Building a Phonological Inventory

Veer, B.M. van 't

Author: Veer, B.M. van 't
Title: Building a phonological inventory : feature co-occurrence constraints in acquisition
Issue Date: 2015-01-07

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]$\rightarrow$[Φ] and [Φ]$\rightarrow$[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

s̢ ṣ ʃ ɕ s̠ sʲ s s̪
2000 Hz ⟶ 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).

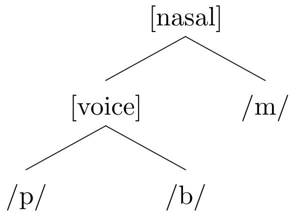
This criticism only holds, however, in so far as the Modified Contrastive Hierarchy aims to be a model of phonological computation. If this is not so, we see that the theory rather is a theory of information structure, which must somehow be encoded in a model of online computation. It is here that the 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.

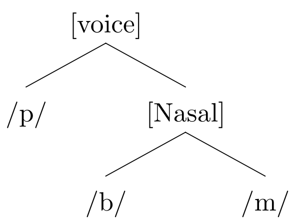
S_1: [±nasal] is acquired. The child has learned that some segments behave in a distinct way, to the exclusion of other segments. The contrast between [-nasal] and [+nasal] is acquired. The mutual exclusivity of both features is expressed by a c-constraint: *[-nasal, +nasal]. The members of the [+nasal] 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.