

Formal Phonology*

David Odden

OSU

Abstract

Two problematic trends have dominated modern phonological theorizing: over-reliance on machinery of Universal Grammar, and reification of functional properties in grammar. The former trend leads to arbitrary postulation of grammatical principles because UG “has no cost”, which leads to a welter of contradictory and unresolvable claims. The latter trend amounts to rejection of phonology and indeed grammatical computation, as a legitimate independent area of scientific investigation. This paper outlines Formal Phonology, which is a metatheoretical approach rooted in an inductive epistemology, committed to seriously engaging the fundamental logic of the discipline, one which demands justification of claims and an integrated consideration of what is known about phonological grammars, eschewing *ad libitum* conjectures and isolated positing of novel claims without evaluating how the claim interacts with other aspects of phonology. Debate over the proper mechanism for apparent segment-transparency in harmony, or the binary vs. privative nature of features, is ultimately doomed if we do not have a clear awareness of what a “grammar” and a “phonology” are. Misconstruing the nature of a phonology as being a model of observed behavior negatively affects theoretical choices, leads to confusion over what could motivate a claim about the nature of grammar, and in general, a lack of developed epistemological foundation leads to confusion over how to approach theory-construction.

1. The object of study in Generative Phonology

An obvious fact about language is that developmentally-normal adults can produce and comprehend an unbounded set of sentences in their language. What is most striking is that speakers can produce and interpret vast numbers of utterances that they have never heard before and could not have learned. This is only possible if speakers use a stock of primitive units plus a system of rules to create utterances, and children learn the primitives and rules rather than learning actual utterances. This then raises two central scientific questions. First, what is the nature of the rule system that enables speakers to create utterances: what does the system do, and how does it do it? Second, how are those rules automatically learned by observation of speech behavior, when the child acquires its language. What is actually *learned*?

A central feature of the theory of generative grammar is the claim that there are special cognitive properties which are particular to the human language faculty.¹ This means that human language has a particular nature, and its nature does not reduce to general statements about human mental ability. A system of rules – a grammar – operates on stored representations, and the fundamental goal of generative grammatical research has been to discover the nature of grammars and representations. The generative enterprise then logically reduces to positing theoretical conclusions in the form of general propositions

* This is a fragment of a draft of a longer work, still in progress. It is the result of numerous influences, and I hope those whose ideas appear here do not object to my co-opting their ideas and not even bothering to give credit where credit is due. I do want to specifically point to the obvious influence of the work of Hale and Reiss. Thanks to Kevin Gabbard, Kati Hout, Martin Krämer, Mike Marlo, Mary Paster, Markus Pöchtrager and two anonymous reviewers for comments on earlier drafts of this paper.

¹ See Chomsky (1965: 4) for a standard characterization of a generative grammar as a perfectly explicit description of the competence of the ideal speaker-hearer.



FORMAL PHONOLOGY

(“language sounds are feature bundles”, “features are attributes with two values”...), and empirically evaluating those conclusions as a model of the language faculty.

Generative grammar (and any good science) carries with it an ontological commitment to reality,² the claim that its theoretical conclusions describe, or “correspond to” facts about a thing. If a theory does describe the nature of language, it is valid or true as a statement about language; if a theory misdescribes the nature of language, it is false as a statement about language. Evaluation of a theoretical conclusion means determining the extent to which a theory describes a fact about what human language is. Comparison of theories means determining which theory is closest to exactly describing the nature of language.

Thus the theoretical conclusion (accepted by most phonologists) that utterances are composed of rule-governed concatenations of simpler units lacking intrinsic meaning – are formed by combining segments – is accepted as an undeniable fact. If a language were discovered whose utterances did not decompose into reusable units, e.g. if utterances in a language were composed of single units varying continuously in pitch or amplitude, then the claim would be refuted. Numerous additional, more sophisticated and specific conclusions have been posited in the course of the development of Generative Phonology.

This brief summary of the relationship between Generative Phonology and its object of study is hardly controversial and hopefully familiar. A very important question about Generative Phonology has, however, been glossed over, namely how theoretical conclusions should get to the intellectual marketplace and be judged. A proper methodology for phonological theory does not rely on random actions or emotional reactions, it requires a rational i.e. logical method of relating ideas to reality. What then is the nature of that method? While certain implicitly methodological terms are widely used in the course of talking about phonological theories (“simple”, “constrained”), there is little discussion of what these terms refer to, and why or even whether they are being used validly in our scientific investigations. The purpose of this paper is to outline Formal Phonology, which is a metatheoretically-driven approach to phonology, focusing on the proper logic of phonological investigations which leads to true statements about the nature of the human language faculty.

This approach holds that creation of scientific knowledge is a unified process where valid concept-creation means that existing knowledge implies a concrete theoretical concept, once focus is placed on the relevant facts. Theoretical conceptualization is not a cycle of arbitrary stipulation of isolated claims followed by a brief search for counterexamples, it is the continuous evaluation of the correspondence between ideas and the totality of what the ideas are about. The FP perspective on the relationship between facts and ideas is that facts are primary because they are what exist independently, and ideas are secondary – they are how we understand reality. A theory is a system of hierarchically-related concepts, and FP holds that foundational concepts must first be firmly established as correct, before erecting higher-order concepts on that base. It follows that positive means of justifying a claim are important, and “not being false” is too weak a criterion of proof. Not only must a claim be uncontradicted by the facts, it must also be the conceptually-simplest proposition that describes the facts. Finally, FP demands that a claim be appropriate to the thing that it is about. FP is about grammatical computations in a phonology, therefore claims about broader sound-system behavior are irrelevant unless they directly prove some fact about that computa-

² The disclaimer “it doesn’t matter”, in answer to a question about the nature of language, is usually false or misleading. It is false on a literal reading, unless the question is really meaningless. The physical shape of a symbol in a theory does not matter, so whether the predicate “becomes” is notated as “=>”, “→”, “⇒”, “→” or “becomes”, the symbols have the same interpretation. It is thus meaningless to ask whether the concept “becomes” is written as “=>”, “→”, “⇒”, “→” or “becomes”, because concepts are not physically-written objects, and it does not matter if rules are written as “a => b” or “a → b”. It *does* matter if features are binary or one thing becomes another, if one cares about the nature of language.

The statement is misleading when intended as “This theory is only a partial theory of the domain, and makes no claims about J – it is an open question what the nature of facts of J are”. An honest and non-dismissive way to express such a claim is to say “This theory is only a partial theory of the domain, and makes no claims about J – it is an open question what the nature of facts of J are”. Very often, the implication is that the only thing that matters is what the system “does”, i.e. the class of string-sets (“languages”) admitted under the theory. But this is false. To repeat: generative grammar is a theory of the language *faculty* – an ability, which is *how* the system does what it does.

tion. These last two desiderata lead to the conclusion that concern about overgeneration is not a valid criterion for evaluating a theory.

2. The nature of Formal Phonology *qua* theory

There are two approaches to reaching theoretical conclusions, which may appear to differ only in style. One, found in much of the earlier research in generative grammar, is the top-down approach. Under this approach, a general hypothesis is proposed, consequences of the hypothesis are deduced, and to the extent that the specific consequences seem to describe facts about language, the general hypothesis is eventually judged true, or rejected if the consequences describe falsehoods. In this approach, great emphasis is placed on hypothesis-testing, the deduction of specific claims that follow from the hypothesis, and matching empirical observation to such claims. An important question arises regarding the hypotheses posited in this approach: how are such hypotheses created? This question has received relatively little attention in the literature. This leaves open the possibility that hypotheses might validly be the result of day-dreams, as Kekulé claimed was the source of his idea about the structure of the benzene ring, though it is unlikely that hypotheses with overtly irrational bases would be given much attention in linguistics. In the top-down approach, the process of postulating a linguistic hypothesis is basically arbitrary, and what matters is whether the hypothesis has been disproven because it makes a false prediction.

An alternative approach, one gaining more attention in current theoretical linguistics, is the bottom-up inductive approach, which emphasizes that which the top-down approach ignores, namely hypothesis-creation. The inductive approach methodologically rejects arbitrary postulation of hypotheses, and rejects unjustifiably far-reaching claims.³ Instead, the approach requires that proposed hypotheses be integrated into a system of factual knowledge. The inductive approach relies on the fact that there already is a substantial empirical foundation, which is a precondition for positing a universal generalization worthy of further consideration. Since greater emphasis is placed on valid hypothesis-formation – the integration of an idea with existing knowledge to yield a concrete proposition – deductive inference plays a relatively minor role in scientific progress, though it remains useful for revealing unappreciated consequence of a hypothesis. A linguistic corollary of the insistence that the scope of a claim should not exceed that for which there is evidence, is that principles cannot be added to Universal Grammar *ad libitum*: there must be compelling evidence for positing the addition of a new principle or entity in UG. In part, this paper outlines the logic and practical application of this inductive approach to phonology.⁴

A second important metatheoretical issue about the nature of and arguments about grammars pertains to modularity and the explanatory scope of phonological theory. Approaches to language, since before generative grammar, have tended in two directions regarding the scope of grammar. For some, phonology is very broad and includes all aspects of linguistic and communicative behavior, and for others, the scope of investigation is more narrowly defined. This paper follows the principle that phonology is a specific and narrowly-defined domain, as advocated in numerous works by Hale and Reiss, where a phonological grammar is a formally-statable system of symbolic computations on “sound representations”, and the goal of phonological theory is to discover the nature of those representations and computations.

Autosegmental Phonology, Metrical Phonology and Optimality Theory are concrete claims about the form of phonological grammars, and carry no methodological commitments to what a phonology is or how a theory should be constructed; they also inherit the strong-nativist practice of freely attributing devices to Universal Grammar because UG is not learned (it is known a priori) and is thus considered to be

³ This does not mean that the inductive approach only allows extremely specific hypotheses with a tiny range of empirical application – it misconstrues the nature of theory-building to only state what has been observed. Rather, the approach rejects claims made without proper justification and scrutiny with respect to conceptual simplicity.

⁴ Formal Phonology says nothing about how to create an idea which leads to a theoretical concept, rather it addresses the logical relation that such an idea should have to existing knowledge, and how the resulting concept should be evaluated.

FORMAL PHONOLOGY

cost-free.⁵ Formal Phonology does not start with a body of a priori technical claims about grammars. FP is a metatheoretical approach which defines the nature of the enterprise, and it embodies an epistemology which allows valid hypotheses to be advanced regarding the nature of phonology. Such substantive claims as are made herein are claimed to be those which are justified by what is known. Thus the conclusion that phonological rules are formalized with featural expressions (not lists of atomic phonemes) would follow from the methodology of FP and the facts of human language, but would not be a foundational stipulation of the theory. The next section discusses the scope of inquiry in phonology, and the nature of inductive epistemology.

3. Domain restriction and falsification: the significance of modularity

Before delving extensively into the logic of theorizing in FP, it is necessary to be explicit about what the object of study is – it is futile to ask questions about the methodology of investigating an object, without first identifying what the object is. FP claims that an autonomous phonological component in grammar is necessary in order to understand linguistic behavior. FP holds that the concern of theoretical phonology is modeling the form of computations in the phonological component, including the things that computations are performed on. The kind of fact that FP is about is that Kimatuumbi /mu-wíkili-ε/ surfaces as [ɲɪwíkili] ‘2pl should cover’, Classical Arabic /tawāḍiʕu/ becomes [tawāḍu] ‘she lays’, and Karanga Shona /mu-á-ka-mú-bíkira/ is realized as [makámubíkira] ‘2pl cooked for him’. A study of “language sound” would include not just the study of grammars, but also a lot of speech behavior that is not about grammar. No theory of “language sound” can be correct without a theory of phonological grammars.⁶

As observed in Hale (2007), Hale & Reiss (2008) and elsewhere, production and perception of English “cat” involves more than the grammatical representation and computation of the output [k^hæt] from the input /kæt/. The entire chain of events involved in speech transmission or reception involves many non-grammatical and non-linguistic, indeed non-biological factors, which are outside the concern of phonological theory. The grammatical mapping from /kæt/ to [k^hæt]⁷ is squarely in the domain of a theory of phonology. FP is “Galilean”, in abstracting away from matters of performance (speaker attention, error, etc.) or acoustic differences between “cat” uttered on a cold, dry day versus a hot, humid day. FP is also concerned only with the grammatical properties of the *phonological* component, meaning that a formal phonology is not also responsible for accounting for what the syntactic, morphological, semantic or phonetic components of a grammar do.

Phonology being just one aspect of a general theory of language behavior (which is itself one aspect of cognition), not all facts about language sound are in the explanatory domain of phonological theory. Even facts which are “about language” and not environment may be outside the domain of a formal theory of phonology. For example, some languages have a process of post-nasal voicing (Kimatuumbi), and others have a process of post-nasal devoicing (Setswana). It is the responsibility of a formal theory of phonology to enable the description of both states of affairs, since both exist. Post-nasal devoicing is extremely rare (being found in only one group of Bantu languages), but post-nasal voicing is relatively well attested, being found in many Bantu languages as well as Greek, Japanese, Zoque, Maasai, and Imbabura Quechua. A formal theory of phonology is not held responsible for encoding this statistical generalization, since there already exist substantive theories of linguistic behavior including language acquisition, physiology and physics which account for this distributional asymmetry (see Hayes & Stivers 1995), and therefore the observation that a certain fact is “rare” or “marked” is irrelevant. FP only asks what the *form* of a

⁵ E.g. Yip (1988: 76) “The outside trigger is of course the OCP, a universal principle and thus free of charge”.

⁶ The converse is not true. The reason for the asymmetry is the simple fact that “language sound systems” properly includes “phonology”, and not vice-versa.

⁷ This assumes that the output of the phonology is in fact [k^hæt], which may not actually be the case.

phonological computation is, not what its likelihood is. Even a zero practical probability of occurrence does not put an observational gap in the domain of grammatical theory.

The potential correctness of a theory of phonological grammars is also not impugned by observing that the elements of the theory can combine in a way that correspond to so-far unattested sound-pattern phenomena. To restate the point, FP only asks what the form of a phonological computation *is*, not what it isn't. The situation where a theory can describe patterns that have not been observed is often seen as evidence that the theory is "too powerful", that it "overgenerates". For instance, numerous formal theories of phonology would allow the description of the hypothetical rule /p b f v m/ → [t d s z n]/__[y]. Such a process is not yet known to exist in any language. The fact that we have not encountered this event is almost certainly outside the purview of FP, since the process is formalizable in any general, empirically-adequate theory of phonology. It is obviously expressible as a feature-changing rule:

- (1) [+ant,+cons] → [+cor] / __ [+hi,-back,-syl]

It is also expressible as cross-planar spreading in a multilinear representational theory where front vowels and glides are [+coronal] – see Hume (1994) for the treatment of analogous changes of velars, triggered by front vowels and glides. Such a rule would not be expressible in a theory where all rules must be expressed as deletions under featural identity or as spreading if [y] is [-coronal], but such a theory is empirically untenable, since attested rules of the type /k g x/ → [č ĵ š] / __ [y] would then also be unformalizable. The fact that labial-coronalization is unattested is, then, not the result of the phenomenon being intrinsically unformalizable, therefore the gap is outside the scope of what a theory of phonology must explain.⁸ (Rice 2007 discusses a distinct sense of "gap" which *is* squarely in the purview of grammatical theory, where affixation is unexpectedly blocked in a defined environment, e.g. in Norwegian where imperatives do not exist for verbs whose roots end in unsyllabifiable clusters).

It might seem to be within the scope of FP to explain why a rule /p z ŋ/ → [ɣ ʌ t] / __ [y] is unattested. The reasoning could be that the collection of consonants /p z ŋ/ cannot be described using known tools for referring to subsets of a segmental inventory, and no phonetic property can be abstracted to describe the mappings {p→ɣ, z→ʌ, ŋ→t}. This argument logically depends on two premises – ones which must be previously established. The first, which is valid and indispensable in Formal Phonology and can be considered to have been established, is that segment classes and phonological changes are defined via conjunctions of features. It also requires features to be defined in terms of substantive universals so that the set {p,z,ŋ} and the respective structural changes could not be formally expressed. The latter is a questionable claim, not necessarily accepted in FP (it is a *possible* claim under FP), requiring justification.

A formal theory of phonological computation cannot be refuted by phenomena, and phonological phenomena are computationally epiphenomenal. A theory of phonological computations is refutable only by facts about phonological computations that contradict the theory. Phonological phenomena become relevant to a theory of computation only when there is a compelling pairing of a theory of representations and a theory of computation which renders the grammatical description of an attested phenomenon impossible, given those theories (the concern being, of course, that a theory which cannot describe actual facts is wrong, *qua* theory of language).

It is the proper concern of a Formal Phonological account of a specific language to say whether there is a rule of intervocalic voicing, or post-vocalic spirantization, or final devoicing in that language. If underlying /apa/ maps to surface [aba], then the phonological grammar must contain a rule or similar formal object which performs that mapping. It is also the proper concern of a Formal Phonology metatheory of grammar to determine whether phonological mappings involve string-changing mappings or string-filtering constraints. Therefore, this question must be decided empirically given the metatheoretical re-

⁸ One could imagine that the lack of a rule /py/ → [t] could be predicted by a well-motivated theory of phonological computation (though nothing presently known suggests that any such theory is possible). But such a prediction would be an accident. The theory would not be founded on the desideratum of formally precluding a rule of labial coronalization: instead, the principles which hypothetically yield this result would be independently justified on the basis of other facts.

FORMAL PHONOLOGY

quirements of FP, and not stipulated arbitrarily as a theoretical postulate. Other valid concerns of Formal Phonology are whether the operation embodied in a rule applies to just a single segment or can simultaneously apply to multiple (perhaps unbounded) segments; whether rule or constraint statements include universal and existential quantifiers or just universal quantifiers; whether references to substrings identified by such statements involve conjunction and disjunction or just conjunction (these questions must be decided empirically *given* the metatheoretical requirements of FP, and not stipulated arbitrarily as a theoretical postulate). These are matters about the form of rules, which is the concern of Formal Phonology.

The fact that rules of intervocalic devoicing or post-stop spirantization have yet to be uncovered in grammars does not justify adding new theoretical concepts to prohibit such rules, since the nonexistence of such rules is already explained via theories of learning and historical change, and duplicating functional reasons for the non-existence of intervocalic devoicing in the computational apparatus would be otiose. The lack of examples of intervocalic devoicing could imaginarily be “explained” in grammar by positing some complex of added notions about feature changing, the context “between vowels”, and values of voicing; but such a complication would be inferior to the simple phonology-external fact that intervocalic devoicing requires unlikely phonetic mechanisms to bring it about. Adding formal principles to phonology to say that intervocalic devoicing is not a computationally-possible rule contributes nothing, since there is nothing in the *form* of the computation that is “impossible”. It is thus a basic principle of Formal Phonology that the lack of instances of a certain kind of rule does not compel complication of the computational theory.

Likewise, grammatical reification of non-phonological explanations for the existence of common rules, in the form of grammatical “benefits” for rules like intervocalic voicing or *g*-spirantization, does not contribute anything to our understanding of grammars. It is known that across languages, the voiced velar obstruent stop [g] has a greater tendency to change to something else – voiceless, fricative, or sonorant – than do labial or alveolar voiced stops. The explanation for this derives from non-linguistic facts about airflow, human anatomy, and the physiology of vocal fold vibration. Since the explanation for the propensity of /g/ to change already has an explanation (Boyle’s Law is an independent fact of physics, Bernoulli’s Principle is an independent fact of physics, the location of the constriction in a velar is an independent fact of articulation, the mass of the vocal folds is an independent fact of anatomy), re-stating the sum of these factors as an autonomous principle of grammatical computations is entirely redundant, adding nothing to our knowledge of the universe. See Hale & Reiss (2000, 2008) for extended discussion of the logical problems with duplicating principles of phonetics and learning within grammatical theory.⁹ We return to the question of post-hoc “functional rationalization” of grammatical principles in section 5.

It should be clear from the preceding discussion that considerations of overgeneration in the language-enumerating sense play a minor role in theory-evaluation in Formal Phonology (see the next section for discussion of the proper role of overgeneration concerns in theory construction). FP does not thereby open up the theoretical floodgates and say “Everything is formally possible; the explanation for all unattested patterns lies in functional factors”. Such a move would be equivalent to denying the existence of phonology, which FP does not do. FP *does*, however, deny that phonological theory shoulders the sole explanatory responsibility for the facts of speech behavior. It is fair to say, though, that Formal Phonology is not particularly concerned over the fact that a theory allows unattested language “types”, when the required theoretical devices are well-justified. The concern of FP is, instead, over what theoretical devices are required to describe the nature of phonological grammars.

⁹ This is not to imply that phonetics is entirely irrelevant to phonology. First, it is relevant on practical grounds because a theorist has to know if a generalization about language has an independent, non-phonological explanation. Second, a theory of grammar must ultimately mesh with a theory of physical implementation as part of a grander theory of the mind, and if some theory of phonology patently contradicts what is clearly true about physical implementation, then that theory of phonology cannot be correct. It is crucial, though, that the theory of physical implementation be “clearly true”, not just “somewhat supported” or “the current belief”.

4. The nature of theoretical concepts

As outlined above, Formal Phonology adopts a bottom-up inductive approach to theory formation. The main epistemological fact which the bottom-up approach focuses on is the transition from knowledge of language facts to conceptual conclusions about human language, which is to say, the *justification* for theoretical conclusions. Befitting its status as a scientific theory, FP requires a logically-organized system of conceptual knowledge.

4.1. *The role of concepts in theorizing*

Conceptual knowledge is knowledge that extends beyond a concrete instance such as “this is Daddy”, “this is Fluffy” or “Fluffy broke the vase”. Conceptual knowledge refers to humans’ ability to identify characteristic properties of classes of things, as embodied in such identifications as “this is a person”, “this is a dog”, “Fluffy breaks vases”; and to further unify such identifications via higher-order identifications like “this is a mammal”. Positive attributes are a cognitive fact which humans can focus on, allowing them to identify instances of a conceptual category, by winnowing out the vast amount of stuff that is not relevant. To be useful as tools of knowledge, concepts must be simple, i.e. graspable by the mind in terms of a limited set of things that are already understood.

For instance, the phonological concept “spread” simply means “create an association relation between two elements”: a term encapsulates a simple definition. Suppose, however, that a theorist wanted to re-define the term “spread” to mean “add an association relation between an element dominated by X_i' and an immediately adjacent element X_j' to the right, provided that X_j' does not dominate anything; and otherwise, add a specification of X under X_j' which satisfies the OCP applied to X , scanning to the left”. Whether or not some other advantage accrues to such a redefinition (eliminating “default specification” as a separate concept), it would be malformed as a concept, because of its considerable complexity.

Reaching a conceptual conclusion requires integrating and differentiating concrete instances to establish the referents of the concept. Integration means recognizing an essential similarity between those instances, and seeing that certain differences are unimportant. Differentiation means recognizing that some instances of a broader concept are, as a whole, distinct from other instances. The individuals referred to by the concept “dog” (this dog, that dog, the dog over yonder...) are not only “the same” in some essential respect, but they are all different from the things referred to by the concept “sheep”, “cow” or “jackal”. Likewise, the various things referred to by “jackal” are “the same” in certain essential ways, and are as a whole different from the things referred to by “dog”, “sheep” and “cow”. “Jackal” and “dog” may also be unified, based on their similarities, into a higher-level concept “canid”, which, as a whole, refers to different things from “bovid”.

Theory construction is, at its heart, the enterprise of discovering valid concepts in a domain. The physical concept “atom” refers to unnumerable actual things, including “this hydrogen atom”, “that hydrogen atom”, “that hydrogen atom over yonder”, “this carbon atom”, and so on. Its validity rests on the fact that it applies to an open-ended collection of instances, and the concept interacts with other concepts to accurately describe many aspects of nature, for instance the Law of Multiple Proportions in chemistry. Needless to say, the Law of Multiple Proportions cannot simply be arbitrarily postulated in chemistry, it must be and was empirically established. To posit a scientific concept is to claim “This is true about the nature of the universe”, and the claim must be justified. Justifying a claim requires showing that it explains some fact, and that it is necessary in the light of alternatives. The fundamental necessity-basis for phonological concepts is that they capture the generalization “grammars do this” – there is no salvation for a theory which does not allow grammars to do what they actually do. Justification requires going beyond simply conjecturing that such-and-such might be the case, or finding an example that is consistent with a claim. It requires showing that there *is* a fact, and that upon consideration of alternatives, we must conclude that the facts are not already explained by existing concepts.

Modern formal theories of phonology are based on certain previous conclusions about the nature of the phonological computation, ones that have been amply justified by observing grammars in human lan-

FORMAL PHONOLOGY

guages.¹⁰ Non-phonological aspects of grammars (syntax and morphology) concatenate abstract elements (morphemes) into linguistic expressions, yielding the input to the phonology, and phonological computations map that input to a physically implementable form, by changing representational properties of the input. At an early stage in the development of a theory of the mapping, conceptual knowledge about the computation will be limited, but it will expand by establishing new concepts. Early on, it would be recognized that phonological computation can involve more than one rule either within or between languages, which leads to the questions “how is this rule different from that rule?” and “what do rules, as a whole, have in common”. We observe that rules change segments in a systematic way in a defined context (“between vowels”, “after a nasal”, “before *t,s,ʃ,l,n*”), and rules have three essential elements: the class of segments that are changed (the target), the class of segments that cause the change (the trigger), and the class of segments that result from the rule (the structural change).¹¹

It then becomes an empirical question what the nature of those elements is. Prior to determining what that nature is, it is metatheoretically known that their natures should be assumed to be the same. Only a single concept – “feature bundle” – is needed to grasp what a target, trigger or structural change is, and unless there is compelling justification for distinguishing their natures, one should not entertain the possibility that the nature of “target” and “trigger” are distinct.¹² Observation of rules leads to the conclusion that multiple segments can define a triggering context, thus {p,t,k} versus all others; {m,n,ŋ} versus all others; {p,b,m} versus all others, and this leads to the conclusion that rules are stated in terms of orthogonal properties (“features” – though the exact nature of those features, be they SPE-style, Government Phonology elements, or abstract minimalist structures in the fashion of the Parallel Structures Model or Radical Substance-Free Phonology, is a separate empirical question). The prior conclusion that rules have three formal elements (target, trigger, structural description) combines symmetrically with the new conclusion that rule elements are defined in terms of features, to yield the conclusion that targets are defined in terms of features, triggers are defined in terms of features, and structural changes are defined in terms of features. In the face of observational evidence from phonological rules, the simpler conclusion that a phonological grammar contains rules is elaborated by saying more precisely what a rule is, resulting in concepts identifying the components of a rule, and the understanding of these components is elaborated via the concept of “feature” which says how classes of segments are referred to.

Likewise, once we know that a rule may require the trigger to precede the target vs. follow the target, then we know that a large class of rules is possible – rules referring to {m,n,ŋ} before the target as well as {m,n,ŋ} after the target; rules referring to {p,m,b} before the target as well as {p,m,b} after the target. Each conceptual addition interacts with existing concepts to enable classes of rules to be formalized. What is added is a general concept, that rules can distinctively specify that the target precedes the trigger or follows the trigger, *not* a list of specific target-trigger pairs (“{m,n,ŋ} before {p,t,k}; {m,n,ŋ} before {a,e,i,o,u}; {m,n,ŋ} after {p,t,k}; {m,n,ŋ} after {a,e,i,o,u}...”).

Symmetry is the automatic but defeasible consequence of the requirement for conceptual simplicity. The conclusion that a rule may require one of two precedence relations (“the target precedes the trigger”, or, “the target follows the trigger”) combines with the method-concept for identifying elements in a rule

¹⁰ This discussion should be read as a normative account of how a theory of phonology should develop, not a historical account of how the theory did develop. It is framed in terms of a system of “rules”, but the same account might, in principle, be framed in terms of “constraints”.

¹¹ This discussion focuses on advanced theorizing, where we have already established through observation of languages that there are segments, that the shape of morphemes varies as a function of phonological context, that there are rules, and so on. These facts were established by prior observation, and we are now interested in higher-level conclusions about the nature of those rules.

¹² This is a simple application of Occam’s Razor, the Newtonian version being “We are to admit no more causes of natural things than such as are both true and sufficient to explain their appearances. Therefore, to the same natural effects we must, so far as possible, assign the same causes”. This version of Occam’s Razor is an essential principle of the epistemology of FP.

(featural definition) to yield numerous possible contexts – “after a vowel” and “before a vowel”; “after a nasal” and “before a nasal”. The definitions of the constituent concepts do not contain complex exceptions e.g. “unless the feature matrix is a target, and precedes the trigger, and refers to nasal segments”, and therefore there is no means, using just simple concepts, to exclude such configurations from the set of predicted structures.

Were it to appear to be necessary to introduce an asymmetry, a new concept might be added and justified to achieve this result, for instance “the feature [nasal] cannot serve as the sole trigger in a rule if the trigger precedes the target”. What then could justify adding such a concept to the theory of phonology? One might be tempted to do so upon the discovery that there were no rules which apply after nasal consonants and vowels, especially if there are adequate numbers of rules triggered by a following nasal consonant or vowel. But even given such a fact, and even if we had a vastly larger sample of phonological systems than we presently do, this would not compel the addition of a complicating concept.

Recalling that phonological theory does not bear sole explanatory responsibility for all of speech behavior, proper justification for adding a complicating concept to phonological theory requires showing that non-grammatical explanations fail. The rareness of rules which devoice consonants post-nasally is a consequence of physical tendencies which favor the output \updownarrow amba \updownarrow (the brackets refer to the physical output of the body, see Hale & Reiss 2007) even from [ampa] over \updownarrow ampa \updownarrow even from [amba], and the grammatical asymmetry in treatment of voiced versus voiceless post-nasal consonants is explained by the fact that the data which form the inductive basis for grammar acquisition is asymmetrically distributed because of this extra-grammatical factor. Nothing needs to be added to phonology to explain these facts. The apparent total non-existence of intervocalic devoicing in human language can likewise be explained by understanding the physical mechanism of vocal fold vibration, which renders a physical output \updownarrow bapa \updownarrow from [baba] a virtual impossibility – therefore, the theory of grammar does not need to say anything about why there is no intervocalic devoicing. In general, the lack of attestation of a certain language pattern is not a compelling argument for theory-complication in FP.

As noted in fn. 3, FP’s rejection of unsupported claims does not reduce theory construction to listing the known instances, because a theory is not a list of specific observations, it is a system of concepts which imply existing observations and predict future observations. A theoretical restriction is the addition of a complicating concept – a restriction on a theory is undesirable – and such an addition requires full justification, just as adding any computational mechanism requires justification.

4.2. *Evaluation of competing concepts*

The concern of a theoretical phonologist is identifying and selecting between domain-internal alternatives, which is to say, making theoretical choices about phonological grammars. For the sake of illustrating the logical analysis entailed by FP, a brief comparison will be made between two theories of feature-variables, and the matter will be pursued in greater depth in section 6. It is clear that some such mechanism is necessary in phonology, given multi-feature assimilations and other notions regarding segmental “identity”. Very many languages have rules assimilating nasals in place of articulation to a following consonant, and without some variable concept referring to “the set of feature values that pertain to place of articulation”, the grammar of a language having N places of articulation would require N separate rules to implement the notion “assimilates place”. A familiar theory allowing this to be expressed in rules was articulated in SPE, via the use of feature variables – $[\alpha F, \beta G, \gamma H]$. McCawley (1973) proposes an alternative mechanism limited to specifying the notion “is the same as” w.r.t. a feature. Reiss (2003) proposes a third theory of identity references; finally, autosegmental representation theory offers representational concepts which may cover the same ground. The question is, how should the choice between these grammatical theories be made? The crucial steps are clearly identifying the underlying concepts of these theories, and judging those concepts for how well they match the facts.

The SPE theory of feature variables does not just import the general mathematical notion “variable” and apply it to a domain where only two values exist. Although feature variables somewhat resemble

FORMAL PHONOLOGY

general mathematical variables x, y , found in equations such as $y=mx+b$, such a superficial resemblance is not sufficient justification for claiming that this same concept is being used in phonological theory.¹³ The concept of “feature variable” must be *linguistically* motivated and defined. While assimilatory processes of the general (SPE) form $[+X] \rightarrow [\alpha F_i, \beta F_j, \gamma F_k] / _ [\alpha F_i, \beta F_j, \gamma F_k]$ demonstrate the need for *some* such concept, they do not automatically justify the choice of the particular formal mechanism. Explicit comparison of alternatives is required. The most important and difficult first step is scrutinizing the structure of the claims, stating the theories explicitly, using well-justified theoretical concepts.

A feature-value variable in SPE is a random symbol (drawn from an unbounded vocabulary) which refers to a disjunction of values that features may have, viz. $\{+, -\}$. The important formal¹⁴ claims entailed by SPE variable notation are that:

- (2) Rules refer to feature values via the values that exist in representations, + and -, or via a variable. The vocabulary of distinct variables is unbounded.

The theory says nothing about the relationship between the feature and the variable associated with the feature. The attributes and values which are “features” are only accidentally related, and the particular pairings of value and attribute in rules are a distinctive property of individual rules, thus $[... \alpha \text{cont}, \beta \text{voice}...] \dots [... \alpha \text{cont}, \beta \text{voice}...] \dots [... \beta \text{cont}, \beta \text{voice}...]$ does not say the same thing as $[... \alpha \text{cont}, \alpha \text{voice}...] \dots [... \beta \text{cont}, \beta \text{voice}...]$. Since a particular variable can be assigned to any token of any feature and there is no bound on the number of feature-tokens in a rule, an unbounded set of variables is necessary. This theory will be referred to as Value-Variable theory.

An alternative theory, to be referred to as “Identical-Value theory”, is that the relevant phonological concept is “the value of feature X”, which presumes a tight bond between value and attribute. Ideas along these lines are found in the work of McCawley and Reiss. The first clause of (2) is also assumed as a statement of the form of rules in this theory. The value of a variable is automatically computed from the fact that it is specified on a given feature, and the comparison is between all variably-specified instances of that feature. Since values are not independent of attributes in this theory, the formal vocabulary only requires a single symbol, written here as “=”. A rule containing the condition $[=F_m]_i \dots [=F_m]_j$ matches a string $\dots S_i \dots S_j \dots$ if and only if S_i and S_j are both $[+F_m]$ or both $[-F_m]$. Thus the requirement that a pair of triggering segments have the same place features is expressed in Identical-Value theory as $“/... [=ant, =cor, =back] [=ant, =cor, =back]”$, meaning “the value of [anterior] in segment 1 is the same as the value of [anterior] in segment 2, and the value of [coronal] in segment 1 is the same as the value of [coronal] in segment 2 ...”.

These theories can be compared in terms of conceptual simplicity. Identical-Value theory has a single variable and the “variable” is not an autonomous thing, it is an additional kind of specification relationship “is the same”, to be added to “is plus” and “is minus”. Value-Variable theory has an unbounded collection of variables which must be treated as things separate from feature attributes. *Ceteris paribus*, a theory with a single added vocabulary item is to be selected over a theory with a larger (especially unbounded) added vocabulary. An empirical argument for Identical-Value theory derives from the fact that Value-Variable theory makes a broader – and unjustified – claim which Identical-Value theory does not make. Strong empirical evidence for feature variables is limited to a well-defined class of references, of the following general form (SPE notation), where each variable is associated with a single feature.¹⁵

¹³ The SPE notation differs syntactically from general numeric variables. A bare variable is meaningful in a numeric equation but not in a phonological rule. Variables can be multiplied and added in a numeric equation, but $[\alpha \beta F_i]$ is undefined in the SPE theory of notation.

¹⁴ I take for granted an interpretation of the notations, because a detailed development of variable interpretation presupposes a theory of string-to-rule matching and then says what is special about variables. Since it does not appear that one theory entails a substantially more complex interpretation algorithm, such discussion is orthogonal to the purpose of comparing the complexity and justification of two theories.

¹⁵ Other uses of feature variables are discussed in 6.3.

$$(3) X \rightarrow [\alpha F_1, \beta F_2, \gamma F_3] / \text{---} [\alpha F_1, \beta F_2, \gamma F_3]$$

Value-Variable theory makes an additional claim, that rules may also include a requirement that instances of different features have the same value, for example that the roundness of one segment must be the same as the voicedness of another. A rule of the form

$$(4) [\alpha \text{round}, \beta \text{back}] \rightarrow [\gamma \text{hi}, \chi \text{tense}] / [\delta \text{round}, \epsilon \text{back}, \gamma \text{hi}] \text{---} [\alpha \text{voice}, \delta \text{nas}, \beta \text{son}] [\epsilon \text{cont}, \chi \text{ant}]$$

is well-formed in Value-Variable theory. Every claim made by Identical-Value theory is also made by Value-Variable theory, and the converse is not true. We have now identified the difference between the concepts making up two theories of variables. Which sets of concepts best correspond to reality?

If the additional claim of Value-Variable theory were factually justified, the concept embodied in Value-Variable theory would be superior to that of Identical-Value theory – Value-Variable theory would be necessitated. There being no evidence for detachment of values from attributes, Value-Variable theory must be rejected in the face of the alternative theory, which is conceptually simpler,¹⁶ and which also does not make this additional unjustified claim.¹⁷ Any argument for Value-Variable theory would therefore have to focus on the empirical differences between the theories – showing for example that grammars do in fact impose conditions on rules such as “takes the same value of nasal as the trigger has for round”.

This discussion reveals the proper role for concerns about overgeneration. The right concern is not whether one concept interacts with other concepts to yield unobserved languages (“intervocalic devoicing”). Indeed, the ability of concepts to interact so as to describe things that have not yet been observed is a positive attribute of science – it is the power to predict. The proper concern is whether the correct concept was identified in the first place: or, was an unjustified claim made. Worry over overgeneration is never valid in isolation. Applied to competing concepts, proper concern with overgeneration is about going *beyond* necessity in positing concepts. The theoretical concept “feature variable” is not necessary, in the face of the alternative “identical value”.

Unfortunately, Occam’s Razor, which is wielded frequently in linguistic argumentation, is often construed the wrong way. Often, Occam’s Razor is interpreted to refer to the extension of a science, that is, to say that the logically preferred theory is the one that claims that there are fewer entities in the world. According to that logic, a representational theory allowing 3,159 distinct segments to be described is held to be superior to a theory allowing 3,160 segments (as long as there aren’t more than 3,159 known segments). But Occam’s Razor is not a metaphysical claim about the nature of the universe, that there are few entities, it is a normative statement about the proper form of *theories* of the universe. A theory is a system of concepts, not a collection of things-in-the-universe, and Occam’s Razor is a statement about systems of concepts. The wording of Aristotle (*Posterior Analytics*) reveals the original intent behind Occam’s¹⁸ Razor: “We may assume the superiority *ceteris paribus* of the demonstration which derives from fewer *postulates or hypotheses*” (emphasis added), that is, the fewer theoretical *concepts*, the better. Likewise, Aquinas holds that “If a thing can be done adequately by means of one, it is superfluous to do it by means of several; for we observe that nature does not employ two instruments where one suffices”, restated by Occam as “It is futile to do with more things that which can be done with fewer”, again emphasizing the centrality of the “instrument” – theoretical concepts – and not the things that theoretical

¹⁶ Phonological epistemology has not progressed to the stage that numerical measurement of simplicity can be undertaken. In numerically-quantifiable physical sciences, formal evaluation of the simplicity of a theory is more meaningful, since the applicable concepts have been made so explicit that they can be represented as a single symbol in an equation. One of the points of this essay is that phonological epistemology *must* progress so that we can better identify the individual logical claims embodied in a metatheoretical conclusion about grammars.

¹⁷ A further question that should be raised is whether a theory is overall consistent with what is known about human cognition. See section 5 for discussion.

¹⁸ The principle is simply named after a prominent Aristotelian scholastic philosopher, William of Ockham, who distilled a millenium of thinking on the topic.

FORMAL PHONOLOGY

concepts are about. The Newtonian statement (*Principia Mathematica*) mentioned in fn. 12 – “We are to admit no more causes of natural things than such as are both true and sufficient to explain their appearances” – focuses on admitting *causes* (explanatory concepts), not effects (entities in the world). There is no justification for elevating “economy of existents” to the status of a methodological principle. The tack taken in Hayes (1985, 1987) of complicating the concepts relevant to foot construction in order to preclude (presumed) non-existent languages is thus contrary to the methodology of Formal Phonology.

Put simply, when evaluating two grammatical principles of (apparently) equivalent conceptual complexity, the logically preferred theory is the one which most closely describes what is known to be true, because it does not make an unjustified claim.

5. Relevant evidence

Evaluating a theoretical concept is conceptually simple, since it amounts to determining whether the concept describes the relevant facts, and in comparison to alternative concepts gives the simplest description of “what is”. One aspect of theory-evaluation is simple: if a concept identifies known grammatical facts and its competitor denies the facts, the denying competitor is wrong. Any theory of grammatical structure which denies the linguistic fact “counter-feeding” is simply wrong (though whether or not the specific mechanism of rule ordering is required to describe that fact depends on what non-ordered alternatives there are). A theory of phonological grammars must be held accountable for what grammars do: they map inputs to outputs, thereby generating the strings that are the language, in the extensional sense.

In positing an argument for one theory over another, the relevance of supposed evidence must be evaluated. A common mode of argumentation in contemporary phonology involves looking outside of grammatical competence to find “confirming” evidence regarding grammatical competence, which often involves performing a behavioral test with speakers of a language, and conjecturing that the results of such a test provide evidence for a specific grammatical theory. Since FP is part of Generative Grammar, it is a mentalist theory which makes claims about how the mind operates – FP computations are claimed to be mentally real and intensional, not Platonic extensional abstractions unbound by the nature of human cognition. As a reputedly real aspect of the mind, it is not unimaginable that psychological tests could bear on the task of finding the correct theory of grammars. We must therefore review the nature of the phonological enterprise to determine what kinds of evidence would be valid for judging theories.

The phonological component maps from input to output (not speech), so facts about those representations and mappings are relevant. The object of study is *not* “the ability to use sound symbolically” or “the cognitive capacity of humans”, therefore facts from those domains do not gain automatic validity for answering questions about grammars. The question that has to be asked about potential evidence is whether it does indeed answer questions about phonological representations or input-output mappings; or does it answer a question about some other faculty, which interacts with phonology only indirectly?

An experiment which determines that oral air-pressure rises more rapidly during production of a voiced velar stop than it does during production of a voiced bilabial stop is not relevant to understanding the formal nature of phonological representations or mappings. No phonological rule implies anything about the rate of air-pressure buildup, and no formal theory of representations implies anything about the rate of air-pressure buildup in segments, thus such an experiment does not provide relevant evidence regarding the theory of grammar, not even “supporting” evidence. The experiment might provide an indirect rationale for a place-asymmetrical stop devoicing rule in some language, where /g/ devoices and /b/ does not, but that rationale is outside the theory of phonological computation and representation. What mediates between air-pressure rise and extant grammatical rules is historical change – see Hale (2007) for extensive discussion. The rules in a language’s grammar are induced on the basis of surface mental representations, which themselves are based on physical sound. The nature of the physical sound that results from the output of the grammar in producing a form such as /abaga/ is thus sensitive to whatever influences the rate of air-pressure rise, and may produce a body output of the type † abaka †, which may be transduced by the language learner as [abaga] or [abaka], depending on whether or not the physically-

predicted effect on vocal fold vibration is phonologized as categorial devoicing. Since grammar does not refer to rate of air-pressure rise, experiments about air-pressure rise tell us nothing about grammar.

It has been thought that evidence for a theory of foot-construction can be found via a forced-choice perceptual experiment to see whether humans have a propensity to organise streams of synthesized tones so that the acoustically prominent member is judged to be last in a binary pairing of beats. As expressed in Hayes (1985), “prominence contrasts based on duration lend themselves to iambic grouping, while prominence contrasts based on intensity lend themselves to trochaic grouping”, a generalization referred to as the “Iambic-Trochaic Law” in Hayes (1987). Under certain experimental conditions, listeners prefer the grouping (X X:) where the underscored beat is judged “strongest” in case one beat is longer, but (X X) in case the beats have equal duration. The prediction of the psychological-test driven theory is that a durationally-asymmetric organization (X: X) is not a possible parsing of beats, since according to the Iambic-Trochaic law, an initially-prominent beat should not be longer than the following beat.

Suppose that we assume the empirical correctness of the perceptual claim.¹⁹ Suppose furthermore that no language actually had Heavy-Light trochees.²⁰ Following the inventory-driven nature of certain claims about foot structure (that Universal Grammar provides a list of foot types which excludes the parse (HL)), the linguistic question at stake would be whether (HL) is indeed a computationally-possible representation. If the assumptions embodied in the Iambic-Trochaic Law are correct, putative non-existence of Heavy-Light trochees is logically explicable on two bases. One, the basis that FP insists on, is that Heavy-Light trochees are computationally possible, though functionally unattestable or rare. The other, which typifies the substance-driven approach to grammar, holds that the theory of grammar contains a restatement of the Iambic-Trochaic Law, perhaps in the form of a list of allowed foot types.

If the results of psychological testing are to be relevant for deciding whether (HL) is a computationally-possible representation, then the underlying mechanism causing the behavior of experimental subjects must be a principle of grammar. However, it is patently obvious that the behavioral patterns said to support the law are *not* the result of a linguistic principle at all, and are most relevant to music, since the stimuli are synthetically manipulated tones with no significant resemblance to speech. The cause of the behavior is the result of something external to language, perhaps reflecting a strategy for reacting to the requirement to find a “strong” beat, when there is no independently perceivable rhythmic structure to the beats. When a (supposed) fact has an extragrammatical cause, that fact does not constitute evidence for adding a grammatical principle – the evidence may even show that grammatical theory should say nothing about the matter.

Experimental findings might in principle constitute evidence for a linguistic theory, if experimental evidence convincingly demonstrates a general fact about the nature of human cognition which directly

¹⁹ There are reasons to doubt the claim. The literature cited by Hayes indicates that sequences with even duration and uneven amplitude tend to get a trochaic parse. Rice (1992) demonstrates the same preference for beats of even duration and amplitude – amplitude turns out to be irrelevant. Rice also demonstrates that higher pitch with equal duration tends to receive an iambic parse. Thus duration-difference is not the trigger for the parsing difference, more generalized “prominence” is. In pilot experiments, I have found that inversely correlating pitch and duration, where sequences of beats have the shape “long-low + short-high”, also yields a trochaic parsing judgment, again suggesting that inequality of prominence (not duration) is the triggering factor for iambic parsing. In sequences of “long-low + short-high”, both beats have some prominence.

Since linguistic stress usually correlates with higher pitch, stressed syllables are, on the surface, prominent. It follows, then, that (X X:) i.e. classical iambic length-distribution but foot-initial stress would also be a “natural” type of trochee – linguistically speaking, it seems to be a non-existent pattern. Finally, pilot experimental evidence indicates that the results depend crucially on a particular experimental setup, where beats are evenly separated and amplitude is tapered so that listeners have no idea how the sequence begins or ends. When the setup is changed so that listeners know whether the sequence starts (short-long) or (long-short) – as is the uniform case in natural language – then English speakers, at least, identify the long beat as being “strong”, and correctly place the strong beat group-initial or group-final, depending on how the sequence begins.

²⁰ See Rice (1992), Mellander (2003) for languages with (HL) feet.

FORMAL PHONOLOGY

implies a choice between linguistic theories. If a hypothesized grammatical principle fundamentally *contradicts* a basic fact about human cognition, the grammatical principle cannot be correct since grammar is an aspect of cognition. If a proposed grammatical principle is simply *different* from what has been seen in other areas of cognitive science, then we may have discovered something interesting about language.

Here is an imaginable scenario of that type. The previous section has discussed competing theories of the nature of feature variables, Identical-Value where values and features are inextricably linked, and Value-Variable theory where value can be factored out and applied to a different feature. The applicable psychological question is whether the mind actually abstracts value from attribute, and can graft one value onto another attribute. It is obvious that humans can sensibly compare the weight of two objects, or their colors or temperatures – we can compare the value of one given attribute between two entities. It is not sensible to say, except jocularly, that “This stone is as heavy as that book is blue”. Experimental psychological evidence might imaginably establish that the mind does not treat values as a floating abstraction detached from an attribute.²¹ If such a result regarding nonlinguistic cognition could be established, then the results of psychological testing could in principle show that a hypothesized linguistic concept contradicts what is known about the mind, giving evidence for Identical-Value over Value-Variable Theory. Arguments based on properties of the mind have to be treated cautiously, as indicating a potentially fruitful source of extralinguistic evidence about cognitive foundations, *if* the foundations can be firmly established. Those theoretical foundations are not yet firmly established, so arguments based on properties of cognition may be suggestive, but not probative.

Certain experiments might provide evidence about grammar, namely those which directly call on grammar. The classical example is the wug-test, where subjects are manipulated to create a certain linguistic input, and then an output form is elicited. That output tests some theory about an aspect of the grammar. Thus when an English-speaking subject is presented with an object named [wʌg] and asked (indirectly) for the plural, the form [wʌgz] is the response, and likewise [lʌp] should be found to have the plural [lʌps]. The wug-test indirectly taps into the grammatical system, by giving the subject an opportunity to combine a conjectured form (/wʌg, lʌp/) with a highly-probable strategy for forming plurals (affix /-z/). The underlying forms are virtual certainties. In English, [wʌg, lʌp] could only derive from /wʌg, lʌp/ though in some languages, an output [bunt] might come from /bunt/ or from /bund/ so that producing [bunt] does not provide the subject with enough information to uniquely select the underlying form), and there are only a few lexical alternatives to the plural affix /-z/, exemplified by *mice*, *sheep*, *children*.

Results from such tests must be used cautiously. An unpredictable output may reflect a fact about the grammar, or simply a problem with the subject’s ability to cope with a counterfactual research method (for instance, stipulating that there is such a bird-like thing with that English name). Wug tests carry the added burden that the subject must effortlessly adopt a new underlying form, and must actually unconsciously apply the phonological rules of the language (does not semi-consciously compute a response based on their memory of spelling and grammar rules from elementary school). In normal language use, we rightly assume that speakers are unconsciously calling on their internalized grammar to generate and interpret utterances. In an experimental setup where subjects are being quizzed on their ability to form plurals of non-words, we cannot assume that production is unalloyed by subjects’ strategies for not looking like they don’t know how to spell or talk right, therefore the experimental setup needs to be subtle.

Interpreting wug-data is similar to interpreting elicited regular-language data, which field workers do all the time. It is well-known to field workers that individuals vary in their ability to generate forms in response to a stimulus, and it may take some practice at performing the task for a speaker to actually understand what the scientist is looking for. In a field-work context which lasts for months or years, these

²¹ A relevant experiment would have to test whether such an ability varies between humans, and requires more than ordinary inductive reasoning to acquire. The ability to automatically acquire language by observation of one’s surroundings is uniform in humans, whereas the ability to construct mathematical and scientific theories is a special talent possessed by a small fraction of the population.

research start-up effects have negligible impact on the resulting data. In the context of half-hour long psychological tests, start-up effects will be quite substantial, and will always cast doubt on the claim that the test data reflect a fact about the language, rather than an effect of the experimental setup. Just as field-workers know that speaker productions have to be evaluated critically in terms of the question whether aspects of production result from competence versus performance, “laboratory phonologists” also need to evaluate speech behavioral evidence critically, and not assume that grammar and behavior are the same thing. Grammar *underlies* behavior, and is not the sole contributing factor.

Another kind of potentially valid grammar-external data comes from language games *a.k.a* ludlings. See Bagemihl (1988, 1995), Vaux (2011) for phonological overviews. Such games have, in the past, revealed a number of interesting facts about phonological structure by validating abstract underlying representations, the existence of certain phonological rules, or supporting a representational claim regarding prosody versus segments. The characteristic operation defining the game is, apparently universally, a transformation of a linguistic form that resembles word-formation processes (movement, infixation) but one which is never employed in that form in ordinary language (infixation after *every* syllable; random transposition of segments; long-distance segment movement *a la* Pig Latin).

It is not clear whether the fundamental operation defining the language game is within the domain that phonology is responsible for, in part because it isn’t even clear what the proper analysis of morphological metathesis, infixation and reduplication are.²² The fact that language games often involve insertion of a CV sequence everywhere does not per se mean that the phenomena are beyond the reach of phonological grammars, since phonological grammars need to account for the insertion of CV sequences in specific locations: the peculiarity of language-game formation seems to reside in the extent to which the operation takes place, not in what kind of operation takes place.

The most uncontroversial valid evidence from language games lies in how a game-transformation interacts with the grammar, thus it is important to distinguish the mechanism of the change from the consequences of such a change. Al-Mozainy (1981) documents a language-game in Bedouin Hijazi Arabic where root consonants are freely transposed (thus /dff/ → [fdʕ], [ʕdf], [ffʕ] etc). The transformation itself is not evidence that phonology includes random segment moving as an operation.²³ The relevance of the language-game facts lies in how that transformation interacts with independently-motivated aspects of the grammar. For example, regular-language /ðarab/ surfaces as [ðarab] ‘he hit’, where a height-dissimilation rule of the language does not affect initial /a/ because the intervening consonant is /t/, which regularly blocks raising. The language game’s transposition can alter the intervening consonant, and in the language game, the form appears as [ribað], [biðar], [riðab], [ðibar] and [barað], exactly as predicted by applying the independently motivated rules of the phonology to the output of the language-game transformation. Similar evidence from a Tigrinya ludling presented in Bagemihl (1988) provides confirming evidence for the underlying form of the root-final consonant and for the reality of the postvocalic velar spirantization process (which does not affect geminates). In the regular language, /sanduk’-ka/ becomes [sandukka] ‘your m.s. box’ via a laryngeal-assimilation rule which creates geminates from /k’+k/, which blocks spirantization. The ludling inserts /gV/, resulting in [saganigidugux’igikkaga], independently showing the reality of underlying /k’/ and the spirantization rule.²⁴ The value of such evidence is that a simple operation feeds into the system of phonological computations in a revealing way.

²² To take reduplication as the best-known example, there are numerous mutually incompatible theories of what object is concatenated with a stem, to trigger phonological copying, and how it is concatenated. The classical templatic approach posits that reduplicants are partially-defined phonological strings such as “σ”, “F”, “CV”, whereas the OT approach posits a single diacritic entity “RED” whose shape is governed by constraints.

²³ Independent of language games, synchronic metathesis demonstrates that segments move.

²⁴ This is not to imply uncritical acceptance of Bagemihl’s analysis of the Tigrinya ludlings or his theory of language games. The point is solely to indicate *potentially* useful evidence from language games.

FORMAL PHONOLOGY

One of the reasons why “independent evidence” has become important in evaluating phonological theories is that theorizing has tended to be an *ad libitum* enterprise where argumentation often reduces to saying “Here’s my position, show me the counterexample”. Extragrammatical evidence has become a necessary arbiter in theorizing, since there are few counterexamples (the inventory of theoretical concepts has become vast and there are many degrees of freedom in constructing analyses), and theoretical concepts tend to be only thinly attached to perceptible reality. Coupled with confusion over what the domain of phonology is (phonology is not the same as “language sound systems”, nor is it “what speakers know or can do w.r.t. sounds in language, or not even in language”), this has meant that important questions cannot be resolved except in a partisan manner that reflects arbitrary assumptions. Formal Phonology is a metatheoretical framework that guides the formation of theoretical concepts, thus limiting the potential for creating complex systems of unrelatable constructs that are in competition.

6. The construction of a hierarchy of phonological concepts

The essential characteristic of the bottom-up approach of Formal Phonology is that theoretical claims are to be built on an existing solid foundation. This section lays out the methodological issues involved in answering one question, namely what is the proper account of feature-variable behavior, initiated in section 4. I phrase the question in terms of “feature-variable behavior”, not just feature-variables themselves, since the very question whether variables are needed is itself a matter to be addressed. I give the benefit of the doubt to the SPE theory of value variables, so “feature-variable behavior” will refer to “facts that would be formally treated using Greek letter variables”. Since, historically speaking, phonological research has not rigorously pursued the bottom-up formal approach, this section illustrates how construction, evaluation and repair of theories can take place “in the middle” of a conceptual hierarchy.

Certain concepts can be assumed as already established. For example, the claim that grammars contain rules that map representations onto representations is taken to be beyond reasonable doubt. As clarified in fn. 10, by “rule”, I do not necessarily mean an ordered, same-type representational mapping of derivational rule-based phonology, but in order to simplify the discussion, I only discuss how the conceptual analysis works in a derivational rule-based theory, leaving exploration of analogous constraint-based analysis as an exercise for the reader. I also take it to be established that representations include “features”, and rules are stated as operations on features.

Other unresolved questions impinge on theories of feature-variable behavior. Those issues and their logical relations must be identified, since they bear on how feature-variable facts should be modeled. The most important question is whether features are binary attribute-value pairs, or privative present / not-present distinctions. A formal theory based on the assumption that features are binary attribute-value pairs could not be correct if they are actually privative present / not-present distinctions. On the other hand, it may be possible to articulate a theory requiring only minor changes in the structure of the theory to accommodate a determination that features *are* binary, or *are* privative (or a mix of the two): in fact, such a theory is articulated below.

6.1. Privativity²⁵

In the development of theories of feature underspecification, it has been proposed that at least some features are single-valued, “privative” in the terminology of Trubetzkoy. An argument for the privativity of [voice] is advanced in Mester & Ito (1989), to resolve a contradiction between Radical versus Contrastive Underspecification theories. The two most-compelling arguments for privativity are transparency and

²⁵ I ignore pseudo-privativity, where each binary-theory feature metatheoretically maps to two privative-theory mutually-exclusive attributes such as “voiced” and “voiceless”, “oral” and “nasal”, and so on, which translates value-attribute pairs into two single lexical items. I also assume the standard principle of privative analysis that rules cannot refer to the lack of a specification. If this principle is abandoned, the resulting theory would be a notational variant of binary feature theory, where “F” maps to [+F] and “lacking F” maps to [-F].

asymmetry. The asymmetry argument is the claim that one value, usually [-F], is never referred to in rules, and segments that are [-F] fail to undergo rules that they might be expected to undergo (or, are maximally susceptible to being specifically targeted). Tendency towards asymmetrical behavior is not a valid basis for deriving a theoretical construct in FP, especially when the supposed asymmetry is only marginally valid from an empirical perspective – there are no hard-core asymmetries where e.g. labial is *never* the sole target for place of articulation assimilation (Hume & Tserdanelis 2002); both [-high] and [+high] do actually spread in vowel harmony; [-round] spreads in Hungarian vowel harmony, etc.

The transparency argument is more credible, since it depends on a strong formal claim, that no element can intervene (by definition, on the same tier) between S_i and S_k which are rule-adjacent. Lyman's Law in Japanese voices an initial obstruent, provided that no voiced obstruent follows within the stem. The scan for a blocking voiced obstruent would be thwarted by a voiceless obstruent between the voiced obstruent on the right and the target if voiceless obstruents are [-voice]. Since voicing is contrastive in Japanese obstruents, they cannot be underspecified for voicing, according to Contrastive Underspecification (which Mester and Ito argue for). The combination of transparency evidence for nonspecification of [-voice] plus the principles of Contrastive Underspecification can only be resolved, argue Mester and Ito, if *voice* is universally assumed to be a privative feature. This at least renders the hypothesis of privativity worthy of further consideration, hence a matter relevant to the theory of feature-variable behavior.

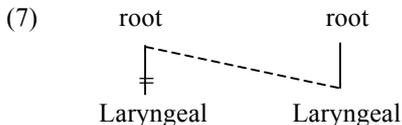
It is uncontroversial that languages (for example Hungarian) have voicing assimilation rules whereby voiced obstruents become voiceless before voiceless obstruents, and voiceless obstruents become voiced before voiced obstruents, standardly stated in SPE notation as (5).

$$(5) \quad [-\text{son}] \rightarrow [\alpha\text{voice}] / ___ [-\text{son}, \alpha\text{voice}]$$

This formulation presupposes that [t] is [-voice] and [d] is [+voice], which contradicts the presupposition that features are privative. If segments are unorganized sets of (unvalued) attributes, regressive voicing assimilation cannot be formalized. Because Hungarian exists, then either features are (at least) binary-valued, or else segments are not unorganized sets of attributes. Since Clements (1985), there has been reasonable evidence for imposing some form of organization on features, which opens the possibility that the distinction [t] / [d] is represented as in (6).



Given this representational possibility, rule (5) can be modeled, at least given autosegmental rule formulation, as (7).²⁶

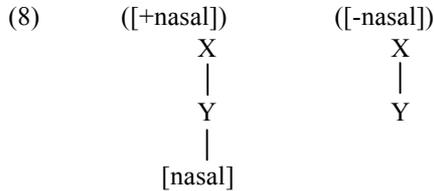


Underlying /gt/ would surface as [kt] by spreading the Laryngeal node of /t/. Since /t/ dominates nothing, when its Laryngeal node replaces that of /g/, both segments would be phonetically voiceless unaspirated, following a widely-assumed interpretive convention on representations. The same kind of formalization

²⁶ This option implies that all laryngeal features assimilate, not just voicing. Symmetrical assimilation of voicing but not glottalization or aspiration would challenge this rule, but might also spur the postulation of an additional intervening node between Laryngeal and voice. This point is discussed below.

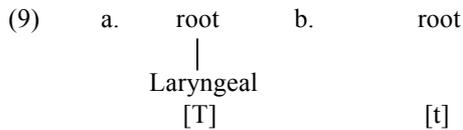
FORMAL PHONOLOGY

would be possible for all features, as long as there are sufficient dominating nodes, thus spread of nasality and orality can be accomplished within a privative theory, if nasality is structured as in (8).



Depending on the actual existence of rules apparently spreading both values of all features, the correctness of Privative theory may depend on whether such intermediate nodes exist. As a case in point, Sagey (1986) postulates that a non-branching Soft Palate node intervenes between [nasal] and the Supralaryngeal node, thus Y in (8) is “Soft Palate”.

Lombardi (1991), who argues for the privative theory of voicing, actually rejects this kind of alternative for voicelessness, and presumes (p. 38) that “the proper representation [of voiceless obstruents] is one with no Laryngeal node”. We will refer to this theory as “constrained privativity”, since it limits grammatical theory to including a subset of the representations that would otherwise be possible. Such a representation is automatically possible given the general formal principles of theories of feature geometry, so to enforce this desideratum, a theoretical concept must be added to the theory. The motivation for the claim is that “(t)here is no phonological contrast between a representation with a bare Laryngeal node and no Laryngeal node at all. Therefore any theory should allow only one type of representation”. The simplest formal system allows a language to distinguish (9a) and (9b), but no such language had been identified.²⁷ Lombardi’s claim is that there is a grammatical principle which renders (9a) impossible.



As discussed in section 4, Formal Phonology rejects “economy of things” in theory-evaluation, embracing instead “economy of concepts”. FP has no methodological commitment to dealing with the claim that languages with such a contrast are lacking: the proper concern would be over the conceptual systems admitting or prohibiting representations, and a formal prohibition of such a difference is otiose in lieu of positive evidence for a prohibition. Indeed, if features are privative, such a distinction could be invaluable in harmonizing the model to the facts of language (symmetrical assimilation of features is not unknown), and one could reasonably posit representation (9a) for a voiceless consonant which causes devoicing. Stronger evidence than the lack of compelling examples of such languages is required to justify adding a complicating principle to the theory.

Unfortunately, Lombardi does not identify a specific theoretical principle to yield this restriction. An FP analysis of the question requires not just a phenomenological statement “this is how it should be done”, it requires a specific concept which has that result. Ordinary application of simple feature-structure building rules would allow both representations in (9), so to rule one of them out (and prohibit a contrast within a language), an additional theoretical concept must be added to force a choice of representation-types. Being generous to the proposal, that concept would likely be a node-pruning convention to the effect that non-terminal node N_i not dominating a feature is prohibited, and is eliminated if derivationally created. Hence one complication necessitated to implement constrained privativity is the inclusion of an otherwise unnecessary pruning convention (a “concept of action”).

²⁷ A three-way contrast in Turkish stops w.r.t. final voicing is analyzed in this way (Inkelas 1994).

In order to implement this convention, it also becomes necessary to distinguish typically terminal nodes i.e. “features” from pre-terminal nodes such as Laryngeal and Place. Under any theory of feature organization, a distinction can be computed from the principle that says “Laryngeal dominates [voice], [s.g.], [c.g]”, but apart from rules stating what Laryngeal etc. dominate (where the notion “preterminal” is implicit), a formal distinction between terminal and pre-terminal nodes has been unnecessary. In order to implement the desideratum of not having empty Laryngeal nodes, a permanent representational distinction between “nodes” and “features” must be added to the theory, and a principle prohibiting empty nodes must be added.

This distinction cannot be read off of unalloyed representations. In (6), the actually-terminal node of [d] is the feature [voice], and the terminal node of [t] is the Laryngeal node: only the latter is prohibited. The distinction cannot be computed by referring to representational potential,²⁸ namely nodes that may dominate material (subject to pruning) versus ones that may not. In feature-geometric work since Sagey (1986), the features [anterior] and [distributed] have been dominated by Coronal, but it is standardly assumed that Coronal can itself be a terminal node (when a language has a single coronal series, thus no specifications of [anterior] and [distributed]). Of course it is sometimes assumed that Coronal itself and possibly Place are unspecified when there are no [anterior], [distributed] distinctions e.g. Avery & Rice (1989), Paradis & Prunet (1989), but this assumption is contradicted by the wide-spread assumption that laryngeal consonants are the ones that are unspecified for place, e.g. Steriade (1987). The specific claim of Avery & Rice does not contradict the theory of placeless laryngeals, because they do not assume the convention against empty “nodes” which Lombardi’s account requires.

It would be possible to reject claims of possible coronal underspecification and the theory that [anterior] and [distributed] are dominated by Coronal in order to preserve the node-pruning convention at issue (the central idea necessary to express constrained privativity), but that convention must then shoulder a substantial extra burden of proof and re-conceptualization, since the idea of constrained privativity leads to the denial of a number of other claims that seem empirically plausible. Taken in isolation, the claim that languages do not make contrasts like those in (9) seems innocuous, but a logical analysis of the consequences of the claim reveal numerous countervailing considerations from domains not necessarily connected to voicing and, as FP insists, claims about the nature of phonological grammar cannot be considered in isolation from the totality of knowledge about phonology.

To summarize the issue of privativity, the first question is whether privative feature theory is *possible*, that is, is logically consistent with what it known. Assimilation of [-voice], [-hi], [-nasal] etc. appears to show that universal privativity is actually *not* possible. An analytical tack to harmonizing such rules with privativity is to assume intermediate nodes as in (6) and (8), so resolution of the privativity question must defer to resolution of the intermediate-node question. If empty intermediate nodes are allowed, a privative analysis is possible. If they are not, then some principle must prevent them (and that principle must be justified). The best option for such a principle seems to be a ban on formally-terminal organizing nodes, which requires adding a further representational distinction between “organizing node” (e.g. Laryngeal, Place) and “feature” (e.g. Coronal, Nasal) which is orthogonal to the formal terminal / non-terminal property. Adding such a distinction then requires justification. Saying “doing so allows a principle to be added to ban terminal Laryngeal without banning terminal Coronal” is not proper justification – it is a circular appeal to the presumed *correctness* of the very principle whose *possibility* is in question. FP demands not just that the claim be shown to be possible, it must be shown to be *necessary*.

The reason why it matters whether features are privative or binary is that the notion of a “variable ranging over values” is meaningless with privative representations. If Privative Theory is correct, neither Value-Variable nor Identical-Value theory can be right, since the notion “value” which these theories depend on is meaningless in Privative Theory. However, a concept which is *similar* to Identical-Value

²⁸ Justifying a specific principle that appeals to “representational potential” would also require prior demonstration of the claim that rules or meta-grammatical principles can refer to “representational potential”.

FORMAL PHONOLOGY

theory, call it Identical-Presence theory, is definable on privative representations. Rather than comparing the plus-or-minus values of a feature F_k in segments S_i, S_n , one could compare the existence of F_k in S_i, S_n , so that the condition “[$=F_k$] ... [$=F_k$]” is satisfied iff F_k exists in both S_i, S_n or is lacking in both S_i, S_n . A third theory, call it Identical-Presence-or-Value theory is definable on mixed privative and binary representations, where “[$=F_k$] ... [$=F_k$]” is satisfied iff F_k exists in both S_i, S_n or is lacking in both S_i, S_n when F_k is a privative feature; and the condition is satisfied for binary-valued F_k iff F_k has the value “-” in both S_i, S_n or the value “+” in both S_i, S_n .

What these theories have in common is the most general concept, call it Feature-Identity theory, which says that comparisons are made between instances of the same feature, and both theories state that rules can require segments to be “the same” with respect to a feature, abstracting away from the exact nature of the feature *qua* entity or attribute-value pair. In short, the correct concept would not be Identical-Value, Identical-Presence or Identical-Presence-or-Value theory, it would be that simplest justified concept, Feature-Identity theory, which is that rules refer to the sameness of features, not saying whether features are privative or binary. By the logic of FP, the concept of variable would not be so specific as to be framed in terms of values, presence or a combination therein – it would abstract away from that orthogonal question. Privativity is a separate question.

At this point, it should be clear that the correctness of Privativity has no impact on the correctness of Feature-Identity theory, but it and the facts of phonological systems are logically incompatible with Value-Variable theory, on the assumption that no features are immune to symmetrical assimilation.

6.2. Representations or variables?

A partially independent question bearing on variable behavior is whether the effect can also be explained on the basis of representations. Under certain assumptions, feature-value references might be accounted for representationally, without variables. This refers in part to the potential relevance of the OCP in rule operation, and in part to the theory of assimilation as spreading.

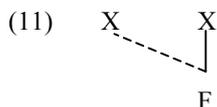
The potential for capturing identical-feature reference via autosegmental representations derives from the fact that a surface same-valued sequence potentially (but not obligatorily) has a unique characteristic. Non-autosegmental representations allow one representation of a different-valued sequence and one similar representation of a same-valued sequence in (10a). Autosegmental representations allow one representation of a different-valued sequence and two representations of a same-valued sequence in (10b), one of them being quite different in nature.

(10)	a.	[X, +F]	[X, -F]	Segmental non-identical sequence
		[X, +F]	[X, +F]	Segmental identical sequence
	b.	[X]	[X]	Autosegmental non-identical sequence
		[+F]	[-F]	
		[X]	[X]	Autosegmental identical sequence ₁
		[+F]	[+F]	
		[X]	[X]	Autosegmental identical sequence ₂
		\	/	
		[+F]	[+F]	

If an identical-value sequence is represented as one specification with a multi-segmental domain, the fact of value-identity could be referred to indirectly, via the fact that such sequences are single features with multiple linkages. The literature on the OCP makes it clear, though, that there is no guarantee that a multi-linked representation prevails over the identical-value sequence representation. The fact that a rule needs

to identify adjacent identical feature values is often simply taken to be sufficient proof that an identical-value sequence representation must be compressed into a multi-linked representation.

The standard autosegmental theory of rules of the form $X \rightarrow [\alpha F] / _ [\alpha F]$ is that they are spreading rules of the form:



It might appear that this provides a variable-free alternative to Value-Variable and Feature-Identity theories. However, that appearance is deceptive, and the autosegmental account simply uses a different notation for expressing what all of these theories can express. The three theories allow the following types of assimilation rules:

(12)	<i>Value-Variable</i> $X \rightarrow [+F] / _ [+F]$	<i>Feature-Identity</i> $X \rightarrow [+F] / _ [+F]$	<i>Autosegmental</i>
	$X \rightarrow [-F] / _ [-F]$	$X \rightarrow [-F] / _ [-F]$	
	$X \rightarrow [\alpha F] / _ [\alpha F]$	$X \rightarrow [=F] / _ [=F]$	

The autosegmental convention of not mentioning a feature value and inferring “whatever value” is nothing but a covert variable, saying exactly what Feature-Identity theory says (though not Value-Variable theory, where rules can have distinctive value variables).

As observed in Odden (1988) and Reiss (2003), there is more to the facts of identity reference than assimilatory spreading plus reference to multiple-linkage. A wide range of references to “value for F_i ” exists, for example syncope in Syrian Arabic is blocked when the surrounding consonants have (certain) same feature values, a phenomenon termed “antigemination” by McCarthy (1986). Likewise, there are “Anti-antigemination” rules such as Koya syncope, where vowels are deleted just in case they are flanked by homorganic consonants, which apply only when doing so creates a violation of the OCP. The proffered explanation for Antigemination is that the OCP prevents the rule from applying, when doing so would create a violation of the OCP. In the case of Anti-antigemination rules, output-violation of the OCP would be the crucial triggering factor. There are rules of Geminate Epenthesis, where a vowel is inserted between homorganic consonants, such as in Lithuanian. Such sequences cannot be identified as “homorganic” via the representational property of multiple linkage, since multiple linkage would prevent epenthesis in the first place – see Hayes (1985), Schein & Steriade (1986). The representational theory must apparently admit two formal mechanisms to derive identity-references: the fact of multiple linkage, and the fact that the OCP is violated, which is not a strictly representational property. It is clear that a purely representational account of identity references is untenable, since there is no representational property that signals “identical sequence” in Antigemination, Anti-Antigemination and Geminate-Epenthesis effects.

Since a pure representational account of identity is untenable, the question is whether a pure OCP-based account might work (if not, then a theory of feature variables would be necessitated). Numerous issues underlie an OCP-based analysis, such as what exactly the OCP says. In the formulation of McCarthy (1986), “at the melodic level, adjacent identical elements are prohibited”, but there is no straightforward interpretation of “element” and “identical” especially degree of identity required in this context, as discussed in Odden (1988). There is no general answer to the question of when this principle holds (it is clearly not universal for all languages, rules and representational objects). Yip (1988) pursues the idea

that all sorts of phonological effects can be attributed to the OCP, and in light of the current view that many putative universals are “violable” in a language-specific way, it would be reasonable to consider a unified OCP-based account, one which does not depend (solely, therefore at all) on a multi-linked representation. Simply saying that identity references can be referred to the OCP does not eliminate the need for some grammatical mechanism to express identity references: a complex mechanism attributed to UG is still a complex mechanism, and attributing a mechanism to UG does not immunize the proposal from the need for critical scrutiny.

An OCP-based account must be parametrized in terms of the representational units that it holds of (tone, voicing, place of articulation etc.), and in that respect there is no meaningful difference between a rule which mentions the feature properties of a feature or node (two features must not be the same), versus specifying a feature or node as the argument of the OCP and applying a rule on the basis of an OCP evaluation. There is a real difference between presuming a theory of blocking and triggering, along the lines of Yip (1988), versus a theory such as that of Reiss (2003) which employs explicit identity and non-identity conditions on standard rule formulations, but the issue at stake there is what the proper method is for formalizing “triggering” and “blocking” conditions, which is orthogonal to the question of how identity references are made in grammars. OCP conditions could be imposed on inputs or outputs, and could serve as blocking or triggering conditions, and we would derive the range of known identity effect. The Antigemination effect derives from blocking outputs that violate the OCP; the anti-Antigemination effect (Koya) where syncope only applies between homorganic consonants is where creating an OCP violation in the output is what triggers the rule; homorganic epenthesis and deletion derive from input triggering by the OCP (the rule only applies if the input violates the OCP); homorganic integrity, i.e. the failure of epenthesis in homorganic clusters is the case of input OCP blockage.²⁹

The relevant question is whether “[=F_i] ... [=F_i]” and “OCP(F_i)” are anything other than notational variants. Since the condition “OCP” is false (violated) on F_i iff [αF_i] is identical to [βF_i], and the requirement “[=F_i] ... [=F_i]” is true (satisfied) iff [αF_i] is identical to [βF_i], then “OCP(F_i)” is nothing more than the opposite of [=F_i]. Therefore, a choice between Feature-Identity variables and an OCP condition on a rule is a false dichotomy – they are the same thing. Needless to say, Feature-Identity/OCP theory *is* different from Value-Variable theory, since in the latter theory a condition may also compare the plus or minus values of different features, which cannot be done in a theory where the central concept is “sameness w.r.t. a given feature”.

6.3. *The independence of variables*

Analysis of alternative theories has led to the conclusion that there is one real issue at stake, namely whether variable behavior always involves comparison on a feature-to-feature basis, as claimed by Feature-Identity theory, or are there also rules of the type [αF_m] → X / ___ [αF_k] as claimed by Value-Variable theory. This is in part an empirical question, but a precursor to inspecting the facts is scrutinizing the logic behind establishing the existence of such a rule.

Chomsky & Halle (1968) raise the question whether variables should be formally restricted,³⁰ concluding on the basis of Viennese German that rules matching same value of different features do exist. They point to the fact that in Viennese German, the words in Standard German *vier* [fi:v] ‘four’ and *für* [fy:ɐ] ‘for’ are both pronounced [fi:r], and the words in Standard German *vieler* [fi:lə] ‘many’ and *fühler* [fy:lə] are both pronounced [fy:lə]. They postulate the following rule to account for these facts.

²⁹ This is not to imply that look-ahead output conditions are actually required to express the full range of identity conditions. Rather, the point is to use conceptual vocabulary that is familiar from discussions of the OCP, to make it clear that Identical-Value theory and OCP theory say the same thing, as long as the scope of the OCP is restricted in a manner befitting the facts.

³⁰ FP would say “emphatically not”, and would focus on the fact that a better theory of variables does not require any restriction.

$$(13) \quad \begin{bmatrix} + \text{voc} \\ - \text{cons} \\ - \text{back} \end{bmatrix} \rightarrow [\alpha\text{round}] / _ \begin{bmatrix} + \text{voc} \\ + \text{cons} \\ \alpha\text{lateral} \end{bmatrix}$$

To account for the fact that the six vowels in Tashkent Uzbek are either front unround or back round, rule (14) is posited.

$$(14) \quad \begin{bmatrix} + \text{voc} \\ - \text{cons} \\ \alpha\text{back} \end{bmatrix} \rightarrow [\alpha\text{round}]$$

Since the variable α appears on different features, no translation of either of these rules into Feature-Identity theory is possible

The question is whether it is justified to claim that these rules are in the grammars of these languages. In the case of Viennese German, data on the phenomenon is so limited that it is impossible to determine whether this is an active phonological process, or just a historical sound change. No evidence shows that the underlying forms of Viennese ‘four’ and ‘for’, or ‘many’ and ‘feel’ are distinct, and without such evidence, the claim that there is a rule at all cannot be accepted. We are not even given evidence that front round and non-round vowels contrast in this dialect. Without something stronger than historical relations between dialects, the claim for there being a synchronic rule is arbitrary.³¹

The correlation between backness and roundness in vowels, on the other hand, is known to exist in a number of languages, such as various Romance languages, virtually all Bantu languages, Basque, Czech and Modern Greek. Redundant correlations between phonetic properties in the phonemic inventories of languages exist, so that voiceless stops are aspirated and voiced ones are unaspirated in a number of Bantu languages; consonants are voiced if and only if they are sonorant in Cuzco Quechua, inter alia. There is no denying the factual generalization implied by (14), but the question that has to be answered is whether the generalization is a phonological one or a phonetic one. It is invalid to declare a priori that physical tongue retraction or lip protrusion in Uzbek or Spanish *entails* the feature specifications [+round] vs. [-round], or [+back] vs. [-back]. While [i,e] contrast with [u,o], only a single feature is needed to make that contrast. Additional phonological evidence is needed to support the claim that *both* rounding and backness are phonologically specified in these languages. See Dresher (2009) for discussion of how languages with this classical triangular vowel system can differ in the featural basis for the contrast. The burden rests on a proponent of the theory that both rounding and backness are phonologically specified in a language claimed to have a rule like (14).

An additional problem underlying the claim that (13) and (14) are rules in grammars is the lack of argument that the operations are formally subsumed under one rule. Suppose that a language were uncovered with a clear, active phonological process, where the value of a feature in a segment is the same as or opposite of another feature (in that segment or another). The argument must still be made that a single rule is at work. As Alan Prince observed at the 1989 MIT Conference on Feature and Underspecification Theories, if a language can have backness harmony and if a language can have roundness harmony, then a language can have backness and roundness harmony. Specific evidence is required to establish that the rounding of back vowels (backing of round vowels) and the unrounding of front vowels (fronting of non-round vowels) derive from a single rule. Arguments of that type can, in principle, be made – see Odden (1991) for arguments in support of the single-rule status of harmonies of vowel-height features, and of back and round, based on shared unlikely restrictions. The fact that Value-Variable theory allows two processes to be expressed as a single rule is not proof that the processes result from a single rule; the fact that Value-Variable theory allows two processes to be expressed as a single rule is therefore not proof that

³¹ Markus Pöchtrager and John Rennison inform me that there is no evidence for a rounding distinction, and that like Bavarian in general, front round vowels have simply been changed to unrounded vowels, but can be historically re-rounded before original */, which is often synchronically deleted or changed to [j].

FORMAL PHONOLOGY

Value-Variable theory is correct. An *independent* demonstration that (13) is a rule is what would show that Value-Variable theory is correct. Until that is done, Feature-Identity theory stands as the only theory of feature-variable behavior consistent with FP.

7. Summary

To make real progress in phonological theory, we must focus on what phonological theory is a theory of – it is a theory of grammatical computations. Then in order to construct a solid hierarchy of theoretical concepts that describes the nature of human phonological grammars, there must be a firm logical connection between observable facts and theoretical conclusion about the facts. Firm logical connections are established by setting high standards for justification of claims – postulation of arbitrary conjectures awaiting ‘testing’ does not constitute a valid method of theory construction. When the referents of phonological concepts are well-understood, it is possible to express generalizations about phonologies in very simple terms, where a concept translates into an unambiguous symbol, and we can sensibly discuss the *form* of phonological computations. Reaching this goal is the purpose of the theory Formal Phonology.

References

- Avery, Peter and Keren Rice. 1989. Segment structure and coronal underspecification. *Phonology* 6: 179-200.
- Bagemihl, Bruce. 1988. *Alternate phonologies and morphologies*. Doctoral dissertation, University of British Columbia, Vancouver.
- Bagemihl, Bruce. 1995. Language games and related areas. In J. A. Goldsmith, ed., *The handbook of phonological theory*, 697-712. Blackwell Publishers, Cambridge.
- Chomsky, Noam and Morris Halle. 1968. *The sound pattern of English*. Harper & Row, New York.
- Chomsky, Noam. 1965. *Aspects of the theory of syntax*. MIT Press, Cambridge.
- Clements, George N. 1985. The geometry of phonological features. *Phonology* 2. 223-50.
- Dresher, B. Elan. 2009. *The contrastive hierarchy in phonology*. Cambridge University Press, Cambridge.
- Hale, Mark and Charles Reiss. 2000. Substance abuse and dysfunctionality: current trends in phonology. *Linguistic inquiry* 31: 157-169.
- Hale, Mark and Charles Reiss. 2008. *The phonological enterprise*. Oxford University Press, Oxford.
- Hale, Mark. 2007. *Historical linguistics: theory and method*. Blackwell Publishing, Oxford.
- Hayes, Bruce. 1985. Iambic and trochaic rhythm in stress rules. In M. Niepokuj et al., eds., *Proceedings of the eleventh annual meeting of the Berkeley Linguistics Society*, 429-446.
- Hayes, Bruce. 1986. Inalterability in CV phonology. *Language* 62: 321-351.
- Hayes, Bruce. 1987. A revised parametric metrical theory. *Proceedings of the Northeastern Linguistics Society* 17: 274-289.
- Hayes, Bruce and Tanya Stivers. 1995. Postnasal voicing. Ms. University of California, Los Angeles. <http://www.linguistics.ucla.edu/people/hayes/phonet/ncphonet.pdf>.
- Hume, Elizabeth and Georgios Tserdanelis. 2002. Labial unmarkedness in Sri Lankan Portuguese Creole. *Phonology* 19: 441-458.
- Hume, Elizabeth. 1994. *Front vowels, coronal consonants and their interaction in nonlinear phonology*. Garland Publishing, New York.
- Inkelas, Sharon. 1994. The consequences of Optimization for Underspecification. *ROA* 40-1294
- Lombardi, Linda. 1991. *Laryngeal features and laryngeal neutralization*. Doctoral dissertation, University of Massachusetts, Amherst.
- McCarthy, John. 1986. OCP effects: gemination and antigemination. *Linguistic inquiry* 17: 207-263.
- McCawley, James. 1973. The role of notation in generative phonology. In M. Gross, M. Halle, and M.-P. Schutzenberger, eds., *The formal analysis of natural languages*, 51-62. Mouton, The Hague.

DAVID ODDEN

- Mellander, Evan. 2003. *A prosodic theory of prominence and rhythm*. Doctoral dissertation, McGill University, Montreal.
- Mester, Ralf-Armin and Junko Itô. 1989. Feature predictability and underspecification: palatal prosodies in Japanese mimetics. *Language* 65: 258-293.
- Al-Mozainy, Hamza. 1981. Vowel alternations in a Bedouin Hijazi Arabic dialect: abstractness and stress. Doctoral dissertation, University of Texas, Austin.
- Odden, David. 1988. Anti antigemination and the OCP. *Linguistic inquiry* 19: 451-475.
- Odden, David. 1991. Vowel geometry. *Phonology* 8: 261-289.
- Paradis, Carole and Jean-François Prunet. 1989. On coronal transparency. *Phonology* 6: 317-348.
- Reiss, Charles. 2003. Quantification in structural descriptions: attested and unattested patterns. *The linguistic review* 20: 305-338.
- Rice, Curt. 1992. *Binarity and ternarity in metrical theory: parametric extensions*. Doctoral dissertation, University of Texas, Austin.
- Rice, Curt. 2007. Gaps and repairs at the phonology-morphology interface. *Journal of Linguistics* 43: 197-221.
- Sagey, Elizabeth. 1986. *The representation of features and relations in non-linear phonology*. Doctoral dissertation, MIT, Cambridge.
- Schein, Barry and Donca Steriade. 1986. On geminates. *Linguistic inquiry* 17: 691-744.
- Steriade, Donca. 1987. Locality conditions and feature geometry. *Proceedings of the Northeastern Linguistics Society* 17: 595-617.
- Vaux, Bert. 2011. Language games and speech disguises. In John Goldsmith, Alan Yu and Jason Riggle, eds., *The handbook of phonological theory*, second edition, 722-750. Blackwell Publishing, Oxford.
- Yip, Moira. 1988. The Obligatory Contour Principle and phonological rules: a loss of identity. *Linguistic inquiry* 19: 65-100.