Radical Substance Free Phonology and Feature Learning

Journal:	*Revue canadienne de linguistique/Canadian Journal of Linguistics*
Manuscript ID	CJLRCL-SA-2020-0001
Manuscript Type:	Thematic Article
Keywords:	Phonology, substance-free grammar, features

Radical Substance Free Phonology and Feature Learning

This paper argues that phonological features have no substantive properties, instead, segments are assigned features by learning strategies set to the task of devising a computational system consistent with the requirements of UG. I address two problems for such a substance-free model. The first is the Card-Grammar problem, which has been suggested to argue for universal features, on the premise that otherwise language data cannot be stored in a fashion necessary to correct learning errors. The Card Grammar problem disappears in a suitably modular theory of mind with learned interfaces, where the mind still retains information not parsed in a particular grammar. The second is the need for a demonstration, not just an assertion, that a reasonable theory of grammar and learning lacking phonetic substance will actually yield a coherent system of feature assignments. This is accomplished by modeling the the learning of features necessary for the phonology of Kerewe.

Phonology; features; substance-free grammar; learning

1. Introduction

One of the most influential ideas of phonological theory, originating with Jakobson (1939), is that speech sounds are defined by the conjunction of a set of autonomous features. Jakobson proposes that “features” (as they are now known) are binary, and have acoustic manifestations. Rather than viewing the sounds standardly symbolized as [t], [ð], [s] as unanalyzable atoms, Distinctive Feature theory defines such segments, following the widely-used version proposed in Chomsky & Halle (1968), in terms of the specific universal feature specifications.[1]

(1)

$$\underset{[\text{t}]}{\begin{bmatrix} -\text{sonorant} \\ -\text{continuant} \\ +\text{coronal} \\ +\text{anterior} \\ -\text{voice} \\ -\text{strident} \end{bmatrix}} \qquad \underset{[\text{ð}]}{\begin{bmatrix} -\text{sonorant} \\ +\text{continuant} \\ +\text{coronal} \\ +\text{anterior} \\ +\text{voice} \\ -\text{strident} \end{bmatrix}} \qquad \underset{[\text{s}]}{\begin{bmatrix} -\text{sonorant} \\ +\text{continuant} \\ +\text{coronal} \\ +\text{anterior} \\ -\text{voice} \\ +\text{strident} \end{bmatrix}}$$

This system for segment definition allows classes of sounds to be defined via the features which the members of the class have in common. All three of the above three segments have in common the features [–sonorant, +coronal, +anterior], the segments {t,s} have in common [–sonorant, +coronal, +anterior, –voice], {t,ð} have in common [–sonorant, +coronal, +anterior, –strident], and {ð,s} have in common [–sonorant, +coronal, +anterior, +continuant]. Phonological rules are (by hypothesis) stated in terms of sets of segments defined by the features.

[1] This is only a partial list of applicable feature specifications.

There have been numerous changes within the theory of features from models such as Jakobson, Fant & Halle (1952) up to geometric models such as Clements & Hume (1995), especially regarding how features relate to one another, and to the question of feature privativity. Theories of features have, for the most part, tacitly accepted two basic premises. First, it is generally assumed that features are defined in terms of phonetic properties – articulatory, acoustic, or both. Second, it is assumed that all languages draw on the same pre-defined set of features, which are provided by Universal Grammar. This paper sets forth an alternative view, that features are not defined in terms of physical substance and are not listed in UG. Instead, the features employed in the grammar of any language are arrived at inductively, strictly on the basis of the representational requirements of a grammar. There are two basic requirements of phonological grammars. First, every sound-type of a language is represented by a unique configuration of features and prosodic elements: if [b] and [p] are both sounds of a language, they have different representations. Second, rules of a phonological grammar apply in configurations that are identified via representational differences. When a rule applies differently in the context of [b] versus [p], the difference comes from an interaction between how the rule is stated, and the representational differences between the segments. This theory of features is formal, in that the inferred features are a consequence of the form of phonological rules which refer to features. Features are not based on physical substance.

The precision with which features have been defined has varied over the course of generative phonology. The SPE theory of features is one of the most phonetically fine-grained theories, having at least 26 defined features. Subsequent theories, especially within feature-geometric trends, generally posit fewer features which are more abstract. For example, the distinction between front and back vowels in SPE theory is governed by the feature [back] where front vowels are [–back] and the distinction between alveolars and labials is governed by [coronal], whereas in Unified Features Theory (Clements & Hume 1995) they are governed by the same feature. The phonetic definition of [coronal] in UFT is less specific, in abstracting away from exact details of tongue-raising. Likewise, in Bradshaw (1999), voicing and L tone are represented with a single feature. The Parallel Structures Model (Morén 2003) abstracts away from phonetic definitions even further, so that the features [open] and [closed] may distinguish different tone registers, vowel height, laryngeal constriction, or fricative vs. stop, depending on what structure these features are predicated of. Element Theory (Kaye, Lowenstamm & Vergnaud 1985 *et seq*.) likewise abstracts away from the fine-grained details of articulation and acoustics so that the unary element "H" may be realized as High tone, aspiration, or frication.

Odden (2006) and Blaho (2008) advance the claim that phonological theory does not require or allow features to have any substantive definition at all. That theory, Radical Substance Free Phonology (RSFP), holds that phonological primitives have no intrinsic phonetic content, and no aspect of phonological grammar refers to the phonology-external substance of the segments referred to by the primitives. In RSFP [coronal] is a formal label which, in conjunction with other formal labels, gives every segment a unique identifier, and these identifiers can be exploited to refer to classes of segments in operations of grammars – the rules. The feature [coronal] makes no claim about the tongue or F_2.

This paper outlines how RSFP "works" in a system of phonological computations. The discussion of the theory of phonological computations (section 2) is brief because the theory adds nothing to existing theories of computation, and aims to take away much. The main focus of the paper is showing how feature learning is possible. Section 3 addresses the card-grammar parsing problem raised by Hale & Reiss (2008), which might be seen as presenting a strong argument in favor of detailed phonetic definitions of features as part of phonology. Closer scrutiny of the argument shows that H&R correctly identify the fact that UG must encode the formal concept "feature", and the syntax of features and computations must be in UG so that these facts can serve as the basis for learning phonological rules, which constitute the primary evidence for assignment of features to segments. But that is all that is needed in UG. The problem with the card-grammar argument is that it does not distinguish between the information available to the extra-phonological interface device that parses sounds into grammatical segments which are operated on by the grammar, and the information actually available within the entire grammar or the mind. Putative phonetic properties of features are at most an aspect of *phonetic* computations, and only under certain assumptions about what the grammar of phonetic implementation is.

If phonological grammars have no recourse to substantive properties of language sounds, how can a system of phonological rules be learned, if it is not known in advance whether [t] is voiced, or continuant, or coronal? If (as I assume) rules in grammars are stated in terms of sets of segments identified by feature expressions, and the feature expression for a segment is based on how a segment behaves with respect to a rule, which comes first, the rule or the feature assignment, and how is this apparent circle escaped? The answer, elaborated on in subsequent sections, is that while the end result of language acquisition is a system of rules and representations in a grammar *qua* computational device, acquisition is not carried out by grammar – it is carried out by a separate cognitive mechanism, one which is aware of the formal requirements of grammar, but which operates on pre-phonological and possibly pre-linguistic cognitive objects. What comes first is that which is perceived – a sound. In order for a sound to be integrated into a grammar as part of a language, some faculty of learning must attempt to create rules which are consistent with the requirements of the grammatical module that the rules are in. Those rules may be right or wrong. If the system of rules does not correctly describe which segments become which other segments, the rules must be corrected. This is, in fact, general automatic learning, and not a special form of phonological learning. What makes "phonological learning" at all identifiable is that the object which is learned is specifically a phonological computation or representation.

2. The Formal Theory of Features and Rules

As indicated above, the formal theory of features in RSFP is minimal. RSFP is, at the level of grammatical theory, part of an answer to the question "What are phonological representations?", answered in a way that is consistent with the principles of Formal Phonology (FP: see Odden 2013). In this paper I only address segmental features, leaving analogous questions about suprasegmental representation to separate investigation. A segment is, formally, a set of features[2] and dominance relations. A segment is formally the root of a feature tree, and everything under it. This is, of course, a simplified description of various autosegmental theories of features, including Clementsian (1985) or Sageyan (1986) feature geometry, Unified Features Theory, and PSM.[3] It is simpler than Sageyan geometry in eliminating certain stipulations such as the claim that Coronal, Labial, Dorsal must be immediately dominated by Place. Any dominance relation possible in one of these representational accounts is, in lieu of compelling theoretical reasoning to the contrary, also possible in RSFP.

2.1. Privativity

An important formal question arises as to the nature of features in RSFP: are features privative, or are they value-attribute pairs, that is, binary (or more)? Or, for that matter, is this a fact that needs to be learned by the child? At this point, I apply the logic of FP to the question, concluding that features are privative. First, the alternative that a child must also learn whether features in a language are binary versus privative will be dismissed on grounds based on the card-grammar logic: there has to be some fixed fact that serves as the innate basis for learning. That fixed fact is the formal nature of phonological computations. A child does not learn from scratch what it means to be "a rule in the grammar", the child knows that. What the child does need to learn is, what are the rules of a particular grammar are.

Still, we could assume that features are all binary, or all privative – what fact of the substance-free formalist framework says that features are privative? Suppose a child is learning a language with the segmental inventory of American English, and a rule does something to a set of vowels before the consonants [t d tʃ dʒ θ ð s ʃ n l r], but not before any other consonants. This class of segments has some characteristic that makes them similar in some way, to the exclusion of any other segment in the language, and we label that characteristic [coronal]. We could also use a juxtaposition of value and attribute, [+coronal]. The latter theory of the syntax of features implies that the complement of this class is automatically inferred: [p b m f k g ŋ] are automatically [–coronal]. The defect of the theory of binary features is subtle: it makes a claim for which there is insufficient evidence. There is no evidence that the complement of [t d tʃ dʒ θ ð s ʃ n l r] has such a similarity in phonological behavior. In not labelling [p b m f k g ŋ], the privative theory does not put "lack of evidence" on a par with "evidence".

Privative theory is formally simpler, because it posits one concept – a feature – whereas binary feature theory requires three concepts – "value", "attribute", and "feature" (the conjunction of value and attribute), and it requires unnecessary theoretical propositions to the effect that a value cannot exist independent of an attribute (there are

---

[2] Conventionally, there is at least a two-way distinction drawn between necessarily terminal features such as [voice] and potentially pre-terminal nodes like Place. This distinction is superfluous in the present substance-free account and can be replaced with the more general term "node", see Author (in progress b). In this paper I will simply talk of "Place" or "Laryngeal" as being features.

[3] Dependency theories of segment representation are sufficiently different from autosegmental models that I will not pursue the question of whether element-learning is likewise possible in that approach.

no floating plusses), nor can an attribute exist independent of a value. As emphasized in Odden (2013), Occam's Razor is an essential tool of theorizing in FP. If there were sufficient evidence for introducing all three concepts into phonological theory, it would be possible to do so, but in lieu of such evidence, the simplest system of causal concepts is to be adopted. That system is the theory of privative features.

A reasonable counter-argument against privative features is that it seems to imply – incorrectly – that voicelessness cannot spread. This potential prediction might follow if voicedness and not voicelessness were the universally-assumed specified value, where voicelessness is the result of not specifying a segment with [voice]. It is generally assumed[4] that a rule can refer to the fact that a segment has a particular feature, but cannot refer to the lack of a specification for a feature, from which it would follow that a rule deleting [voice] before a segment lacking the specification [voice] would be a formally impossible rule – thus, voicelessness cannot spread.

There are at least two reasons why behavioral asymmetry is not mandated by privative features, especially in RSFP. First, the presumption that voicing is universally implemented via the feature [voice] is untenable. The premise that all languages employ the specification [voice] is directly counter to the premises of RSFP, where it is possible for a language to specify the laryngeal distinction as [voiceless] (names are arbitrary, features have no intrinsic interpretation). Indeed, nothing in RSFP precludes having [voice] and [voiceless] coexist in a language. When voiceless segments act as a class under a rule, that behavior motivates the existence of a feature [voiceless]. Nothing prevents a language from having a fact pattern motivating a feature [voice] as well as a fact pattern motivating a feature [voiceless]. The second reason why privativity does not mandate the phonological inertness of the presumed "unmarked" member of an opposition is that alleged spreading of a supposedly unspecified terminal node can be accomplished by spreading of a dominating preterminal node.

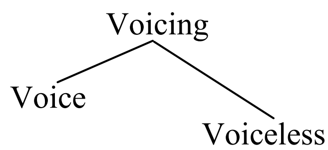
The logic of this argument is made clear in Lombardi (1991), who shows the illusivity of the asymmetry claim. There may exist a dominating node (such as Laryngeal), and when voicelessness seems to spread, it is not the terminal voicing feature that spreads, it is the dominating Laryngeal node that spreads, even when there is no lower-level feature corresponding to voicelessness. Spreading of voicelessness is formally just as possible as spreading of voicedness.

(2)

[4] This is a statement about assumptions made in the field, not an acceptance of the validity of the assumption, which needs independent scrutiny and validation, and that is beyond the scope of the present paper. However, I do in fact adopt that assumption here.

Moreover, if a language can have both the features “voiced” and “voiceless”, and two features can be dominated by a node, then a representation like (3) is possible.

(3)



This structure is analogous to the widely-accepted hypotheses that Laryngeal dominates constricted glottis, spread glottis and voice, and that Place dominates Labial, Coronal, Dorsal, Radical. Since the technical device of nodes dominating nodes can apparently accomplish everything that is accomplished by binary feature specifications, and nodes dominating nodes is a representational device with independent usage (at least in any autosegmental theory), it follows that without other evidence to support treating features as value-attribute pairs, features should be treated privatively.[5]

2.2. Order of features

A characteristic of many theories of feature geometry is that dominance relations are prespecified by UG: Place dominates Coronal and not the other way around, Laryngeal dominates Voice, but not Nasal. Because RSFP denies that specific features are in UG, it follows that dominance relations are not in UG. RSFP allows language-specific conditions, so if there is evidence for a node “Place” and for a node “Coronal”, there may also be evidence for a condition on representations saying that Coronal may (or must) be dominated by Place. Such rules are learned on the basis of pertinent evidence – the requirements for setting up a system of grammatical computations.

2.3. The theory of rules

The RSFP analysis of feature learning depends on having a theory of phonological computations combined with general principles of conceptual learning. That means: we have to have a theory of phonological computations. My argument is conducted in a minimalist FP-friendly rule-based model, but this is not a prerequisite for an RSFP analysis of feature learning. See Blaho (2008) for an OT-based instantiation of RSFP, though a computational model where constraints are learned. In Blaho’s model, UG does not contain specific constraints like Ident[Nasal], it contains a constraint form: Ident[__], where existence of the target feature must be learned. In that account, when the existence of a fea-

[5] Lombardi’s analysis also depends heavily on an invariant and phonetically-defined set of features and dominance relations, plus various stipulations regarding well-formed structures, which result in the prediction that if voicelessness spreads by some rule, glottalization and aspiration must also spread. Those premises, which also characterize UFT, are not valid in FP and RSFP. RSFP primarily posits the device of domination, allowing a language recourse to conditions on domination such as “laryngeal dominates voice” if there is evidence for such a condition.

ture is learned – [coronal] for example – the existence of a class of constraints Ident[coronal], also OCP[coronal], *[coronal] and so on is thereby learned.

It is well beyond the scope of the paper to articulate and defend a complete theory of phonological rules, nor is it necessary to do so. No novel assumptions about rule theory are required to facilitate feature learning. I follow the standard assumption that a rule combines two representation, the structural description which identifies the class of strings that undergo the rule, and a structural change which describes how the string is changed. A fundamental desideratum of autosegmental rule theory has been that the structural change be reduced to a single operation. I assume, specifically, that a rule is limited to the insertion or deletion of a single dominance relation or node in the representation. That is, a rule can insert of delete an association line, or a segment, feature, or other phonological constituent. I do not assume abbreviatory schemata as encountered in the SPE theory of phonology, or rule-independent repairs or limitation constraints as encountered in parametric versions of autosegmental rule theory. There are a number of independent questions about rule theory that need to be answered, irrespective of the theory of features assumed. For example, is “structure preservation” a valid phonological concept? are rules subject to “blocking” conditions, and if so, what is the syntax of such expressions?

A salient unanswered question will remain unanswered for lack of relevant evidence: what is the theory (if any) of feature realization. Under the premise that *s* in some language, perhaps English, is [coronal, anterior, continuant], a legitimate and interesting question is, by what mechanism is [coronal, anterior, continuant] realized as *s* in the language? The general answer is, it happens in and after the phonetic component, and we need to at least understand the nature of phonetic computations (see discussion in section 3.3). We need a fully-articulated theory of phonetic implementation. There seem to be some differences in how *s* is pronounced in American English, Andalusian Spanish, Basque, Icelandic, Korean and Modern Greek, and except in the case of Basque, these differences do not seem to reflect phonological patterns, instead, they are simply language-specific details about how a sound is pronounced. Does the phonetic grammar directly control the superior longitudinal, inferior longitudinal, transverse and genioglossus muscles? Or, more likely, does the phonetic component produce a more abstract symbolic output which is the input to some other cognitive model? We simply don’t know how phonological features are physically implemented, we just know that they are, perhaps indirectly.

For the sake of concreteness, I assume that feature matrices like [coronal, anterior, continuant] are somehow interpreted as “*s*, as pronounced in this language”, and I use transcriptional symbols to stand for a lower-level cognitive representation related to a grammatical feature matrix – perhaps an auditory image, a set of articulatory instructions, or maybe an autonomous linguistic phonetic object. A feature specification is thus a direct or indirect index to other cognitive entities, some of which are outside of grammar.

3. The Card Grammar argument for pre-defined features

Hale & Reiss (2008, ch. 2) put the question of feature innateness into sharp focus, pointing to the necessity of some innate basis for learning. As they put it, "ya gotta start with something". That is, not everything about language can be learned. This is a restatement of the Innateness of Primitives Principle (Pylyshyn, Fodor, Jackendoff), expressed in Jackendoff (1990: 40) as:

> In any computational theory, 'learning' can consist only of creating novel combinations of primitives already innately available[6]

In order for a child to learn that *s* has specification X and *θ* has specification Y, the child must be able to store the fact that the language has a segment *s* which is distinct from *θ*. This section scrutinizes (and rejects) the implication that this entails phonetically-defined innate features as part of Universal Grammar. The argument to that effect ultimately depends on a premise that is not self-evidently true, that phonological feature assignments arise deterministically from an unlearned direct interface between the auditory system and the phonological component. Given the alternative that features are assigned by a learned interface which relates pre-phonological cognitive representations to representations suited to phonological computation, there is no logical impediment to learning phonological features.

3.1. The Card Grammar argument

H&R advance the "card grammar" argument for innateness of phonological features based on the supposed impossibility of learning contrastive feature assignments without the specific features being already available in UG. Here I review the argument, laying bare the required assumptions. The argument holds if you make certain assumptions, and not otherwise.

The argument form explores the consequences of different models of UG for what could be learned, using a stripped-down model of language, Card Grammar, where a grammar is a set of conditions on cards. Each model of card UG provides a set of primitive features and operators defined for those features. A card *c* is grammatical with respect to grammar G iff *c* satisfies the conditions imposed by G. In example UG_1, the primitives are NUMBERCARD (henceforth "#"), the suits ♣, ♦, ♥ and ♠, and the operator AND. Grammar G_1 is the rule [#], which means that only a card with the property [NUMBERCARD] is grammatical. A physical card |K♦|[7] is ungrammatical, because it does

---

[6] I assume that the intended claim is that the thing-combinations which constitute "learning" are *reducible* to innate primitive, not that they *are* themselves innate primitives, but it is possible that the authors believe that learned concepts like "apple", "fork", "canid", "mammal" and so on are themselves innate cognitive primitives.

[7] By "physical card", I mean a mind-external actual card, notated with "||", and not a mental representation of a card. This is analogous to H&R's use of body brackets as in 🯅kʰæt🯅 to refer to a physical production of

not have the property [#]. Such an input is parsed by UG_1 simply as [♦] – the physical property identifiable as “K” is not assigned any representation by UG_1.[8] The cards |6♦|, |3♣| are parsed as [#♦], [#♣] (likewise |3♦|, |6♣| are parsed as [#♦], [#♣]). These cards are grammatical since they follow the rule requiring [#]. G_3 has the rule [♠], which means that any of |2♠…A♠| are grammatical. The physical cards |2♠| to |10♠| are parsed by UG_1 as the same thing, [#♠], and |J♠| through |A♠| are parsed as [♠]. Following the rule of G_3, |2♠…10♠| are all judged to be grammatical [#♠], and |J♠| through |A♠| are judged to be ungrammatical *[♠]. The important point is that given the particular primitives provided by UG_1, there are only eight possible mental card representations: [#♣, #♦, #♥, #♠, ♣, ♦, ♥, ♠].[9]

UG_3 has a richer set of representational primitives, including [picturecard] = [P], individual numbers [2, 3, 4, 5, 6, 7, 8, 9, 10], and [±red]. This allows |2♥| and |3♥| to be parsed distinctly as [2 +red] and [3 +red], but does not allow |2♥| to be distinguished from |2♦|, which are both represented as [2 +red]. On the other hand, UG_4 has the very impoverished inventory of features, containing only [♦]. Accordingly, physical cards are all parsed as either [♦] or are not parsed at all. The upshot of this analysis is that unless UG has a very rich representational inventory, all physical inputs will be parsed as the same thing, or not parsed at all, and therefore there would be no basis for learning that |2♥| is in fact distinct from |2♦|. Thus, the existence of particular features could not be learned. As they say p. 38:

> It should now be obvious that we are heading toward the conclusion that children must “know” (i.e. have innate access to) the set of phonological features used in all of the languages of the world. This is how the IofPP will be extended in this chapter; but it is equally clear that the same conclusion holds for primitive operators like the AND and OR of card languages, or whatever are the operators of real grammars (in both phonology and syntax).

H&R do not actually claim that the set of innate phonological features has to be phonetically defined, instead, there has to be a set of features available in UG. The feature [coronal] must by this argument be available in UG, but it need not have anything to do with the tongue. Other assumptions (see Hale, Kissock & Reiss 2007) could have the consequence that features have their traditional phonetic consequences.

“cat”. I believe my use of these brackets clarifies their original intent, which is to say how a mind-external stimulus maps to a mental representation.

[8] An alternative would be that J, Q, K, A are parsed as [#]. A detailed study of human perception of parrot “speech” might shed light on how humans parse stimuli that are well outside the norms of human production, but clearly parsing sounds as speech is not very strict.

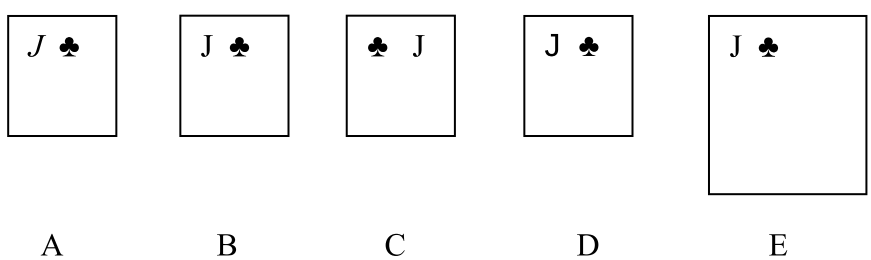
[9] In fn. 3 p. 32, H&R say that only one of these suit features can characterize any given card, thus [♣♦] or [♣♣] are not considered. This could either be because of the nature of the physical inputs, or because this is a property of all versions of card UG. For our purposes it does not matter what restricts inputs, but in language it obviously matters substantially.

3.2. What parses?

The concepts of representation and input relied on above require a cognitive domain, since there are no absolute inputs, there are inputs *to* something and representations *in* something. In saying that a sound is parsed, we mean that it is assigned a featural representation in phonology.

An intended speech sound enters the human body, is mechanically transduced to a pressure wave in the cochlea, causing neuronal excitation which are the neural basis for creating a first representation of a sound. This signal is parsed into other representations, proceeding to the auditory cortex, and perhaps ultimately the phonological component. In cognitively processing an instance of *s*, B♯, or  (the sound of a car crash), different representations are created, but at early stages of processing, the same kind of representation is created – a raw acoustic image (the structure of the ear does not sort out whether a sound is linguistic). The physical sound |s| may be interpreted as a language sound, but it can also be interpreted as escaping steam, or as a person or other thing making a snake noise. It is an empirical question exactly how this happens: what is clear is that UG does not directly convert the pattern of neuronal excitation from the cochlea into phonological features. The fundamental question for understanding how physical sound maps to phonological repesentations is, what is the sequence of representation-to-representation mappings that takes place prior to phonology?

Modular theories of grammar typically hold that each mental module has a distinct set of primitives. As a mental object passes through the various modules, it is subject to that many symbol-to-symbol translations: these translation devices are called interfaces. To clarify the consequences of interfaces for the Card Grammar argument, I set forth the Revised Card Grammar argument, which introduces multiple modules and interfaces in a theory of mind, MT_1, demonstrating that phonolocal features can be learned. Assume the following physical inputs.



Perhaps because of the nature of the sensory apparatus, MT_1 does not process differences in card size at all, so in transducing physical inputs to a first mental form, the difference between |E| and the remaining inputs is irretrievably lost. This is analogous to the loss of information in an acoustic signal above or below certain frequencies. This physical transduction provides a representation in module Mo_1 where |A-D| are all distinct, and results in $[_{\mathrm{Mo1}}J\clubsuit]$, $[_{\mathrm{Mo1}}\mathrm{J}\clubsuit]$, $[_{\mathrm{Mo1}}\clubsuit\mathrm{J}]$ and $[_{\mathrm{Mo1}}\mathbf{J}\clubsuit]$ respectively. |E| is not represented distinctly in Mo_1: we will say that only the size aspect of the signal is lost so it is parsed as $[_{\mathrm{Mo1}}\mathrm{J}\clubsuit]$, but it is possible that an input is entirely rejected. After the computations of Mo_1 are performed, the output is passed through interface $I_{1\text{-}2}$ which converts and discards some information, in this case the typeface difference, with the result that Mo_2 receives $[_{Mo2}\mathit{J}\clubsuit] \leftarrow |A|$, $[_{Mo2}\mathrm{J}\clubsuit] \leftarrow |B,D,E|$ and $[_{Mo2}\clubsuit\mathrm{J}] \leftarrow |C|$. With respect to language, this discarding of earlier information is similar to discarded spatial information in auditory processing. Although we can hear that "hat" uttered close to the right ear is not the same as "hat" uttered close to the left ear, that difference is thrown away on the path to grammatical computation.

Since our goal is to understand how features could be learned, we now consider the case where some aspect of interface $I_{2\text{-}3}$ between Mo_2 and Mo_3 depends on experience: the mapping may be learned. Learning constitutes the postulation of a hypothesis as to the nature of the system which operates on the relevant data. As new data become available, hypotheses may be reinforced, or they may be subject to correction when the hypothesis is contradicted by known facts. To draw a phonological analog, a hypothesis could be initially advanced by a German-learning child, based on known facts, that the German word [bunt] 'federal' is underlying /bunt/. That hypothesis will eventually be overridden in light of complicated facts regarding the German word [bunt] meaning "colorful", and other forms of the words 'federal' and 'colorful'. Eventually the child corrects the system of rules and representations such that 'federal' is /bund/ and there is a devoicing rule.

Based on initial data, a child might postulate that $[_{Mo2}\mathrm{J}\clubsuit] \rightarrow [_{Mo3}\mathrm{J}\clubsuit]$ and $[_{Mo2}\clubsuit\mathrm{J}] \rightarrow [_{Mo3}\clubsuit\mathrm{J}]$, and also that $[_{Mo2}\mathit{J}\clubsuit] \rightarrow [_{Mo3}\mathrm{J}\clubsuit]$, that is, the distinction between $[_{Mo2}\mathrm{J}\clubsuit]$ and $[_{Mo2}\mathit{J}\clubsuit]$ may be eliminated. At this point, Mo_3 contains only $[_{Mo3}\clubsuit\mathrm{J}, \clubsuit\mathrm{J}]$: but the basis for retrieving the third distinction still exists in Mo_2. Subsequent experience can establish the incorrectness of the interface mapping, for example it might be discovered that computations in Mo_3 treat supposed instances of $[_{Mo3}\mathrm{J}\clubsuit]$ differently, depending on whether they derive from $[_{Mo2}\mathrm{J}\clubsuit]$ or $[_{Mo2}\mathit{J}\clubsuit]$. The mind still stores a distinct representation of |A|, namely $[_{Mo2}\mathit{J}\clubsuit]$, and the interface mapping can easily be corrected so that $[_{Mo2}\mathit{J}\clubsuit] \rightarrow [_{Mo3}\mathit{J}\clubsuit]$. Although computation within M_3 is limited to the primitives of M_3, learning about the interface from M_2 to M_3 is not an operation in M_3. The mind which is still learning the rules still has access to information that was erroneously discarded. The key to solving the Card Grammar problem is that the phonological component of a grammar does not learn, the mind learns about the phonological component.

H&R use the Card Grammar argument to claim that the traditional Subset Principle of learning theory is wrong. Given the Card Grammar argument, it would be impossible to correct the hypothesis that a language only has the vowels [i a u] to the hypothesis that the language has [i ɪ ɛ a ɔ ʊ u]. If parsing of inputs is absolutely limited to just [i a u], then a child could not gain awareness that [ɪ ɛ ɔ ʊ] also exist. As shown above, a child can learn that the inventory [i a u] is an error, since phonological feature assignment is not performed by the cochlear nucleus.

3.3. What part of the mind would have universal features?

The card-grammar argument is not specifically about phonological features, it is presumptively an argument about the broad language faculty. Even though the language faculty must have some primitive properties that form the basis for learning, we cannot conclude that the relevant primitives are specifically a list of features, as standardly envisioned in phonological theory. Still, H&R propose that universal features are an innate part of UG, which raises two important questions: what is UG, and where in UG do these claimed universal features exist?

3.3.1. What is UG?

Chomsky (1965: 6) advances the concept of UG as an architectural mechanism regulating the operation of specific grammars:

> The grammar of a particular language, then, is to be supplemented by a universal grammar that accommodates the creative aspect of language use and expresses the deep-seated regularities which, being universal, are omitted from the grammar itself. Therefore it is quite proper for a grammar to discuss only exceptions and irregularities in any detail. It is only when supplemented by a universal grammar that the grammar of a language provides a full account of the speaker-hearer's competence.

The historical source of the concept UG is rationalist philosophy of the Renaissance period, which Chomsky is attempting to put on a cognitive footing. Chomsky & Halle (1968: 43) also characterize UG:

> A universal grammar is a system of conditions that characterize any human language, a theory of essential properties of human language. It is reasonable to suppose that the principle of the transformational cycle and the principles of organization of grammar that we have formulated in terms of certain notational conventions are, if correct, a part of universal grammar rather than of the particular grammar of English.

Roberts (2016) summarizes the theory of UG as "the scientific theory of the genetic component of the language faculty", that it is "the theory of that feature of the genetically given human cognitive capacity which makes language possible, and at the same time defines a possible human language". This architectural view is the one which I assume.

H&R p. 2 state an alternative view:

> Once we accept the existence of a language faculty, it is also uncontroversial that this faculty has an initial state, before any experience with a particular language. Under this view Universal Grammar, the theory of this initial state, is a topic of study, not a hypothesis.

This follows the view of Chomsky (1980), seeing UG to be the initial state of the child or the language faculty:

> In a highly idealized picture of language acquisition, UG is taken to be a characterization of the child's pre-linguistic initial state. p. 7
>
> These and many other questions must be considered in the development of a comprehensive theory of UG, as a characterization of the initial state of the language faculty. p. 138

Under this view, it is hard to see how UG has potency for regulating computation of linguistic forms, or how UG could have an effect on an already quadralingual child learning a fifth language.[10] Obviously, these radically different views of the nature of UG substantially affect arguments about whether UG must contain pre-specified features. Especially relevant to the debate is the extent to which a special language-specific theory of learning is necessary, or can the universal properties of language be explained by the interaction between the fixed architecture of grammar and general mechanisms of automatic learning such as employed in learning eating, walking or visual tracking. As proposed here, the theory of grammar (UG) specifies what are possible representations and computations in the phonological component. Language acquisition results from such general learning strategies, which are set to the task of inducing a set of UG-consistent computations and representations given the primary linguistic data. Even if there exists specific learning strategies for language, it does not follow that the learning device needs to be pre-coded with the specific features of a language. What needs to be known is simply that sounds are represented and operated on in phonology using features. The features resulting from learning must integrate with the grammar that is also being learned.

3.3.2. Where are the universal features, if anywhere?

Even if universal substantive sound properties are somewhere encoded in UG, it is an open question what those properties are, and how they relate to distinctive features as exist in phonology. It is possible that the linguistic phonetic component contains a collection of genetically predetermined primitives, which influence what objects are presented to the phonological component, thus indirectly determining how physical sounds map to feature matrices. For example, language sounds in the linguistic phonetic component could be single symbols analogous to IPA letters,[11] and the theory of linguistic phonetic sounds might reduce to being whatever expressions can be constructed in the IPA – e.g. [$_{\text{phonetic}}$ i ɪ i ï i̞ i̝]. If this is the alphabet with which phonetic forms are represented and how phonetic computations are carried out, we would have a basis for postulating defeasible interface hypotheses about what phonology receives. A child could learn, based on evidence, that [$_{\text{phonetic}}$ ï i̞] behave phonologically like distinct objects, and could therefore recover from the error of assuming that [$_{\text{phonetic}}$ ï i̞] both map to [$_{\text{phonological}}$ ɪ]. A child could

[10] Experience with language begins *in utero* starting at around 30 weeks. It remains completely unclear at what stage the brain has developed to the point that the language organ is fully developed: arbitrarily, I assume that the language faculty is fully formed prior to 30 weeks.

[11] I also intend that [ɪ̃] be interpreted as one symbol, not as the composition of two symbols with independent interpretations.

then learn a phonological contrast between tense and lax vowels, even when that hypothesis had been previously rejected.

In order to argue about the nature of the phonological component based on the phonetic component, we need a theory of that component as a computational device, and obviously we need to determine whether there even exists such a component. See Hamann (2011) for an overview of issues regarding the relation of phonetics to phonology. As observed by Hamann (2011: 203),

> the nature of the interface cannot be unearthed by experimental studies alone. It depends to a considerable part on the theoretical assumptions we make, and on the aim we have in mind with our phonological and phonetic descriptions.

If devoiced and underlyingly voiceless obstruents in German and Dutch are physically or perceptually distinct, this fact has no implications for the theory of phonological computation in the event that final devoicing is a phonetic rule, but has substantial implications for phonology if this is a phonological rule.

The question of whether a phonetic component exists must be determined based on how the assumption solves what would otherwise be severe problems for the theory of phonological computation and representation. Chomsky & Halle (1968) deny that there is a component of phonetic computation, holding that phonetic and phonological features and computations are indistinguishable. The SPE approach to phonetics reduces all language-specific differences in phonetic properties to either non-contrastive features (for instance [suction], invoked for phonetic differences in the production of labiovelars between Guang languages such as Late versus those in Yoruba and Ibibio), or to rules assigning integer values to a particular feature, where in SPE [1F_i] is the maximum degree of property F_i and [nF_i] for some value *n* substantially greater than 1 is the smallest degree of the property, total lack. In that framework, the slight difference between "phonetics" and phonology is (by hypothesis) that features do not have whole number values in the lexicon, they only have the values {*u,m*,+,–,0}, and phonological outputs (which are directly interpreted by some motor control device) always have numeric feature coefficients assigned to features.[12] The consequences of this approach for the theory of phonological computation are not trivial to assess, and the literature on phonetic interpretation *qua* phonology within the "no-phonetics" tradition is not extensive.[13] It is clear that within this approach, the theory of phonological computations would have to be expanded to include arithmetic operations and integer or fractional variables, which have no phonological justification.

[12] Postal (1968: 61-2) claims that "given some dictionary representation (systematic base form), this will automatially be mapped onto some phonetic representation (in general of course not a fully correct one), even without the application of any particular rules of the grammar, by virtue of universal principles (rules) for interpreting the binary phonological values as n-ary phonetic values", so that all phonological matrices are guaranteed to be physically interpretable with some default scalar value associated with "+" and "–".

[13] The primary works within that tradition are King (1969), Fromkin (1972), Peters (1973), Anderson (1974), Johnson (1975) and Clifton (1976).

Ascertaining whether there is a component of linguistic phonetics is necessary, so that we can know what constitutes the input to the transducer yielding phonological representations as outputs. Indeed, we need a theory of transducers into grammatical components. One possibility, similar to the SPE program, is suggested in Hale, Kissock & Reiss (2006): there are two transducers between phonology and lower-level body functions, a "transduction of features (the input) to some gestural score… transduction of a percept (the input) to features (the output)", where "these two transducers are innate and invariant—they are identical in all humans (barring some specific neurological impairment) and do not change over time or experience (i.e., they do not 'learn')". It is proposed that

> we can consider the transduction process, too, as invariant in that the *relationship* or *mapping* between a particular feature bundle and a particular gestural score is a deterministic (and thus consistent) conversion process and, similarly, that the relationship or mapping of a particular perceptual input to a feature bundle is deterministic.

Additionally, they propose that transductions can be context-sensitive. Clearly, there is a substantial difference of opinion regarding the nature of phonological transductions. In the perspective advanced here, transductions between grammatical components are learned, and are also probably context-free symbol-replacement lists.

We cannot expect to resolve questions about the theory of transduction into / from grammar or within grammar here. We can, however, ask what kinds of facts could be relevant to answering these question. The main question of interest is whether there is a linguistic component of phonetics. The potential evidence for a linguistic phonetic component falls into three categories. First, phonological features refer to ranges of physical facts, not precise measurements of such facts. The range of facts related to a feature may depend on which language the feature is instantiated in. For example, languages with the feature [spread glottis] on consonants differ in how a segment is realized when specified or not specified with that feature, via language-specific target VOT values. Second, the realization of a feature may vary in a language-specific way that depends on surrounding context. For example, F_0 values of a H or L tone may be adjusted upwards or downwards from a target value, depending on surrounding tonal context (e.g. H before L may be subject to a language-specific process of pitch-raising). Such crosslinguistic variation is extremely informative, since it provides a basis for constructing a theory of phonetic computations, which allows us to evaluate the adequacy of competing theories of phonetic grammar. Finally, the time course of the realization of a feature may be language-specific, thus lip protrusion associated with a rounded consonant might, on a language-specific basis, be timed close to the release of a consonant, or is might be timed earlier and be easily detectable on a preceding segment. Below we consider specific facts that lend prima facia credibility to the claim that there is a language-specific component of phonetics. Needless to say, the argument cannot be properly evaluated without a theory of phonological computations, since the alternative hypothesis (as set forth in SPE) is that by definition, all linguistic variation comes from phonological rules, so the theory of phonology must include whatever devices are necessary to enable such computations.

As an example of language specific degree in the realization of a feature, consider how the feature [spread glottis] (the feature underlying the phenomenon of aspiration) is realized as differences in voice onset time, since this is a well-known example of variation between languages. Cho & Ladefoged (1999) document an example of degree-variation between languages in the voice onset time lag which implements an aspiration contrast, in languages with a phonemic contrast.

(4)

	[k]	[kʰ]	N speakers
	msc	*msc*	
Jalapa Mazatec	23	80	6
Gaelic	28	73	11
Apache	31	80	8
Khonama Angami	20	91	6
Tlingit	28	128	4
Hupa	44	84	3
Navaho	45	154	7

As Cho & Ladefoged note, crosslinguistic comparison of VOT requires "a body of data from a number of widely different languages, all of which have been collected and analyzed in the same way". There is no information on token variability within and across these languages, so it is impossible to know whether the measured mean VOT of 23 msc in Jalapa Mazatec [k] is statistically different from the 28 msc of Tlingit. Based on the magnitude of differences, it is reasonable to conclude that [k] falls into at least 2 different subtypes, with Khonama Angami having the shortest VOT and Navaho having the longest, and [kʰ] falls into at least 3 subtypes with Gaelic having the shortest VOT, Navaho having the longest and Tlingit being in the middle – though Khonama Angami probably represents a 4th type. Within the SPE model of phonetic implementation, this suggests language-specific integer targets along the following lines, where e.g. Navaho [kʰ] is assigned the value [1spr.gl.] and Gaelic [kʰ] is [4spr.gl].[14]

(5)

	[k]	[kʰ]
Jalapa Mazatec	7	4
Gaelic	7	4
Apache	7	4
Khonama Angami	7	3
Tlingit	7	2
Hupa	6	4
Navaho	6	1

[14] The lack of [5spr.gl.] corresponds to a lack of language in the above subset with VOT in the neighborhood of 50-70 msc VOT for velars in these example languages. Cho & Ladefoged indicate a mean VOT of 56 msc for Yapese, which has no aspiration contrast.

These phonetic differences between languages are fodder for a theory of phonetic computation. Similar language differences in vowel realization are well-known, see for example Disner (1983), who finds that [i] in German and Norwegian differ in that [i] is higher in German (has lower F_1) – this result holds for bilingual speakers, and depends on which language the individual is speaking. See Vaux & Samuels (2015) for further discussion of language-specific segment target differences and the problem which they pose for dispersion theory.

Contextual differences in phonetic realization of segments can also be language-specific, though this issue has not been as well studied. Lebanese Arabic, Logoori, Finnish and North Saami all have a four-way contrast between VCV, VC:V, V:CV and V:C:V structures. All four languages differ in how these prosodic subclasses are realised in terms of segment duration.[15] For example vowel length in Finnish involves much more prolongation, compared to what is found in Logoori. In addition, the languages differ in how durations are determined in V:C:V. In Logoori, there is no interaction between consonant and vowel duration, whereas Finnish and North Saami have a durational trade-off where neither vowels nor consonants in V:C:V are lengthened as much as expected.

(6)

Language	*Prosodic context*	*V duration*	*C duration*
Lebanese Arabic	VC	78	84
	VC:	77	182
	V:C	166	99
	V:C:	149	181
Logoori	VC	76	99
	VC:	81	201
	V:C	146	97
	V:C:	140	217
Finnish	VC	109	95
	VC:	121	227
	V:C	240	97
	V:C:	215	178
North Saami	VC	79	73
	VC:	76	235
	V:C	169	90
	V:C:	148	181

15 Data from North Saami and Logoori come from my own research. Lebanese Arabic data are presented in Khattab & Al-Tamimi (2014), who give mean durations in the four relevant contexts, along with information on the statistical significance of different durational means. Data for Finnish derives from Dunn (1993), who presents numeric data in Appendix 2 separated according to speaker, an additional extraneous factor of V_2 length, and distinguishing /p/ from /m/. The Finnish duration values in (6) are the mean of reported V and C durations within the 4-way prosodic subclasses studied here, averaging across speakers, segmental differences in C, and disregarding vowel length in σ_2.

Taking the context before __CV to best reveal the duration target for long versus short vowels and the context V__V to best reveal the duration target for long versus short consonants, we arrive at the following differences between languages in the long-to-short ratio.

(7)

	V:/V	C:/C
Lebanese Arabic	2.13	2.17
Logoori	1.92	2.03
Finnish	2.20	2.39
North Saami	2.14	3.22

Arabic and North Saami have essentially the same degree of prolongation associated with vowel length. Finnish lengthens long vowels more than Arabic and North Saami do, and Logoori lengthens vowels less than those languages do. Similarly, North Saami lengthens long consonants most substantially, and Logoori the least. These differences can be expressed as a cross-linguistically variable ratio determining how much longer long vowels or consonants are compared to short vowels and consonants.

Beyond differences in target duration associated with contrastive length for vowels or consonants, there are also language-specific contextual differences in how consonant and vowel length is computed. One pattern found in these data (and in comparable data from other studies) is that long vowels have greater duration before a single consonant, and shorter duration before a geminate. A second effect is that long consonants have greater duration after short V than they do after long V:. The degree of shortening of long vowels and consonants next to long consonants and vowels depends on the language. (8) gives the degree of shortening of long vowels and consonants associated with adjacent long segments. The first column is the duration of V: before long C: divided by the duration of V: before short C in the language, and the second column is the duration of C: after long V: divided by the duration of C: after short V in that language.

(8)

	V: $_{C:/\_C}$	C: $_{V:\_/V\_}$
Lebanese Arabic	0.90	0.99_{ns}
Logoori	0.96_{ns}	1.08_{ns}
Finnish	0.90	0.78
North Saami	0.88	0.77

The Logoori pattern is the simplest: segment length is realized as simple doubling of a segment's duration target, and there is no significant interaction between V: or C: duration as a function of following C-length or preceding V-length, respectively. The Arabic pattern is similar to that of Logoori, but a long vowel is somewhat shorter before C: than it is before C, and that difference is statistically significant.[16] Contextual differences for

[16] In Logoori, the duration differences between long consonants associated with long-V versus short-V context are not statistically significant, likewise the differences between long vowels associated with long-

both V-length and C-length in Finnish and North Saami are statistically significant. Most obviously, duration of long consonants in those two languages is strongly influenced by the length of the preceding vowel – a long consonant has only about 3/4 of its expected duration, based on the duration after a short vowel.

3.4. What segments does phonology receive, and from where?

The present theory of feature learning assumes that children mentally categorize tokens of sounds, from which a system of computations and a feature analysis of those sounds emerges. The model in section 4 is based on the logic that if segments $\{s_1 \ldots s_i\}$ computationally group together and $\{s_{i+1} \ldots s_j\}$ are excluded, a feature expression selects $\{s_1 \ldots s_i\}$ but not $\{s_{i+1} \ldots s_j\}$. The child must therefore know what the segments of the language are. In hypothesizing a feature assignment based on a rule where [b,d,g] become [p,t,k] before an obstruent, it is highly relevant to know if any of [dʒ,z,bʰ,pʰ] also exist in the language. Feature learning presupposes prior analysis of the stream of speech – a child must have knowledge of words like [bunt], [bundə]. It is insufficient to experience acoustic waveforms: there must already be analysis of continuous speech into segments.

Segmentation of physical sound has two aspects, reduction to a sequence of discrete mental units, and analysis of units into types. Discretization and categorization are not specifically linguistic functions. To understand the pre-grammatical prerequisites for phonological analysis, we must know how a child learns that physical signals {I,II,III} are discretized into sequences $\{(a_1,a_2,a_3),(b_1,b_2),(c_1,c_2,c_3),\ldots\}$, and we must know how a child determines that discrete tokens $\{a_1,b_1,c_1\ldots\}$ are subsumed under one category X, and tokens $\{a_2,b_2,c_2\ldots\}$ are subsumed under a separate category Y.

A continuous signal must obviously be physically converted from a form of external energy to something in the mind, whose nature is determined by the sensory apparatus (e.g. the inability of human eyes to detect x-rays, the inaudibility of a 5 Hz sound at normal, safe amplitudes). Physical mechanisms within the head (including the outer ear) cause inner hair cells of the cochlea to transduce physical sound to a pattern of electrical impulses, resulting in a tonotopic map which is interpreted in the primary auditory cortex. Much cognitive processing takes place, the end point of interest for us being that the continuous signal is converted into a series of things and relationships between things.

Categorization of things means discerning that parsed tokens {a,b} are similar in a perceivable way, and are distinguishable from {c} in that same way. A sine wave at 100 Hz is similar to a square wave at 100 Hz, and distinct from a sine wave at 105 Hz. A sine wave at 100 Hz is, likewise, similar to a sine wave at 105 Hz, and distinct in that respect from a square or triangular wave at 100 Hz. A sine wave at 10,000 Hz might not, on the other hand, be perceptibly distinguishable from a sine wave at 10,005 Hz. Speech-sound categorization is carried out at a much higher level than is involved in detecting differences and similarities between constructed signals differing in F_0 and amplitude distributions, indeed spectral properties rarely relate to categories (conventional groupings with

C versus short-C context are not significant. Khattab & Al-Tamimi (2014) report that the difference degree of vowel shortening associated with following C: is statistically significant.

an attached label such as "brassy"), and frequency properties relate to labels not available to most people ("middle C", "high C", "C", "B sharp"). Though we don't know how it happens, we know that it does happen, that children learn that a certain range of physical sounds are "the same thing" in their language.

The aspect of sound-to-symbol conversion most relevant to phonology is that which is specifically about language. While it might be interesting to know how people can learn to identify different musical instruments or notes based on sound, this is not the same task as learning speech segments. Starting from the assumption (apparently made by Hale, Kissock & Reiss) that humans can perceive very subtle distinctions in speech sounds, learning to correctly categorize speech sounds involves learning which differences are inconsequential, and which ones are important. The range of variation in the phonetic properties of segments observed in a collection of tokens of the word 'cat' is an example of inconsequential difference. So too, probably, is the difference observed in a collection of tokens from a number of speakers, at least to the extent that they are speaking the same language form. Such variation has in common that it is not observed to correlate with anything linguistically relevant. A case where variation in sound realization might correlate with a property of the grammar is when it systematically correlates with other sounds – the appearance of perceptible [ɪᵊ] before [q] and [i] anywhere else could signal a rule-governed distinction. It could also signal a nonlinguistic physical necessity imposed by the nature of the articulatory gestures required to utter [i] followed by [q]. This raises the question of what is a child's genetically-dictated "knowledge" of anatomy and the acoustic consequences of that anatomy. Do we know in advance of experience that it takes N milliseconds to move the tongue from point A to B, for all possible points? Or do we learn this in the course of babbling, by observing that attempts to produce perfect [iq] always result in something like [ɪᵊq]? Alternatively, do we learn that it is possible to produce both [ɪᵊq] and [iq], but in the language of the environment, one nevers encounter [iq] even though one encounters [ik]?

If a child has learned that a fact about a language sound is linguistically relevant (is not an unavoidable physical requirement of doing something else), this does not say whether it is expressed in the phonetic grammar, as opposed to in the phonological grammar. A child is in a privileged position to answer the question because it is not prejudiced by the presumptions of a transcription, which linguists must overcome. In reporting that a word is pronounced [ɪᵊq], linguists build a phonological analysis into the lowest level of data reporting, rather than using a non-linguistic system of tongue-movement notation. A classical example of building phonological analysis into data reporting is the case of Marshallese vowels. Marshallese has been said to have 12 pure vowels and 24 diphthongs at the phonetic level. Research leading up to Bender (1968) revealed a rich system of vowels and consonants, including previously non-obvious consonant qualities (rounding and palatality) and correlations between vowel and consonant quality. Bender (1968) sets forth a phonological analysis which reduces the set of phonological vowels from 12 to 4 – a set of central vowels differing in height, which are neutral for frontness / backness and roundness. The richer set of apparent surface vowels such as [o] can treated as arising from a phonetic process interpretating vowel phonemes specified only for height. Bender (1968) explicitly recognized the problem of earlier impressionistic transcriptions like /jok/, /koj/, stating (p. 20) that

> relistening to an item which had been transcribed as /jok/ revealed the actual quality to move from front to back with increasing rounding, all at mid height: [tʸᵉᵊokʷ]. Similarly, /koj/ came to be perceived as [kʷᵒəᵉtʸ]. And as the phonetic facts of other mixed environments were reexamined, each proved to be capable of similar interpretation as resulting from competing consonantal influences on a less fully specified vowel

A benefit of attributing some surface vowel qualities to rules involving surrounding consonants is a simplification in rules for computing the surface forms of affixes, as noted by Bender.

Choi (1992) investigates this question acoustically, arguing that there is a continuous interpolation from one consonantal articulatory state to another in cases like [lʲᵉəᶺtˠ] 'well-sifted', [pˠᶺəᵒkʷ] 'wet'. Apparent short monophthongs like [o, e] only appear between consonants of the same secondary articulation (all consonants are analyzed as palatalized, rounded, or velarized), e.g. [rʷorʷ] 'bark', [tʲɛpʲ] 'cheek', [pˠɯpˠ] 'triggerfish'. Choi posits that vowels are unspecified for the features [palatal] and [velar] (and presumably [round]), and remain unspecified going into phonetic interpretation (there are no universal default rules in phonology that fill in values for every missing feature – this is a common assumption in current feature theories). The surface variation in realization of /ə/ results from interpolation between consonantal targets. In the phonetic component, vowels are assigned a target for F_1, which instantiates the phonological height specification, but they have no F_2 target. F_2 instead derives by a phonetic interpolation function between C_1 and C_2. When the flanking consonants are of the same vocalic type, the interpolation function returns a constant F_2 value for all times between C_1 and C_2, but when the consonants differ, the function returns continuously varying values of F_2. This continuously varying F_2 path is often discretized in linguistic transcriptions as a sequence of micro-vowels, in forms like [lʲᵉəᶺtˠ] = /lʲətˠ/.

The relevant acquisitional question is, what are possible forms for a child to contemplate as the output of phonology, given the classes of physical things to be modeled? There is, arguably, only one possibility for "lʲᵉəᶺtˠ": the phonology produces [lʲətˠ]. The alternative of /lʲətˠ/ → [lʲᵉəᶺtˠ] is factually arbitrary in a manner not supported by any consideration. As a symbolization of phonetic fact, it is arbitrarily imprecise, skipping many intermediate steps in the vocalic continuum which suffer no pre-theoretical disadvantage. A more accurate symbolization of pronunciation would be something like [lʲⁱᶦᵉəᶺᶤᵚtˠ]. Either alternative faces serious problems in receiving a coherent phonological representation – what system of computation maps /lʲətˠ/ to [lʲⁱᶦᵉəᶺᶤᵚtˠ]? Letter-strings like [lʲᵉəᶺtˠ] or [lʲⁱᶦᵉəᶺᶤᵚtˠ] are phonologically incoherent, without a theory of the "micro-segments" represented by raised letters or the even finer-subdivided "nano-segments" in [lʲⁱᶦᵉəᶺᶤᵚtˠ]? As emphasized above, we need a model of phonetic grammar to evaluate such claims. Given a phonetic model such as set forth by Choi, there *is* a theory of phonetic interpretation and a model of Marshallese phonetic grammar that generates correct physical outputs, operating on a phonological output [lʲətˠ].

In the case of [rʷorʷ], when combined with the independently necessary place-interpolation rule of phonetics, either [rʷərʷ] or [rʷorʷ] *qua* phonological output can be credibly related to the physical facts (the possibility that rounding of the vowel derives from interpolation between consonants does not preclude the possibility that rounding is also present in the input). When no phonetic fact dictates which of two (or more) phonetic forms are the output of the phonology, the resolution of the question must fall to inspection of the resulting systems of phonetics and phonology. Whichever system is the simplest, that is the system learned, under the premise of FP. Obviously, we need much more information about Marshallese phonology, in order to advance any argument about whether it is simpler to assume [rʷərʷ] or [rʷorʷ].[17]

There are also two prima facie plausible accounts of certain patterns of English consonant allophony: perhaps they result from late phonological rules, or perhaps they are from processes of phonetic implementation. Even assuming that underlying representations do not contain all of /p, b, pʰ/ (a claim that cannot be stipulated arbitrarily), their systematic presence in pre-physical mental representations is not in serious doubt. Is there a phonological rule assigning aspiration, or is it part of phonetic implementation? What kind of process derives [ʔ] ([hɪʔ] 'hit', [kaʔn̩] 'cotton', in some dialects) or [ɾ] (*write* ~ *writing* vs. *ride* ~ *riding*, [ɹʷaɪɾɪŋ] in both cases). If these processes are not phonological, what does that entail about the nature of phonetic computations (can phonetic processes perceptually neutralize a distinction between words)? Apparently, phonetic implementation must be able to refer to syllable or foot structure in a phonetic account of consonant allophony, which is not a strikingly controversial claim. Is there some logical principle regarding the nature of grammars that dictates that aspiration should *not* be treated phonologically, or phonetically? This is the kind of knowledge required to decide what segments are present in phonological output. Different knowledge – phonological knowledge – is required to figure out whether certain segments are missing from underlying forms.

To summarize this section, the argument that phonological features are logically unlearnable and must be built into UG depends on a number of assumptions that have not been established. It relies on the claim that phonological features are provided by an invariant direct interface with non-linguistic articulatory and perceptual systems. This precludes the existence of a linguistic component of phonetic interpretation. The consequences of that move for the theory of phonological computation and representation are quite significant, since this amounts to expanding the scope of phonology in a manner that renders phonology less coherent, not more coherent.

---

[17] An even less-phonological account is suggested by H&R (2008), to the effect that grammar does not operate on /lʲətˠ/ and /rʷərʷ/ at all, instead, interpolation is a non-linguistic universal interpretation of whatever the phonetics feeds into motor control. One reason to doubt this position is that the triggering segment is not always physically instantiated, thus the glide in /tˠɐɰɐpˠ/ 'although' is not physically realised. This is not a problem under a phonological analysis, and may also be consistent with a theory of language-specific phonetic implementation.

4. How are features learned?

It does not suffice to just assert that the features of a language are learned. The purpose of this section is to show how feature-learning procedes. In the first subsection, I sketch the basic logic of feature acquisition using a simple constructed phonology. The second subsection gives an account of feature learning in the Bantu language Kerewe.

4.1. Introduction

The greatest challenge for showing that features need not be innate and can be learned is that there is little by way of explicit logical framework for discussing the acquisition of phonological grammars. See Hale & Reiss (2008) for extensive discussion of this point.

Any language has a set of segments $A=\{s_x \ldots s_z\}$, where each segment is instantiated in the grammar with a distinct feature matrix M_i which is the structured conjunction of features $\{F_i, F_j \ldots F_n\}$. The fact that a segment $[s_x]$ exists in a language is sufficient reason for it to be assigned some set of features which renders it distinct from any other segment in the language. The specific features posited for a segment are motivated by how the segment functions in phonological computations. I assume a theory of learning where assignment of features to segments is made in such a way that the set of feature specifications in the rule system is the simplest possible, and the set of features invoked for the language is minimal. More generally, I assume the simplicity-driven model of phonological theorizing Formal Phonology (Odden 2013). The theory of UG predetermines what is a possible computational system for generating the data, and the theory of learning tells us how (and to what extent) the set of possible grammars is further narrowed down – the simplest system is the one that is learned.

Take an simple hypothetical language, with the segments /p t k m n ŋ i a u/. This language has the following phonological fact, which motivates a rule that is being learned.[18]

(9) $\{p,t,k\}_i \rightarrow \{m,n,ŋ\}_i$ / __ {m,n,ŋ}

The child has awareness of the facts represented as (9), and automatic learning creates a rule which has the effect that when the sounds *p, t, k* stand before any of *m, n, ŋ*, the former become respectively *m, n, ŋ*. Knowledge such as (9) is outside of grammar (grammars do not contain the data that they generate): it is the basis on which a grammatical rule like (10) can be posited. The learning algorithm therefore evaluates the hypothesis that the grammar has the following rule:

(10) $[F_i] \rightarrow [F_j]$ / ___ $[F_j]$

[18] The hypothetical language has a rich and transparent morphology so underlying sequences of segments can be easily motivated and alternations are quite general. In shorthand notations for rules, $\{a,b,c\}_i$ is always paired with at least one other instance of $\{e,f,g\}_i$, and means "*a* in the context *e*, *b…f*, *c…g*, respectively", but {a,b,c} means "any of *a, b*, or *c*".

If (10) is correct, it follows that *p,t,k* are $[F_i]$ and *m,n,ŋ* are $[F_j]$. There are other UG-consistent statements of the rule, for example:

(11) a. $[F_i] \rightarrow [F_j] / \_\_ \begin{bmatrix} F_j \\ F_k \end{bmatrix}$

b. $[F_i] \rightarrow [F_j] / \_\_ [F_k]$

Rule (11a) is rejected because it is more complex, compared to (10) – it posits an extra feature F_k, and it employs three features in the rule rather than two.[19] Analysis (11b) is rejected compared to (10) because it posits an extra feature F_k. From rule (10) we learn the featural analogy p:m::t:n::k:ŋ. Stops are $[F_i]$ and nasals are $[F_j]$. We also know that vowels are not $[F_i]$, since vowel segments do not change before a nasal.

Another phonological fact of the language is that

(12) i → u / {p,k,m,ŋ}__

This implies a rule of the form:

(13) $[F_a] \rightarrow [F_b] / [F_b] \_\_$

(13) allows us to identify $[F_a]$ and $[F_b]$ as features of [i] and [u], respectively, and it also identifies $[F_b]$ as a feature common to labials and velars. We now know that $[F_b,F_j]$ identifies the class [m,ŋ]. Again, the class {p,k,m,ŋ} could be analyzed as $[F_c]$, but that invokes an additional unnecessary feature.

Finally, in this language, i → a / __{k,ŋ}. Thus

(14) $[F_a] \rightarrow [F_c] / \_\_ [F_c]$

Based on segment behavior in phonological rules, we have sufficiently learned the features of this language to the point that all segments are distinctively represented.

(15)

p	t	k	m	n	ŋ	i	a	u
i	i	i						
						a		
b		b	b		b			b
			j	j	j			
		c			c		c	

The point of this brief artificial sketch is to illustrate the reasoning of RSFP feature learning, and not to solve all problems in the theory of rules or representations. Put simply, when a class of segments functions together, that is because they have a shared property

[19] More specifically, that hypothesis would not be entertained, unless there is a specific reason to consider the possibility of such a rule.

in the grammar. The task of feature acquisition is finding the simplest system of properties that accounts for those cases of grammatical functioning-together that can be observed in the primary linguistic data.

It should be obvious that RSFP depends on having a well-defined and simple theory of rule formalism. The broad syntax of rules (10) and (14) is the same: why, then, not posit one rule that does both things?

(16) $$\begin{Bmatrix} \text{p t k} \\ \text{i} \end{Bmatrix}_i \rightarrow \begin{Bmatrix} \text{m n ŋ} \\ \text{a} \end{Bmatrix}_i / \_\_ \begin{Bmatrix} \text{m n ŋ} \\ \text{k ŋ} \end{Bmatrix}_i$$

This might be a possible rule using SPE notations such as braces and angled brackets, since the SPE theory of rules has relatively little to say about what constitutes a formally-possible rule.[20] Embedded in a theory with minimal rule machinery, such an expression is not a formally-possible rule, so would not affect the computation of the interface rules mapping segments to feature matrices by presenting another analysis – (16) is never entertained as a hypothesis.

Insofar as the choice of feature assignments is based on how segments function in the system of phonological rules, we have to consider the possibility that the rules do not always uniquely identify each segment. This brings us into the contentious area of claims about what rules a grammar must contain. For example, it is not clear whether there are any segmental alternations in Vietnamese,[21] but there are around 2 dozen consonants and 9 vowels which need a phonological representation. Some system of features is needed to represent the following words.

(17) ti ‘bureau’ tɯ ‘fourth’ tu ‘to drink’

In the worst case, the distinctions can be represented with a system of 6 features assigned arbitrararily to individual segments. We must also consider whether it is a fact of Vietnamese grammar that no words begin with two consonants, or end with two consonants.[22] This might justify rules constructing syllables, which may refer to a distinction “vowel” versus “consonant”. Similar gap-filling considerations could lead to discovering a rule

[20] Much of the problem lies in the fact that SPE does define the formal properties of a simple rule – ZXAYW → ZXBYW – but virtually never employs simple rules, instead proffering rule schemata which are expressions for evaluating infinite sets of simple rules in the grammar. RSFP, in contrast, does not allow a grammar to have an infinite set of rules, and does not employ schemata such as braces, angled brackets, and English-language conditions.

[21] It should be obvious that encoding historical dialect data in synchronic phonology is not a valid basis for positing a phonological rule. It is perhaps less obvious that segment-minimalization in underlying forms is also not a valid basis for positing phonological rules. I take it to be an open question whether there are “allophonic” rules in Vietnamese such as /ɔŋ/ → [ăwŋ͡m], and on what basis a child would learn this putative phonological change, rather than storing the invariant surface form.

[22] It is clear that this is a fact about the language, the question is whether this is part of the grammar of Vietnamese. As has been repeatedly pointed out in the substance-free literature, there has been an excess of assumptions made in phonology to the effect that grammar contains reflexes of all forms of human sound-related behavior. Grammars do not encode all stateable observations about their languages.

excluding [ɓ, ɗ, γ] and other consonants from syllable codas. Ultimately, there seems to be no fact of Vietnamese phonology determining what features distinguish [ɓ, ɗ, γ]. Analgously, Hawaiian appears to have no phonological alternations and co-ocurrence restrictions seem to only suggest identifying "consonant" in order to say that there are no consonant clusters or final consonants.

It is possible that the final featural analysis is influenced by learning artifacts. For example, [p] as distinct from [ɓ] is marginal in Vietnamese, so at an early stage of acquisition, both sounds may be mapped to the same feature representation – they are not yet understood to be distinct segments. Further exposure to data may correct the analysis, whereby some instances of assumed /P/ are distinguished from others: thus [p] and [ɓ] could have the same features save for one distinction, because they were initially treated as *being* the same segment. Longitudinal evidence regarding development of segment perception in infants is sparse, so it is not possible to be more specific than to point out that if a phonological distinction is recovered – one assumed segment turns out to be two – remnants of earlier learning patterns may remain in later grammar. Ontogeny may partially recapitulate phylogeny.

One methodological point remains, regarding the simple case of feature learning considered above. In the analysis posited in (15), [t] is specified [F_i] and [n] is specified [F_j]. But some instances of [n] derive from /t/ by (10), which adds the specification [F_j], so [n] from /t/ would be represented as [F_i,F_j] but underlying /n/ would be represented as [F_j]. One solution to this problem (derivational history being maintained in the representation) would be to presume that all [F_j] segments are made to be [F_i] (via language-specific rule). An alternative solution is simply that that [F_i,F_j] and [F_j] need not be pronounced differently: both are produced as [n]. There is an analogous problem in vowels, where in the proposed system [i] is [F_a], [u] is [F_b] and [a] is [F_c]. Since [a] can derive from /i/, some instances of [a] are expected to be [F_a,F_c]; and since /i/ also becomes [u], some instances of [u] are expected to be [F_a,F_b]. In other words, all vowels would be [F_a]. Now the problem is that [F_a] no longer identifies [i], what identifies [i] is the fact of being [F_a] with no other feature specification. As noted in section 2, there has been an assumption that a rule cannot refer to the lack of a feature.

The underlying issue is one that transcends privativity vs. binarity, and features by universality vs. by learning, namely the widely-adopted but unproven assumption that languages have rules pertaining to licensed combinations of primitives – structure preservation. If a rule would create a structure that is illicit in a language, it has been conjectured that the illicit structure is brought into conformity with the rules defining the object in question (segment or syllable). Whatever the theory of primitives and feature values is, it is at least credible to contend that some mechanism in the grammar of Arabic (most dialects) indicates that [p] and [v] are not segments of the language, even though free combination of independently necessary primitives might allow their existence. Or, there simply happen to be no lexical items or derivational results containing such a segment, but the segments are not grammatically illicit.

The concept of structure preservation has historically been fraught with problems, such as the premise that it is defined in part by the questionable notion "contrast" (so-called allophonic rules are not limited by a structure-preservation requirement). The

RSFP view of segments and features denies the significance of the taxonomic phoneme, holding that if [pʰ tʰ kʰ ɾ] are segments in the phonology of English, then [pʰ tʰ kʰ ɾ] are segments that have to be assigned features.[23] RSFP does not make an a priori holding as to what the segments of a language are: this is an empirical question (where competing answeres are a significant contributor to language change).

One analytic trend in coping with structure preservation has been to make gaps fall out from the featural analysis of segments – a perfectly valid approach to the problem. If a language has a voicing contrast in obstruents but not in sonorants, and since feature dominance relations are not predetermined by UG, it may be that the feature for voicing is a structural dependent of the feature that distinguishes obstruents from sonorants. If a language has [i e a o u] and not [ø, ɯ], the language may simply not employ the feature [round], and lip protrusion is an articulatory fact about back vowels, having no phonological significance. I take the matter of structure preservation to be a real issue that ultimately needs to be addressed, but also, persistent feature-coocurrence relations are not a theoretical artifact of RSFP's making. Whether there is any real problem pertaining to asymmetries and gaps in feature specification remains to be seen.

4.2. Acquisition of Kerewe features

This subsection demonstrates the logic of feature acquisition to segments in Kerewe, a Bantu language spoken in Tanzania. The goal is not only to show that it is possible to arrive at a feature specification of the segments of the language based only on phonological behavior, but also to exemplify the dependence of this analysis on a logic of acquisition and a theory of rule formulation. Since the point of the discussion is to demonstrate how features are learned, I forego extensive empirical discussion and draw on a deeper analysis of the data (Author, in progress), only providing basic illustrative examples.

As a starting point, the surface segments of Kerewe are as follows.

(18) p t tʃ k f s β v z b d dʒ g l h m n ɲ ŋ j w
i u e o a

Besides these obvious segmental distinctions, Kerewe is a tone language, but tonal properties are not analyzed here. There is also a robust lexical contrast in the length of vowels (*ekisibo* 'tether', *ekisiibo* 'fasting', *emboga* 'vegetable', *embooga* 'infected eyes', *kuhata* 'to dislike', *kuhaata* 'to peel' etc.), and a limited but productive nonlexical surface contrast between single and geminate nasals in verbs in utterance-initial position, deriving from /n+C/ sequences, e.g. [naahúla ~ nnahúla] 'choose me!', cf. [kujáhúla] 'to choose'. There is good evidence for the classical autosegmental treatment of length as involving a many-to-one mapping between segments and higher prosodic units – moras or skeletal positions. It has also been classically assumed that syllabicity and length are represented with suprasegmental non-featural prosodic objects. RSFP does not have a principled commitment to including or excluding higher prosody from the set of learned representa-

[23] To state the point somewhat differently, "contrast" in RSFP means "is a segment in the language".

tional objects, though obviously it would be advantageous if prosody could also be entirely learned. Because so many extraneous issues would arise, I ignore prosody and presuppose a standard moraic phonology account of syllabicity and length. The main consequence for the analysis of segmental features is that the distinction between [j,w] and [i,u] might be just prosodic, or it might also be featural (but there is no evidence in Kerewe for a featural difference). When [j,w] act differently from [i,u], that could be because their prosody is distinct. In the course of ignoring prosody, it may be useful to know the pattern of segment sequencing which is usually handled by a set of rules of syllable structure. As in most Bantu languages, words are of the form $(C^3V)^*$, between 1 and 3 consonants[24] in the onset followed by a vowel, which defines a syllable. The optional left margin of the syllable is always a nasal (homorganic with the following consonant) and the optional right margin is a glide [j,w].

Some consonants have very limited distribution in Kerewe. Two of them are so limited that it is unclear whether they are actually segments of the language. Those segments are [v] and [dʒ], which each exist in under a half-dozen recent loan words. The fact of being very low frequency does not per se justify ignoring them. What is not clear is whether they actually exist as the result of normal language acquisition. They may be analogous to [ø,y,x,ǁ] in English, which are sounds that educated adult speakers can make in pronouncing *milieu, Übermensch, Bach, Xhosa*, but which are not acquired or represented in the same way that *p, t, θ, ɹ* are. It is simply unknown what the acquisitional facts are surrounding foreign phonemes in Kerewe. It is possible that [v] is more generally nativized as [β] but is pronounced by educated speaker in certain words such as "driver" as [v], using extragrammatical information. It is known that [dʒ] is an originally Jita phoneme, and most or all Kerewe speakers are bilingual in Jita. There is a widespread strategy of nativising Jita [dʒ] as [zj] which may be resisted in certain words (especially those coming from Swahili). Because [v,dʒ] do not appear in a context where they clearly undergo or are excluded from a phonological rule, there is little evidence for what their feature assignment is.

[ŋ] appears robustly before [k,g,ŋ], never before any other consonant, and exists in very few words in root-initial position. Although the distribution of [ŋ], like [v], is highly limited, the evidence suffices to provide a feature assignment to [ŋ]. In like fashion, [b] (distinct from [β]) is distributionally limited. After a nasal, [β] always becomes [b], and this rule is the main source of instances of [b]. But there are a few instances of [b] which are not so derived.[25]

4.2.1. Consonant features

The first rule to be considered is nasal place assimilation. Certain prefixes (1s subject or object prefix; cl. 9-10 noun prefix) have /n/ which assimilates in place to a following con-

[24] In utterance-initial position, a short V syllable, one without an onset, is also possible.
[25] It is also unclear how stable underived [b] is. Examples seem to come from loanwords, and sometimes [b] is pronounced [β], for example in *kubómóla ~ kuβómóla* 'to tear down', influenced by Swahili /bomoa/.

sonant. This is illustrated in (19) with the 1s SP /n/. A process of hardening is also attested in some examples.

(19)	m-bazílé	'I counted'	a-βazílé	'he counted'
	m-boomílé	'I was dull-witted'	a-boomílé	'he was dull-witted'
	m-mazílé	'I finished'	a-mazílé	'he finished'
	m-pweezílé	'I helped'	a-hweezílé	'he helped'
	m-pekesílé	'I made fire'	a-pekesílé	'he made fire'
	m-fuzílé	'I cleaned	a-fuzílé	'he cleaned
	n-naaβílé	'I bathed'	a-naaβílé	'he bathed'
	n-dabílé	'I passed'	a-labílé	'he passed'
	n-tegílé	'I fished'	a-tegílé	'he fished'
	n-ziikílé	'I planted'	a-ziikílé	'he planted'
	n-sonílé	'I sewed'	a-sonílé	'he sewed'
	ɲ-ɲamwíílé	'I ruined'	a-ɲamwíílé	'he ruined'
	ɲ-tʃumisílé	'I stabbed'	a-tʃumisílé	'he stabbed'
	ŋ-ganílé	'I told a story'	a-ganílé	'he told a story'
	ŋ-kwaasílé	'I touched'	a-kwaasílé	'he touched'
	ŋ-ŋoloosílé	'I groaned'	a-ŋoloosílé	'he groaned'

There are no clear examples of initial /d/ in verb roots – [d] is historically the post-nasal allophone of /l/ – but there is a synchronic initial contrast in noun roots as shown by forms with a vowel-final noun prefix such as [aka-lezu] 'little beard', [aka-dege] 'little airplane'. Compare [en-dege] 'airplane', [en-dezu] 'beard'.

The underlying form /n/ of the 1s SP is motivated in (20) before the past prefix *-a-*.

(20)	*Simple past*		*Perfective (hesternal)*	
	n-a-gáβá	'I divided'	ŋ-gaβílé	'I divided
	w-a-gáβá	'you divided'	u-gaβílé	'you divided'
	tw-aa-gáβá	'we divided'	tu-gaβílé	'we divided'

In general, any nasal can appear before any vowel, but a nasal before a consonant is always homorganic with the consonant.[26] The fact pattern that the child will be aware of is as follows:

[26] There is a non-phonological process of phonetic implementation whereby the vowel /u/ may have reduced duration when preceded by /m/ and followed by a consonant, so that the vowel in *omuβígi* 'trap fisherman' is noticeably shorter compared to *oβuβígi* 'act of trap-fishing'. In my data, the vowel is present in words like *omuβígi* most of the time, but it is plausible to expect that under Swahili influence (where there is categorial vowel deletion) the process is becoming phonologized among younger speakers.

(21)	n → m	/	___	p,b,f,m (v) (w) <β h>
	ɲ		___	tʃ,ɲ (dʒ)
	ŋ		___	k,g,ŋ
	n		___	t,d,s,z,n <l> <j>

The significance of parenthesized (v,dʒ) is that there is an attestation gaps for these segments. The consonant /v/ never appears after a nasal. Root-initial /dʒ/ is also unattested, so we cannot unambiguously observe alternating prefixes with final /n/ appearing before /dʒ/. However, /dʒ/ does appear root-internally after a nasal in a very few words – *kʊβúúɲdʒa* 'to peddle' and the noun *embúúɲdʒá* 'jigger', as well as *emooɲdʒo* 'toy top'. It is far from certain that the nasal in these words is the result of applying an assimilation rule: the underlying roots may well simply be /βúúɲdʒ, mooɲdʒo/. Still, this constitutes weak evidence pointing to a possible place of articulation specification for /dʒ/. To avoid building a theory of features on a weakly supported conclusion, we temporarily set aside the question of place feature for /dʒ/, until we come to the point of obligatorily distinguishing /dʒ/ from other sounds.

/w/ also never appears in the relevant context (root-initial position), underlyingly or in derived forms, where it could be preceded by an alternating nasal. Angled brackets in (21) indicate segments which have been independently removed from rule inputs by another rule. The glide /j/ does underlyingly appear in this context but because of a prior deletion rule, it is not present when assimilation takes place. Likewise /β,l,h/ are present in underlying forms, but have become [b,d,p], respectively.

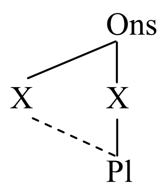
There is some freedom in the statement of the rule in question. All and only consonants are triggering segments – no consonants must be explicitly excluded from the rule, but all vowels are excluded. However, all consonants are within the onset of the syllable, so provided that the domain of the rule is stated as being the onset, all segments in that domain trigger the rule. Because the language has no sequences of obstruents at any stage of the derivation, no consonants have to be excluded from the input as non-alternating. As far as trigger segments are concerned, because of the restricted combinatorics of complex onsets in Kerewe, the trigger consonant is always the unrestricted obligatory consonant, C_2 in the (C_1)C_2(C_3) onset template.

(22)	(N_1)	C_2	(G_3)		
	m	b	w	á	'dog!'
		n	w	á	'drink!'
	ŋ	k		é	'few$_9$'

Given the option of referring the identification of target and trigger segments to positions within the syllable, picking out target and trigger may not directly provide direct evidence for features (nasal and something similar to consonantal). But the rule whereby C_1 position is only filled with a nasal provides evidence for a feature Nasal, and the rule whereby C_3 position is only filled with a glide provides evidence for something that identified glides (Vplace, see section 4.2.2).

The change is more informative regarding features. The structural change is conditional, in that the output segment depends on which subset of segments is on the right. In an autosegmental interpretation of representations, Place spreads to a preceding nasal. The underlying formal premise is that each phonological rule operates on one phonological unit: one thing of the input segment is changed.[27] Since there are four distinct outputs depending on the subset of segments which follows, there must be four distinct feature configurations under Place (Pl). We observe that the set *p,f,b,m* functions as one class (labelled Lab) conditioning [m], *tʃ, ɲ* and possibly *dʒ* as another (Pal) conditioning [ɲ], and *k, g, ŋ* as a third (Vel) conditioning [ŋ]. Possibly, *t,s,z,d,n* are unified with a feature (Cor); or perhaps they are not specified with Pl, and instead underlying /n/ is unchanged in that context, just as it is unchanged before a vowel. Since there is other evidence for a positive specification of Cor, we will at least initially adopt a 4-feature account of consonantal place. The place assimilation rule is a standard autosegmental spreading rule.[28] X refers to the segmental root node. The reason why there is no assimilation in /nw/ (*kúnwá* 'to drink') is that glides have no Place node, they have a Vplace node.

(23)



On the basis of these alternations, we also gain evidence that some feature unifies {m,n,ɲ,ŋ}, namely Nasal, which is a property of the input *n*. That feature is inherited from /n/ in the case of derived *m,ɲ,ŋ*: the reason why /n/ becomes [m] before labials, and not some other consonant, is that the only difference between *n* and *m* is that the former is Labial, and [n,m] have in common the property Nasal. At this point, we have learned the existence of the feature Place (the property which spreads), four nodes dominated by Place namely Labial, Palatal, Velar and Coronal, as well as Nasal (inherited from /n/ and unifying {m,n,ɲ,ŋ}).

The second rule to consider is post-nasal Fortition, whose effect, descriptively, is:

(24) $\{\beta,\text{l},\text{h}\}_i \rightarrow \{\text{b},\text{d},\text{p}\}_i$ / {m,n}<ŋ,ɲ> ___

Examples of this process are seen in (19), and we can add examples of the present tense where the 1s SP stands right before the verb root.

[27] There is a competing scenario, proposed in Reiss (2003), whereby a rule may contain expressions quantifying over subsets of the features, e.g. "all of the features of the set [coronal, anterior, back]". I will not pursue that approach here, but assuming such a rule statement, one would simply gain information about the individual features and their set-theoretic unification *qua* "place".

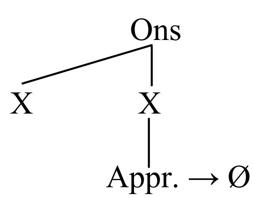
[28] It is beyond the scope of this work to satisfactorily address the highly relevant question of how to interpret autosegmental notation in RSFP-FP. (23) should not be taken to assert that X immediately dominated Pl. At most, the rule asserts that a segment which is C receives Pl from a following segment that is C. Such issues are explored in Author (in progress b).

(25)	n-dímá	'I cultivate'	a-límá	'he cultivates'
	m-béézá	'I carve'	a-βéézá	'he carves'
	m-panaantúká	'I descend'	a-hanaantúká	'he descends'

Not all consonants can appear post-nasally in underlying forms. There is no clear example of rare underlying /b/ appearing here, but there are no problematic cases either.[29] The consonants [f s z p t tʃ k b d dʒ g] do appear post-nasally and are not changed, so these segments must either be explicitly excluded, or the effect as realized on these consonants must be vacuous. No vowels trigger Fortition, and no nasals have to be excluded as non-triggers. Again, the only morphemes which demonstrably trigger this process have the underlying form /n/. There are no input cases of *ɲ, ŋ, m* which precede /β l h/, so those segments could be excluded or included as necessary in order to achieve a more economical grammar (no savings results from excluding any nasals). As in the case of place assimilation, these facts can be expressed by limiting the rule to affecting C in the onset.

From the fact-pattern in (24) we conclude that /β l h/ have a feature in common that distinguishing them from /f s z t tʃ k dʒ g/. The pairs β vs. b, l vs. d, h vs. p are themselves distinguished by a feature: the members of the pairs {β,b}, {l,d}, {h,p} are the same except for some feature, which is the feature that is changed by this rule. The simplest solution is that the target-identifying feature and the changing feature are the same. We call this feature Approximant (Appr.), and the rule turns Approximants into non-Approximants. Thus β/b are the same (the features of [b] are inherited from /β/), save for the feature Approximant, likewise l/d and h/p. The change performed by the rule can be formalized as deletion of Approximant when preceded by an onset segment.

(26)



By specifying only /β l h/ as Approximant, we guarantee that only /β l h/ are affected. Based on these two rules, we have a partial assignment of features to segments.

(27)	p	t	tʃ	k	f	s	β	z	b	d	g	l	h	m	n	ɲ	ŋ
	Pl	?	Pl	Pl	Pl	?	Pl	?	Pl	?	Pl	?	Pl	Pl	?	Pl	Pl
	L	Cor	Pa	Ve	L	Cor	L	Cor	L	Cor	Ve	Cor	L	L	Cor	Pa	Ve
							Ap					Ap	Ap				
														N	N	N	N

[29] It is empirically unanswerable whether the underlying form of roots such *gaamba* 'say' is /gaamβ/ or /gaamb/. In the latter case, no rule is required to derive the surface form, and in the former case, the independently motivated rule Hardening will derive [gaamb]. No motivated mechanism of the grammar turns /Nb/ into anything other than [mb].

A third rule, Spirantization, changes /t tʃ d l/ to [s s z z] respectively before three of six morphemes which underlyingly begin with [i].

(28) $\{t,tʃd,l\}_i \rightarrow \{s,s,z,z\}_i$ / ___ [i]

Hesternal perfective examples with the triggering suffix -*ile* illustrate this change.

(29)				
	a-has-ílé	'he disliked'	ku-hat-a	'to dislike'
	a-fuz-ílé	'he washed'	ku-ful-a	'to wash'
	a-laanz-ílé	'they interlaced'	ku-laand-a	'to interlace'
	a-pekes-ílé[30]	'he started a fire'	ku-pékétʃa	'to start a fire'

The following are examples of other consonants which do not change.

(30)	*1s hest. perf*		*Infin*
	n-dom-ílé	'I made a speech'	ku-lom-a
	n-doβ-ílé	'I fished with hooks'	ku-lóβ-a
	ŋ-kop-ílé	'I borrowed'	ku-kóp-a
	ŋ-gon-ílé	'I snored'	ku-gon-a
	m-bis-ílé	'I concealed a fact'	ku-βís-a
	m-boj-ílé	'I fought'	ku-βój-a
	ŋ-koɲ-ílé	'I cooked improperly'	ku-kóɲ-a
	n-dog-ílé	'I bewitched'	ku-log-a
	m-bik-ílé	'I announced a death'	ku-βík-a
	m-boh-ílé	'I tied'	ku-βóh-a

This rule applies before the suffixes for nominalization /-i/, causative[31] /-i-/ and perfective /-ile/, but not stative /-ik-/, applied /-il-/ or causative /-isj-/. The reason for the behavioral divergence in instances of [i] is that the trigger morphemes in Proto-Bantu had the vowel [i] but the non-trigger vowels had [ɪ]. Finding the proper mechanism for identifying these specific morphemes as triggers is not essential to our present goal (a solution is discussed at the end of 4.3).

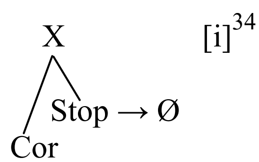
Rule (28) provides evidence that /t,tʃ,d,l/ have something in common. Any consonant (aside from marginal /v,b,dʒ/) does appear in the triggering context, and notably /p,β,k,g,m,n,ɲ/ do not change. There is no need to explicitly exclude /s,z/ from the input class, since vacuous application of the rule to these segments also yields the correct out-

[30] Most words with apparent root-final /tʃ/ actually derive surface [tʃ] from /kj/ and have a substantially different form in the perfective, cf [ku-βwáátʃ-a] 'to greet', [m-bwaak-íízjé] 'I greeted' root has underlying /tʃ/.

[31] This suffix appears as [j] since it always undergoes glide formation, but there is sufficient evidence that it is underlyingly /i/.

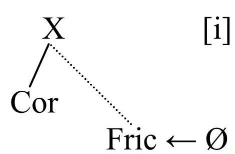
put. The segments targetted by the rule are definable by the feature Coronal (Cor),[32] and the simplest analysis is that Cor is unaffected by the rule (hence *s,z* are also Cor, the output segment having inherited that specification from its input). We also observe that /t/ and /d/ map to distinct outputs – some distinguishing feature (a form of voicing) is preserved from the input. One option is that /t/ has a voicing specification and /d,l/ do not: or the converse.[33] The nature of the shared change also establishes that the relationship of *t* to *s* is the same as the relationship of *d,l* to *z*, that is, the rule adds (or subtracts) something in /t,tʃ,d,l/ to give [s,z]. We will eventually see evidence for a feature Voiceless, necessitated by the fact that /k/ is targeted by palatalization while /g/ is not.

As for the change performed by the rule, it could be that the input segments gain the specification Fricative, or else an existing feature Stop is deleted. The overall strategy is to equate [t,tʃ,d,l] versus [s,z,n] via feature specification, so that none of [t,tʃ,d,l] would change to [n]. Assume first that [t,tʃ,d,l] have a feature that is deleted, Stop:

(31)


If /n/ does not have the specification Stop, it will not undergo (31). If /n/ is specified Stop, and assuming another feature which positively identifies nasals (e.g. Nasal), either (31) requires an additional condition whereby the target cannot have Nasal (a complication which is theoretically problematic – requiring a rule to refer to the lack of a property), or an additional rule is required to reinsert Stop on Nasal segments. Another option is that the grammar generates Stop and Non-stop nasals, which are phonologically and phonetically indistinguishable.

Alternatively, a feature Fricative could be assigned by rule to otherwise unspecified /t,d,l/ in this context.

(32)


Again, the matter of how /n/ is treated is a prominent problem: what unifies /t,d,l/ excluding /n/? Must the rule be formalized to exclude /n/, or can the rule apply vacuously? Assuming (32), if /n/ is underlyingly specified as Fricative, applying (32) to /n/ does not change /n/.

[32] The previous identification of Palatal as independent of Coronal now gives way to an analysis of Palatals as having a combination of Coronal and Velar. This decomposition of Pal is partially motivated by *kj*-fusion to be discussed, and the present fact that *tʃ* acts as part of a class that includes [t,d,l].

[33] It is also possible for all segments to be characterized by two such features, but such a hypothesis would not be considered without compelling evidence for its necessity.

[34] Here, "[i]" refers to whatever properties identify the triggering vowels.

We are not done with structure-preservation type problems with Spirantization, since both /d/ and /l/ become [z]. From Fortition (26) we know that *l* is Approximant and *d* is not. /z/ is also not affected by Fortition, therefore it lacks Approximant which identifies targets. In the case of /d/ → [z], we could equally treat the change as Fricative-insertion or Stop-deletion. In the case of /l/ → [z], we also need a mechanism for removing Approximant (since [z] is not specified Approximant). When the spirantization change takes place, Approximant is also removed. This indicates that there is a more general feature-deletion process – removal of Stop, and Approximant if it is present.

This dependency relationship has a straighforward solution, using representational machinery well-motivated in phonology which has been universally assumed in autosegmental theories of representation. If Approximant is a dependent of (dominated by) Stop, Approximant will always delete when Stop deletes. The proposed representation of relevant segments is given in (33).

```
(33)   X            X          X
       |            |
      Stop         Stop
       |
      Approx

      [l]          [d,t,tʃ]    [s,z,n]
```

Under this analysis, it follows that /h,β/, the other two Approximant segments, are also Stop – not an expected outcome based on substantive associations suggested feature names, but RSFP features have no substantive associations, and names are arbitrary.

A detail regarding spirantization needs attention: both /t,tʃ/ become [s]. The difference between [t] and [tʃ] is place of articulation – some feature distinguished these stops (initially identified as Palatal but now analyzed as the combination of Coronal and Velar). Since [t,tʃ] have different place features, the fricative versions ought to inherit that difference. Yet both stops become the phonetically identical fricative, [s]. There are three obvious responses to this problem, a problem which is not overwhelming, but it is also not obvious which solution is correct. First, it could be that the unmodified outputs – [s] which is the fricative coming from /t/, and what we could symbolize as [ʃ] for the fricative coming from /tʃ/ – are simply not pronounced differently, even though they are featurally different. We cannot evaluate a phonetic-interpretation account without having a substantially-supported theory of linguistic phonetic computation that addresses this option, which we presently lack. Second, we could appeal to the phonological notion of structure-preservation, meaning that there are well-formedness rules governing allowed feature configurations on segments (probably the same mechanism generates syllable well-formedness conditions). In that approach, the additional feature for palatals cannot appear on an oral non-fricative – the language does not have [ʃ,ʒ], and any rule which creates such a configuration automatically repairs the ill-formed structure. Saying how the repair is automatically brought about is complex (why not restore Stop? or make the segment Nasal?). This brings us to the third option, that there simply is a rule in the language which deletes Velar from an oral non-stop (there is also no [x,γ] in the language). The problem with this approach is that "non-stop" is an appeal to lack of specification. We will not attempt to resolve this problem here.

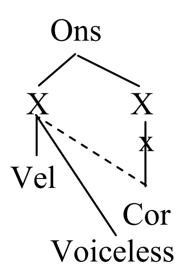
One last alternation provides phonological evidence for the analysis of Kerewe consonants, namely a Palatalization rule where the sequence *kj* becomes [tʃ]. While the marginal phoneme /ŋ/ does not appear in a context where we can test whether it undergoes an analogous change (e.g. root-finally in a verb), /g/ does, and /gj/ is not changed. This process can be motivated with examples of the short causative *-j-*,[35] which stands between the root-final C and the final suffix, *-a* in the following examples.

(34)

Infinitive	*Causative infinitive*	
kuβuuka	kuβuutʃa	'comb'
kuhika	huhitʃa	'arrive'
kuseleka	kuseletʃa	'hide'
kuhiiŋga	kuhiiŋgja	'exchange'
kuloga	kulogja	'bewitch'
kulima	kulimja	'cultivate'
kuβííha	kuβííhja	'be bad'
kulóβa	kulóβja	'fish with hooks'
kukópa	kukópja	'borrow'
kuβóna	kuβónja	'find'
kuβísa	kuβísja	'conceal'

Thus there is a feature, Voiceless, which [k] has and [g] lacks, and the Palatalization rule refers to this feature in selecting the target. The rule also identifies [k] owing to it being specified Velar, which excludes other places of articulation. By feature inheritance it follows that [tʃ] is also Voiceless. The simplest account of the rule *kj→tʃ* is that the features of the input segments are merged into one segment: meaning that [tʃ] is Velar plus whatever place feature characterizes [j]. That feature is Coronal, a fact which we knew from the outcome of Spirantization applied to *tʃ*.[36] Because [ki] does not change (*eki-tóóke* 'banana', cf. *ech-áála* 'finger' ← /eki-ála/) it is essential that the input sequence be limited in domain to the Onset.

(35)



[35] This suffix is also one of the triggers of spirantization, so examples of final *l,d,t* are changed for an independent reason.

[36] A side effect of this merger is that [tʃj] should not exist in the language, which is true.

It is now appropriate to take stock of our feature analysis. We have identified Place as the item operated on by Place-assimilation, and the features L(abial), Ve(lar) and Coronal as the dependents of Place which are carried along in this assimilation. Previously-assumed Palatal has been supplanted with the combination of Coronal and Velar within a segment, on the evidence of the Palatalization rule. Cor(onal) is also motivated in the target-selection aspect of the Spirantization rule, and Velar is likewise motivated via the Palatalization rule. The consonants [v w j dʒ] have been removed from the inventory below, because so far nothing in the phonology tells us anything about the make-up of these consonants. Question mark indicates uncertainties, situations where we cannot so far tell whether the segment has the feature in question.

(36)	p	t	tʃ	k	f	s	β	z	b	d	g	l	h	m	n	ɲ	ŋ
	Pl	Pl	Pl	Pl	Pl	Pl	Pl	Pl	Pl	Pl	Pl	Pl	Pl	Pl	Pl	Pl	Pl
	L	Cor	Pa	Ve	L	Cor	L	Cor	L	Cor	Ve	Cor	L	L	Cor	Pa	Ve
	?	St	?	?	?		St		?	St	?	St	St	?		?	?
							Ap					Ap	Ap				
	?	Vl	Vl	Vl	?	Vl	?		?		?		?	?	?	?	?
														N	N	N	N

Not all consonants are fully differentiated at this point, but few consonants remain undifferentiated. In particular, [p,f,b] are not yet distinguished from each other (they are all "labials"), nor are [h,β] (both are "labial approximants"), and [v,dʒ] are not yet distinguished from any other consonant. As the question marks indicate, it is possible that [p] is Stop, or Voiceless – we have not seen positive evidence in the form of phonological rule behavior that would justify an assignment. We can call on the fact that /h/ becomes [p] after a nasal by deletion of Approximant, to fill in missing features. Noting that /h/ bears Stop,[37] the simplest analysis is that the output [p] does as well. The assignment of Stop to [b] also follows from the analysis that fortition deletes Approximant *qua* dependent of Stop. In other words, [p] and [b] are both Stop, and [f] is not. We cannot derive direct evidence from rule behavior as to whether /k g/ are Stop or not.

We have not provided any direct evidence for Labial – no rule specifically refers to labials. Now observe the place representation of the attested consonantal places.

(37)	k	t	tʃ	p
	Pl	Pl	Pl	Pl
	Ve	Co	Ve,Co	La

The 4-way distinction in types of place can, in fact, be derived without the invocation of a separate feature Labial. Instead, labials may be segments with a Place node, but no other feature under Place, i.e.

[37] This follows from the fact that we have direct evidence that [h] is Approximant, a feature which is dominated by Stop.

(38)	k	t	tʃ	p
	Pl	Pl	Pl	Pl
	Ve	Co	Ve,Co	

There being no specific reason to posit Labial in addition to Place, Labial can be eliminated from the feature inventory.

4.2.2. Vowel features

Now we turn to the analysis of vowels. As a class, vowels have a non-featural property in common, that they are dominated by one or two moras, and the difference between [jw] versus [iu] is at least a difference in moraic status. Since we are only investigating the acquisition of putatively phonetically-based features and not all representational entities, we will freely exploit the existence of contrastive prosodic properties as a means of avoiding postulation of features, making it harder to justify the invocation of a feature when a prosodic distinction is available.

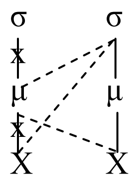
For example, vowel segments are always syllable-final (there are no codas or diphthongs) and never syllable initial (except in short utterance initial syllables). The grammar does not express these facts in terms of the concept "vowel", it does so via rules regulating moras. One relevant rule is Glide Formation, which turns vowels other than /a/ into corresponding glides.[38] These examples also illustrate an optional rule of glide-deletion which deletes root-initial glides which are not word-initial (only /j/ appears in the relevant context).

(39)	tu-janíká	twaaníká	'we spread'
	o-játʃá	wáátʃá	'you sg. light'
	mu-jégá	mwéégá	'you pl. rejoice'
	gu-jolóβá	gwoolóβá	'it (cl. 3) is soft'
	βi-jologozíβwá	βjoologozíβwá	'they (cl. 8) are cleaned'
	li-jóká	ljóóká	'it (cl. 5) burns'
	e-jóká	jóóká	'it (cl. 9) burns'

The Glide Formation rule is potentially statable without reference to any features at all, given that non-application of the rule to /a/ is due to an independently motivated rule which eliminates /a+V/ sequences. The question of how to properly express rules of de-syllabification with compensatory lengthening is a matter of longstanding controversy which we will not enter into: (40) is stated to explicitly perform all of the relevant prosody-to-segment relations, and it is a separate question how such rules are properly formalized.

[38] The mid vowels [e,o] do not generally appear in prefixes, except in absolute word-initial position. It is possible, but not guaranteed, that the cl. 9 and 2sg prefixes are underlyingly /i,u/ and are subject to word-initial lowering. I assume a more concrete underlying representation, /e,o/ which are the actually-observed vowels, but do not crucially rely on that.

(40)



What we learn from this alternation is that [j] and [i] are prosodic variants of each other, likewise [w] and [u]: that is, they are featurally the same (unless we find that there is an additional feature which correlates with the prosodic difference). Given that /e,o/ also become [j,w], we would also conclude that there is a featural analogy between /e,j/ and /o,u/ – independently we know that fact from a rule of Vowel Fusion, to which we now turn.

Word-internally, /a/ fuses with a following vowel, so that /ai, ae/→[e:], /au, ao/→ [o:], /aa/→[a:]. We can draw on the optionality of root-initial *j*-deletion to generate suitable vowel sequences which undergo Vowel Fusion, looking at examples of the present tense.

(41)	βa-jéémbá	βéémbá	'they sing'
	βa-jíβá	βééβá	'they steal'
	βa-joléká	βooléká	'they point'
	βa-jahúlá	βaahúlá	'they choose'

Sequences with /a+u/ are hard to come by, but the example [elj-úúla] 'shaving', [amóóla] 'shavings' from /ama-úla/ shows that the rule applies at /au/ as well.

There are at least two plausible approaches to the relationship between Glide Formation and Vowel Fusion, in terms of specifying which vowels undergo which process. One is that Glide Formation is explicitly restricted so that it does not apply to /a/, and Vowel Fusion subsequently applies to any vowel sequences that remain (/a+V/). The other is that Vowel Fusion explicitly applies only to /a+V/, and Glide Formation subsequently applies to any remaining vowel sequences. In the former analysis, the formalization of Glide Formation requires that target vowels be specified with a feature A which is lacking in /a/, and in the latter analysis Vowel Fusion requires the first vowel in the sequence to have a feature B which is found in /a/ but not /iu/. Either of (42a) or (42b) would seem to be possible specifications at this point.

(42)	a.	a	e	o	i	u
			A	A	A	A
			X	Y	X	Y
	b.	a	e	o	i	u
		B				
			X	Y	X	Y

Equally relevant is the fact that /a, e, o/ have a feature in common, which sets these vowels apart from /i, u/. The merger of /ai, au/ into [ee, oo] indicates that /a/ intrinsically bears the feature that distinguishes mid vowels from high vowels. We can identify that feature as Mid, whose existence is necessary regardless of how the targets of Vowel Fusion and Glide Formation are identified. Thus we select between the following feature asignments, where X and Y are whatever distinguishes front from back / round vowels.

(43) a.

a	e	o	i	u
	A	A	A	A
	X	Y	X	Y
Mid	Mid	Mid		

b.

a	e	o	i	u
B				
	X	Y	X	Y
Mid	Mid	Mid		

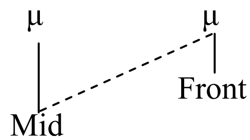
The choice between these analyses would be arbitrary, unless some independent evidence exists for the A grouping or the B grouping.

We turn now to evidence which supports the A grouping, coming from vowel harmony. There is theoretical evidence for two such rules, one of which turns /i/ into [e] after [e,o] (skipping over any consonants), with a second rule (likewise skipping over any consonants) turning /u/ into [o] after [o] only. The theoretical premise behind the conclusion that there are two harmony rules is that rule formalism does not contain expressions encoding dependencies like “applies to X only if the trigger is also Y”, as could be expressed using SPE angled brackets notation. In the case of a suffixal front vowel, a mid vowel causes /i/ to become [e]. The last three examples show that no vowel can stand between the target and trigger vowels.

(44)

plain Verb	*V+applied*	
kuβúga	kuβúgíla	‘to paddle’
kugaβa	kugaβila	‘to divide’
kuβója	kuβójéla	‘to fight’
kuhooŋga	kuhooŋgela	‘to bribe’
kumweemweesa	kumweemweesela	‘to smile’
kugeenda	kugeendela	‘to go for’
kugelula	kugelulila	‘to deduct’
kubónékana	kubónékanila	‘to appear’
kulémála	kulémálila	‘to be lame’

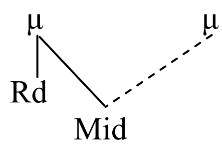
The triggering segments can be identified as Mid, and the structural change is that the target becomes Mid. Because, as we will see, the vowel /u/ does not become [o] after [e] (it only lowers after [o] by a more specific rule), we must also restrict the target to a vowel that is front. We can tenatively state the rule as follows.[39]

(45)


The reversive /ul/, often doubled, serves to motivate a second rule lowering /u/ only after [o].

(46)				
	kusiβ-ul-a	'to untie an animal'	kusiβ-ik-a	'to be tethered'
	kuhan-ul-a	'to take in laundry'	kuhan-ik-a	'to be hung up'
	kuseemb-ulul-a	'to unwrap'	kuseemb-a	'to wrap up'
	kuβóh-ólol-a	'to untie'	kuβóh-a	'to tie

To identify the more restricted trigger [o], this rule requires specification of a feature present in [o] and lacking in [e] – [o] is Rd.

(47)


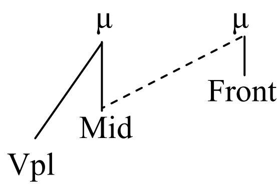
The obvious question that arises from these rules is, why does /a/ not trigger application of (45), bearing in mind that /i u/ also do not cause lowering of /i/? As contemplated in (43) it is possible that /e o i u/ have a shared feature, or else /a/ has a feature that is unique to it. The fact that /a/ does not condition either vowel harmony rule motivates the decision that /e o i u/ have something in common, a feature which is lacking from /a/, even though it has the feature which spreads. Now assigning mnemonic labels to the features that we have identified, the feature which unifies vowels other than /a/ is Vplace (Vpl), which correlates with presence of either Front or Round.

(48)	a	e	o	i	u
		Vpl	Vpl	Vpl	Vpl
		Fr	Fr	Rd	Rd
	Mid	Mid	Mid		

(45) is therefore revised as follows, to include the restriction that only Mid vowels bearing Vplace trigger lowering of /i/.

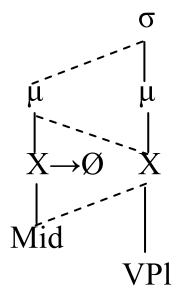
[39] The rule spreads the feature Mid from vowel to vowel, where "vowel" is captured non-featurally via reference to moraicity. This formalization should not be interpreted as saying that Mid is immediately dominated by μ.

(49)



With vowel feature specifications now resolved, we can now adopt the rule contemplated in (40), with the provision that the target bears VPl. Vowel Fusion simply merges the content of the sole remaining prevocalic vowel /a/, which has just a Mid specification, with the feauture specifications of the following vowel. The formulation below explicitly states all of the rearrangements involved, and it is left to separate theorizing within the theory of rule formulation to determine how this rule should be expressed.

(50)



At this point, the vowels have been fully distinguished from each other and from the glides.

(51)

i	u	e	o	a	j	w
μ	μ	μ	μ	μ		
VP	VP	VP	VP		VP	VP
Fr	Rd	Fr	Rd		Fr	Rd
		M	M	M		

These feature assignments must be integrated with the feature assignments for consonants, and simplifications are possible. The first, for which there is direct evidence, is that the feature Front invoked for vowels is in fact identical to Coronal invoked for consonants. It was previously found that Coronal, which unifies [t d s z l n tʃ], can come from the merger of [k] and [j], justifying the conclusion that [j] is a non-moraic Coronal. We also know from Glide Formation that front vowels become the glide *j*, therefore front vowels have the feature Coronal. The alternation *ki-jóká* ~ *tʃóóká* ‘it (cl. 7) burns’ shows the combined effect of Glide Formation and Palatalization which supports the equation of Coronal with “Front” as applied to vowels. The desideratum of simplicity motivates exploiting the existing feature Velar, necessitated for consonant phonology, as the place feature underlying *o,u,w*.

Finally we should considered whether it is necessary to posit Mid as a distinct feature, or is there already some feature in the consonantal inventory, which could be exploited to take its place? The feature Approximant is one possibility, as are Nasal and Voiceless. These are features motivated for consonants, which have found no other place in vowel phonology, and are therefore available for exploitation. Let us consider the possibility that "Mid" is simply the feature Approximant on moraic segments. That results in the following feature specifications for vowels.

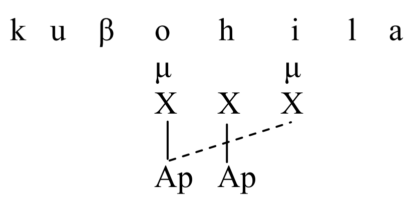
(52)	i	u	e	o	a	j	w
	μ	μ	μ	μ	μ		
	Pl	Pl	Pl	Pl		Pl	Pl
	Co	Ve	Co	Ve		Co	Ve
			Ap	Ap	Ap		

This analysis has two consequences which have to be empirically evaluated, but which cannot be resolved in this paper.

The first consequence relates to the dominance account of Approximant and the fact that when Stops lose that feature, they also lose the feature Approximant – because Stop dominates Approximant. We have not found evidence for assigning any equivalent of Stop to vowels. Under the theory that "Mid" is really Approximant, we might conclude that all vowels have the feature Stop, and mid vowels additionally have Approximant thereunder. An alternative is that the dominance relationship between Stop and Approximant is rule-governed (as indeed it must be by the logic of RSFP), and the question is, what is the rule? It might be "Approximant must be dominated by Stop", but it might also be "If a segment has Stop and Approximant, Stop must dominate Approximant". The formal theory of structural rules needs deeper investigation, before drawing form conclusions regarding the relevance of Stop to the theorized equation of Approximant with "Mid". The alternative that the feature in question is Nasal or Voiceless does not face this issue.

The second matter of some greater concern is that consonants, including approximants /l h β/, do not block vowel harmony: /kuβóhíla/ → [kuβóhéla] 'to tie (applied)'. Given the features motivated here, harmony has the following effect.

(53)



The No-Crossing Constraint, if it is part of grammatical theory, would prohibit vowel harmony from applying across an Approximant, thus /l h β/ should block harmony (but they do not). As formalized in (49), the input string satisfies the structural description of the rule because the moras are adjacent. This same issue arises whether we equate "Mid" with Approximant, Nasal or Voiceless. Clearly, the status of the No-Crossing Constraint is a very important question within this framework, one which we will not attempt to resolve here. Put simply, the evidence for No-Crossing within a minimalist, substance-free theory of phonological representations and computations must be re-evaluated, just as many other assumptions carried over from non-minimalist, substance-dependent frameworks must be re-evaluated. The alternative, should it turn out that No-Crossing is indispensible to the theory, is that harmonizing vowel features are disjoint with respect to consonant features.[40]

4.3. Summarizing Kerewe

We now summarize the feature assignments for Kerewe which have so been far motivated by the facts of the grammar. Evidence has been found from phonological behavior for the eight features Place, Vplace, Coronal, Velar, Stop, Approximant, Voiceless and Nasal, as well as the prosodic property μ. The designations α, –β, β, –β indicate that we can determine that [p,h] and [b,β] are the same in voicing, but we cannot tell if [p,h] is Voiceless or not Voiceless.

(54)

(54)	[p	t	tʃ	k	f	s	β	z	b	d	g	l	h	m	n	ɲ	ŋ]
	Pl	Pl	Pl	Pl	Pl	Pl	Pl	Pl	Pl	Pl	Pl	Pl	Pl	Pl	Pl	Pl	Pl
		Co	Co, Ve	Ve		Co		Co		Co	Ve	Co			Co	Co, Ve	Ve
	St	St	?	?			St		St	St	?	St	St	?		?	?
							A					A	A				
	α	Vl	Vl	β	?	Vl	-α		–α		–β		α	?	?	?	?
														N	N	N	N

[i	u	e	o	a	j	w]
μ	μ	μ	μ	μ		
VPl	VPl	VPl	VPl		VPl	VPl
Co	Ve	Co	Ve		Co	Ve
		?	?	?		
		A	A	A		

The grammatical facts have not yet given us a unique assignment of features to segments, even though it does determine the vast majority of features.

We must entertain the possibility that undecidable features are assigned at random from the grammatical perspective. This is especially obvious in the case of [v,dʒ] – if we assume that these segments are within the language's segmental inventory. The only phonological information that we can glean from these segments is that they are not moraic since they appear after vowels (*edʒaaházi* 'ship', *omundeléévwa* 'driver'). Barring an accidental-gap analysis, they are not Nasal since they do not appear preconsonantally. We have noted that in the rare instances where a non-alternating nasal appears before [dʒ], it is palatal. Positing that [dʒ] is palatal (Coronal + Velar) is consistent with this fact, which

[40] A further solution is available, analogous to the treatment of consonant transparentcy in UFT, that specific vowel and consonant features are the same, but may be dominated by distinct nodes for consonants versus vowels. However, there is no independently motivated other node which is exclusive to vowels that Approximant could be dominated by – /a/ is not specified with Vplace, as argued above.

could be sufficient evidence to assign that place specification to /dʒ/ rather than another specification.

We have yet to (clearly) distinguish *f, v* from other consonants. Let us compare what features are detectable among the phonological labials, plus phonologically undetermined [v]. We see that, within that set, every possible combination of features is exploited (although we are not certain whether [p,h] are phonologically Voiceless or not). With respect to possible combinations of Place not dominating Coronal or Velar, and also lacking the specification Stop, there are only 4 remaining possibilities: Voiceless or non-Voiceless, Nasal or non-Nasal.

(55)

p	β	b	h	m	f	v
Pl	Pl	Pl	Pl	Pl	Pl	
St	St	St	St	?		
	A		A			
α	-α	–α	α	?	?	
				N		

Considerations of syllable structure suggest that neither [f] nor [v] are Nasal. In fact, owing to above reasoning based on feature inheritance, a complete specification of [f] is available: it has a bare place specification, and an uncertain voicing specification. Obviously, [f] and [v] could be exactly the same except for voicing specification. Phonological reasoning does not relate the voicing of [f,v] to that of [p,b,h,β], so we would logically assign [f,v] the voicing values [β,–β], i.e. distinct from each other but not relatable to the voicing of [p,b,h,β]. It is also possible that [v] is assigned Velar place (but not Coronal, since the combinatorics of Stop and Approximant for Coronals has been exhausted). Thus the role for phonologically random feature assignment in Kerewe is very small.

A final issue regarding Kerewe features is the problem of the trigger of Spirantization: half of the suffixes which have initial /i/ trigger the rule. The question is, how do we distinguish those instances of /i/ which do not trigger the rule from those which do? Those which do not trigger the rule (/-ik-/ 'stative', /-il-/ 'applied', /-isj-/ 'causative') are derivational suffixes with the shape VC(G), and those which do trigger the rule (/-i-, -j-/ 'causative', /-i/ 'nominalization', /-ile/ 'perfective') are derivational of inflectional suffixes that do not have that shape. There is no obvious phonological generalization that makes this distinction, nor is there a morphosyntactic generalization, so unless the rule simply enumerates the specific triggering suffixes, some arbitrary representational property is required to either trigger the rule or block it.

In substance-based phonological theories, this kind of problem is either resolved by invoking a diacritic feature such as [+D] which has no phonetic interpretation and only serves to distinguish those segments that trigger the rule, or else by invoking an abstract phonetic-featural distinction where e.g. non-trigger [i] is underlyingly [ɪ] and trigger [i] is /i/ (or vice-versa). In RSFP, phonological features are *all* abstract, and we only require phonologically distinctive behavior to justify positing a feature. We have such behavior here – then what feature distinguishes the /I/ which triggers spirantization from the /I/ which does not? There are plenty of gaps in feature combinatorics which allow two kinds of /i/ to be distinguished, for example Nasal, Stop, or Velar. Nothing in the grammar favors one feature over the other, so speakers may be assumed to assign some feature at random.

5. Conclusions

It has been shown here that there is no logical requirement for UG to contain primitives expressing physical facts of speech – features need not be defined in terms of articulation or acoustics. There is no evidence that features are "defined" in grammar, nor that they are automatically assigned by any aspect of the language faculty. Instead, features are *used* distinctly to refer to how one segment is the same as or different from another segment in the grammar. There is no evidence for a UG limit on the number of (undefined) features, and any limit observed in a language emerges from the limited need for more features. However, UG clearly must contain formally-defined representational potentials. "Feature" is a concept of UG with a fixed formal nature. Likewise, grammatical computations have a fixed formal nature set by UG. These two aspects of UG, and the non-phonological ability to identify segments of a language, give rise to language-specific feature assignments, via the learned symbolic interface between the phonetic and phonological components. A key to a completely substance-free theory of phonology is recognizing that such computations are performed by a specific highly-symbolic module in the mind. Processing and retaining physical inputs is performed by separate mental modules, which are not exclusively linguistic. The interfaces between those aspects of the mind and the phonological component are not part of UG and, I argue, are learned based on the formal requirements of creating a grammatical system that is accountable for perceptible facts which are outside of grammar.

We have seen that phonological features for Kerewe can be learned simply by reference to two considerations. First, when a set of sounds is identified by a rule, those sounds have a feature in common – if another set of sounds are excluded by a rule, those sounds lack the feature. Second, even in lieu of class behavior in rules, the fact that a phonology contains distinct objects [p] and [b] means that some arrangement of features distinguishes those objects. Vietnamese seems to have no synchronic phonological alternations of the type /e/ → [ɛ] /__ X, /e/ → [ɤ] / Y__, though there may be rules governing segment combinatorics whereby [uən, uɛm] are possible syllable-final sequences but [ion, uɤm] are not. Even if there are no rules in Vietnamese which treat [v] and [z] differently, they are independent sounds of the language, so must be represented with different features in the grammar. Since phonological class behavior is the primary driving force behind feature assignment, when there is no class behavior but a distinction in sounds is still made, an interesting question arises. Is there any discernable pattern to feature assignment when the grammar is silent? Are available gaps in combinatorics exploited randomly (as presumed here), or are phonetic properties called on as a fall-back method to reaching a uniform analysis, given a particular fact pattern which constitutes the primary linguistic data? The obvious theoretical question to address is: given a corpus of data constituting the basis for acquisition of a given language, must the theories of grammar and learning be expanded so that it is guaranteed that there is only one analysis of that data?

FP certainly does not impose such an a priori requirement on either grammatical theory or the theory of learning, and it certainly does not achieve uniformity of analysis by stipulating lists of substantive default assumptions. RSFP as a theory of features only maintains that the responsibility of the LAD is to construct the most economical possible grammar. To the extent that extragrammatical facts of speech such as acoustic similarity could be known to a child, such facts might influence the outcome of the formally random coin toss performed by learning theory in acquiring the feature-assigning interface.

References

Anderson, Stephen. 1974. The organization of phonology. New York: Academic Press.
Author. in progress a. *The Kikerewe language.*
Author. in progress b. Phonological ontology.
Bender, Byron. 1968. Marshallese Phonology. *Oceanic Linguistics* 7: 16–35.
Blaho, Sylvia. 2008. *The Syntax of Phonology: A Radically Substance-Free Approach*. Ph. D. dissertation, University of Tromsø.
Bradshaw, Mary. 1999. *A Crosslinguistic Study of Consonant-Tone Interaction*. PhD dissertation, OSU.
Cho, T. & P. Ladefoged. 1999. Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics* 27: 207-229.
Choi, John. 1992. *Phonetic Underspecification and Target Interpolation: An Acoustic Study of Marshallese Vowel Allophony*. UCLA Working Papers in Phonetics 82. Los Angeles: UCLA.
Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper and Row.
Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
Chomsky, Noam. 1980. Rules and representations. Behavioral and Brain Sciences 3: 1–61.
Clements, G.N. & Elizabeth Hume. 1995. The internal organization of speech sounds. In John Goldsmith (ed.), *Handbook of Phonological Theory* (pp. 245-306). Oxford: Basil Blackwell
Clements, G.N. 1985. The geometry of phonological features. *Phonology Yearbook* 2:225-52
Clifton, John. 1976. Downdrift and rule ordering. *Studies in African Linguistics* 7:175-194.
Disner, Sandra. 1983. *Vowel quality: the relation between universal and language-specific factors*. UCLA Working Papers in Phonetics 58
Dunn, M.H. 1993. *The phonetics and phonology of geminate consonants: A production study*. Doctoral dissertation, Yale University.
Fromkin, Victoria. 1972. Tone features and tone rules. *UCLA Working Papers in Phonetics* 21: 50-75
Hale, Mark & Charles Reiss. 2008 *The Phonological Enterprise*. Oxford University Press, Oxford.

Hale, Mark, Madelyn Kissock & Charles Reiss. 2007. Microvariation, variation, and the features of Universal Grammar. *Lingua* 117: 645-665.

Hammann, Silke. 2011. The phonetics-phonology interface. In: N. Kula, B. Botma & K. Nasukawa (eds.) *Continuum Companion to Phonology*. London: Continuum, 202–224.

Jackendoff, Ray. 1990. *Semantic Structures*. MIT Press, Cambridge, MA.

Jakobson, Roman, C. Gunnar M. Fant & Morris Halle. 1952. *Preliminaries to Speech Analysis: the Distinctive Features and their Correlates*. Technical Report 13. Massachusetts: Acoustics laboratory, MIT.

Jakobson, Roman. 1949. On the identification of phonemic entities. *Travaux du Cercle Linguistique de Copenhague V*, 205-213.

Johnson, Robert. 1975. *The role of phonetic detail in Coeur d'Alene phonology*. Doctoral dissertation, Washington State University.

Kaye, Jonathan, Jean Lowenstamm & Jean-Roger Vergnaud. The internal structure of phonological elements: a theory of charm and government. *Phonology* 2: 305-328.

Khattab, Ghada & Jalal Al-Tamimi. 2014. Geminate timing in Lebanese Arabic: the relationship between phonetic timing and phonological structure. *Laboratory Phonology* 5: 231 – 269

King, Robert. 1969. *Historical linguistics and generative grammar*. Englewood Cliffs, NJ: Prentice-Hall.

Lombardi, Linda. 1991. *Laryngeal features and laryngeal neutralization*. Doctoral dissertation, University of Massachusetts, Amherst

Morén, Bruce. 2003. 'The Parallel Structures Model of feature geometry'. *Working Papers of the Cornell Phonetics Laboratory* 15: 194-270.

Odden, David. 2006. Phonology ex nihilo. Presentation at University of Tromsø phonology meeting, 6 Dec. 2006.

Odden, David. 2013. Formal phonology. In *A Festschrift on the occasion of X Years of CASTL phonology and Curt Rice's L^{th} birthday*, ed. by Sylvia Blaho, Martin Krämer and Bruce Morén-Duolljá. *Nordlyd* 40.1: 249-273

Peters, Ann. 1973. A new formalization of downdrift. Studies in African Linguistics 4: 139-153

Postal, Paul. 1968. Aspects of phonological theort. New York: Harper & Row.

Reiss, Charles. 2003. Quantification in structural descriptions: attested and unattested patterns. *Linguistic Review*, 20, 305–38.

Roberts, Ian. 2016. Introduction. *The Oxford Handbook of Universal Grammar*, ed. by Ian Roberts. Oxford Handbooks Online: Oxford University Press.

Sagey, Elizabeth. 1986. *The representation of features and relations in non-linear phonology*. Cambridge, MA: MIT dissertation

Vaux, Bert & Bridget Samuels. 2015. Explaining vowel systems: dispersion theory vs. natural selection. *The Linguistic Review* 32: 573–599