostendorp (in preparation). The facts of Turkish vowel harmony, very briefly, are that all suffix vowels agree with the stem vowel in terms of backness, and high vowels additionally agree in roundness (remember that Turkish has a perfectly symmetrical eight-vowel inventory in stems). Building on substantial evidence, Padgett (2002) argues that both forms of harmony should be analysed as one and the same process, dubbed 'color' harmony, where [color] stands for the natural class formed by [back] and [round] vowels.[5] The fact that low vowels do not harmonise in roundness is captured by the Feature Co-occurrence Constraint *[low, round]. Padgett's proposal is couched in Optimality Theory, and the co-occurrence constraint outranks the constraint enforcing [color] harmony. Low round vowels do occur, however, albeit that their distribution is limited to stems. Hence, a set of faithfulness constraints for [low] and [round] outranks both the harmony and the co-occurrence constraint.

Linke and van Oostendorp (in preparation) employ a set of constraints that is minimally different from the set in (1). Using the same format of c-constraints, the definition is slightly different (the formulation in (5) is my own paraphrase of Linke and van Oostendorp (in preparation)'s constraint definition):

(5) a. *FG
assign a violation mark for every feature [F] that co-occurs with feature [G] (*c-constraint*)

Although the different formulation is not important to the proposal made in the current theory, it is crucial to Linke and van Oostendorp (in preparation) that violations are assigned on a per-feature base. This is because features that violate a c-constraint in the stem, but do this to avoid violating a higher ranked faithfulness constraint, do not incur additional violations when spreading to suffix vowels. Under the segment-based definition in (1), the c-constraint in a harmonic form would be violated twice: once by the stem vowel, and once by the suffix vowel. Under the alternative definition, conversely, the fact that the

[5] A more extensive discussion of Padgett's proposal will be provided in chapter 3. The reader is reminded that in Element Theory (see, e.g., Harris, 1994; Harris & Lindsey, 1995; Backley, 2011) both backness and roundness are represented by one and the same element, |U|.

feature in the stem vowel is linked to two segmental slots is of no effect to the amount of violations: there is only one offending feature, and the number of association lines is irrelevant for the c-constraint.

1.2.3 Feature Co-occurrence Constraints and specificational variation

Laterals are notoriously variable in their typological behaviour. In her overview, Yip (2011) proposes to capture this behaviour by means of two universal ranking schemes involving simple feature co-occurrence constraints referring to the feature [lateral]; one for place (6a), one for major class (6b):

(6) Universal ranking schemas for [lateral]

a. Place of Articulation
[Lateral, Labial]≫[Lateral, Dorsal]≫*[Lateral, Coronal]

b. Major classes
[Lateral, Obstruent]≫[Lateral, Sonorant]

By interspersing these universal rankings of c-constraints with faithfulness constraints, different types of inventories are derived. For example, if faithfulness is ranked below the three constraints in (6a), the resulting inventory will have either no laterals at all (as in Maori), or placeless laterals only (Cambodian). Ranking faithfulness between *[Lateral, Dorsal] and *[Lateral, Coronal] rules out labial and dorsal laterals, but allows for laterals with a coronal place specification (Yip gives English as an example). Interaction with Share-F constraints results in a typology of lateral spreading behaviour.

Although Yip (2011)'s constraints are formally very similar (if not equal) to our c-constraints, it must be noted that the way in which they are employed is conceptually somewhat different. The main difference is that the constraints are ranked with respect to each other (or, more correctly, with respect to faithfulness or spreading constraints) in a universal hierarchy. Nevertheless, Yip (2011) shows that c-constraints can be used to derive inventories.

So far we have mostly seen variants of c-constraints, or i-constraints to harness the effects of c-constraints (as in Kiparsky & Pajusalu, 2003). An early OT-based use of i-constraints is Itô, Mester, and Padgett (1995), who focus on the effects of the constraint [sonorant]→[voice]. Apart from stating that the presence of [sonorant] implies the presence of [voice], Itô et al. (1995) note that the constraint states that voicing is redundant, or non-contrastive, for sonorants: sonorants are always voiced. From this, they deduce that voicing is not *licensed* in sonorants; and in interaction with the constraint License[voice], which requires that the feature [voice] always be licensed, the implicational constraint drives surface underspecification of sonorants for [voice]. At the same time, there is no incompatibility between [voice] and [sonorant] (which would follow from a c-constraint such as *[sonorant, voice]), which would equally derive phonologically voiceless sonorants at the surface.

The discussion in Itô et al. (1995) takes place in the context of Yamato Japanese, where the initial obstruents of the second member of compounds becomes voiced (a process known as Rendaku) *unless* there is a voiced consonant elsewhere in the same morpheme (a generalisation known as Lyman's Law). An example of Rendaku is given in (7a), whereas (7b) shows how Lyman's Law interferes.

(7) Compounds in Yamato Japanese
  a. yama tera → yama+dera 'mountain temple'
  b. ore kugi → ore+kugi 'broken nail' (*ore+gugi)
  c. ori kami → ori+gami 'paper folding'

Crucially, example (7c) shows that nasals do not block Rendaku; hence, they must be underspecified for [voice].

The problem is that Yamato Japanese has a different property involving nasals and [voice], which leads to what Itô et al. (1995) call an 'underspecification paradox'. This property is post-nasal voicing, as in (8).

(8) Post-nasal voicing in Japanese
  a. tombo 'dragonfly' (*tompo)
  b. /yom+te/ → yonde 'reading'

Example (8a) shows that post-nasal voicing holds true of monomorphemic forms, whereas (8b) shows that it is active in derived forms, as well. The paradox is that whereas nasals act as voiceless for Rendaku and Lyman's Law, they seem to act as phonologically voiced in post-nasal voicing.

Without doing justice to the full analysis in Itô et al. (1995), the solution lies in the interpretation of the implicational constraint [sonorant]→[voice] (which, for reasons we will not go into here, only pertains to nasals in their analysis). The implication entails a 'licensing cancellation': as indicated above, sonorants do not independently license [voice], but they are not incompatible with [voice] either. This leads to the non-specification of sonorants for [voice] in the general case. Obstruents, however, are able to license [voice]. In NC clusters, therefore, the nasal acts as if it is voiced, because it is linked to a [voice] feature. That same feature is also linked (by spreading) to the obstruent, which licenses it (remember that high-ranked Licence[voice] prohibits that the optimal candidate contains unlicensed instances of [voice]).

Although Itô et al. (1995) employ an implicational constraint that is formally similar to the i-constraint in the current thesis, there are important differences in interpretation. We will see, in chapter 3, that the description of the Dutch consonant inventory involves one i-constraint: [approximant] → [continuant]. We shall follow Itô et al. (1995) in observing that this constraint (when unviolated) states that [continuant] is redundant on [approximant], but we do not adhere to their notion of 'licensing cancellation', nor will we explicitly state any licensing requirements for the realisation of features in the surface form. Hence, in Feature Co-occurrence Theory, [apprx]→[cont] does not entail that approximants are un(der)specified for [continuant]. Incidentally, an interesting learnability issue arises under the Itô et al. (1995, et passim) interpretation of implicational constraints: if sonorants are surface-underspecified for [voice] except in NC clusters, then how will the learner know that [sonorant]→[voice]?

1.2.4 Feature Co-occurrence Constraints and the inventory

We have now seen a number of specific applications of feature co-occurrence constraints in Optimality Theory, from vowel harmony to typological variation and questions regarding (under)specification. Before we turn to the an overview of earlier approaches to the acquisition of the inventory, it must be mentioned that from the early days of OT, constraints on feature combinations were understood to govern the structure of the adult inventory (see, for example, Kager, 1999 who uses feature co-occurrence constraints in various forms). At the same time, employing constraints on feature co-occurrences to model the consonant inventory is by no means a practice exclusive to Optimality Theory; the reader is reminded of the discussion on Van der Hulst (2012) in section 1.2.2.

In section 4.6, we will discuss the theory of Inductive Grounding (Hayes, 1999) in some detail. What is important to our present discussion is that Hayes (1999) uses both paradigmatic and syntagmatic feature co-occurrence constraints, where I take 'paradigmatic' to mean infra-segmental, and 'syntagmatic' to mean intersegmental.[6] Four types of constraints are proposed to be emergent:[7]

(9) a. *[F][G]
assign a violation mark for every output where [F] precedes [G]

b. *[F]
assign a violation mark for every [F]

c. *[F, G]
assign a violation mark for every co-occurrence of [F] and [G]

d. *[F] unless [G]
assign a violation mark for every [F] that occurs without [G]

The latter two constraints are directly related to our c-constraints and i-constraints, respectively; the second will be shown in chapter 3 to be a special version of a c-constraint, and the first is a syntagmatic constraint, pertaining

[6] Kiparsky and Pajusalu (2003) follow the same nomenclature.

[7] Hayes does not provide formal definitions, and to avoid going beyond a general impression into a distracting discussion about technicalities the definitions given here are admittedly somewhat sloppy.

to phonotactic restrictions (which we are not primarily concerned with in the current thesis).

Hayes' constraints live in an OT ecosystem, and function much to the same end as the FCCs proposed in the current thesis, namely to limit the range of grammatical combinations of features. An important difference is that Hayes' proposal builds on a notion of phonetic difficulty, represented in a 'difficulty map'. Of all possible feature co-occurrence constraints, only those are selected to participate in the actual grammar that are more 'effective' than constraints that are minimally different, where effectiveness is measured by how well the constraint predicts the relative difficulty of two points on the difficulty map (the reader is referred to section 4.6 for a more elaborate discussion of Inductive Grounding).

1.3 Two perspectives on the acquisition of the segment inventory

The acquisition of language is of interest to science in at least two ways: first, the process it worth studying in its own right. Acquiring a language implies interacting with the world, uncovering knowledge and codifying it, *et cetera*. At the same time, any theory of the final state, that is, the adult grammar, must show that it is learnable for the theory to be a candidate model of human linguistic competence. These two perspectives do not always converge, and consequently Ingram (1989) distinguishes between the 'developmental problem of language acquisition', which concerns the study of language acquisition in the first sense, and the 'logical problem of language acquisition', concerning learnability issues.

In this section, we will review some of the past literature on the acquisition of the segment inventory, classifying studies in one of Ingram's two categories, or, when appropriate, as aiming for a synthesis. It is perhaps to the latter category that the current study belongs.

We have already mentioned Jakobson (1941/1968), the study that directly inspired much of modern day language acquisition research. As mentioned by Ingram (1989, section 6.5.2) in his thorough review of Jakobson's proposal, the Jakobsonian theory of language acquisition is incomplete and imprecise at some points. In fact, Ingram proposes to read Jakobson (1941/1968) as the *outlines* of a theory rather than a theory in its own right – a proposal which I think is correct. Nevertheless, *Kindersprache* set the agenda for language acquisition research for decades after its publication, or after the publication of the English translation in 1968. One of the most spectacular claims made by Jakobson is that the acquisition of *oppositions* ('contrasts') is fixed and universal in terms of precedence. That is to say, the Jakobsonian hypothesis with respect to the acquisition of oppositions is that the order of acquisition proceeds such that the greatest opposition is acquired first, and subsequent oppositions are acquired in decreasing orders of magnitude. The point of departure for the child is the CV syllable, meaning that the opposition between consonants and vowels is prime. For consonants, the first two oppositions are nasal vs. oral and labial vs. dental. The first vocalic opposition is high versus low, or /i/ versus /a/. The following vowel stage introduces either a front-back opposition (resulting in /i, u, a/) or a further refinement of the height dimension (such that the child has /i, e, a/).

This claim is a rather strong one at first glance, and it has triggered a host of reactions. However, as pointed out by Ingram (1989), it is not a straightforward hypothesis to test, for being somewhat underdeveloped and not very specific in some aspects. The Jakobsonian hypothesis holds that the order of acquisition of phonemic contrasts is related to typological markedness statements, or, in the words of Jakobson (1941/1968), 'laws of irreversible solidarity'. A result of this claim is that a great deal of importance is ascribed to the order of acquisition of the segment inventory, and even though the idea that the same principles govern both the order of acquisition and typological distribution patterns was not always acknowledged explicitly in subsequent research, the order of acquisition question remained high on the agenda.

In this thesis, we will not have much to say about the order of acquisition. We will make little or no claims about the relation of order of acquisition and the inherent content of features, or their (relative) markedness. Rather than focusing on the order of acquisition of the segment inventory, we will extend the Jakobsonian hypothesis to the domain of the *phonological mechanisms underlying acquisition.* In short, one of the central claims of the current thesis is that the acquisition of the segment inventory can be described in the same terms as the structure of the adult inventory, where these terms are formalised as the Feature Co-occurrence Constraints in example (1) (chapter 1).

Having said that, it must be acknowledged that the (universality of) order of acquisition of the inventory remains a research topic in its own right. Perhaps the main reason for that is that the variation that has been shown to exist between children (both in the same and in different language communities) indicates that the order of acquisition of the inventory is not the most promising avenue for investigating the Jakobsonian claim; in addition, the order of acquisition question needs to be tackled employing a uniform methodology for analysing acquisition and adult phonology; a goal of the current thesis is to contribute to the development of such a methodology.[8]

1.3.1 The developmental problem of language acquisition

A highly influential study of the development of the segment inventory is Ferguson and Farwell (1975), who introduce the notion of Phone Classes, which

[8]With respect to the order of acquisition of the inventory, this is perhaps the moment to note that Levelt and van Oostendorp (2007) did not find that any of the three measures of distributional frequency is a good predictor, either. Hence, the factors determining the order of acquisition remain to be determined.

can be ordered into Phone Trees. Phone classes are groupings of variants of sounds (the authors focus on word-initial consonants, as do we), such that all variants occurring in the same word are grouped together in a phone class, together with all variants of words with the same initial consonant. Example (10) gives a phone class for one of Ferguson and Farwell (1975)'s subjects, T.:

(10)	*Phone class*	*Lexical items*
	[b ∼ β ∼ bw ∼ p^h ∼ ɸ ∼ ∅]	baby, ball, blanket, book, bounce, bye-bye, paper

As we can see, the adult target forms are all labial-initial words, and so are all the variants in the phone class (with the obvious exception of the null-realisation). Phone classes are construed for the entire lexicon, and by listing the phone classes on a horizontal axis, and subsequent recording sessions on a vertical axis, the longitudinal development of a child's inventory can be plotted. Phone classes in successive recording sessions are connected when they contain the same words, the result of which is a so-called Phone Tree (the graphic representation of Phone Trees is rather complex, or, in the words of Ingram (1989), they "look like the wiring diagram of a television set" (p. 201). For this reason, the reader is referred to either Ferguson & Farwell, 1975 or Ingram, 1989, p. 202 for examples).

Ingram (1989) notes a number of problems with phone trees, the most important of which is their sensitivity to variation of one single lexical item. Going back to example (10), there is no way of telling whether variation in the first column corresponds to different tokens of each type in the second column, or whether tokens belonging to one type exhibit more variation than those belonging to other types. Furthermore, with its focus on surface variability, it is difficult to see what phone trees are meant to represent other than exactly that, surface variability. With respect to the latter criticism, it should be noted that Ferguson and Farwell (1975) explicitly reject the continuity hypothesis and propose that the early lexicon (at least up to about containing 50 items) is organised no further than individual word forms (i.e., without any type of sub-segmental representational units or rules). Phonological organisation and representation is constructed by the individual during the course of linguistic development. If this is true, however, the question arises whether it is warranted to group into phone classes surface variants of adult targets if these occur in different lexical items.

Nevertheless, one of the successes of Ferguson and Farwell (1975) and successive applications of the phone tree methodology was that it showed systematically that individual differences in the order of acquisition exist within learners of the same language, and hence, that the strong Jakobsonian hypothesis must be amended.

A different method for analysing the acquisition of the segment inventory is developed by Ingram (1981, 1988, 1989). The aim of Ingram's program is virtually opposite to that of Ferguson and Farwell (1975): whereas the latter incorporate as much surface variability as possible, and reject the hypothesis that this reflects some underlying unity (although their method suggests otherwise), Ingram explicitly aims to filter out variability in order to obtain insights in the child's phonological competence.

Ingram (1989) lists a number of criteria for selecting types on which to base the assessment of the inventory. These are reproduced below:

1. If a phonetic type occurs in a majority of the phonetic tokens, select it.
2. If there are three or more phonetic types, select the one that shares the most segments with the others
3. If there are two phonetic types, select the one that is not pronounced correctly
4. If none of the above work, select the first type listed

Using these criteria as a heuristic, a representative sample of words can be selected, without incorporating incidental performance-induced variants. This is something that Ferguson and Farwell (1975) do not attempt, or consider to be relevant. The method proposed in Ingram (1981, 1988, 1989) is similar in spirit to the one we employ in the current thesis (see section 4.2). The developmental problem of language acquisition is also tackled frequently from the perception point-of-view. We will discuss a large body of literature below (see chapter 2.3)

1.3.2 The logical problem of language acquisition

As pointed out by Levelt (1994, p. 3), the logical problem of language acquisition can in principle be resolved without reference to actual child language data whatsoever. As noted before, in order to obtain cognitive credibility, any theory of the final state (adult grammar) must provide an account of learnability. Learning algorithms for Optimality Theory, for example, are proposals to resolve the logical problem of language acquisition (Tesar & Smolensky, 2000; Boersma, 1998; Boersma & Hayes, 2001), even if they are shown to be adequate predictors of actual language acquisition (Boersma & Levelt, 1999).

In section 2.2 we will see the learnability argument of Boersma and Hamann (2008), which pertains to the acquisition of the segment inventory. Perhaps the most important and influential solutions to the logical problem of language acquisition is the Modified Contrastive Hierarchy (Dresher, 2009), which employs the Successive Division Algorithm to arrive at the underlying structure of the segment inventory. We will devote considerable attention to both theories throughout the thesis, but a brief introduction is in order at this point.

The Modified Contrastive Hierarchy holds that only contrastive features are specified, in other words, that rules or constraints can only refer to contrastive features. Hall (2007, p. 20) goes even further, and formalises the 'contrastivist hypothesis':

(11) Contrastivist Hypothesis
The phonological component of a language L operates only on those features which are necessary to distinguish the phonemes of L from one another.

Superficially, the Modified Contrastive Hierarchy is similar to various forms of Underspecification Theory, and to be sure, it does incorporate underspecification. It goes beyond Underspecification Theory (Archangeli, 1984, 1988, among others), however, in that features may not be filled-in at some stage in the derivation to allow for certain rules to operate on them. The entirety of representational information is codified in the lexicon, and no more is codified than what is necessary to uniquely specify each member of the inventory.

The logical problem of language acquisition in this respect can be rephrased as the problem of algorithmically arriving at the correct feature specifications for the forms in the lexicon. After showing that the so-called 'pairwise algorithm' (the practice of determining contrast based on the comparison of minimal pairs) fails to deliver, Dresher (2009, footnote 3) proposes that contrast and feature specification is determined via the Successive Division Algorithm:

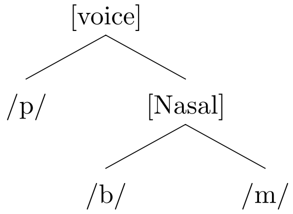
(12) Successive Division Algorithm
a. In the initial state, all tokens in inventory I are assumed to be variants of a single member. Set I = S, the set of all members.
b. i. If S is found to have more than one member, proceed to (c).
   ii. Otherwise, stop. If a member, M, has not been designated contrastive with respect to a feature G, then G is *redundant* for M
c. Select a new n-ary feature F, from the set of distinctive features. F splits members of the input set, S, into n sets, $F_1 - F_n$, depending on what value of F is true of each member of S.
d. i. If all but one of $F_1 - F_n$ is empty, then loop back to (c). (That is, if all members of S have the same value of F, then F is not contrastive for this set.)
   ii. Otherwise, F is *contrastive* for all members of S
e. For each set F_i, loop back to (b), replacing S by F_i.

In prose, the Successive Division Algorithm assumes an initial state in which all members of the inventory I are present, but allophonic to each other. Next, a feature is selected (e.g., [±voice]), and the set of inventory members is exhaustively assigned to either [+voice] or [-voice]. If the inventory contains no more than two members, the resulting subsets (designated by [+voice] and [-voice] respectively) each contain one uniquely specified member. However, if I contains more than two members, the resulting subsets must be further subdivided.

Take, for example, the subset inventory of labial stops in French (we adopt this example from Dresher, 2009, who adapts it from Jakobson & Lotz, 1949), which consists of three members: /p, b, m/. This subset can be specified by using two features: [±voice] and [±nasal], and hence, two different feature orders are possible [±voice]≫[±nasal], or [±nasal]≫[±voice]. Example (13) gives a graphical representation of the first option, together with the resulting feature specifications, and example (14) does the same for the second possibility (where branching to the left implies the negative value of the feature, and branching to the right implies the positive value).

(13) [±voice]≫[±nasal]

a. Feature hierarchy

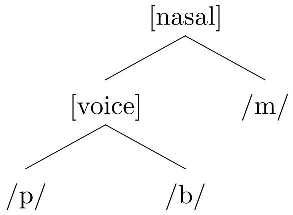


b. Contrastive specification

	p	b	m
[voice]	-	+	+
[nasal]		-	+

(14) [±nasal]≫[±voice]

a. Feature hierarchy



b. Contrastive specification

	p	b	m
[voice]	-	+	
[nasal]	-	-	+

For every set of n features, there are $n!$ possible orderings. The (hypothesised) learner and the linguist are thus faced with the question which feature to select first, which second, *et cetera*. Multiple strategies are possible (trial-and-error, parallel evaluation, to name but a few), but the crucial point is that the source of the evidence is always the same: contrast and phonological activity. In terms of examples (13) and (14), let us assume that the learner has found evidence that the three members of the subset inventory are contrastive. If there is evidence that /p/ functions with /b/ to the exclusion of /m/, the order in [±nasal]≫[±voice] is the correct one, as it assigns [-nasal] to both obstruents, whereas under the ordering [±voice]≫[±nasal] /p/ remains unspecified for [nasal] and hence does not form a natural class with /b/.

1.3.3 Syntheses

The deductive methods employed to tackle the developmental problem of language acquisition and the theory-driven approaches to the logical problem are both valid and valuable scientific enterprises in their own right, but ideally, the two should converge. We briefly mentioned Boersma and Levelt (1999) who show a convergence of the acquisition of syllable structure in Dutch and the predictions made by the Gradual Learning Algorithm (Boersma, 1998; Boersma & Hayes, 2001). Dresher (2009), too, mentions that the Successive Division Algorithm predicts the same feature hierarchy when applied to the inventory of Dutch, as the order of acquisition of distinctive features reported in Fikkert (1994), but it should be noted that we have no *a priori* reason to assume that the hierarchies derived by the Successive Division Algorithm represent the developmental order as uncovered in longitudinal child language studies. We will have more to say about this in chapter 5.

Having said that, a number of studies do tackle the developmental problem of language acquisition informed by hypotheses derived from the logical problem of language acquisition (Smith, 1973; Fikkert, 1994; Pater, 1997; Rose, 2000, among many others). Two of these specifically concern issues pertaining to the acquisition of the segment inventory and hence merit some discussion. Incidentally, both also concern the acquisition of Dutch.

Focusing on place of articulation (PoA) features exclusively, Levelt (1994) develops an analysis of the developing phonology of Dutch by means of a system of output conditions. In her system, features are specified only for words in their entirety, whereafter the left word-edge may become specified independently, and finally, the nucleus and right edge may receive independent specifications for PoA. Under this analysis, combined with a feature system in which vowels and consonants are specified by the same feature set (Levelt employs the model of Lahiri & Evers, 1991) Levelt (1994) shows that consonant harmony, a process previously analysed as long-distance assimilation (Vihman, 1978; Stemberger & Stoel-Gammon, 1989, 1991), is in fact subject to strict locality (i.e., adjacency) restrictions: the harmonising features are always shared between both the harmonising consonants and the intervening vowel. At the same time, specific output constraints limit the distribution of features over the various positions that become available to the child stepwise: initial dorsals are forbidden, and [labial] must be aligned with the left edge.

As we can see, Levelt (1994) is not concerned primarily with the acquisition of the inventory, but with her focus on the developing system of place of articulation features, she does provide a contribution. Furthermore, the investigations in Levelt (1994) are theoretically informed and data oriented, and as such, the author tackles both the developmental and the logical problem of language acquisition.

In a cross-sectional study, Beers (1995) analyses the acquisition of the Dutch inventory in both normally developing and language-impaired children. Like Levelt (1994), Beers assumes a model of feature geometry, and predicts that the acquisition of a system of contrasts that characterises the inventory proceeds in terms of a top-down fashion along the lines of that geometry. Higher-level features indicate more coarse contrasts (e.g., [labial] ∼ [dorsal] contrast is represented higher in the geometry then the [nasal] ∼ [lateral] contrast, and hence, it is predicted to be acquired earlier). Acquisition patterns are grouped in three categories: 'expected', 'unlikely', and 'abnormal', in descending degrees of convergence with the feature geometry. The multitude of children in the study display patterns that fall under the 'expected' ledger.

A major difference with the current study is that Beers (1995) focusses on the system of contrasts, rather than on the inventory proper. What this means is that when evidence of a contrast (say, [labial] ∼ [coronal]) at any manner of articulation reaches the pre-set criterion, this is sufficient for that contrast to be considered acquired. In other words, Beers focusses less on gaps, and more on the underlying representation. The difference with the current study is mainly that we propose, closer in spirit to Levelt (1994), that the acquisition of the segment inventory is best described in terms of a tandem development of underlying representations and output constraints, rather than focusing on the representational aspect of acquisition exclusively.[9]

As we discussed above, Levelt and van Oostendorp (2007) first proposed the idea that the acquisition of the inventory can be described in terms of the acquisition of features and constraints on feature combinations. We aim to contribute to resolving both the developmental and the logical problem of language acquisition – limited to the consonant inventory, that is.

1.3.4 Levelt and van Oostendorp (2007)

The current study is based on a pilot study published as Levelt and van Oostendorp (2007). Before exploring the formal properties and practical application to child language data in following chapters, we must briefly reflect on the similarities and differences between the Levelt and van Oostendorp study and the theory proposed here.

Apart from the theoretical coverage, which is both deeper and broader in the current study, and some minor differences such as the different feature sets, there are two more fundamental differences.

First, the difference in the features employed is more than minor in at least one respect, namely that Levelt and van Oostendorp adhere to a full three-way place specification, whereas we have adopted coronal underspecification. As we will see, this choice leads us to predict the early acquisition of /t/. It should be noted that Levelt and van Oostendorp assume underspecification of non-continuancy, as we do in the present thesis (that is to say, there is no feature

[9] Another difference is that we will have little to say about the order of acquisition.

[stop]). The issue of underspecification is not fully explored in the pilot study, presumably due to reasons of space.

Another difference relates to assumptions about the role and scope of FCCs. In the current thesis, we have adopted the minimalist view that there is no difference between language-specific gaps and universal restrictions on feature co-occurrence. Both are governed by FCCs. Levelt and van Oostendorp, however, note that there need be no FCCs referring to both [coronal] and [labial], as "... place features [...] exclude each other" (Levelt & van Oostendorp, 2007, p. 168).

In addition to the differences mentioned here, there are also some similarities. Both studies, for example, assume that features are privative. Both studies, too, adopt a definition of the Feature Co-occurrence Constraints that is essentially segment-driven. Consider the definitions given in Levelt and van Oostendorp (2007, ex. 5):

(15) *Assumptions about Feature Co-occurrence Constraints (FCCs)*
Where F, G denote features; FCCs in the constraint set Con are of the following type (only):

a. *[F,G]: No segment has both F and G

b. [F]⊃[G]: If a segment has F, it should also have G (no segment has [F] without having [G])

If we compare these definitions to the definitions in 1, we see that both assume the segment as the domain of the constraints. Levelt and van Oostendorp are less explicit about this than chapter 3, but in their case, too, we can read 'segment' as 'dominating the root node'. That is to say, the 'segment' in this thesis is nothing more than the simultaneous expression of phonological features, namely those that are dominated by the same root node.

The alternative would be to define the constraints with the feature as 'subject'. The definition for c-constraints would then read along these lines: "assign a violation mark to every feature [F] that is dominated by the same root node as feature [G]." The difference may seem one of little importance, but while the current, segment-driven definitions do not allow for asymmetrical c-constraints, the alternative would. One area where this becomes visible is feature spreading. Say a given form has two segments, A{[F,G]} and B{[F]}, and a constraint that enforces spreading of [G], such that [G] is dominated by both A and B, the constraint *[F, G] is violated by both A and B under the current definitions. A hypothetical *[G, F] behaves exactly the same and hence is redundant. Under the feature-driven view, however, even though there are two root nodes dominating both [F] and [G], the constraint *[F, G] is violated by both A and B, but its mirror image, *[G, F] is only violated once (remember there is only one instance of [G]). A more thorough exploration of the feature-driven view can be found in Linke and van Oostendorp (in preparation).

1.4 A brief history of thought on the segment inventory

The subject matter of the current thesis is the acquisition of the segment inventory, as we have seen. In the previous section, we discussed some earlier approaches to this problem, but before turning to the main matter of the thesis, it is important to reflect for a moment on the history of the study of the segment inventory *per se*.

The systemic, synchronic study of language begins, in modern times, with de Saussure. Distinguishing *langue* and *parole*, the *phonème* (speech sound) and sound images, de Saussure opened the door to an abstract approach to phonology. For our present purposes, which concern the phonological inventory, the most important contribution of de Saussure is his emphasis of language as a system: "*dans la langue, il n'y a que des differences [. . .] sans termes positifs*". In other words, the segment inventory is defined in terms of the contrasts it expresses, although Anderson (1985) emphasises that for de Saussure, the study of the speech sounds themselves is an important prerequisite to linguistic analysis, for it is only by knowing the positive definitions of speech sounds that we can study their relations. Thus contrast, or *oppositions*, became an important focus of phonology.

The inventory was of central concern for (American) structuralism. For American structuralists, in fact, the study of language was to construct inventories, not merely of phonemes, but also of morphemes, constructions, etc. (Anderson, 2000). The construction of these inventories was regulated by strict principles, of most concern to us at this point, the segregation of levels, and the bi-uniqueness criterion. According to the latter, for a speech sound to be considered a phoneme, it was required to stand in a special relation to its phonetic counterpart: For any phonetic event X, corresponding to phoneme A, every such event X corresponds to phoneme A. In other words, there is a unique relation between phonetic event and phoneme. Conversely, for any phoneme B, corresponding to phonetic event Y, every instance of B corresponds to event Y. Again, the relation is unique, hence the term bi-uniqueness. The bi-uniqueness criterion is a strict requirement, and various effects follow from it. It must be understood, however, that it does not entail that every phonetic detail or contrast is represented phonemically.

The importance of Jakobson's *Kindersprache, Aphasie und Allgemeine Lautgesetze* (Jakobson, 1941/1968) cannot be underestimated. Several important traits of modern linguistics come together in the work of Jakobson: first, it signals the beginning of feature theory (see also Jakobson, Fant, and Halle (1952); Jakobson and Halle (1956)). The introduction of feature theory not only allowed phonologists to study the internal structure of the phoneme; it also provided an apparatus to formalise natural classes, where a natural class is defined as the set of phonemes having a feature in common. Secondly, in addition to the Saussurian paradigm of linguistics as the synchronic study of language as a system, Jakobson proposed that the systems of individual languages were in fact not unrelated, but governed by universal principles, such as the set of laws of irreversible solidarity. These laws are implicational universal statements over the structure of the inventory, and can be re-interpreted as relative markedness generalisations ("if a language has X, it must also have Y"). Thirdly, Jakobson proposed to integrate the study of child language (and aphasic language) in the domain of general linguistics. Specifically, the laws of irreversible solidarity were intended to govern not only the sound systems of the world's languages (typology), but also (order of) acquisition, and loss of contrast in aphasia. Thus, Jakobson proposed an integrated theory of the segment inventory as a structured set of phonemes defined by distinctive features, and in addition, that language (or, at least, phonology) is driven by universal principles that make themselves known in typology, language acquisition, and language attrition.[10]

A true shift away from the primacy of the inventory was introduced with the advent of Generative Phonology (Chomsky, 1951; Halle, 1957, 1959; Chomsky & Halle, 1968). In (early) generative phonology, a view was propagated that the optimal way to explain linguistic structures was fundamentally derivational, or, in the words of Goldsmith and Laks (to appear),

> Of the four major tenets [...] of the SPE model, the most important was the view that the best explanation was algorithmic explanation. An algorithmic explanation is one which provides an account of the data which satisfies the conditions for being an algorithm: it is a fully explicit description of a process that can be carried out in a finite amount of time on a computational device such as a Turin machine or its equivalent.

A similar account is given by Anderson (2000), in his evaluation of the impact of Halle's The Phonetic Rules of Russian (Halle, 1957), later published as The Sound Structure of Russian (Halle, 1959). With the predominance of generativism in the ensuing years, phonology became derivationally oriented. In fact, SPE placed so much importance on rules and derivations, that positing new underlying segments was not an issue, as long as the system of rules would generate the attested surface form, and no more. The importance of the internal structure of the inventory was pushed to the background.

[10]With respect to aphasia, it should be noted that it is an abrupt condition, resulting from a stroke or lesion in the brain. It is not a progressive condition, and the mirror of child language is thus not proposed to be an individual's development (as it is for child language) but rather the states of individual patients ranked with respect to severity of the condition. In that respect, it could prove interesting to apply the Jakobsonian hypothesis to Altzheimer's disease, which is a progressive degenerative disease of the nervous systems, of which the progression of cognitive and functional symptoms is often described as mirroring the cognitive and functional development of early childhood, and which effects linguistic abilities as well as other cognitive and physical capacities (Blair, Marczinski, Davis-Faroque, & Kertesz, 2007).

This is not to say that the inventory no longer was considered; it was merely less prominent as an object of phonological study. In addition, it appears important to note that the introduction of feature theory, in combination with the developments in derivational analyses and the abandonment of the segregation of levels that was adhered to by structuralism, enabled phonologists to define segments and their contrastive and allophonic relations no longer with reference to phonetic correlates and systemic properties (contrasts), but also with respect to their phonological behaviour (morphophonology). To this day, phonologists do not agree on how to evaluate the different metrics to arrive at feature specification. Dresher (2009)'s Contrastive Hierarchy, for example, relies much on contrast, where contrast is defined in a specific way. For Morén (2003)'s Parallel Structures model, behaviour and economy are paramount, and the phonetic correlates are close to irrelevant. Conversely, much work in Element Theory (see for example Backley (2011)) holds that the behaviour and phonetic (acoustic) signature of a segment determine its subsegmental make-up, and contrast is almost epiphenomenal.

1.5 Overview of the thesis

Chapter 2 deals with the inventory, its description and its parts (features). We will begin by examining studies of the acquisition of the inventory and of features. A great deal of work has been done in this field, often focussing on the order of acquisition. This latter issue is of relatively minor concern to the current thesis; rather than testing a hypothesis about the order of acquisition, we provide a theory about the (former) mechanisms of acquisition. On the other hand, while the focus is often different, previous studies also had to provide some assumed or hypothesised formalism.

After discussing the inventory from an acquisitionist perspective, the chapter continues by exploring some aspects of the inventory *per se*. We will make a distinction between the *shape* of the inventory, and its *structure*, where the latter refers to the phonological organisation of the inventory, and the former to its phonetic implementation. Key aspects relating to the structure of the inventory are discussed: these include contrast, (under-)specification, symmetry and economy. Next, several theories about the shape and the structure of the inventory are reviewed: Dispersion Theory (Flemming, 2004), Parallel Bidirectional Phonetics and Phonology (Boersma & Hamann, 2008), the Modified Contrastive Hierarchy (Dresher, 2009; Hall, 2007) and expanding on that the theory of Dispersedness proposed in Hall (2011). These theories are elected because a) they explicitly concern themselves with the inventory and b) they are representative for the breadth of the phonological spectrum: from the fully functionalist to the formalist, from emphasising the shape of the inventory to focusing on its structure, and, via Hall (2011), back again. Furthermore, they each differ from the Feature Co-occurrence Constraint Theory in that the latter is anti-holistic: the inventory is fully epiphenomenal, it has no ontological status and no reference can be made to it by phonological rules or constraints. It is simply emergent from the system of freely combining distinctive features and filtering constraints. At the same time, we explore areas of overlap and conclude that some of the proposals reviewed are indeed compatible with Feature Co-occurrence Constraint Theory.

Next, we turn to the question of the developmental origins of features. The innateness debate is neither new nor settled, but after we acknowledge that the question is difficult to resolve empirically, we review a large body of literature. Rather than assuming that emergence is the default hypothesis and hence should be adopted in the absence of conclusive evidence indicating otherwise, we adopt a working definition of innateness, which states that a property is innate if the child exhibits knowledge of that property at or before the first encounter with it. This definition allows us to systematically examine the question of the innateness of features from three different perspectives, based on the different functional roles of distinctive features: categorisation/distinction, lexical organisation and phonological patterning. We find enough evidence to assume innate features.

Chapter 3 introduces the theory of Feature Co-occurrence Constraints, by outlining the Final State of the acquisition process of Dutch: the adult inventory. Only a small set of constraints is shown to be necessary, and no possible segments are either over- or underpreditcted. With these results, we proceed to an exploration of the formal properties of Feature Co-occurrence Constraints. It is demonstrated that the two constraint types are sufficient; and that it is neither necessary nor desirable for constraints to refer to more than two features. Single feature constraints can be modeled by assuming constraints such as *[F, F] and [F]→[F]. The latter is obviously always vacuously satisfied, but the former has a its own sets of effects, which we explore.

Next, we turn to the assumptions and implications regarding feature theory. Monovalency is defended, and so is universality. The constraint set introduced above is unable to rule out the empty segment, and hence the empty segment must be interpreted phonetically. There is a large body of literature on the non-specification of coronality and stopness; we assume indeed that both traits are not represented by distinctive features. The same goes for major class features. Rather than adopt Feature Geometry, we opt for Feature Classes. First, the latter is designed with Feature Co-occurrence Constraints in mind; secondly, and more importantly, it allows us to do away with major class features while retaining the insight that there are different *types* of features. Whereas Feature Co-occurrence Constraints do not reference these types but rather individual features directly (as per Feature Classes), they must be visible to the grammar or at the phonology-phonetics interface in order to avoid the situation where a segment containing, for example, [continuant] is also interpreted as a stop. In other words, feature classes are necessary to block the default interpretation when it is unnecessary. Also, it is shown how the Feature Co-occurrence Constraint Theory provides a way of implementing both inventory symmetry and

Feature Economy.

Acquisition of the inventory is the theme of **Chapter 4**, using data from the CLPF corpus of Dutch (Fikkert, 1994; Levelt, 1994). The main difference between the adult grammar and acquisition is that the latter entails a chronological component, whereas the former is assumed to be stable. Chronology is segmented into *stages*, which are defined by a change in the inventory. Constraints are invoked and can be revoked, but the time at which they are invoked is severely limited: no later than the time at which both features that the constraint refers to are acquired. This is borne out in the data. Next, the methods by which the data for this chapter are obtained are explained. As an example, we go through every stage of Noortje's actual productions. Her development is shown to be representative for the other children, and, other than in the case of Adult Dutch, some overpredictions occur. Considering all the children in the study, only six different segments (feature combinations) are ever overpredicted. This is not borne out for every child. Importantly, however, most instances of overpredictions concern cases where the segment is in fact in the child's inventory, but not represented in the sample because it is not robust enough to meet the inclusion criteria, which are subsequently discussed. Only four different segments are ever actually overpredicted. One of these is /t/, which, due to being represented as the empty segment, is predicted to be present in every inventory from the first stages. Three definitions are given for the contexts in which segments can be overpredicted, and the featural contents of overpredicted segments are discussed.

We conclude the chapter by considering the innateness of constraints, a difficult issue. For one thing, there is no way of knowing whether a constraint is present in the constraint set if its structural description has not been acquired yet. For example, the constraint *[F, G] has no effect on the grammar until both [F] and [G] are acquired. We conclude that while it is not unreasonable to assume that constraint templates are innate, there is no evidence in favour of innate feature co-occurrence constraints.

The main body of the thesis is concluded by **Chapter 5**. Some issues remain, which are dealt with there. We will explore to what degree the theory developed in this thesis is compatible with existing theories, such as Parallel Bidirectional Phonetics and Phonology (Boersma & Hamann, 2008), Inductive Grounding (Hayes, 1999) and the Modified Contrastivist Hierarchy/Successive Division Algorithm (Dresher, 2009; Hall, 2011). The theory of Feature Co-occurrence Constraints that is proposed owes much to these, and we see that especially in the first two cases, a high degree of compatibility exists. Finally, we will discuss the relation between emerging constraints and negative evidence, the different frameworks that the theory can be implemented in (mainly: constraint-ranking and non-constraint-ranking frameworks), and the role of perception.

Finally, **Appendices A, B** and **C** list for each child the succession of stages (A), the different FCCs that are activated during acquisition (B), and the inventories, features, constraints and overpredictions per child, per stage and per level of description (C).

CHAPTER 2

The Shape and the Structure of the Inventory[1]

Even as the current thesis is concerned with acquisition, this chapter is devoted entirely to the segment inventory, and, to be more precise, various ways of how it has been studied. In section 2.1, we will consider some key concepts in the study of the phonological inventory, after which, in section 2.2, we discuss three phonological theories dealing specifically with the segment inventory. In section 2.3, we will discuss whether the material of the inventory, distinctive features, are learned or innate.

Every phonological theory that adopts a non-atomic view of the segment (or phoneme) must somehow provide a means to constrain the combinations in which the subsegmental particles may appear. A theory that lacks such a restriction is limited in its predictions to fully symmetrical systems, a prediction that is not borne out, since many phonological inventories contain gaps. The fact that this may seem a trivial observation underlines its fundamental importance: inventories are shaped and structured, and they are not shaped and structured randomly. In this chapter, we will consider some proposals that have been put forward, from functionally driven to formal.

A number of important concepts stand out in the literature on segment inventories: contrast, dispersion, symmetry, and economy. The term 'contrast' is so central to phonological theory, that its meaning seems immediately clear. However, as we will see, much hinges on a precise definition of contrast, and how it is integrated in the theory of grammar. Dispersion refers to the centrifugal tendencies of the members of an inventory. Regardless of the number of segments in the inventory, its members will generally occupy the entire phonetic

[1]Parts of this chapter are based on van 't Veer (2008).

space and populate it evenly. In addition, inventories often display a degree of symmetry; that is to say, given a number of features in a language, a fully symmetrical system contains the full permutation of feature combinations. A well-known example of a symmetrical system is the Turkish vowel inventory: it can be described using the features [±high], [± front] and [±round], and all eight possible combinations ($2^3 = 8$) correspond to actual vowels in the language. Feature Economy (Clements, 2003) formalises the notion that languages tend to make full use of distinctive features – an economical system creates as many segments as possible out of as little a number of features as possible. In section 2.1, each of these terms will be explored in more detail.

Besides the question of the shape and the structure of the inventory, there is the matter of its ontological status, to which we turn in section 2.2. Logically, one could see the segment inventory either as a phonological primitive, or as epiphenomenal. Much related to this is the ontological status of the segment or phoneme; is it a primitive itself? Is /p/ a phonological object that rules (or constraints) can refer to? Or can we only refer to the features that make up /p/? In the current thesis, a segment is no more than the (phonologically) simultaneous actualisation of a number of features; it has no further status as phonological object. Not all features can or may be simultaneously realised, and restrictions on feature combinations are regulated by Feature Co-occurrence Constraints. Thus, the inventory is purely epiphenomenal in the sense that it exists only because some feature combinations are allowed and some are not. There are no rules or constraints referring to the inventory, only to features and prosodic categories. This is preferable in at least two ways: first, it is more parsimonious than a theory that assigns a phonological status to the inventory. Secondly, if the (phonological) grammar is a machine that maps sound onto meaning and *vice versa*, it is unclear where in the chain the inventory has its place. This is perhaps the main argument against so called 'holistic' theories of segment inventories. There are, however, theories in which the inventory, either implicitly or explicitly, is seen as a primitive. Dispersion theory (Flemming, 2004, 2006) is a clear example, which we will discuss below. However, we will also see how Parallel Bidirectional Phonetics and Phonology (Boersma & Hamann, 2008, for example) and a theory of contrast and enhancement (Hall, 2011) run into similar problems.

2.1 Some key concepts in the study of the inventory

2.1.1 Symmetry and feature economy

Two concepts are important in the study of the structure of the inventory: symmetry and Feature Economy (Clements, 2003), which predicts that phonologies should make use of the features they employ as much as possible (employment means exploitation). The two are intimately linked, but in prose, we might express the relationship as follows: In a fully economical system, all feature combinations are possible, whichever features are active in the language. In a fully symmetrical system, for every feature that is active, its counterpart is also active (this works only with binary features, but see below for an elaboration). In other words, every intersection between the phonetic dimensions described by features is populated by a segment (feature co-occurrence). So, whereas symmetry has something to say about which features are active, economy makes predictions about how they combine. It should be noted that a fully symmetrical system is also maximally economic (every feature combination is used), whereas a maximally economical system is not necessarily symmetrical – there might be gaps. At first glance, it would appear that Feature Co-occurrence Constraints are inherently anti-economical. This is because FCCs punish feature combinations, whereas Feature Economy would be improved if more combinations were made while keeping the number of features constant. Below, however, we will see that the situation is more nuanced.

Even though Clements (2003) rightly argues that Feature Economy and Symmetry are not the same thing, the two are very similar. For example, a maximally symmetric inventory is also maximally economic. Both Feature Economy and Symmetry are decreased when gaps appear in the system, but they are sensitive to different types of gaps. Consider Clements' examples 4a-c (Clements, 2003, p. 292), reproduced below:

(16) Three sound systems differing in symmetry and economy

a. System A

p	t	c	k
b	d	ɟ	g
f	s	ʃ	x
v	z	ʒ	ɣ

b. System B

p	t	c	k
b	d	ɟ	g
f	s	ʃ	x

c. System C

p	t	c	k
b	d	ɟ	g
f	s		x
v		ʒ	

System A is fully economical, as well as fully symmetric. System B, on the other hand, is less economical, but no less symmetrical: even though it contains gaps, the gaps occur symmetrically in their own right. System C, finally, is more economical then System B by virtue of having more segments without employing more features; at the same time, it is less symmetric because the gaps occur in less regular places.

In terms of FCCs, System A can be described with three constraints: *[lab]-[dist], *[lab][dors], and *[dors][dist]. System B is more 'economical' than C, because it requires only a single constraint in addition to the three PoA-constraints: *[cont][voice]. System C is predicted not to exist under the FCC approach as adopted here. It requires the exact same constraints as System A, while overpredicting precisely those segments that Clements (2003) removed for expositional purposes (see section 4.5 for an analysis of overpreditions). Clements makes no claims with respect to the degree of likelihood for the three systems, either under Feature Economy, or in general; they remain hypothetical.

By way of further illustration, the Turkish vowel inventory can be used as an example of both Symmetry and Economy. It is both fully symmetrical (there are no gaps) and fully economical – there is no restriction on the combinations of features (at least not in roots). In terms of Feature Co-occurrence Constraints, this means that all constraints banning vocalic features to be simultaneously actuated are inactive (or low-ranked, in OT terms). In this example, we clearly see a property of the c-constraint class of Feature Co-occurrence Constraints: they are anti-economical, by enforcing to combine as little features per segment as possible (every combination will yield a violation). The i-constraints, on the other hand, promote the use of as many features per segment as possible. For every constraint F→G, a segment with F but not G will result in a violation.

The examples do serve to show that FCCs implement a degree of symmetry: the combination of two features ([continuant] and [voice] in the example above) is banned regardless of any other feature in the segment. Put differently: the Feature Co-occurrence theory is a theory of inventorial gaps, and it predicts that these gaps are symmetrical to a degree.[2]

At the same time, the examples show that rather than being anti-economical, FCCs in fact implement economy, under the assumption that a system with less constraints is by some metric better than a system with more constraints. This is easily demonstrated: a fully economical system employs every possible feature co-occurrence. In other words, no FCCs are required.[3] Every gap that appears in such a system requires at least one FCC to be introduced in the grammar. An example of this is System B above. Hence, we could say that a 'meta-goal' of acquisition is to have as little active Feature Co-occurrence Constraints as possible. As we shall see in chapter 4, an important part of the theory is that constraints can be revoked/demoted during acquisition.

Another way to phrase the above is to say that Feature Economy is a driving

[2] A large-scale typological study would have to demonstrate to what degree this prediction is correct; this is clearly beyond the scope of the current thesis, but see van 't Veer (2008) for a first attempt.

[3] Note that Clements is incorrect in stating that System A is "fully economical" (p. 292), as it does not allow for segments with complex place of articulation. The economy of System A (under Clements' feature assignment the inventory is described by 5 features) is $E = 16/5 = 3.2$. System A' which includes every logically possible feature combination has a much higher economy: $E = 5!/5 = 120/5 = 24$.

Stage	*Inventory*															
1	p	b	t	k	m	n	l									
2	p	b	t	k	m	n	l	d	j							
3	p	b	t	k	m	n	l	d	j	z						
4	p	b	t	k	m	n	l	d	j	z	s					
5	p	b	t	k	m	n	l	d	j	z	s	r				
6	p	b	t	k	m	n	l	d	j	z	s	r	ʋ			
7	p	b	t	k	m	n	l	d	j	z	s	r	ʋ	v		
8	p	b	t	k	m	n	l	d	j	z	s	r	ʋ	v	ŋ	
9	p	b	t	k	m	n	l	d	j	z	s	r	ʋ	v	ŋ	f
10	p	b	t	k	m	n	l	d	j	z	s	r	ʋ	v	ŋ	f	x

Table 2.1: Development of word-initial onsets in Catootjes actual productions

force during acquisition. We see this exemplified in the data from the children in the current study, and although a detailed examination of this hypothesis is beyond our current goals, a brief exploration is promising. Take, for example, the developing segment and feature inventories of Catootje (tables 2.1 and 2.2).

Catootje's inventory at the first stages has a fairly high degree of symmetry, especially if we consider the first two stages together: the major places of articulation are occupied by stops and nasals, and also by voiced stops. There are also some anti-symmetrical aspects in Catootje's inventory development, for example the absence of the dorsal nasal and voiced stop, both of which are not in Adult Dutch onsets (the occurrence of /ŋ/ later is idiosyncratic). Furthermore, the development of continuants is not as symmetrical as the development of non-continuants.

More importantly, comparing tables 2.1 and 2.2, we see that whereas the segmental inventory expands in a gradual fashion, every necessary feature has been activated by the second stage. This can mean only one thing: from the second stage onwards, Catootje's inventory is becoming more and more economical. Catootje's inventory is a striking example, but the same process can be seen in other children's developing phonologies.[4]

In chapter 4 we will see that the process through which this economisation takes place is the demotion/revocation of Feature Co-occurrence Constraints (we will also go deeper into the methodology of the data analyses, and the definitions of notions such as 'stage' in that chapter). For now, we will leave it at the observation that Feature Economy appears to be a characteristic of child phonology, and that the Feature Co-occurrence Constraint theory is capable of describing the process of increasing economy.

[4]If Feature Economy is indeed a driving force, the question arises as to why the developing inventory – or the adult inventory for that matter – has gaps in the first place. In the case of final state (adult) inventories, the answer is diachrony. Children do not process the inventory on an inventory-level; rather, they are processing words and extracting sounds and features from these forms. We will come back to the issue of input processing in chapter 4.

Stage	*Features*
1	[labial], [voice], [dorsal], [nasal], [liquid]
2	[labial], [voice], [dorsal], [nasal], [liquid], [continuant], [approximant]
3	[labial], [voice], [dorsal], [nasal], [liquid], [continuant], [approximant]
4	[labial], [voice], [dorsal], [nasal], [liquid], [continuant], [approximant]
5	[labial], [voice], [dorsal], [nasal], [liquid], [continuant], [approximant]
6	[labial], [voice], [dorsal], [nasal], [liquid], [continuant], [approximant]
7	[labial], [voice], [dorsal], [nasal], [liquid], [continuant], [approximant]
8	[labial], [voice], [dorsal], [nasal], [liquid], [continuant], [approximant]
9	[labial], [voice], [dorsal], [nasal], [liquid], [continuant], [approximant]
10	[labial], [voice], [dorsal], [nasal], [liquid], [continuant], [approximant]

Table 2.2: Development of features in word-initial onsets in Catootjes actual productions

2.1.2 Contrast, minimal pairs and feature specification

Contrast is central to phonology – perhaps more so than any other concept. In the words of de Saussure, *"dans la langue, il n'y a que des differences."* The *raison d'être* of phonology is, broadly speaking, to map sound structures to lexical items so that further computation may ensue, and vice versa. Phonology acts as an *addressing system* for the lexicon. That is, it decodes the phonetic signal and outputs a symbolic code (underlying form), according to a non-random derivational algorithm. This code is then used to locate the relevant entry in the lexicon.[5] As in any other addressing system, each address must be unique. In other words, each address must be contrastive with respect to all the others. It is easy to see that a system with too much ambiguity will not function effectively.

The textbook conception of contrast, and the one underlying much work tacitly, is that the relevant contrasts in a language can be derived from sets of minimal pairs. As an example, let us look at the relevance of contrast in the theory of vowel harmony proposed in Nevins (2010).

In his approach to vowel harmony, Nevins (2010) argues that harmony is not driven or initiated by the source of the harmonising feature, but rather by the target. Target vowels are 'needy', which means that in order to be phonetically interpreted they require one or more features that are not present in the underlying representation. In order to acquire the needed feature (say, [F]), a Search algorithm is initiated, which evaluates whether neighbouring segments contain [F]. If so, it is copied onto the needy vowel. There are certain parameters governing the search algorithm, such as directionality, limits on the domain of application, and conditions on the source (in parasitic harmony, for example).

The 'needy vowel' model of vowel harmony offers an interesting approach

[5] I was first introduced to this metaphor by Jonathan Kaye and Markus Pöchtrager.

to the subject. It is both descriptively and explanatorily adequate, and at the conceptual level, comes with the boon of cross-modular uniformity: vowel harmony is parallel to Agree in syntax, and the search algorithm works the same in both modules. In other words, syntax and phonology are, to a degree, procedurally equivalent. Nevins shows that the model is able to account for a wide variety of vowel harmony cases – as long as we know which vowels are needy, and what they are needy for. In Nevins (2010) proposal, segments are defined as bundles of binary features, and all vowels contain the same features (and thus differ only in the values of these features). Any vowel lacking a feature (or a value for a feature) is thus needy.

As mentioned above, the Search algorithm is parametrised for a number of dimensions. One of those dimensions, contrastiveness, is an important parameter in determining what counts as a potential source, for example in transparency in vowel harmony. Taking Finnish as an example, Nevins adapts the Search algorithm in such a way that it only considers *contrastive* occurrences of [±back] as potential sources. In other words, locality is relativised to contrastiveness.

This is not a trivial step, because it implies that the Search algorithm knows or is able to know when and where a feature is contrastive. The point is that contrastiveness must be marked somehow. Contrastiveness is not a property of features, but of segments: two segments contrast if they differ in the specification of at least one feature. Going back to the example of Finnish, the Search algorithm must know for each instance of [-back] it encounters whether it is contrastive, or not. The same feature (value) can be both contrastive and non-contrastive in the same language; contrastiveness becomes a property of the feature.

A possible solution to this is that the Search algorithm evaluates the entire inventory in each search pass, or that features or feature bundles are somehow marked for contrastiveness. The latter option would add a new element (sub-features? non-phonological features?) to the phonological alphabet (as per minimalist tradition), the former dramatically increases the computational load of the Search algorithm.

Underspecification

In order to get around the problem of marking contrastiveness, feature theories have often made use of underspecification. By specifying only those features (or values) that are contrastive, no reference need to be made to contrastiveness as independent part of the phonological grammar: contrast has become epiphenomenal, emerging from specification (but see below on Dresher, 2009's critique of practices of underspecification). Contrastive underspecification is not an option for Nevins, however, as the formal mechanism of underspecification is already used to denote neediness. To underspecify neither needy features, nor redundant features would lead to hyper-harmony: all segments would be needy for whatever feature values they are not contrastive for, initiating the Search algorithm. In conclusion, in the Needy Vowel approach to vowel harmony, diacritics are unavoidable to mark either contrastiveness, or neediness. It appears that any theory of full specification needs to either not refer to contrastiveness as a functional element of grammar, or mark it diacritically.

Some elements in phonology display a behaviour that is different from what one would expect on the basis of their surface form. For example, in a process like vowel harmony (see above), some vowels are 'transparent'. Although at the surface they appear to form a natural class with the triggering vowels (or their complement), they have no effect on the harmony process whatsoever. Consider Finnish, which has a [±back] harmonising requirement:

(17) Transparent vowels in Finnish [±back] harmony (taken from Hall (2011, example 6))

a.	[grøtsi+næ]	'porridge + ESSIVE'
	[tsaari+na]	'tsar + ESSIVE'
b.	[syyte+ttnæ]	'action + ABESSIVE'
	[suure+na]	'entry + ABESSIVE'
c.	[væitel+lyt	'dispute + PAST PART'
	[ajatel+lut	'think + PAST PART'
d.	[værttinæ+llæ+ni+hæn]	'with spinning wheel, as you know'
	[palttina+lla+ni+han]	'with linen cloth, as you know'

Situations like these lead to the idea that not all surface features are 'visible' to phonology; in fact, that some segments are (lexically) underspecified. The question of what may be left unspecified and what not has been answered in a number of ways. In Radical Underspecification (Kiparsky, 1982; Archangeli, 1984), for all features [±F], only one value, either + or –, can be stored. Every language has one 'default' vowel (for instance, the vowel that shows up in epenthesis), and it is assumed that this vowel is underspecified for all its features, and that these features are filled in in the course of derivation. Because the fill-in rules are there by necessity, it becomes possible to erase from the lexicon *all* feature specifications that the default vowel will receive eventually. That is, if the default vowel is [+high], [+high] is underspecified in the lexicon. Other patterns of phonological activity can also be used to motivate which value of a feature can be left unspecified, but the general theme is that somehow *predictable* feature values are unspecified in the lexicon and filled-in derivationally.

A serious problem for Radical Underspecification is not underspecification, but the point at which a given feature becomes specified. This point, at which value α is assigned to F, is defined as '. . . automatically ordered prior to the first rule referring to [αF]' (Archangeli, 1984, cited in Dresher, 2009), thus defying any falsification. In Contrastive Underspecification (Steriade, 1987) the criterion for specification is not only unpredictability but rather contrastiveness: only feature values that are contrastive in the language are specified. The problem is that Contrastive Underspecification does not supply a principled way of determining contrasts, and the algorithm that became identified with it (Dresher, 2009, p. 203), the Pairwise Algorithm (or minimal pairs, see above) is inadequate, as we shall see in section 2.1.2. A third possibility to achieve underspecification is simply to identify one feature and label it 'underspecified' universally. This is the route taken by proponents of the Featurally Underspecified Lexicon (FUL, see Lahiri & Reetz, 2002, for example; see also section 2.3.2), who claim that [Coronal] is always underspecified, a proposal which we shall adopt in chapters 3 and 4.

In section 1.3.2 above, we have discussed a number of theories that approach the acquisition of the inventory as a logical problem. In other words, they posit an abstract, hypothetical learner to which certain capacities are ascribed, and demonstrate that the model they propose is learnable, or even emerges from a learning mechanism. One of the most prominent of such theories, as we have seen, is known as the 'Modified Contrastive Hierarchy' (Dresher, 2009; Hall, 2007, 2011, among others). As the name indicates, the Modified Contrastive Hierarchy (henceforth, MCH) puts a high burden on the contrastive status of features. In fact, the 'Contrastivist Hypothesis' holds that *only* contrastive features can be phonologically active in a language (Hall, 2007). Hence, MCH represents a special instance of underspecification theory. What is more, the MCH represents an alternative to the minimal pair hypothesis, and Dresher (2009) effectively shows how the minimal pair hypothesis faces a learnability problem if it is used to derive underspecification.

Minimal pairs in acquisition

Interestingly, criticism of the minimal pair hypothesis has come from both the perspective of the 'logical problem of language acquisition' (Dresher, 2009), and the 'developmental problem of language acquisition' (see Ingram (1989) and section 1.3 for an exploration of these terms).

Much of the literature in first language acquisition reports research on a remarkable transition that children go through in their first year. Infants start out being able to discriminate between all possible speech sounds, and gradually lose this ability while improving the recognition of sounds relevant to their own language. Many models have been proposed and/or adapted to interpret this transition (e.g., Kuhl (1991); Best (1995)), but generally, this development is seen to be due to a 'reorganisation in perceptual biases' (Werker & Pegg, 1992). The process of acquiring native language speech sound categories proceeds from universal yet categorical discrimination at a very young age (Eimas, Siqueland, Jusczyk, & Vigorito, 1971), to an adult-like perception of native vowels (Kuhl, 1991) at around six months, and finally to a manner of discriminating consonants similar to that of adults before the end of the first year (Werker & Tees, 1984).

In the developmental perspective, the Minimal Pair hypothesis assumes a top-down developmental process, in which the child first stores whole word forms, and only later, as the lexicon is large enough, begins to analyse these forms in smaller units of representation (see, for instance, MacKain (1982); Jusczyk (1985); Werker and Pegg (1992); Best (1995)). Crucially, the Minimal Pair hypothesis relies on the assumption that the acquisition of a phonology is based on linguistic factors (more specifically, some part of the chain between sound and meaning), not simply on bare statistical generalisations (e.g., Werker & Pegg, 1992:p. 299).

Maye, Werker, and Gerken (2002) argue that the Minimal Pair hypothesis cannot provide a model of phonetic categorisation, as it depends on the occurrence of minimal pairs in the lexicon. We will come back to Maye et al. (2002) in more detail below. However, from earlier studies (see, e.g., Eimas et al., 1971; Werker & Tees, 1984; Kuhl, 1991)[6], we know that native phonemic categories are formed robustly before a reliably large lexicon has been acquired that is, native phonemic categories are largely in place before the end of the first year of life. What is more problematic, hardly any minimal pairs occur in the child's receptive lexicon when it contains of 50 words (Maye et al., 2002). For these reasons, Maye et al. (2002) state that the child's lexicon cannot serve as the basis for phonemic development as it is not large enough to contain the required number of minimal pairs. We may have to relativise their claim somewhat, as it seems that their conception of the lexicon is the lexicon in its final, adult-like state. The acquisition of the adult lexicon has two major precursors: retention of word forms, and formation of concepts (the third ingredient of an adult-like lexicon being syntactic information). These elements are then integrated in the adult lexicon. Thus, the child must have storage of phonological forms before the onset of a meaningful lexicon. That this is so, at least at eight months of age, has been shown in Jusczyk and Hohne (1997). Again, this is not taken to mean that indeed, children use minimal pairs in their (proto-)lexicon to form phonological categories, but it does serve to indicate that the characterisation of the minimal pair hypothesis sketched in Maye et al. (2002) may be overly restrictive (see also Swingley, 2008). Even so, the Minimal Pair hypothesis is a poor candidate as a model of acquisition – it seems to suffer from a (fatally) severe case of poverty of the stimulus.[7]

---

[6]Although Werker and Tees (1984) note that "It is probably no accident that this decline, or tuning, occurs at about the age that the child is beginning to understand and possibly produce sounds appropriate to his/her native language."

[7]In contrast to the Minimal Pair hypothesis, many studies propose that children have direct access to a great deal of fine-grained statistical information present their linguistic environment, and that they are able to exploit this information in order to construct phonological categories. According to these studies, acquiring the features of the native languages proceeds through distributional learning (see, for example, Maye et al. (2002)). However, one problem for a Distributional Learning hypothesis is, as Peperkamp (2003) notes, that distributional data is notoriously unreliable in itself: one phoneme may have multiple allophones that overlap in phonetic space, and, thus, one surface phoneme may belong to a number of underlying segments. Furthermore, it is not *a priori* clear what distributions the child should be sensitive to: type or token distributions, word-level, phoneme-level or feature level? This is a serious problem for distributional learning studies, because a) post-hoc correlations do not carry much meaning, and b), while many distributional learning studies assume a causal

2.1.3 Holisticity, contrast and specification in Feature Co-occurrence Constraint Theory

As we have seen, a theory of the inventory must either supply an external metric of contrast (Nevins, 2010), or somehow make contrast epiphenomenal to the features that are specified in the lexicon (Archangeli, 1984; Steriade, 1987; Dresher, 2009). In the latter case, it is crucial to supply a principled way of deciding which features are contrastive in what contexts; this is the goal of the Modified Contrastive Hierarchy (Hall, 2007; Dresher, 2009).

The approach advocated in Feature Co-occurrence Constraint Theory is in principle antagonistic to the first approach. In the FCC theory, there is no inventory – at least not as a primitive or otherwise as an entity. The inventory is fully epiphenomenal. The only elements of computation are features and prosodic units; the only computational devices are the constraints themselves.

This begs the question whether the FCC approach can be combined with underspecification. This would – at first sight – be a difficult enterprise. Any underspecified feature is, as it happens, underspecified, and hence, there can be no constraint against it. However, we shall see in chapter 4 that a certain degree of underspecification (that is, both coronality and stopness are not represented by features in the current approach) makes largely correct predictions.

The Modified Contrastive Hierarchy (Dresher, 2009) is not free of a hint of 'holisticity' (defined as the idea of the inventory being an independent entity). Take, for example, the two possible specifications for a three-vowel system /i a u/:

(18) Possible specifications for /i a u/

a. [high] >> [round]

	i	a	u
[high]	+	-	+
[round]	-		+

b. [round] >> [high]

	i	a	u
[round]	-	-	+
[high]	+	-	

The issue is that underspecification is defined per segment, rather then per feature. In online computation, whether a feature is distinctive can only be decided if the other features are known; for example, in (18b), [+round] is all that is necessary to specify /u/. The reason we know this is that there is no other segment specified [+round]. However, the only way to know this is to compare /u/ to all the other members of the inventory. Hence, the inventory must have some independent status.

relation between input distribution and learning output, this is an unwarranted step (see also Fikkert and Levelt (2008) for a rare observation as to the possible reverse directionality of any causal relation).

In the Feature Co-occurrence Constraint Theory, the inventory has no independent status. It is simply epiphenomenal to a 'generator-and-filter' model of phonology, where the generator supplies permutations of acquired features, and the filter (consisting of the FCCs) bans those that are ungrammatical. In the final chapter of this thesis, we will tentatively explore to what degree it is possible to combine the minimality of the Feature Co-occurrence Constraint Theory with the insights from the Modified Contrastive Hierarchy.

2.2 Phonetics and phonology of the inventory

When describing the inventory of a language, it is useful to distinguish between the *shape* and the *structure*. Whereas the two are intimately linked, they are not the same (see also Hall, 2011). The *structure* of the inventory refers to the featural representation of its members; the *shape* refers to the way the members are distributed in phonetic space. With respect to the latter, a common observation, often made concerning vowels, is that the members of the inventory are equally spaced throughout the phonetic expanse (conversely, the structural property of symmetry is often observed for consonants). Several theories have been constructed to account for this observation; from phonetically oriented to phonologically based; from teleological to epiphenomenal. In this section, we will discuss three of the most important proposals about the *shape* of the inventory: Dispersion Theory (Flemming, 2004, henceforth: DT), Parallel Bidirectional Phonetics and Phonology (Boersma & Hamann, 2008, henceforth: PBPP), and Hall (2011)'s proposal on dispersedness without dispersion.

These theories were selected for a number of reasons. First of all, they are well-developed, and, at least in the case of Dispersion Theory and Hall (2011)'s elaboration of the Modified Contrastive Hierarchy, have a long history in phonology. More importantly, these three frameworks *explicitly* deal with the phonological inventory, which is not something that can be said of many other theories. Furthermore, each is intimately linked to ideas and assumptions about the *structure* of the phonological inventory. Also, they are representative of different approaches to phonology: from the highly functionalist (Dispersion Theory) to the formalist (Hall, 2011), and the computationalist (PBPP). Most importantly, however, each of the theories has deep connections to learning algorithms.

There are, of course, other frameworks that could have been selected. In the first chapter, we briefly discussed Inductive Grounding (Hayes, 1999), to which we will come back in section 4.6 and in the final chapter. We have already seen variants of underspecification above (Kiparsky, 1982; Archangeli, 1984; Steriade, 1987), but these are no longer under active development. Work on substance free phonology and the parallel structures model of phonological representation (Morén, 2003; Morén, 2007) deals with the inventory, but seems more aimed towards developing a theory of features, rather than focusing on the inventory *per se*. With its heavy emphasis on representations, it is no surprise that some work in element theory/government theory focuses on the segment inventory (e.g., Harris (1994), Charette and Göksel (1998), Backley (2011). These authors are not, however, generally concerned with acquisition or learnability.

2.2.1 Dispersion Theory

Dispersion theory (DT) is a version of functionalism: it takes as its starting point the premise that languages (in this case, inventories) are shaped by the conflicting interests of speakers and hearers: whereas the former aim to minimise the effort of producing speech sounds, the latter demand a maximal degree of perceptibility. That the two conflict, can be demonstrated with respect to the phonetic vowel space: the speaker prefers to keep the energy investment per vowel down, and thus deviate only minimal from the resting position, roughly schwa. However, such an attitude would render the vowels in the system quite similar, making it harder for the listener to distinguish between the vowels, and thus between words. As every speaker is also a listener, s/he is familiar with these conflicting forces, and allows them to actively influence her grammar – such is the view of functionalism. In this section, we will briefly discuss the dynamic model of dispersion introduced by Liljencrants and Lindblom (1972), before turning to a more elaborate evaluation of modern incarnations of DT, mainly the one proposed in Flemming (2004) and related work. We will follow the criticisms expressed before in Boersma and Hamann (2008); van 't Veer (2008); Dresher (2009); Hall (2011). In particular, we will conclude that Dispersion Theory encodes more in the grammar then is needed (teleology), and that it makes unclear and strange predictions about the nature of human linguistic competence.

Early versions of Dispersion Theory

Liljencrants and Lindblom (1972) propose a dynamic computational model of the vowel inventory, in which vowels are endowed with mutually repelling forces, much as electrical particles with equal charge are. Starting from a circle around the centre of a two-dimensional phonetic space, the vowels move, generally, to more extreme positions in response to the repellent influence of the other vowels. At each step, the 'energy' of the entire system is measured (the energy decreases as the distance between the vowels becomes larger, as the strength of the repelling force decreases with distance – analogous to the familiar Inverse Square Law in physics), and the model reaches its final state at the point at which no further reduction in total energy can be obtained. Although the model is relatively successful in some respects, there are some empirical and conceptual problems. Liljencrants and Lindblom (1972) show results obtained for vowel inventories of varying sizes, from thee to twelve members. These results are compared to descriptions of human languages, and the authors conclude that their model performs rather well: "The model produces about nine clear errors in a comparison involving 75 vowel qualities." As Hall (2011) points out, however, the model as it is presented is incapable of generating schwa in inventories smaller then ten vowels, nor does it generate any non-high front rounded vowels. Also, it predicts an unattested five-way back-front contrast among the high vowels in inventories of more than nine vowels. Furthermore, although the authors state that the initial state of the model is a "random" organisation, the vowels are initially highly ordered (at equal distances on a circle with a radius of 100 mel). As demonstrated by Hall (1999, 2007), however, the outcome of the model is highly dependent on the arrangement at the initial state, while there is no principled reason to assume the shape of the initial state used by Liljencrants and Lindblom (1972).

Dispersion Theory in OT

A contemporary incarnation of the spirit of Liljencrants and Lindblom (1972) can be found in Dispersion Theory. Flemming (2004) takes as his point of departure that phonology functions to 'minimally distinguish words'.[8] From this it follows, he argues, that phonology should favor larger contrasts over smaller ones. Adopting Optimality Theory as model for the phonological grammar, Flemming (2004) argues that in order to fulfil this task, the grammar provides constraints that do not concern individual segments, but that rather evaluate contrasts between pairs of phonemes. According to Flemming (2004), phonology shapes its inventories under the influence of these three 'functional goals' (p.7):

1. Maximise the distinctiveness of contrasts
2. Minimise articulatory effort
3. Maximise the number of contrasts

The idea that these principles play a role in phonology and/or phonetics is, of course, not new (Liljencrants & Lindblom, 1972, among others). What is specific to Dispersion Theory, however, is that they are directly encoded in the grammatical formalism. In Flemming (2004), for example, they are recorded in the following Optimality Theoretical constraints (or constraint families):

Mindist = *D:n*
: assign a violation mark for any contrast that has a distance less then n on dimension D

*Effort
: minimise articulatory effort

[8] Although similar to our conception of phonology as an addressing system, there is a crucial difference between the two ideas. An addressing system must involve the lexicon and relate (or translate) phonetic input to addresses in that lexicon; to 'minimally distinguish words' is an important step, but much more than that. We shall see below that this different conception of phonology has deep implications.

MAXIMISE CONTRAST
Assign a satisfaction mark for every member in the candidate inventory

The first constraint, MINDIST, evaluates for each pair in the inventory the distance on a scale. For vowels, Flemming (2004) proposes a somewhat abstract distance scale on three dimensions, corresponding to the first three formants. The second constraint against articulatory effort is not discussed in much detail; it remains entirely unclear how 'effort' is defined and/or quantified. Flemming (2004) goes no further then to assert that to minimise effort '... appears to be a general principle of human motor behaviour not specific to language.' The third constraint is a positive constraint, that is satisfied by the candidate containing the most members.

This brings us to a rather peculiar characteristic of Flemming (2004)'s Dispersion Theory, briefly hinted to above: although it employs the formal apparatus provided by Optimality Theory, it does not perform the task usually ascribed to grammar, namely to map input forms to output forms. Dispersion theory is not a theory of derivation, which leaves its ontological status, or place in the grammar rather unclear. This is not necessarily a problem, as long as it is acknowledged. It does, however, beg the question as to what exactly is modeled.

The difficulty that DT encounters with derivations is extended to its incompatibility with faithfulness constraints. As Flemming (2004) notes, faithfulness 'subverts the intended effect of the MINDIST and MAXIMISE CONTRAST constraints', by enforcing a relation between input and output. Take, for example, the tableaux in (19) (examples 9 and 23 in Flemming (2004)), representing a part of the vowel inventory of Italian. Dispersion constraints correctly predict the optimal inventory to be [i-e-a], as we can see in 19a: the two-member inventory [i-a] is excluded because other candidates fare better on the MAXIMISE CONTRAST-constraint, whereas the four-member candidate [i-e-ɛ-a] fails on a MINDIST-constraint. By placing four segments on the F1 continuum, the distance between the individual segments is smaller than enforced by the high-ranked MINDIST=F1:3.

If a faithfulness constraint enters the evaluation, however, the system derails easily (19b). Remember that Richness of the Base holds that the correct output must be selected regardless of the input, and that no constraints hold at the level of the input. Hence, /ɪ/ is a valid input. In 19b, candidates a. and d. are discarded based on their unfaithfulness to the input on the F1 dimension. Candidate b. retains faithfulness and does better on MAXIMISE CONTRAST than candidate c. Candidate b. is minimally different from the actual subinventory, but it crucially fails on MINDIST=F1:3.

(19) a. Dispersion constraints without faithfulness

Input: //ɪ//			Mindist = F1:2	Mindist = F1:3	Maximise contrasts	Mindist = F1:4	Mindist = F1:5
a.		i-a			✓✓!		
b.	☞	i-e-a			✓✓✓	*!*	**
c.		i-e-ɛ-a		*!***	✓✓✓✓	***	*****

b. Dispersion constraints with faithfulness

Input: //ɪ//			Ident [F1]	Mindist= F1:3	Maximise contrasts	Mindist= F1:4
a.		i̲-e-a	*!		✓✓✓	**
b.		ɪ̲-e-a		*!	✓✓✓	**
c.	☞	ɪ̲-a			✓✓	
d.		i̲-a	*!		✓✓	

Flemming (2004) solves this problem by relegating the burden of faithfulness effects to constraints on output resemblance, such as output-output correspondence constraints (Benua, 1997) or constraints enforcing paradigm uniformity. It is not shown how this is done exactly, however, and the problem bears directly on the issue of the Italian vowel inventory outlined in example (19). Whereas the tableaux presented by Flemming derive the intended inventories for stressed and unstressed positions, the theory does not predict which vowels will neutralise to what position. The inability to deal with faithfulness restrictions leads to another complication: whereas Flemming stated the goal of phonology to be to 'minimally distinguish words,' this apparently must be read as 'minimally distinguish surface forms.'

One of the more conceptual objections one might have raise against Dispersion Theory is that it is overly teleological. Although Flemming (2004) is correct in arguing that phonology distinguishes between words, it is a non-sequitur to then assert that this entails that the phonological grammar must encode a preference for more extreme contrasts over lesser ones. As shown by, among others, Boersma and Hamann (2008) and Hall (2011), dispersive effects can be modeled very well as epiphenomenal to a system designed to map lexical forms to surface forms, but which is not aimed at these effects per se. A related issue is that of degree, or gradience. While it is true that the phonological grammar must conserve certain contrasts (but also neutralise them elsewhere), this by no means automatically implies that this same grammar is concerned with the *degree* of differentness. Again, although dispersion in vowel inventories is a real observation, this maw well be an epiphenomenal effect of enhancement (Hall, 2011), or of diachrony.

Summary

Whereas the model of conflicting interests of speakers and hearers is often successfully used to model diachronic change (Blevins, 2007, for example), it is by no means a given that it is an active ingredient of synchronic grammars. For one thing, both articulatory effort and perceptual ease are difficult to quantify. Related to this is the question whether it is desirable for a theory of grammar to encode physical properties (other than perhaps instructions to the speech organs at the phonetics–phonology interface). Doing so effectively duplicates the explanatory burden for some effects both in the physical realm, and in the grammar. Parsimony is not served well. Note that this does not hold to the same degree for Boersma and Hamann (2008), because although physical properties in their model are treated with the same apparatus as grammatical properties, they are computed in parallel, and do not replace grammar.

In this thesis, we entertain a minimal view on phonology, where it computes only on features and their paradigmatic (co-occurrence) and syntagmatic (phonotactic) relations.[9] There are other elements in the phonological alphabet, most notably prosodic categories, but they serve mainly to describe these relations. Systemic relations (such as contrast) are not part of phonological computation; even in a framework that promotes contrast to the center stage (such as the Modified Contrastive Hierarchy, see Dresher (2009)), contrast is not seen as part of grammatical knowledge. Instead, it is epiphenomenal (or emergent) to the proposed algorithm to arrive at the featural specification of the segment inventory. The reason why systemic relations are not usually thought to be part of linguistic knowledge is because it is unclear when such relations are relevant for on-line computation of linguistic forms – either in production or perception. Phonology takes care of mapping lexical forms of morphemes to surface forms, forms which are interpretable by phonetics. This mapping function is, as shown extensively by Hall (2011), thrown overboard by Dispersion Theory, and in any case in the version presented in (Flemming, 2004). In Flemming (2004) systemic relations *are* part of what the grammar computes. In evaluating inventories rather than individual forms, the theory is no longer a theory of derivation, which is what Optimality Theory is usually conceived as.

Coming back to the main matter of this dissertation, which is the acquisition of the segment inventory, there are two questions we may ask with respect to Dispersion Theory. The first is whether there is any evidence that children evaluate entire inventories during the course of acquisition, the second is whether dispersion is an active force in it.

As with any theory of phonology framed in Optimality Theory, Dispersion Theory is in principle learnable (Tesar & Smolensky, 2000). That is to say, constraints are rankable when the learner is faced with learning data. The issue, however, is with the learning data themselves: what is it that the child evaluates? Does she evaluate entire phonological inventories or individual forms?

[9] Flemming (2006) uses the term ‘paradigmatic’ for his constraints on contrast. I will not adopt that use, but rather take ‘paradigmatic’ to mean ‘referring to distributional properties’ (what fits in what position), ‘syntagmatic’ to denote sequential relations (phonotactics), and ‘systemic’ to relations that hold between elements regardless of their context in lexical or derived forms, but rather in abstract constellations such as the inventory.

To my knowledge, there is no evidence that children perform the former kind of computation. Furthermore, although there is some evidence that larger contrasts facilitate learning (Stager & Werker, 1997), this has been brought into question (Fikkert, 2008; White & Morgan, 2008).

2.2.2 Parallel Bidirectional Phonetics and Phonology

In the introduction to this section, we proposed that inventories have both a 'shape' and a 'structure'. Whereas the latter refers to the abstract characteristics of the members, the former denotes their place in phonetic space. The model proposed in Flemming (2004) does not distinguish between these, which is one of the ways if differs from the model of Parallel Bidirectional Phonetics and Phonology (PBPP), (Boersma, 2006c, 2006a, 2006b, 2007b, 2007a; Boersma & Hamann, 2008). Whereas the constraints in DT refer to only one level of representation (the output level – remember that DT is not compatible with faithfulness), PBPP consists of multiple levels. The figure in (20) gives a graphical reflection of the levels of representation in PBPP, and the type of constraints that act on – or between – these levels.

(20) Graphic representation of levels and constraints in Parallel Bidirectional Phonetics and Phonology

"context"	⟵	situational constraints
	↖	
		reference constraints
	↙	
Morphemes'	⟵	morphemic constraints
	↖	
		lexical constraints
	↙	
|Underlying Form|		
	↖	
		faithfulness constraints
	↙	
/Surface Form/	⟵	structural constraints
	↖	
		cue constraints
	↙	
[Auditory Form]		
	↖	
		sensorimotor constraints
	↙	
[Articulatory Form]	⟵	articulatory constraints

As can be seen, the constraints relevant to the |underlying form| and /surface form/[10] correspond to the familiar Markedness and Faithfulness constraints (although Markedness constraints are dubbed Structural constraints in PBPP). Presumably, this is where the structure of the inventory arises (for further discussion, see below). However, PBPP goes one step further in also accounting for the shape of the inventory – this is done in the interactions between the /surface form/ and [phonetic form]. This approach ensures that PBPP does not run into some of the problems that we encountered in our discussion of Dispersion Theory. First, as we have seen, dispersion effects remain effects on the shape of the inventory (although they can diachronically influence the structure of the inventory as well). Secondly, the model assumes no teleology at the level of the language agent, as dispersion effects are derived from the interaction between non-teleological constraints (although this is not so clear in the case of the *Articulation constraints – see discussion below). Importantly, dispersion effects are derived without constraints on contrasts. Contrast enhancing or preserving (in a diachronic sense) effects are shown to arise as the result of the constraints, the grammar in which they live, and the learning algorithm employed. As such, PBPP is a model of derivation (of perception and production), whereas it is unclear what Dispersion Theory is aimed to model. In what follows, we will see how Parallel Bidirectional Phonetics and Phonology accounts for dispersion and stabilising effects, taking Boersma and Hamann (2008)'s case study on sibilant inventories as an example.

The main ingredients in PBPP are Stochastic Optimality Theory (Boersma & Hayes, 2001) for constraint weighting and the Gradual Ranking Algorithm (*idem*) to model learning. Parallel Bidirectional Phonology and Phonetics proposes to analyse all of speech sound processing (phonetics and phonology) in six levels. Interaction between the levels is parallel, and the same levels are used in both perception and production (bidirectionality). Each level is (part of) an OT grammar, and is defined by the constraints which act upon it.

The upper two levels describe the interface between the semantics and phonology of a language; this interface accounts for the interaction between phonology and the lexicon, where situational, reference and morphemic constraints will for instance allow the speaker/listener to decide between homonyms, and lexical constraints account for lexicon retrieval. Moving down in the model, we come to what might traditionally be considered 'core' phonology: the relation between underlying and surface forms. This relation is determined by the interaction between faithfulness and structural constraints, where the term 'structural constraints' is roughly equal to the more traditional term of 'markedness constraints'. Going down further, we immediately encounter the 'phonology-phonetics interface' (Boersma, 2006c, §3)). This interface is regulated by cue constraints, which map phonetic (ie. gradient) values to phonological (ie. discrete) elements (phonemes in this case). It might be thought that the next two

[10] In this section, use of bracketing is copied from Boersma and Hamann (2008) and differs slightly from the usual convention.

levels need not be considered in parallel because one either speaks or listens, but as the model is bidirectional, both are indeed simultaneously processed. This is motivated by the understanding that a speaker constantly monitors her own speech. There is an asymmetry here, however. Although a speaker is also a listener, the opposite does not hold: in PBPP, it is not necessary to employ sensorimotor and articulatory instruments (which, in PBPP, are directed by constraints in an OT fashion) to hear, recognise and process speech (e.g., Boersma, 2006c; Boersma, 2007a, figure 4).

For present purposes, only the lower four constraints in PBPP are of importance. In Parallel Bidirectional Phonology and Phonetics, the shape of inventories is regulated by the interactions between cue constraints and articulation constraints. Cue constraints map phonetic values (e.g., values of formants, spectral means, voice onset times) to phonemes. Thus, a low-ranking constraint that prohibits the mapping of a spectral mean value of, say, 7000hz to an /s/, will both increase the likeliness of noise of 7000hz to be interpreted as an /s/ in perception, and in articulation for the speech organs to be instructed to produce noise with a spectral mean of 7000hz when producing an [s] (Boersma & Hamann, 2008).[11] Theoretically, the number of these constraints is equal to the product of the number of entities to map to and the resolution of the phonetic dimension. For example, in vowels, if the range of frequencies for a given formant is dividable in x steps of JND (Just Notable Difference), the number of constraints referring to that constraint would be x * [the number of vowels in the inventory].

In their 2008 article, Boersma and Hamann show that indeed, given the size of a portion of the inventory (the portion being the class of sibilants in this case), an artificial learner will learn to map the correct frequencies to the two sibilants in English (/s/ and /ʃ/). Furthermore, the learner is able to handle any inventory size: the correct mapping between frequency and phoneme is achieved in all cases (Boersma & Hamann, 2008). The learning algorithm that is used is the Gradual Learning Algorithm (GLA, Boersma, 1997; Boersma & Hayes, 2001). This algorithm assigns all constraints a ranking value, and ranks them on a continuous scale. Each time the grammar is evaluated for either production or perception, a small amount of evaluation noise distorts the ranking values, such that two constraints that have only a small difference in ranking value between them may actually vary in their respective ranking from one instance to another. This way, GLA accounts for variation. A result of this is that perceptual decisions vary, and they do so in a probability matching way. For example, if the distributions of two phonemes overlap with respect to a certain phonetic dimension, the listener, when equipped with GLA, is able to

[11] The mapping between auditory form (a collection of phonetic values) and articulatory form (a set of instructions for the speech organs) is directed by a set of sensorimotor constraints. A possible example of such a constraint is given in Boersma (2006c): an auditory high F1 does not correspond to an articulatory raised jaw. More formally, these constraints would directly address individual muscles. It is not clear why they should be of interest for the grammar, as they merely implement articulations.

decide between the two phonemes even if the input she receives may be mapped to either one. She does so in a ratio analogous to the distribution ratio in her input.

What is important about the notion of inventory espoused in Boersma and Hamann (2008), is that it strikes a balance between the functional goals of 'Ease of Articulation' and 'Perceptual Clarity', without using teleological or imprecisely defined constraints. Following Boersma and Hamann (2008) in their application of Occam's Razor, I propose that Parallel Bidirectional Phonology and Phonetics is preferable to Dispersion Theory.

In the above mentioned study, Boersma and Hamann make a number of simplifying assumptions. Most of these are relatively uncontroversial, but in paragraph 5.3, one is made that reveals one of the most serious challenges for PDBB as a theory about inventories. A telling quotation is given below:

> We describe here the situation when a child already has correct lexical representations, but not yet an adult-like prelexical perception. That is, she already knows which lexical items have an underlying /ʃ/ and which have an underlying /s/...

In other words, PBPP can predict the phonetic contours of phonemes, but it cannot independently decide which phonemes are part of a language's inventory, and which are not. In order to reach a stable state, the number and labeling of phonemes must be known to the learner. The phonetic identities of members of the inventory are determined by constraints interacting on the /surface form/, [auditory form] and [articulatory form]. Thus, although the *shape* of the inventory is emergent and non-teleological, PBPP leaves open the question of phonological representations: inventories of abstract entities segments that are composed of features. The *structure* of the inventory is not the subject matter of Boersma and Hamann (2008). Because PBPP operates on an inventory of phonemes whose members are determined elsewhere, it is still a holistic theory: given a phonemic inventory, it can predict its phonetic shape. It cannot independently predict the phonological identity of the inventory members. Note, however, that it is very well possible to integrate the theory of Feature Co-occurrence Constraints developed in the current thesis into the Parallel Bidirectional Phonetics and Phonology model. This will be pursued further in chapter 5, after we have properly investigated the Feature Co-occurrence Constraint theory. It should be mentioned, however, that the GLA has been successfully applied to modelling real world language acquisition (Boersma & Levelt, 1999, for example).

2.2.3 Shape arises from structure: Dispersedness through contrast and enhancement

The two theories outlined above primarily concern themselves with the *shape* of the inventory, in particular, with dispersion effects. Whereas the Dispersion

Theory does not concern itself with the structure of the inventory (in fact, it is questionable whether the structure of the inventory can be described in DT, with its emphasis on phonetic representations and antagonistic attitude towards input-output mappings), it aims to describe dispersion effects, rather than derive or explain these. Parallel Bidirectional Phonetics and Phonology, on the other hand, distances itself from the overly teleological perspective of DT, and demonstrates that dispersion effects can be seen to emerge once a rich, integrated model of phonetics and phonology is assumed. Still, however, it does not say much about the structure of the inventory, and in fact, must pre-suppose knowledge of phonemic categories in the learner.

Taking a different route, Hall (2011) shows that dispersion effects arise naturally and predictably when we mix two pre-existing theoretical concepts: Contrast and Enhancement. The theory of phonetic enhancement (as expressed in, for example, Keyser & Stevens, 2006) holds that, while underlying phonological representations are specified only for a limited, language-specific set of features (that is, languages select a sub-set from the set of universally available features), in the course of speech production these are supplemented with features that enhance the features of segments that are otherwise in danger of losing perceptibility. As an example, let us assume a three-member vowel inventory /a, i, u/. Here, /u/ need only be specified for [+back], and rounding is supplied later to enhance the effect of backness. Similarly, speakers of English add rounding to /ʃ/ to enhance the effect of [-anterior].

The problem with enhancement theory is that it does not, by itself, supply a principled means to decide which features are primary, phonological, and stored lexically, and which features are enhancements. At the same time, the Modified Contrastive Hierarchy has no means of deciding by itself how redundant (non-contrastive) features are filled in at the surface. Hence, Hall (2011) proposes to combine the two, according to the following schema (Hall, 2011, example 10):

(21) *Elements of a theory of contrast and enhancement*

a. Phonological feature specifications are assigned by the Successive Division Algorithm

b. Only these contrastive feature specifications are phonologically active

c. In phonetic implementation, redundant properties of segments tend to be filled in in ways that enhance the auditory impression of their contrastive features

d. Phonetic enhancement is variable across languages, speakers and contexts,and the distinctness of phonemes is sometimes reduced by other factors, such as auditory overlap (Stevens & Keyser, 2010, §4)

Hall (2011) then proceeds to list some ways in which (21c) can take place, ranging from the amplification of the phonetic dimensions inherent in a feature specified as per (21a), to enhancing a contrastive (and thus lexically specified) feature's phonetic correlate by a separate phonetic event that increases the salience of the first.

To illustrate how the contrast and enhancement approach derives dispersed inventories, a very simple demonstration suffices (but see Hall, 2011 for many more). Consider again the inventory /i a u/, which we discussed in section 2.1.3 above. To characterise this inventory, two binary features are minimally necessary (and, since we are dealing with contrastive specifications, also maximally). Two features means two ways of ordering divisions, which means two possible hierarchies. Example (18) is repeated here in (22) below, for clarity.

(22) Possible specifications for /i a u/

a. [high] >> [round]

	i	a	u
[high]	+	-	+
[round]	-		+

b. [round] >> [high]

	i	a	u
[round]	-	-	+
[high]	+	-	

Now consider a sub-optimally dispersed inventory /ɨ ɘ ʉ/. The question is, why such an inventory is unattested. If we apply the Successive Division Algorithm to this inventory, we again need two features. /ɨ/ and /ʉ/ are [+high], to the exclusion of /ɘ/. /ʉ/ is [+round], to the exclusion of /ɘ/ and /ɨ/. It follows that we need [±high] and [±round] minimally (and maximally). Again, there are two possible sequences of applying the binary divisions, listed in (23):

(23) Possible specifications for /i ɨ ɘ ʉ/

a. [high] >> [round]

	ɨ	ɘ	ʉ
[high]	+	-	+
[round]	-		+

b. [round] >> [high]

	ɨ	ɘ	ʉ
[round]	-	-	+
[high]	+	-	

As becomes clear immediately, there is no difference in the feature specification of /i a u / and /ɨ ɘ ʉ/, and, as enhancement can only apply to the specified features, it is unlikely for any of the two possible schemata in (23a) or (23b) to surface as /ɨ ɘ ʉ/: /a/ is lower then /ɘ/; for example, and the roundedness of /u/ is more enhanced (by backness) then that of /ʉ/. Furthermore, the possible feature specifications for /i a u/ are a subset of the possible feature specifications of /ɨ ɘ ʉ/ (for example, a system with [±low] and [±back] could describe the former, but not the latter). Thus, /ɨ ɘ ʉ/ cannot be distinguished from /i a u/, but the reverse does hold. Hall (2011) shows that the contrast and enhancement theory excludes other unattested vowel inventories (such as horizontal or diagonal ones) through a similar logic, and that the same holds for consonant inventories. The key to the success of the contrast and enhancement approach appears to be that, rather than mix and confound phonetic and phonological effects, it separates phonetics from phonology in a principled way, and therewith separates phonetic and phonological tendencies. It has long been observed that inventories tend to be dispersed through phonetic space (shape), but also that they tend to be symmetrical in their phonological specification (structure). By allowing dispersion effects to act at the level of the feature, and allowing only contrastive features to be specified, the contrast and enhancement approach appears to have found a equilibrium between phonetic and phonological forces, which makes correct predictions, too.

There, are, however, some issues with the contrast and enhancement theory. First of all, there is the matter of underspecification. Phonetic enhancement applies at the feature level, and only contrastive features are specified. In a standard feed-forward model of the phonology-phonetics interface, however, the phonetic implementation module has no access to underlying structures. This means that underspecification must be permanent, in the sense that it holds at the phonological output level. Redundant features are thus not filled in by the phonology, and cannot be supplied by the phonetics unless they enhance one or more contrastively specified features. Whether this is a real problem must be tested empirically.

A second problem is that enhancement does not *really* hold at the feature level only: only those correlates of features are enhanced that result in more contrast. Hence, not only the phonology must have access to the entire inventory in online processing, the same holds for the phonetic implementation module. It appears, thus, that the contrast and enhancement theory is, in some sense, holistic.

Finally, the Successive Division Algorithm is successful in deciding the underlying representation of entire inventories; it remains to be seen, however, whether it is a suitable model for a learner whose inventory grows over time. In other words, whereas it is an excellent solution to the logical problem of language acquisition, it remains to be seen how well it fares in the light of the developmental problem of language acquisition. This is an issue we will address in chapter 5.

2.3 The Origin of features

In this thesis, a theory of Feature Co-occurrence Constraints is developed that governs the acquisition of the segment inventory. Given that the constraints we propose concern distinctive features above all else, we must consider the developmental origin of these features.

If we want to assess whether features are innate or not, we must consider them in all their facets. Generally speaking, features serve three major functions: a) to categorise or distinguish contrasts, b) to organise the lexicon, c) to provide classes of triggers/targets of phonological processes. Arguments for innateness must then ideally present evidence in all three areas, but this requirement runs into trouble immediately: for example, children are born without knowing any words. The lexicon is something that must obviously be learned. How to assess, then, whether the stuff that it is composed of is innate? In fact, similar problems exist for all arguments for and against innateness. What we need is a precise definition of innateness, so that the claims can be tested and the evidence evaluated. For this reason, we shall adopt a definition proposed by Elizabeth Spelke (Spelke, 2012):

Innateness A given property P is innate, if a child shows evidence of knowledge of P before or at the first encounter with P

We must be careful in applying Spelke's definition because of the complicated situation of language, in which, for one thing, it is not always clear what counts as a 'first encounter'. Related to that, there is a danger of circularity. For example, if the targeted rules in a rule-learning task are defined as operating over distinctive features, it is difficult if not impossible to say whether the results say anything about the features *per se*. The question is whether children exhibit knowledge of rules or alternations that are defined in terms other then feature classes. More literature is available when we consider the storage of lexical items and phonetic categorisation, for the simple reason they have been studied more intensively. Nevertheless, it could be rewarding to set up a research program based on the questions derived from Spelke's definition. A first word of caution is in order, then (another one will follow below), that the current paragraph should be read as a tentative survey rather then a conclusive study.

In the end, the question about innate features is one that may have deep conceptual import, but at the same time little consequences for specific phonological theories. Virtually all theories of phonology adopt some kind of subsegmental abstract generalisations, and whether they are called features, elements or something else, the evidence for sub-segmental organisation is overwhelming. Restricting ourselves now to feature theories (as opposed to alternatives, such as Element Theory, see e.g. Harris & Lindsey, 1995; Backley, 2011), the convergence of different feature sets is often remarkable and independent of whether features are assumed to be innate or not. In fact, Mielke (2008, p. 76) argues that only the most radical innatist position, wherein 'innate' means 'substantively innate', is interesting, as it is difficult to imagine different empirical predictions between learned features and features that must be 'filled in'. The problem is that it is not immediately clear what the different predictions are, even if we juxtapose the most radical innate position with the most extreme emergentist one. As we have seen above, emergentist features resonate well with usage-based, strongly functionalist approaches to phonology, but there is no *a priori* reason why they should not be part of more formalist proposals, where features are taken to be universal: universal does not logically entail innate. It seems then, that we must be content with a discussion that, to a large degree, has 'merely' philosophical ramifications, and less empirical relevance – but this does not make it less important.

Before we go on to considering some of the available evidence, it is necessary to stop and defend the strategy that we will employ. A standard argument in favour of innateness is Poverty of the Stimulus, which holds that not everything that is necessary to acquire language is present in the input. The argument has been proposed most vehemently in syntax, and not so much in phonology. With respect to features, it is often said that enough data is present in the input, so that there is no Poverty of the Stimulus argument, and therefore that there is no evidence for innateness.

This line of reasoning is problematic in a number of ways. To begin with, remember the old adage 'absence of evidence is not evidence of absence'; the fact that there is no evidence for a Poverty of the Stimulus argument does not mean that there is no evidence for innateness. Related to that is the problem of the null hypothesis. Implicit in the line of reasoning sketched above is that non-innateness is the null hypothesis, and hence that positive evidence must be provided for the competing alternative. Rejection of that alternative hypothesis (whatever it is) by no means entails acceptation of the null hypothesis.

Another issue concerns the so-called *Innateness of Primitives Principle* which holds that "... learning can consist only of creating novel combinations of primitives already innately available" (Jackendoff (1990), cited in Hale & Reiss, 2003). One corollary of the principle is that language acquisition, which entails applying a linguistic interpretation to a linguistic input (the notion 'linguistic input' in itself presupposes knowledge in the learner of what counts as linguistic and what does not). But in order to interpret ('parse') the input, to assign it a mental representation, the learner must have available at least *some* primitives to construct this initial representation. Colloquially phrased, Hale and Reiss (2003) mention that "... YA GOTTA START WITH SOMETHING!" (their capitals). Putting it somewhat differently, YA CANNOT MAKE SOMETHING FROM NOTHING (my capitals). We cannot form generalisations over input data, not even stochastic or distributional generalisations, if we cannot parse (interpret) that input data. We cannot map an input onto a mental representation *if we have nothing to map onto.*

We must concede that the innateness of primitives principle and the arguments against the anti-PoS reasoning in and by themselves are a far cry from supporting the claim that phonological features are innate.[12] Hence, we do not

[12]One complication often introduced in the innateness debate is the separation between underlying mechanisms and the units of knowledge that emerge from these mechanisms. It is proposed that the mechanisms are innate, rather than the knowledge that they create. With respect to phonological features, the mechanisms could be a) a bias to pay attention to speech, b) a stochastic pattern detector, c) a generalisation hypothesiser and so forth. However, such a view is inconsistent with the innateness of primitives principle. Consider a). For such a bias to work, the child must have knowledge of what is language and what is

claim here that phonological features as we know them in phonological theory (theories) are substantively innate. It does seem warranted to keep the possibility available, however, and to assess the results of decades of studies keeping in mind that for learning to occur, *something* must be innate (cf. Exemplar Theory, see also section 4.7.1).

An important term that will guide us in this section concerns the Continuity Hypothesis, which holds that the linguistic system of children is not *substantively* different from that of adults. In other words, acquisition does not alter the material of which representations are construed, or the manner in which linguistic input is processed. For example, according to the Continuity Hypothesis, if the adult lexicon is analysed in terms of features, so must the child's lexicon, and not, for instance, in terms of holistic phonetic *gestalts*. Also, if mapping from input to output in the adult phonology proceeds by rules, then so it must be for children. The content of the rules may differ, and so may the underlying representations, but there is continuity in the formal mechanisms that are available.

Section 2.3.1 discusses some of the extensive literature that exists in psycholinguistics, in which the early (that is, first year) linguistic categories are investigated. It turns out that young children have a remarkably precise ability to distinguish speech sound categories, and gradually tune in to the relevant categories of their ambient language in the second half of the first year of their live. Section 2.3.2 then deals with the matter of how adult-like young children's underlying representations are. Broadly speaking, there are two camps in the literature: first, there are those who assume that the early lexicon is essentially adult-like, and second, there are those who propose that the early lexicon is holistic and only becomes analysed in terms of features when the need arises (for example, for reasons of economy). We will also encounter a third view, which holds that whereas the early lexicon is not adult-like from the start, it is not *substantively* different. That is to say, the lexicon is organised in terms of distinctive features from the very start, but for children, the features have a different domain of application (the word, rather then the segment/root node). In the next section, we will discuss some work on how children deal with alternations. Before attempting to conclude this section, a brief side step will be made, in which we will discuss a case of segmental unfaithfulness, motivated not by markedness or ease of articulation or frequency effects, but by deeply rooted phonological (prosodic) knowledge. It turns out that cases like these are not hard to find once one knows what to look for, and they shed a different light on the innateness debate.

---

not. For patterns to be recognised, they must be parsed, which brings us back *linea recta* to the innateness of primitives principle. Despite the apparent appeal, the mechanisms versus knowledge idea of innateness is thus somewhat problematic.

2.3.1 categorisation and distinction

Early perception studies

Infant perception studies (as opposed to production studies, whose history goes back much further, for example, Preyer, 1895) can be argued to originate in 1971, by Eimas et al. (1971). Using the High Amplitude Sucking paradigm (a variation on the theme of Habituation-based research paradigms), Eimas et al. (1971) present results that show that voicing categories are universal. In the experiments, synthesised speech tokens were presented to one month old and four month old infants. The speech stimuli were constructed such that the VOT varied from pre-voicing to aspiration, on 20ms. intervals, straddling the three 'adult' phonemic categories (pre-voicing, short lag, long lag). Both age groups were subdivided in three groups, according to the type of test stimulus they would receive (control, same category, different category). For the control group, the test stimulus was the same as the stimulus they had been habituated on. For the other two groups, the test items were different, in that the VOT had shifted by 20ms. The difference between the two experimental groups is that for one, the test stimulus and habituation stimulus belonged to the same 'adult' VOT category, whereas for the other, the difference straddled a category boundary. Thus, although physical difference was the same for both groups, perceptual difference was predicted to be different. Indeed, children in the Different group dishabituated, whereas children in the Same and Control groups did not. These results show that children, from a very young age are able to detect very minute differences in speech sounds, and furthermore, that they categorise speech sounds along the same boundaries we find in adult language cross-linguistically.

Subsequent studies showed that these results could be replicated on other phonetic dimensions, indicating that infants are able to discriminate between all the speech sounds (categories) found in adult languages, regardless of their ambient language. In two major studies, it was shown that infants tune in to relevant language specific categories before they reach the end of the first year of their lives (Werker & Tees, 1984; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992).

In a series of three experiments, with both infants (English-learning) and adults (English and Thompson Salish native speakers), Werker and Tees (1984) sought out to investigate the time path of native language speech sound category formation. In the first test, 7 month old English learning infants, English adults and Thompson Salish adults were tested on the contrast between the velar and uvular voiceless stops, in the syllables [ki] versus [qi]. While non-distinctive in English, the pair is contrastive in Thompson Salish. Predictably, the Thompson Salish adults reliably identify the contrast, but the English adults fail to do so. The infants, however, performed as good as the Thompson Salish adults, indicating that at 7 months, infants' perception has not yet been specialised for the language environment. The second experiment tested at what age specialisation begins. Three groups of infants were tested, at seven months of age, at nine months, and at 11 months. In addition to the Thompson Salish contrast, the children were also tested on the Hindi contrast between alveolar and retroflex voiceless stops ([ta] versus [ʈa]. As expected, the youngest group performed well on both contrasts, as did the middle group. The oldest children, however, were unable to detect the differences, indicating that their phonetic categories already conform to those of the language they are acquiring. These results were further strengthened in experiment 3, in which the younger children from experiment 1 were re-tested at 11 months. Now they, too, failed to detect the contrast, ruling out individual differences as cause of the results in experiment 2.

Having established that consonantal contrasts generally become language specific between nine and eleven months of age, the question remains whether the same applies to the vocalic system. One might predict that vowels, due to their inherently greater salience, are acquired earlier, and this is precisely what was found by Kuhl et al. (1992). In an earlier study, Kuhl (1991) had shown that both American adults and six month old infants display a *perceptual magnet* effect for vowels, meaning that prototypical tokens (those tokens rated by native adults as being 'good' exemplars of their category) warp perceptual space. In other words, non-prototypical tokens are less likely to be judged as 'different' when presented in conjunction with a prototype, then when the competing stimulus is a different non-prototype. Testing both American and Swedish infants on /i/ (as in the English word *fee*, thus prototypical for (American) English but not for Swedish) and /y/ (as in the Swedish word *fy*, thus prototypical for Swedish but not for English), and 32 non-prototypical tokens per category (all tokens, including the prototypes, were synthetically generated), Kuhl et al. (1992) show that by six months of age, infants display a stronger prototype effect for their native language prototype then for the non-native prototype. This implies that by six months of age, the perception of vowels has become language specific.

So far, the results of the studies mentioned are compatible with innate features. Even if they do not explicitly support or assume the notion of innate features, they are strikingly compatible with the innateness of primitives principle: children from a very early age display knowledge of linguistically relevant categories. The general picture that emerges is that children grow from being universal speech perceivers to language-specific speech perceivers within the time span of a year; and they do so by erasing phonetic category boundaries that are irrelevant for their native language. Although it might seem counter-intuitive that acquiring a language means to become less precise, it is worth bearing in mind that ignoring irrelevant categories greatly enhances the robustness of the perceptual system – *and* that more precise means more restrictive (as per Hale & Reiss, 2003).

Emergent features and distributional learning

Much of the later psycholinguistic literature on language acquisition focused on statistical learning mechanisms. This found a resonance with phonologists, for example Mielke (2004). The initiation of this shift is perhaps best exemplified by Saffran, Aslin, and Newport (1996), whose subject is speech segmentation rather than the inventory.

As Maye and colleagues (Maye & Gerken, 2000; Maye et al., 2002; Maye & Weiss, 2003) argue, at the point at which the native language phonetic categories take shape (i.e., between six and twelve months of age), the (receptive) lexicon is too small to contain enough minimal pairs to compare. In a series of experiments, Maye and Gerken (2000); Maye et al. (2002); Maye and Weiss (2003) argue that in stead, children learn their native language categories by means of 'distributional learning'; that is, they pay attention to the frequency with which certain categories are produced.

As we have seen above, results from earlier studies indicate that within the first year, infants go from being 'universal listeners' to language specific perceivers. They do so by 'unlearning' categories that are not distinctive in their ambient language, but the mechanisms by which they accomplish this are largely unknown. In their 2002 article, Maye et al. (2002) investigated whether exposure to different types of input frequency distribution would aid children in breaking down the phonetic barriers between non-native language categories. Six- and eight month old infants were presented with resynthesised speech tokens. The stimuli (CV syllables) differed solely on the VOT of the onset, such that the range went from a voiceless unaspirated onset in [ta] to a voiced one in [da] in eight steps. The infants were assigned to one of two groups, which differed in the distribution of the stimuli, such that half of the infants were presented with a monomodal input distribution, and the other half with a bimodal input distribution. During test, both groups were tested on discriminative abilities on items that were near the extremes of the range, and, as predicted, children who had been exposed to the bimodal distribution performed significantly better then the children who had been in the monomodal group.[13]

This result lead the authors to conclude that indeed, children capitalise on the input frequencies of speech token to determine whether a given category boundary is irrelevant. In a follow-up study, the question was investigated whether the reverse also holds: does exposure to input distribution aid discriminative abilities? English-learning eight month olds were tested on the same type of stimuli as in the earlier study, but a second set of stimuli was added, which differed only on place of articulation. Thus, there were two groups of stimuli; one ranging from voiceless unaspirated [t] to voiced [d], and one from voiceless

[13]Similar results were obtained with adult subjects in a study reported in Maye and Gerken (2000). The test consisted of trials in which two stimuli were presented in conjunction with a visual display. In half of the test trials, the same item was repeated, whereas in the other half, two different tokens were used. Differential looking reactions to non-alternating versus alternating trials was seen as an indication of discrimination.

unaspirated [k] to voiced [g]. Three experiments were run, the results of which lead Maye and Weiss (2003) to conclude that being exposed to a bimodal distribution facilitates discrimination; furthermore, discrimination occurs at the subsegmental level, as evidenced by the results from experiment 3: infants in this experiment were able to discriminate changes in VOT, even after being familiarised on stimuli with a different PoA.

Convincing though the results may be, some words of caution are in order. First of all, the amount of exposure was extremely limited (2.30mins), and occurred immediately prior to testing. What is measured is not the children's knowledge of language but rather their ability to process speech input (as was the goal of the studies). Whether these results have any bearing on what happens outside the laboratory, where input is both less clear, less concentrated, and exposure is much more prolonged, is still an open question. Secondly, even if the model is correct, it does not say much about *phonological* category acquisition. Statistical calculation over the input may give information about the surface structure of the language, it does not help much when constructing underlying representations. For one thing, the distinction between phonemic and allophonic relations cannot be read off the input distribution. Also, input frequency has been found to be a poor predictor of the order of acquisition of segments in production (Levelt & van Oostendorp, 2007). All in all, the cited studies paint a credible picture of a model of early category formation, if not somewhat unsurprising: if children go from universal discriminators to language specific contrast detectors, how else then through exposure to the ambient language can they do so? The most interesting result is that even in these early stages, children make sub-segmental generalisations.

All in all then, it is not straightforward to reason from theses studies that features must be emergent. In fact, the results of experiment 3 in Maye and Weiss (2003) indicate that generalising from independent phonetic parameters to classes of segments – which is made possible by features – is not something young children have any trouble with. Even if features are not substantively innate, the ability to analyse speech in a featural manner must be a very fundamental capacity.

2.3.2 Early representations

We have now seen a number of studies that deal with early linguistic development, focusing on phonetic/phonemic categories. Some seem to accept innate features, some are compatible with it (Eimas et al., 1971), some question the idea (Maye & Gerken, 2000), and some explicitly reject it (Mielke, 2008, see below). However, if we want to talk about features, we need to go beyond mere categories. As features are the building blocks of the lexicon (the status of the segment as an independent phonological prime is not clear, see Lahiri & Reetz, 2002), we need to know if and how these categories are used to build the (early) lexicon.

A long standing question in the research of child language is whether the underlying representation of children is adult-like or not. Early generative studies in child phonology assumed adult-like underlying representations (Smith, 1973; Ingram, 1989), but this point-of-view was criticised in the work of, amongst others, Ferguson and Farwell (1975). Under the first view, a child's phonological system must at least be adult-like in the types of symbols it manipulates (representations), but whether the type of manipulations (=derivations) are adult-like is an open question. According to those who oppose this view, the child's early representations can be very different from those of adults. One popular view, for example, is that children store words holistically, as phonetically unanalysed acoustic units, until the lexicon reaches a certain size at which such rote memorisation becomes untenable (or at least sub-optimal). At this point, the lexicon will be analysed and generalisations will be made, resulting in a more adult-like system.

Lexical organisation is one of the three pillars of features, and as such, the nature of the early lexicon bears directly on the question of innateness of features. In this section, we will examine some of the literature on early representations to see to what degree we can say the child's lexicon mirrors that of the adult.[14]

An important question concerns the amount of phonetic detail that children store in their early lexicon. In this respect, an important study is Stager and Werker (1997). Using the newly developed Switch method (a variation on the habituation/dishabituation theme), Stager and Werker argue that although infants are able to distinguish fine phonetic detail, they are incapable of storing such detail in the lexicon. The reason for this is that the task of word learning places such high demands on processing resources, that these can no longer be allocated to phonetic distinction.

In the first of a series of four experiments, 14 month olds were presented with the stimuli *bih* and *dih*. Both stimuli were presented in combination with a visual stimulus (a picture of an unknown, brightly colored object), while looking time was measured. After a pre-set criterion was met the test phase commenced. Here, the same visual display is presented, in combination with either the original stimulus (Same), or the other stimulus (Switch), such that word-object pairings are switched in half of the test trials. Looking time is measured and dishabituation in response to the Switch stimulus is taken to be a sign of discrimination. In this two-by-two design, the children failed to dis-

[14] It should be noted that some have argued that this question has become obsolete with the rise of Optimality Theory: due to Richness of the Base, the adult form must be part of the candidate set. In my opinion this line of reasoning is incorrect or at least incomplete, as it foregoes the possibility that the actual *substance* of lexicon at the two stages differ. Under the assumption that children store their early forms holistically, it is hard to imagine how GEN could create a range of featurally specified candidates, what type of constraint could decide between the two types of forms and how the non-specified form could be the optimal form under Lexicon Optimisation. Furthermore, what type of evidence would drive the learner to rearrange the constraints in CON such that the featurally specified forms become the preferred underlying forms? In other words, the argument holds only within a set of competing theories that both assume that the *substance* of the lexicon remains constant.

criminate. In a follow-up, the task load was lightened by only including a single word-object pair in the habituation phase. Fourteen month olds did not dishabituate, but eight month olds did. To test whether the null results obtained so far are due to the stimuli, rather then the design of the study, experiment two was repeated with more distinct stimuli: *lif* and *neem*. Indeed, 14 month olds noticed the difference. Finally, experiment four was a repetition of experiment two, but with the visual stimuli replaced by a display of a nondescript, boundless image of a checkerboard. Such a display, the authors argue, is not interpreted as an object by young children, and therefore, the task changes from a word-learning task to a discrimination task.

The reason for the difference between the age groups in experiment 2, the authors argue, is that the younger children are not yet building a meaningful lexicon, which means that for them, the task is not a word-learning task, but rather a discrimination task. Thus, the effort of word-learning does not interfere with phonetic discrimination. This is also the motivation behind experiment four: to show that the problem for the older children does not lie with the discriminability of the stimuli *per se*, but rather with the demand of having to discriminate and learn words at the same time. Experiment 3 showed that the task in itself is solvable, when the stimuli are more favourable (more distinct).

A potential problem with Stager and Werker (1997) is that in English (North-American English at least), forms such as [bɪ] and [dɪ] are not possible words; another issue is that only one dimension (PoA) was tested. These issues were taken up in Pater, Stager, and Werker (2004), who replicated the original study with the following adaptations: the stimuli were changed to conform to English phonotactics: *bin* and *din* (experiment 1); a voicing contrast was tested: *bin* versus *din* (experiment 2); and finally, a two-feature change was tested: *pin* versus *din* (experiment 3). in all three experiments, the original results from Stager and Werker (1997) were replicated: children at fourteen months of age are unable to detect the change in stimulus, seemingly reinforcing the interpretation that infants are unable to encode phonetic detail when learning words. This, of course, implies that the early lexicon is substantially different from the adult lexicon, where fine phonetic detail is stored in so far as it is contrastive in the language; in other words, in so far as it concerns the phonetic correlates of distinctive features.

Stager and Werker (1997) was not accepted without criticism. Two types of reaction can be found in the literature: first, it is proposed that children *are* able to store features, but that not all features are stored equally. The failure of the older children in the Stager and Werker (1997) study, then, is due to a problem with the stimuli (Fikkert, 2008). A different response also proposes that children are able to store sub-segmental details, and that the failure in SW's experiments has to do with the task design (White & Morgan, 2008).

Developing a representation

In a series of studies, Fikkert and colleagues (e.g., van der Feest (2007); Fikkert (2008); Fikkert and Levelt (2008)), working in the FUL paradigm (Lahiri & Reetz, 2002) propose that the underlying representations of children are not adult-like from the start, but their proposal is still consistent with the Continuity Hypothesis: the building blocks of the child's phonological system are no different from those of the adult's: features and (OT) constraints. What is different is the domain of application for features: children start out in a one-word-one-feature stage, after which the word becomes increasingly more segmented.

The FUL model (Featurally Underspecified Lexicon, Lahiri & Reetz, 2002) proposes that items in the phonological lexicon consist of features, but not all features: [coronal] is not represented (note that this does not mean that FUL denies the existence of [coronal]. It is perceived, but not stored, meaning that while it can be part of phonological processing, it is never part of the lexical representation). The model proposes that in lexical recognition, all word-forms are activated and compared with the perceived form. For each feature, there are three possibilities:

- Match: the lexical item remains a candidate for the perceived form, the next feature is compared
- Mismatch: the feature in the lexical item does not match with the feature in the perceived form. The lexical item is discarded as a candidate
- No Mismatch: the feature in the perceived form neither matches nor mismatches the feature in the lexical form. The lexical form remains a candidate.

The latter situation occurs, for example, if a listener hears the form [pukæn] for /tukæn/; Starting with the first segment, the listener compares the feature [labial] of the [p] to the stored PoA of the /t/: ∅. There is neither a match nor a mismatch. On the other hand, if the listener hears [tæɹət] for /pæɹət/, the feature [coronal] is perceived in the first stop, and a mismatch with [labial] in the underlying form is the result. In this way, the model is able to account for variation in lexical retrieval.

FUL in acquisition

Fikkert (2008) and Fikkert and Levelt (2008) propose that the failure of children in the Stager and Werker (1997) study is not due to task demands, but to two other factors: first, in the early stages of the lexicon, children only store one feature per word: the feature of the stressed vowel. Secondly, the feature [coronal] is (permanently) unspecified in the lexicon. The upshot of this is that the 14 month olds in the Stager and Werker (1997) and Pater et al. (2004) studies never stood a chance, because the habituation items contained a coronal vowel. Hence, nothing could be stored, although the items could be discriminated. Under this view, the success on the discrimination task is not because of the lesser task demands, but rather because the lexicon was not involved to begin with. Remember that [coronal] is *perceived*, even though it is not stored. A crucial notion here is *staged segmentation:* words are initially stored only by the features of their stressed vowel only (vowels and consonants have the same place features), even though other features (of onsets) are perceived.

FUL in word-learning: Fikkert (2008)

Fikkert (2008) reports on a number of experiments testing the FUL model and the hypothesis that this specific model of lexical representation can explain the null results reported by Stager and Werker (1997). Experiment 1 in Fikkert, Levelt, and Zamuner (2005) aims to replicate a version of experiment 2 of Pater et al. (2004), with *bin* and *din* as test items. The method is the Switch, with one word-object pair in habituation. The prediction is a null result, because when learning *din*, the children will perceive [coronal][coronal], and hence store ∅ (remember that in the early stages, only the feature for the vowel is stored). Then, when confronted with the Switch *bin*, children perceive [labial][coronal], which will result in a No Mismatch mapping with ∅. Similarly, when children learn *bin*, they perceive [labial][coronal], and store ∅. Again, mapping *din* [coronal][coronal] results in a No Mismatch situation. Further experiments test various permutations of syllables with labial or coronal onsets and /ɪ/ and /ɔ/ nuclei (and /n/ codas). Table 2.3.2 summarises the experiments:

Learned Word	*Stored Representation*	*Perceived Form in Test*	Matching
bin/din	*null*	labial coronal (bin)	No Mismatch
		coronal coronal (din)	No Mismatch
Learned Word	*Stored Representation*	*Perceived Form in Test*	Matching
bon	[labial]	labial labial (bon)	Match
		coronal labial (don)	Mismatch
don	[labial]	coronal labial (don)	Mismatch
		labial labial (bon)	Match
Learned Word	*Stored Representation*	*Perceived Form in Test*	Matching
din	*null*	coronal coronal (din)	No Mismatch
		coronal labial (don)	No Mismatch
don	[labial]	coronal coronal (din)	Mismatch
		coronal labial (don)	Mismatch
Learned Word	*Stored Representation*	*Perceived Form in Test*	Matching
bin	*null*	labial coronal (bin)	No Mismatch
		labial labial (bon)	No Mismatch
bon	[labial]	labial coronal (bin)	Mismatch
		labial labial (bon)	Match

Table 2.3: Summary of the conditions reported in Fikkert, 2008

It turns out that, as predicted, children show a significantly different reaction to the Switch test trial compared to the Same test trial on experiments 2 and 3, but not experiment 1. The initial Stager and Werker (1997); Pater et al. (2004) results are replicated, but shown to be more complicated than assumed earlier. The FUL model has made the correct predictions in this series of experiments. In experiment 4, when habituated on *din*, ∅ is stored, so both the Same and the Switch will result in a No Mismatch. On the other hand, when habituated on *don*, [labial] is stored. For both the Same and the Switch, [coronal] is perceived, resulting in a Mismatch. In both conditions, the matching procedure has equal results for both Same and Switch, so children are predicted to fail on this experiment. In the final experiment, when habituated on *bin*, ∅ is stored, so both the Same and the Switch will result in a No Mismatch (as in experiment 4). When habituated on *bon*, [labial] is stored. The Same test trial will result in a match ([labial][labial] mapped onto [labial], but the Switch will result in a Mismatch ([labial][coronal] mapped onto [labial]). In the *bin* condition, the children are predicted to fail, whereas in the *bon* condition, the infants are predicted to succeed in distinguishing the Same and the Switch. Again, the results were as predicted. To sum up, table 2.4 gives the results of all five experiments.

	experiments	Contrast	Vowel or Consonant	Longer Looking Times to Switch
exp. 1	bib-din object	b - d	ɪ	no
exp. 2	bon-don object	b - d	ɔ	yes
exp. 3	bin-din checkerboard	b - d	ɪ	yes
exp. 4	din-don object	ɪ - ɔ	d	no
exp. 5	bin-bon object	ɪ - ɔ	b	yes, but only when habituated on *bon*

Table 2.4: Results in the first five experiments reported in Fikkert, 2008

Further experiments show evidence for *staged segmentation*, in the sense that the onset feature are being represented by older children (17m.o.s., see Fikkert (2008) for details). The important thing for us to remember at this point, is that the FUL-inspired experiments assume that features are available for young children both in perception and storage, and second, that underlying representations do get more detailed, *but that the material of which they are made does not change.* In her study of known word representations, van der Feest (2007) found similar effects; this is important with respect to the section on detailed representations below. Furthermore, Fikkert and Levelt (2008) showed evidence for FUL and Staged Segmentation in production, too.

The featural lexicon in production: Fikkert and Levelt (2008)

Fikkert and Levelt (2008) contributes to the debate about child language idiosyncrasies, and does so on two issues: consonant harmony and underlying representations. It does so via a study of the development of Place of Articulation. Consonant Harmony has long been the focal point of debates among acquisitionists. It is a phenomenon wherein during some phase in the phonological development, consonants in a word agree along some phonological dimension. Crucially, there is no (surface) adjacency restriction, such as there would be in cluster assimilation. Usually, Consonant Harmony is described with respect to Place of Articulation.

The reason that Consonant Harmony features so centrally in the literature is that it does not occur in adult Language (save some relatively rare occurrences of palatalisation harmony within the realm of coronals, and of some forms of nasal harmony). The existence of Consonant Harmony thus appears to challenge the Continuity Hypothesis. Earlier accounts of Consonant Harmony appeal to mechanisms of Spreading or copying, enforced by markedness or alignment constraints (e.g., Repeat (Pater, 1997)) or higher-order licensing constraints (Rose, 2000). Consonant Harmony has been described for many languages, among which Dutch (Levelt, 1994), English (Smith, 1973; Cruttenden, 1978; Menn, 1978; Goad, 1997; Pater, 1997; Rose, 2000; Pater & Werle, 2001, 2003), French (Rose, 2000) and German (Berg & Schade, 2000). Cases in more languages are reported in Vihman (1978), but as Levelt (2011) notes, it is unclear whether these cases represent systematic patterns. For this reason, Levelt (2011) concludes that Consonant Harmony is a phenomenon not as wide spread as sometimes is believed. Nevertheless, some cases remain, and thus it remains a topic of theoretical significance, because of the challenges it poses to the continuity hypothesis.

In Fikkert and Levelt (2008), five children were chosen from the CLPF database (Levelt, 1994; Fikkert, 1994), and from their utterances a selection was made: only CVC and CVCV forms were considered. Each word was coded along the following schema:

Feature	*Code*
[labial]	P
[coronal]	T
[dorsal]	K
round vowels	O
front vowels	I
low vowels	A

Table 2.5: Coding scheme for the Fikkert and Levelt (2008) study.

For example, a word like *brood* /bʀot/ ‘bread’, was coded POT[15], and the produced form [bop] was coded as POP. This was done at the level of actual productions, but also for target forms and faithful forms. Thus, the study covers three levels, or ‘tiers’ (as does the current study, see chapter 4). Next, the order of acquisition of these abstracted word forms was established, and plotted on a

[15] clusters were simplified in coding to their least sonorous member.

Guttman scale. The Guttman scales line up, from which the authors conclude that acquisition proceeds in discreet stages:

1. Whole word stage
2. C–V disintegration
3. $C_1 - C_2$ disentanglement 1: PvT ('labial-left')
4. $C_1 - C_2$ disentanglement 2: PvK, TvK ('dorsal right')
5. $C_1 - C_2$ disentanglement 3: TvP, KvT, KvP (anything goes)

Roughly the same stages were found in the Actual and Target forms. From the general results, five generalisations arise:

- Whole-word stage
- Staged Segmentation
- Emerging Constraints
- Coronal underspecification
- Input frequency effect

In the first stage, the whole-word stage, words are either POP, TIT, PAP, TAT (and KOK, KAK). The following adage holds: one word, one feature (/a/ has no PoA, just height). The authors argue that this is indicative of incomplete storage. After this holistic stage, staged segmentation sets in. The first step is for the consonants to behave different from the vowels, even if they are still identical to each other. When consonants receive an individual specification, variation is limited to the PvT pattern. In other words, only labials may occur at the left edge. The result of this is, that the child's lexicon is populated to a large degree with labial initial words. Fikkert and Levelt (2008) propose that this situation drives the child to make a generalisation: [Labial (assign a violation market for every initial consonant that is not labial). In the next stage, dorsals appear, but they are banned from C_1 position. Hence, the child hypothesises *[dorsal (assign a violation mark for every initial consonant that is a dorsal). Finally, all positions may be occupied by all places of articulation. Coronals, being underspecified in the lexicon, are always free to occur anywhere.

The upshot of this developmental pattern is that Consonant Harmony is epiphenomenal to the way the lexicon is structured and constraints emerge. First, Consonant Harmony is due to the fact that only one PoA feature is responsible for the entire word (stage 1) or the consonants in the word (stage 2). Next, [labial creates apparent harmony in words in which C_2 is also a labial.

The analysis in Fikkert and Levelt (2008) and the research reported in Fikkert (2008) demonstrate that children make use of the same grammatical instruments (features, constraints) as are present in the adult phonological grammar. Under this proposal, the child populates the lexicon with features from the very early lexicon. The way in which it is different, then, is that the words in the lexicon are not yet segmentalised to the degree that they are in the adult lexicon. To summarise, the representational symbols are adult-like (features), the derivational system is adult-like (OT), but what is different is the domain of application (initially words, then staged segmentation sets in). With regard to the current definition, Fikkert and Levelt (2008) provide evidence for innate features to the extent that the function of features in lexical storage is concerned.

The premise of the work cited above is that children's initial lexical representations are different, but not *substantively* different: there are no holistic representations in the sense of unanalysed stored chunks of speech signal. Children can use the same set of features as adults in both recognition and storage. Features thus pre-exist lexical storage.

Fine detail in the lexicon after all?

The proposal in Fikkert and Levelt (2008) and Fikkert (2008) demonstrates the possibility that the null results obtained in Stager and Werker (1997) and Pater et al. (2004) are due to properties of the stimuli, rather than the task load inherent in the experimental design. White and Morgan (2008) take the other route, and show that with a different design, it can be demonstrated that the early lexicon is capable of representing fine detail after all.

A central question for White and Morgan (2008) is to find out how adult-like children's lexical representations are. Earlier studies (Swingley & Aslin, 2000) using the intermodal preferential looking paradigm (IPLP) had shown that children are indeed sensitive to small mispronunciations of known objects. However, the magnitude of the reaction was not in proportion to severity of the mispronunciations, while such 'graded sensitivity' has been found in comparable experiments with adults. Various possible explanations are compatible with this finding. For example, a ceiling effect holds that every deviation in the stimulus (independent variable) beyond a given threshold (the ceiling) is of no – or less – influence on the reaction of the child (dependent variable). Here, the ceiling is very low, at only one feature distance. A different interpretation is that children have a more holistic representation. Here, holistic is not meant to mean 'one-word-one-feature' but rather representations consisting of unanalysed, monolithic phonemes. Thus, *ball* is as different from *shawl* as it is from *gall*, even though the distances are unequal on a feature metric. According to White and Morgan (2008), the hypothesis of holistic representations entails that early lexical information is just enough to distinguish the lexical entries (this reminds us of the minimal pair principle). New lemmas thus exert pressure to re-analyse the entire lexicon time and again.

White and Morgan (2008) argue that rather than a reflection of the child's competence, the null results for graded sensitivity were due to performance failure, induced by task effects. The standard set-up of IPLP is that the child is presented with two pictures of known objects. One object is named, either correctly or with a mispronunciation. The dependent variable is the time the child looks at the two objects.

A crucial innovation is that White and Morgan (2008) pair a known object with an unknown object. In the original IPLP set up, both objects are expected to be familiar to the child. This means, however, that both object exert an effect on the child's looking behaviour: the target object causes a so-called attractor effect; it attracts the child's attention. At the same time, the distractor object exerts a repeller effect: if the child knows the object and its name, the mismatch between the phonological form of the (mispronounced) target name and the phonological form of the the attractor's name makes it difficult for the child to accept the former as a candidate for the latter. This is repeller effect is taken to mimic real-world situations, in which children, when faced with a hitherto unheard string of speech sounds, face the choice of either mapping it to a known word (de facto interpreting the string as a mispronunciation of a known word), or creating a new lexical entry. Hence, White and Morgan (2008) argue, the standard IPLP paradigm is not sensitive enough, because it pushes the subjects in the right direction.

Three experiments are reported. In experiment 1, children were tested on one, two and three feature differences, where a mispronunciation involved a change in voicing, place, or manner (continuancy). Single-feature mispronunciations involved a change in PoA, two-feature changes combined PoA and Voice, three feature mispronunciations added Manner.

The results (White & Morgan, 2008, fig. 2 p. 120) clearly show an effect that is compatible with graded sensitivity to mispronunciations. Next, in order to rule out a possible alternative explanation, which holds that the graded results found in experiment 1 is due to graded sensitivity to mispronunciation *type* rather than mispronunciation *magnitude*, experiments 2 and 3 were run. The issue in experiment 1 is that single feature mispronunciations always involved place of articulation only; the possibility exists that children are not sensitive to PoA mispronunciations as much as they are to errors in Manner and Voice.

Qua set-up, experiment 2 is much like experiment 1, but all mispronunciations concerned single-feature deviations in one of the three dimensions. The results again show a gradient sensitivity: all single-feature changes were interpreted as mispronunciations. Furthermore, there was no significant difference between the types of mispronunciation.

Experiment 3 was designed to show that the results were truly graded, and that dimension of mispronunciation is irrelevant. In this experiment, every combination of 2-feature changes were tested. The results show that all two-feature changes were interpreted as mispronunciations; and again, there was no significant difference between the types of mispronunciation.

From this study we can conclude a number of things: first, White and Morgan (2008) interpret the difference between their own finding of graded sensitivity and earlier null results with respect to the same in terms of task effects: in the their study, children were presented with known and novel items. The interpretation is that children have different levels of tolerance, depending on referential context. Second, most importantly, these experiments show that children are sensitive to differences in pronunciation from what they have stored, and that their sensitivity is graded. In that respect, they look a lot like adults, and appear to have adult-like representations.

2.3.3 Phonological activity and phonotactics

Having considered features in acquisition from the point of view of categorisation/distinction and lexical representations, we will now consider whether there is any evidence for innate features from the perspective of phonological activity. For this, we ask ourselves (at least) two questions. First, how do children treat phonological activity (i.e., alternations)? and second, do we find evidence of substantively innate features at the typological level?

Rules and constraints operate over natural classes, defined in terms of features. We will see how children make generalisations over classes of segments, and to what degree phonological naturalness determines learnability. We will also take some time to consider Mielke (2004)'s study in which the notion of 'natural class' is deconstructed in to phonetic naturalness, phonological naturalness and phonological activity. The crucial question for Mielke is to what degree these overlap.

In this section, we will consider static, distributional generalisations (phonotactics) in addition to alternating patterns. However, as Seidl and Buckley (2005) argue, the difference between the two has never been defined clearly (cf. the 'duplication problem', Kenstowicz & Kisseberth, 1977), and furthermore, one of the original goals of Optimality Theory is to extinguish the distinction (constraints are agnostic as to how their violations are repaired, hence every [markedness] constraint is a somehow a morpheme structure constraint. Remember that the 'old' rule format X → Y / A __ B rewrites to AXB → AYB, whereby AXB stands for an illegal sequence, and AYB for its repair. A great deal of markedness constraints coincide with the structural description of rules). Both phonotactic restrictions and alternations concern legal and illegal structures in the grammar, and what is of most interest to us at this point is that if they are to be any more general than statements over individual segments, they must be defined in terms of (sets of) features.

Alternations

A key publication in the study of how children process alternations is Jusczyk, Smolensky, and Allocco (2002), discussed in more detail in section 4.6.2 below. In short, Jusczyk et al. (2002) test the presence and relative ranking of markedness and faithfulness constraints in infants (aged 10 and 4.5 months).

Their results indicate that indeed, the children show evidence of both markedness and faithfulness constraints, and what is more, that in the initial state, markedness constraints outrank faithfulness constraints. These results imply that features are innate, as is the only way in which a class of nasals could be separated out and targeted by a constraint that disallows them to have an independent place of articulation in coda position (or by whatever other constraint that enforces nasal cluster assimilation). We will return to this matter in section 4.6 below.

The alternation that was the subject of the Jusczyk et al. (2002) study concerns nasal place assimilation, in which [m] is an allophone of /n/, when the latter is followed by a labial obstruent. Hence, in English, the coronal and labial nasal stand in both a contrastive and allophonic relation. Not much is known about the acquisition of phonological rules such as allophony, but Peperkamp, Calvez, Nadal, and Dupoux (2006) propose a possible learning algorithm. Employing a metric of dissimilarity in the distribution of pairs of segments, the study shows that their algorithm can detect allophony in a corpus of pseudolanguage. The algorithm compares the distribution of two segments, and ascribes a score that correlates with complementarity. In 'real' language, however, complementarity of distribution is not a reliable cue for allophony; the authors give the example of the French semivowel [ɥ] and its vocalic counterpart [œ], but many more examples of these pseudo-allophones exist; consider, for example, [h] (never in coda) and [ŋ] (never in onset) in Dutch.

Peperkamp et al. (2006) run an allophony-detecting algorithm over a corpus of French child-directed speech, where all segments are represented as a numerical vector correlating with phonetic or phonological features. The number of pseudo-allophones (false positives) detected by the algorithm far outran the number of hits, unless the possibility space for real allophones was constrained by imposing additional, linguistic requirements on allophony. Two such constraints were employed: first, a pair of segments could only be considered allophones if no third, intermediate segment exists (in other words, allophones must differ minimally), and second, the pair was only considered if the allophone was more similar to the conditioning environment then the base segment.[16] Thus, it could be shown that probabilistic, distributional analysis of the input can lead to the detection of allophonic rules, but only if the learner is guided by prior linguistic knowledge. Crucial for our present discussion, that knowledge was encoded in a way that is very similar to distinctive features.[17]

White, Peperkamp, Kirk, and Morgan (2008) set out to experimentally investigate whether distributional learning is a viable strategy for acquiring alternations, to the degree that the allophones are in a (relative) complementary distribution. The authors tested two groups of English-learning 12 month olds, and two groups of 8.5 month olds, using the head-turn preference paradigm.

[16]See Peperkamp et al. (2006) for a functional definition of 'allophone' and 'default segment'.

[17]That is to say, the algorithm was successful only when it was constrained by limits on sub-segmental (featural) generalisations.

Children in each of the four groups were divided in two conditions, STOP and FRICATIVE. All children were familiarised on strings of a single syllable ('determiner') followed by a sequence of two CV syllables ('noun', where C was invariably a voiced or voiceless obstruent). There was no pause between the first and second syllable, and neither between the second and third. The first syllable was either *rot* or *na*. The complementary distribution was always in the onset of the second syllable (thus in the onset of the 'noun'); In the STOP group, the initial consonant of the two syllable 'word' was voiced following *na* and unvoiced following *rot* if the consonant was a stop, but not if it was a fricative. Thus, the distribution of initial stops favored an analysis in which voiced stops agree in voicing with preceding obstruents. The situation was reversed in the FRICATIVE condition. During test, children in either condition heard the same stimuli: sequences of either *rot* or *na* followed by novel disyllables, that were voiced-obstruent-initial following *na* and voiceless-initial following *rot* (experiments 1 and 2, 12m.o.s. and 8.5m.os. respectively). Experiments 3 and 4 (8.5m.o.s. and 12m.os. respectively) were similar, but the 'determiners' were removed in the test phase. This way, the authors reason it could be tested whether children learn context-sensitive assimilation patterns or actually group different surface phonemes in a single functional category.

White et al. (2008) reason that children in STOP condition would parse the stop-initial test stimuli as 'determiner+noun' pairs, as they obey the distributional generalisation they had been exposed to. For example, *rot pevi* and *na bevi* would be parsed as the same 'noun'. The fricative-initial words however, should be parsed as separate lexical items depending on the voicing of the initial consonants. In other words, the 'nouns' in *rot sobi* and *na zomi* should be treated as minimal pairs, if the children had learned the generalisation. Hence, a difference in looking time was expected. This was indeed found in experiments 1, 2 and 4. They conclude that both 12 month olds and 8.5 month olds are able to use distributional information to construct phonological rules, if the phonological context is present (experiments 1 and 2) and that 12 month olds generalise that rules to when there is no conditioning context (experiment 4, a repetition of experiment 1 but without the 'determiners' in the test phase). In experiment 3, however, the younger children failed to generalise in the test phase. Hence, it is likely that the younger children learn a phonological context sensitive rule rather then true allophonic functional categorisation. Interestingly, however, children in both age groups are sensitive to generalisations of voicing over obstruents of different places of articulation. This implies that the rule they hypothesise during the experiment (whatever the rule is specifically) is a rule over *features*, rather then individual words, syllables or segments.

Although this approach to phonological rule learning yields interesting results, it remains a simplification. Above, we noted that the alternation investigated by Jusczyk et al. (2002, see also below) involves a pair of segments that stand in allophonic as well as a contrastive relation to each other. Final Obstruent Devoicing is a phonological process that yields a similar situation, and there are many more. The real-world situation is thus more complicated than sketched in the work of Peperkamp cited above, and in a way that points to the necessity of *more* linguistic knowledge, rather then less.

The general picture that arises from this collection of studies is that phonological rules are encoded in features from the start of their acquisition. With respect to the definition of innateness we adopted, it would seem that as far as allophonic rule-learning is involved, the definition stands.

Phonotactic patterns

After showing that nine month old infants are able to induce generalisations about syllable structure in a laboratory condition, Saffran and Thiessen (2003, experiment 1) continued to investigate whether the same holds for phonotactic patterns (experiment 2). Nine month olds were assigned to one of two groups, both of which were familiarised to CVCCVC words. In the first group, the onsets were voiceless and the codas voiced; in the other group, the pattern was reversed. After familiarisation, the infants were tested with a speech segmentation task: would they be able to separate out the familiar patterns from a continuous speech stream? It turned out that they did; showing a novelty preference (that is, they listened longer to the test stimuli that deviated from the familiarisation pattern). Crucial to our present purposes, Saffran and Thiessen also ran a follow-up experiment using the exact same experimental paradigm, but with different stimuli (Saffran & Thiessen, 2003, experiment 3). In this experiment, the stimuli were constructed so that the only possible pattern that could be induced was based on individual segments, rather then generalised features such as [voice]. In this experiment, nine month old infants failed to discriminate between the two patterns at test. Hence, the results reported in Saffran and Thiessen (2003) indicate that as young as at nine months of age, children use features to learn about phonotactic patterns.[18]

Naturalness and learnability

In a series of two experiments, Seidl and Buckley (2005) set out to test whether children are biased to learn phonetically grounded rules more easily then phonetically arbitrary rules. Seidl and Buckley (2005) employ a version of the Head-turn Preference Paradigm, in which eight month old children are familiarised to sets of strings of words. For one group, the words follow a phonetically grounded rule, for the other, the rule is phonetically arbitrary. Experiments 1 and 2 differ in that the rule in 1 concerns the first consonant in a bisyllabic, trochaic word, whereas in experiment 2 the rule restricts the first CV sequence in words with the same structure. In the first experiment, familiarisation stimuli randomly constructed from a set of segments containing only coronal frica-

[18] Incidentally, this is younger then the age of ten months, at which the native language consonant categories are said to be acquired (Werker and Tees (1984); see also section 2.3.1 above).

tives and affricates, and coronal and labial non-continuants. In the test phase, words containing labial fricatives and affricates, and dorsal non-continuants were added. In this way, it could be tested whether children generalise over the stimuli and analyse them in terms of features, rather then as phonetic images or some similar construct. In experiment 2, the place of articulation of the first consonant and the first vowel was either the same (natural) or different (arbitrary). Again, novel consonants were added to the pool from which the stimuli were generated for the test phase.

Although the children learned the generalisations in both experiments, they did not show a preference for the natural pattern. As Seidl and Buckley (2005) mention, this is not unsurprising from the point of view that phonetically ungrounded rules exist in the world's languages; hence, they must be learnable. However, these experiments go beyond that in two ways: first, they show that (at eight months of age) they are *equally* learnable. Second, they show that children make abstract subsegmental generalisations and apply these to novel stimuli. In other words, children appear to employ features when encoding the rules of the ambient language.

Emergent features in typology

Subsegmental generalisations are very close to, if not the same as, the identification of natural classes. At the same time, features are used to define sets on which rules operate. Thus, features have a double role[19]. This double role captures the observation that rules apply not to individual segments, but rather to natural classes. However, if it can be shown that rules do not follow natural classes, this double role collapses. If rules operate over unnatural sets of segments, we must either abandon the idea that rules apply over sets (instead, then, a set of very similar rules applies to a set of individual segments) or, we must abandon the notion that features denote natural classes. The former case is extremely unappealing as it introduces a host of redundancy and randomness in the theory. The other option implies that features are acquired by analyses over input structures, and thus cannot be innate. This is, in a brief description, the motivation behind Emergent Feature Theory: if theories of innate features fail to capture the structural descriptions of rules, then features must be emergent.

In a large study, Mielke (2004) put this idea to the test. In contrast to the UPSID database,[20] which aims to reflect the genetic relations of language families (and thus counter overrepresentation of any one language group or family), the resulting P-base was compiled opportunistically, by aggregating all language descriptions available to its author. Furthermore, not only inventories

[19] In fact, a triple role, as they also serve to identify contrasts.

[20] The original UPDSID files can be obtained from `http://www.linguistics.ucla.edu/faciliti/sales/software.htm`, a web-interface by Henning Reetz can be found here: `http://web.phonetik.uni-frankfurt.de/upsid_info.html` (both websites last visited 05-08-2014).

are encoded, but also alternations. This leads to a database containing 628 language varieties (549 languages). For each of these language varieties, the 'phonologically active classes' were extracted, whereby 'phonologically active class' is defined as such (Mielke, 2008, p. 49):

Phonologically Active Class A group of sounds within the inventory of a language which to the exclusion of the other members of the inventory

- undergo a phonological process; or
- trigger a phonological process; or
- exemplify a static distributional restriction

Every segment inventory was coded according to three feature theories: *Preliminaries to Speech Analysis* (Jakobson et al., 1952), *The Sound Pattern of English* (Chomsky & Halle, 1968), and Unified Feature Theory (Clements, 1990; Hume, 1994; Clements & Hume, 1995). The result is a set of feature matrices; one per feature theory per language variety. In these matrices, phonologically active classes were plotted. A feature theory is said to be able to characterise a phonologically active class if it is also a natural class according to the following definition (Mielke, 2008, p. 12):

Natural Class (Feature theory-dependent definition)
A group of sounds in an inventory which share one or more distinctive features *within a particular feature theory*, to the exclusion of all other sounds in the inventory

That is to say, the phonologically active class can be described as a conjunction of features, a disjunction of features, or subtraction of features. Then, for each feature theory, it was computed how many of the phonologically active classes were also natural classes in that theory. Of the 6,077 phonologically active classes in the database, the number (and percentage) of natural classes per feature theory are listed below in table 2.6: as we can see, the highest score of an individual feature theory is almost 71% overlap between phonologically active and phonologically natural (within that theory) classes, whereas the highest degree of overlap for *any* feature theory is just over 75%.

According to Mielke, these results indicate that the idea of innate features (or at least the universal features proposed by the three tested theories) cannot account for all phonologically active classes, as there is always a significant proportion of phonologically active classes that is unnatural according to any theory. As an alternative, Mielke proposes that features emerge during acquisition, as the result of generalisations learners make over the sound patterns they encounter. Features, under this view, have an indirect relation to phonetic correlates; they are merely handles to characterise groups of sounds.[21] The distinction between phonologically natural classes and phonologically unnatural

[21] Note that similar ideas have also been proposed by proponents of generative phonology, most notably Hale and Reiss (2008)

classes disappears; in fact, by definition there are no phonologically unnatural classes. Phonetically, the members of a class may be more similar or less similar, but this is of no consequence to the phonological naturalness of the class.

Feature System	characterisable (Natural)		Non-characterisable (Unnatural)	
Preliminaries	3,640	59.90%	2,437	40.10%
SPE	4,313	70.97%	1,764	29.03%
Unified Feature Theory	3,872	63.72%	2,205	36.28%
ANY SYTEM	4,579	75.33%	1,498	24.65%

Table 2.6: Natural Classes in three feature theories (Mielke, 2008, p. 118)

There are a number of points that we can raise against Mielke (2008)'s analysis and conclusion. The first is methodological. The 628 language varieties reflect all descriptive grammars available from the Ohio State University and Michigan State University library systems.[22] Although the majority was published in the seventies, eighties and nineties of the twentieth century, the publication date ranges from 1906 to 2002 (with one outlier even at 1854 (Koelle, 1968[1854], cited in Mielke, 2008)). Needless to say, the referenced grammars were compiled and written by a vast range of authors, all of whom inescapably brought their own perceptions, prejudices, education, and preferences to the act of transcription (in itself an imperfect abstraction) and grammar writing. In other words, the variability in the P-base data sample is of necessity considerable (see also Hall, 2011, §5.1 for an explicit warning about taking phonetic transcriptions at face value). Whether this may account for the number of cases where none of the feature theories could describe the relevant class (39% of all classes) is highly doubtful, but at the same time, it is not unreasonable to suppose that the variety in the source causes some muddiness in the outcome.

Secondly, Mielke (2008) assumes that feature theories apply to the inventory as a fully specified feature matrix. Although it makes no sense to assume that underspecified features may be phonologically active (and thereby constituting a *phonologically active class* (Mielke, 2008, p. 13), phonological activity might determine which features are specified and which remain underspecified, and the scope of their specification (Hall, 2007; Dresher, 2009, among others). As discussed elsewhere in this thesis, the Modified Contrastive Hierarchy proposes that learners arrive at their phonological representations by recursively dividing the phonetic inventory according to binary choices, while applying features to the resulting sets. The main criterion for division is phonological activity: if a group of segments behaves in some specific way to the exclusion of another group, then the two groups must be contrastive, and the learner assigns two values (+ and −) of a feature to the two groups. This is repeated until each member of the inventory has a unique feature specification. Importantly, the feature assignment does not retroactively apply to the segments that have

[22] With the additional restrictions that only grammars written in English were considered.

already been uniquely defined.

With respect to the results in Mielke (2008), this means a number of things. First, although *features* may be universal and substantively innate, their application does not have to be the same in every language (note that nothing in the Modified Contrastive Hierarchy prevents that features are substantively innate; it might very well be the feature's substance that drives the learner in deciding which feature to apply to which subdivision). This greatly undermines the universal feature matrices with which Mielke's study set out. In this way, the problem of ambivalent segments (Mielke, 2008, chapter 4) is also solved (although ambivalence *within* a language remains problematic; consider the status of the high front vowel [i] in Finnish vowel harmony versus (transparent) versus its behaviour in assibilation (trigger)). Finally the fact that Mielke's study concerns a synchronic state of each language should make us not expect a complete overlap between phonological activity and featural naturalness in the first place.

Patterns that are generalisable in terms of features are learned, whereas random patterns are not (Saffran & Thiessen, 2003). On the other hand, the phonetic naturalness of these patterns appears to be of much less concern (Seidl & Buckley, 2005). Phonetically unnatural patterns occur readily in the world's languages, and for Mielke (2008), this is a reason to assume that features cannot be innate – not all phonologically active classes can be defined using feature theories. On the other hand, the criterium for innateness employed by Mielke (2008) is a rather limited one: the ability to account for all phonologically active classes.

The most severe critique of Mielke's argument is that it concerns natural *rules*, rather then natural *classes*. In other words, the Emergent Feature theory considers features only in their role as 'handles' for phonological rules and ignores the other roles we have been discussing in this chapter.[23] We know that during the life cycle of phonological rules, they tend to become less phonetically motivated, and more morphologically conditioned (quote from Hyman (1975, 181f)):

> Although sound changes are sometimes blocked by considerations within a paradigm [...] no corresponding force has been discovered which would strive to keep rules natural. Instead, the above examples show the great tendency for rules to become unnatural [...] that is, to lose their phonetic plausibility and become morphologically conditioned.

It is thus reasonable to ask the question whether Mielke's definition is too narrow to warrant the conclusion that features cannot be innate. Although the underlying motivation for feature theory is that there is such a thing as 'natural class', and the aim for feature theories is or should be to achieve the greatest

[23]Unless we were to adopt a fully substance-free emergent feature set, in which case it would seem that the falsifiability of the Emergent Feature theory becomes problematic.

possible coincidence between phonetically natural, phonologically natural and phonologically active classes, the nature of language change prohibits that that goal will never be reached (see again Hyman, 1975). The fact that the infants in the Saffran and Thiessen (2003) study were unable to induce generalisations based on random groups of segments indicates that contrary to Mielke's predictions, 'crazy classes' are difficult to learn.

2.4 Summary

This chapter began with an overview of the literature on the developing inventory, distinguishing between approaches focusing on the logical problem of language acquisition versus those who emphasise the developmental problem of language acquisition. Next, we considered three major theories about adult inventories. One, Dispersion Theory, is a *bona fide* theory *about* the Inventory, whereas the other two (Parallel Bidirectional Phonetics and Phonology and the contrast and enhancement theory of Dispersedness) view the inventory as epiphenomenal.

Dispersion Theory (DT), at least as it is presented in the work of Flemming (Flemming, 2004, 2006), is antagonistic to feature theory. While this in itself may not be an argument against a theory that does not claim to be a model of subsegmental structure, it is nonetheless important to note that DT is fundamentally phoneme-centric.

Parallel Bidirectional Phonetics and Phonology provides some conceptual advantages over Dispersion Theory. Most notably, it dispenses with the latter's high level of teleology, and unlike it, does not require such a radical re-thinking of the architecture of grammar. Put differently, its ontology is compatible with the standard view of grammar, in which phonology maps output forms on input forms – and vice versa. On the other hand, it appears to require a presupposed inventory, and here, too, it is unclear how it interacts with features. One could conceive of a model in which the Feature Co-occurrence Constraints proposed in the present thesis and elsewhere (Levelt & van Oostendorp, 2007; van 't Veer, 2008) are incorporated in the PBPP model at the level of Structural constraints. It is conceivable that such a combined model would be able to derive both the shape and the structure of the inventory, but we will not pursue this idea further at this point.[24]

In the next section, we saw that the proposal of phonetic enhancement of phonological contrasts introduced by Hall (2011) can account for the shape of segment inventories without teleology (contra DT), without incorporating

[24]One of the possible pitfalls of such a model would be learnability. The PBPP model proposes that the precise phonetic correlates of phonemes – and thus presumably of features – is only learnable when the child is aware of the number and identity of the members of the inventory. On the other hand, the FCC approach proposes that knowledge of features comes to the learner through either innate or emergent features (see next section), but that these pre-exist the emergence of the co-occurrence constraints. Hence, the two appear to put different requirements on the learning algorithm.

gradient phonetics into the grammar (Contra DT and PBPP), and within what can be presumed to be the limits of linguistic competence (contra DT). Whereas the *shape* of the inventory is the result of phonetic enhancement of contrast, the *structure* of the inventory, the contrasting elements in relation to each other, is derived by means of the Successive Division Algorithm (Hall, 2007), an extension of the Modified Contrastive Hierarchy model (Dresher, 2009). The contrast and enhancement theory of dispersedness does not come without its own issues, though; first of all, it is unclear what the ontological status of the Contrastive Hierarchy is. By relying so much on the system-level property of contrast, it appears that MCH is or needs to be holistic in its attitude towards the segment inventory. This is also expressed in the learning algorithm: it only works for a highly idealised one-pass-does-all type of learner.

In the last section, we discussed the innateness of features, based on a definition of innateness proposed by Elizabeth Spelke. We then proceeded to investigate whether distinctive features reach the criterium set by this definition, differentiating between features in their contrastive/categorising, lexical and phonotactic uses. From each of these three perspectives, the concept of innate features cannot be discarded and often, the evidence supports it.

In the present thesis, we propose a model of the segment inventory in which many of these issues are avoided. First of all, the ontological status of the inventory is clear: it has none. Or, to put it more nuanced, it is entirely epiphenomenal. The means by which it is derived, furthermore, are entirely implementable within existing theories of phonology; in fact, in one of the dominant models of the phonological inventory (Optimality Theory), feature co-occurrence constraints have been dormantly present since its inception. Hence, learning is not thought to be categorically different from learning of other language-specific properties of phonology. The theory will be explored in detail in chapter 3, and a demonstration of how it functions in acquisition is provided in chapter 4.

CHAPTER 3

The Final State: Feature Co-occurrence Constraints in the adult grammar

3.1 Introduction

The general claim of this thesis is that the acquisition of the segment inventory can be accurately described with a system of monovalent features and co-occurrence restrictions on those features. It aims to provide insights in phonological acquisition, in feature theory and – to some extent – constraint theory. For any theory of phonological acquisition, it is imperative that it is able to describe the end point of the acquisition process: the adult grammar, or, in more technical terms, the Final State. For our present purposes, the Final State is the segment inventory of Dutch, and in this chapter, we will see how the Feature Co-occurrence Constraint theory accurately and adequately describes the Dutch segment inventory, and what we can learn from that concerning the two major ingredients of the theory: features and constraints.

The theory developed here takes as its central apparatus the Feature Co-occurrence constraint. An FCC is structurally very simple: it refers to no more and no less then two features, and comes in either of two types: c-constraints ban a combination of the two features within one and the same segment, whereas i-constraints demand a feature to be present, given the presence of another feature. Definitions of the two are given below, in standard OT fashion for convenience:

(24) a. *[F, G]
assign a violation mark for every segment Σ iff [F] is dominated by

Σ and [G] is dominated by Σ (*c-constraint*)

b. [F→G]
assign a violation mark for every segment Σ iff [F] is dominated by Σ and [G] is not in Σ(*i-constraint*)

Feature Co-occurrence Constraints in this thesis are defined as segment-based, even though the segment is seen as no more than the coincidence of features under a root node. As we will see later, constraints can be seen as functions that take as their input a phonological representation and output a number: 0 or 1 in non-OT type frameworks (that is to say, violated or not violated), and an integer equal to the number of violations in OT. There are different metrics to decide the outcome of such a constraint-function. In this thesis, violations are assigned to *segments* that contain an illicit structure. This is in line with the proposal in McCarthy (2004). The alternative would be to assign violation marks for every *feature* that co-occurs with an illegal partner. The practical difference between these metrics is negligible for our present purposes; it is manifested mostly in cases where the autosegmental behaviour of features and segments come into play. In the current thesis, we are not concerned with spreading, deletion and epenthesis. Hence, we will assume the segment-based definition as given above.

In this chapter, we will explore the formal and linguistic properties of FCCs and the implications thereof for feature theory, beginning with an illustration of an FCC account of the segment inventory of Dutch. In the next chapter, we will see how Feature Co-occurrence Constraints provide insights in phonological acquisition.

One important aspect of the FCC-approach is that is entails a division of labour in defining the shape and content of the inventory between representational and derivational (grammatical) devices. Both are obligatory ingredients in any phonological theory (Scheer, 2011). The representational side of FCC theory is formalised as the set of (possible) segments, the derivational side is the set of FCCs and their filtering effects. As in OT, this division takes the following shape: the representational device proposes, the derivational device disposes. Many authors have proposed – or tacitly assumed – that the members of the segment inventory are determined by a set of feature co-occurrence constraints (Hayes, 1999; Kager, 1999). Whether, for the current proposal, Optimality Theory is the most suitable framework to be adopted will be discussed in chapter 5. In other frameworks too, the ungrammaticality of certain feature combinations has been formalised by means of feature co-occurrence constraints. In principle, this holds for every theory that adopts a generator-and-filter model of phonology; a model where possible structures are generated by one module, and filtered by another (constraints). The generator is most explicitly defined in Optimality Theory, where we know it as GEN. In many other models, the generator function is performed by rules, where constraints may be active to filter unwanted rule output. The type of generator we assume in the current model is much like GEN, in that once a feature is acquired, it can in principle combine with any other feature. Whether co-occurrence restrictions hold only at the surface level or also underlyingly is a question that pertains directly to the framework in which one chooses to implement the theory. We will come back to this issue in chapter 5.

For the current discussion, suffice it to say that constraints on feature combinations are an effective way to delimit the segment inventory in a number of frameworks, but that a precise discussion of their shape is often lacking (notable exceptions being Hayes & Wilson, 2008; Hayes, 1999), and that, prior to Levelt and van Oostendorp (2007), the possibility of using FCCs in a theory of acquisition was never investigated.

3.1.1 Overview of the chapter

In section 3.2 we will provide a descriptive picture of the Dutch segment inventory. Here, we will be introduced to the feature set that was arrived at to complete the analysis in terms of Feature Co-occurrence Constraints presented in section 3.2.3. The assumptions and consequences involved in defining the constraints are motivated in section 3.3. Next, section 3.4 explores the assumptions and consequences of the chosen feature set.

Before we continue, however, it is important to note that in a theory which characterises acquisition as the simultaneous development of representational and derivational devices, the characteristics of these devices are interlocked. That is to say, the feature set is not trivial. An important benefit that comes from approaching a single subject (the segment inventory) from both the representational (features) side and the derivational (constraints) side, is that we can learn something not just about both, but also how choices on one side have consequences for the other. As an example (this will be discussed in much greater detail below), major class features and feature geometric grouping nodes are not easily combined with Feature Co-occurrence Constraints.

3.2 Dutch

As mentioned above, any theory of acquisition must provide an account of the final state: the adult grammar. A proof-of-concept of the FCC theory is thus a good way to illustrate how the theory works. In this section, we will show how FCCs account for the inventory of Dutch, the language that the children discussed in chapter 4 are acquiring. We will begin with a description of the segment inventory of the language, then show which features are involved. The section ends with a discussion of the constraints that are needed. Using this illustration, the next sections will explore the properties of Feature Co-occurrence Constraints in more detail, as well as discuss the implications for feature theory.

3.2.1 The segment inventory of Dutch

The inventory of Adult (Standard) Dutch, as proposed in Booij (1995), is given in table (3.1). The voiced velar plosive, /ɡ/, is represented in brackets, because it is only marginally present in standard Dutch; that is, it only occurs in loan words. The contrast between voiced and voiceless fricatives, although represented in Booij's analysis, is subject to neutralisation for many speakers (in favor of the voiceless members of the opposition). This neutralisation occurs most often in word-initial position, although for some speakers it applies in intervocalic position too; final obstruent devoicing (FOD) ensures that no voiced obstruents occur word-finally (but see – among others – Ernestus, 2000 and van Oostendorp, 2008 for a nuance of the absoluteness of FOD). The phonemic status of the /x–ɣ/ contrast is defended by Booij on the basis of the phonologically conditioned allomorph selection of the past tense suffix (-tə/-də): compare *vlagde* [vlɑɣdə] 'flagged' with *lachte* [lɑxtə] 'laughed'. It is otherwise absent in the language, and Ernestus and Baayen (2001) argue that the *-te/-de* allomorph selection is not based on the underlying voicing specification, but rather to what they dub 'systematic analogy'; they report an experiment that shows that if the *-te/-de* alternation were due to a rule (or constraint) of phonologically conditioned allomorph selection, that rule is frequently violated. Furthermore, some descriptions of Dutch – such as Gussenhoven (1999) – do not list /ɣ/ in the Dutch consonant inventory, whilst including a voicing contrast for labial and coronal fricatives. There is a general tendency for the voicing contrast to disappear in Standard Dutch fricatives, and the resulting neutralisation is most obvious for the velars. We will leave the matter at that, and for our current purposes, we will not consider /ɣ/ to be part of the core inventory of Dutch. For many speakers, the /s–z/ and /f–v/ contrasts are still phonemic, however, even in word-initial position. Finally, we do not include /h/ in our sample.

	Bilabial	Labiodental	Alveolar	Palatal	Velar
Plosives	p, b		t, d	c	k, (ɡ)
Fricatives		f, v	s, z	ʃ	x
Nasals	m		n		ŋ
Liquids			l, r		
Glides		ʋ		j	

Table 3.1: The consonants of Dutch (after Booij, 1995)

A number of observations must be made. The rhotic in Dutch has a very wide range of phonetic realisations on all three major dimensions: place, manner and voicing. That is to say, it can be coronal, velar or even uvular; it can be a tap, a trill or (in some dialects) a fricative. In addition, post-vocally it is often realised as a retroflex approximant (in one dialect, that of Leiden, the approximant occurs in all positions). Furthermore, Gussenhoven (1999) lists – as marginal – a number of post-alveolar segments: [c, ʃ, ʒ, ɲ]. The first two surface mostly as the result of contraction of a coronal-final stem with the diminutive suffix *-je*: *bootje* [bocə] = /boːt+jə/ 'little boat'; *visje* [vɪʃə] = /vɪs+jə/ 'little fish'. /ʃ/ occurs in select lexical items such as *sjorren* [ʃɔrə] 'to pull, fasten'. /c/ also occurs in the allomorph of the diminutive suffix following coronal nasals: *boontje* [boncə] = /bon+cə/ 'little bean'. /ʒ/ occurs mostly in loan words, such as *garage* [xaraʒə] 'garage', and /ɲ/ is limited to older loans, such as *oranje* [ɔrɑɲə] '(the color) orange'.

3.2.2 The features of Dutch

As we have discussed in chapter 1, only privative features are employed for the current study. The inventory of Dutch, as discussed above, can then be described with the features and feature combinations given in table (3.2).

	p	b	t	d	c	k	f	v	s	z	ʃ	x	m	n	ŋ	l	r	ʋ	j
[continuant]							+	+	+	+	+	+					+	+	+
[nasal]													+	+	+				
[approximant]																		+	+
[liquid]																+	+		
[voice]		+		+				+		+									
[distributed]					+						+								
[labial]	+	+					+	+					+					+	
[dorsal]						+						+			+				

Table 3.2: Feature specifications for Dutch.

It is sometimes proposed that the difference between /l/ and /r/ is one of continuancy, where /r/ is continuant and the status of /l/ differs between languages; I will not go into this here (but see Roca & Johnson, 1999; Gussenhoven & Jacobs, 2005). Sonorants are phonologically voiceless, and nasals are non-continuants. Where it is necessary to refer to /g/ (for example, in loan words), it will be specified like /k/, with the addition of [voice]. The rhotic is assumed to be placeless, coronals are underspecified (see, e.g., Fikkert and Levelt (2008) for a motivation for coronal underspecification in child phonology).

It deserves attention that in this system, there are no major class features: all features are phonetically motivated (with the possible exception of [liquid]). There is no place node, nor a root node feature [sonorant]. One implication of this is that the organising effect of feature geometry (e.g., Clements & Hume, 1995) is not employed; in fact, in section 3.4 we will see that organising nodes are fundamentally at odds with Feature Co-occurrence constraints.

3.2.3 Constraints on the segment inventory of Dutch

The system of nine features yields a large number of possible segments (9! = 362 880), whereas only twenty are actually used in adult Dutch. The number of Feature Co-occurrence Constraints that are active to separate out the illegal segment combinations, however, is limited to 19, out of a possible 108 (1.5•N(N-1)). They are listed below in (48a) (c-constraints) and (48b) (i-constraints).

(25) a. C-CONSTRAINTS
*[lab, dors]
*[lab, dist]
*[lab, liq]
*[voice, dors]
*[voice, dist]
*[voice, nas]
*[voice, liq]
*[voice, apprx]
*[dors, dist]
*[dors, liq]
*[dors, apprx]
*[cont, nas]
*[dist, nas]
*[dist, liq]
*[dist, apprx]
*[nas, liq]
*[nas, apprx]
*[liq, apprx]

b. I-CONSTRAINTS
[apprx]→[cont]

The constraints listed in example (25) above are those that are not violated; every other possible constraint is. From this list, we can make a few observations, which we will discuss below.

Place First, there is a number of c-constraints that ban segments with double articulations: *[lab, dors], *[lab, dist], *[dors, dist]. This is, of course, entirely expected given the feature definitions in (3.2). Furthermore, voiced dorsals and distributed coronals are banned (*[voice, dors], *[voice, dist]). Furthermore, distributed coronals are limited in their distribution in that they only occur in obstruents: *[dist, spread], *[dist, nas], *[dist, liq] and *[dist, apprx].

Manner The features [nasal], [liquid] and [approximant] cannot occur together, as per *[nas, liq], *[nas, apprx] and *[liq, apprx]. Liquids are always placeless: *[lab, liq], *[dors, liq]. Furthermore, there is no dorsal glide (*[dors, apprx]) and nasals are non-continuants (*[cont, nas]). The only i-constraint that is not violated by the inventory of Dutch, [apprx]→[cont], expresses that glides are always continuant, or, as stated explicitly by Itô et al. (1995), glides are non-contrastively continuant. This constraint also illustrates the necessity of i-constraints in a system with only monovalent features: there is no element that expresses non-continuancy. If there were, for example, a feature [-continuant] or [stop], the constraint [apprx]→[cont] could be replaced by a constraint such as *[apprx, stop].

Laryngeal The feature [voice] does not occur with any of the manner features, barring [continuant]: *[voice, nas], *[voice, liq], *[voice, apprx]. This is commensurate with the observation that sonorants are inherently – rather then contrastively – voiced.

A number of these observations reflect cross-linguistic generalisations, whereas others address co-occurrence restrictions particular to Dutch. It is important to note that there is no formal difference in Feature Co-occurrence Constraint Theory; as we will see in the next chapter, all constraints are posited and ranked by the learner. This seems a desirable feature of the theory: no substance-related generalisations need to be hard-wired in UG. On the other hand, constraints are posited that rule out phonetically implausible constellations, effectively repeating phonetic effects in the grammar. Furthermore, if any of these typological generalisations do indeed reflect hard universals, the theory is unable to express this – at least, that is, by way of Feature Co-occurrence Constraints.

The list in (25) is generated by an automated procedure, ensuring that transparency in the procedure is maximal. In what follows, the steps of the procedure are explained in some detail. It should be noted that the same procedure is used for all constraint lists reported in chapter 4.

The algorithm relies on two tables: one listing the segments present in the inventory, the other indicating for each segment its constituent features in a matrix– essentially as in table (3.2). Based on the features that are active,[1] a list of all possible feature combinations is generated. Every entry on the list represents a possible segment, and hence the task of the Feature Co-occurrence Constraints is to divide the list into two groups: legal segments and illegal segments. The FCCs are also generated based on this feature list. The constraints may be seen as templates, which are populated by features in actual grammars. For each combination of two features, then, one c-constraint and two i-constraints are generated; the difference being that the former is symmetrical, whereas the latter are not.

The next step is to generate a table listing all features as column headers and row headers. Regular table cells are then filled: "1" for legal combinations, "0" for illegal combinations, and "–" for those cells in the diagonal axis where row header and column header are identical. An example of such a table is given in table (3.3). As every "0" indicates an illegal feature combination, we can read the column header and the row header (both features), and list their combinations as a c-constraint banning illegal combinations; in other words, a c-constraint that is not violated. By going over the entire matrix, we can separate

[1]In chapter 4, where the development of the segment inventory is investigated, we will encounter many developmental stages at which not all of the features are active. Naturally, this does not hold for adult Dutch, as we will assume that all features are active. On the matter of the acquisition of features, please see section 2.3 above.

the violated c-constraints from the c-constraints that are not violated.[2]

	[cont]	[nas]	[apprx]	[liq]	[voice]	[dist]	[lab]	[dors]
[cont]	–	0	1	1	1	1	1	1
[nas]	0	–	0	0	0	0	1	1
[apprx]	1	0	–	0	0	0	1	0
[liq]	1	0	0	–	0	0	0	0
[voice]	1	0	0	0	–	0	1	0
[dist]	1	0	0	0	0	–	0	0
[lab]	1	1	1	0	1	0	–	0
[dors]	1	1	0	0	0	0	0	–

Table 3.3: Legal and illegal feature combinations in Dutch

Splicing the list of i-constraints involves slightly more. In contrast to c-constraints, i-constraints ([F]→[G]) deal with feature combinations that are legal, but only in a restricted way: one feature cannot occur without the other. Hence, with respect to table (3.3), we are interested in the cells listing "1". For each of these cells, every possible segment is checked. If [F] (row header) is not in that segment, it is ignored. If [F] is present, the segment is then checked for the presence of [G] (column header). If the latter requirement ([G] is present) is true for all segments that satisfy the former requirement ([F] is present), the i-constraint [F]→[G] is not violated; if there is at least one segment which contains [F] but not [G], F→[G] is violated.

This procedure yields two lists of constraints: violated constraints and satisfied constraints. To ensure whether the constraints do indeed ban every illegal segment and allow all legal segments to surface, a checking procedure is built into the algorithm. This procedure makes visible any overpredicted or underpredicted segment.

The checking procedure works by taking the list of all possible segments, dividing them into a list of segments that are legal by virtue of the constraint set as derived by the procedure described above, and a list of illegal segments. The list of legal segments is then compared to the list of attested segments. Any segment that is unattested but legal is flagged as 'overpredicted', whereas any segment that is attested but illegal is marked 'underpredicted'. The list of possible segments is the same as the one generated in the first step of the algorithm; it is a list of feature bundles.

For every constraint that is unviolated, the two features it refers to are listed (F and G). Every segment in the array of possible segments is then checked for the presence of [F]. If it contains [F], the algorithm proceeds to look for [G] in that same segment. If both are found, the segment is deemed illegal by virtue

[2]The matrix being symmetrical, each c-constraint is generated twice. The excess constraints are naturally discarded.

of the constraint *[FG]. If every segment has been checked, the next constraint (e.g., *[HI]) is selected and the procedure is repeated. This yields lists of legal and illegal segments as determined by c-constraints only, so that it is necessary to check again in order to add the effect of the i-constraints.

Again, the two features in the first unviolated i-constraint listed ([F]→[G]) are extracted. Every possible segment is then checked for the presence of [F], and, if [F] is found, for the presence of [G]. If neither or both of the criteria are met, the segment is deemed legal with respect to [F]→[G], and the algorithm moves to the next segment. However, if [F] is found but not [G], the segment is marked illegal. If every segment has been checked, the algorithm proceeds to the next constraint ([H]→[I]).

Given the inventory of Dutch as given in table (3.1), and the feature specification as given in table (3.2), the algorithm yields the constraints in (25) as unviolated. More importantly, no segments are under- or overpredicted, indicating that the procedure is valid, and that the feature specifications are plausible at the very least. In section 3.4, we will go deeper into the choices that were made to arrive at this particular feature set and feature specifications; we will see that adopting a Feature Co-occurrence Constraint methodology to derive the inventory has some significant implications for feature theory. In this light, it is important to note again that FCCs in one form or another are ubiquitous in the OT literature, but certainly also in other theories. Before turning to a discussion of features, however, it is necessary to further explain and motivate the choice and design of the two types of Feature Co-occurrence Constraint.

3.3 Feature Co-occurrence Constraints

In the previous section, we saw that a system of Feature Co-occurrence Constraints is able to accurately derive the inventory of Dutch. This was accomplished by adopting two constraint templates, and filling the templates with the features active in the inventory. The two types of templates are c-constraints and i-constraints, repeated here:

(26) a. *FG
assign a violation mark for every segment Σ iff [F] is dominated by Σ and [G] is dominated by Σ (*c-constraint*)

b. F→G
assign a violation mark for every segment Σ iff [F] is dominated by Σ and [G] is not in Σ(*i-constraint*)

In this section, we will motivate the choice for these two templates, and discuss some of the implications thereof. We will begin with a discussion of the logical form of c-constraints and i-constraints.

3.3.1 The logical form of FCCs

The two constraint types in (26) are very simple, both referring to exactly two features, and making use of very basic logical operators (NOT, AND and IF-THEN). This makes the system rather transparent and appealing. However, one could conceive of other possible forms for Feature Co-occurrence Constraints, for example, constraints referring to one or three features; or constraints making use of different logical operators. In this section, we will motivate the choice for exactly the two types listed in (26), focusing first on arity (the number of elements the constraint refers to) and next on the connectives.

Constraint arity

In many papers in the OT literature, we find constraints such as *⟨segment⟩ (such as, for example, *g). Given that contemporary phonology sees segments as feature combinations rather then monolithic phonemes, it seems reasonable to assume that such constraints are shorthands for constraints forbidding the co-occurrence of the features that make up the segment. In other words, such *(segment) constraints are shorthands for feature co-occurrence constraints. However, it is often unclear what the exact form of the constraint is. For example, it could be a single constraint banning the co-occurrence of exactly those features that distinctively make up the segment; it could be a conjunction of the type of binary c-constraints proposed here, or something else entirely. In this section, we will discuss constraints that refer to less or more than two features.

Single-feature constraints Another type of constraint that we often encounter is *[F], banning the realisation of a single feature. Often, this is related to positional restrictions on the distribution of features. Single feature constraints appear to have no place outside positional restrictions, under the assumption that a grammar makes use of only those features that have been acquired or activated due to positive evidence – that is, the grammar can only refer to those features that are active (if even inactive features are accessible to the grammar, we might want to use single feature constraints to prevent them from being realised).

In the current proposal, single feature constraints have no place: constraints are bound to strict binary reference. This is not to say that we cannot harness the effects of single feature constraints, as these effects can be derived from a special kind of c-constraint: *FF. This constraint punishes any segment that contains [F] and that contains [F]; in other words, any occurrence of [F] period. Conversely, although a single feature i-constraint is conceivable, it is not useful in any sense: F→F is always, under any interpretation, satisfied: it is violated only by segments that contains [F], but does not contain [F], which is a logical impossibility.

In an OT-style grammar, another motivation for single feature constraints might be in interaction with other constraints. Examples of such interactions are given in (27)

(27) a. i-constraint outranking single feature constraint

Input: /[F], [G]/	F→G	*FF	*GG
a. F	∗!	∗	
b. ☞ G			∗
c. FG		∗!	∗

b. Anti-antecedent constraint ranked highest

Input: /[F], [G]/	*FF	F→G	*GG
a. F	∗!	∗	
b. ☞ G			∗
c. FG	∗!		∗

c. Anti-consequent constraint ranked highest

Input: /[F], [G]/	*GG	F→G	*FF
a. ☞ F		∗	∗
b. G	∗!		
c. FG	∗!		∗

d. i-constraint > anti-consequent > anti-antecedent

Input: /[F], [G]/	F→G	*GG	*FF
a. F	∗!		∗
b. ☞ G		∗	
c. FG		∗!	∗

The i-constraint in the examples above is the only constraint which is violated by a single candidate, and hence it is only decisive if ranked highest: switching the two lower-ranked constraints in (27b) and (27c) makes no difference, because the decision has been made by the first constraint. Looking at all the tableaux in (27), we can see that there is no way that [FG] can win when single feature constraints are active.[3] In the mini-grammars above, where an i-constraint ranks with two single-feature constraints, [FG] is harmonically bound by [G], for the reason that [G] violates a proper subset (*[GG]) of the

[3]Not all possible rankings are shown in (27). The two omitted tableau are those where the i-constraint is ranked lowest. It should be clear that here, too, [FG] cannot win and that the optimal candidate is decided by the two single-feature constraints.

set of constraints that are violated by [FG] (*[GG], *[FF]). The segment [F], however, not only violates a subset of c-constraints, but also the i-constraint. Neither *[FF] nor *[GG] can provide a decision about [FG], so it is up to the i-constraint (see (27d)). This constraint removes [F] from the evaluation, leaving [G] and [FG] to compete. However, we have already seen that [FG] is harmonically bound by [G], given the two remaining constraints. Hence, [FG] cannot win.

In combination with single feature constraints, regular c-constraints are superfluous, in that the candidate that they act against (for example, [FG]), is always harmonically bound by simpler candidates ([F], [G]). An example is given in (28).

(28) Regular and single-feature c-constraint: harmonic bounding

Input: /[F], [G]/	*FG	*FF	*GG
a. [F]		∗!	
b. ☞ [G]			∗
c. [FG]	∗!	∗	∗

The segment [FG] violates every constraint, leaving it unnecessary to consider other the other possible rankings given this specific mini-excerpt of CON. Clearly, [FG] can never win.

In closing, under the assumption that features are only activated in the grammar after the child has acquired it based on positive evidence,[4] single-feature c-constraints act to ban, quite surprisingly perhaps, more complex segments, for the reason that simpler segments harmonically bind more complex ones. They have no place in the regular inventory, but may be employed to express positional markedness effects. Single-feature constraints can be derived, but single-feature i-constraints are always satisfied; single-feature c-constraints may yield specific effects.

Multi-feature constraints In addition to single-feature constraints, one could imagine a system of FCCs where each constraint can refer to three or more features. This is undesirable for at least two reasons: first, the restrictiveness of the theory is compromised, and secondly, there is very little evidence that cognitive systems (such as language) ever count beyond two, and there are good reasons to assume that phonology does not compute recursively. This latter reason deserves discussion to the extent impossible to fit in the current thesis.

The more features a constraint can refer to, the more specific the situation (segment) it can ban. Hence, the more features a constraint can refer to, the more powerful the grammar, the less restrictive and hence the less predictive

[4] An alternative assumption would be for all features to be active in the grammar from the first stages, but that their expression is inhibited by single-feature c-constraints. The issue of the innateness of features is discussed in more detail in 2.3.

the theory. Take, for example, an artificial language with three features: F, G, and H. All possible segments are listed below:

(29) /F/
/G/
/H/
/F,G/
/F,H/
/G,H/
/F,G,H/

In a grammar that allows for only two features per constraint, it is impossible to ban the maximally complex segment /F,G,H/, if the combinations of subsets of these features are not illegal. In other words, if /F,G/ and /G,H/ and /F,H/ are legal segments in the language, there is no way of banning /F,G,H/. This is because *[F, G] is violated by /F,G/, *[G, H] by /G,H/ and *[F, H] by /F,H/. I-constraints cannot save the day because there is no implication that is not violated: F→G is violated by /FH/, [G→F] by /GH/, and so forth. If every possible subset of features constitutes a legal segment, then so must the full set, and conversely, a more complex segment entails the presence of its subset segments. These implications of complexity relate to the issue of overprediction, where a segment cannot be banned from the inventory even though it is not attested. A more detailed exploration of the contexts in which overpredictions occur is given in section 4.5.

This situation is compromised as soon as we allow for ternary Feature Co-occurrence Constraints. Now, we can use the constraint *[F,G,H] which is not violated by either /FG/ or /GH/, nor by /FH/, but, when ranked high enough, it will rule out /FGH/.

We have seen that constraints referring to a single feature outright ban its expression (potentially mitigated, in OT and other constraint ranking systems, by the interactions between ranked constraints). Also, we have seen that if the effects of single-feature constraints are needed, they can be modeled by n-ary c-constraints, where each position refers to the same feature. Hence, the most restrictive theory only allows for a minimum of features per constraint, but at the same time, the minimum must exceed one. Two is then the only option.[5]

Let us now briefly consider the formal argument against multiple-feature constraints. Following Samek-Lodovici and Prince (1999), and in the spirit of the algorithm discussed above, we can regard constraints as functions taking an input set of candidates and yielding two output sets: legal and illegal candidates (where the union of the output sets is equivalent to the input set). The question

[5]It should be noted with respect to multiple-feature constraints that similar results can be obtained with constraint conjunction. We will not go into the arguments concerning conjunction here, other then noting that allowing it greatly diminishes the restrictiveness of the theory.

what a possible constraint is, can then be rephrased as the question what a possible function is.

So far, we have treated the logical representation of the Feature Co-occurrence Constraints in a more or less sloppy way, which was sufficient for our purposes. A more precise definition is give in (30):

(30) Definitions of FCCs, where Φ is the set of features such that {f, g, ...} $\in \Phi$, and $D(x, h)$ is a predicate such that x is a root node, $h \in \Phi$, and x dominates h, and α, β are variables ranging over Φ

a. c-constraint
$\lambda\alpha\lambda\beta.\forall \text{x}\ \neg(D(x,\alpha) \wedge D(x,\beta))$

b. i-constraint
$\lambda\alpha\lambda\beta.\forall \text{x}\ \neg(D(x,\alpha) \wedge \neg D(x,\beta))$

The point here is that the functions can be seen as a function of f and g. Crucially, it is posited that there is no recursion in phonological functions. In other words, both f and g are of the same type, namely, elements of the set Φ. Constraints referring to more than two features require that recursion be allowed: instead of g being in Φ, it could also be another function.

Logical connectives in FCCs

Given that Feature Co-occurrence Constraints refer to exactly two features, we must now ask ourselves in what manner this is formalised. In other words, what is the relation between the features that is banned? Logically, two connectives are available: $\wedge$ (AND) and $\vee$ (OR), both of which can be negated ($\neg$). In addition, negation can take scope over one of the members only. Since constraints are restrictions, they must be formulated negatively. In the current proposal, only two logical relations are posited to be necessary (which is empirically borne out): AND and NOT.

As we have seen, c-constraints are formulated negatively, as is common practice in OT and many other frameworks employing formalised constraints: NOT(F AND G). Conversely, i-constraints are formulated as a positive proposition: IF F THEN G. However, the IF-THEN relation is formally equivalent to the negative statement NOT(F AND NOT G):[6]
As we can see in (3.4), although i-constraints seem at first glance to be positive constraints, in their effect, they are not.

C-constraints employ the negation of conjunction: two operands may not co-occur. This is taken to be the most basic expression of what constraints are: negatively formulated restrictions on some structure at some level of representation. A simple AND connective would constitute a positive requirement, see table (3.5).

[6] In the truth tables that follow, the column headers should be read as propositions of the type $\exists$F, $\exists$X & X$\Re$F where F is a feature and X is a root node and $\Re$ expresses dominance, as in example 30 above.

F	G	F→G	¬(F ∧ ¬G)
0	0	1	1
0	1	1	1
1	0	0	0
1	1	1	1

Table 3.4: Truth table for i-constraints

F	G	F∧G	¬(F ∧ G)
0	0	0	1
0	1	0	1
1	0	0	1
1	1	1	0

Table 3.5: Truth table for c-constraints

In tables (3.4) and (3.5), 1 means that the constraint is satisfied, whereas 0 indicates a violation. The trouble with positive requirements becomes apparent when we look at the first row of table (3.5): a violation mark would have to be assigned for any candidate that has neither F nor G. On the issues that are introduced with positive constraints in general, see Prince (2007). The negated counterpart of the AND constraint has no such issues. In fact, it is this constraint that is most often implied when feature co-occurrence constraints appear in the literature (Kager, 1999, for example).

A second look at table (3.4) reveals the necessity of i-constraints, given the assumption that features are monovalent (see section 3.4.1 below) in the current proposal. The formulation ¬(F ∧ ¬G) crucially makes use of a negation of a single operand: ¬G. Carried over to feature theory, the only way to express this is by way of negative features, which entails binarity of features. Naturally, this necessity only holds in so far as we need to express the constraint ¬(F ∧ ¬G), which in turn hinges on the question whether we need to able to express ¬G.

There are good reasons to assume that it is not desirable to include negative feature values in our feature system; we will discuss these at some length in section 3.4.1 below. The reason why it is necessary to refer to the negation (in monovalent feature systems interpreted as the absence) of a feature is that it enables us to express asymmetric necessity of co-occurrence. Returning to the list in (25), we see that one i-constraint is active: [apprx]→[cont]. With what we have discussed in this section, we know that this is equivalent to stating *[apprx][-cont], but we also know that [-cont] is not available to us. The crucial point here is that the interpretation of [-cont] is its absence; in other words, *[apprx][-cont] expresses that [approximant] may not co-occur with the complement of [cont], i.e., the absence of [cont].

It is important to note that the interpretation of [-cont] is *not* the presence of all other features: a list of constraints of the form *[apprx][G], where G is every feature in our set except for [cont] or [apprx] would admittedly yield a grammar in which [approximant] must co-occur with [continuant] (if it co-occurs with anything), but it also entails that [approximant] cannot co-occur with any other feature. In other words, in such a grammar, [approximant] can only occur in the segments [apprx], and [apprx, cont]. This is obviously not the desired result. Also, would we not have had i-constraints, there would be no way to ban a segment constituted solely by [approximant], without having to invoke the all-approximant-banning constraint *[apprx][apprx].

In this short discussion, we have made use of an actual constraint, referring to actual features, rather than the more abstract F and G. Again, an alternative to the i-constraint [apprx]→[cont] is *not* to introduce the feature [stop], allowing for the constraint *[apprx][stop]. This is because [stop] is not the complement of [cont]; it may be so materially, but not logically. Put differently, [-cont] and [stop] are not the same thing; [-cont] is the absence of [cont], whereas [stop] is the presence of a thing that happens to be phonetically the exact opposite of [cont].

With the two constraint types that are proposed in this thesis, the logical operators AND and NOT are the only two that are necessary. Of course, many more relations can be described using these operators than the ones employed here. For example, we might ask ourselves the question why, if it is necessary to express the absence of a feature, the statement $\neg(\neg F \wedge \neg G)$ is not a constraint. As we can see in table 3.6, such a constraint would only be violated if neither [F] nor [G] is present. In other words, it expresses the logical relation of inclusive disjunction: a candidate is good if it contains [F], or it contains [G], or it contains both. This entails that it is a *de facto* positive statement, and hence, it is not included as a constraint.

F	G	$\neg F \wedge \neg G$	$\neg(\neg F \wedge \neg G)$
0	0	1	0
0	1	0	1
1	0	0	1
1	1	0	1

Table 3.6: Truth table for c-constraints

For the moment, we will leave it at the stipulation that only AND and NOT are necessary in the constructions that constitute c-constraints and i-constraints. Furthermore, we have seen that both constraint types are necessary: one to express simple co-occurrence restrictions, and the other to express minimal but not exclusive requirements of co-occurrence.

3.3.2 Some implications of FCCs

The violability of FCCs

So far, we have discussed the internal structure of Feature Co-occurrence Constraints, but we must also consider their outward properties: what does it mean for a constraint to be violated? At first glance, this question seems intimately related to the framework to implement FCCs that we envision. This is because violability is minimally permitted in OT, whereas it is categorically prohibited in other frameworks (barring 'family members' of OT, such as Harmonic Grammar (Smolensky & Legendre, 2006)).

It is important to note that in principle, FCCs are not and need not be ranked with regard to each other other than being violated or not. In the words of Tesar and Smolensky (2000), they exist in two strata: each possible constraint is either violated by some candidate, or it is not. It is this property that allows the theory to be applied in frameworks that employ constraints, yet do not crucially rely on a ranking hierarchy. As noted earlier, this includes virtually all phonological frameworks.

In OT, as opposed to other frameworks, constraints can be distributed over multiple strata. The fact that the theory proposed here does not make use of this property of OT makes it translatable to other theories, but at the same time, it must be shown why this is the case. In OT tableaux, candidates are eliminated from the candidate set if they violate a constraint that is not violated by some other candidate, and so on, until a single candidate is left – this is the winning candidate. In the case of the segment inventory, however, there is no single winner, as per definition the inventory is a set of segments. The only way to have multiple winners in OT is by not ranking crucial constraints. A simple example will demonstrate this.

Suppose a language has an inventory of two segments, namely the following: /[F]/, /[F,G]/. Three constraints are relevant to this inventory: *[F,G], [F]→[G], and [G]→[F]. Any ranking over more than two strata (Prince, 2002) yields an inventory of just one segment:

(31) OT tableaux with complete ranking

a. Ranking: [F]→[G] ≫ *[F,G] ≫ [G]→[F]

input: F, G	[F]→[G]	*[F,G]	[G]→[F]
a. /F/	∗!		
b. ☞ /G/			∗
c. /F,G/		∗!	

b. Ranking: [F]→[G] ≫ [G]→[F] ≫ *[F,G]

input: F, G		[F]→[G]	[G]→[F]	*[F,G]
a.	/F/	∗!		
b.	/G/		∗!	
c. ☞	/F,G/			∗

c. Ranking: [G]→[F] ≫ *[F,G] ≫ [F]→[G]

input: F, G		[G]→[F]	*[F,G]	[F]→[G]
a. ☞	/F/			∗
b.	/G/	∗!		
c.	/F,G/		∗!	

As we can see, the winner is solely decided by the lowest constraint, the order of the higher constraints is irrelevant. In order to generate the inventory of our imaginary language, then, we need to avoid decision making between the lower two constraints, yielding a two-stratum grammar (in effect, combining 31a and 31b to generate both winners, or, alternatively, to eliminate only the single loser):

(32) Two-strata tableau

input: F, G		[F]→[G]	*[F,G]	[G]→[F]
a.	/F/	∗!		
b. ☞	/G/			∗
c. ☞	/F,G/		∗	

With this in mind, we start to see that learning in OT, that is, constraint demotion (Tesar & Smolensky, 2000), yields a two-stratum grammar. In the initial state, all constraints are ranked in the same statum (or, in some versions of constraint demotion, ranked randomly),[7] meaning, all candidates are equally good or bad. Faced with evidence to the contrary (i.e., a specific candidate is actually encountered in the surrounding language), the child demotes every constraint that acts against that winning candidate, to ensure that is is a winning candidate indeed. The next step would normally be to repeat this process to differentiate between the demoted constraints, such that a single winning candidate is left. This is blocked, in our case, since more than one winner makes up the inventory.

Constraint interaction and FCCs

In Optimality Theory, constraints can be violated by a candidate output form if this is the price to pay for that candidate satisfying a higher ranked constraint.

[7] In some versions of constraint demotion, markedness constraints outrank faithfulness constraints. For our present purposes, this is of no immediate relevance since we are concerned exclusively with markedness constraints.

Violations in and of themselves, however, are hard and discrete (contra Harmonic Grammar, where weaker constraints can 'gang up' in order to overrule a stronger constraint, by adding up their combined weight). In other words, it is not the case that a constraint can be violated to a greater or lesser degree (see also McCarthy, 2003 for an argument against gradient constraints), but the impact of the violation depends on the place in the ranking of constraints.

An important but often neglected concept of Optimality Theory, designed to restrict the number of constraints in CON, is the principle of the Factorial Typology. The Factorial Typology states that every permutation of CON should coincide with the grammar of a (possible) human language. This principle poses an important barrier for ad hoc constraints, as it is meaningless to state that constraint X is only active in language Y, or, alternatively, that it is ranked so low as to be inactive in every language *except* in Y. Factorial Typology entails that for a CON containing n constraints, the number of permutations is $n!$. Importantly, however, this does not mean that the number of predicted grammars must be $n!$. This is because not every constraint interacts with every other constraint. In other words, quite a number of the rankings derived in the Factorial Typology are functionally equivalent, meaning that they correspond to one and the same grammar, or, at least, yield the same output.

If we want to assess the violability of FCCs within an Optimality Theory-type grammar, the crucial question is with what constraints the FCCs interact. FCCs are violated by feature combinations, and hence it is reasonable to assume that interacting constraints refer to the same level in the phonological hierarchy (that is, the sub-segmental level). An example where FCCs interact with other constraints is Linke and van Oostendorp (in preparation), who analyse Turkish vowel harmony in terms of Feature Co-occurrence Constraints in interaction with specific faithfulness constraints.

To conclude, constraints in Feature Co-occurrence Constraint theory are either violated, or not. No crucial use is made to the type of graded violability that we see in, for example, Optimality Theory – at least, not when considering only the inventory, the subject of the current thesis.[8]

3.3.3 The candidates

A problem for a theory in which things are stated in terms of what is not allowed, apart from the subset problem (see section 4.6), is that what is not allowed is *a lot*. It must be decided how much exclusion is to be done through the formal mechanism proposed by the theory (here: Feature Co-occurrence Constraints), and how much exclusion we want to do by assumption. The domain of the theory must be stipulated, as must the data. More specific to our present concerns, we must define what a possible segment is. The definition adopted here is the most neutral definition available: any combination of features in the feature system is a possible segment, and it is up to the set of active Feature

[8] The issue of constraint ranking will be discussed in more detail in chapter 5.

Co-occurrence Constraints to sift the grammatical from the ungrammatical feature combinations.

In the Optimality Theoretical model attempted in van ‘t Veer (2008),[9] the candidates are represented as strings of all segments in the language.[10] The current model takes over this practice. In chapter 2 above, however, Dispersion Theory was criticised for (ab-)using Optimality Theory to compare languages, rather than individual candidates within a single language. In other words, in Dispersion Theory, OT is no longer considered a model of linguistic competence. This leaves our current position to be defended.

The string of segments used in van ‘t Veer (2008) is taken to be a compression of all the words in the language, and the candidates are taken to be a pseudo-random subset of candidates generated by GEN (candidates are matched for inventory size). The reason why this simplification is justified in the FCC model and not in Dispersion Theory, is in the constraint set. The FCC constraints penalise the simultaneous realisation of phonological features in individual segments. Each of these segments will be evaluated by the grammar at some point, because each segment is contained in at least one output form (by definition; if it were not, it would not be part of the language). The segments in the string are not compared to each other; the evaluation for any segment would not be different if it were presented in a word, or in the string of segments used as a shorthand for all words, or in isolation. It would violate the exact same constraints. In Dispersion Theory, however, what is evaluated is not individual segments, but rather phonemes *in relation to other phonemes*. A vowel, say /a/, will violate a different set of constraints when it is in the same candidate as /ɑ/, than when it is presented in conjunction with /i/ (in particular, the former pair would not fare well on MinDist constraints). In contrast, /a/ will always violate the same Feature Co-occurrence Constraints, regardless of what other segments are present in the inventory (given the same featural representation). Thus, in Dispersion Theory, the evaluation does not concern a *set* of segments, but rather a *system* of segments; it evaluates contrasts. As we have already discussed, it is highly unlikely that such evaluation belongs to the realm of human linguistic competence (see also Hall, 2011; Boersma, 1998). In the OT incarnation of FCC theory, what is evaluated is merely combinations of features.

3.4 Features

Although the idea of FCCs governing the shape of the inventory is not new, implementing it in such a strict way as done in the present work brings with

[9] Due to technical issues, the model was never fully run successfully. As such, van ‘t Veer (2008) should be seen as a proof-of-concept rather than a full-fledged model.

[10] Only candidates with equal sized inventories were candidates in any evaluation, to avoid the interference of harmonic bounding. Inventory size was considered an independent, possibly diachronic factor.

it some implications for feature theory. In this section, we will discuss the assumptions and implications concerning sub segmental representation in detail. We will begin with feature valency, or, the issue of privativity versus binarity. We have already seen that this issue is closely related to the types of constraint in our theory: i-constraints are necessary because features are monovalent, an assumption that is well motivated. Secondly, we will consider the degree of feature specification. Some degree of underspecification also follows from the effects of FCCs as developed here, specifically, that empty segments cannot be banned. Following an important body of work in the phonological literature, we will assume that there is no feature [coronal]. From this specific type of underspecification, which is closer to the notion of underspecification in the Featurally Underspecified Lexicon (Lahiri & Reetz, 2002) than to the notion in work in Radical- or Contrastive Underspecification, we will move to the redundancy of major class features. The status of these features in main stream phonology is somewhat unclear, as pointed out in Padgett (2002); in a system in which the inventory is regulated by Feature Co-occurrence Constraints, the inclusion of major classes as features on a par with terminal nodes in a feature geometric model is shown to have disruptive effects. We will briefly consider an existing alternative to feature geometry known as Feature Classes (Padgett, 2002) in section 3.4.4.

3.4.1 Privativity

Early feature theory (Jakobson, 1941/1968, for example) proposes an equivalence between a feature and its negative counterpart. This practice persists in modern phonology, with perhaps its most sophisticated exposition being the theory known as the Modified Contrastive Hierarchy (e.g., Dresher, 2009; Hall, 2007). In MCH, features are binary, but they are only specified in the feature combinations in which they are contrastive. This system is derived by successively dividing the inventory in two, such that one subset is assigned [+F], and the other [-F], until each segment has a unique specification.

In the current thesis, all features are assumed to be monovalent. A direct comparison of binary and privative systems is complicated, as demonstrated by Harris and Lindsey (1995), among other things because the same phenomena receive a different treatment under the two systems, not only in terms of representation, but also in the choice of operation. Consider, for example, vowel height harmony.[11] A lowering harmony system, such as we find in Chichewa (see Harris & Lindsey, 1995 for examples and references for both the Chichewa and Pasiego Spanish cases), can be analysed by spreading in both accounts ([-high] in the binary version, |A| in Harris and Lindsey's element-based analysis). Raising harmony (as in Pasiego Spanish), conversely, receives a spreading analysis of [+high] in binary feature theory, but instead a delinking analysis

[11] The examples cited here are from Harris and Lindsey (1995), and following them, we shall adopt elements as representatives of the privative assumption. See Backley (2011) for a concise introduction to Element Theory.

- driven by the lack of a licensor for |A| - in the privative, elemental account. A fundamental difference between the representational system adopted in the current thesis and the one in Harris and Lindsey (1995) is that we do not subscribe to elements. With respect to the height harmony cases mentioned above, a privative feature account is forced to adopt a feature [high] and [low]; even withstanding the harmony cases, such a duplet is the only way to arrive at contrastive mid vowels (which are neither [high] or [low]). However, the main point stands: comparison on empirical grounds is complicated by a number of external factors.

A number of conceptual arguments can be brought forth, however: first, binary feature systems contain a certain amount of derivational redundancy. For example, feature deletion processes can either be established by the deletion of feature [+α], or by deletion and subsequent insertion of [-α]. Often, it cannot be decided which of the two has taken place (Harris, 2009). In a monovalent feature system, no such redundancy exists.

Next, including a negative counterpart for every feature doubles the featural inventory. Let us return to the example of vowel inventories, where it can be shown that the count between monovalent and binary feature systems is the same: [high] versus [high] and [low].[12] Here, the privative account has no economic advantage, but for the binary approach to mitigate the numerical disadvantage, it would have to be shown for each individual feature that under a privative approach, two (opposing) features would be necessary. Exactly such a discussion has been ongoing with respect to the valency of the feature [voice]. Whereas the feature [voice] is usually treated as privative, Wetzels and Mascaró (2001), among others, argue that it should be treated as binary. The arguments in this case refer to the purported phonological activity of [-voice], and we will come back to them in more detail below. Importantly however, the negative counterparts of place features, for example, are often absent from analyses.

With respect to place features, complementary classes are rarely to be found. Rice and Avery (2004), for instance, propose that in their model based on the Contrastive Hierarchy (Dresher, 2009), there should be a feature [peripheral], complementary to [coronal] and hence somewhat equivalent to [-coronal]. This feature unites labials and dorsals, and arguments in favour of it rest on natural class behaviour where the latter two pattern together, to the exclusion of coronals. However, it is hard to find convincing arguments in favour of complementary classes for each of the primary place features, leaving the numerical advantage of the privative approach in tact. Furthermore, joint class behaviour of labials and dorsals can be represented by a union of the two individual classes.

From a representational economy perspective, the odds are in favor of the privative approach. An argument for binary features must then show that a) there is evidence for a natural class that is precisely the complement of the class

[12] The fact that we need both [high] and [low], and not a single featural counterpart to |A| is due to the different ways in which features and elements are interpreted.

described by an established feature, and b) there are no compelling reasons to assign to that complementary class a different, positively framed and monovalent feature. For example, whereas there is little difference between [low] and [-high], some of the cases brought forward in favor of [-voice] have been successfully re-analysed referring to [spread glottis] – a feature for which independent phonological and phonetic evidence exists (Honeybone, 2001/2002, 2005).

Features, as a theoretical apparatus, exist mainly for two reasons: first, some groups of segments display behaviour to the exclusion of others, and second, certain properties of segments cause the segment to contrast with other segments in the inventory. Feature binarity does not follow from either of these arguments, nor is it a necessary assumption for either. Leaving aside the issue of how to arrive at the relevant contrasts, a contrast can be expressed by the presence versus the absence of a feature just as effectively as it can be expressed with a binary prefix.

The other motivation for feature specification, phonological activity, is more problematic for binary feature theory. This is because there is no ontological difference between [+F] and [-F]; both are features in their own right. Their distribution may be complementary, but there is no way in which it can be represented that [+F] is somehow more 'feature-like' then [-F]. From this, it follows that both counterparts should describe phonologically active classes of segments. A large body of literature on exactly this issue exists with respect to the feature [voice], as already briefly touched upon above.

The existence of a [-voice] feature hinges – at least in part – on the question whether there is [-voice] spreading. A first example of such a case provided by Wetzels and Mascaró (2001) is Yorkshire English, where as in other dialects of English, no final devoicing exists. Word-final voiced obstruents, however, loose their voicing when followed by a voiceless obstruent (e.g., bed-time → be[tt]ime). Crucially, there is no analogous spreading of [+voice] (white book → *whi[db]ook). Examples involving English are problematic, however, as the language (and many Germanic languages with it) has been described as having a [spread glottis] rather then a [voice] contrast - a point which is phonetically plausible as English stops are often aspirated (Honeybone, 2001/2002, 2005). Hence, additional examples are given from typologically distinct languages (Dutch, French,Ya:thê, Bakairi). It appears that all examples can be re-analysed in privative terms, with the additional condition that [voice] must be licensed from adjacent [voice] (this is reminiscent of the harmony analysis of Harris & Lindsey, 1995 discussed above). Take French, for example, which shows obligatory devoicing of codas in word-internal clusters, if the following onset is voiceless. Rather then analysing this as spreading of [-voice], van Oostendorp (to appear) proposes that codas in French cannot independently license [voice]. The word-final voicing opposition in French exists simply because word-final consonants are not codas but rather onsets of empty nucleus syllables.

As we discussed in the previous section, binary feature systems are unable to express the difference between [-F] and [∅F]. This raises the question whether every segment that is not specified for [+F] should be specified for [-F]. For Dresher (2009), this is clearly not the case: here, the absence of specification for a feature indicates that it is not contrastive in the context of the combination of other features that specify a given segment. Hence, Dresher introduces a tripartite distinction: a feature is contrastively present, contrastively absent, or non-contrastively absent. Since the natural class of segments specified for [-F] no longer is the complement of the set [+F], but rather a *relativised* complement, Dresher predicts three natural classes per feature: one in which [+F] segments display some behaviour, one in which [-F] segments jointly act, and one in which both are irrelevant. In other systems, [∅F] must be supplied with a value during derivation.

Before we conclude, a third – hybrid – option must be considered: it is entirely possible to think of a system (in fact, many have assumed such a schema) where some features are privative (e.g., place features) and some are binary (e.g., [continuant]). However, in light of the above, it should be clear that the privative approach has a number of advantages, and should therefore be preferred unless compelling counter-evidence is provided. In addition, it seems rather arbitrary to restrict binarity, once it has been admitted to the system, to certain features only. Considering what has been brought forward in this section, monovalency is the approach adopted in the current proposal.

3.4.2 The non-specification of Coronal and Stop

In a system where the segment inventory is structured by constraints on feature co-occurrence, there is one segment that no feature co-occurrence constraint can ban, for the reason that it vacuously satisfies all constraints.[13] This is the empty segment. In other words, in every system in which the inventory is regulated by feature co-occurrence constraints and no other stipulations, underspecification is a necessary part of the representational toolbox. This, in turn, entails that the empty segments must receive a phonetic interpretation. Hence, a prediction of the Feature Co-occurrence Constraint theory is that every inventory has an empty segment that has a phonetic interpretation. For Dutch, we assume this interpretation is that of /t/.[14]

In the current thesis, there is no feature [coronal], and neither is there a feature [stop].[15] Coronals have a special status in phonological theory, as testified by – among others – the contributions to Paradis and Prunet (1991). Furthermore, in the literature in the Featurally Underspecified Lexicon tra-

[13]A specific i-constraint, $\emptyset \rightarrow$F could be envisioned to ban the empty segment, but this constraint has comes with some additional problems – for one thing, it is a positive constraint (like HavePlace). Secondly, every segment fits the structural description of the constraint, as there is no upper limit to the number of features per segment.

[14]Whether this interpretation is universal or language specific is a matter of another debate.

[15]The reader may have noticed that the feature [distributed], often treated as a coronal dependent in feature geometry, is present in the feature inventory. This is justifiable because no feature geometry is assumed; in fact, as we will see, fundamental issues arise when combining FCCs and feature geometry.

dition (Lahiri & Reetz, 2002, for example), [coronal] is underspecified in the lexicon. This is also the position that is successfully adopted in their treatment of Dutch child phonology by Fikkert and Levelt (2008, see also section 2.3.2 above).

As we have seen in chapter 2, the chief argument for Fikkert and Levelt (2008) to assume coronal underspecification is that coronals are free to occur in all syllable positions in children's productions, even at stages where this clearly does not apply to other places of articulation. In this, they follow the literature on Consonant Harmony that assumes that coronals are underspeficified (Spencer, 1986; Stemberger & Stoel-Gammon, 1991). In a range of psycholinguistic experiments they report on, Lahiri and Reetz (2002) show a clear asymmetry in the degree to which mispronunciations are noted, depending on whether the target sound is coronal or of a different place of articulation. Lahiri and Reetz report on a number of studies with adult participants, but these effects are replicated in different studies with children as well (the reader is referred to chapter 2 for references to and descriptions of these studies).

In his treatment of Consonant Harmony, Rose (2000) argues against coronal underspecification on the grounds that coronal appears to be the trigger of consonant harmony in the development of one of his subjects, Clara. Tellingly, however, the cases he reports on are uniformly instances of right-to-left harmony, where [dorsal] is the target (e.g., *gâteau* 'cake': target [gɑto], actual pronunciation [tæto]; example from Rose (2000, p. 171)). If we adopt the perspective of Fikkert and Levelt (2008), cases like this provide no evidence for an underlying feature [coronal], for two reasons: first, Rose (2000) analyses Consonant Harmony in terms of assimilation, requiring a trigger [coronal] and a target [dorsal], whereas Fikkert and Levelt embed Consonant Harmony in a broader theory of restrictions on output forms in child language. Furthermore, the environment in which this purported coronal assimilation takes place is exactly the same environment in which Fikkert and Levelt (2008) show a general ban on dorsals to be in effect – at least in the Dutch cases they describe – proposing the constraint *[dorsal ("no initial dorsals", p. 242).

An argument against coronal underspecification could be that in some languages, harmony effects hold for features that are traditionally seen as dependents of the feature [coronal] (an example would be Sanskrit, where coronals harmonise for retroflexion). In the current proposal, there is no feature geometry and hence no dependency in the feature geometric sense. Since there is no hierarchical difference between features, the so-called dependent features are on a par with major place features. This might seem problematic at first, as it appears to imply that there is no way to prevent [retroflex] to co-occur with [labial], for example, resulting in an unpronounceable retroflex labial. There are two responses to this apparent problem: first, during acquisition, the relevant Feature Co-occurrence Constraints will emerge (remember that FCC theory is blind to whether restrictions on co-occurrence are universal or language specific), and second, features are labeled for class membership (following Padgett, 2002, see section 3.4.4 below).

Coronal is not the only phonological entity that is underspecified (and hence cannot in fact be dubbed a true phonological entity); the same applies for non-continuancy. Here, too, the literature is full of examples where stops are either tacitly or explicitly underspecified for manner (but see van de Weijer, 1994). In binary feature systems, stops may be specified for [-continuant], but a feature [stop] is a rarity.[16] One reason for this is the observation that stops are the least marked among consonants. In addition, phonological activity of a feature [stop] or [-continuant] is, as far as I am aware of, rare at best.

We conclude that non-specification in the FCC approach as we present it here follows from the fact that the empty segment cannot be banned, and thus must receive an interpretation. One could conceive, however, of a constraint *[], which penalises empty segments. This in itself would be of no avail, as segment such as [continuant] would still be generated, and not filtered out. Under the assumption of coronal underspecification, [continuant] is interpreted as /s/; without underspecification, we would need (for example) an additional feature [place], and a constraint [cont]→[place], enforcing every continuant to be specified for place. This would, however, reintroduce a hierarchical relation between features, and furthermore, it means a segment [place] would be generated. It is entirely unclear what interpretation that segment would receive if no place of articulation is underspecified. No i-constraint can be of help here, as it would ban the other places of articulation.

Summarising, empty segments receive a default interpretation consisting of non-continuancy and coronality, where the choice of these traits is empirically motivated (plus, see Fikkert & Levelt, 2008 for a scenario of how precisely this type of underspecification arises during acquisition). The question then becomes how this should be represented. In the previous section we have defended monovalent features against binary features on different grounds, but it is with respect to underspecification that the binary approach has an initial advantage. The three-way distinction that is possible in binary feature theory allows for underspecification not so much of a feature, but rather of a feature value. This can then be filled in at some point (e.g., at the phonology-phonetics interface) by some mechanism, such as Default Fill-in Rules (example from Archangeli, 1988).

(33) Default fill-in rules in Radical Underspecification

a. Inventory and specifications

	i	e	ɑ	o	u
[high]		-		-	
[low]			+		
[back]				+	+
[voice]					

[16] A possible counterexample being the element |?| in element theory (Backley, 2011; Harris, 1994; Harris & Lindsey, 1995, for example). However, there is no direct translation between elements and features both in phonetic and phonological terms.

b. Default fill-in rules

i. [+low] → [-high]
ii. [+low] → [+back]
iii. [   ] → [-low]
iv. [   ] → [+high]
v. [   ] → [-back]
vi. [   ] → [+voice]

The point is, that it is immediately clear from the representation where underspecification applies.

A similar solution is unavailable to monovalent feature theory. That is to say, it cannot be read off the representation directly why a segment [] should receive a coronal interpretation, and a segment [lab] should not. Whatever mechanism is responsible for assigning the default phonetic interpretation must be sensitive to the *type* of features that are present in the representation of a segment and know when to assign a default interpretation, and when not to. Staying close to the previous example, the mechanism must know that a segment consisting of only [cont] must receive a coronal interpretation, whereas a segment consisting of [lab] must not.[17] Obviously, a metric of specificity or arity is not going to be effective here. This may appear to be a trivial matter, but it is not.

In feature geometry, a rule could be foreseen that triggers the default interpretation when there is no dependent feature present under the [place] node. However, there are reasons to assume that feature geometry may be too rigid an instrument (Padgett, 2002; Yip, 2011), and in the next section, we shall see that the concept of major class features is difficult to unify with the strict implementation of Feature Co-occurrence Constraints as developed in the current thesis. Hence, we need a different formal apparatus to represent the *kind* of class a feature belongs to. For this purpose, we shall adopt a theory known as Feature Classes, first proposed in Padgett (2002) and further elaborated in Yip (2011). We will return to this theory in more detail in section 3.4.4

3.4.3 Major class features

In section 3.2.3 above, at various times references were made to phonological terms that are not present in the feature inventory. Place, manner, laryngeal, sonorant, all of these are not phonological primes in the current proposal. The reason for this is that their inclusion inevitably leads to overprediction of segments; segments that are not attested in Dutch, and yet cannot be excluded by FCCs – at least not in the minimalistic way proposed here.

Consider again the Dutch consonant inventory, reproduced below in table (3.7).

[17] Incidentally, it must also know that the former must not receive a stop interpretation and that the latter must do so.

	Bilabial	Labiodental	Alveolar	Palatal	Velar
Plosives	p, b		t, d	c	k, (g)
Fricatives		f, v	s, z	ʃ	x
Nasals	m		n		ŋ
Liquids			l, r		
Glides		ʋ		j	

Table 3.7: The consonants of Dutch (after Booij, 1995)

A standard way of analysing this inventory would be with the feature set in example (34).

(34) Example feature set:
Son, Cont, Nas, Rhot, Voice, Dist, Lab, Dors

The constraints needed to generate the inventory in table (3.7) are given in (35a). The issue is that this set is unable to filter our all unattested segments; the list of overpredicted segments is given in (35b).

(35) a. Constraint set:
*[lab, dors]
*[lab, dist]
*[lab, rhot]
*[voice, dors]
*[voice, dist]
*[voice, son]
*[voice, nas]
*[voice, rhot]
*[dors, dist]
*[dors, rhot]
*[cont, nas]
*[dist, son]
*[dist, nas]
*[dist, rhot]
*[nas, rhot]
[dist]→[cont]
[nas]→[son]
[rhot]→[cont]
[rhot]→[son]

b. Overgenerated segments
[lab, son]
[dors, son]
[dors, cont, son]

The three overpredicted segments have in common that they all include the feature [sonorant], and that they cannot be excluded because they are either a subset of an attested segment (the first two) or because they are the subset of the union of two attested segments (the third). Overpredictions will be discussed in more detail in section 4.5.

The underlying reason why major class features, as well as feature geometric grouping nodes are inherently incompatible with the minimalistic system of FCCs proposed here, is that they are redundant in specifying unique segments.[18] Every major class feature is further refined; sonorants are either liquids, nasals or glides; [laryngeal] breaks down into [voice], [spread glottis], [constricted glottis], [place] breaks down into the primary PoAs. Not including grouping features in the feature set amounts to acknowledging that they have always had an ontologically somewhat unclear status; is [place] a feature? or is it something else? If so, how is the difference expressed? In non-linear phonology, these entities are treated in the same fashion as terminal nodes, in that they occupy the same position in rules. This is motivated by the desire to capture in phonological rules and representations the fact that segments often display group behaviour that transcends classes defined by terminal nodes. Furthermore, certain features, like [coronal], behave like terminal nodes in some sense, but act as grouping features in that they can have dependent features.

To avoid the problems of overgeneration posed by major class features, and to avoid the problems of having to decide which features are grouping features, and what that would entail for the status of that feature, the current proposal rejects the idea of Feature Geometry. It should be noted that this is not a unique position. Element theory and dependency phonology, for example, do not employ feature geometry. Furthermore, it is not entirely clear if and how the Modified Contrastive Hierarchy (Dresher, 2009) can be unified with Feature Geometry. The proposal in Nevins (2010) is an outright rejection of Feature Geometry. At the same time, the observation that segments tend to behave in classes of various levels of generality still stands; we would be throwing out the baby with the bath water if we loose the ability to capture this fact.

3.4.4 Theoretical precursor: Feature Classes

The rejection of major class nodes/features and the adoption of an explicit system of Feature Co-occurrence Constraints is reminiscent of earlier work on Feature Classes by Padgett (2002) and Yip (2011). Padgett rejects feature geometry and develops a system of Feature Classes motivated by partial class behaviour of segments. Padgett (2002) notes that Feature Geometry is right in attempting to capture class behaviour, but at the same time, it is too rigid in forcing features to fit into classes defined by single class nodes. Take, for example, the feature [back], which contrastively denotes back vowels. There are con-

[18]This is, of course, reminiscent of the observation that major class features are never contrastive.

vincing phonetic (e.g., second formant enhancement effects) and phonological reasons to assume that it forms a class with [round]. Much of the phonological evidence comes from vowel harmony in Turkic, Altaic and Uralic languages, many of which display both backness and roundness harmony, which occurs jointly (in the sense that suffixes that are exceptions to backness harmony are also exceptions to roundness harmony, and vice versa). Padgett takes this, in standard autosegmental fashion, to be evidence for a class consisting of [back] and [round], which he dubs COLOR.

Having established the class, Padgett proposes that rather than two different rules or constraints driving roundness and backness harmony, there is a single mechanism referring to COLOR. So far, the account is compatible with Feature Geometry, but the picture becomes complicated when we take into account that roundness harmony in Turkish, for example, is limited to high vowels. Rather then retreating from the position that COLOR is a class, Padgett proposes that rules and/or constraints in the theory he coins Feature Classes refer to individual features, rather than to class nodes. Hence, class nodes cease to exist, and are replaced by set-theoretic classes: COLOR = {[back], [round]}, PLACE = {[labial], [dorsal]}, et cetera. The fact that non-high vowels fail to harmonise on roundness is captured in an OT grammar by a high-ranking feature co-occurrence constraint *[-high][+round]. Summarising, although harmony drives both [back] and [round] by referring directly to them jointly, via the COLOR class, it fails to apply because non-high round vowels are ruled out by an FCC.[19] More generally, Padgett shows that Feature Class Theory is able to account for partial class behaviour.

Other evidence that features are targeted directly even when behaving as a class exists, too. Padgett mentions dissimilation, which, when seen as an OCP-effect, is problematic for Feature Geometry without extra stipulations. Although it is entirely possible to posit a rule or constraint OCP(PLACE), such a mechanism, when taken seriously, would ban sequences such as in (36):

(36) Unwanted OCP(PLACE)-effects

*root root
| |
PLACE PLACE

In other words, adjacent roots (either absolutely adjacent or on some tier) specified for place are banned. This is obviously not the desired result. The goal of OCP(PLACE) is to rule out identical *features* to the degree that both belong to the class of place features – not identical classes.[20]

---

[19] It should be noted that Turkish allows for non-high round vowels, but only in roots; that is to say, not as the result of vowel harmony. Hence, the analysis in Padgett (2002) also involves a faithfulness constraint.

[20] Yip (2011) applies this framework to the question of the representation of laterals, which do not show typologically uniform behaviour with respect to continuance, but also with

Partial class behaviour is not something we are particularly concerned with in the current thesis, but there are at least two aspects of Feature Class Theory that make it an appealing companion for the Feature Co-occurrence Constraint Theory. First, it rejects the notion of overarching class nodes and/or features. All sub-segmental entities are created equal and all are targeted by rules or constraints individually. This resonates well with the finding that major class features and grouping features are incompatible with our Feature Co-occurrence Constraint Theory. The other striking point is that Feature Class Theory, in order to express the things that set it apart from traditional Feature Geometry, relies heavily on feature co-occurrence constraints.

This raises the question of how Feature Classes are to be formalised. Padgett (2002) proposes a set-theoretic set of definitions, such as in (37) (adapted from Padgett (2002, exx. 24, 25)).

(37) a. Simplex classes
Nasal $=_{def}$ {[nasal]}
Voice $=_{def}$ {[voice]}
Dorsal $=_{def}$ {[dorsal]}
continuant $=_{def}$ {[continuant]}
b. Complex classes
Laryngeal $=_{def}$ {[voice] ∪ [constricted glottis] ∪ [spread glottis] }
Color $=_{def}$ {[back] ∪ [round] }
Place $=_{def}$ {[labial] ∪ [distributed] ∪ [dorsal] }

Although this formalisation is logically adequate, it opens the question as to the ontological status of the list of definitions. That is to say, where in the mental representation of speaker and hearer does it reside? Therefore, I propose to go one step further and rather than assigning features to classes, to assign classes to features: features are 'labeled' or 'indexed' for the classes to which they belong. This results in a (non-exhaustive) list as in (38):

(38) [dorsal]$_{place}$
[voice]$_{laryngeal}$
[back]$_{colour}$

Although formally equivalent to Padgett's proposal, this fashion of representation immediately displays the classes to which a feature belongs.[21]

Earlier a concern was expressed regarding the indexing of features that is implicit in Nevins (2010). It should be noted that the indexes proposed here

respect to sonorancy and coronality. It is possible, technically, that the constraint in 36 is present in the grammar, but must be dominated. This is problematic in two ways: first, it runs counter to the notion of Factorial Typologies, and second, unless constraints are innate, it is hard to see how such a constraint can be learned.

[21] One open question is whether the equivalence between the two definitions extends to the simplex classes. That is to say, I will remain agnostic for the time with respect to the question whether [dorsal]$_{place}$ should really be represented as [dorsal]$_{dorsal,place}$.

refer to inherent properties of features (laryngeality is an important aspect of [voice], for example), rather than a system-dependent and non-substantive label such as 'contrast'. The degree to which this difference is important must be explored further.

A full exploration of this version of Feature Class Theory is far beyond the scope of the current thesis, for which it bears no direct relevance. For example, the precise status of class indices must be investigated; they are not features and hence do not behave like features. The question then arises whether they are mere shorthands for groups of features. Remember that major class features must be discarded; this hints to a rejection of Feature Geometry. Feature Class Theory has the obvious advantage of retaining crucial insights from Feature Geometry, namely class behaviour. Finally, its reliance on feature co-occurrence constraints resonates well with the current proposal.

3.5 Summary

We began this chapter with the observation that any theory of acquisition must provide an account of the Final State. In the case of this thesis, the Final State is the segment inventory of Dutch, and we have introduced the theory of Feature Co-occurrence constraints by applying the system to the inventory of Dutch. Feature Co-occurrence Constraints are extremely minimalistic and simple constraints, which are compatible with many if not all phonological frameworks. Since no language employs the full permutation of its sub-segmental particles, every theory of phonology must account for the restrictions on feature combinations. Constraints have been the instrument of choice for phonologists since long before OT (see Cavirani (2010, 2012) for a historical overview of the phonological constraint).

We have seen that a simple set of two constraint types (c-constraints and i-constraints) is sufficient to describe the inventory, and we have explored the motivation behind the choice of constraint types, their formal properties, and the predictions that follow from that. The inclusion of i-constraints is closely linked to the choice for monovalent features; a choice that was motivated in section 3.4.1. In addition to monovalency, the proposed theory of Feature Co-occurrence Constraints entails a degree of non-specification. Together, these factors drive the system to use a minimal amount of features. Furthermore, Feature Co-occurrence Constraints do not sit well with major class features and grouping features; this is a confirmation of the unclear ontological status of these features. This observation led to the rejection of Feature Geometry and the subsequent adoption of Feature Classes, a framework that retains the insight that segments display class behaviour, yet sidesteps the rigidity of geometric structures.

Having established the theory of Feature Co-occurrence Constraints, the theory will be applied to developing inventories in language acquisition in the next chapter.

CHAPTER 4

Feature Co-occurrence Constraints in Acquisition

4.1 Introduction

In the previous chapter we have explored the Feature Co-occurrence Constraint theory in great detail, and we have seen how the theory enables us to analyse the final state of acquisition of an actual language (Dutch) and what implications follow from the rigorous application of the assumption that the segment inventory is derived solely by a set of feature combinations, and two types of minimalistic feature co-occurrence constraints. While we have seen that features may be innate (section 2.3), in the present chapter we will see that there is little evidence that actual, substance-containing constraints are part of UG (4.6). Rather, we assume that language learners are endowed with constraint templates, which are employed as necessary during acquisition. In this chapter, we will demonstrate the use of Feature Co-occurrence Constraints in language acquisition, specifically the acquisition of the Dutch consonant inventory. We will focus on the system at the level of actual productions of word onsets.

If the actual, feature-referring constraints are not innate but rather constructed from templates and features during acquisition, the question immediately arises at what point in time this happens. The earliest logically possible time is when the second of the two features to which the constraint refers is acquired. The latest logical possibility is anytime after. Of these two options, the first is the only one that can be *a priori* linked to a developmental event; it is hard to predict an event later in acquisition that would trigger the activation of a Feature Co-occurrence Constraint. For this reason, we assume that constraints are actuated no later then when both its features become active in the child's grammar. For every pair of features [F, G], the full set of constraints is activated automatically as soon the second of [F] and [G] is acquired. This is posited to be an automatic process, meaning that *all* constraints involving [F] and [G] are activated, even those whose effects are not seen (these are immediately either demoted or deactivated). Alternatively, we could assume that only those constraints are activated for which the child finds evidence of in the language she is learning. The predictions following from either option with respect to the data are not actually different, but the second option does presuppose more knowledge on the part of the learner: under the first assumption, all possible constraints are evaluated with respect to the data that is in the child's uptake. Under the second option, the learner must evaluate the constraints before they are even activated. Hence, the more logical assumption is to assume that constraint generation is automatic and that constraints for which the child finds no evidence are immediately demoted or deactivated. We will come back to this assumption, especially regarding the question whether constraints are not introduced at a later stage, and see whether it is tenable in section 4.4.

In section 4.2 we will elaborate on the methods that were employed to obtain the data for the current study, and we will discuss some definitional choices that were made (e.g., with respect to developmental stages). Data from one of the children, Jarmo, will serve as an illustration at various points in this section. We shall see that the theory describes his developmental data remarkably well. In section 4.3, we focus on another child, Noortje. Every stage that is specified in her data is treated in detail. At first glance, Noortje's data appear to present some deviations from the theory, but a close examination of these reveals that none of them poses a serious problem. Section 4.4 summarises the main findings, while section 4.5 is devoted to the possible underlying causes for overpredictions, that is, segments that the theory predicts should be present in the inventory at some stage yet remain unattested at that time. Most cases of overpredictions will be found to result from artefacts of data treatment, but some remain. Finally, section 4.7.1 concludes the chapter.

Before we continue, however, let us briefly pause to reflect on the assumptions and predictions of the Feature Co-occurrence Theory.

Most basic assumptions have been noted and motivated in the previous chapter, such as the monovalence of features, the bivalence (and other properties) of Feature Co-occurrence Constraints, the assumption that features may be innate, and the non-specification of coronality and stopness. With respect to the application of the theory to acquisition data, a number of assumptions are added. First of all, we mentioned above that we predict that FCCs are activated at a specific time during development. Furthermore, we will see in section4.6 that we will not assume that the Feature Co-occurrence Constraints are substantively (i.e., specific constraints with individual features) innate. Rather, we will assume that the two types that we use are somehow innate as templates, that learners fill with the features they acquire.

One of the great questions of cognition, one that we will not attempt to solve here, is the question why acquisition proceeds gradually. The child is surrounded by the entire language; what is it that prohibits the entire language from being acquired at once? Or, if we restrict ourselves to the acquisition of the segment inventory, why is there an order of acquisition to begin with? We will discuss a number of theories that have approached the problem of the order of acquisition in the final section of this chapter. Arguments have been proposed based on markedness, frequency and lexical diffusion. The order of acquisition is not our primary concern here, but we must acknowledge that there *is* an order.

Somewhat counterintuitively perhaps, the minimal view of phonology we adopted, which holds that the inventory is epiphenomenal, implies that the inventory is *not* what is acquired. Rather, the child acquires words, which we assume, are analysed in terms of features (contra, for example, Fikkert and Levelt (2008), who propose that featural analysis is not present in the very early stages of development. Whether triggered by frequency, markedness or some other property, the child learns that segments are part of her inventory in a gradual, step-by-step way. She does so, we assume, based solely on positive evidence: only the presence of a segment in the surrounding language can force her to adopt that segment in her inventory. A ban on a given segment (feature combination) cannot follow from lack of exposure to that segment. This is why *all* Feature Co-occurrence Constraints are automatically activated as soon as a feature is acquired; it allows for maximum restrictiveness.

A number of predictions can be made, too, with respect to the application of Feature Co-occurrence Constraint theory acquisition data. For example, in section 3.4, we motivated the non-specifaction of coronality and stopness. This, as was demonstrated, is intimately linked with the use of the constraint types that are part of the current theory, and the ban on positive constraints. The typological prediction is that every inventory has a featurally empty segment; it cannot be banned and hence must receive a phonetic interpretation (*in casu*, /t/). With respect to acquisition, the prediction is that /t/ is the first segment that is acquired (or, more precisely, it is present in the inventory from the first stage on).

A related, more fundamental prediction is a specific instantiation of the Continuity Hypothesis. We do not assume continuity at the surface level, which would entail that the inventory at every stage of development should correspond to an inventory of an existing language (this type of continuity has been shown to exist in the case of the acquisition of syllable types by Levelt, Schiller, and Levelt (1999/2000)). Rather, we predict continuity in the sense that the same type of constraints can be used to model the adult inventory, and the child's inventory alike.

Finally, although it was noted that the current proposal makes no specific predictions with respect to the order of acquisition of features, it does limit the amount of possible acquisition paths by making predictions about possible and impossible feature combinations.

Name	Sex	First session	Last session	Nr. of utterances
Catootje	F	1;10.12	2;7.4	2210
Eva	F	1;4.12	1;11.8	895
Jarmo	M	1;4.18	2;4.1	1544
Noortje	F	1;7.14	2;11.0	1867
Robin	M	1;4;14	2;4.28	2283
Tirza	F	1;6.14	2;6.12	1681
Tom	M	1;0.10	2;2.2	1761

Table 4.1: General information about the selected children

4.2 Methods

The data used to test the theory of Feature Co-occurrence Constraints in development are taken from the CLPF corpus of Dutch child language (Fikkert, 1994; Levelt, 1994). The corpus contains longitudinal recordings of twelve children, all acquiring Dutch. Not all of the children in the CLPF database are represented in the current study, because in order to be able to say something about the consonantal development of the child, we need to have a time-window that does not begin too late, nor end too soon. Seven children are included in the final selection, because they provide enough data and because the recordings took place at a time interval in which we can observe segmental development. The seven children are Catootje, Eva, Jarmo, Noortje, Robin, Tirza and Tom. Table 4.2 gives some general information about the children (based on Table (I) in Levelt (1994)).

The data for the current study was collected from the CLPF database (Fikkert, 1994; Levelt, 1994), as available from PhonBank.[1] The database was loaded and searched in the Phon software application (Rose et al., 2006; Rose & MacWhinney, to appear), using a script designed primarily by Greg Hedlund.[2] In the Phon design structure, the database consists of Projects, Coropora, Sessions and Records. The Project is the entire CLPF database, in which every child represents a Corpus. The Corpora are further subdivided into Sessions, and Sessions into Records. Records contain utterances, consisting of word groups. These are represented on a number of 'Tiers', most important of which for our present purposes is the Actual Tier. This contains the IPA transcription of the utterance; in addition to segmental information, stress and syllabic affiliation and position also are coded.

One of the strongest assets of the Phon interface for phonological databases is its built in search capabilities. The data from the seven children were searched

[1] See `http://childes.psy.cmu.edu/phon/`

[2] A free and open source copy of Phon can be obtained from `http://phon.ling.mun.ca/phontrac/wiki/Downloads`

for each of the members of the inventory of Dutch, as given in table 4.2 (see also section 3.2 on the segments and features of Dutch).

	Bilabial	Labiodental	Alveolar	Palatal	Velar
Plosives	p, b		t, d	c	k, (g)
Fricatives		f, v	s, z	ʃ	x
Nasals	m		n		ŋ
Liquids			l, r		
Glides		ʋ		j	

Table 4.2: The consonants of Dutch (after Booij, 1995)

The results of these searches were tabulated in a spread sheet, after which they were processed to exclude false positives, by the following criteria:

1. First, all single occurrences of any segment in any session were deleted
2. Next, any session that contained precisely two occurrences of a given segment, but were followed by at least two sessions with no occurrence of that segment, were discarded with respect to the segment
3. Then, early occurrences of segments were checked to see over how many lexical items the occurrences were distributed. The number of occurrences was replaced by the number of lexical items, to exclude high hits for segments that only occur in single words.[3]
4. Finally, steps 1 and 2 were repeated.

It may seem somewhat indirect to start the search for features by searching for segments, but there is one compelling reason to do this; had the search been performed for individual features, it would have been very difficult – if not impossible – to find restrictions on feature co-occurrences. To illustrate, looking for [labial] returns all labials, without discrimination for which features they co-occur with (or do not). It teaches nothing about co-occurrences; the returned results will have to broken down further. Conversely, simply looking for all labial consonants allows one to know not only when the first labial occurs, but also how the feature combines in various constellations. This advantage outweighs the obvious fact that looking for all labial segments limits the possibilities to those labial segments that are included as search parameters; i.e., the list of segments given in table (4.2) above. It is highly unlikely that the children systematically produce segments that are not in the Dutch inventory, first of all because they are not in the input, and secondly, because the inventory contains relatively unmarked segments. Systematic substitutions are thus likely to be structure preserving.

[3]If a child were to say *bal* [bɑl] 'ball' 6 times in a session but no other word starting with [b], only 1 [b] was counted, for example.

A number of exceptions were made to this procedure. Most notably, the search for /r/ was broken down into searches for the feature [rhotic][4] and the segment [ʁ] (which, in the Phon system, is not coded for rhoticity – even though it is a frequent phonetic variant of rhotics). This was done to capture all occurrences of /r/ which, in Dutch, has a wide range of surface shapes; from trill to fricative, from apical to uvular, voiced and voiceless (Sebregts, ms.). The reason why in this case, the search for a feature rather then for individual segments is warranted, is fairly straightforward: underlyingly, [rhotic] only describes one segment. Another exception was made for the dorsal fricative, because of variations in the transcription. This segment is sometimes transcribed as uvular, sometimes as velar; in the latter case, sometimes as voiced, sometimes as voiceless. There are no lexical contrasts between these options. Hence, searches were performed for all of {x, χ, ɣ}, which were then pooled into one segment labeled /x/.

After the procedure outlined above, the spread sheets were rearranged to derive Guttman scales (Torgerson, 1963). Guttman scales are a tool to reveal an order in data; originally, they were designed to test for hierarchical order in response patterns for questionnaires, but they have been shown to be suitable to reveal temporal orders, too, for example, in language acquisition research (Barton, 1976; Fikkert & Levelt, 2008; Levelt et al., 1999/2000). The Guttman scales revealed an order of acquisition for *segments*, but in this thesis segments have no other ontological status then as the epiphenomenal surface reflection of feature co-occurrences. An example of such a scale can be seen below in (39):

(39) Guttman Scale for Jarmo's onsets

[4] The features in PhonBank are defined as pre-theoretically as possible, and are somewhat different from the definitions used for our analysis. The feature [rhotic] is not in our feature set, but it is in PhonBank.

Jarmo	k	t	d	p	b	m	ʋ	l	n	f	s	x	r	ʃ
1989-09-22														
1989-10-06														
1989-10-31	2													
1989-11-17	2													
1989-12-01	2	6	7											
1989-12-19	5	8	8	8	2									
1990-01-02	2	9	4	11	4									
1990-01-16	5	9	2	9	5									
1990-01-30	3	9	2	12	6									
1990-02-13	4	7	2	9	6	12								
1990-02-27		2	3	4	8	4								
1990-03-13	10	11	2	8	6	3	2							
1990-03-27	11	9	3	9	9	2								
1990-04-10	7	6		5	2	2	2							
1990-04-24	18	30	4	4	5	3	2	5						
1990-05-08	16	38	8	17	6	7	3	10	5	2				
1990-06-01	24	16	6	10		3	4	5	3	3				
1990-06-12	17	23	6	9	2	4		14		3				
1990-06-26	21	40		15	2	8		12	8		3	3		
1990-07-10	19	32	11	20	8	2	5	9	10		3	4		
1990-07-31	5	15	4	11	6	2		11	4		2	3		
1990-08-13	24	25	13	12	24	14	9	16	16	2	11	3	2	2
1990-09-05	24	10	20	11	12	2	10	22	10	4	11		2	

The Guttman scales list the inventory of each child at each stage. From these segmental inventories, inventories of active distinctive features were derived according to the schema in table 4.3.

	p	b	t	d	c	k	f	v	s	z	ʃ	x	m	n	ŋ	l	r	ʋ	j
[continuant]							+	+	+	+	+	+					+	+	+
[nasal]													+	+	+				
[approximant]																+	+	+	+
[liquid]																+	+		
[voice]		+		+				+		+									
[distributed]					+						+								
[labial]	+	+					+	+					+					+	
[dorsal]						+						+			+				

Table 4.3: Feature specifications for Dutch.

Before turning to the procedure with which the emergence and demotion of the individual Feature Co-occurrence Constraints is determined, let us take a brief moment to discuss the way in which the notion ‘stage’ is used in the current thesis.

4.2.1 Stages

Although language acquisition is a gradient process which proceeds with rushes and delays, it is common practice in language acquisition research to segment the time during which a given phenomenon develops into segments, usually called 'stages'. Two general reasons underly this practice: first, it is extremely difficult if not impossible to collect continuous data, as it implies continuously collecting data. Although diary studies (such as Preyer (1895); Röttger (1931); Leopold (1939 – 1949)) can be seen as approximations of continuous data, these are often limited to data from a single child. Therefore, data is often collected at regular intervals; examples of results of this are the CLPF corpus used in the current study (Fikkert, 1994; Levelt, 1994), or the Goad/Rose corpus of Canadian French (Rose, 2000). The second reason is that it is unclear whether continuous data yields dramatically more insights then interval data, especially if the intervals are short enough.

Even if data are collected at regular intervals, these data points do not automatically coincide with developmental stages. It might be the case that a child's progress at phonological property A stagnates for some months to then rapidly develop to the adult form. In such a case it makes no sense to segment the first period into several stages; at least not with respect to property A. This example also shows that there is no *a priori* need for stages of different aspects of grammatical acquisition to coincide.

There are several ways to define stages. One such way is developed in Ingram (1989). Ingram proposes to segment development into three types of stages: continuous stages, in which change occurs at a steady rate, acceleration stages, in which development proceeds rapidly, and transition stages, at which no change occurs: it represents a transition between two maturational phases. Naturally, the Final State is a special case of a transition stage, even if the terms 'Final' and 'transition' are somewhat at odds with each other. Hence, Ingram mentions a fourth stage, the plateau stage, at which competence has reached the level of the Final Sate and no change is necessary. A typical s-shaped developmental curve, for example, consists of a continuous stage, followed by an acceleration stage, which is succeeded by yet another continuous stage. Although Ingram's stages have proved to be of great use and influence in language acquisition research, they require some type of "goodness" to be measurable. For example, van 't Veer (2012) describes the actual realisations of target liquids by Catootje (one of the subjects in the CLPF corpus, see below), and finds that the realisation of her liquids are scattered over various possibilities (including $\emptyset$) and only slowly improve initially, after which the child goes through a stage of rapid improvement (of the rate of target-to-non-target substitutions). Finally, when production is nearly always correct, the rate of improvement drops to a lower pace – the final continuous stage.

In the current study, stages are defined differently. For one thing, the rate of acquisition is not our primary concern, and furthermore, stages defined on the basis of the rate of development are not fine-grained enough. To determine a rate in the first, place, a stage needs to contain at least two different changes in the child's observable size (for example, additions of segments to the inventory). We are interested in the changes *per se*, and hence it was decided to define a stage as the period of one recording session up to but not including the next recording session in which the inventory has grown. This allows us to discard data points at which no change occurs, yet gives us a fine-grained enough look at development.[5]

4.2.2 Finding FCCs

To find the relevant Feature Co-occurrence Constraints, the featural inventories as described above are fed into the algorithm described in section 3.2.3 above. The steps of the algorithm are briefly reproduced below. The definitions of the FCCs are reproduced in (40).

(40) a. *FG
assign a violation mark for every segment Σ iff [F] is dominated by Σ and [G] is dominated by Σ (*c-constraint*)

b. F→G
assign a violation mark for every segment Σ iff [F] is dominated by Σ and [G] is not in Σ(*i-constraint*)

The first step in the algorithm is to list the features that are active for the current stage. Based on the segments in the inventory and the data in table (4.3), every combination of two features in the matrix is assigned "–" where both features are the same, "0" where the combination is not attested, and "1" where it is. As every cell containing "0" refers to a combination of two features that is logically possible in the phonological grammar at that stage, and yet not legal, that combination is listed as a c-constraint.

Next, a list containing every possible feature combination at the current stage is compared to every cell containing "1", where the feature heading the row is [F], and the feature heading the column is [G]. If [G] is present in every segment that contains [F], [F]→[G] is satisfied. If a segment is found that contains [F] but not [G], [F]→[G] is listed as being violated. Finally, a checking procedure is included to list any over- and undergenerated segments; overgenerated segments are attested segments that are not ruled out by the FCCs that the algorithm found, whereas undergenerated segments are feature combinations that are ruled out, but which are attested.

The result of this procedure is that we have, for each stage of every child, a list of features that are acquired, a list of violated constraints, and lists of over- and undergenerated segments. Ideally, the latter two lists are empty; the attested inventory is then congruent with the state of the grammar. It should be noted that each stage is treated individually by the algorithm, even if this is not the case for the child (who, while having no knowledge of the future, builds on knowledge already acquired). Hence, we can make a number of predictions: if the algorithm that is proposed here is a good model of acquisition, it will

[5]Note, too, that stages are always defined with respect to a specific trait or behaviour. In our case, it is the development of segmental material in the word-onset position.

converge, from one stage to the next, on those parts of the inventory that do not change. In other words, gaps that remain in the inventory remain there because of the same constraints.

It is worthwhile to pause for a moment and consider the limits of the procedure; the type of scenario that is predicted to be impossible. For one thing, the theory cannot rule out segments for which it is necessary to posit co-occurrence restrictions for more than two features. In this case, the theory predicts more segments than are attested. Related to this issue is that no conditional restrictions can be posed. It is impossible to ban, for example, [continuant] combining with [labial], but only when [voice] is involved:*[cont, lab, voice]. Such a constraint would be necessary in the case where /f/, /b/ and /z/ are legal but not /v/. In other words, the following (sub-)inventory is predicted not to occur:

(41)

b	f	*v
		z

The segment /v/ has precisely the three features given as an example above: it is [cont, lab, voice]. In terms of c-constraints, there are three possibilities to rule it out:

(42) *[cont, lab]
*[lab, voice]
*[cont, voice]

The first constraint in 42 is violated by /f/, the second by /b/ and the third by /z/. With all of the relevant c-constraints independently violated by other segments, none can be used to exclude /v/.[6] Given the active features in our example inventory, the following i-constraints are active:

(43) [cont]→[lab]
[lab]→[cont]
[lab]→[voice]
[voice]→[lab]
[cont]→[voice]
[voice]→[cont]

Here, /z/ violates the first and fourth constraint in 43, as it is not a labial. The second constraint is violated by /b/, as is the sixth, a non-continuant. The third and fifth constraints are violated by /f/, because it is not voiced. Again, we see that each available constraint is violated by a segment other than /v/, and hence, it cannot be ruled out. In section 4.5 below, we will come back in more detail to this type of impossible inventory, which is related to set-subset relations of feature combinations (segments). For now, however, it should be noted that we predict that no 'conditional gaps' occur in the data.

[6]Note that constraint conjunction as proposed in OT would allow a constraint *[cont, lab]&*[lab, voice]. Constraint conjunction is an incredibly powerful tool, and we shall not employ it here.

In the introduction, we adopted an assumption that constraints over [F,G] are activated as soon as the second member of the pair F and G is acquired, an assumption that we will see below is not contradicted. In the face of positive evidence, that is to say, after having learned that the segment is part of the surrounding language, the child can revoke constraints immediately or later during development. However, although constraints impose co-occurrence restrictions, it is possible to describe a situation where segment A is allowed at stage S, but no longer so at S+1. The way in which this can be done is highly limited however. Let us first consider the type of shrinking inventory that *cannot* be described, using the following scenario:

(44) Impossible scenario: the shrinking inventory (1)

Stage	*Features*	*Inventory*
S-1	[F]	/f/
S	[F], [G]	/f/, /f͜g/ /g/
S+1	[F], [G]	/f/, /g/

In this scenario, stage S heralds the acquisition of the feature [G]. This feature is uninhibited in its combinations: it is expressed in the segment /g/ as well as in /f͜g/. At stage S+1, it is no longer permitted; to account for this, the constraint *[F, G] must be activated. This is not possible, however, as both [F] and [G] were acquired before stage S+1. The second of the two, [G], was acquired at stage S, which means that the constraint *[F, G] was activated at that same stage, but demoted/de-activated immediately, due to the presence of /f͜g/ in the inventory.[7]

The same argument holds with respect to i-constraints. Consider the marginally different scenario below:

(45) Impossible scenario: the shrinking inventory (2)

Stage	*Features*	*Inventory*
S-1	[F]	/f/
S	[F], [G]	/f/, /f͜g/ /g/
S+1	[F], [G]	/f/, /f͜g/

Here, it is the segment /g/ that is lost after it is acquired. To account for this, a constraint [G]→[F] must be activated, but we run into the same timing problem as we did before.

The only possible way to describe a inventory that disallows a segment after allowing it at an earlier stage is when three features are involved:

[7] It should be noted that under most OT theories of acquisition constraints are never promoted, only demoted. Moving *[F, G] up in the hierarchy is no option.

(46) Possible scenario: the shrinking inventory

Stage	*Features*	*Inventory*	*Active constraints*
S-1	[F]	/f/	
S	[F], [G]	/f/, /f͜g/	[G]→[F]
S+1	[F], [G], [I]	/f/, /g͜i/	[G]→[I]

Here, /g/ is ruled out by [G]→[F] at stage S. At S+1, a new feature is acquired: [I]. It is only present in one segment, /g͜i/. This means that the constraint [G]→[F] must be revoked: it is violated by the new segment. However, the non-existence of /i/ means that a new constraint, [G]→[I], is introduced. This constraint bans the previously legal segment /f͜g/.

In practice, this situation is not encountered because of the way the Guttmann scales are interpreted: gaps are interpreted as accidental in the sense that they are considered artefacts of the sampling methods. Once a segment passes the inclusion criteria described above, it is considered to be acquired permanently. It would be interesting to see whether the gaps found in the Guttmann scales conform to the predictions outlined here, but that is beyond the scope of the current thesis.

Before turning to a detailed exploration of the acquisition of an individual child, we must consider two additional aspects of the proposed theory: constraint inactivity and constraint redundancy.

Let us consider the inventory of Jarmo at early stages.[8] As can be seen in (47a), Jarmo quickly acquires the full range of places of articulation (by stage 3, every major PoA is represented). At stage 4, /m/ is acquired, and with it, the feature [nasal] (see 47b). This is the only nasal for some time, however; only at stage 7 is it accompanied by /n/. The acquisition of the feature [nasal] triggers the activation of two constraints: *[nasal, dorsal], and [nasal]→[labial] (table 4.4 lists all constraints that are active in Jarmo's development). The effect of the first is to ban /ŋ/, whereas the second bans /n/. Since we are looking at (word) onsets only, it is to be expected that the former constraint remains active; this is indeed the case (see table 4.4). At stage 7, however, something must change; /n/ is now an admissible segment in Jarmo's phonological grammar. For this reason, [nasal]→[labial] must be revoked, or, in OT-terms, demoted to a place where it no longer has any influence. The important thing to remember here is that, while OT-type grammars offer a readymade solution for rendering constraints less influential (demotion), a mechanism for constraint de-activation (be it partial or complete) is a necessity for every theory that aims to combine constraints and acquisition data: children's grammars are simply more restrictive at some stages than they might be at a later stage.

(47) a. Jarmo's inventory of segments

[8] The inventories, acquired features and active constraints for each child are listed in Appendix C

Stage	*Inventory*													
1	k													
2	k	t	d											
3	k	t	d	p	b									
4	k	t	d	p	b	m								
5	k	t	d	p	b	m	ʋ							
6	k	t	d	p	b	m	ʋ	l						
7	k	t	d	p	b	m	ʋ	l	n	f				
8	k	t	d	p	b	m	ʋ	l	n	f	s	x		
9	k	t	d	p	b	m	ʋ	l	n	f	s	x	r	ʃ

b. Jarmo's inventory of features

Stage	*Features*
1	[dors]
2	[dors], [voice]
3	[dors], [voice], [lab]
4	[dors], [voice], [lab], [nas]
5	[dors], [voice], [lab], [nas], [cont], [apprx]
6	[dors], [voice], [lab], [nas], [cont], [apprx], [liq]
7	[dors], [voice], [lab], [nas], [cont], [apprx], [liq]
8	[dors], [voice], [lab], [nas], [cont], [apprx], [liq]
9	[dors], [voice], [lab], [nas], [cont], [apprx], [liq], [dist]

The example of Jarmo's nasals is useful to illustrate another aspect of the theory as it is proposed here. At stage 4, when Jarmo acquired the feature [nasal] used only in the segment /m/, two constraints are activated, as we have seen: [nasal]→[labial] to ban /n/ and *[nasal, dorsal] to prevent /ŋ/ from surfacing as a legal segment. The reader may have noticed, however, that the former constraint in fact does both jobs: by requiring [nasal] to only co-occur with [labial], it effectively bans both /n/ *and* /ŋ/. Yet, an additional constraint is activated. At first glance this may seem like an unnecessary complication, but it is important to consider that the child has no knowledge of the future. If only [nasal]→[labial] were to be activated at stage 4, there learner would be forced to reanalyse the inventory and posit *[nasal, dorsal] after all. We assumed that all FCCs would be acquired no later than the stage at which its second feature is acquired (see also section 4.4); if constraints were not redundantly activated, this position would not be tenable.

With this in hand, we can now investigate the acquisition of the segment inventory, focussing on the word-onset. Although it is interesting to consider other positions, as well, the word-onset is usually the first consonantal slot in which segmental knowledge is acquired. To illustrate, the development of the coda starts at a later point, when the child already has some knowledge of sub-segmental phonology; hence, the development of other positions is 'contaminated' by earlier knowledge. In addition, not many children in our database provide a reasonable amount of data from other syllable positions, acquiring only a very limited inventory during the time window of the recordings. For

Stage	1	2	3	4	5	6	7	8	9
Constraints					[apprx]→[cont]	[apprx]→[cont]	[apprx]→[cont]	[apprx]→[cont]	[apprx]→[cont]
					[apprx]→[lab]	[apprx]→[lab]	[apprx]→[lab]	[apprx]→[lab]	[apprx]→[lab]
					[cont]→[apprx]	[cont]→[apprx]			
					[cont]→[lab]	[cont]→[lab]	[cont]→[lab]		
									[dist]→[cont]
									*[apprx, dist]
				[nas]→[lab]	[nas]→[lab]	[nas]→[lab]			
						*[apprx, liq]	*[apprx, liq]	*[apprx, liq]	*[apprx, liq]
						*[cont, liq]	*[cont, liq]	*[cont, liq]	
					*[dors, apprx]	*[dors, apprx]	*[dors, apprx]	*[dors, apprx]	*[dors, apprx]
					*[dors, cont]	*[dors, cont]	*[dors, cont]		
									*[dors, dist]
			*[dors, lab]	*[dors, lab]	*[dors, lab]	*[dors, lab]	*[dors, lab]	*[dors, lab]	*[dors, lab]
						*[dors, liq]	*[dors, liq]	*[dors, liq]	*[dors, liq]
				*[dors, nas]	*[dors, nas]	*[dors, nas]	*[dors, nas]	*[dors, nas]	*[dors, nas]
		*[dors, voice]	*[dors, voice]	*[dors, voice]	*[dors, voice]	*[dors, voice]	*[dors, voice]	*[dors, voice]	*[dors, voice]
									*[lab, dist]
						*[lab, liq]	*[lab, liq]	*[lab, liq]	*[lab, liq]
									*[liq, dist]
					*[nas, apprx]	*[nas, apprx]	*[nas, apprx]	*[nas, apprx]	*[nas, apprx]
					*[nas, cont]	*[nas, cont]	*[nas, cont]	*[nas, cont]	*[nas, cont]
									*[nas, dist]
						*[nas, liq]	*[nas, liq]	*[nas, liq]	*[nas, liq]
					*[voice, apprx]	*[voice, apprx]	*[voice, apprx]	*[voice, apprx]	*[voice, apprx]
					*[voice, cont]	*[voice, cont]	*[voice, cont]	*[voice, cont]	*[voice, cont]
									*[voice, dist]
						*[voice, liq]	*[voice, liq]	*[voice, liq]	*[voice, liq]
				*[voice, nas]	*[voice, nas]	*[voice, nas]	*[voice, nas]	*[voice, nas]	*[voice, nas]

Table 4.4: Feature Co-occurrence Constraints in Jarmo's Actual word onset productions

these reasons, we will limit the discussion to the word onset position.

4.3 Constraints in the child's grammar

In the previous section, we have illustrated some of the aspects of the current theory with data from one child, Jarmo. His data represents an almost ideal case, in which all predictions are borne out (save for the overprediction of /t/ in the first stage - but see below). In this section, we will examine every stage of the development of another child, Noortje. She is selected because her data present the most cases of apparent anomalies. As we will see however, the issues that seem to arise will all be resolved under closer inspection.

4.3.1 Noortje

Stage 1

Inventory /m/
Features [nasal], [labial]
Constraints

Active constraints	*Inactive constraints*
[nas]→[lab]	*[nas,lab]
[lab]→[nas]	

In Noortje's first stage, her onset inventory consists only of [m], for which two features are necessary: [nasal] and [labial]. Two constraints are activated: [nas]→[lab] and the reverse, [lab]→[nas]. The former rules out a segment consisting of just [nasal], which would be interpreted as /n/. The second constraint rules out a segment [labial], which corresponds to /p/. Finally, we see that it is not possible to rule out the empty segment (interpreted as /t/). We will explore the issue of overprediction in full detail in section 4.5.

Stage 2

Inventory /m, p, t, k/
Features [nasal], [labial], [dorsal]
Constraints

Active constraints	*Inactive constraints*
*[nas, dors]	*[nas, lab]
*[dors, lab]	[nas]→[lab]
[lab]→[nas]	[nas]→[dors]
	[dors]→[nas]
	[lab]→[dors]
	[dors]→[lab]

At the second stage, Noortje's inventory grows substantially, adding the voiceless stops to the set. As we mentioned above, /t/ was already predicted to be in the inventory. The feature set in stage 1 is already sufficient to allow for /p/, the only thing that needs to change is the deactivation (or demotion) of the constraint requiring labials to be nasal: [lab]→[nas] is revoked. Finally, [dorsal] is added to allow for /k/, but it is accompanied by two constraints to restrict its combinatorics: *[lab, dors] prevents doubly articulated segments and *[nas, dors] prevents /ŋ/. Since we are considering the onset inventory only, both constraints will remain active in the grammar.

Stage 3

Inventory /m, p, t, k, n/
Features [nasal], [labial], [dorsal]
Constraints

Active constraints	*Inactive constraints*
*[nas, dors]	*[nas, lab]
*[dors, lab]	[nas]→[lab]
	[lab]→[nas]
	[nas]→[dors]
	[dors]→[nas]
	[lab]→[dors]
	[dors]→[lab]

Stage 3 heralds the addition of /n/. The feature needed to produce this segment, [nasal], is already in the inventory. By revoking [nas]→[lab], /n/ is now a legal segment.

Stage 4

Inventory /m, p, t, k, n, b/
Features [nasal], [labial], [dorsal], [voice]
Constraints

Active constraints	*Inactive constraints*
*[nas, dors]	*[nas, lab]
*[nas, voice]	*[lab, voice]
*[lab, dors]	[nas]→[lab]
*[dors, voice]	[lab]→[nas]
[voice]→[lab]	[nas]→[dors]
	[dors]→[nas]
	[nas]→[voice]
	[voice]→[nas]
	[lab]→[dors]
	[dors]→[lab]
	[lab]→[voice]
	[dors]→[voice]
	[voice]→[dors]

Voicing is introduced at stage 4, but it is restricted to occurring with [labial]. Hence, three new constraints are activated: [voice]→[lab], *[dors, voice], and *[nas, voice]. The i-constraint forces [voice] to be with [labial], banning both /d/ and /ɣ/. One might think, then, that the constrain *[dors, voice] is unnecessary, and in fact, it is redundant. However, the requirement that only labials be voiced must be revoked at some point to allow for /d/, and it is important to note that the learner has little or no way of knowing what will happen in the future. Hence, the most prudent strategy is to activate constraints regardless of whether they are redundant, or not. It should be noted that this strategy is also in accordance with the subset principle, which can be paraphrased to hold that the most restrictive grammar should be preferred by the learner.

Stage 5

Inventory /m, p, t, k, n, b, d/
Features [nasal], [labial], [dorsal], [voice]
Constraints

Active constraints	*Inactive constraints*
*[nas, dors]	*[nas, lab]
*[nas, voice]	*[lab, voice]
*[lab, dors]	[nas]→[lab]
*[dors, voice]	[lab]→[nas]
	[nas]→[dors]
	[dors]→[nas]
	[nas]→[voice]
	[voice]→[nas]
	[lab]→[dors]
	[dors]→[lab]
	[lab]→[voice]
	[voice]→[lab]
	[dors]→[voice]
	[voice]→[dors]

At stage 5, this is indeed what happens: [voice]→[lab] is revoked or demoted, because /d/ is added to the inventory. /ɣ/ is still ruled out by *[dors, voice], and continues to be so.

Stage 6

Inventory /m, p, t, k, n, b, d, s/
Features [nasal], [labial], [dorsal], [voice], [continuant]
Constraints

Active constraints	*Inactive constraints*
*[nas, dors]	*[nas, lab]
*[nas, voice]	*[lab, voice]
*[nas, cont]	[nas]→[lab]
*[lab, dors]	[lab]→[nas]
*[lab, cont]	[nas]→[dors]
*[dors, voice]	[dors]→[nas]
*[dors, cont]	[nas]→[voice]
*[voice, cont]	[voice]→[nas]
	[nas]→[cont]
	[cont]→[nas]
	[lab]→[dors]
	[dors]→[lab]
	[lab]→[voice]
	[voice]→[lab]
	[lab]→[cont]
	[cont]→[lab]
	[dors]→[voice]
	[voice]→[dors]
	[dors]→[cont]
	[cont]→[dors]
	[voice]→[cont]
	[cont]→[voice]

No constraints are revoked at stage 6, but [continuant] is acquired because /s/ is added to the inventory. Since no other continuants are in the inventory (yet), the constraints *[dors, cont], *[lab, cont], *[voice, cont] and *[nas, cont] remain activated. Again, one of these constraints, *[nas, cont], will remain active in the mature grammar.

Stage 7

Inventory /m, p, t, k, n, b, d, s, ʋ/
Features [nasal], [labial], [dorsal], [voice], [continuant], [approximant]
Constraints

Active constraints	*Inactive constraints*
*[nas, dors]	*[nas, lab]
*[nas, voice]	*[lab, voice]
*[nas, cont]	*[lab, cont]
*[nas, apprx]	*[lab, apprx]
*[lab, dors]	*[cont, apprx]
*[dors, voice]	[nas]→[lab]
*[dors, cont]	[lab]→[nas]
*[dors, apprx]	[nas]→[dors]
*[voice, cont]	[dors]→[nas]
*[voice, apprx]	[nas]→[voice]
[apprx]→[lab]	[voice]→[nas]
[apprx]→[cont]	[nas]→[cont]
	[cont]→[nas]
	[nas]→[apprx]
	[apprx]→[nas]
	[lab]→[dors]
	[dors]→[lab]
	[lab]→[voice]
	[voice]→[lab]
	[lab]→[cont]
	[cont]→[lab]
	[lab]→[apprx]
	[dors]→[voice]
	[voice]→[dors]
	[dors]→[cont]
	[cont]→[dors]
	[dors]→[apprx]
	[apprx]→[dors]
	[voice]→[cont]
	[cont]→[voice]
	[voice]→[apprx]
	[apprx]→[voice]
	[cont]→[apprx]

The first glide appears at stage 7, when the feature [approximant] is acquired. Severe co-occurrence restrictions hold for this feature, with c-constraints being activated for [approximant] with nearly every other feature that has been acquired up to now: *[dors, apprx], *[nas, apprx], *[voice, apprx]. This does not suffice to rule out /j/ [cont, apprx], however, so the i-constraint [apprx]→[lab] is activated as well. Approximants must be continuant, and so we see the activation of [apprx]→[cont].

One constraint must be revoked in order to allow for /ʋ/, which has the features [lab, cont, apprx], to be in the inventory: *[lab, cont]. This results in another overprediction, namely that of /f/, which is [lab, cont].

Stage 8

Inventory /m, p, t, k, n, b, d, s, ʋ, f/
Features [nasal], [labial], [dorsal], [voice], [continuant], [approximant]
Constraints

Active constraints	*Inactive constraints*
*[nas, dors]	*[nas, lab]
*[nas, voice]	*[lab, voice]
*[nas, cont]	*[lab, cont]
*[nas, apprx]	*[lab, apprx]
*[lab, dors]	*[cont, apprx]
*[dors, voice]	[nas]→[lab]
*[dors, cont]	[lab]→[nas]
*[dors, apprx]	[nas]→[dors]
*[voice, cont]	[dors]→[nas]
*[voice, apprx]	[nas]→[voice]
[apprx]→[lab]	[voice]→[nas]
[apprx]→[cont]	[nas]→[cont]
	[cont]→[nas]
	[nas]→[apprx]
	[apprx]→[nas]
	[lab]→[dors]
	[dors]→[lab]
	[lab]→[voice]
	[voice]→[lab]
	[lab]→[cont]
	[cont]→[lab]
	[lab]→[apprx]
	[dors]→[voice]
	[voice]→[dors]
	[dors]→[cont]
	[cont]→[dors]
	[dors]→[apprx]
	[apprx]→[dors]
	[voice]→[cont]
	[cont]→[voice]
	[voice]→[apprx]
	[apprx]→[voice]
	[cont]→[apprx]

In stage 8 the overprediction introduced in stage 7 is resolved: the segment /f/ is now in the attested inventory. No new features are added, and therefore no constraints are introduced, nor is any constraint revoked.

Stage 9

Inventory /m, p, t, k, n, b, d, s, ʋ, f, l, j, x/
Features [nasal], [labial], [dorsal], [voice], [continuant], [approximant] , [liquid]
Constraints (see next page)
At the final stage within the time span of the recordings, Noortje has not acquired every segment yet. Notably missing are voiced fricatives and /r/. The other liquid, /l/, is acquired at this point. This entails the acquisition of the feature [liquid], which may not co-occur with any of the place features (*[dors, liq], *[lab, liq]), nor with any of the other manner features (*[apprx, liq], *[cont, liq], where the latter rules out /r/), nor with voice (*[voice, liq]). *[dors, cont] is lifted to allow for /x/, and the revocation of [apprx]→[lab] accounts for the acquisition of /j/.

Constraints

Active constraints	*Inactive constraints*
*[nas, dors]	*[nas, lab]
*[nas, voice]	*[lab, voice]
*[nas, cont]	*[lab, cont]
*[nas, apprx]	*[lab, apprx]
*[nas, liq]	*[dors, cont]
*[lab, dors]	*[cont, apprx]
*[lab, liq]	[nas]→[lab]
*[dors, voice]	[lab]→[nas]
*[dors, apprx]	[nas]→[dors]
*[dors, liq]	[dors]→[nas]
*[voice, cont]	[nas]→[voice]
*[voice, apprx]	[voice]→[nas]
*[voice, liq]	[nas]→[cont]
*[cont, liq]	[cont]→[nas]
*[apprx, liq]	[nas]→[apprx]
[apprx]→[cont]	[apprx]→[nas]
	[nas]→[liq]
	[liq]→[nas]
	[lab]→[dors]
	[dors]→[lab]
	[lab]→[voice]
	[voice]→[lab]
	[lab]→[cont]
	[cont]→[lab]
	[lab]→[apprx]
	[apprx]→[lab]
	[lab]→[liq]
	[liq]→[lab]
	[dors]→[voice]
	[voice]→[dors]
	[dors]→[cont]
	[cont]→[dors]
	[dors]→[apprx]
	[apprx]→[dors]
	[dors]→[liq]
	[liq]→[dors]
	[voice]→[cont]
	[cont]→[voice]
	[voice]→[apprx]
	[apprx]→[voice]
	[voice]→[liq]
	[liq]→[voice]
	[cont]→[apprx]
	[cont]→[liq]
	[liq]→[cont]
	[apprx]→[liq]
	[liq]→[apprx

Having now fully explored the developmental path of Noortje's onset inventory in terms of features and Feature Co-occurrence Constraints, let us investigate how she and Jarmo, whose data we examined in section 4.2, compare to the other children in the database.

4.3.2 Constraints: Noortje, Jarmo and the other children

The total number of constraints employed by Noortje is far lower than the number that is logically possible (for nine features, the number of possible constraints is 1.5n(n-1) = 108); the theory is thus highly economic in terms of the number of constraints that remain activated. Table 4.5 lists, for the actual productions, all constraints that are active in the current study, and the children in whose grammars the constraints are activated.

It can be seen that there is a high degree of overlap between the constraints that the children are using, especially in the set of c-constraints. Comparing the children's constraints to the ones active in Dutch, we see that there are some that are active in the children's grammars, but not in the Adult grammar; these are transient constraints that will be revoked or demoted at some point after our window on development ends. More interestingly, there are also constraints that are active in Dutch, but not in the children's grammars. First of all, it is important to note that there are no constraints that are exclusive to the mature grammar. Secondly, the vast majority of constraints in this category involve the feature [distributed]. These constraints are only active in the grammars of Jarmo and Tirza. This is because Jarmo and Tirza are also the only two children who have acquired the feature by the end of the developmental window.[9]

Given that we have nine features, and that for every combination of two features, three constraints are possible (one c-constraint, and two i-constraints), the number of possible constraints is quite large (108). In chapter 3, we saw that for adult Dutch, only 19 FCCs are required to be active: 18 c-constraints and one i-constraint.These are represented in the final column in table 4.5 repeated in (48) below:

(48) a. C-CONSTRAINTS
*[lab, dors]
*[lab, dist]
*[lab, liq]
*[voice, dors]
*[voice, dist]
*[voice, nas]
*[voice, liq]
*[voice, apprx]
*[dors, dist]
*[dors, liq]
*[dors, apprx]

[9]The one other constraint that is in the mature grammar but not in all of the children's, is *[liq, apprx], which only Catootje, Tirza and Tom have. It should be noted that this constraint is formally and functionally equivalent to *[apprx, liq], which is in all of the children's systems. The fact that some children have a redundant formalisation of this co-occurrence restriction, whereas others have only a single constraint, reflects a difference in the order of acquisition of both features. Catootje, Tirza and Tom, who have both versions, acquire [liquid] before [approximant], whereas the other children have the opposite acquisition order.

	Catootje	*Eva*	*Jarmo*	*Noortje*	*Robin*	*Tirza*	*Tom*	*Adult Dutch*
[apprx]→[cont]	✓	✓	✓	✓	✓	✓	✓	✓
[apprx]→[lab]			✓	✓	✓		✓	
[cont]→[apprx]	✓	✓	✓					
[cont]→[lab]			✓				✓	
[dist]→[cont]			✓			✓		
[lab]→[voice]					✓			
[nas]→[lab]			✓	✓				
[voice]→[lab]	✓					✓	✓	
*[apprx, dist]			✓			✓		✓
*[apprx, dors]	✓	✓	✓	✓	✓	✓	✓	✓
*[apprx, liq]	✓	✓	✓	✓	✓	✓	✓	✓
*[cont, dors]	✓	✓	✓	✓	✓	✓		
*[cont, liq]	✓	✓	✓	✓	✓	✓	✓	
*[dors, dist]			✓			✓		✓
*[dors, lab]	✓	✓	✓	✓	✓	✓	✓	✓
*[dors, liq]	✓	✓	✓	✓	✓	✓	✓	✓
*[dors, nas]	✓	✓	✓	✓	✓	✓	✓	
*[dors, voice]	✓	✓	✓	✓	✓	✓	✓	✓
*[lab, apprx]	✓					✓		
*[lab, cont]	✓			✓	✓	✓		
*[lab, dist]			✓			✓		✓
*[lab, liq]	✓	✓	✓	✓	✓	✓	✓	✓
*[lab, nas]		✓			✓			
*[liq, apprx]	✓					✓	✓	✓
*[liq, dist]			✓			✓		✓
*[nas, apprx]	✓	✓	✓	✓	✓	✓	✓	✓
*[nas, cont]	✓	✓	✓	✓	✓	✓	✓	✓
*[nas, dist]			✓			✓		✓
*[nas, liq]	✓	✓	✓	✓	✓	✓	✓	✓
*[nas, voice]	✓	✓	✓	✓	✓	✓	✓	✓
*[voice, apprx]	✓	✓	✓	✓	✓	✓	✓	✓
*[voice, cont]	✓	✓	✓	✓	✓	✓	✓	
*[voice, dist]			✓			✓		✓
*[voice, liq]	✓	✓	✓	✓	✓	✓	✓	✓

Table 4.5: Constraints activated by the children in the current study

*[cont, nas]
*[dist, nas]
*[dist, liq]
*[dist, apprx]
*[nas, liq]
*[nas, apprx]
*[liq, apprx]

b. I-CONSTRAINTS
[apprx]→[cont]

Table 4.6 condensed the data in table 4.5 by listing the number of different constraints activated per child. In some cases, the number for the children is smaller than the number of constraints for the adult grammar, but the opposite is the more frequent situation. The list in (48) concerns the inventory at all positions, whereas for the children, we are only considering word onsets exclusively. /ŋ/, for example, is in the inventory of Dutch, but it is not allowed in onsets. To account for this difference, the adult grammar has one constraint *less*: we find *[nas, dors] in the grammars of every child (and at every level), but it is not part of the list in (48).

More importantly, a look at the inventories in appendix C reveals that by the end of the recording sessions, not all children have acquired the full inventory (see also our discussion of Noortje's inventory, above). The two largest numbers in table 4.6 are found for the inventories of Jarmo and Tirza. These are also the only two inventories where the feature [distributed] is acquired (see above). This feature causes seven of the constraints in (48), its absence in most of the children's inventories accounts for much of the observed difference in the amount of constraints.

	Constraints
Catootje	22
Eva	19
Jarmo	28
Noortje	21
Robin	21
Tirza	27
Tom	20

Table 4.6: Number of different constraints used per child

In table 4.7, we see that there are 34 constrains in the current study. Appendix B presents the constraints per child, in the same fashion as table 4.5 does. Of interest are those constraints that do not remain active, that is, those that are unique to the children's grammars. One possible way in which the method developed here can be used is to study temporary constraints in child language, in a way similar to Fikkert and Levelt (2008). Example (49) lists the constraints that are unique to the inventories of the children studied here.

(49) a. C-CONSTRAINTS
*[cont, dors]
*[cont, liq]
*[dors, nas]
*[lab, apprx]
*[lab, cont]
*[lab, nas]
*[voice, cont]

b. I-CONSTRAINTS
[apprx]→[lab]
[cont]→[apprx]
[cont]→[lab]
[dist]→[cont]
[lab]→[voice]
[nas]→[lab]
[voice]→[lab]

The first thing that we see is that while for the mature inventory only a single i-constraint is needed, the children use i-constraints far more frequently. This is likely to be due to the nature of FCCs. For every feature, c-constraint specify a single co-occurrence restriction, but i-constraints ban every possible restriction *except* for the specified co-occurrence. Hence, i-constraints are much more powerful, and perhaps too powerful for a full-fledged inventory.

It is worth repeating, furthermore, that the number of constraints is much smaller than it would be, if every possible constraint were activated and subsequently ranked. It is unclear whether there is, at this point, any evidence that suggests a correlation between the number of constraints in a grammar and its efficiency, fitness, learnability or any other measure of goodness (note that in the original conception of OT, there is no pressure to keep the number of constraints to a minimum; the goal of factorial typologies is to reduce the number of ad-hoc constraints, not the number of constraints *per se*.

4.3.3 Summary

In this section, we have seen a demonstration of Feature Co-occurrence Constraints in acquisition, by carefully observing the development of Noortje's inventory of onset productions. In addition, we have seen that over all, only a small number of possible FCCs become activated. I take this to be a good

	Active
[apprx]→[cont]	✓
[apprx]→[lab]	✓
[cont]→[apprx]	✓
[cont]→[lab]	✓
[dist]→[cont]	✓
[lab]→[voice]	✓
[nas]→[lab]	✓
[voice]→[lab]	✓
*[apprx, dist]	✓
*[apprx, dors]	✓
*[apprx, liq]	✓
*[cont, dors]	✓
*[cont, liq]	✓
*[dors, dist]	✓
*[dors, lab]	✓
*[dors, liq]	✓
*[dors, nas]	✓
*[dors, voice]	✓
*[lab, apprx]	✓
*[lab, cont]	✓
*[lab, dist]	✓
*[lab, liq]	✓
*[lab, nas]	✓
*[liq, apprx]	✓
*[liq, dist]	✓
*[nas, apprx]	✓
*[nas, cont]	✓
*[nas, dist]	✓
*[nas, liq]	✓
*[nas, voice]	✓
*[voice, apprx]	✓
*[voice, cont]	✓
*[voice, dist]	✓
*[voice, liq]	✓

Table 4.7: Constraints active in the current study

thing, and even though the procedure by which constraints are derived results in some redundant constraints, the number of constraints remains very manageable. In the rest of this chapter, we will go deeper into some of the findings illustrated by Noortje above.

4.4 General observations

Based on the discussions above (the data for every child are given in full in appendix C), a number of general observations can be made concerning the assumptions and predictions discussed in the introduction to this chapter.

4.4.1 Continuity

A fundamental prediction we made at the beginning of this chapter is that the inventory at each stage of development should correspond to a possible adult inventory. This does not necessarily have to hold at the surface level (especially since early child inventories are often smaller than most adult inventories), but the underlyingly, the structure should follow the same principles. These principles, we assumed, are modeled in terms of monovalent features, i-constraints and c-constraints. We have seen that this prediction holds true with only a very limited number of exceptions.

4.4.2 Order of acquisition

Since a segment that has no features (an empty root node, for example) cannot be banned by Feature Co-occurrence Constraints, it was predicted that /t/ is present from the very first stage on. This prediction is borne out with the exception of Noortje, whose first stage contains only /m/. Furthermore, related to the discussion on the Continuity Hypothesis above, the Feature Co-occurrence Constraint can rule out certain impossible orders of acquisition (e.g., those leading to conditional gaps. The theory makes no claim as to determining the order of acquisition of individual features, but given the acquisition of those features, some combinations are ruled out.

4.4.3 Timing

In the introduction to this chapter, the assumption was adopted that constraints would emerge exactly at the point in time at which the second of the two features to which it refers is acquired. This assumption is proves to be tenable; no counterexamples were encountered. This raises the question whether what we are dealing with is in fact not the confirmation of an independent assumption, but rather an artefact resulting from some mechanism in the methodology. Such an artefact could arise if the algorithm by which constraints are uncovered had built-in to it some device to explicitly ban the introduction of constraints at any other stage than the one at which the second of both features is introduced, or, alternatively, if it had some way of knowing what constraints would be necessary at later stages.

The important point is that the algorithm has no memory beyond a single stage; each stage is treated as if it were the inventory of an individual language. Rather than evaluating what is different, the algorithm approaches each individual inventory anew. This approach ensures that the algorithm has no knowledge of what constraints were active at previous stages, and hence, it cannot know whether a constraint it activates is new or not. This, in turn, means that the timing of the activation of the constraints is not an artefact of the algorithm. An additional implication of the fact that no new constraints are created without the introduction of a new feature is a further reinforcement of the robustness of the FCC approach.

4.4.4 Converging stages

In this chapter, we approached the acquisition of the segment inventory as the successive application of the constraint generating algorithm introduced in chapter 3. This means that the model diverges from actual acquisition, as children do not likely reanalyse the structure of their inventory each time a new segment (feature combination) is added (but see section 5.2.3). At the same time, if the model converges on the same constraints for parts of the inventory that do not change form one stage to the next, this reinforces the claims it makes further. Indeed, we have seen that convergence is the norm without exception.

4.4.5 Progression of stages

Before considering individual stages and their content, it is important to consider how stages might differ from each other. There are three general ways in which a stage may develop into the next stage: first, a new feature may be acquired, secondly, a constraint is revoked, and third, no change occurs in either the number of features or constraints; in this case, a segment that was previously predicted but not attested is now in the inventory. We will come back to this phenomenon (which we will call 'overprediction') in detail. In each case, the inventory grows; this is trivial, as stages are defined as containing at least one additional segment in the inventory. The three different stage progressions are given schematically below.

(50) Stage progressions

	Feature	*Constraint*	
A	New	i.	New
		ii.	No new
B	No new		Lifted/Revoked
C	No new		No change

Situation A is a straightforward way to expand the inventory: to add a new feature to the system. Such a feature may be immediately subject to co-occurrence restrictions (A-i), or it may occur freely (A-ii). The latter situation is rare, and usually found in earlier stages, when there are few features to be combined. In practice, we only encounter situations in which a feature is free to combine with others in the first stages. Consider for example Eva, whose stage 1 has the inventory in (51a):

(51) Unconstrained feature combinations in Eva's first stage

a. Inventory

Stage	*Inventory*			
1	p	b	t	d

b. Features

Stage	*Features*
1	[labial], [voice]

This inventory is the perfectly symmetrical result of the features in (51b) (remember that there is no feature 'coronal'), and hence, no FCCs are introduced.

Situation Ai is default, and exemplified by Eva's Actual productions when she transitions from stage 1 to stage 2, as can be seen in example (52a).

(52) Eva's early stages

a. Stages

Stage	*Inventory*				
1	p	b	t	d	
2	p	b	t	d	n

b. Constraints

Stage	1	2
Constraints		*[lab, nas]
		*[voice, nas]

The feature [nasal] becomes active, but it is subject to restrictions: it may not co-occur with [labial], nor with [voice] (the former requirement remaining active in the final state of the grammar). This is enforced by the constraints in (52b).

Examples of situation B, where the inventory grows by revoking/demoting a constraint, are found in the data from virtually all children. Let us consider Noortje's inventory, and zoom in on the progression from stage 2 to stage 3. Example (53a) lists Noortje's inventory at stages 2 and 3. The inventory grows by adding /n/. In (53b), we see that no new features are acquired; the set in stage 2 suffices to generate all segments in the inventory at stage 3.

(53) Lifted constraints in Noortje's inventory

a. Inventory

Stage	*Inventory*				
1	m				
2	m	p	t	k	
3	m	p	t	k	n

b. Features

Stage	*Features*
1	[nasal], [labial]
2	[nasal], [labial], [dorsal]
3	[nasal], [labial], [dorsal]

c. Constraints

Stage	1	2	3
Constraints	[lab]→[nas]	[lab]→[nas]	
	[nas]→[lab]		
		*[lab, dors]	*[lab, dors]
		*[nas, dors]	*[nas, dors]

The inventory at stage 3 is generated simply by abandoning the requirement that nasals should be labial.

We see that a subset of the inventory in stage 2, leaving /k/ (and thus [dorsal]) out of the equation, also exemplifies situation B: the nasal inventory grows not by adding a new feature, but by lifting the constraint that restricts [nasal] to only co-occur with [labial].

The final manner in which stages may progress is listed in 50C, where the inventory grows despite the fact that no new features are added, and no constraints are revoked. It may seem contradictory at first that the inventory can change without a change in the system generating it, i.e., the inventory changes whilst there is no change in the grammar. The context in which this occurs is one where, at a given stage, the combination of features and FCCs predicts a larger inventory than is attested. In a next stage, attested and predicted are aligned once more. Remember that there are no overpredicted segments in the system of features and FCCs generating the adult inventory (see chapter 3). We will discuss the properties and implication of overpredictions in section 4.5, but before doing this, let us clarify what overprediction is by an example.

For this, we will return to Eva's developing inventory. Example (54) gives every stage of Eva's inventory (54a), the features she employs (54b), and the segments that the system predicts, but which are not attested (54c). Finally, the constraints that govern her inventory are listed in table (4.4.5).

(54) a. Development of Eva's inventory

Stage	*Inventory*												
1	p	b	t	d									
2	p	b	t	d	n								
3	p	b	t	d	n	m	ʋ	j					
4	p	b	t	d	n	m	ʋ	j	l	s	z		
5	p	b	t	d	n	m	ʋ	j	l	s	z	f	
6	p	b	t	d	n	m	ʋ	j	l	s	z	f	k

b. Features

Stage	*Features*
1	[lab], [voice]
2	[lab], [voice], [nas]
3	[lab], [voice], [nas], [cont], [apprx]
4	[lab], [voice], [nas], [cont], [apprx], [liq]
5	[lab], [voice], [nas], [cont], [apprx], [liq]
6	[lab], [voice], [nas], [cont], [apprx], [liq], [dors]

c. Overpredicted segments

Stage	*Overpredicted*
1	
2	
3	
4	[lab, voice, cont] [lab, cont]
5	[lab, voice, cont]
6	[lab, voice, cont]

Two segments are overpredicted in Eva's onset inventory. First, in stage 4, the segment [lab, cont] (/f/) is legal according to the combination of features and constraints in her grammar, but it is not yet attested. Secondly, its voiced counterpart, /v/, remains unattested in her recordings. This situation arises because of the early acquisition of /ʋ/, consisting of [labial, continuant, approximant], at stage 3. A subset of this segment, namely [labial, continuant], makes up /f/, which is unattested at this time. This is accounted for by the constraint [continuant]→[approximant]. This constraint must be revoked at stage 4, however, because /s/ is now attested; clearly, [continuant] is no longer restricted to [approximant]. The combination [labial, continuant] can now no longer be ruled out, and hence /f/ is predicted to be in the inventory. Since it is not, it is overpredicted.

The other overpredicted segment, /v/, consists of [labial, continuant, voice]. This combination is predicted to occur because the subset combinations cannot be ruled out: we have already seen how [labial, continuant] is overpredicted. The combination [continuant, voice] is legal and attested, in /z/. Furthermore, [continuant, labial] is present in /ʋ/, and [labial, voice] in /b/. Indeed, we see that /f/ is attested in the next stage (although /v/ remains unattested).

Stage	1	2	3	4	5	6
Constraints			[apprx]→[cont]	[apprx]→[cont]	[apprx]→[cont]	[apprx]→[cont]
			[cont]→[apprx]			
						*[apprx, dors]
				*[apprx, liq]	*[apprx, liq]	*[apprx, liq]
						*[cont, dors]
				*[cont, liq]	*[cont, liq]	*[cont, liq]
		*[lab, nas]				
						*[lab, dors]
				*[lab, liq]	*[lab, liq]	*[lab, liq]
						*[liq, dors]
			*[nas, apprx]	*[nas, apprx]	*[nas, apprx]	*[nas, apprx]
			*[nas, cont]	*[nas, cont]	*[nas, cont]	*[nas, cont]
						*[nas, dors]
				*[nas, liq]	*[nas, liq]	*[nas, liq]
			*[voice, apprx]	*[voice, apprx]	*[voice, apprx]	*[voice, apprx]
			*[voice, cont]			
						*[voice, dors]
				*[voice, liq]	*[voice, liq]	*[voice, liq]
		*[voice, nas]	*[voice, nas]	*[voice, nas]	*[voice, nas]	*[voice, nas]

Table 4.8: Feature Co-occurrence Constraints in Eva's word onset productions

Segment	Catootje	Eva	Jarmo	Noortje	Robin	Tirza	Tom
	10	6	9	9	10	5	9
t			1	1			1
s	3						5
f	6-8	4		7			
z					5		
v	6	4-6					

Table 4.9: Overpredicted segments. Numbers indicate stages at which overprediction occurs.

4.5 Overpredictions

The previous section concluded with the observation that overpredictions occur, where the term denotes segments that are legal with respect to the state of the grammar at some stage, but not yet attested. This is indeed a situation that we encounter at some frequency, and in this section, we will investigate why this is so, when it happens, and what it means. We will conclude the section with a brief discussion of why the reverse situation, where the attested inventory is larger than can be generated, does not occur.

4.5.1 Overpredicted segments

As it turns out, the variety of overpredicted segments is smaller than one might expect. Over all children, only five different segments are ever overpredicted. These are listed in (55) below:

(55) Overpredicted segments

Segment	*Feature combination*
t	[ ]
s	[continuant]
f	[labial, continuant]
z	[voice, continuant]
v	[labial, voice, continuant]

Apart from the fact that only a limited subset of all possible segments is ever overpredicted, all six have in common that they are relatively simple. There is only one segment that requires three features, the others are less complex. This, as we will discuss in more detail below, is because one of the important contextual aspects of overprediction is that it is not possible, under certain circumstances, to rule out feature combinations that form a subset of other combinations.

Table 4.9 lists the five overpredicted segments, and indicate per child and per level of description in which stage(s) it is overpredicted. A number of observations can be made with respect to these tables, and we will discuss the time span between the stage at which the overpredicted segments become available and the stage at which they are attested, and the identity of the overpredicted segments.

4.5.2 Possible causes

Overpredictions are taken to mean that segments that the child *should* be able to produce, are not encountered. This raises the question of what underlies overpredictions. Several possible causes exist:
1 – It could simply be the case that the system of Feature Co-occurrence Constraints as it is proposed here is too permissive. However, if this were the case, it would be difficult to account for the limited number of overpredicted feature combinations we find.
2 – Another possibility is that the child does in fact produce the overpredicted segments, but just happens not to do so during the recording session(s). Considering the generally unmarked status of the segments in (55), this seems an unlikely explanation. For example, van Severen, Molemans, van den Berg, and Gillis (2012) find that while chances of inclusion (avoiding false negatives) are related to sample size, this is true to a lesser degree for more frequent, less marked segments.
3 – A very similar explanation is that the segments are in the inventory, but do not reach criterium yet. The grammar of the child is constantly evolving, and the recordings take place at what are, essentially, random moments. Because of this, a 100% match between predicted and attested inventories is not to be expected in the first place. However, this explanation applies only to those overpredictions that are reasonably quickly resolved.
4 – Considering the limited variation in the table in (55), a possible cause could be that the overpredictions that are encountered are an artefact of the feature system that is used.

The latter option is certainly true of /t/, which cannot be ruled out, being devoid of featural content and thus immune to any FCCs. In fact, in every case in the current survey, /t/ is present at the first stage. In many cases it is in the attested inventory, but on those occasions where it is not, it is in the list of overpredicted segments: it cannot be ruled out and thus is predicted to always be in the inventory. In other words, the prediction is that it is acquired first. This is not always the case, but a look at table 4.9 reveals that where /t/ is overpredicted, this situation is usually resolved quickly.

If we go back to the raw data, before the filters listed in 4.2 are applied, we see clear evidence that the penultimate explanation (overpredicted segments are in the inventory but do not yet reach criterium) is also true. Of all the cases listed in table 4.9 above, only a handful turn out not to be in the inventory at all. These are listed below:

(56) Unattested overpredicted segments
/t/: Noortje, stage 1

/f/: Catootje, stages 6-8
/v/: Catootje, stage 6

This brings the list of overpredicted segments down to three: /t, f, v/. Of these, we know that the first cannot be ruled out, and tellingly, if it is unattested, it is only so in the first stage of Noortje's acquisition.

The other two segments, /f, v/ make up a considerable part of the fricative subset inventory of Dutch. There are two possible explanations why it should be these segments that we encounter here. First, it is a familiar observation that children prefer to avoid fricatives in onsets (Fikkert (1994), see also section 4.6.3 below on initial stopping). Secondly, each segment forms a subset of an approximant: /f, v/⊂/ʋ/. Indeed, in Catootje's stages 6-8, /f/ is not in the inventory, and in stage 6, /v/ is not in the inventory, whereas in these stages, /ʋ/ is.

Criteria and data inclusion

At this point, it is of interest to note that a clear prediction of the Feature Co-occurrence Constraints theory is that /t/ is present in the developing inventory from the start. Noortje's overprediction of /t/ in her stage 1 illustrates an important point. Stage 1 contains only /m/, but only because it minimally reaches criterion. Had it not, or had the criterion been slightly different, stage 1 would have extended over more recording sessions, and it would have included /t/.

Also of interest is that the majority of overpredictions in table 4.9 is only apparent in the sense that the segments are produced, but not in such a way as to reach criterium.

These observations illustrate an important proviso that must be acknowledged with respect to any study of acquisition: inclusion criteria are always somewhat arbitrary, and never without artifactual consequences. The criteria in 4.2 are no exception. They were chosen to resemble those in Levelt and van Oostendorp (2007), and also the criteria proposed in Ingram (1981).

It is, of course, possible to admit every instance of a segment into the data set. In fact, some studies do just this (Ferguson & Farwell, 1975). However, as pointed out by Ingram (1989), this makes any results extremely vulnerable to incidental variation, or even to non-linguistic utterances. In fact, this is illustrated by Noortje. Remember that her stage 1 is defined by containing only /m/. Before stage 1, she produces one word consistently (/mama/ [mama]) and some only once (/ku/ [ku] 'cow'). During the third recording, a new word enters her vocabulary: /χəmakt/ [mɑ] 'made'. Taken together with four instances of 'mama' (counted as a single instance for being tokens of one type), makes it that Noortje just reaches criterion for /m/. The problem is that it is difficult to ascertain whether /mama/ is a 'word' in the sense that it is generated by a phonological grammar, or whether it is a remnant from a previous stage of development.

If including every utterance is too permissive, Noortje's case illustrates that perhaps, our criteria still not restrictive enough. A case can be made to exclude items such as 'mama' and 'papa', onomatopoeia, and some other classes. In the current study, we have opted to restrict ourselves to objective, numerical inclusion criteria. These could have been stricter or laxer, but ultimately a choice must be made.

4.5.3 Context for overpredictions

In the previous section, we discussed possible causes of overprediction. We concluded by observing that many of the overpredicted segments are subset segments of others. Simply being an unattested subset segment of an attested segment is not enough to be overpredicted. This is because i-constraints can force the feature(s) comprising the subset segment to co-occur with another feature. Below, we will see that there are three formal contexts in which overpredictions occur:

(57) Segment (feature combination) A is overpredicted if
- a. A is unattested **and**
- b. A is empty

(58) Segment (feature combination) A is overpredicted if
- a. A is unattested **and**
- b. there is some attested feature combination B such that A⊂B **and**
- c. there is an attested feature combination C such that at least one of the members of A∈C **and**
- d. C≠B

(59) Segment (feature combination) A is overpredicted if
- a. A is unattested **and**
- b. |A| >2 **and**
- c. every subset {F,G} such that {F,G}⊂A is in some segment B, C, *etc.* **and**
- d. B ≠ A and C ≠ A *etc.*

The contexts in (57a) and (58a) are obvious: a segment can only be overpredicted if it is unattested. We have already discussed requirement (57): /t/ is either attested or overpredicted. The requirements in (58) treat cases of non-empty segments. Requirement (58b) demands a superset of the overpredicted segment be present, while requirement (58c) ensures that at least one of the features in A occurs in another combination than the one in B. This is to ensure that there is no i-constraint limiting the subset feature(s) to a single co-occurrence, as the following will illustrate. Take, for example, unattested segment A to be [F,G]. It is a subset of segment B, which is [F,G,H]. In this scenario, which corresponds to (58 a-b), it is still possible to describe the ungrammaticality of segment A by use of the constraint G→H. This possibility no longer exists, however, if segment C ([G,I]) exists, as the i-constraint is now violated by C and hence cannot be employed to rule out A.

Most cases of overprediction are covered under the definitions in (57) and (58), but there are exceptions. These are described in requirement 59. To illustrate, let us look at the overprediction of /v/ in Catootje's stage 6. Below, the inventory is given at the relevant stage and the ones preceding and following it.

(60) Overprediction in Catootje's stages

Stage	*Inventory*													
5	p	b	t	k	m	n	l	d	j	z	s	r		
6	p	b	t	k	m	n	l	d	j	z	s	r	ʋ	
7	p	b	t	k	m	n	l	d	j	z	s	r	ʋ	v

At stage 6, both /f/ and /v/ are predicted to be in the inventory, but are not actually attested. The situation is resolved at the next stage for /v/, but not for /f/ (incidentally, both segments are also not present in the unfiltered inventory – see above). Running /f/ through the requirements listed in the definition in (58), we see the following:

(61) Overprediction of /f/: [cont, lab] is

a. not attested and
b. there is some attested feature combination /ʋ/ [cont, lab, apprx] such that /f/⊂/ʋ/ and
c. there is an attested feature combination /m/ [lab, nas] such that [lab]∈[cont, lab] and
d. /m/ ≠ /ʋ/

In other words, the overprediction of /f/ falls neatly in the context described in (58). The same does not hold for /v/, however.

(62) Overprediction of /v/: [voice, cont, lab] is

a. not attested and
b. there is **no** attested feature combination X [voice, cont, lab, . . .] such that /v/⊂X and
c. there is an attested feature combination /m/ [lab, nas] such that [lab]∈[voice, cont, lab] and
d. /m/ ≠ X

There is no superset segment for /v/, and yet it is overpredicted. Looking at the inventory in (60), we see that every possible subset of /v/ is attested: [voice, continuant] in /z/, [continuant, labial (approximant)] in /ʋ/, and [labial, voice] in /b/. Thus, there cannot be a ban on the combination of [voice] and [continuant], there cannot be a ban on the combination of [continuant] and [labial], and there cannot be a ban on the combination of [labial] and [voice]. Since each of the three constituent features co-occurs with at least two others, i-constraints cannot be of help here, either. Hence, we need additional requirements to the ones in (57) and (58).

(63) Segment (feature combination) A is overpredicted if

a. A is unattested **and**

b. $|A| > 2$ **and**

c. every subset {F,G} such that $\{F,G\} \subset A$ is in some segment B, C, *etc.* **and**

d. $B \neq A$ and $C \neq A$ *etc.*

With these three definitions, we can describe every case of overprediction in table 4.9, where definition (57) describes the overprediction of the zero-feature segment /t/, while definition (58) describes the overprediction of the one- and two-feature segments /s, f, z/, and definition (63) describes the overprediction of the three-feature segment /v/.

4.5.4 Underpredictions

Absent from the findings are underpredictions: feature combinations that are attested, yet not predicted by the set of features and constraints. FCCs allow *at least* the feature combinations that are actually attested. The absence of underpredicted segments follows directly from the procedure by which FCCs are constructed.

A segment can be underpredicted by virtue of either a c-constraint or an i-constraint. In the first case, a segment [F, G] is attested and yet a constraint *[FG] is activated. C-constraints, however, are generated based on a two-dimensional feature co-occurrence matrix where all possible combinations of two features are indicated as attested or unattested (see section 4.2 above). For every unattested feature combination in this matrix, a c-constraint is activated, but not for every attested combination of two features. For underprediction to occur, attested feature combinations must be ruled out. It is clear that the procedure used to derive c-constraints cannot do this.

The other possibility for underprediction is by an i-constraint. In this case, the constraint $F \rightarrow G$ is activated while F is attested to co-occur with at least one other feature H, where $G \neq H$.

The procedure for deriving i-constraints starts out from attested combinations of two features. Hence, it would seem that there is a higher risk for underprediction (as underprediction entails that an attested combination is ruled out). Every possible i-constraint that refers to an attested combination of two features is candidate for activation. Next, for each of these pairs, every other attested combination is checked. If the antecedent occurs in combination with any feature H where H≠G, the i-constraint F→G is no longer a candidate for activation. Hence, the procedure used to arrive at the set of activated i-constraints is unable to yield underprediction – just as the algorithm cannot derive underprediction by c-constraints.

4.5.5 Summary

Although the theory of Feature Co-occurrence Constraints largely yields correct results, we did encounter some examples of overpredicted segments – segments that are allowed by the set of features and constraints at some stage, but not attested. The set of overpredicted segments is very small, compared to the set of possible feature combinations. We discussed several possible causes for overpredictions, where the most important ones were a) a strong prediction that /t/ should be acquired at the first stages, and b) overpredicted segments are often subsets of attested segments (this is, of course, trivially true of /t/, as it is the interpretation of an empty segment). This brought us to a set of formal definitions of the contexts in which overprediction occurs. Finally, the reason that underpredictions do not occur follows from the constraint derivation algorithm. Before concluding the current chapter, let us consider the developmental origin of Feature Co-occurrence Constraints.

4.6 The origin of Feature Cooccurrence Constraints

The theory proposed in this thesis has two main ingredients: features and constraints. In section 2.3, we have investigated the question of the innateness of distinctive features from the perspective of categorisation, lexical storage and phonological rules and generalisations, and in each case, we have seen that the evidence supports the idea of innate features. In this section, we will discuss the innateness of constraints.

We must differentiate between different notions of innateness. One possibility is to claim that humans come into this world with a predetermined, universal and substantive set of constraints (CON). This is the position we find in the 'classical' OT literature (Prince & Smolensky, 1993/2004). On the other hand, it is possible that constraints are formed during acquisition. Given this position, two scenarios are possible: first, the child may posit whatever constraint she finds evidence for, and second, instead of an innate set of substantive constraint, the child comes into the world endowed with a limited set of innate constraint *templates*, that are actuated and/or populated based on experience with the surrounding language.

In this section, we will begin with an overview of the attitudes toward the developmental origin of constraints in the literature, focussing mainly on the most well-known constraint-based framework in phonology, Optimality Theory. Surprisingly little experimental research has been done with regards to this question, with Jusczyk et al. (2002) as a notable exception. We will consider their results and methodology in some detail. Finally, we will see that there are reasons to distinguish between constraints concerning melody, and those concerning prosody. The arguments for substantive innateness are much stronger in the case of the latter.

4.6.1 Ideas and Assumption in the Literature[10]

In the original conception of Optimality Theory, constraints were proposed to be both universal and innate. The *only* aspect of grammar that was thought to be language specific was the constraint ranking. In this respect, OT signaled a departure from generative phonology in at least two ways. First, as we shall see, constraints in earlier generative phonology (morpheme structure constraints, surface phonetic constraints) were language specific and learned. Secondly, constraints were originally invoked to *limit* the generative potential of the grammar. In OT, constraints *define* the generative potential.

Prince and Smolensky (1993/2004) are very clear in stating that the goal of the Optimality Theoretic program to provide a set of universal constraints. This ambition is made possible by the architecture of the theory; in previous thought, a conflict existed between the ambition to make universal statements and the requirement of these statements to be 'surface- or level true'. In Optimality Theory, the latter requirement no longer plays a role, because the theory provides a constraint-external mechanism of conflict resolution: strict domination. In Optimality Theory, constraints can be universal because they are *always* true: if a constraint X does not correspond to some output structure in language A, this does not mean that X is falsified by facts about A, but rather that in A some constraint Y takes precedence. The upshot is that if the theory is going to make predictions, it *must* assume substantively universal constraints; if language-specific constraints are admitted, overgeneration and post-hoc argumentation take over. Prince and Smolensky (1993/2004) do not specifically mention innateness of constraints but seem to imply that they are; the acquisition of grammar in OT is equal to ranking constraints, and a learnability problem arises if constraints must be learned, and then ranked – especially it they are to be ranked low (low ranking entails little influence and hence little overt evidence). Innateness of constraints (or constraint templates) is also the premise in later work on learning algorithms in OT (Boersma & Hayes, 2001; Tesar & Smolensky, 2000).

In addition, a subset principle violation poses a problem for emergent constraints. A well-known position in the learnability literature is that children learn only from positive evidence – that they make no generalisations based on the *absence* of structure. Hence, the most conservative grammar must corre-

[10]I would like to express my gratitude to Edoardo Cavirani, for parts of the following section originated in discussions with him and in attending his presentation (Cavirani, 2012)

spond to the initial state. The development of the grammar is informed only by positive evidence. This idea, roughly sketched here, is known as the 'subset principle' (first formalised in phonology by Dell (1981)). It is easy to see that the subset principle is difficult to combine with emergent constraints (that is, non-innate constraints that are emergent in the child's developing grammar). Constraints are, generally speaking, negative statements, banning structures that are not allowed, like NoCoda, which disallows codas. If the language surrounding the learner has only CV syllables (at the surface level), there is no way for the constraint to emerge in the child's grammar – even though the evidence favouring it is abundant. A grammar without NoCoda is perfectly capable of generating CV syllables, and the lack of coda consonants in the child's input or uptake could just as well be an accident. Conversely, consider the same learner but now surrounded by a language in which CVC syllables are quite frequent. Here, too, there is no reason for the child to posit a constraint NoCoda, because codas are allowed. See, for a similar argument, Hale and Reiss (2003), who explain the issue in clear terms. The problem is that lower-ranked constraints in OT can still be active sometimes, a process known as The Emergence of the Unmarked (TETU): unmarked structures sometimes arise in defiance of higher-ranked faithfulness constraints, if there is nothing to be faithful to (e.g., in reduplication, loan-word adaptation, etc.).[11]

For emergent constraints to work, then, there must be some external way to motivate their presence in the grammar. One way to do this is with a 'phonetic difficulty map', as has been proposed by Hayes (1999). Here, the learner maps the articulatory effort in phonetic space, and assigns a numerical value to each phoneme, corresponding with its 'difficulty'. Next, potential feature co-occurrence constraints[12] are profusely hypothesised by the learner, and then evaluated with respect to the segments on the map. So, for a constraint C and any two segments X and Y, four possibilities exist:

(64) a. Both X and Y violate C
b. Both X and Y satisfy C
c. X violates C, Y satisfies C
d. X satisfies C, Y violates C

The first two cases are uninteresting, as the outcome is the same for both segments. The important comparison is in the latter two situations. Roughly

[11] A way out in Optimality Theory could be to resort to the principle of Richness of the Base: since any conceivable output form is a candidate for any given input, GEN will propose forms that have a coda. The fact that these never surface could be interpreted as a form of positive evidence. Note that for this argument to hold the candidates with coda must not be harmonically bound by the forms without in any other way, as this would destroy the evidence for NoCoda. Another issue is that such a proposal to work still needs to refer to innateness, if not in CON, but in GEN: the learner must still consider the candidate with coda as a candidate, and hence, in the face of no surface evidence, know that there is such a thing as a 'coda' – as per the *innateness of primitives principle* (see section 2.3.

[12] Hayes (1999) proposes both paradigmatic and syntagmatic feature co-occurrence constraints.

speaking, if X violates C and Y satisfies C, C is a ‘good’ constraint if X is more difficult then Y. Thus, in this example, if the situation in (64c) holds, *and* Y is easier then X, potential constraint C makes a correct prediction. However, if X is easier then Y, the prediction is incorrect. For any pair of two segments, predictions are made and subsequently, the ‘goodness’ of constraints is determined by the following metric:

(65)

$$Effectiveness = \frac{Correct\ predictions}{Correct\ predictions + Incorrect\ predictions}$$

This yields an index of ‘groundedness’ for all constraints. All potential constraints are mapped in a constraint space, whereby constraints whose structural description differs minimally are neighbours. In this multi-dimensional space, local maxima of constraint effectiveness are determined, and the corresponding potential constraints are those that will constitute the grammar. In this way, Hayes (1999) sketches how constraints can be emergent. The necessary ingredients (or presupposed knowledge) for his learner are a) knowledge of phonological features, b) their phonetic content and c) the (relative) difficulty involved in expressing these features in different contexts. Hayes’ proposal thus presents us with a less severe version of the duplication problem: even though articulatory knowledge is not directly encoded in the grammar, it *must* be directly encoded in the algorithm that gives the grammar substance. The subset problem is circumvented, however, because the evidence for constraints is not the absence of certain pattern (i.e., negative evidence), but stems from the knowledge the learner has about relative phonetic difficulty.[13]

The constraints proposed by Fikkert and Levelt (2008), discussed in section 2.3.2, are not proposed to be innate either, but in contrast to Hayes’ inductive grounding, they rather emerge as generalisations over the lexicon. This is illustrated with the production patterns of Noortje, who has initial dorsals initially, then substitutes them for coronals, and only later reverts back to faithful productions of dorsals in C_1. These generalisations, being generalisations over the lexicon of words that the child has stored after being exposed to them, reflect input frequency. Thus, the authors are fairly explicit in the way in which they allow input frequency to play a role in acquisition. As is often the case when invoking input frequency, however, the argument runs the risk of being circular: if learners posit grammatical devices (constraints) to formalise generalisations over the input, it could be expected that these devices will leave a trace in the adult language. In fact, first language acquisition is often hypothesised to be the mechanism of language change!

[13]It remains a question whether the proposal in Hayes (1999) is in fact more parsimonious then a theory in which all constraints are innate and ineffective constraints are quickly demoted, as every possible constraint must be generated and evaluated for its effectiveness. The difference is that for Hayes, only effective constraints are admitted to CON.

A more serious problem for emergent constraints of the type proposed by Fikkert and Levelt (2008), as discussed above, is that they are not learned from positive evidence: it is the absence of non-coronals in initial position that leads the child to posit [LABIAL, and later the absence of dorsals in that position to posit *[DORSAL. Superficially, it would appear that the same problem haunts the analysis proposed in the current thesis, but the learning datum for Feature Co-occurrence Constraints, is a new segment; if for that segment a new feature must be activated or acquired, it is accompanied by FCCs to keep the grammar as conservative as possible. In contrast, the learning datum for Fikkert and Levelt (2008) is introspection and generalisation over the lexicon.

4.6.2 Experimental evidence

One of the few studies that explicitly tackles the question of constraint innateness from an experimental point of view is Jusczyk et al. (2002). Here, the cross-linguistically highly frequent process of nasal cluster place assimilation was the subject of investigation. Children were exposed to triads of non-words of the form X... Y... XY, where XY is a concatenation of X and Y. X has a nasal coda, and Y has a obstruent onset. In this way, the effects of markedness (cluster assimilation) and faithfulness (no assimilation) could be tested. An example of a non-assimilating triad would be *on... pa... onpa*, an assimilating triad would be *on... pa... ompa*.

In the first experiment, 10 month olds were tested to see if they have a preference for either faithful or unmarked outputs. In this version, all stimuli were non-assimilating, but some concatenations were marked, and some were unmarked. An example of a marked triad is *un... ber... unber*, whereas its unmarked counterpart is *um... ber... umber*. Hence, although all triads represented faithful concatenations, half violated the markedness constraint agains non-homorganic nasal-obstruent sequences. The children showed a preference for the unmarked triads, indicating that they display evidence of a markedness constraint.

Experiment 2 sought to investigate whether 10 month old infants have faithfulness constraints, as well. To achieve this, the stimuli were manipulated so that half contained faithful concatenations (*um... ber... umber*), whereas the other half consisted of triads in which the latter member was phonologically unrelated to the former two: *um... ber... iŋgu*. An added benefit of this set-up was that it checked whether the results in experiment 1 were not due to a general novelty preference, and indeed, the children preferred the faithful sequence. Finally, in experiment 3, faithfulness and markedness were pitted against each other. One type of stimuli violated markedness but respected faithfulness (*un... ber...* ***unb****er*), whereas the other violated faithfulness but respected markedness (*un... ber...* ***umb****er*). As expected, the 10 month olds showed a preference for the alternating, unmarked sequences. In order to rule out the possibility that the results were due to native language knowledge (10 month olds are sensitive to phonotactic properties of the environment language

(Friederici & Wessels, 1993; Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993)), the same set of experiments was repeated with $4\frac{1}{2}$ month olds. The results were identical, indicating that native language experience may not have been the decisive factor.

According to Jusczyk et al. (2002), these results are consistent with a nativist OT account, in which constraints are substantively innate, and markedness constraints outrank faithfulness constraints in the initial state. With respect to the framework that we are developing, this implies that Feature Cooccurrence Constraints are innate too, and what is more, that they are all initially high-ranked.

However, the interpretation of Jusczyk et al. (2002) crucially relies on the assumption that the infants actually parse the stimuli as X... Y... XY, rather then X... Y... Z. In other words, it depends on whether children take the latter member of the triad to be derived from a concatenation of the earlier two remains uncertain. Hence, the results may not be as conclusive as they are presented to be.

4.6.3 Two types of constraints

In their discussion about emergent constraints, Fikkert and Levelt (2008) propose that some constraints might be innate, whereas others are not. The division, they propose, might coincide with the domains of (sub) segmental phonology on the one hand, and suprasegmental phonology on the other. That is, constraints concerning features are emergent (possibly as generalisations over the developing lexicon), whereas constraints governing prosodic structure might be innate. One possible argument for this is that cross-linguistically, a considerable variation exists in (the phonetic expression of) phonemic categories, while the typology of prosodic structures appears to be more restricted (to such a degree that some deny its existence *per se* (Scheer, 2004, for example) – but this goes beyond our present concerns).

This idea is also found in Inkelas and Rose (2008). Inkelas and Rose describe the phonological development of E., a child learning English. The two phenomena described are Positional Lateral Gliding and Positional Velar Fronting, two cases of what the authors call 'strong merger': neutralisation of a contrast in a prosodically strong position. This is quite unusual from the perspective of adult phonologies: both Positional Faithfulness (J. Beckman, 1998) and Positional Markedness (Lombardi, 1995; Zoll, 1998) are designed to capture the generalisation that positional effects in adult languages are Weak Merger (positional neutralisation in weak positions) and Strong Enhancement (making strong positions more salient). Weak Merger is taken to be the result of Positional Faithfulness: certain faithfulness conditions may only apply to strong positions, whereas Strong Enhancement is taken to be affected by Positional Markedness: certain contrasts may only appear in strong positions. Also, positional neutralisation rarely targets primary PoA. In short, if neutralisation is positionally limited, it is limited to weak positions. Hence, Strong Merger poses a challenge to the Continuity Hypothesis.

In Positional Velar Fronting (PLG), underlying velars are produced as coronals in the onset of words and stressed syllables. In Positional Lateral Gliding (PLG), all laterals are glided. Laterals in onsets of words and stressed syllables become palatal [j], laterals in weak positions become [w]. Some examples are given in (66) (representing a subset of the examples given in Inkelas and Rose (2008, ex. 1):

(66) Strong Merger in E.'s speech

a. Positional Velar Fronting:

[tʰʌ]	'cup'	1;09.23
[ˈtuwɔ]	'cool	' 1;11.02
[əˈdɪn]	'again'	1;10.25

b. i. Positional Lateral Gliding: strong positions:

[jæmp]	'lamp'	1;10.0
[jɪdiə]	'Lydia' (adult: [ˈlɪdiə])	2;1.8
[jivan	'Livan' (adult: [liˈvan])	2;8.19

ii. Positional Lateral Gliding: weak positions:

[hæwətʰʌəkɚ]	'helicopter'	1;11.10
[æwədɛɾɚ]	'alligator'	2;1.18
[hɨwdə]	'Hilda'	1;11.10
[bejgu]	'bagel'	1;9.24

c. PVF and PLG interacting:

[ʧiːn]	'clean'	1;11.0
[ˈʤæsəs]	'glasses'	2;2.1

Positional Velar Fronting is a systematic part of E.'s speech for a fourteen-month period beginning early in his second year, whereas Positional Lateral Gliding starts eight months later. Hence, during some period, the two processes interact, providing further evidence for the hypothesis that the processes are productive parts of E.'s phonological grammar (see (66c)). What we see here is that in target /Kl/ onset clusters, PVF ensures that the dorsal is fronted, whereas PLG glides the lateral. The resulting /Tj/ cluster is then merged to a single affricate. Interaction and merger (palatalisation) does not occur in all cases; in roughly one third of the eligible clusters, the lateral is simply omitted.

The apparent contradiction to the Continuity Hypothesis posed by Strong Merger is solved by Inkelas and Rose (2008) by appealing to the notion of 'phonologisation'. Phonologisation refers to two related phenomena: first, it can mean the diachronic adoption into the phonological grammar of a phenomenon that was already phonetically present; second, it can mean the adoption (into the phonological grammar of individual children) a phenomenon that was already phonetically present, applying rather in the time scale of individual language acquisition. In the case of E., what this means concretely is that he is aware of the fact that strong positions require enhancement of articulatory gestures (in other words, he is aware of 'syntagmatic contrast'). Laterals contain the orchestrated movement of two articulators: tongue tip and tongue body. The relative prominence of these two movements is how clear (tongue tip) /l/ and dark (tongue body) /l/ are distinguished. In gliding his laterals, E. retains this double identity of laterals, resulting in a coronal (tongue tip) glide in the positions where clear /l/ is usually found, and a labio-velar (tongue body) glide in positions where we usually encounter dark /l/. These positions, of course, coincide with strong and weak positions respectively. We can see that E. has phonologised this pattern when we consider intervocalic positions: pretonic intervocalic consonants are glided to [j], whereas post-tonic laterals are glided to [w].

In PVF, E. also remains faithful to the articulatory gestures involved in producing velars. However, children's production of stops is less refined than adults' productions, and physiologically, their tongue body is much larger with respect to the oral cavity. These two things result in greater difficulty producing velars. In addition, in English, stops in strong position are articulated more extremely than stops in weak positions. Furthermore, velar stops in strong positions are somewhat more fronted. The combined result of these factors is that "... [i]n the context of imperfect articulatory control, bigger tongue size, when combined with a relatively shorter palate, implies that even a slight increase of vertical tongue movement, required in the enhanced articulations in prosodically strong positions, will have direct consequences for the childs production of target velars. The greater emphasis on the dorsal articulator expands tongue contact into the coronal region, yielding the coronal release that characterises fronted velars." (Inkelas & Rose, 2008, p. 724)

Similar cases have been reported. For example, Chiat (1989) reports on a child who stops fricatives word-initially and in the onset of strong syllables word-medially, but produces them faithfully elsewhere.[14] Similarly, in a report on two experimental studies on the consonantal substitution patterns in different positions by English learning children with delayed phonological development, Rvachew and Andrews (2002) note a general pattern that '... consonants that occur at the beginning of stressed, word-internal syllables will be produced in the same manner as consonants that occur at the beginning of words', that the syllable final position patterns alike regardless of position in the word, and, with respect to fricatives, for two of the three children the intervocalic (before an unstressed nucleus) position and the syllable final position patterned alike. Although Rvachew and Andrews (2002) note that substitution patterns are generally alike for syllable initial positions, regardless of the adjacency of a word boundary, and that the same holds for syllable final positions, they treat intervocalic consonants as a distinct category ('ambisyllabic'). In other words, the children's consonantal substitution patterns were influenced by foot structure.

In his study on truncation in child language, Pater (1997) underlines the importance of recognising the syllable as a unit in the developing phonology,

[14]The same child also exhibits positional velar fronting, see Chiat (1983).

next to the foot. The general truncation pattern for trisyllabic (and longer) words in English and Dutch (Fikkert, 1994; see Pater, 1997 for further references) is for the stressed and final syllable to survive. Together, they satisfy a minimal word requirement, where the minimal prosodic word is a trochaic foot. For example, Pater reports on examples where *cinnamon* truncates to [sɪmɛn], *Allison* to [æ:sʌn], and *museum* to [zi:ʌm]. However, there are forms that deviate from this pattern: *banana* can become both [nænə] and [bænə], *marakas* becomes [ma:kas], and *delicious* becomes [dɪʃəs]. The crucial point is that the onset of the stressed syllable is replaced by the onset of the syllable that is otherwise deleted, when the latter is of lower sonority. These cases, then, indicate that children are aware of the relation between sonority and syllable structure.

The crucial point in these studies is that the learners are very much aware of prosodic structure and are willing to risk sacrificing segmental contrast in order to express syntagmatic contrast (prosodic strength). Here, we see an early activity of prosodic well-formedness, at the expense of (sub)segmental faithfulness.

4.6.4 About the innateness of Feature Co-occurrence Constraints

The arguments reviewed so far are equivocal; on the one hand, there seems to be good reason to oppose the idea of innate Feature Co-occurrence Constraints – even if features themselves might be innate. A hint in this direction was made by Levelt and van Oostendorp (2007) and Fikkert and Levelt (2008). Substantial backing for the idea that segmental constraints , as opposed to prosodic constraints (Inkelas & Rose, 2008), need not be innate is provided by Hayes (1999). The two domains are very different, where the suprasegmental domain is more restricted then the sub-segmental domain. This intuition is also captured by Scheer (2004), who divides phonology in UPPER and LOWER. Relations in UPPER are much more restricted: government and licensing are both bound by directionality, for example. Although this difference in restrictedness is not in itself an argument about innateness or emergence, it does show that there is an important difference.

Emergent constraints do not fare well in light of the subset principle. Constraints are statements on what is *not* allowed, but by definition, there is no positive evidence for illegal structures. Note that this argument is independent of the argument about constraint ranking (or deactivation). As soon as there *are* constraints, positive evidence must be available to determine their place in the grammar. Both extreme positions about constraints, namely substantive innateness and absolute emergentism are not viable, then – at least for constraints governing sub-segmental phonology. Hence, the third option, of constraint templates, remains as the best option. Note that the issues posed by the subset principle do not categorically rule out emergent constraints, as long as some independent mechanism can be found to generate them. This is the route we will adopt here, because in light of the studies and arguments cited above, it would seem that there is no evidence for innate Feature Co-occurrence Constraints.

If the set of Feature Co-occurrence Constraints is not innate, then, we must somehow account for the limitations of their form. Only constraints referring to two features are allowed, and only two connectives occur. The motivation of these limitations were discussed in chapter 3, but for now, I propose that rather than an innate set of substantive constraints, the child is equipped with a much smaller set of innate constraint templates.

4.7 Summary

The theory of Feature Co-occurrence Constraints and the segment inventory as developed in chapter 3 is shown to account for the acquisition of the segment inventories of different children in the present chapter. Acquisition is modeled as the succession of stages, defined by an addition to the inventory. Although these stages are not independent from the perspective of the features that are acquired (i.e., features are acquired in a monotonically increasing manner), they are independently approached by the constraint derivation procedure. Even though the procedure has no memory of what constraints it generated at a previous stage, we saw that the assumption that the activation time of constraints is severely limited is borne out fully: all constraints are activated no later than the moment at which the second feature that they refer to is acquired. More importantly, the algorithm converges on those parts that do not change from one stage from the next: a gap described by constraint C remains a gap described by constraint C.

The demonstration of Noortje's development showed that the theory is largely correct in describing the inventory in terms of features and constraints on feature combinations. However, the theory overshoots at some points: a total of six different segments were found to be overpredicted in the data from the children included in the study. One of these, /t/, cannot be excluded, and hence, it must be present from the first stage; either as an attested segment, or as an overprediction. For the other segments, /p, f, s, v, z/, the contexts in which they can be overpredicted were defined. However, overpredictions were found to be largely an artefact of the inclusion criteria for segments to be accepted as 'acquired'; a look at the unfiltered initial data revealed that only a number of cases of overpredictions actually concern segments that are not produced during the recording sessions. Underpredictions are not attested, which was attributed to properties of the constraint derivation procedure.

Finally, we turned to the question of innateness of constraints, a matter which is separate from the question of innateness of features (see section 2.3). We found that there are good reasons to assume that constraint templates, but not substantive constraints, are innate.

All in all, the model was able to account for the attested inventories, for each child and for each stage. This means that an important prediction, namely that Continuity holds at the level of the structure underlying the inventory (constraints and features), is borne out.

4.7.1 Final thoughts

The question of how the acquisition of the segment inventory can be modeled in a satisfactory way has divided researchers for a long time. Approaches that are based on a strong continuity with the adult grammar and utilise distinctive features have great appeal. However, some feature-driven theories of phonological acquisition are too restrictive. Jakobson (1941/1968), for example, proposes a universal order of acquisition based on oppositions between features. Such a universal order has, however, never been uncovered. On the contrary, considerable variation in order of acquisition exists both between and within languages. At the same time, as pointed out by Levelt and van Oostendorp (2007), some theories are too permissive. One example is Beers (1995), who describes the order of acquisition in terms of the unfolding of a feature geometry, where mother nodes are acquired before daughter nodes, and variation is restricted to the relative order of acquisition of sister nodes. Importantly, however, the theory makes no attempt at restricting the combinatorics of features once they have been acquired: every feature is free to combine with others once it is part of the child's system. Gaps in the segment inventory remain unaccounted for, so at best, the theory is incomplete.

Non-feature-driven theories have been put forward, in part as answer to the difficulties described above. Some eschew the notion of distinctive features altogether (exemplar theory), others hold that phonological features are emergent properties of an ever more densely populated lexicon (lexical diffusion theories). Each make specific predictions with respect to the acquisition path.

Exemplar Theory (M. Beckman, Yoneyama, & Edwards, 2003; Zamuner, Gerken, & Hammond, 2005) holds that children acquire their language based on whole forms – words, segments – rather than abstract categories such as distinctive features (various exponents of Exemplar Theory differ in the degree to which they allow abstraction to take place later in development). An important aspect of the theory is that acquisition proceeds by general (acoustic, statistical) processing of the input; i.e., there is no language specific competence. This entails that input frequency is an important, if not the sole, predictor of the path of acquisition. For example, Gonzalez-Gomez, Poltrock, and Nazzi (2013) find that children acquiring French learn CVCV sequences at an earlier stage when C1 is labial and C2 is coronal, than when the PoA specifications are the reverse. This, they argue, is due to the higher frequency of labial-coronal than coronal-labial words in French.

Other studies, too, have found evidence in favour of input frequency effects on phonological acquisition, but many have failed to do so. Most important to our present subject, Levelt and van Oostendorp (2007) found no correlation between the order of acquisition of Dutch consonants and the relative frequency in a corpus of Dutch child-directed speech (van de Weijer, 1998).

Apart from a lack of empirical evidence, the frequency approach suffers from a number of other problems. First of all, it is entirely unclear what the precise relation is between input frequency and acquisition. Even if a correlation is found, cause and effect are rarely scrutinised. Does a higher frequency cause earlier acquisition, or is an independent learning bias the cause of both early acquisition and higher frequency in the adult language (see, for a rare comment on this issue, Fikkert and Levelt (2008))? What is the relevant measure of frequency? What constitutes a relevant corpus? Furthermore, the frequency approach appears ill equipped to deal with individual variation, such as the variation in acquisition order encountered in the current study. The tacit assumption is that large corpora provide an accurate representation of the input frequency for a given language community. However, in order to account for individual paths of acquisition, individual input corpora should be used (see again Fikkert and Levelt (2008)).

In theories of lexical diffusion, again the unit of acquisition is the word, rather then the feature. Lexical items are subject to finer degrees of specification (word $\rightarrow$ segment $\rightarrow$ feature), the more the lexicon becomes populated with 'neighbours': words that are highly similar (see, for example, M. Beckman and Edwards (2000)). Again, Levelt and van Oostendorp (2007) found no evidence for lexical diffusion, in the sense that the transition to correct production of a segment is independent of the word it is in.

The current theory adopts the classical, feature-based, Jakobsonian view of a high degree of continuity from child to adult language (see also chapter 3. Whereas earlier theories focused on the order of acquisition of segments, contrasts or features (Jakobson, 1941/1968; Beers, 1995), and failed in the sense that they were either too stringent (Jakobson) or too lenient (Beers), the primacy of the *order* of acquisition has been abandoned here – even thought it is possible to predict impossible orders. Rather, by focussing on the *mechanism* employed by learners of a language, we have developed a theory that accurately describes the segment inventory at every stage, for different individual children, maintaining continuity at both the level of the material (features) and the mechanism (constraints).

CHAPTER 5

Conclusion and discussion

In the previous chapters, the Feature Co-occurrence Constraint Theory and its constituent assumptions about features and constraints were developed, and an application was illustrated. The theory is by no means complete, however, as many factors of it remain unexplored. Such is inevitable, but in this chapter, we will address a number of remaining questions. Section 5.2 aims to sketch a number of outlooks on how Feature Co-occurrence Constraint Theory fits in the contemporary phonological landscape, and how it may complement existing theories. We will look at the same frameworks that we discussed in chapter 2: Inductive Grounding (Hayes, 1999, section 5.2.1 below), Parallel Bidirectional Phonetics and Phonology (Boersma & Hamann, 2008, section 5.2.2 below), and the Modified Contrastive Hierarchy (Dresher, 2009, section 5.2.3 below). Section 5.3, finally, discusses some residual issues. For example, we have hardly touched upon perception and the perception-production relation in the previous chapters. Far from developing an answer to the problems raised in relation to perception, section 5.3.3 aims to at least outline the issues. Also, we have repeatedly mentioned that our constraints are compatible with both Optimality Theory and strict violation frameworks, but we have deferred much of the discussion (although in section 3.3.2 we did demonstrate that this is the case). Section 5.3.2 takes up this issue further. Finally, there is the question of how emerging constraints relate to the hypothesis that children learn only from positive evidence. A brief discussion of this is presented in section 5.3.1. First, however, section 5.1 briefly summarises the theory and the main conclusions.

5.1 Summary of the main findings

This thesis was devoted to developing a minimalist theory of the consonant inventory, and how it is acquired. The point of departure is that phonology functions as an addressing system: it assigns a unique representation to lexical items. Perception and production consist of mapping perceived surface forms to these underlying representations, and *vice versa*. This seems an uncontroversial view, and from it, we derived that ideally, a theory of the inventory should not be 'holistic', where the term is taken to mean that the entire inventory must be assessed in online computation, rather than merely the segments present in surface and/or underlying forms.

The system that is proposed consists of features (and some temporal ordering mechanism such as root nodes, ×-slots or the like), an unspecified generator function that proposes feature combinations (segments), and output constraints on feature combinations. We found evidence for the innateness of features in chapter 2, but not to the same degree for constraints. Hence, we assume that while features are innate, segmental markedness constraints such as our Feature Co-occurrence Constraints are emergent. These Feature Co-occurrence Constraints come in two types:

(67) a. *[F, G]
assign a violation mark for every segment Σ iff [F] is in Σ and [G] is in Σ (*c-constraint*)
b. [F→G]
assign a violation mark for every segment Σ iff [F] is in Σ and [G] is not in Σ(*i-constraint*)

With these in hand, we demonstrated in chapter 3 that the consonant inventory of Dutch can be described with only a limited set of constraints.

Feature Co-occurrence Constraints are exactly binary in their reference (i.e., the constraints can refer to no more and no less than one feature), although single-feature constraints can be derived: *[F, F][1] This design characteristic is motivated by reference to the non-recursive nature of phonological computation, and it was shown that logically, the set can be described with the logical connective AND and the negation operator NOT.

The choice of constraint types is intimately linked to the feature system employed. For example, i-constraints are necessary in monovalent feature theory, because it is impossible, with monovalent features, to representationally express the complement set of the set denoted by the presence of a feature. This is possible in binary feature theory, where the complement of a set denoted by feature [+F] is simply labeled [-F]. Monovalency appears to be a valid, if not preferable, option.

Furthermore, we assumed that not all phonological traits are represented by distinctive features. Following, among others, Fikkert and Levelt (2008) we

[1] The corresponding i-constraint [F]→[F] is always vacuously satisfied.

specifically assumed the non-specification of coronality and non-continuancy. The featurally empty segment cannot be ruled out, and must thus receive a phonetic interpretation. In this case, the interpretation is that of /t/. Major class features are also rejected, as is Feature Geometry. Instead, we adopt Feature Classes, a non-hierarchical system of expressing natural classes in terms of features. Feature Classes allows us to do away with major class features, without also dispensing the Feature Geometric insight that features are of different types (i.e., place, manner, *et cetera*).

In chapter 4, we applied the Feature Co-occurrence Constraint Theory to the acquisition of the Dutch consonant inventory. Feature Co-occurrence Constraints are posited to emerge automatically and no later than at the point in acquisition when the child's system reaches the criterion that both features in the constraint's structural description are activated. This is equivalent to saying that whenever a new feature F is activated, it is automatically accompanied by a set of constraints *[F, Φ], [F]→[Φ] and [Φ]→[F], where Φ stands for any other feature in the child's system. This assumption is not contradicted in the data: no constraint was introduced later. This assumption allows the child to remain maximally restrictive in her acquisition and respond only to positive evidence: the presence of a segment in the language she is acquiring can trigger her to adopt it; the absence of a segment cannot (and need not) trigger her to configure her grammar such that it is excluded, because in principle, *everything* is excluded.

It was predicted that continuity would hold at a structural level: even if not every child inventory coincides with a typologically attested inventory (especially at the earliest stages, where child inventories are generally quite small), it is generated by the same mechanism: privative features, and a system of two constraint types: i-constraints and c-constraints. This was indeed what was found.

In most cases, the constraints that were derived predicted an inventory of possible segments that coincides exactly with the attested inventory, but in a number of cases, the constraint set was too permissive. Only six different segments were ever overpredicted in the data set involving longitudinal recordings of seven Dutch monolingual children. However, many of these cases were only apparent overpredictions: careful re-examination of the raw data revealed that in most cases, the overpredicted segments *were* present in the attested inventory, but were not included in the analytical sample due to not reaching criterion (e.g., they were not produced often enough, or not in enough different lexical items). Hence, overprediction is in part an artefact of the sampling method.

Some real cases of overpredictions occur, however, but only involving four different segments. One of these is /t/, which in our system is represented as a featurally empty segment. By virtue of not having any featural content, no FCC can ever forbid /t/, and hence, it is predicted to be in the earliest inventories. This prediction is borne out in almost every case, but in some cases (particularly in the inventory of one child, Noortje, whose data are notably different than that of the other children), it is not.[2] Two other contexts in which overpredictions occur were identified, both involving in-excludable subset segments.

After their introduction, constraints are divided in two groups, according to whether they are violated or not. This binary division, due in Optimality Theoretic terms to lack of further ranking arguments (when no other constraint families are considered) allows the theory to be compatible with both OT and non-ranking phonological frameworks. In the next section, we will discuss how the Feature Co-occurrence Constraint Theory relates to a number of different theories, and in section 5.3 we will come back to the question of implementation.

5.2 Compatibility

Throughout this thesis, we have made reference to a number of other theories about the shape and the structure of the inventory. These theories, however, turn out to be not all competing theories. In this section, we will briefly investigate whether a symbiosis of Feature Co-occurrence Constraint Theory and three other frameworks is feasible, and if so, whether pursuing it further could be a beneficial enterprise. We will begin with Inductive Grounding (Hayes, 1999), followed by Parallel Bidirectional Phonetics and Phonology (Boersma & Hamann, 2008), and we will conclude with the Modified Contrastive Hierarchy (Dresher, 2009). These frameworks were chosen earlier (see chapter 2 because they are among the few that are specifically about the inventory, and because they illustrate issues concerning the shape (Dispersion Theory, Parallel Bidirectional Phonetics and Phonology) and the structure (Modified Contrastive Hierarchy, Parallel Bidirectional Phonetics and Phonology) particularly well. Most importantly, however, they are theories with particular concern for learnability and acquisition.

5.2.1 Feature Co-occurrence Constraints and Inductive Grounding

One of the earliest explicit theories about feature co-occurrence constraints in an Optimality Theory setting we find in Hayes (1999). As we have seen in chapter 1, Hayes' aim was to replace innate markedness constraints on feature combinations (and sequences) by a set of principled and phonetically 'grounded' emergent constraints. In this section, we will briefly discuss whether there is any degree of compatibility between Hayes' Inductive Grounding and Feature

[2]A different possibility is that in those cases where /t/ is overpredicted, the child actually has ∅ in her grammar, but the empty segment receives a different interpretation. This could be due to two factors: first, articulation is immature and imprecise at such an early age, and second, the *absence* of phonological material in the empty segment leaves room for variation in its expression, especially in an inventory with few contrasting segments. We will not pursue this option further at this point.

Co-occurrence Constraint Theory, and if so, whether the combination is worth pursuing.

The aim of the Inductive Grounding program can be summarised as follows. Grammar(s) tend to strike a balance between functional motivation and formal simplicity. It would be a mistake, Hayes notes, to derive phonological mechanisms (such as constraints) directly from the phonetics (Hayes focuses on articulation, and the difficulty thereof), because grammars often deviate from a perfect phonetic fit. The important point is, Hayes argues, that *if* a grammar (or a language, for that matter) deviates from phonetic fit, it does so in the direction of formal simplicity. Hayes gives the example of voicing in labial obstruents; generally speaking, voicing is easier to maintain the further forward the place of articulation (this is the mirror effect of the typologically frequent ban on voicing in dorsal obstruents; Dutch is an example of the latter ban). Hence, Egyptian Arabic has a gap in its inventory, */p/ while /b/ is allowed. At the same time, voicing is more difficult to maintain in geminate obstruents than in singleton obstruents. This is reflected in the phonology of Japanese, which bans any voiced geminate, including */bb/, whereas /pp/ is permitted. An interesting juxtaposition arises between Egyptian Arabic and Japanese when it comes to geminates: in the former, the ban is reversed. In Egyptian Arabic, /bb/ is allowed by virtue of involving the voiced labial stop /b/, whereas */pp/ is banned for involving the illegal voiceless labial stop. Hence, Egyptian Arabic deviates from what would be phonetically 'better', namely, a ban on /bb/, but this deviation is formally motivated by being an extension of the functionally motivated ban on voiceless labials *per se*.

Inductive Grounding provides an algorithm for selecting the phonetically grounded constraints in the space of all formally possible constraints (we have discussed the algorithm in some detail in section 3.2.3 above). A comparison with Feature Co-occurrence Constraint Theory yields a number of similarities and differences.

Emergence In both theories, constraints are not innate but rather learned or activated. Feature Co-occurrence Constraint Theory explicitly states, in addition, that constraints templates *are* innate; Hayes (1999) makes no such explicit claim, but appears to adhere to the same perspective.

Optimality Theory In section 5.3.2 below, we will address the question of what framework Feature Co-occurrence Constraints are most suitable for. Both OT and non-OT frameworks can be combined with Feature Co-occurrence Constraints. For Inductive Grounding, the matter is more transparent: the theory is anchored in Optimality Theory by relying on constraint ranking (beyond two strata). This is intimately connected to the next point.

Constraint activation In Feature Co-occurrence Constraint Theory, only those constraints are activated for which the following holds: both features are acquired. No other restrictions on constraint activation hold; save for the timing criterion. Substance is irrelevant to constraint activation. The situation is somewhat different in Inductive Grounding, where all formally possible constraints are constructed by the learner, but not all are included in the grammar (ranking). The metric of constraint effectiveness evaluation assures that only those constraints are selected which are better predictors of phonetic difficulty than their immediate neighbours (of equal or lesser complexity). Hence, Inductive Grounding appears to incorporate some degree of redundancy: first, all possible constraints are constructed and divided into two subsets: grounded and non-grounded constraints. Next, the grounded constraints are selected and subjected to the OT ranking style of strict domination, where they are ranked based on evidence from the surrounding language (Tesar & Smolensky, 2000). The question immediately arises why not all possible constraints are fed into the grammar in the first place; unmotivated constraints would not be ranked high in any case.

Constraint Sets While not stated explicitly, Inductive Grounding appears to adhere to the notion of the universal constraint set, or at least a weak version thereof. This is because constraints are selected, ultimately, on the basis of generalised phonetic difficulty maps, which we may assume are not substantially different from one language to the next (as long as anatomy may be considered universal). So, even though constraints are not innate, they appear to be universal. There is no such universality in Feature Co-occurrence Constraint Theory, where constraints are activated based on the actual features active in the language. The difference may not be so large in the end, as in Inductive Grounding only language-relevant constraints are expected to be ranked high.

As we can see, Inductive Grounding and Feature Co-occurrence Constraint Theory have some degree of similarity, and some differences. Whether they make the same empirical predictions is a matter for further research. The degree of compatibility seems less promising, however, mostly because of the different aims (satisfying formal and functional demands in Inductive Grounding versus only formal criteria in Feature Co-occurrence Constraint Theory), and the vastly different roles ascribed to learners.

5.2.2 Feature Co-occurrence Constraints and Parallel Bidirectional Phonetics and Phonology

In chapter 2, we noted that the model of Parallel Bidirectional Phonetics and Phonology (as summarised in Boersma & Hamann, 2008) provides an interesting perspective on the phonetics-phonology interface. Remember that the model assumes a multitude of representational levels, from the semantic to the articulatory, which are characterised by the constraints to which they are subjected. These constraints are ordered on a continuous scale (contra 'classical' OT, where ranking is discrete), and the ranking values are learned through the application of the Gradual Learning Algorithm.

The problem with PBPP as a theory of the structure of the inventory is primarily that it does not explicitly state a manner in which the structure of the inventory is derived; rather, it gives a principled solution to the problem of how, given an inventory and its structure, the shape arises. It does, however, provide room for markedness constraints ('structural constraints' in the words of Boersma & Hamann, 2008). At the same time, the cue constraints that map acoustic values to phonological structures, act on whole phonemes rather than on features, indicating that the segment has some independent ontological status other then the timing of simultaneous feature actualisation. These observations raise the question whether Feature Co-occurrence Constraint Theory and Parallel Bidirectional Phonetics and Phonology can benefit from each other.

For our present purposes, only two of the levels in PBPP are of interest. These are repeated in (68) below.

(68) Relevant levels and constraints in Parallel Bidirectional Phonetics and Phonology

|Underlying Form|
↖
faithfulness constraints
↙
/Surface Form/ ⟵ structural constraints
↖
cue constraints
↙
[Auditory Form]

We are not currently concerned with perception, nor with faithfulness; the constraints which we shall discuss are the cue constraints and the structural constraints, where we will assume the further simplification that only FCCs populate the level of structural constraints.

Learning is crucial in PBPP to the degree that it is almost meaningless to make observations about the final state without showing how it is emergent given an input and the Gradual Learning Algorithm. Let us now briefly illustrate the GLA, following the example of (English) sibilants given in Boersma and Hamann (2008). Example (69) presents a cursory display of sibilants and their spectral noise mean values (adapted from Boersma & Hamann, 2008).

(69) Spectral noise mean values for sibilants

2000 Hz —ʂ ṣ ʃ ɕ s̠ sʲ s s̪⟶ 7500 Hz
spectral mean

Before we continue, however, let us recapitulate the main ingredients of Parallel Bidirectional Phonetics and Phonology: as said before, constraints are not ranked discretely but are rather assigned a value on a continuous ranking scale. Evaluation is noisy, meaning that at each evaluation moment, the ranking value is distorted in a random way. Constraints are assigned a ranking probability on the continuous scale, which takes the shape of a normal distribution, and where the ranking value corresponds to the mean. From this it follows that for any pair of two constraints C1 and C2, where the ranking value V1 of C1 is higher than the ranking value V2 of C2, the likelihood of the ranking C1≫C2 is dependent on the difference between |V1-V2| and the standard deviation of the ranking distribution. The closer these two numbers are, the higher the probability that at any evaluation point C2≫C1.

Learning proceeds through a (large) number of iterations through a cycle: First, the listener hears and recognises a given word. For this input, she assumes an underlying form. Then, the learner takes the underlying form as input to her current grammar, and an optimal candidate arises. This candidate is then compared to the perceived input. If they are identical, nothing happens, but in the case of a mismatch, ranking values are adjusted. So, values of constraints that prohibit the perceived input to be the optimal candidate, are lowered by a small amount ('plasticity'), while the values of constraints that critically act against the current grammar's winner are raised. This increases the likelihood of the perceived winner to be equal to the learner's optimal candidate the next time the same form is encountered. Because this is done in every iteration of the learning cycle, differences between ranking values of cue constraints are small when the input is equivocal, and larger where no confusion exists. In other words: where evidence is stronger, ranking values differ more, and the ranking is less likely to be overturned by evaluation noise.

Let us look at some tableaus of the ranking for the correct spectral mean values for /s/ and /ʃ/. Ranking values are omitted.

(70) Perception tableau for classifying tokens with a spectral mean in English (taken from Boersma & Hamann, 2008)

Input: /[26.6 Erb]/	*[26.5]/s/	*[26.6]/s/	*[26.7]/s/	*[26.7]/ʃ/	*[26.6]/ʃ/	*[26.5]/ʃ/
a. /s/		*!				
b. ☞ /ʃ/					*	

In this tableau, we see that the input is mapped to /ʃ/, not /s/, because the cue constraint acting against such mapping outranks the constraint that militates against the mapping that is correct. Now let us look at a production tableau, also taken from Boersma and Hamann (2008).

(71) Preliminary production tableau for /s/

Input: //s//		*[30.6]/s/	*[30.7]/s/	*[30.8]/s/	*[31.5]/s/	*[30.9]/s/	*[31.4]/s/	*[31.3]/s/	*[31.0]/s/	*[31.2]/s/	*[31.1]/s/
a.	[30.6 Erb]	∗!									
b.	[30.7 Erb]		∗!								
c.	[30.8 Erb]			∗!							
d.	[30.9 Erb]					∗!					
e.	[31.0 Erb]								∗!		
f. ☞	[31.1 Erb]										∗
g.	[31.2 Erb]									∗!	
h.	[31.3 Erb]							∗!			
i.	[31.4 Erb]						∗!				
j.	[31.5 Erb]				∗!						

This tableau predicts that the optimal mean spectral value for /s/ is 7100Hz, while in fact the optimal values is 7000Hz (that is, in the simulations performed by the authors). This is due to the stochastic nature of GLA, which "...causes cue constraints to end up ranked lowest in auditory regions where the learner has heard the largest number of least confusable tokens" (Boersma & Hamann, 2008, p18). To counteract this so-called prototype effect (see Boersma and Hamann (2008) for references), the authors add articulatory constraints, which act against the articulation of any value, and are roughly analogous to the *Gesture constraints in Boersma (1998). The resulting tableau is given in 72:

(72) Full production tableau for /s/

Input: //s//		*31.2	*31.1	*31.0	*30.9	*[30.6]/s/	*30.8	*[30.7]/s/	*30.7	*[30.8]/s/	*30.6	*[30.9]/s/	*[31.0]/s/	*[31.2]/s/	*[31.1]/s/
a.	[30.6 Erb]					∗!					∗				
b. ☞	[30.7 Erb]							∗	∗						
c.	[30.8 Erb]						∗!			∗					
d.	[30.9 Erb]				∗!							∗			
e.	[31.0 Erb]			∗!									∗		
f.	[31.1 Erb]		∗!												∗
g.	[31.2 Erb]	∗!												∗	

What is crucial in this tableau, is that the cue constraint militating against the optimal candidate is outranked by the articulatory constraints acting against its competitors. Note that the relative ranking of the cue constraint and articulatory constraint acting against 30.7 Erb (the optimal candidate) is irrelevant. What matters is that both are outranked by the articulatory constraint acting against the other candidates. Although I will not go in to this much further, Boersma and Hamann (2008) show that both the perception tableau and the full production tableau are learnable, and that they remain stable over generations (or will evolve into a stable state, when the initial state is suboptimal somehow).

To see how interaction between Parallel Bidirectional Phonology and Phonetics and the model presented in this study might take place, let us go back to the matter of English sibilants, discussed above. As we have seen, English has two: /s/ and /ʃ/. These two segments are defined as in 73.

(73) Feature definitions for /s/ and /ʃ/

	/s/	/ʃ/
[continuant]	✓	✓
[ant]	✓	
[distributed]		✓

In other words, English is a language in which the two co-occurrence patterns shown above are optimal, so that the two c-constraints acting against these segments (*[cont, ant] and *[cont, dist]) are ranked low.

As for the cue constraints, there are many; as many as the product of the number of phonological entities and the number of phonetic Just Notable Differences. So, just as there is a constraint acting against a mapping of noise with a spectral mean of 30.7 Erb to [ant], there is also one against mapping the same noise to [cont], and so on for every step along the noise spectral mean scale. With respect to the /s/ ∼ /ʃ/ contrast, only the ranking of constraints acting against the mapping of 30.7 Erb to either [ant] or [dist] is relevant: it is undesirable that the relative ranking of constraints mapping noise values to the feature [cont] to indicate anything more specific than the fact that we are, in fact, dealing with continuants.

This, however, naturally follows from the nature of the Gradual Learning Algorithm. Both /s/ and /ʃ/ map a different band of values of noise spectral mean to [cont], which means that the evidence that any of the two noise values is a cue to continuancy is equivocal. The reader is reminded that differences in ranking values are smaller when the evidence is less univocal, and hence the ranking values of cue constraints acting against these mappings will not differ greatly.

The two segments differ in being either [ant] or [dist], however, and this is where the difference in noise spectral mean is realised: the difference between a cue constraint acting against a mapping of 30.7 Erb to [ant] and one acting against a mapping of 30.7 Erb to [dist] will be much greater, in a fashion much like the difference that Boersma and Hamann (2008) found between /s/ and /ʃ/. In other words, noise spectral mean values are not very informative in identifying a segment for the feature [cont], but are crucial when it comes to the difference between [ant] and [dist]. More generally, only phonetic values that correlate with features that distinguish one segment from another - given a certain inventory will be able to induce significant differences in cue constraint ranking values. The more distinctive a feature is, the more information it carries. This seems to be a desirable outcome.

A combination of Feature Co-occurrence Constraint Theory and the Parallel Bidirectional Phonetics and Phonology model is promising, we may conclude – with the caveat that the outline above is extremely sketchy and informal. One outcome concerns the contrastive nature of features, which brings us to the Modified Contrastive Hierarchy.

5.2.3 Feature Co-occurrence Constraints and the Modified Contrastive Hierarchy

Throughout this thesis, we have repeatedly discussed properties of the Modified Contrastive Hierarchy (Hall, 2007; Dresher, 2009). One of the main questions that arose concerns the ontological status of the inventory: in the process of online computation, where phonology maps perceived forms to the lexicon or lexical forms to output representations, where is the hierarchy located or represented? Our claim has been that what we have called *holisticity* is an undesirable characteristic for a phonological theory, where holisticity is taken to mean that in the process of mapping forms, the phonology must incorporate representational information other than that which is represented in either form. Take, for example, a listener who perceives a form that contains /m/. How is she to know whether the periodicity in this signal is to be encoded by a feature [+voice] or not? And if not, should the segment be assigned a non-contrastive [-voice] feature, or is it contrastively underspecified? Depending on the order of successive divisions in the application of the Successive Division Algorithm, each of these options corresponds to a real possibility. In order to decide, the listener must refer to the hierarchy, or the feature specification matrix that is the result thereof, in each computation. Hence, the MCH is not free of a hint of holisticity, as we noted earlier (chapter 2).

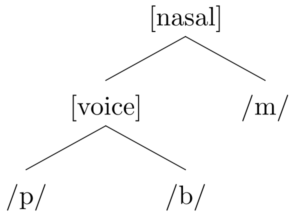
This criticism only holds, however, in so far as the Modified Contrastive Hierarchy aims to be a model of phonological computation. If this is not so, we see that the theory rather is a theory of information structure, which must somehow be encoded in a model of online computation. It is here that the Feature Co-occurrence Theory presents itself as a possible companion.

In this section, we will briefly outline a possible way in which Feature Co-occurrence Constraints can encode the information represented by the contrastive hierarchy of features. Within the limits of the sketch outlined below, we can still conceive of two different ways in which the Successive Division Algorithm applies: chronologically or non-chronologically. Dresher (2009, p. 327) briefly discusses the convergence of the SDA as applied to Dutch and the order of acquisition of Dutch initial consonants as reported in Fikkert (1994), so we will follow him by adopting a chronological perspective on the SDA. Below, however, we will note that there are reasons to doubt the validity of this approach.

For our example, we will consider the labial consonant sub-inventory in French, which we already encountered in chapter 2. This sub-inventory consists of the three members /p, b, m/. The hierarchy and feature specifications corresponding to the final state are given in (74). Two features are acquired, so including the initial state we must account for three successive states of the grammar. For reasons of conciseness, we will temporarily suspend our choice for monovalent features and adopt the binary system that is used in Dresher (2009).

(74) [±nasal]≫[±voice]

a. Feature hierarchy



b. Contrastive specification

	p	b	m
[voice]	-	+	
[nasal]	-	-	+

S_0: Initial state. In the initial state, Dresher assumes, all phonemes are allophones of each other. There is no contrast, and hence, no features, and hence, no FCCs.

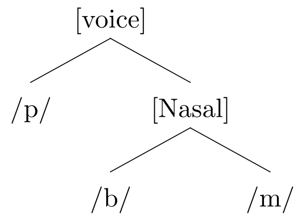
S_1: [±nasal] is acquired. The child has learned that some segments behave in a distinct way, to the exclusion of other segments. The contrast between [-nasal] and [+nasal] is acquired. The mutual exclusivity of both features is expressed by a c-constraint: *[-nasal, +nasal]. The members of the [+nasal] class number exactly one: only /m/ is [+nasal]. Hence, /m/ is now uniquely specified.

S_2: [±voice] is acquired. There are two non-nasal segments in the inventory, and there is no way to tell them apart, up to S_2, that is. At this point, the learner acquires [±voice]. The [+voice] value is assigned to /b/, the [-voice] value to /p/, whereas /n/, already being uniquely specified, receives no voicing feature. The contrastive nature of [±voice] is expressed by the c-constraint *[-voice, +voice], like we saw in the previous state for [±nasal.] Furthermore, the [+nasal] segment may not bear a voicing specification, and thus the following c-constraints are activated: *[+nasal, +voice] and *[+nasal, -voice].

This interpretation of the Successive Division Algorithm asserts that it describes a process that unfolds during phonological acquisition. There are reasons to doubt, however, that this is the best way to interpret the SDA. For example, the procedure is extremely sensitive to differences in the order of acquisition of features, or rather, of contrasts. Beers (1995) notes a considerable degree of individual variation in populations of both normally developing and language impaired children. By virtue of the Successive Division Algorithm, differences in order of acquisition have direct ramifications in the realm of feature specifications. Consider the inventory in (74). As we discussed in chapter 2, this is not the only possible final state for the SDA: if the order of acquisition would be reversed, the resulting inventory would the one in (75).

(75) [±voice]≫[±nasal]

a. Feature hierarchy



b. Contrastive specification

	p	b	m
[voice]	-	+	+
[nasal]		-	+

The chronological interpretation of the SDA thus makes strong predictions about the order of acquisition of contrasts, which must at some point be empirically tested. However, the interpretation where the SDA works by monotonically adding contrasts is not its only possible conception. For example, we might envision the SDA as an iterative process that re-applies every time the inventory grows. It would go too far to flesh out the different perspectives and their predictions at this point, but the brief exposé above seems promising.

It would appear, thus, that Feature Co-occurrence Constraints can express the information represented by the contrastive hierarchy. The presence of constraints now receives the interpretation listed below. It remains a question for further research whether this non-trivial extension is fully compatible with the ideas presented in the current thesis, and to what degree the constraint types below are still capable of expressing the functions described in chapter 3.

***[-F, +F]** This c-constraint simply expresses that feature [F] is contrastive. It must be activated for all features that are acquired, to divide the inventory in two subsets; one for each value of [F].

***[αF, -G], *[αF, +G]** If both constraints are activated, this entails that [$\pm$G] is not contrastive for [αF]. The relativised non-contrastiveness of [$\pm$G] is thus not expressed by a single constraint, but rather by two.

***[αF, βG]** (where α can be + or -, and β can be + or -) The presence of a c-constraint indicates a gap in the inventory. Hence, the interpretation of a single c-constraint is not different. In fact, the interpretation of the two cases above can be said to be similar to 'regular' FCCs: they, too, describe gaps. What is special is not the constraints, but rather the gaps.

i-constraints In this brief demonstration we have not used i-constraints. The reader is reminded that the need for i-constraints was intimately linked with the choice for monovalent features (see chapter 3), and in this section we followed Dresher (2009)'s binary feature system. One possible role for i-constraints in a combined theory of both MCH and FCCs could be to express the enhancement features proposed by Hall (2011). We will leave this issue for further investigations.

The previous note does raise the question whether the sketch above can be cast in monovalent feature theory. Dresher (2009, § 2.7.2) claims that the Modified Contrastive Hierarchy is compatible with privative features, and we know that the Feature Co-occurrence Constraint Theory is. The problem with monovalent features in the Modified Contrasitive Hierarchy is, Dresher notes, that the theory makes use of the logical three-way contrast ∅F, +F and -F, where the former expresses non-contrastivity, and the latter two express opposing contrastive values for F. In monovalent feature theory, the difference between -F and ∅F is conflated; both are expressed by the absence of a specification for F. Dresher notes that "some machinery in addition to the representations itself" is then required, but does not remark on the possible nature of that machinery.

In this section, we have shown one possible way in which the Modified Contrastive Hierarchy and the Feature Co-occurrence Constraint Theory can complement each other. Although the sketch seems promising, we have not chosen to adopt it throughout the thesis, among other reasons because it entails specific conditions for the process of acquiring features. The current thesis aims to be as broad as possible, and at the same time as specific as possible, and liaising explicitly with the Modified Contrastive Hierarchy would jeopardise these aims.

5.3 Residual Issues

No theory is complete. There are always remaining questions and issues. In this section, we will consider a number of those, without the pretension or ambition of exhaustiveness. Three questions that we will address concern the nature of evidence in a theory of non-innate constraints, the choice of framework for implementation of Feature Co-occurrence Constraints, and the relation between perception and production.

5.3.1 Emergent constraints and evidence

In language acquisition, the subset principle holds that the most conservative grammar corresponds to the child's linguistic system; that is, if two competing grammars A and B are available, where the set of grammatical forms generated by each grammar is $|A|$ and $|B|$ respectively, the child will select grammar A if $|A| \subset |B|$. Furthermore, grammars are only adjusted based on positive evidence: the child does not learn from the absence of a certain input/intake structure.

The picture painted above, in one form or another, is rather uncontroversial in generative linguistics. However, it does present a difficulty for any theory that proposes emergent constraints. The reason for this is that constraints are generalisations about what is *not* allowed, and hence, can be seen as representing knowledge of absent input structures.

In early Optimality Theory works, the positive evidence hypothesis was encoded by positing that in the initial state, markedness constraints collectively outrank faithfulness constraints (Gnanadesikan, 1995; Levelt, 1995; Smolensky, 1996; Stemberger & Bernhart, 1998; Jusczyk et al., 2002, but see ; Hale & Reiss, 2008 for an opposite perspective). This initial ranking entails that the grammar is maximally conservative: only unmarked forms can ever be optimal. Only upon exposure to positive evidence (i.e., forms that deviate somehow from markedness requirements) do children learn that sometimes, faithfulness trumps markedness, and the relevant markedness constraint is demoted (see Tesar & Smolensky, 2000 for a constraint demotion algorithm).

Emergent constraints distort this scenario, by introducing an additional step in the acquisition process. In order for a constraint to emerge, the child must somehow gain awareness of the absence of a possible output form, and codify this awareness in a negative statement (i.e., the constraint). Take, for example, the constraints proposed in Fikkert and Levelt (2008): *[DORS and LABIALLEFT. The authors propose that these are not innate, but rather emergent as "...generalisations over the learner's production lexicon" (p. 238). This is problematic in the sense that the generalisations are not about positive data, but rather about the absence of other structures.

The problem is less severe for Feature Co-occurrence Constraint Theory, however. This is because FCCs are posited *automatically*, that is, without reflection. Whenever the conditions for a constraint are met by the state of acquisition of the feature set, the FCCs are activated and ranked, or labeled as 'violated' or 'satisfied'. Even though the constraints ban structures the child has never encountered, they do so to retain maximal conservatism: the unrestricted addition of a feature to the feature system results in massive overprediction of grammatical segments.

5.3.2 Constraint demotion versus constraint revocation

In chapter 3, we demonstrated that Feature Co-occurrence Constraint Theory is compatible with both constraint ranking (OT) and non-ranking frameworks of phonology. The reason for this, we argued, is that among themselves, the set of FCCs will not be ranked beyond two strata, for lack of further ranking arguments. Hence, Feature Co-occurrence Constraint theory does not crucially employ Optimality Theory-specific machinery. The two strata translate to 'violated' and 'not violated' in non-ranking frameworks. Furthermore, we argued that some form of feature co-occurrence constraints must be incorporated in any theory that has an unbounded generator function (be it GEN or a set of rules, or something entirely different).

Having established that this is the case, the question remains as to which type of framework is best suited for Feature Co-occurrence Constraints (or *vice versa*). In a sense (or: in essence) the question is independent of Feature Co-occurrence Constraint Theory and is largely up to the analyst. For example, we have seen two analyses of vowel harmony that rely on FCCs, one couched in an OT framework (Linke & van Oostendorp, in preparation), and one in a non-ranking framework of phonology (Van der Hulst, 2012). What remains, then, is a choice based on the role of constraints in the grammar (see Cavirani, 2010 for a historical and comparative overview of the different roles and applications of constraints in phonological theorising).

One (arguably non-crucial) argument in favor of an Optimality Theoretic implementation of FCCs is that the theory inherently provides the mechanism for a crucial point of Feature Co-occurrence Constraint Theory: the fact that some FCCs are only active for a limited time. In OT, this can be modeled quite straightforwardly by constraint demotion; in non-ranking frameworks, a provision must be introduced to allow for transient constraints.

The main difference for our current purposes is that in Optimality Theory, constraints may exert an influence far beyond their structural description, whereas in non-ranking frameworks, the effect of the constraint is more immediately transparent. This is not to say that in OT, the interpretation of constraints is not transparent, but rather that due to the ranking and violability of constraints, constraint interaction may be more complex and far-reaching (see, e.g., Pater, 1997).

Another important difference is that in OT, constraints may only affect output structures. In the words of Kager (1999, p.19): "no constraints hold at the level of underlying forms." Such a limitation of the domain of application of constraints is to my knowledge specific to Optimality Theory and related frameworks, such as Harmonic Grammar (Smolensky & Legendre, 2006). If it can be shown that the Feature Co-occurrence Constraint Theory makes crucial reference to underlying forms, we can exclude Optimality Theory as a possible framework (at least for the theory as-is). It is important to note that the requirements stated by Feature Co-occurrence Constraints are requirements of output realisations. The 'picking and choosing' of suitable input forms is most probably due to other factors; reflection on the current state of the grammar is one such possible factor, but the grammar itself remains an output oriented device.

Ultimately, then, the choice of framework is dependent on factors other than the theory developed here. The two crucial questions then are a) is it necessary for the Feature Co-occurrence Constraints to engage in complex interactions with other (types of) constraints? If so, Optimality Theory is the way to go; and b) do we want to further expand the theory to apply to underlying forms? If so, Optimality Theory is of no avail. The theory as it is developed in the current thesis remains compatible with both options.

5.3.3 Perception

There is one additional issue that deserves mentioning, and this is the issue of perception. Throughout this thesis, we have treated Feature Co-occurrence Constraint Theory as a theory of a production grammar, and although this is common practice in the general phonological literature (where it is assumed that the production grammar and the perception grammar are one and the same), matters are not so simple when we talk about acquisition.

In chapter 2.3 we discussed a number of studies in phonological acquisition, both production studies and perception studies. We saw that perceptual development begins rather early, and, for example, the relevant language-specific phonemic/phonetic categories are in place before the end of the first year of life. Thus, while children make mistakes in the production of targeted segments up to two or even three years of age, and typically do not start production before their first birthday, perceptual categories seem to be well in place before that. In other domains, too, perception precedes production.

Again, the problem is not unique to Feature Co-occurrence Constraint Theory. Any theory of phonological acquisition should aim to provide an account for the perception/production mismatch, although to my knowledge no satisfactory answer has been formulated yet. We can name three possible sources for the disparity: different performance factors, differences in the perception and production grammar (or differences in the interpretation of elements of the grammar), or immature underlying representations.[3]

One possible source of the disparity is the difference between the performance factors involved in either perception or production. Memory plays an important role in both, but production is also affected by issues relating to motor control, articulatory planning, difference in vocal tract anatomy, *et cetera*. As we have seen in previous chapters, Inkelas and Rose (2008) provide an analysis of positional velar fronting and positional lateral gliding by reference to (phonologised) performance factors, where in fact the deviation in production is attributed to an attempt at remaining faithful to the (correctly) perceived form.

One solution, concerning the acquisition of the segment inventory, is proposed by Pater (2004), and further expanded in Li (2007). Pater starts form the assumption that the child grammar can be identical to adult forms, or that there is a lag between perception and production. In contrast to Smolensky (1996), who argues that the role of grammar in perception (and lexical representation selection) is limited to faithfulness constraints, Pater (2004) proposes that all constraints are involved in "... regulating the structure, and therefore the complexity, of the representation(s) used in perception." In other words, the mapping of incoming words to underlying representations is subject to the same grammar (constraint ranking) as is the mapping of underlying representations to surface forms. The difference is that Pater proposes two sets of faithfulness constraints: one requiring identity from surface form to lexical form (MaxSL), and one requiring identity in the opposite direction (MaxLS).[4]

In 'standard' Optimality Theory, faithfulness constraints are bidirectional. By dissecting each faithfulness constraint into a surface-lexical and a lexical-surface form, Pater (2004) is able to account for the apparent lag of production

[3] A fourth option, that we will not further review here, is that the lexicon contains different perception and production representations for any lemma. This view is pursued in Hemphill (1998) and critically reviewed in Menn and Matthei (1992), for example.

[4] Pater (2004) goes beyond these two sets by adding identity requirements for acoustic form to surface form, but for reasons of conciseness we will not further go into this extension.

relative to perception. It remains to be investigated, however, whether there is independent evidence supporting this move.

Another view, stated in Fikkert (2008) and underlying other work on phonological acquisition (see, e.g., Levelt, 1994; Fikkert & Levelt, 2008), points to the difference between perception and representation. This view holds that the results from perception studies that are taken to indicate that perceptual development precedes productive development is too simplistic; Fikkert argues that even if discriminative perception is adult-like, it does not follow that underlying representations are, too. In chapter 2 we have discussed these works, which hold that even though features are active from very early stages, they are not yet specified in the adult-like way: rather than attaching features to a segmental (root) node in the prosodic hierarchy, children begin with specifying whole words for a given feature. The representation becomes ever more segmentalised, up to the point where individual root nodes are available for features to attach to. We have seen that this hypothesis leads to interesting results in both perception (Fikkert et al., 2005) and production studies (Fikkert & Levelt, 2008).

5.4 Conclusion

The segment inventory is central to phonology, and its acquisition has been the subject of study for more than a century. However, a satisfactory theory that describes both the acquisition of the individual segments but also the gaps that the developing inventory contains, does not yet exist. The present thesis aims to fill that lacuna.

Earlier studies of the maturing inventory tended to focus on mostly on the representational aspect of phonology (features) with some notable exceptions (Levelt, 1994; Levelt & van Oostendorp, 2007). The main innovation of Feature Co-occurrence Constraint theory is that it approaches acquisition as a tandem development of both the representational and computational (constraints) sides of the phonological grammar.

Furthermore, Feature Co-occurrence Constraint theory is unique in its ability to account for gaps in the inventory. We have devoted considerable space to discussing the type of gaps that the theory predicts.

Feature Co-occurrence Constraint theory is built on two main ingredients: distinctive features and constraints on feature combinations. Both are part of many if not all phonological theories, but the specifics of Feature Co-occurrence Constraints remained largely understudied up to now. Feature Co-occurrence Constraint theory contributes to our understanding of the workings of phonological constraints by exploring the consequences of applying strict restrictions on the structure of constraints.

Finally, it is worth noting that Feature Co-occurrence Constraint theory is a theory as much as a tool for further research.

APPENDIX A

Stages

The analysis in chapter 4 rests on the segmentation of developmental time into stages. Sometimes, these stages coincide with the interval between two recording sessions (see section 4.2). Often, however, a stage spans multiple sessions. In this appendix, we list the dates of the recording sessions together with the stages they fall into. Note that stages are defined by a change in the inventory (see section 4.2 for a more detailed description), which means that there is no logical necessity that for the same child, the stages at the different levels of description should coincide.

<table>
<tr><th></th><th>Actual</th><th>Target</th><th>Faithful</th></tr>
<tr><td>Catootje 1990-09-05</td><td>Stage 1</td><td>Stage 1</td><td>Stage 1</td></tr>
<tr><td>Catootje 1990-09-19</td><td rowspan="2">Stage 2</td><td rowspan="2">Stage 2</td><td rowspan="6">Stage 2</td></tr>
<tr><td>Catootje 1990-10-04</td></tr>
<tr><td>Catootje 1990-10-17</td><td rowspan="4">Stage 3</td><td rowspan="10">Stage 3</td></tr>
<tr><td>Catootje 1990-10-31</td></tr>
<tr><td>Catootje 1990-11-14</td></tr>
<tr><td>Catootje 1990-11-28</td></tr>
<tr><td>Catootje 1991-01-09</td><td>Stage 4</td><td>Stage 3</td></tr>
<tr><td>Catootje 1991-01-23</td><td>Stage 5</td><td>Stage 4</td></tr>
<tr><td>Catootje 1991-02-20</td><td rowspan="2">Stage 6</td><td rowspan="2">Stage 5</td></tr>
<tr><td>Catootje 1991-03-06</td></tr>
<tr><td>Catootje 1991-03-20</td><td>Stage 7</td><td rowspan="4">Stage 6</td></tr>
<tr><td>Catootje 1991-04-03</td><td rowspan="2">Stage 8</td></tr>
<tr><td>Catootje 1991-04-17</td><td rowspan="3">Stage 4</td></tr>
<tr><td>Catootje 1991-05-01</td><td>Stage 9</td></tr>
<tr><td>Catootje 1991-05-29</td><td>Stage 10</td><td>Stage 7</td></tr>
</table>

Table A.1: Catootje's stages

<table>
<tr><th></th><th>Actual</th><th>Target</th><th>Faithful</th></tr>
<tr><td>Eva 1989-10-19</td><td>Stage 1</td><td>Stage 1</td><td>Stage 1</td></tr>
<tr><td>Eva 1989-11-02</td><td rowspan="4">Stage 2</td><td rowspan="2">Stage 2</td><td rowspan="2">Stage 2</td></tr>
<tr><td>Eva 1989-11-29</td></tr>
<tr><td>Eva 1989-12-08</td><td>Stage 3</td><td rowspan="2">Stage 3</td></tr>
<tr><td>Eva 1989-12-18</td><td rowspan="7">Stage 4</td></tr>
<tr><td>Eva 1990-01-22</td><td rowspan="3">Stage 3</td><td rowspan="3">Stage 4</td></tr>
<tr><td>Eva 1990-01-29</td></tr>
<tr><td>Eva 1990-02-19</td></tr>
<tr><td>Eva 1990-03-15</td><td>Stage 4</td><td rowspan="3">Stage 5</td></tr>
<tr><td>Eva 1990-03-29</td><td rowspan="2">Stage 5</td></tr>
<tr><td>Eva 1990-04-10</td></tr>
<tr><td>Eva 1990-05-15</td><td>Stage 6</td><td>Stage 5</td><td>Stage 6</td></tr>
</table>

Table A.2: Eva's stages

<table>
<tr><th></th><th>Actual</th><th>Target</th><th>Faithful</th></tr>
<tr><td>Jarmo 1989-09-22</td><td rowspan="2"></td><td>Stage 1</td><td></td></tr>
<tr><td>Jarmo 1989-10-06</td><td rowspan="3">Stage 2</td><td rowspan="3">Stage 1</td></tr>
<tr><td>Jarmo 1989-10-31</td><td rowspan="2">Stage 1</td></tr>
<tr><td>Jarmo 1989-11-17</td></tr>
<tr><td>Jarmo 1989-12-01</td><td>Stage 2</td><td>Stage 3</td><td rowspan="2">Stage 2</td></tr>
<tr><td>Jarmo 1989-12-19</td><td rowspan="4">Stage 3</td><td>Stage 4</td></tr>
<tr><td>Jarmo 1990-01-02</td><td rowspan="3">Stage 5</td><td rowspan="3">Stage 3</td></tr>
<tr><td>Jarmo 1990-01-16</td></tr>
<tr><td>Jarmo 1990-01-30</td></tr>
<tr><td>Jarmo 1990-02-13</td><td rowspan="2">Stage 4</td><td rowspan="2">Stage 6</td><td rowspan="2">Stage 4</td></tr>
<tr><td>Jarmo 1990-02-27</td></tr>
<tr><td>Jarmo 1990-03-13</td><td rowspan="3">Stage 5</td><td rowspan="2">Stage 7</td><td rowspan="6">Stage 5</td></tr>
<tr><td>Jarmo 1990-03-27</td></tr>
<tr><td>Jarmo 1990-04-10</td><td>Stage 8</td></tr>
<tr><td>Jarmo 1990-04-24</td><td>Stage 6</td><td>Stage 9</td></tr>
<tr><td>Jarmo 1990-05-08</td><td rowspan="3">Stage 7</td><td rowspan="5">Stage 10</td></tr>
<tr><td>Jarmo 1990-06-01</td></tr>
<tr><td>Jarmo 1990-06-12</td><td rowspan="2">Stage 6</td></tr>
<tr><td>Jarmo 1990-06-26</td><td rowspan="3">Stage 8</td></tr>
<tr><td>Jarmo 1990-07-10</td><td rowspan="2">Stage 7</td></tr>
<tr><td>Jarmo 1990-07-31</td><td>Stage 11</td></tr>
<tr><td>Jarmo 1990-08-13</td><td rowspan="2">Stage 9</td><td rowspan="2">Stage 12</td><td rowspan="2">Stage 8</td></tr>
<tr><td>Jarmo 1990-09-05</td></tr>
</table>

Table A.3: Jarmo's stages

<table>
<tr><th></th><th>Actual</th><th>Target</th><th>Faithful</th></tr>
<tr><td>Noortje 1989-09-25</td><td rowspan="2"></td><td rowspan="3">Stage 1</td><td rowspan="3">Stage 1</td></tr>
<tr><td>Noortje 1989-11-06</td></tr>
<tr><td>Noortje 1990-02-21</td><td>Stage 1</td></tr>
<tr><td>Noortje 1990-03-28</td><td rowspan="3">Stage 2</td><td rowspan="3">Stage 2</td><td rowspan="6">Stage 2</td></tr>
<tr><td>Noortje 1990-04-18</td></tr>
<tr><td>Noortje 1990-05-02</td></tr>
<tr><td>Noortje 1990-05-18</td><td rowspan="3">Stage 3</td><td rowspan="2">Stage 3</td></tr>
<tr><td>Noortje 1990-06-01</td></tr>
<tr><td>Noortje 1990-06-15</td><td>Stage 4</td></tr>
<tr><td>Noortje 1990-08-03</td><td>Stage 4</td><td rowspan="10">Stage 5</td><td>Stage 3</td></tr>
<tr><td>Noortje 1990-08-16</td><td rowspan="4">Stage 5</td><td rowspan="10">Stage 4</td></tr>
<tr><td>Noortje 1990-08-30</td></tr>
<tr><td>Noortje 1990-09-13</td></tr>
<tr><td>Noortje 1990-09-28</td></tr>
<tr><td>Noortje 1990-10-12</td><td rowspan="4">Stage 6</td></tr>
<tr><td>Noortje 1990-10-26</td></tr>
<tr><td>Noortje 1990-11-09</td></tr>
<tr><td>Noortje 1990-11-13</td></tr>
<tr><td>Noortje 1990-11-23</td><td>Stage 7</td></tr>
<tr><td>Noortje 1990-12-07</td><td>Stage 8</td><td rowspan="2">Stage 6</td></tr>
<tr><td>Noortje 1991-01-11</td><td>Stage 9</td><td>Stage 5</td></tr>
</table>

Table A.4: Noortje's stages

<table>
<tr><th></th><th>Actual</th><th>Target</th><th>Faithful</th></tr>
<tr><td>Robin 1989-11-08</td><td rowspan="3">Stage 1</td><td rowspan="2">Stage 1</td><td rowspan="5">Stage 1</td></tr>
<tr><td>Robin 1989-11-22</td></tr>
<tr><td>Robin 1989-12-07</td><td rowspan="2">Stage 2</td></tr>
<tr><td>Robin 1989-12-20</td><td>Stage 2</td></tr>
<tr><td>Robin 1990-01-10</td><td rowspan="2">Stage 3</td><td>Stage 3</td></tr>
<tr><td>Robin 1990-01-24</td><td>Stage 4</td><td rowspan="4">Stage 2</td></tr>
<tr><td>Robin 1990-02-07</td><td rowspan="3">Stage 4</td><td rowspan="2">Stage 5</td></tr>
<tr><td>Robin 1990-02-21</td></tr>
<tr><td>Robin 1990-03-07</td><td rowspan="2">Stage 6</td></tr>
<tr><td>Robin 1990-03-21</td><td rowspan="3">Stage 5</td><td rowspan="3">Stage 3</td></tr>
<tr><td>Robin 1990-04-04</td><td rowspan="10">Stage 7</td></tr>
<tr><td>Robin 1990-04-18</td></tr>
<tr><td>Robin 1990-05-04</td><td>Stage 6</td><td rowspan="5">Stage 4</td></tr>
<tr><td>Robin 1990-05-18</td><td rowspan="3">Stage 7</td></tr>
<tr><td>Robin 1990-06-01</td></tr>
<tr><td>Robin 1990-06-15</td></tr>
<tr><td>Robin 1990-07-04</td><td>Stage 8</td></tr>
<tr><td>Robin 1990-07-23</td><td rowspan="4">Stage 9</td><td>Stage 5</td></tr>
<tr><td>Robin 1990-08-24</td><td>Stage 6</td></tr>
<tr><td>Robin 1990-09-07</td><td>Stage 7</td></tr>
<tr><td>Robin 1990-09-19</td><td rowspan="3">Stage 8</td><td rowspan="3">Stage 8</td></tr>
<tr><td>Robin 1990-10-05</td><td rowspan="2">Stage 10</td></tr>
<tr><td>Robin 1990-10-26</td></tr>
</table>

Table A.5: Robin's stages

<table>
<tr><th></th><th>Actual</th><th>Target</th><th>Faithful</th></tr>
<tr><td>Tirza 1989-11-23</td><td colspan="3"></td></tr>
<tr><td>Tirza 1989-12-19</td><td rowspan="2">Stage 1</td><td rowspan="2">Stage 1</td><td rowspan="2">Stage 1</td></tr>
<tr><td>Tirza 1990-01-09</td></tr>
<tr><td>Tirza 1990-01-25</td><td rowspan="7">Stage 2</td><td>Stage 2</td><td rowspan="2">Stage 2</td></tr>
<tr><td>Tirza 1990-02-22</td><td>Stage 3</td></tr>
<tr><td>Tirza 1990-03-08</td><td>Stage 4</td><td rowspan="2">Stage 3</td></tr>
<tr><td>Tirza 1990-03-22</td><td rowspan="3">Stage 5</td></tr>
<tr><td>Tirza 1990-04-02</td><td rowspan="2">Stage 4</td></tr>
<tr><td>Tirza 1990-04-19</td></tr>
<tr><td>Tirza 1990-05-02</td><td rowspan="2">Stage 6</td><td rowspan="2">Stage 5</td></tr>
<tr><td>Tirza 1990-05-16</td><td rowspan="4">Stage 3</td></tr>
<tr><td>Tirza 1990-05-31</td><td rowspan="9">Stage 7</td><td>Stage 6</td></tr>
<tr><td>Tirza 1990-06-14</td><td rowspan="2">Stage 7</td></tr>
<tr><td>Tirza 1990-06-26</td></tr>
<tr><td>Tirza 1990-07-09</td><td rowspan="5">Stage 4</td><td rowspan="5">Stage 8</td></tr>
<tr><td>Tirza 1990-07-26</td></tr>
<tr><td>Tirza 1990-08-10</td></tr>
<tr><td>Tirza 1990-09-19</td></tr>
<tr><td>Tirza 1990-10-05</td></tr>
<tr><td>Tirza 1990-10-26</td><td>Stage 5</td><td>Stage 9</td></tr>
</table>

Table A.6: Tirza's stages

<table>
<tr><th></th><th>Actual</th><th>Target</th><th>Faithful</th></tr>
<tr><td>Tom 1989-10-30</td><td></td><td></td><td></td></tr>
<tr><td>Tom 1989-11-13</td><td></td><td></td><td></td></tr>
<tr><td>Tom 1989-11-27</td><td></td><td></td><td></td></tr>
<tr><td>Tom 1989-12-20</td><td></td><td></td><td></td></tr>
<tr><td>Tom 1990-01-02</td><td>Stage 1</td><td></td><td></td></tr>
<tr><td>Tom 1990-01-16</td><td rowspan="2">Stage 2</td><td></td><td></td></tr>
<tr><td>Tom 1990-01-30</td><td></td><td></td></tr>
<tr><td>Tom 1990-02-20</td><td rowspan="2">Stage 3</td><td>Stage 1</td><td>Stage 1</td></tr>
<tr><td>Tom 1990-03-06</td><td>Stage 2</td><td>Stage 2</td></tr>
<tr><td>Tom 1990-03-20</td><td>Stage 4</td><td>Stage 3</td><td>Stage 3</td></tr>
<tr><td>Tom 1990-04-03</td><td rowspan="2">Stage 5</td><td>Stage 4</td><td>Stage 4</td></tr>
<tr><td>Tom 1990-04-17</td><td>Stage 5</td><td rowspan="3">Stage 5</td></tr>
<tr><td>Tom 1990-05-01</td><td rowspan="2">Stage 6</td><td rowspan="6">Stage 6</td></tr>
<tr><td>Tom 1990-05-15</td></tr>
<tr><td>Tom 1990-05-29</td><td>Stage 7</td><td>Stage 6</td></tr>
<tr><td>Tom 1990-06-12</td><td rowspan="7">Stage 8</td><td rowspan="5">Stage 7</td></tr>
<tr><td>Tom 1990-06-26</td></tr>
<tr><td>Tom 1990-07-10</td></tr>
<tr><td>Tom 1990-08-14</td><td rowspan="2">Stage 7</td></tr>
<tr><td>Tom 1990-09-04</td></tr>
<tr><td>Tom 1990-09-18</td><td>Stage 8</td><td rowspan="5">Stage 8</td></tr>
<tr><td>Tom 1990-10-02</td><td rowspan="4">Stage 9</td></tr>
<tr><td>Tom 1990-10-23</td><td rowspan="3">Stage 9</td></tr>
<tr><td>Tom 1990-11-20</td></tr>
<tr><td>Tom 1991-01-08</td></tr>
</table>

Table A.7: Tom's stages

APPENDIX B

Constraints

In the table below, every attested constraint is listed, with checkmarks in the columns headed by the children's names indicates whether the constraint is active in the child's grammar at some point.

	Catootje	*Eva*	*Jarmo*	*Noortje*	*Robin*	*Tirza*	*Tom*
[apprx]→[cont]	✓	✓	✓	✓	✓	✓	✓
[apprx]→[lab]			✓	✓	✓		✓
[cont]→[apprx]	✓	✓	✓				
[cont]→[lab]			✓				✓
[dist]→[cont]			✓			✓	
[lab]→[voice]					✓		
[nas]→[lab]			✓	✓			
[voice]→[lab]	✓					✓	✓
*[apprx, dist]			✓			✓	
*[apprx, dors]	✓	✓	✓	✓	✓	✓	✓
*[apprx, liq]	✓	✓	✓	✓	✓	✓	✓
*[cont, dors]	✓	✓	✓	✓	✓	✓	
*[cont, liq]	✓	✓	✓	✓	✓	✓	✓
*[dors, dist]			✓			✓	
*[dors, lab]	✓	✓	✓	✓	✓	✓	✓
*[dors, liq]	✓	✓	✓	✓	✓	✓	✓
*[dors, nas]	✓	✓	✓	✓	✓	✓	✓
*[dors, voice]	✓	✓	✓	✓	✓	✓	✓
*[lab, apprx]	✓					✓	
*[lab, cont]	✓			✓	✓	✓	
*[lab, dist]			✓			✓	
*[lab, liq]	✓	✓	✓	✓	✓	✓	✓
*[lab, nas]		✓			✓		
*[liq, apprx]	✓					✓	✓
*[liq, dist]			✓			✓	
*[nas, apprx]	✓	✓	✓	✓	✓	✓	✓
*[nas, cont]	✓	✓	✓	✓	✓	✓	✓
*[nas, dist]			✓			✓	
*[nas, liq]	✓	✓	✓	✓	✓	✓	✓
*[nas, voice]	✓	✓	✓	✓	✓	✓	✓
*[voice, apprx]	✓	✓	✓	✓	✓	✓	✓
*[voice, cont]	✓	✓	✓	✓	✓	✓	✓
*[voice, dist]			✓			✓	
*[voice, liq]	✓	✓	✓	✓	✓	✓	✓

APPENDIX C

Data

This appendix contains the data for each child included in the study described in chapter 4. It is divided into three different sections; one each for the three levels of description: actual, target, and faithful. For every child, the word onset inventories are listed per stage, followed by the features that are active at each stage, and finally, the feature co-occurrence constraints.

C.1 Catootje

C.1.1 Inventories

Stage	*Inventory*																
1	p	b	t	k	m	n	l										
2	p	b	t	k	m	n	l	d	j								
3	p	b	t	k	m	n	l	d	j	z							
4	p	b	t	k	m	n	l	d	j	z	s						
5	p	b	t	k	m	n	l	d	j	z	s	r					
6	p	b	t	k	m	n	l	d	j	z	s	r	ʋ				
7	p	b	t	k	m	n	l	d	j	z	s	r	ʋ	v			
8	p	b	t	k	m	n	l	d	j	z	s	r	ʋ	v	ŋ		
9	p	b	t	k	m	n	l	d	j	z	s	r	ʋ	v	ŋ	f	
10	p	b	t	k	m	n	l	d	j	z	s	r	ʋ	v	ŋ	f	x

C.1.2 Features

Stage	*Features*
1	[labial], [voice], [dorsal], [nasal], [liquid]
2	[labial], [voice], [dorsal], [nasal], [liquid], [continuant], [approximant]
3	[labial], [voice], [dorsal], [nasal], [liquid], [continuant], [approximant]
4	[labial], [voice], [dorsal], [nasal], [liquid], [continuant], [approximant]
5	[labial], [voice], [dorsal], [nasal], [liquid], [continuant], [approximant]
6	[labial], [voice], [dorsal], [nasal], [liquid], [continuant], [approximant]
7	[labial], [voice], [dorsal], [nasal], [liquid], [continuant], [approximant]
8	[labial], [voice], [dorsal], [nasal], [liquid], [continuant], [approximant]
9	[labial], [voice], [dorsal], [nasal], [liquid], [continuant], [approximant]
10	[labial], [voice], [dorsal], [nasal], [liquid], [continuant], [approximant]

C.1.3 Overpredictions

Stage	*Overpredicted*
1	
2	
3	[cont]
4	
5	
6	[lab, voice, cont] [lab, cont]
7	[lab, cont]
8	[lab, cont]
9	
10	

C.1.4 Constraints

Stage	1	2	3	4	5	6	7	8	9	10
Constraints	[voice]→[lab]									
		[apprx]→[cont]	[apprx]→[cont]	[apprx]→[cont]	[apprx]→[cont]	[apprx]→[cont]	[apprx]→[cont]	[apprx]→[cont]	[apprx]→[cont]	[apprx]→[cont]
		[cont]→[apprx]								
		*[dors, apprx]	*[dors, apprx]	*[dors, apprx]	*[dors, apprx]	*[dors, apprx]	*[dors, apprx]	*[dors, apprx]	*[dors, apprx]	*[dors, apprx]
		*[dors, cont]	*[dors, cont]	*[dors, cont]	*[dors, cont]	*[dors, cont]	*[dors, cont]	*[dors, cont]	*[dors, cont]	
	*[dors, liq]	*[dors, liq]	*[dors, liq]	*[dors, liq]	*[dors, liq]	*[dors, liq]	*[dors, liq]	*[dors, liq]	*[dors, liq]	*[dors, liq]
	*[dors, nas]	*[dors, nas]	*[dors, nas]	*[dors, nas]	*[dors, nas]	*[dors, nas]	*[dors, nas]			
		*[lab, apprx]	*[lab, apprx]	*[lab, apprx]	*[lab, apprx]					
		*[lab, cont]	*[lab, cont]	*[lab, cont]	*[lab, cont]					
	*[lab, dors]	*[lab, dors]	*[lab, dors]	*[lab, dors]	*[lab, dors]	*[lab, dors]	*[lab, dors]	*[lab, dors]	*[lab, dors]	*[lab, dors]
	*[lab, liq]	*[lab, liq]	*[lab, liq]	*[lab, liq]	*[lab, liq]	*[lab, liq]	*[lab, liq]	*[lab, liq]	*[lab, liq]	*[lab, liq]
		*[liq, apprx]	*[liq, apprx]	*[liq, apprx]	*[liq, apprx]	*[liq, apprx]	*[liq, apprx]	*[liq, apprx]	*[liq, apprx]	*[liq, apprx]
		*[liq, cont]	*[liq, cont]	*[liq, cont]						
		*[nas, apprx]	*[nas, apprx]	*[nas, apprx]	*[nas, apprx]	*[nas, apprx]	*[nas, apprx]	*[nas, apprx]	*[nas, apprx]	*[nas, apprx]
		*[nas, cont]	*[nas, cont]	*[nas, cont]	*[nas, cont]	*[nas, cont]	*[nas, cont]	*[nas, cont]	*[nas, cont]	*[nas, cont]
	*[nas, liq]	*[nas, liq]	*[nas, liq]	*[nas, liq]	*[nas, liq]	*[nas, liq]	*[nas, liq]	*[nas, liq]	*[nas, liq]	*[nas, liq]
		*[voice, apprx]	*[voice, apprx]	*[voice, apprx]	*[voice, apprx]	*[voice, apprx]	*[voice, apprx]	*[voice, apprx]	*[voice, apprx]	*[voice, apprx]
		*[voice, cont]								
	*[voice, dors]	*[voice, dors]	*[voice, dors]	*[voice, dors]	*[voice, dors]	*[voice, dors]	*[voice, dors]	*[voice, dors]	*[voice, dors]	*[voice, dors]
	*[voice, liq]	*[voice, liq]	*[voice, liq]	*[voice, liq]	*[voice, liq]	*[voice, liq]	*[voice, liq]	*[voice, liq]	*[voice, liq]	*[voice, liq]
	*[voice, nas]	*[voice, nas]	*[voice, nas]	*[voice, nas]	*[voice, nas]	*[voice, nas]	*[voice, nas]	*[voice, nas]	*[voice, nas]	*[voice, nas]

C.2 Eva

C.2.1 Inventories

Stage	*Inventory*												
1	p	b	t	d									
2	p	b	t	d	n								
3	p	b	t	d	n	m	ʋ	j					
4	p	b	t	d	n	m	ʋ	j	l	s	z		
5	p	b	t	d	n	m	ʋ	j	l	s	z	f	
6	p	b	t	d	n	m	ʋ	j	l	s	z	f	k

C.2.2 Features

Stage	*Features*
1	[labial], [voice]
2	[labial], [voice], [nasal]
3	[labial], [voice], [nasal], [continuant], [approximant]
4	[labial], [voice], [nasal], [continuant], [approximant], [liquid]
5	[labial], [voice], [nasal], [continuant], [approximant], [liquid]
6	[labial], [voice], [nasal], [continuant], [approximant], [liquid], [dorsal]

C.2.3 Overpredictions

Stage	*Overpredicted*
1	
2	
3	
4	[lab, voice, cont] [lab, cont]
5	[lab, voice, cont]
6	[lab, voice, cont]

C.2.4 Constraints

Stage	1	2	3	4	5	6
Constraints		*[lab, nas]	[apprx]→[cont]	[apprx]→[cont]	[apprx]→[cont]	[apprx]→[cont]
			[cont]→[apprx]			
						*[apprx, dors]
				*[apprx, liq]	*[apprx, liq]	*[apprx, liq]
						*[cont, dors]
						*[cont, liq]
				*[cont, liq]	*[cont, liq]	
		*[lab, nas]				*[lab, dors]
				*[lab, liq]	*[lab, liq]	*[lab, liq]
						*[liq, dors]
			*[nas, apprx]	*[nas, apprx]	*[nas, apprx]	*[nas, apprx]
			*[nas, cont]	*[nas, cont]	*[nas, cont]	*[nas, cont]
						*[nas, dors]
				*[nas, liq]	*[nas, liq]	*[nas, liq]
			*[voice, apprx]	*[voice, apprx]	*[voice, apprx]	*[voice, apprx]
			*[voice, cont]			
						*[voice, dors]
				*[voice, liq]	*[voice, liq]	*[voice, liq]
		*[voice, nas]	*[voice, nas]	*[voice, nas]	*[voice, nas]	*[voice, nas]

C.3 Jarmo

C.3.1 Inventories

Stage	*Inventory*													
1	k													
2	k	t	d											
3	k	t	d	p	b									
4	k	t	d	p	b	m								
5	k	t	d	p	b	m	ʋ							
6	k	t	d	p	b	m	ʋ	l						
7	k	t	d	p	b	m	ʋ	l	n	f				
8	k	t	d	p	b	m	ʋ	l	n	f	s	x		
9	k	t	d	p	b	m	ʋ	l	n	f	s	x	r	ʃ

C.3.2 Features

Stage	*Features*
1	[dorsal]
2	[dorsal], [voice]
3	[dorsal], [voice], [labial]
4	[dorsal], [voice], [labial], [nasal]
5	[dorsal], [voice], [labial], [nasal], [continuant], [approximant]
6	[dorsal], [voice], [labial], [nasal], [continuant], [approximant], [liquid]
7	[dorsal], [voice], [labial], [nasal], [continuant], [approximant], [liquid]
8	[dorsal], [voice], [labial], [nasal], [continuant], [approximant], [liquid]
9	[dorsal], [voice], [labial], [nasal], [continuant], [approximant], [liquid], [distributed]

C.3.3 Overpredictions

Stage	*Overpredicted*
1	[]
2	
3	
4	
5	
6	
7	
8	
9	

C.3.4 Constraints

Stage	1	2	3	4	5	6	7	8	9
Constraints					[apprx]→[cont]	[apprx]→[cont]	[apprx]→[cont]	[apprx]→[cont]	[apprx]→[cont]
					[apprx]→[lab]	[apprx]→[lab]	[apprx]→[lab]	[apprx]→[lab]	[apprx]→[lab]
					[cont]→[apprx]	[cont]→[apprx]			
					[cont]→[lab]	[cont]→[lab]	[cont]→[lab]		
									[dist]→[cont]
									*[apprx, dist]
				[nas]→[lab]	[nas]→[lab]	[nas]→[lab]			
						*[apprx, liq]	*[apprx, liq]	*[apprx, liq]	*[apprx, liq]
						*[cont, liq]	*[cont, liq]	*[cont, liq]	
					*[dors, apprx]	*[dors, apprx]	*[dors, apprx]	*[dors, apprx]	*[dors, apprx]
					*[dors, cont]	*[dors, cont]	*[dors, cont]		
									*[dors, dist]
			*[dors, lab]	*[dors, lab]	*[dors, lab]	*[dors, lab]	*[dors, lab]	*[dors, lab]	*[dors, lab]
						*[dors, liq]	*[dors, liq]	*[dors, liq]	*[dors, liq]
				*[dors, nas]	*[dors, nas]	*[dors, nas]	*[dors, nas]	*[dors, nas]	*[dors, nas]
		*[dors, voice]	*[dors, voice]	*[dors, voice]	*[dors, voice]	*[dors, voice]	*[dors, voice]	*[dors, voice]	*[dors, voice]
									*[lab, dist]
						*[lab, liq]	*[lab, liq]	*[lab, liq]	*[lab, liq]
									*[liq, dist]
					*[nas, apprx]	*[nas, apprx]	*[nas, apprx]	*[nas, apprx]	*[nas, apprx]
					*[nas, cont]	*[nas, cont]	*[nas, cont]	*[nas, cont]	*[nas, cont]
									*[nas, dist]
						*[nas, liq]	*[nas, liq]	*[nas, liq]	*[nas, liq]
					*[voice, apprx]	*[voice, apprx]	*[voice, apprx]	*[voice, apprx]	*[voice, apprx]
					*[voice, cont]	*[voice, cont]	*[voice, cont]	*[voice, cont]	*[voice, cont]
									*[voice, dist]
						*[voice, liq]	*[voice, liq]	*[voice, liq]	*[voice, liq]
				*[voice, nas]	*[voice, nas]	*[voice, nas]	*[voice, nas]	*[voice, nas]	*[voice, nas]

C.4 Noortje

C.4.1 Inventories

Stage	*Inventory*												
1	m												
2	m	p	t	k									
3	m	p	t	k	n								
4	m	p	t	k	n	b							
5	m	p	t	k	n	b	d						
6	m	p	t	k	n	b	d	s					
7	m	p	t	k	n	b	d	s	ʋ				
8	m	p	t	k	n	b	d	s	ʋ	f			
9	m	p	t	k	n	b	d	s	ʋ	f	l	j	x

C.4.2 Features

Stage	*Features*
1	[nasal], [labial]
2	[nasal], [labial], [dorsal]
3	[nasal], [labial], [dorsal]
4	[nasal], [labial], [dorsal], [voice]
5	[nasal], [labial], [dorsal], [voice]
6	[nasal], [labial], [dorsal], [voice], [continuant]
7	[nasal], [labial], [dorsal], [voice], [continuant], [approximant]
8	[nasal], [labial], [dorsal], [voice], [continuant], [approximant]
9	[nasal], [labial], [dorsal], [voice], [continuant], [approximant], [liquid]

C.4.3 Overpredictions

Stage	*Overpredicted*
1	[]
2	
3	
4	
5	
6	
7	[lab, cont]
8	
9	

C.4.4 Constraints

Stage	1	2	3	4	5	6	7	8	9
Constraints							[apprx]→[cont]	[apprx]→[cont]	[apprx]→[cont]
							[apprx]→[lab]	[apprx]→[lab]	
	[lab]→[nas]	[nas]→[lab]							
	[nas]→[lab]								
				[voice]→[lab]					*[apprx, liq]
									*[cont, liq]
							*[dors, apprx]	*[dors, apprx]	*[dors, apprx]
						*[dors, cont]	*[dors, cont]	*[dors, cont]	
									*[dors, liq]
				*[dors, voice]	*[dors, voice]	*[dors, voice]	*[dors, voice]	*[dors, voice]	*[dors, voice]
						*[lab, cont]			
		*[lab, dors]	*[lab, dors]	*[lab, dors]	*[lab, dors]	*[lab, dors]	*[lab, dors]	*[lab, dors]	*[lab, dors]
									*[lab, liq]
							*[nas, apprx]	*[nas, apprx]	*[nas, apprx]
						*[nas, cont]	*[nas, cont]	*[nas, cont]	*[nas, cont]
		*[nas, dors]	*[nas, dors]	*[nas, dors]	*[nas, dors]	*[nas, dors]	*[nas, dors]	*[nas, dors]	*[nas, dors]
									*[nas, liq]
				*[nas, voice]	*[nas, voice]	*[nas, voice]	*[nas, voice]	*[nas, voice]	*[nas, voice]
							*[voice, apprx]	*[voice, apprx]	*[voice, apprx]
						*[voice, cont]	*[voice, cont]	*[voice, cont]	*[voice, cont]
									*[voice, liq]

C.5 Robin

C.5.1 Inventories

Stage	*Inventory*															
1	b	t	d													
2	b	t	d	p	n	s										
3	b	t	d	p	n	s	m									
4	b	t	d	p	n	s	m	f	ʋ							
5	b	t	d	p	n	s	m	f	ʋ	v						
6	b	t	d	p	n	s	m	f	ʋ	v	k	z				
7	b	t	d	p	n	s	m	f	ʋ	v	k	z	x			
8	b	t	d	p	n	s	m	f	ʋ	v	k	z	x	l		
9	b	t	d	p	n	s	m	f	ʋ	v	k	z	x	l	j	
10	b	t	d	p	n	s	m	f	ʋ	v	k	z	x	l	j	r

C.5.2 Features

Stage	*Features*
1	[voice], [labial]
2	[voice], [labial], [nasal], [continuant]
3	[voice], [labial], [nasal], [continuant]
4	[voice], [labial], [nasal], [continuant], [approximant]
5	[voice], [labial], [nasal], [continuant], [approximant]
6	[voice], [labial], [nasal], [continuant], [approximant], [dorsal]
7	[voice], [labial], [nasal], [continuant], [approximant], [dorsal]
8	[voice], [labial], [nasal], [continuant], [approximant], [dorsal], [liquid]
9	[voice], [labial], [nasal], [continuant], [approximant], [dorsal], [liquid]
10	[voice], [labial], [nasal], [continuant], [approximant], [dorsal], [liquid]

C.5.3 Overpredictions

Stage	*Overpredicted*
1	
2	
3	
4	
5	[voice, cont]
6	
7	
8	
9	
10	

C.5.4 Constraints

Stage	1	2	3	4	5	6	7	8	9	10
Constraints				[apprx]→[cont]	[apprx]→[cont]	[apprx]→[cont]	[apprx]→[cont]	[apprx]→[cont]	[apprx]→[cont]	[apprx]→[cont]
				[apprx]→[lab]	[apprx]→[lab]	[apprx]→[lab]	[apprx]→[lab]	[apprx]→[lab]		
	[lab]→[voice]									
						*[apprx, dors]	*[apprx, dors]	*[apprx, dors]	*[apprx, dors]	*[apprx, dors]
								*[apprx, liq]	*[apprx, liq]	*[apprx, liq]
						*[cont, dors]				
								*[cont, liq]	*[cont, liq]	
								*[dors, liq]	*[dors, liq]	*[dors, liq]
		*[lab, cont]	*[lab, cont]							
						*[lab, dors]	*[lab, dors]	*[lab, dors]	*[lab, dors]	*[lab, dors]
								*[lab, liq]	*[lab, liq]	*[lab, liq]
		*[lab, nas]								
				*[nas, apprx]	*[nas, apprx]	*[nas, apprx]	*[nas, apprx]	*[nas, apprx]	*[nas, apprx]	*[nas, apprx]
		*[nas, cont]	*[nas, cont]	*[nas, cont]	*[nas, cont]	*[nas, cont]	*[nas, cont]	*[nas, cont]	*[nas, cont]	*[nas, cont]
						*[nas, dors]	*[nas, dors]	*[nas, dors]	*[nas, dors]	*[nas, dors]
								*[nas, liq]	*[nas, liq]	*[nas, liq]
				*[voice, apprx]	*[voice, apprx]	*[voice, apprx]	*[voice, apprx]	*[voice, apprx]	*[voice, apprx]	*[voice, apprx]
		*[voice, cont]	*[voice, cont]	*[voice, cont]						
						*[voice, dors]	*[voice, dors]	*[voice, dors]	*[voice, dors]	*[voice, dors]
								*[voice, liq]	*[voice, liq]	*[voice, liq]
		*[voice, nas]	*[voice, nas]	*[voice, nas]	*[voice, nas]	*[voice, nas]	*[voice, nas]	*[voice, nas]	*[voice, nas]	*[voice, nas]

C.6 Tirza

C.6.1 Inventories

Stage	*Inventory*														
1	p	b	t	k	m	n	s								
2	p	b	t	k	m	n	s	d	l	x					
3	p	b	t	k	m	n	s	d	l	x	ʃ	j			
4	p	b	t	k	m	n	s	d	l	x	ʃ	j	r		
5	p	b	t	k	m	n	s	d	l	x	ʃ	j	r	f	ʋ

C.6.2 Features

Stage	*Features*
1	[labial], [voice], [dorsal], [nasal], [continuant]
2	[labial], [voice], [dorsal], [nasal], [continuant], [liquid]
3	[labial], [voice], [dorsal], [nasal], [continuant], [liquid], [distributed], [approximant]
4	[labial], [voice], [dorsal], [nasal], [continuant], [liquid], [distributed], [approximant]
5	[labial], [voice], [dorsal], [nasal], [continuant], [liquid], [distributed], [approximant]

C.6.3 Overpredictions

Stage	*Overpredicted*
1	
2	
3	
4	
5	

C.6.4 Constraints

Stage	1	2	3	4	5
Constraints			[apprx]→[cont]	[apprx]→[cont]	[apprx]→[cont]
			[dist]→[cont]	[dist]→[cont]	[dist]→[cont]
	[voice]→[lab]				
		*[cont, liq]	*[cont, liq]		
			*[dist, apprx]	*[dist, apprx]	*[dist, apprx]
			*[dors, apprx]	*[dors, apprx]	*[dors, apprx]
	*[dors, cont]				
			*[dors, dist]	*[dors, dist]	*[dors, dist]
		*[dors, liq]	*[dors, liq]	*[dors, liq]	*[dors, liq]
	*[dors, nas]	*[dors, nas]	*[dors, nas]	*[dors, nas]	*[dors, nas]
			*[lab, apprx]	*[lab, apprx]	
	*[lab, cont]	*[lab, cont]	*[lab, cont]	*[lab, cont]	
			*[lab, dist]	*[lab, dist]	*[lab, dist]
	*[lab, dors]	*[lab, dors]	*[lab, dors]	*[lab, dors]	*[lab, dors]
		*[lab, liq]	*[lab, liq]	*[lab, liq]	*[lab, liq]
			*[liq, apprx]	*[liq, apprx]	*[liq, apprx]
			*[liq, dist]	*[liq, dist]	*[liq, dist]
			*[nas, apprx]	*[nas, apprx]	*[nas, apprx]
	*[nas, cont]	*[nas, cont]	*[nas, cont]	*[nas, cont]	*[nas, cont]
			*[nas, dist]	*[nas, dist]	*[nas, dist]
		*[nas, liq]	*[nas, liq]	*[nas, liq]	*[nas, liq]
			*[voice, apprx]	*[voice, apprx]	*[voice, apprx]
	*[voice, cont]	*[voice, cont]	*[voice, cont]	*[voice, cont]	*[voice, cont]
			*[voice, dist]	*[voice, dist]	*[voice, dist]
	*[voice, dors]	*[voice, dors]	*[voice, dors]	*[voice, dors]	*[voice, dors]
		*[voice, liq]	*[voice, liq]	*[voice, liq]	*[voice, liq]
	*[voice, nas]	*[voice, nas]	*[voice, nas]	*[voice, nas]	*[voice, nas]

C.7 Tom

C.7.1 Inventories

Stage	*Inventory*													
1	p													
2	p	t												
3	p	t	b											
4	p	t	b	k	m	n	l	f						
5	p	t	b	k	m	n	l	f	ʋ	x				
6	p	t	b	k	m	n	l	f	ʋ	x	s			
7	p	t	b	k	m	n	l	f	ʋ	x	s	j		
8	p	t	b	k	m	n	l	f	ʋ	x	s	j	d	
9	p	t	b	k	m	n	l	f	ʋ	x	s	j	d	r

C.7.2 Features

Stage	*Features*
1	[labial]
2	[labial]
3	[labial], [voice]
4	[labial], [voice], [dorsal], [nasal], [liquid], [continuant]
5	[labial], [voice], [dorsal], [nasal], [liquid], [continuant], [approximant]
6	[labial], [voice], [dorsal], [nasal], [liquid], [continuant], [approximant]
7	[labial], [voice], [dorsal], [nasal], [liquid], [continuant], [approximant]
8	[labial], [voice], [dorsal], [nasal], [liquid], [continuant], [approximant]
9	[labial], [voice], [dorsal], [nasal], [liquid], [continuant], [approximant]

C.7.3 Overpredictions

Stage	*Overpredicted*
1	[]
2	
3	
4	
5	[cont]
6	
7	
8	
9	

C.7.4 Constraints

Stage	1	2	3	4	5	6	7	8	9
Constraints					[apprx]→[cont]	[apprx]→[cont]	[apprx]→[cont]	[apprx]→[cont]	[apprx]→[cont]
					[apprx]→[lab]	[apprx]→[lab]			
				[cont]→[lab]					
			[voice]→[lab]	[voice]→[lab]	[voice]→[lab]	[voice]→[lab]	[voice]→[lab]		
					*[dors, apprx]	*[dors, apprx]	*[dors, apprx]	*[dors, apprx]	*[dors, apprx]
				*[dors, cont]					
				*[dors, liq]	*[dors, liq]	*[dors, liq]	*[dors, liq]	*[dors, liq]	*[dors, liq]
				*[dors, nas]	*[dors, nas]	*[dors, nas]	*[dors, nas]	*[dors, nas]	*[dors, nas]
				*[lab, dors]	*[lab, dors]	*[lab, dors]	*[lab, dors]	*[lab, dors]	*[lab, dors]
				*[lab, liq]	*[lab, liq]	*[lab, liq]	*[lab, liq]	*[lab, liq]	*[lab, liq]
					*[liq, apprx]	*[liq, apprx]	*[liq, apprx]	*[liq, apprx]	*[liq, apprx]
				*[liq, cont]	*[liq, cont]	*[liq, cont]	*[liq, cont]	*[liq, cont]	
					*[nas, apprx]	*[nas, apprx]	*[nas, apprx]	*[nas, apprx]	*[nas, apprx]
				*[nas, cont]	*[nas, cont]	*[nas, cont]	*[nas, cont]	*[nas, cont]	*[nas, cont]
				*[nas, liq]	*[nas, liq]	*[nas, liq]	*[nas, liq]	*[nas, liq]	*[nas, liq]
					*[voice, apprx]	*[voice, apprx]	*[voice, apprx]	*[voice, apprx]	*[voice, apprx]
				*[voice, cont]	*[voice, cont]	*[voice, cont]	*[voice, cont]	*[voice, cont]	*[voice, cont]
				*[voice, dors]	*[voice, dors]	*[voice, dors]	*[voice, dors]	*[voice, dors]	*[voice, dors]
				*[voice, liq]	*[voice, liq]	*[voice, liq]	*[voice, liq]	*[voice, liq]	*[voice, liq]
				*[voice, nas]	*[voice, nas]	*[voice, nas]	*[voice, nas]	*[voice, nas]	*[voice, nas]

References

Adriaans, F., & Kager, R. (2010). Adding generalisation to statistical learning: the induction of phonotactics from continuous speech. *Journal of Memory and Language*, *62*, 311 - 331.

Anderson, S. R. (1985). *Phonology in the twentieth century. theories of rules and theories of representations.* Chicago and London: The University of Chicago Press.

Anderson, S. R. (2000). Reflections on 'On the phonetic rules of Russian'. *Folia Linguistica*, *34*, 11-28.

Archangeli, D. (1984). *Underspecification in yawelmani phonology and morphology.* Unpublished doctoral dissertation, MIT, Cambridge, MA. (Revised version published 1988 by Garland, New York (*Outstanding Dissertations in Linguistics*))

Archangeli, D. (1988). Aspects of underspecification theory. *Phonology*, *5*, 183-207.

Backley, P. (2011). *An introduction to element theory.* Edinburgh: Edinburgh University Press.

Barton, D. (1976). *The role of perception in the acquisition of phonology.* Unpublished doctoral dissertation, Unitversity of London, Bloomington.

Beckman, J. (1998). *Positional faithfulness.* Unpublished doctoral dissertation, University of Massachusetts, Amherst.

Beckman, M., & Edwards, J. (2000). The ontogeny of phonological categories and the primacy of lexical learning in linguistic development. *Child Development*, *71*, 240249.

Beckman, M., Yoneyama, K., & Edwards, J. (2003). Language specific and language universal aspects of lingual obstruent productions in Japanese acquiring children. *Journal of the Phonetic Society of Japan*, *7*.

Beers, M. (1995). *The phonology of normal developing and language-impaired children.* Unpublished doctoral dissertation, University of Amsterdam. (IFOTT)

Benua, L. (1997). *Transderivational identity.* Electronic Doctoral Dissertations for UMass Amherst. Retrieved from `http://scholarworks.umass.edu/dissertations/AAI9809307` (Paper AAI9809307)

Berg, T., & Schade, U. (2000). A local connectionist account of consonant harmony in child language. *Cognitive Science*, *24*, 123-149.

Best, C. T. (1995). Learning to perceive the sound pattern of English. *Advances in infancy research*, *9*.

Blair, M., Marczinski, C. A., Davis-Faroque, N., & Kertesz, A. (2007). A longitudinal study of language decline in alzheimer's disease and frontotemporal dementia. *Journal of the International Neuropscychological Society*, *13*, 237-245.

Blevins, J. (2007). What is evolutionary phonology? In *Evolutionary phonology - the emergence of sound patterns.* Cambridge: Cambridge University Press.

Boersma, P. (1997). How we learn variation, optimality and probability. *Proceedings of the Institute of Phonetic Sciences*, *21*.

Boersma, P. (1998). *Functional phonology.* Unpublished doctoral dissertation, University of Amsterdam.

Boersma, P. (2006a). A constraint-based explanation of the McGurk-effect. *manuscript*. (ROA: 869)

Boersma, P. (2006b). *Parallel bidirectional phonology and phonetics.* (Paper presented at III congreso internacional de fonetica e fonologia, Belo Horizonte)

Boersma, P. (2006c). A programme for bidirectional phonology and phonetics and their acquisition and evolution. *manuscript*.

Boersma, P. (2007a). Cue constraints and their interaction in phonological perception and production. *manuscript*. (ROA 944)

Boersma, P. (2007b). *The emergence of auditory contrast.* (Paper presented at GLOW 30, Tromsø)

Boersma, P., & Hamann, S. (2008). The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology*, *25*, 217-270.

Boersma, P., & Hayes, B. (2001). Empirical tests of the gradual learning algorithm. *Linguistics Inquiry*, *32*.

Boersma, P., & Levelt, C. (1999). Gradual constraint-ranking learning algorithm predicts acquisition order. *Proceedings of Child Language Research Forum*, *30*.

Booij, G. (1995). *The phonology of Dutch.* Oxford: Clarendon Press.

Cavirani, E. (2010). *Il vincolo nella fonologia generativa.* Unpublished master's thesis, Università de Pisa, Pisa.

Cavirani, E. (2012). *Brief history of the concept "constraint" in generative phonology.* Oral presentation at the Leiden University Center for Linguistics Phonology Reading Group.

Charette, M., & Göksel, A. (1998). Licensing constraints and vowel harmony in turkic languages. In E. Cyran (Ed.), *Structure and interpretation* (Vol. 4).

Lublin: PASE Studies and Monographs.

Chiat, S. (1983). Why *Mikey's* right and *my key's* wrong: The significance of stress and word boundaries in a child's output system. *Cognition*, *14*, 275-300.

Chiat, S. (1989). The influence of syllable position on children's production of consonants. *Clinical Linguistics and Phonetics*, *3*, 223-242.

Chomsky, N. (1951). *Morphophonemics of modern hebrew.* Unpublished master's thesis, University of Pennsylvania, Philadephia.

Chomsky, N., & Halle, M. (1968). *The sound pattern of english.* New York: Harper and Row Publishers.

Clements, G. N. (1990). The role of the sonority cycle in core syllabification. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology i, between the grammar and the physics of speech* (p. 283-333). Cambridge: Cambridge University Press.

Clements, G. N. (2003). Feature economy in sound systems. *Phonology*, *20*.

Clements, G. N., & Hume, E. (1995). The internal organization of speech sounds. In J. A. Goldsmith (Ed.), *The handbook of phonological theory.* Oxford: Blackwell.

Cruttenden, A. (1978). Assimilation in child language and elsewhere. *Journal of Child Language*, *5*, 373-378.

Dell, F. (1981). On the learnability of optional rules. *Linguistic Inquiry*, *12*, 31-37.

Dresher, B. E. (2009). *The contrastive hierarchy in phonology.* Cambridge: Cambridge University Press.

Eimas, P., Siqueland, E., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*.

Ernestus, M. (2000). *Voice assimilation and segment reduction in casual dutch.* Utrecht: LOT Dissertation Series.

Ernestus, M., & Baayen, H. (2001). Choosing between the dutch past-tense suffixes *-te* and *-de*. In T. van der Wouden & H. Broekhuis (Eds.), *Linguistics in the netherlands* (Vol. 18, p. 77-88). Amsterdam: John Benjamins Publishing Company.

Ferguson, C., & Farwell, C. (1975). Words and sounds in early language acquisition. *Language*, *51*.

Fikkert, P. (1994). *On the acquisition of prosodic structure.* HIL dissertations in linguistics 6. (Leiden University)

Fikkert, P. (2008). Developing representations and the emergence of phonology: evidence from perception and production. In *Laboratory phonology x.* Berlin: Mouton de Gruyter.

Fikkert, P., & Levelt, C. (2008). How does place fall into place? the lexicon and emergent constraints in children's developing phonological grammar. In P. Avery, B. E. Dresher, & K. Rice (Eds.), *Contrast in phonology. theory, perception, acquisition* (p. 231). Berlin: Mouton.

Fikkert, P., Levelt, C., & Zamuner, T. S. (2005). *Perception through pro-*

duction? Talk presented at NELS 2005 and ms. Radboud University of Nijmegen, Leiden University and University of British Columbia.

Flemming, E. (2004). Contrast and perceptual distinctiveness. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *The phonetic bases of markedness.* Cambridge: Cambridge University Press.

Flemming, E. (2006). The role of distinctiveness constraints in phonology. *manuscript*.

Friederici, A., & Wessels, J. (1993). Phonotactic knowledge of word boundaries and its use in infant speech perception. *Perception and Psychophysics*, *54*, 287-295.

Gnanadesikan, A. (1995). *Markedness and faithfulness in child phonology.* New Brunswick. Retrieved from `roa.rutgers.edu/files/67-0000/67-0000-GNANADESIKAN-0-0.PDF` (Manuscript: ROA 67)

Goad, H. (1997). Consonant harmony in child language: An optimality-theoretic account. In S. J. Hannahs & M. Young-Scholten (Eds.), *Focus on phonological acquisition* (p. 113142). Amsterdam & Philadelphia: John Benjamins Publishing Company.

Goldsmith, J., & Laks, B. (to appear). Generative phonology: its origins, its principles, and its successors. In L. Waugh, J. Joseph, & M. Monville-Burston (Eds.), *Cambridge history of linguistics.* Cambridge: Cambridge University Press.

Gonzalez-Gomez, N., Poltrock, S., & Nazzi, T. (2013). A "bat" is easier to learn than a "tab": Effects of relative phonotactic frequency on infant word learning. *PlosOne*, *8*.

Gussenhoven, C. (1999). Illustrations of the IPA: Dutch. In *Handbook of the International Phonetic Association.* Cambridge: Cambridge University Press.

Gussenhoven, C., & Jacobs, H. (2005). *Understanding phonology.* London: Hodder Arnold.

Hale, M., & Reiss, C. (1998). Formal and empirical arguments concerning phonological acquisition. *Linguistic Inquiry*, *29*, 656-683.

Hale, M., & Reiss, C. (2003). The Subset Principle in phonology: why the *tabula* can't be *rasa*. *Journal of Linguistics*, *39*, 219-244.

Hale, M., & Reiss, C. (2008). *The phonological enterprise.* Oxford: Oxford University Press.

Hall, D. C. (1999). On the geometry of auditory space. *Toronto Working Papers in Linguistics*, *17*, 134-157.

Hall, D. C. (2007). *The role and representation of contrast in phonological theory.* Unpublished doctoral dissertation, University of Toronto.

Hall, D. C. (2011). Phonological contrast and its phonetic enhancement: Dispersedness without dispersion. *Phonology*, *28*, 1-54.

Halle, M. (1957). *On the phonetic rules of russian.* unpublished paper read to the Annual Meeting of the Linguistic Society of America, Chicago.

Halle, M. (1959). *The sound pattern of russian.* The Hague: Mouton & Co.

Harris, J. (1994). *English sound structure.* Oxford: Blackwell.

Harris, J. (2009). Why final devoicing is weakening. In K. Nasukawa & P. Backley (Eds.), *Streng relations in phonology.* Berlin: Mouton de Gruyter.

Harris, J., & Lindsey, G. (1995). The elements of phonological representation. In J. Durand & F. Katamba (Eds.), *Frontiers of phonology: Atoms, structures, derivations* (p. 34-79). Harlow: Longman.

Hayes, B. (1999). Phonetically-driven phonology: The role of optimality theory and inductive grounding. In M. Darnell, E. Moravscik, M. Noonan, F. Newmeyer, & K. Wheatly (Eds.), *Functionalism and formalism in linguistics* (Vol. I: General Papers, p. 243-285). Amsterdam: John Benjamins Publishing Company.

Hayes, B., & Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, *39*, 379 - 440.

Hemphill, R. (1998). Perception/production asymmetries in the acquisition of phonological contrasts. *Journal of the Acoustic Society of America*, *104*, 1757-1758.

Honeybone, P. (2001/2002). *Germanic obstruent lenition: some mutual implications of theoretical and historical phonology.* Unpublished doctoral dissertation, University of Newcastle upon Tyne. (Submitted and defended 2001, graduation 2002)

Honeybone, P. (2005). Diachronic evidence in segmental phonology: the case of obstruent laryngeal specifications. In M. van Oostendorp & J. van de Weijer (Eds.), *The internal organization of phonological segments* (p. 319-354). Berlin: Mouton de Gruyter.

Hume, E. (1994). Representing the duality of glides. In G. Tsonlas & L. Nash (Eds.), *Les actes du congrès: Langues et grammaire i.* Paris: Université Paris 8.

Hyman, L. (1975). *Phonology theory and analysis.* Austin: Holt Rinehart Winston.

Ingram, D. (1981). *Procedure for the phonological analysis of children's language.* Baltimore, Md.: University Park Press.

Ingram, D. (1988). Jakobson revisited: some evidence from the acquisition of Polish phonology. *Lingua*, *75*, 55-82.

Ingram, D. (1989). *First language acquisition. method, description and explanation.* Cambridge: Cambridge University Press.

Inkelas, S., & Rose, Y. (2008). Positional neutralization: A case study from child language. *Language*, *83*.

Itô, J., Mester, A., & Padgett, J. (1995). Nc: Licensing and underspecification in Optimality Theory. *Linguistic Inquiry*, *26*. (ROA: 38)

Jackendoff, R. (1990). *Semantic structures.* Cambridge, MA: MIT Press.

Jakobson, R. (1941/1968). *Child language, aphasia and phonological universals.* The Hague: Mouton.

Jakobson, R., Fant, G., & Halle, M. (1952). *Preliminaries to speech analysis:*

the distinctive features and their correlates. Cambridge, MA: MIT Press.

Jakobson, R., & Halle, M. (1956). *Fundamentals of language.* The Hague: Mouton.

Jakobson, R., & Lotz, J. (1949). Notes on the French phonemic pattern. *Word*, *5*, 151-158.

Jusczyk, P. (1985). On characterizing the development of speech perception. In J. Mehler & R. Fox (Eds.), *Neonate cognition. beyond the blooming buzzing confusion.* Hillsdale, NJ: Lawrence Erlbaum Publishers.

Jusczyk, P., Friederici, A., Wessels, J., Svenkerud, V., & Jusczyk, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of memory and language*, *32*.

Jusczyk, P., & Hohne, E. (1997). Infants'memory for spoken words. *Science*, *277*(5334).

Jusczyk, P., Smolensky, P., & Allocco, T. (2002). How english-learning infants respond to markedness and faithfulness constraints. *Language Acquisition*, *10*, 31-73.

Kager, R. (1999). *Optimality theory.* Cambridge: Cambridge University Press.

Kenstowicz, M., & Kisseberth, C. (1977). *Topics in phonological theory.* New York, San Francisco and London: Academic Press.

Keyser, S. J., & Stevens, K. N. (2006). Enhancement and overlap in the speech chain. *Language*, *82*, 33-63.

Kiparsky, P. (1982). From cyclic phonology to lexical phonology. In H. van der Hulst & N. Smith (Eds.), *The structure of phonological representations* (Vol. 1, p. 13175). Dordrecht: Foris.

Kiparsky, P., & Pajusalu, K. (2003). Towards a typology of disharmony. *The Linguistic Review*, *20*, 217-241.

Koelle, S. W. (1968[1854]). *Outline of a grammar of the vei language: Together with a vei-english vocabulary and an account of the discovery and nature of the vei mode of syllabic writing.* Farnborough, UK: Gregg International.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics*, *50*.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*, 606-608.

Lahiri, A., & Evers, V. (1991). Palatalization and coronality. In C. Paradis & J.-F. Prunet (Eds.), *The special status of coronals* (p. 79-100). San Diego: Academic Press.

Lahiri, A., & Reetz, H. (2002). Underspecified recognition. In *Laboratory phonology 7.* Berlin: Mouton de Gruyter.

Leopold, W. (1939 – 1949). *Speech development of a bilingual child.* Evanston, IL: Nortwestern University Press.

Levelt, C. (1994). *On the acquisition of place.* HIL dissertations in linguistics

8. (Leiden University)

Levelt, C. (1995). *Unfaithful kids: Place of articulation patterns in early child language.* Paper presented at the Department of Cognitive Science Seminar, Johns Hopkins University, Baltimore, Maryland.

Levelt, C. (2011). Consonant harmony in child language. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell Companion to Phonology.* Oxford: Blackwell.

Levelt, C., Schiller, N. O., & Levelt, W. (1999/2000). The acquisition of syllable types. *Language Acquisition*, *8*, 237-264.

Levelt, C., & van Oostendorp, M. (2007). Feature co-occurrence constraints in l1 acquisition. *Linguistics in the Netherlands*, *24*.

Li, Y. (2007). *Differential acquisition of phonemic constrasts by infant word-learners: Does production recapitulate perception?* Unpublished doctoral dissertation, University of North Carolina at Chapel Hill.

Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, *48*, 839-862.

Linke, K., & van Oostendorp, M. (in preparation). *The grammar of vowel inventories - vowel harmony in Turkic languages.*

Lombardi, L. (1995). Laryngeal neutralization and syllable well-formedness. *Natural Language and Linguistic Theory*, *13*, 39-74.

MacKain, K. (1982). Assessing the role of experience on infants' speech discrimination. *Journal of Child Language*, *9*.

Maye, J., & Gerken, L. (2000). Learning phonemes without minimal pairs. In *Proceedings of the 24th annual boston university conference on language development.* Somerville, MA: Cascadilla Press.

Maye, J., & Weiss, D. (2003). Statistical cues facilitate infants' discrimination of difficult phonetic contrasts. In B. Beachley, A. Brown, & F. Conlin (Eds.), *Proceedings of the 27th annual boston university conference on language development.* Somerville, MA: Cascadilla Press.

Maye, J., Werker, J., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*.

McCarthy, J. J. (2003). Ot constraints are categorical. *Phonology*, *20*.

McCarthy, J. J. (2004). Headed spans and autosegmental spreading. In *The selected works of john j. mccarthy.* Retrieved from `http://works.bepress.com/john_j_mccarthy/60`

Menn, L. (1978). Phonological units in beginning speech. In A. Bell & J. B. Hooper (Eds.), *Syllables and segments* (p. 315334). Amsterdam: North-Holland.

Menn, L., & Matthei, E. (1992). The "two-lexicon" approach to child phonology: looking back, looking ahead. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications.* Parkton: York Publishing Company.

Mielke, J. (2004). *The emergence of distinctive features.* Unpublished doctoral dissertation, Ohio State University.

Mielke, J. (2008). *The emergence of distinctive features.* Oxford: Oxford University Press.

Morén, B. (2003). The parallel structures model of feature geometry. *Working papers of the Cornell Phonetics Laboratory*, *15*.

Morén, B. (2007). The division of labor between segment-internal structure and violable constraints. In S. Blaho, P. Bye, & M. Kraemer (Eds.), *Freedom of analysis?* New York: Mouton de Gruyter.

Nevins, A. (2010). *Locality in vowel harmony* (No. 55). Cambridge, MA: The MIT Press.

Padgett, J. (2002). Feature classes in phonology. *Language*, *78*.

Paradis, C. (1988). On constraints and repair strategies. *The Linguistic Review*, *6*, 71-97.

Paradis, C., & Prunet, J.-F. (1991). *The special status of coronals: Internal and external evidence.* San Diego: Academic Press.

Pater, J. (1997). Minimal violation and phonological development. *Language Acquisition*, *6*, 201–253.

Pater, J. (2004). Bridging the gap between receptive and productive development with minimally violable constraints. In R. Kager, J. Pater, & W. Zonneveld (Eds.), *Constraints in phonological acquisition* (p. 219-244). Cambridge: Cambridge University Press.

Pater, J., Stager, C. L., & Werker, J. F. (2004). The perceptual acquisition of phonological contrast. *Language*, *80*, 384-402.

Pater, J., & Werle, A. (2001). Typology and variation in child consonant harmony. In C. Fry, A. D. Green, & R. van de Vijver (Eds.), *Proceedings of hilp* (Vol. 5, p. 119139). Potsdam: University of Potsdam.

Pater, J., & Werle, A. (2003). Direction of assimilation in child consonant harmony. *Canadian Journal of Linguistics*, *48*, 385 - 408.

Peperkamp, S. (2003). Phonological acquisition: Recent attainments and new challenges. *Language and Speech*, *46*.

Peperkamp, S., Calvez, R. L., Nadal, J.-P., & Dupoux, E. (2006). The acquisition of allophonic rules: Statistical learning with linguistic constraints. *Cognition*, *101*, B31-B41.

Piggott, G., & van der Hulst, H. (1997). Locality and the nature of nasal harmony. *Lingua*, *103*, 85-112.

Preyer, W. (1895). *Die Seele des Kindes. Beobachtungen ¨ber die geistige Entwicklung des Menschen in den ersten Lebensjahren.* Leipzig: Th. Grieben's Verlag.

Prince, A. (2002). *Entailed ranking arguments.* Rutgers Optimality Archive, ROA-500.

Prince, A. (2007). The pursuit of theory. In P. de Lacy (Ed.), *Cambridge handbook of phonology.* Cambridge: Cambridge University Press.

Prince, A., & Smolensky, P. (1993/2004). *Optimality theory - constraint interaction in generative grammar.* Oxford: Blackwell Publishing. Retrieved from `http://roa.rutgers.edu` (Original publication date 1993 as technical report at the University of Colorado (CU-CS-696-95) and Rutgers University (RuCCS-TR-2))

Rice, K., & Avery, P. (2004). *The representational issue: the role of contrast in phonology.* (Paper presented at the 12th Manchester Phonology Meeting)

Roca, I., & Johnson, W. (1999). *A course in phonology.* Oxford: Blackwell Publishers.

Rose, Y. (2000). *Headedness and prosodic licensing in the l1 acquisition of phonology.* Unpublished doctoral dissertation, McGill University, Montréal.

Rose, Y., & MacWhinney, B. (to appear). The phonbank initiative. In J. Durand, U. Gut, & G. Kristoffersen (Eds.), *Handbook of corpus phonology.* Oxford: Oxford University Press.

Rose, Y., MacWhinney, B., Byrne, R., Hedlund, G., Maddocks, K., OBrien, P., & Wareham, T. (2006). Introducing phon: A software solution for the study of phonological acquisition. In D. Bamman, T. Magnitskaia, & C. Zaller (Eds.), *Proceedings of the 30th annual boston university conference on language development* (p. 489-500). Somerville, MA: Cascadilla Press.

Röttger, F. (1931). *Phonetische Gestaltbildung bei jungen Kindern.* München: C.H. Beck'sche Verlagsbuchhandlung.

Rvachew, S., & Andrews, E. (2002). The influence of syllable position on children's productions of consonants. *Clinical Linguistics and Phonetics*, *16*, 183-198.

Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8 month old infants. *Science*, *274*(5294), 1926-1928.

Saffran, J., & Thiessen, E. (2003). Pattern induction by infant language learners. *Developmental psychology*, *39*(3).

Samek-Lodovici, V., & Prince, A. (1999). Optima. *Rutgers Optimality Archive*. Retrieved from `http://roa.rutgers.edu`

Scheer, T. (2004). *A lateral theory of phonology - what is CVCV, and why should it be?* Berlin/New York: Mouton de Gruyter.

Scheer, T. (2011). *A guide to morphosyntax-phonology interface theories. how extra-phonological information is treated in phonology since trubetzkoys grenzsignale.* Berlin/New York: Mouton de Gruyter.

Sebregts, K. (ms.) *The sociophonetics and phonology of dutch r.* Utrecht: LOT.

Seidl, A., & Buckley, E. (2005). On the learning of arbitrary phonological rules. *Language Learning and Development*, *1*, 289316.

Shibatani, M. (1973). The role of surface phonetic constraints in generative phonology. *Language*, *49*, 87-106.

Singh, R. (1987). Well-formedness conditions and Phonological Theory. In W. Dressler (Ed.), *Phonologica 1984* (p. 273 - 286). Cambridge: Cambridge University Press.

Smith, N. V. (1973). *The acquisition of phonology.* Cambridge: Cambridge

University Press.

Smolensky, P. (1996). *The initial state and 'richness of the base' in child language* (Tech. Rep. No. JHU-CogSci-96-4). Baltimore, MD: Cognitive Science Department, Johns Hopkins University.

Smolensky, P., & Legendre, G. (2006). *The harmonic mind.* Cambridge, MA: The MIT Press.

Spelke, E. (2012). *Core knowledge and language: Insights from infants, and unanswered questions about children and adults.* Presentation delivered at the Core Knowledge, Language and Culture workshop, Lorentz Center, Leiden.

Spencer, A. (1986). Towards a theory of phonological development. *Lingua*, *68*, 3-38.

Stager, C., & Werker, J. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, *388*.

Stanley, R. (1967). Redundancy rules in phonology. *Language*, *43*, 393-436.

Stemberger, J., & Bernhart, B. (1998). *Handbook of phonological development from the perspective of constraint-based nonlinear phonology.* San Diego: Academic Press.

Stemberger, J., & Stoel-Gammon, C. (1989). *Underspecification and consonant harmony in child phonology.* (Ms. University of Minnesota at Minneapolis)

Stemberger, J., & Stoel-Gammon, C. (1991). The underspecification of coronals: Evidence from language acquisition and performance errors. In C. Paradis & J.-F. Prunet (Eds.), *The special status of coronals: Internal and external evidence* (p. 181-199). San Diego: Academic Press.

Steriade, D. (1987). Redundant values. In A. Bosch, B. Need, & E. Schiller (Eds.), *Cls 23: Papers from the 23rd annual regional meeting of the chicago linguistic society. part two: Parasession on autosegmental and metrical phonology* (p. 339-362). Chicago: Chicago Linguistic Society.

Stevens, K. N., & Keyser, S. J. (2010). Quantal theory, enhancement and overlap. *Journal of Phonetics*, *38*, 10-19.

Swingley, D. (2008). The roots of the early vocabulary in infants' learning from speech. *Current Directions in Psychological Science*, *17*.

Swingley, D., & Aslin, R. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*.

Tesar, B., & Smolensky, P. (2000). *Learnability in optimality theory.* Cambridge, MA: MIT Press.

Torgerson, W. S. (1963). *Theory and methods of scaling.* New York: Wiley.

van de Weijer, J. (1994). *Segmental structure and complex segments.* Unpublished doctoral dissertation, Leiden University, Leiden. (HIL Dissertations; 9)

van de Weijer, J. (1998). *Language input for word discovery.* Unpublished doctoral dissertation, Max Planck Institute for Psycholinguistics. (Max Planck Series in Psycholinguistics 9) van der Feest, S. (2007). *Building a phonological lexicon. the acquisition of the dutch voicing contrast in perception and production.* Unpublished doctoral dissertation, Radboud Universiteit Nijmegen, Nijmegen.

Van der Hulst, H. (1988). The geometry of vocalic features. In H. van der Hulst & N. Smith (Eds.), *Features, segmental structure and harmony processes* (Vol. 2, p. 77-125). Dordrecht: Foris.

Van der Hulst, H. (2012). A minimal framework for vowel harmony. In B. Botma & R. Noske (Eds.), *Phonological explorations: Empirical, theoretical and diachronic issues.* Berlin: Mouton de Gruyter.

Van der Hulst, H., & Smith, N. (1986). On neutral vowels. In K. Bogers, H. Van der Hulst, & M. Mous (Eds.), *The phonological representation of suprasegmentals* (p. 233-279). Dordrecht: Foris.

van Oostendorp, M. (2008). Incomplete devoicing in formal phonology. *Lingua*, *118*, 1362-1374.

van Oostendorp, M. (to appear). Final devoicing in french. In H. van Riemsdijk & M. van Oostendorp (Eds.), . Berlin: Mouton de Gruyter.

van Severen, L., Molemans, I., van den Berg, R., & Gillis, S. (2012). Consonant inventories in the spontaneous speech of young children: A bootstrapping procedure. *Clinical Linguistics and Phonetics*, *26*, 164-187.

van 't Veer, M. (2008). *Grammatical inventories - exploring a constraint based way of describing the world's sound systems.* Unpublished master's thesis, Leiden University.

van 't Veer, M. (2012). *New observations on the acquisition of liquids.* Talk delivered at TINdag 2012.

Vihman, M. M. (1978). Consonant harmony: Its scope and function in child language. In J. H. Greenberg, C. A. Ferguson, & E. A. Moravcsik (Eds.), *Universals of human language* (Vol. 2: Phonology, p. 281334). Stanford: Stanford University Press.

Werker, J. F., & Pegg, J. (1992). Infant speech perception and phonological acquisition. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, and implications.* Parkton: York Publishing Company.

Werker, J. F., & Tees, R. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*.

Wetzels, L., & Mascaró, J. (2001). The typology of voicing and devoicing. *Language*, *77*.

White, K., & Morgan, J. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language*, *59*, 114-132.

White, K., Peperkamp, S., Kirk, C., & Morgan, J. (2008). Rapid acquisition of phonological alternations by infants. *Cognition*, *107*, 238-265.

Yip, M. (2011). Lateral consonants. In *The Blackwell Companion to Phonology.* Oxford: Blackwell.

Zamuner, T., Gerken, L., & Hammond, M. (2005). The acquisition of phonology based on input: A closer look at the relation of cross-linguistic and child language data. *Lingua*, *10*.

Zoll, C. (1998). *"licensing" constraints in Optimality Theory.* (Manuscript, Massachusetts Institute of Technology)

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 representationele 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 Σ iff [F] is in Σ and [G] is in Σ (*c-constraint*)

b. F$\rightarrow$G
assign a violation mark for every segment Σ iff [F] is in Σ and [G] is not in Σ (*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), specificationele variatie (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 continuum 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 uitgebreide 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 verwervingstraject 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: /ʋ/. 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 derivationele 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 (vroege) 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 psycholingüistisch 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 bedoelt 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.

Curriculum Vitae

Burght Marinus (Marijn) van 't Veer werd in 1982 te Amsterdam geboren. Hij doorliep het gymnasium aan het Dr. W. A. Visser 't Hooftcollege te Leiden, waar hij in 2001 zijn VWO-diploma behaalde. Na een propedeusejaar bij de opleiding Spaanse Taal en Cultuur aan de Universiteit van Amsterdam, begon hij in 2002 aan de bovenbouwstudie Algemene Taalwetenschap aan de Universiteit Leiden. Naast zijn studie was hij als student-assistent betrokken bij onderzoeken van Dr. J.A van der Weijer (over complexe articulatie) en Dr. C.C. Levelt (perceptie-onderzoek bij jonge kinderen). In 2008 behaalde hij zijn doctoraaldiploma, waarna hij als promovendus bij het NWO-project "*Kindersprache, Aphasie und Algemeine Lautgesetze* Revisited" begon aan het werk dat zou leiden tot dit proefschrift.