Phonological Ontology

(version 1, Sept. 17, 2021)

David Odden
University of Washington
Ohio State University

Prologue

This is the first general public version of a paper that became too long to be a paper. The intent is to incorporate suggestions for revision from readers, include references (sorely missing from this version), and to compose Part II, which is a more detailed example of theory comparison where I compare the present technical proposals (the theory of computation and representation) with competing theories. The primary challenge for Part II is devising explicit alternative theories, for example how exactly *does* this theory compare to, say, Dependency Phonology in terms of representations? What are alternative and comparable formal theories of computation? Not being a user of DP, I can't necessarily give a fair synopsis of that theory.

1. Introduction

Before getting into the details of this work, I should first say why I find it necessary to write on the topic (whatever that topic is – it's about methods of theorizing in phonology). The reason for starting with the motivation for the work is that knowing the purpose of the paper gives you a standard for judging the extent to which the work has succeeded or failed. More specifically, it was necessary for me to write this so that I could explicitly identify and answer some fundamental questions that I have had about the nature of reasoning and knowledge-creation throughout my career in linguistics. Having partially made these thought-processes objectively available (a consequence of *writing* about a matter, not just thinking about), I thought I would share those results with any colleagues who are interested.

The primary goal of the work is to aid reasoned theory development and comparison, a task which has become increasingly difficult due to the rapid expansion of knowledge of phonological theory and fact. This expansion of knowledge has been largely carried out in the context of an under-developed epistemological framework for forming and evaluating theories, so while it just takes a literature search to find out what people have said, it is very difficult to determine what we collectively *know*, or to even grasp the difference between "what people have said" and "knowledge". Essentially, it has become impossible to absorb (grasp, understand) all of the theoretical or empirical findings of generative phonology from the past 60 years, and even more impossible to also absorb non-generative or pre-generative findings. Even within a single research paradigm such as SPE theory, it is extremely difficult to control all of the technical details of the approach.

There is an explicable tendency towards theoretical insularity which follows from this expansion of phonological knowledge, where every scholar carves out a theoretical domain that they can handle and which they attempt to keep current on, as contrasted with areas that might be investigated if there is spare time. The enterprise of theory-comparison is generally conducted at a relatively low level internal to a broader theory, for example in OT, there can be comparisons of different OT sub-theories of assimilation such as Agree, Align, Copy, Correspondence, but there will not be a comparison of OT methods in general, as compared to SPE-theoretic treatments, Dependency Phonology models, or autosegmental accounts. Comparing a rule-based autosegmental account of a language to one in Dependency Phonology requires knowing two separate theoretical frameworks. Where would such knowledge come from? Comparison implies a metric of similarity and difference. In what way are autosegmental feature trees comparable to or different from dependency relations between elements? It is thus natural (unless compelled by reviewers) to avoid broad theory comparisons, focusing instead on lowest-level most-current and most-related theories of a topic.

One might think that the answer to the question "how can we compare theories" is that we look at the systems of theoretical postulates, and translate between theories given shared primitives. But first, there is no catalog of theoretical postulates for a theory – in any theory. Are the concepts of "ranking" in OT and "ordering" in rule-based phonology the same thing? They

are similar, but we don't know for certain what the properties of "ordering" are in rule-based phonology, and what the properties of "ranking" are in OT.[1] Some theoretical concepts are well-enough established that textbooks may indicate a typical range of uses for the concept – this is not the same as having a definitive collection of theoretical concepts. Second, phonologists frequently employ primitives which share no more than sameness of name. An infamous example of this sort is the concept "marked" (Odden 2016). What it means to be "marked" in Trubetzkovian phonology is not the same as it means in Jakobsonian phonology, Greenbergian phonology, SPE, autosegmental phonology or OT, to the point that the Trubetzkovian and OT concepts seem to only share a name.

A second problem related to theory comparison is that it is frequently very difficult to evaluate the arguments supporting a claim, for two primary reasons. First, arguments often refer to theory-internal reification and encapsulation of facts, using uncertain terms. The class "natural class" is frequently used in generative phonology, but what does it refer to? It is used to refer, for example, to the situation where [p, t, k] are identified in a phonological computation, to the exclusion of [b, d, g, f, s, x, m, n, ŋ, i, u, a]. Are there classes which are not natural, or natural phonological things that are not classes? What is the theoretical relevance of the fact that some number of languages computationally identify [p, t, k] to the exclusion of [b, d, g, f, s, x, m, n, ŋ, i, u, a]? Surely a theory which cannot account for natural class behavior in computations must be wrong – as long as we know what natural class behavior *is*. Knowing how people happen to have used a concept is not the same as knowing what a concept is, or is claimed to be.

Second, having a name for patterns of language fact does not *per se* constitute a valid argument for a theoretical conclusion. The fact well-known to phonologists that sounds group together in some way might enter into a valid argument for some conclusion about representations, provided that they are paired with an already-established set of theoretical conclusions and a logical analysis of how the present conclusion is compelled by these theoretical parts. Concepts like harmonic boundedness, TETU, checking, floating tone, spreading, opacity (of segments? of process-interaction?) or phonological phrase may be meaningful within a specific sub-theory of a theoretical framework, at a certain time, but it is generally non-trivial to make explicit the underlying system of facts, and the logic that connects those facts to a theoretical idea. What is the connection between ideas about language data, and technical claims that constitute a theory?

The reason for this problem arising is not that phonologists are bad or lazy people, it is because phonologists (perhaps linguists in general) are not in the habit of operating in an explicit epistemological framework, mainly because we are more interested in technical results than we are in the metatheory of how we get and evaluate theories. My proposal is that a bit more attention to certain epistemological principles is not an unconscionable burden, and if given, should result in great comparability of phonological ideas, and better ideas.

The fundamental question that science asks is "what exists? and what are the properties of that which exists?". The above issues are about existential claims of theories. "Ordering" is a claim that there exists a certain relationship between grammatical objects – rules. The claim depends on the concept "rule", which is another existential claim (rules exist). Within theories havingrule ordering, ordering relations exist between rules, but not between a rule and a "constraint" – we have another existential claim (there exists a certain property of rules versus constraints ex-

[1] Are rule orderings and constraint rankings formally "strict orders" in the mathematical sense? How do these concepts interact with the notion of cycle or stratum? What is the difference between "dominating" and "preceding", formally or substantively? Do notions of real time enter into these concepts?

ists, which has this consequence). It is not valid to claim without reason that some two rules have an ordering relation – the claim has to be justified by reference to facts of language. We need a reason to believe that an ordering relation exists between two particular rules of a language – for example in Logoori, Glide Formation precedes Palatalization. What facts exist in the language that justify the conclusion? What facts exist in human language that justify the theoretical concept "precede"? Do other concepts exist that could render unnecessary the concept "precede"?

1.1. Overview of the work

This subsection attempts to explain and (weakly) justify the structure of what follows. The main premises guiding the presentation are that knowledge is incremental, it arises by applying logic to observations, and it is about objective reality. A particular kind of knowledge is of interest to scientists, namely theoretical knowledge, a distinction (compared to knowledge of concrete facts) developed in section 1.2. Because this work is very meta, in order to explain the structure of the work, it is first necessary to surreptitiously introduce a theory of "theory".

It is undeniable that we have a lot of knowledge of phonology. I have claimed that this knowledge is not always optimally organized, presented, or justified, and I have claimed that we are not generally aware of the epistemological framework that we adopt in creating knowledge. Given that, I obviously should introduce such a framework, and should relate it to existing phonological knowledge. Since knowledge is incremental (what does that really *mean*?), my presentation should itself be incremental, that is, I should practice what I preach. Therefore, I start with certain issues which I judge to be "most basic", from a practical, cognitive perspective. As a case in point, it is only in section 5 that I explore what "data" are, and yet it would seem obvious that data are so important to science that I ought to start with a definition and theorization of "data". One reason why this would be a mistaken conclusion is that "data" is not an obscure concept that needs advance explication before phonologists can make any sense of a theory that explains how to apply logic to data to arrive at a theory. The section 5 discussion *refines* our understanding of data. Furthermore, in order to be able to detect and correct misunderstandings about the nature of data, we need a firm epistemological foundation, and since the refinements presuppose aspects of the model of phonology and epistemological framework, we cannot simply *assert* those claims in advance of that knowledge. We therefore can and must take for granted that there is a valid concept "data"; we should be aware that we may not know everything that needs to be known about data in order to talk about notions like "grammar", "representation" and "generalization", but we surely know enough about data that our implicit knowledge can serve as a basis for getting a better grasp of "grammar" and so on. In this work, until we are in a position to do better, we will adopt Naïve Data Theory: data are "the facts of language".

The end product of theorizing, according to this present model, is a hierarchical arrangement of ideas, where you can say (following the discussion in section 6) that the concepts mind, language, grammar, phonology and rule have a hierarchical relation (and so on), but it takes five prior sections to have the metatheoretical tools necessary for discovering and justifying that relation. Therefore, this work builds towards an end: it does not know in advance what the answer is and does not spell out that answer in a way that reflects either a top-down deductive subdivision of concepts or an inductive bottom-up unification of elementary facts. Indeed, the structure of the work is largely a result of starting with knowledge that seem to be fairly clear, then discovering and clarifying (or attempting to do so) details about that existing knowledge.

I start with a very brief overview of the epistmological framework that I adopt, then segue into a map of the six sections that constitute part 1. To satisfy the reader's possible curiosity, the intent for current nonexistent part 2 is to engage in detailed non-strawman comparisons of actual phonological theories. Now as I have said, facts and theoretical concepts do not exist in isolation, they are hierarchically related within a system – a theory. Given the title of the paper, you should expect that this paper will develop a preliminary ontology for phonology, which would be a statement of what metathoretical and theoretical things exist in phonology: an identification of the basic concepts required for phonology, and what they relate to. A parochial ontology of phonology might just be a list of claims particular to a certain theory, and there might be thousands of ontologies of phonology given the existence of thousands of theories of phonology. This work deliberately complicates the enterprise by assuming that there is *a* correct answer and not thousands of correct answers, because there is *a* reality and not thousands of realities. Thus theory justification plays a prominent role in this work.

In order to talk about the basic concepts of the correct theory of phonology, and in order to show why they are the right concepts, we have to *have* such a theory of phonology. This brings us to the secondary goal of the paper: to present an actual theory of phonology, founded on the epistemological principles that I am advocating. This theory of phonology is consistent with Formal Phonology (Odden 2013), henceforth FP, and implements the desideratum of FP that the arbiter between competing theories is simplicity – Occam's Razor. A simpler theory has fewer things than a more complex theory, therefore we must be able to figure out what things a theory has.

The overall character of the epistemological framework can be characterized as inductive, incremental, and bottom-up. That is, observations are integrated into simple abstractions, sets of abstractions are integrated into higher-level abstractions, all the while subjecting abstractions to scrutiny in the face of theoretical alternatives – scrutiny which pits observations against various concepts. This is in contrast to the top-down mode of reasoning, where a high-level abstraction is conjectured for some reason, then specific instances are deduced, where we test to see if those instances actually exist. The probative utility of deducing the existence of specific instances and comparing that to observation – searching for counterexamples – still exists in the inductive model, because a counterexample must be integrated with all of the previous observations that justified a theoretical conclusion. If it can't be done, the theory is wrong.

The subsection that follows this, 1.2, discusses what a theory is. In a nutshell, a theory is a compact encapsulation of hierarchical knowledge about a domain of things and their causal properties. This view contrasts with a sometimes-observed characteristic of phonological theories that a theory is an arbitrary collection of definitions and statements about those defined terms, and we accept the theory iff we find the statements to be true.

Section 2 is intended to spell out the epistemological framework in greater detail, so that the idea of a hierarchical arrangement of claims makes more sense (but still, a complete exegesis of "hierarchy" has to await section 6). Building on our previous groundwork about the nature of a theory, I briefly discussing the nature of an ontology as a system of existential claims: a claim that a certain thing or relation exists in the universe. I then discuss the practical problem of (re)structuring a theory in a bottom-up fashion, given that much of the foundation has already been laid, for better or worse. That is, I do not propose to completely erase existing phonological knowledge, I propose that we re-evaluate existing knowledge. As will be obvious, the proposed theoretical framework is, itself, constructed in a bottom-up fashion, starting with knowledge of a mature theory (rule-based autosegmental phonology), so we will identify the technical essentials

of the theory, then we analyze those essentials (necessarily, partially deconstructing the existing theory, to distill out the *essential* concepts).

Thus discussion of formalism *follows* discussion of theoretical essentials, because we cannot possibly formalize a theory until we *have* a theory. Historically speaking, formalism came first, and conceptualization of what the theory was about came last. But the point of *this* approach is not to repeat history, it is to conjecture where we would be if our methodological history had been different. Starting with the formalism and then presenting what the formalism is about inverts the bottom-up logic of theory formation. By the end of Part I, we will have apparently identified all of the essential theoretical details, but not in just one spot. The reader may wonder, why didn't you just present all of the essential claims at once? The reason is, knowledge is not created all at once, and the presentation here reflects how theory is constructed, at least following this particular cycle of focus-and-investigation in theory construction. A consequence of this "incremental" approach to theory is that we may tenatively accept premises because they have in fact been accepted and seem useful, but eventually we discover that a certain claim requires deeper scrutiny. A non-linguistic scientific example is the modern notion of "force", articulated in the 10th (Avicenna) and 14th-16th centuries, which supplanted various theories to the effect that bodies move toward their natural place according to the ratio of the elements which they are composed of, and elements reside on nested celestial spheres. Channeling Avicenna, it would be wrong to demand that the correct theory be intuited and presented as a *fait accompli* for reader approval, instead we seek a method of reasoning that preserves prior knowledge, and tells us why our new theory of the universe is better than the previous theory.

Starting in section 3, I move to specific technical analyses. The first step involves distinguishing concepts of phonology from other kinds of concepts, thus we must distinguish ideas about language data, human behavior, and syntactic or morphological structure. Phonological theory is interested in the theory of phonological computation and representation. The vast array of phonological entities in the literature notwithstanding, representations reduce to nothing more than nodes, which have types, and nodes have one of two relations with other nodes, dominance and precedence. Computation (a rule) is a kind of representation, which acts on representations: rules either insert or delete a representational object. Armed with the barest of theories, we consider how simplicity can be employed to decide between competing theories of a fact, specifically theories of what features are, and the nature of representational ordering. At the end of this section, we will have the theoretical essentials that constitute the core of autosegmental phonology, though not all of the peripheral claims which are less-well justified (such as elaborate theories of automatic repair, and none of the substance-dependent notions such as sonority, markedness).

Section 4 relates the concepts of section 3 to a system of notation – rule formalism. This exercise in formalizing is not purely for the sake of being formalistic. We uncover – and propose theoretical solutions for – subtle problems not previously recognized. This leads to a refinement of the theory of "dominance" (versus "precedence"), and to that aspect of rule theory that relates to "matching" between a rule and a string. We also consider how a rule can say "not", how "identical" can be expressed, and what it means to be "initial" or "final".

Section 5 investigates the logic of justification and theory evaluation. This section focuses on one of the central problems of induction, namely how you can prove a theory without stipulating the truth of the theory. The problem decomposes into two sub-issues: how do you prove, and how do you form an explanatory concept capable of non-circular empirical proof? To show how this is done, we model the ground-up creation of a theory of phonological rules and

representations by repeated theory-construction, centering on exposure to utterances from Damascene Arabic, starting with simple transcription, and culminating in the discovery of idealized generative phonology 1.0. An important key to successful prosecution of an inductive program of theory formation is that we can form valid phenomenological concepts that generalize over observations. "Final devoicing" is a fact of data patterns, devoid of causal interpretation. Based on independent phenomenological concepts which are outside of the theory of phonological grammar, we can then evaluate competing causal theories. Because we are able to independently talk about fact patterns, we can non-circularly evaluate how different theories fare as explanations for these fact patterns. This leads to a simple model of phonology, GP 1.0 which is a subset of SPE theory. We confront the theory with certain well-known facts which cannot be handled in the simplest theory – variable-length string requirements (parentheses and the like), and identity relations (alpha notation). This leads to certain minimalist expansions of the theory – which can be contrasted with the maximalist approach characterizing SPE phonology, namely "any mathematically-expressible relation".

Section 6 focuses on the hierarchical nature of knowledge, which is important in proper theory construction. Put simply, a theory should not re-invent the wheel, it should avail itself of the appropriate available conceptual tools, and a hierarchical theory of knowledge tells you what is available, and what is contributed by the theory. The theory of "rule" in phonology does not have to create an entire conceptual framework for talking about rules, it only needs to say what (if anything) is special about *phonological* rules, otherwise "rule" in the theory of phonological grammar is the same thing as it is in the theory of grammar. Again, throughout this paper, the goal is more to provide a model of how an explicit epistemology can guide theory formation, and is least about making specific theoretical claims. But, concrete theoretical claims are made.

1.2. What are theories?

We must first ask, what is a theory? The question has to be raised because misunderstanding is likely to lead to a mis-constructed ontology. A theory is a man-made intellectual product, not a natural artifact, so the apt question is, why do we construct theories? We do so because as humans, we need knowledge to be compact, yet open-ended. We need simple, understandable principles that encapsulate knowledge of past and future observations and states. A theory of a thing is a set of reasonably-well established propositions stating the essential properties of that thing. A theory has two basic aspects: the thing that the theory is about (the subject matter), and the content of the theory.

This approach to theorizing being existent-centric, a logically prior point to establish in advancing a theory is that a thing exists, and the secondary focus is refining knowledge of its nature. The subject matter of a theory is simply the name of the thing that we are talking about. A theory of gravity is about the thing known as "gravity", a theory of dogs is about the thing(s) known as "dogs", and the theory of rule iteration is about the thing (process) known as "rule iteration". It makes no sense to talk of a "theory of unicorns" if we have no reason to believe that unicorns exist, and it also makes no sense to talk of a "theory of karulization" if we don't know what "karulization" refers to. A basic and minimal definition of the subject-matter word or phrase is necessary, so that we can pin down what thing we are talking about. Sometimes, a definition is self-evident from ordinary language, for example "dog behavior" obviously refers to the behavior (qv.) of dogs (qv.). Dogs are defined ostensively. A more detailed definition may be necessary if a term is used in a specialized way in a field. "Feature" in phonology has a much

narrower definition than it does in ordinary English, as does "value". Some terms are (or originally were) only technical terms, for instance "electron" or "structure preservation".

Definitions are relevant for a theoretical concept only so that we can say which concept we are speaking of. The definition of a concept is not and should not be turned into the entire theory. A theory also has content, which is the system of concepts and propositions that state the essential properties of the thing, a system which allows us to understand its nature. By "property" I mean a fact about a thing's composition that determines what it does. This is not the same as a factual observation about a thing (or class of things), nor is it some observed correlation that is true of the thing. Android is the most popular OS for smartphones, but this is not a property of Android. Plutonium is used in reactors and for making nuclear weapons, but this is not a property of plutonium. A corollary of understanding the properties of a thing is that we can predict what the thing will do when the causal circumstances are present. *Essential* properties are those properties that characterize a thing, and which are not already implied by some other property. Being warm-blooded is a property of humans, but it is not an essential property of humans, since humans are mammals, and "mammal" implies "warm-blooded" (that is, warm-bloodedness of humans is a fact implied by a higher-level concept). Language is a symbolic system, but that is not an essential property of language, it is an essential property of cognition (it *is* essential that language is an aspect of cognition).

This view of theory can be contrasted with the view implicit in numerous works of generative linguistics and phonology in particular, such as embodied in Chomsky & Halle (1968). The appendix on formalism (pp. 390-399) presents a list of primitives and definitions intended to constitute a theory of phonological rules. As such, it is exemplary in that it attempts to be explicit. It would be distracting to undertake a critical analysis of their entire system of primitives and definitions *qua* theory at this moment (we will undertake a more detailed analysis of the concept "rule" in section 3), but it is worthwhile to point to some specific differences between that apparent epistemological system and the present system. Having listed the primitive types that appear in rules – *feature, specification, category* as well as "ø" and "→" – "rule" is defined with the aid of four further definitions:

(1) (a) *specified feature:* αX, where α is a specification and *X* is a feature.
(b) *unit*: $\alpha F_1\beta_1F_{i_1}\ldots\beta_mF_{i_m}$, where α = + or –, β_i is a specification, and $F_1F_{i_1}\ldots F_{i_m}$ is a substring of $F_1\ldots F_q$
(c) *matrix:* $X_1\ldots X_m$, where X_i is a unit.
(d) ***rule****:* ZXAYW → ZXBYW, where *A* and *B* may be ø or any unit; A ≠ B; X and Y may be matrices; Z or W may be C_i, for some *i;* Z, *X, Y,* W may be null; and where these are the only possibilities.

This definition of rule does much more than say what we mean by rule (as contrasted with constraint, ordering, feature or any other phonological concept), it also states numerous putative properties of rules. The condition "A≠B" is an empirical claim about what rules may or may not do – it says that the expression identifying the target cannot be the same as the expression stating the change.[2] The condition "X and Y may be matrices" embodies the empirical claim that in a

[2] Actually, this isn't even a hypothesis, it is an automatic consequence of any reasonable model of learning. The rationale for the condition is that such a rule would do nothing. Since rules are learned based on their effect on input-output relations, if a rule has no effect, it cannot be learned. Such a rule is computable, but it not attestable. In-

rule, the triggering context can be a *sequence* of segments and boundaries ("units"). Whether or not this is a true fact about the triggering context in rules is, again, an empirical hypothesis about the properties of rules, and is not properly part of the explanation "what we mean when we say rule". Supposed "knowledge" has been claimed not because a conclusion is well-supported by facts (discovering from observation that the conditioning environment can be more than a single segment), instead it comes by stipulation. In the present approach, such statements would, at most, be claims about the properties of a thing, such as "rule", and not part of a definition.

Saying that a theory is a *system* of propositions means that the propositions have some logical relation to each other. The ontology of a theory says what exists in the domain of the theory, and includes the logical structure of the system, stating each concept, the identifying definitions of the things, and the concepts and facts that the definitions depend on. To take a simple example (in one well-known theory), an utterance is composed of a sequence of segments. Segments are composed of a set of feature specifications, and feature specifications are pairings of values and attributes. The concepts *utterance*, *segment*, *feature specification*, *value* and *attribute* are hierarchically related.[3] Utterances can be understood in terms of the properties of segments (and possibly other concepts), and segments can in turn be understood in terms of the properties of feature specifications, and so on. Utterances are *not* understood in terms of features or rule ordering, and features are *not* understood in terms of utterances.

The concept *feature specification* is to be comprehended by reference to the concepts *attribute* and *value*, for example "voicing specification" invokes a specific attribute (voicing – an element which a feature theory may recognize), plus whatever properties voicing can have. We do not need to have a complete theory of *value* in order to be able to make valid claims about *feature specifications*, and ideally all claims about *feature specifications* will be stated with such generality that new knowledge about *values* does not disrupt the theory of rules. That is, we will not start by defining "rule" in a way that crucially depends on *feature specifications* having just the values '+' and '–'. Perhaps the answer is that "*values* are '+' or '–'", but equally possible the answer is that "*values* are 'present' or 'not present'". In the minimalist metatheory of theory-building that constitutes FP, generalizations are stated once, in the most general and widely-applicable fashion and context. If a concept "rule" is valid for all aspects of grammar, that concept is stated just once, as a concept of grammatical theory. It will not specify the aspects of "rule" that are only valid for phonological grammar. The theory of phonology further articulates "rule" as applicable to phonology. The theory of phonology thus elaborates the broader theory of grammar, by saying what distinguishes a phonological rule from a morphological, phonetic or syntactic rule. If we mess up our theory of phonological rule, the mess is modularly limited to phonology, and does not require overhauling the theory of rule in syntax or morphology. We only correct the error that relates to how we expanded the concept "rule" for phonology.

To aid the reader's comprehension of the points that follow, I sketch here what kind of structured information constitutes an ontology of phonology. An ontology can be reduced to a set of concepts within a domain ("rule", "feature" etc.), and a set of propositions about those concepts that state the properties of the concept. Concept A may presuppose concept B, which is an important hierarchical relationship between concepts ("ordering" presupposes "rule", because

---

deed, the condition doesn't even achieve the goal, given that a rule "[+nasal,+voice] → [+voice]" satisfies the requirement A≠B but still perform no change, and cannot be learned.

[3] This is not to imply that every theoretical concept of phonology can be arranged into a strict proper-inclusion hierarchy. But hierarchical organization is an important fact of the structure of a theory.

what is ordered is a collection of rules). This structured information constitutes the technical theory of phonology, but there is another dimension to a theory, namely the evidence for it.

For any concept or proposition about a concept in the theory, there are (or should be) articulable alternatives, thus in talking about "features" one might adopt a universal inventory of two-valued features, or a learned collection of unvalued properties. In asserting (as I shall do) the proposition about features that they are a learned collection of unvalued "types", I need to say what the alternatives are. And, I need to say what the evidence is that supports the proposed concept or proposition. That evidence refers either to metatheoretical premises (the appeal to simplicity; or a competing premise that linguistic theory should minimize computational complexity in the computer science sense), or to conclusions about language facts. Such evidence itself usually takes the form of phenomenological concepts and propositions, thus "inalterability" is a phenomenon observed in a number of languages. It is a concept to which a number of propositions can be attached, and "inalterability" conveniently refers to that body of evidence that can be used to support certain theories of grammar.

1.3. Assumptions

This is not an introduction to phonology: it presupposes an introduction to phonology. It assumes that you know about phonological grammars and their analysis at least to the level of most of Odden (2013) e.g. you understand what a rule is, what ordering is, what an underlying form is. It also assumes, at my greater peril, that you know basic autosegmental phonology, as exemplified by Kenstowicz (1994). If you have no idea what a "floating feature" is or what "assimilation" is, you might be well served by getting that background. If you understand the idea of a rule or floating feature but you are skeptical that the ideas are correct, you should just follow along and get more engaged in theory comparison in part II.

2. The epistemology of phonological theorizing

The point of this subsection is (a) to briefly introduce the problem of fundamentality, (b) to explore the logic of justification in theorizing, and (c) to indicate how an ontological framework might reasonably be erected in mid-discussion (60 years into research on generative phonology).

2.1. Fundamentality

Fundamentality relates to the hierarchicality of concepts, which is essential to a theory. A good theory is built on a solid foundation – fundaments. Therefore, one needs to identify the fundamentals. To take a simple (and deliberately misleading) example, we must ask which is more fundamental: feature, segment, or syllable? The question is misleading because fundamentality is determined by the context of the concept. Features are fundamental, from the perspective of the nature of the universe as regards grammatical computations and representations. Segments are the atoms of phonology, and features are the particles that make up those atoms. Syllables are the molecules that can be constructed from atoms. On the other hand, segments are epistemologically more fundamental – their existence is most self-evident. The evidence for features is very strong, but no argument for features can be advanced without presupposing the segment, whereas we have had theories of segments that did not presuppose the existence of features. Likewise, the evidence for the syllable is good enough, but syllables cannot be posited without presupposing the segment (a syllable *is* a unit that organizes… segments). On the third hand, syllables might be fundamental from the perspective of how people produce or perceive language – if so, that would be fundamentality from a phonetic or some non-grammatical perspective. We are concerned here with phonological theory.

Theorists are generally concerned with the nature of the universe, therefore our greatest interest is things that are existentially fundamental – the subatomic particles, or features. Existential fundamentals are not generally immediately available to inspection, instead we may conclude that they exist, given their explanatory value regarding the things that we *can* directly inspect (the epistemological fundamentals). We accept claims of existential fundamentals which are not epistemologically fundamental, when the existential fundamental explains the epistemological fundamentals, that is, the known facts.

2.2. Justification

The justification for the concepts and propositions of phonological theory are not themselves part of the theory of grammar. Still, theoretical claims are not to be arbitrarily stipulated in a rational

scientific framework. All claims require sufficient justification. Justification is part of the epistemology of phonology, constituting a large "appendix" to the theory. For example, positing the concept "feature" is justified by how it allows an account of segment-patterning in rules, with respect to things that undergo or trigger rules, or how segments change. In some languages, /b,d,g/ do similar things because they have a "feature" in common. In proposing or accepting a claim, we should be able to articulate the justification for the claim. The reader *should* be able to put together some set of facts of phonological grammar that justify the claim that segments are made up of features. A complete analysis of how to fully justify a phonological claim is beyond the scope of this paper, and unnecessary, since the fundamental principles can be easily summarized. We scrutinize the logic of justification in more detail in §5.

Justification refers to the strength of evidence for a proposition. If there is no evidence for a proposition, it is arbitrary. For example the out-of-the-blue utterance "this coin will land heads-up on the 37th flipping" is devoid of evidence. There is no rational basis for making that claim, but it does describe a definite situation, so it does refer to an identifiable state.[4] If there is just a little evidence, the proposition is "possible", for example "Charles has a chihuahua".[5] If there is a fair amount of evidence, it is "probable". A proposition can be so well established that it would be irrational to deny it – "my car is yellow" – or so well-established to be "what is not" that it would be irrational to maintain it – "my dog is a chihuahua". Such conclusions are "certain". Extreme precision in the estimation of evidentiary strength is pointless. The cut is (ideally) made in the middle, and corresponds to the question "is it more likely than not that A is true?". Practically speaking, arbitrary claims and certainty claims play little role in phonological theory, at least with a broad theory of what counts as evidence. I cannot say that I have ever seen a truly arbitrary claim posited in phonology, except as a straw-man. It is more difficult to identify a candidate for an unquestionably certain proposition of phonology. The subject is further explored in section 5.

2.3. Starting in the middle of the discussion

"Ideally" (fictionally), scientific research into a domain starts with a clear understanding of the right methodology of theory-construction. We would "ideally" start by identifying the most self-evident facts about language, positing the fewest, simplest necessary concepts and propositions within that domain, then we would further elaborate those statements as we learn more about language. This idealized progression is unrealistic, because we cannot un-know what we have learned about language over the past few millennia, and we cannot start with a fully-articulated metatheory of theory-construction without first constructing good and bad theories, which thereby allows us to discover what distinguishes good theories from bad theories. Still, it is useful to engage in Galilean idealization to understand what we *should* be doing in theory construction. The reader's indulgence is begged, as I attempt, in this subsection, to re-discover some fundamental concepts and propositions leading to the theory of phonology. Especially since any kind of human knowledge presupposes language (the phenomenon which underlies *grammar*), it is somewhat ludicrous to pretend to scientifically "discover" the concept language, yet if we can't establish that "language" is a real thing, it is even harder to establish that some theory of

[4] In case you were thinking "there *could* be some evidence, if we develop a chronoscope that allows us to see future events", I don't entertain fantasy scenarios in this work, not even fantasy constructed by Robert L. Forward.

[5] It is not entirely germane who Charles is, but I know from testimony that Charles has or did have a dog.

language is real. The point of this exercise is to illustrate how essential concepts and relationships between concepts can be discovered, and how knowledge can be created incrementally.

Two basic and undeniable facts about human behavior which lead to the concept *language* are the fact that humans "talk", and the fact that knowledge is encoded and exchanged via this talking behavior. After studying this talking and encoding of knowledge, we conclude that humans have a mental tool, "language", which encodes knowledge. Our initial cognitive theory defines *language* as the cognitive system which is unique to humans and which encodes knowledge. One of the distinctive properties of language is that it is manifested through *speech*.[6]

Given the existence of *language* as a cognitive faculty, our investigation leads to the question "what is the nature of language"? We do not start by defining language as a mapping between matrices of binary phonological feature and some kind of semantic representation that is organized into syntactic substructures, because we have yet to establish that any of these concepts are valid. In the realm of matters that are ultimately relevant to phonology, we can observe that language is manifested as stretches of (sound-)output, resulting in the concept of "utterances" (people say things). Further study reveals that some utterances refer to specific existents (entities, actions, attributes): utterances may be "words". Initially, we observe that |beana|[7] is used to refer to "dog", and |gulan| refers to the situation "I hear". We also observe that |beaʔnaga| is also used to refer to "dog", and |gullaa| refers to the situation where "he/she/it hears" – we initially do not understand what the difference between |beana| and |beaʔnaga| is, we simply know that people can say these things.[8]

We also observe that some utterances are composed of sequences of words: there are "sentences" in language, so we have observed |beana gullaa| being used to refer to the fact that a dog hears something, and |gulan beaʔnaga| is used to refer to me hearing a dog. Deeper investigation into utterances tells us that *sentences* are permutations of one or more words. Not every imaginable permutation of the words of the language constitute sentences of the language, thus *|beana gulan| is not a sentence of the language.[9] This leads to an important pre-theoretical mathematically-expressed factual proposition about the nature of language: *the set of utterances allowed in a language is a proper subset of the set of combinatorially-possible permutations of entities in the language*. Based on observation of the word-sequences which are allowed versus those which are excluded, we conclude that there must be a cognitive system which identifies the allowed arrangements of words (the distribution is not random, there is a pattern). In other words, certain valid propositions about the nature of *language* give evidence for the concept of *grammar*, which is a system that generates the language. The evidentiary conclusion, that the set of allowed outputs is a proper subset of the set of possible permutations of language entities (the general phenomenon of sentence-patterning), is justification for the concept *grammar*. *Grammar* is the alternative to the null hypothesis that every arrangement of language entities is allowed.

---

[6] Eventually, we would refine this observation to not be specific to a modality. What matters is that there is a systematic mind-external realization of language. *Speech* is relevant as the antecedent concept that leads to *phonology*.

[7] This refers to the unanalyzed sound sequence produced when a speaker utters a word, and does not presuppose segmentation or the concept "segment". Vertical bars simple denote the output of a speaker. Illustrative examples come from North Saami.

[8] Later we discover that |beana| is the form used when the word is the "subject" in a sentence and |beaʔnaga| is used when the word is the "object".

[9] Prior to investigating how *sentences* are constructed, we will have arrived at a method of relating particular sequences of words to the behavior of speakers, and understand that some sequences are "not accepted".

Further investigation into the concept of *grammar* gives evidence that a grammar is composed of a set of *rules* which "do" particular things. One prominent feature of what rules do is that they combine classes of words in a certain order, so that e.g. "A B D E" is possible in some language but *"A E B D" is not.

We must, finally, narrow the concept of grammar to the thing that phonologists are interested in. Having previously analyzed linguistic behavior so that we can independently focus on relations between words in utterances (syntax), the interpretation of utterances (semantics) and the manifestation of utterances (phonology), we realize that what a phonology does is create some kind of output that is (ultimately) perceivable, and we want to understand how it is that the product of syntax and semantics results in a perceivable utterance. We almost certainly will discover that to answer this question, we must assume at least one other aspect of grammar, namely "morphology", though we may also retract that conclusion at some point. So we take "what people say" to be the thing that phonology produces.

We have now arrived at a hierarchy of theoretical concepts pertaining to humans. At the highest level we have *cognition*, one aspect of which is *language*. An essential property of *language* is that it is governed by a mechanism known as a *grammar*, and an important concept in the theory of grammar is *rule*. As phonologists, we are specifically concerned with a particular sub-component of grammar. We leave open, for the moment, what exactly a "rule" is, or what other concepts (such as "representation" or "ordering") are necessary for undertanding *grammar*. In leaving open details of the nature of the concept "rule", we also raise many questions about exactly what a rule is or does. Indeed, it is highly probable that the properties of "rule" in phonology are different from the properties of "rule" in syntax, and we want to leave the concept underspecified at the level of general linguistic theory. Even though "rule" in an existential fundamental in the theory of language, it is not a defect that we do not at the moment have a full account of what a "rule" is or does, because that is something that we can specify once we have more knowledge.

3. Basic concepts of phonology

How can we identify the fundamental concepts of phonological theory, along with their essential definitions, and indicate what their justification is? We start by identifying the foundational concepts that phonological theory presupposes, which defines the scope of the theory. When we make any claim about "phonology", what concepts do we presuppose in making those claims?

3.1. Pre-phonological concepts

Most obviously, we presuppose (inherit from prior investigation into language) the concept *grammar*, since "phonological component" is an elaboration of what a "grammar" is. A grammar has *components* (distinctive parts), specifically phonological; morphological and syntactic; and phonetic components.[10] What is most important to phonology about morphology and syntax is that these components create the input to phonology, the *underlying form*. The output of phonology constitutes the input to the phonetic component. Although questions about exactly what actions the morphological or phonetic components perform are not the concern of phonology, having a basis for answering such questions can be important, since we need to first delimit which computations constitute phonological computations (not phonetic or morphological computations). Typically, we do not consider word order to be a matter of phonological computation. Sometimes, the boundaries between components are unclear. When we can identify such unclarities, we have identified an important theoretical lacuna that needs to be filled.

In Semitic languages, certain derivational and inflectional categories are indicated by choice of stem vowel or syllabic pattern (*a* versus *u* for passive, CVCVC versus CVVCVC or CVCCVC), and in many languages, word forms are indicated by inserting a segment (sequence) inside of another morpheme – infixation. A number of affixes in English have a three-way realization scheme with [s] appearing in context A, [z] in context B, and [ɪz] appearing in context C. Is this "allomorphy" (a morphological computation), or is this the product of phonological rules? Are these variants due to phonological computations, or morphological computations? The same question can be asked about alternations like *sing, sang; ring, rang*. On what basis do we decide the component-disposition of such phenomena? In English, the lateral /l/ becomes partially devoiced after voiceless fricatives and syllable-initial voiceless stops (e.g. *play, slip*): is this a phonological computation, or a phonetic one? Likewise, /t/ and /d/ both become the flap [ɾ] in a certain phonologically-defined context – is this a phonological or a phonetic computation? What is the theoretical basis for saying that this is phonology, or morphology, or phonetics?

The answer to such questions has to be based on a theory of phonetic, phonological and morphological computations (also, a theory of how morphology interfaces with phonology, and how phonology interfaces with phonetics). If phonetic computations can absolutely neutralize

[10] Originally, there was only syntax and phonology: a bit of compression of history of the field is necessary.

underlying distinctions, the flapping rule could be phonetic; if phonological rules cannot create segments that are not underlyingly present and if we know that /ɾ/ does not exist in underlying representations,[11] we would know that flapping cannot be phonological – but otherwise, it could be. The proper way to frame the question is not in terms of phenomena, it is in terms of whether a theoretical contradiction arises from proposing a grammar where flapping, or l-devoicing, are treated as phonological rules, versus as phonetic rules. What well-supported premise of phonological theory must be abandoned if we treat flapping as phonological? What well-supported premise of phonetic theory must be abandoned if we treat flapping as phonetic? The problem (which should be obvious) is that we need to be able to model flapping as both a phonological computation and as a phonetic computation (we must be in a position to state what the alternative analyses are), and while phonologists can probably do the former ("intervocalic syllable-final alveolar stops become [+sonorant]", supplemented with various assumptions about syllabification), it is unlikely that phonologists can do the latter – what *is* the theory of phonetic computation, how do you model flapping as a phonetic computation? Would it be a problem for the theory of phonetic computation if distinct inputs to phonetics such as [ˈɹaɪd̬ɹ̩] 'rider' and [ˈɹaɪt̬ɹ̩] 'writer' always have the same physical outputs; or if they sometimes have the same outputs?

In the case of *l*-devoicing and anticipatory rounding in English, we can identify a theoretical argument that these are from phonetic computations. What points to this conclusion is that the process in question continuously implements a physical property over time. In the case of anticipatory rounding (observable in '**s**oup, **st**upid, e**xcl**ude'), lip protrusion gradiently increases within the consonant sequence up to the triggering round vowel, and *s* is not uniformly changed from "nonrounded" to "rounded" in a crisp segmentally-defined context. Instead, lip protrusion increases gradually within the stretch of time comprising the phonological outputs [su], [stu], [ksklu]. This fact fundamentally contradicts the nature of phonological computations, which change segments from one category to another. A phonological computation would change /ksklu/ to [kʷsʷkʷlʷu], missing the gradient nature of the process. The exact phonetic computation is not clear, but would most likely say something like "the onset of the lip protrusion gesture may commence after the preceding vocalic gesture", which has gradient rounding as a consequence, thus anticipatory rounding is comprehensible as a phonetic process but not as a phonological process. In order to render anticipatory rounding as a phonological process, the theory of phonological computation would have to be expanded in some way (admitting fractional values of features and a time-continuum within segments), and without compelling evidence for such an expansion, the logical of theoretical minimalism says that this is not a phonological process.

Two final fundamental pre-theoretical concept of grammar are *representation* and *computation*. What a computation does is relate one representation to another representation. The concepts of representation and computation are not specific to phonology – morphology, syntax, semantics and phonetics also have computations and representations (and there are non-linguistic mental representations and computations), but the precise nature of representations and computations in these components differs to some degree. *Computation* logically presupposes *representation*, that is, there are no computations that are not computations of representations. In fact, it is not clear that there is a valid concept *representation* or *computation* which is specific to language, and which is distinct from the general cognitive notion of representation or computation.

[11] This is not something we "know", but some lines of reasoning at least make that a possible conclusion.

In lieu of evidence to the contrary, phonological theory will elaborate the broader cognitive notions of *representation* and *computation* into **phonological** representations and computations.[12]

The broadest pre-phonological concepts which phonology depends on are *cognition, representation* and *computation*, which are so broad that nothing of technical significance carries over from the theory of grammar, other than the fact that all components of grammar perform computations on representations. The technical concept that "implements" language is *grammar*,[13] which has certain components including a *phonological component*. Although earlier versions of generative grammar assumed significant formal similarity between the components of grammar (there was some overlap in the formalism for rules and the theory of features), it is not clear that phonology has anything in common with other grammatical components, especially in light of the axiomatic premise of modularity theory that components have completely non-overlapping representational vocabulary. Within the theory of the phonological component, we must specify the nature of phonological representation and phonological computation. This then constitutes the enterprise of initializing a theory of phonology. Phonological theory depends on the following concepts of linguistic theory:

(2) grammar
    component
        phonological component
        (morphology, syntax, phonetics…)
    representation
        underlying form
    computation
        rule

Indentation suggests a hierarchy of conceptual relations, but actually this is not a correct organization of concepts – we will sort these concepts properly in the course of modeling the systematic discovery of the concepts of phonological theory, their properties, and their justification.

I have glossed over an important detail regarding theory construction, namely the role of "facts". Exactly what justifies having a theory of language or grammar in the first place? Well, I hinted at that justification in 2.3 by referring to the fact that some utterances identify specific entities and actions, and there is a complex set of rules governing how utterances like |beana|, |beaʔnaga|, |gulaa| and |gullan| can be used. Somehow, behavior of North Saami speakers is encoded in such a way that we can say that |beana| means 'dog' and can be used in certain syntactic positions; |beaʔnaga| means 'dog' but is used in different syntactic positions, and so on. As long as you believe that these forms are trustworthy representations of speaker behavior (I promise, they are), these forms have a special status in linguistic theorizing: they are *data*. Phonological theory is held responsible for accounting for the data. If a theory cannot account for the data, the theory simply is not correct. For the moment, we will assume a simple model of what data is, and

[12] This is an invitation to discover commonalities in all linguistic computations and representations. I suspect that a division can be motivated between symbolic systems and perceptual systems: the latter serve as interface from symbolic systems and sensory organs. Candidate examples of the former kinds of systems include language, music and reasoning. Needless to say, nonlinguistic forms of cognition will be ignored here. Since I don't know of any reason to posit a unified language-specific notion of representation or computation, I don't posit any such notions.

[13] I admit that I am ignoring the lexicon, since I don't understand what the lexicon is, or whether it "does" anything.

will look deeper into the nature of data in section 5. Data result when facts are observed (speakers utter things), and are rendered in objective, inspectable form. A theory is responsible for explaining the nature of the data.

3.2. Concepts of phonological representation

In developing a theory of phonological representations, we must first identify the purported entities that the theory claims to exist, and second, the properties that are attributed to those entities.

3.2.1. PHONOLOGICAL ENTITIES

What is the "most basic" unit of phonological representation? Here are some candidates.

(3) Utterance
Word
Foot
Syllable
Segment
Feature

"Utterance" is the least-analyzed form of linguistic behavior that corresponds to a representation. But there is virtually no reason to think that "utterance" is even a concept of phonological grammar. Utterance is to phonology as Universe is to physics or philosophy: it's all that exists, and there doesn't exist anything that is outside of Utterance which behaves differently from things that are inside Utterance. Maybe there is a higher cognitive unit that spans utterances, but for language, the utterance is everything. Like Universe, the concept does not usefully partition existence into things that are, versus things that aren't. Just because you can define a technical concept "utterance" does not mean that we need it in grammar. "Utterance" may be a valid concept of a theory of human behavior, but it is not necessarily a phonological one, just as "hue" is not a valid phonological concept, even though it is a valid cognitive concept.

"Word" is a relatively useful linguistic concept, but it is primarily a morphosyntactic concept. The word "apple" is not a single atomic sound unit which we could represent with a single symbol [蘋果].[14] Ontologically speaking, "word" *qua* phonological unit is clearly derivative, and is not the best-motivated phonological concept on the market (even though there is good enough evidence for its existence). At this point, we will not even claim that word is a phonological concept, but perhaps we can eventually find evidence for a phonological concept ω. Words, as phenomena of language behavior, are built out of lower-level units from different components (syllables, morphemes, segments, stems). It is only in terms of perceptual self-evidence that words are "basic", that is, the word is one of the most directly-observable forms of linguistic behavior. Since our interest is in phonological theory and not speech behavior, we bypass word and leave for later the question of how "word" in the morphological sense is relevant to phonology.

One of the two most credible candidates for representational primitive in phonology is "segment" – the other is "feature". Using standard conventions for symbolically writing segments, we can say that "apple" is composed of 3 segments: [æpl]. The segment [æ] in "apple" is

[14] Despite appearances, you should treat this as a single sound symbol, and not two or three symbols in a row.

the same in type as the [æ] in "axle" and "bash": [æksl̩], [bæʃ]. If we assume (based on speech-behavior phenomena[15] such as slow speech, as well as some facts of phonological computation) that "syllable" is a valid concept, the concept "segment" provides an enlightening account of the concept "syllable" (which itself needs further motivation). English has about 16,000 possible syllables, which is a huge number of primitives if syllables are primitive. But this can be reduced to a much smaller inventory, if syllables are described as a combinatorial function of a much smaller set of about 40 segments that can be permuted in specific ways. Major generalizations can be captured by stating the inventory of syllables in terms of patterns of segments. In any syllable, the segment [ŋ] is preceded by one of the segments [i ɪ e ɛ æ ɔ o ʊ u ʌ]. The segment [h] is always followed by one of the segments [i ɪ e ɛ æ ɔ o ʊ u ʌ].[16]

It is hard to imagine how any phonological generalization could be stated without presupposing the concept of segment, and at least in the context of a scientific theory of language sound, "segment" is a self-evident[17] concept. That does not mean that it is existentially fundamental, nor does it mean that we always know which thing in phonological theory corresponds to "segment". It simply means that we cannot reasonably question that there is such a thing. The question that we now face is, can we hold that individual segments are the most basic objects of phonological representation, or are segments composed of other *more* basic things, just as in the physical world, actual atoms are not atomic, they are composed of other smaller things? As is now well-known, segments are built up, in some way, from *features*. The justification for features comes from the nature of phonological computations. If we only have unchanging representational strings like [ɪt, æt, ɔt] or [hɪp, hɪt, hɪtʃ, hɪk, hɪd, hɪs, hɪz, hɪm, hɪl], which are distinct lexical items distinguished by the occurrence of [ɪ æ ɔ] or [p t tʃ k d s z m l], we would have no grammatical evidence for further analysis of segments into smaller entities. We can develop a theory of representation and a theory of computation in parallel. We have already seen evidence for something more basic than segments in the discussion of segments: [ŋ] is always preceded by a vowel (a concept that subsumes [i ɪ e ɛ æ ɔ o ʊ u ʌ]), and [h] is always followed by a vowel. The justification for features comes from how they allow a coherent analysis of segment-set behavior in the computational system (so-called "natural classes"). A full reference work on phonological theory would point to much more specific evidence that *feature* is a valid phonological concept (though we will change the terminology once we investigate the nature of "feature").

We should continue the search for the lowest-level phonological units, asking whether features themselves are made up of something else. In earlier versions of generative phonology, it was assumed that features are composed of *attributes* and *values*.[18] The privative theory of features does not include the latter two concepts. Under the simplicity premise of FP, the privative theory of features thus enjoys logical pride of place ahead of binary feature theory (but we will leave open for the moment the question of which theory is in fact correct).

---

[15] Normally, I abjure appeal to speech behavior as evidence for anything in phonological theory (speech behavior is evidence for a theory of speech behavior, and not all speech behavior is about phonology, so raw speech behavior is irrelevant to phonological theory). Our current goal is to motivate – or, systematize the motivation for – the concept "phonological component" and the related concept "phonological computation", so such facts can be reasonably considered as evidence that there is a phonology.

[16] The reader will no doubt have noticed that a generalization is being missed here.

[17] Axiomatic.

[18] Terminology may vary inconsequentially. In SPE, "[+coronal]" is a "specified feature", "coronal" is a "feature", and "+" is a "specification".

At this point, we have at least described a portion of the SPE theory of representations: a segment is some conjunction of features, seen as attribute-value pairs. It is not the intent of this work to forget about research in phonology from the mid 70's onwards, so we will also take for granted the correctness of some version of autosegmental phonology, where segments are *not* just a conjunction of features. Instead, features have some kind of hierarchical relation to segments (a single feature may be a property of some number of segments; features relate to each other within a segment in some hierarchical fashion). Thus the primitives of representation often called "features" are not just properties like "anterior", "voice" and so on which are terminal nodes in a representational tree, they also include grouping properties such as "place", "laryngeal", and "coronal" the latter usually being thought of as being a featural property, one which is not necessarily terminal (it dominates "anterior" and "distributed").

Again, while it is not the purpose of this work to recite the modern history of phonology or to teach Phonology 101, it is useful to consider what kind of facts justify adding something to the theory besides features. Why do we want to be able to talk about groupings of features? A basic fact about phonological grammars is that many languages have rules of assimilation, where some segment becomes the same as some other segment, with respect to a specific set of features. In Matumbi, a nasal (which might be underlyingly *ɲ, n* or *m*) takes on just those features of the following consonant which define "place of articulation", thus {anterior, coronal, back}. This can be seen in (4), where the plural prefix /ɲ/ changes to [ŋ] before a velar, [m] before a labial, and to [n] before an alveolar.

(4)	Singular	Plural	
	lw-ɪ́ímo	ɲ-ímo	'land being weeded'
	lu-golóká	ŋ-golóká	'straight'
	lu-báu	m-báu	'rib'
	lu-dʒiíŋgjá	ɲ-dʒiíŋgjá	'entered'
	lu-laála	n-daála	'pepper'

The claim of autosegmental theory is that there is not a set of separate assimilations at work here, each targetting a particular collection of place-of-articulation specifications, instead there is a single rule which refers to a grouping property, "place", which organizes the specific features which we observe being assimilated.

In addition, prosodic constituents such as "syllable", "mora" and "foot" have the same formal properties as features, at least in the account developed here. An expression like "feature, node or prosodic constituent" would be unwieldy in making theoretical statements about representations, but it would be confusing to subsume all three under the term "feature" since everybody knows that a syllable is not a feature, even if I were to redefine "feature" so that a syllable is now called a feature.[19] An expression like "(Representational) entity" might be more apt, but it is a very neutral term, and it is difficult to keep clear the "feature, node or prosodic constituent" use of "entity" from the broader "thing" meaning. "Element" seems to be an appropriate term, except that it has a specific technical meaning in a particular theory ("element theory"). I solve this problem by referring to "feature, node or prosodic constituent" simply as "node", perhaps because that is what these things are called in graph theory.[20] We have now arrived at the fun-

[19] I hate redefining well-known words, because it is confusing, and I hate confusion.

[20] I will still use "feature" when speaking historically of things that were called "features".

damental existential concept of phonological representation: *node*. In arriving at this fundamental concept, we can also see the role of simplicity in this theory of theory-construction. We do not posit distinct notions of feature, node or prosodic constituent in the theory, since we have no evidence that "node" needs to be further subdivided within the theory. The simplest theory is, definitionally, the one that is correct.

A representation is not just a pile of nodes, it is a pile of nodes with structure. We therefore must also investigate structural concepts as part of the set of things that exist in the theory of representations. As part of the empirical background for this analysis, I take the main factual results of autosegmental and feature-geometry to be established. That does not mean that I assume all of the theoretical claims of autosegmental theory, rather, I take certain fact patterns to be well-enough established that any theory must have some way of accounting for "floating tone" or "onset sequencing" – it is not reasonable to deny that such facts exist, but their analysis is up for grabs. This means, concretely, that the SPE-theoretic model of representation (a segment is an unordered conjunction of features $F_1 \ldots F_j$, for some fixed value of *j*) is virtually untenable as a theory of representations.[21]

What structural concepts exist? Two basic relationships (between nodes) exist as part of a phonological representation: *dominance*, and *precedence*.[22] The notions of dominance and precedence are imported from general mathematics. Mathematical concepts do not gain automatic free entry into phonological theory, but it is legitimate for a theorist to assign a mathematical interpretation to concepts of phonological theory, when such concepts already exist in mathematics. Presumable these are concepts familiar to theoretical linguists and do not require much practical explanation. In terms of the basic representational things that exist in phonology, there are nodes, and there are relations between nodes. In fact, this is not a specifically linguistic or phonological finding: the linguistic (phonological) finding about representations is the nature of nodes and their relationships.

3.2.2. PROPERTIES OF PHONOLOGICAL ENTITIES

We have yet to determine whether any specific nodes are in UG[23] – at the moment, all we have is nodes, not even nodes of different types. Representational nodes do in fact fall into different types, which is an important property of the concept "node" in linguistics. There are two dimetrically opposed theories of representation regarding node types: the universal-substance theory where all nodes and dominance relations are fixed in UG, and the substance-free theory where no nodes or dominance relations are listed in UG. One might entertain mixed models with some nodes and relations fixed in UG and others being learned, if one had good reason to do so. In all models, any node is of some type – in the substance-free model, possible types are not given by

[21] We will, in part II, discuss the best-possible theory of representations where a segment is just an unordered set of feature specifications – that alternative theory puts the whole burden of structure into the theory of computation.

[22] This is an example of concepts that are valid for grammar, not just phonology. However, the properties of 'dominance' and 'precedence' in phonology and syntax may be different.

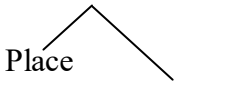
[23] Indeed, we have yet to address the question of whether there is such a thing as UG. I refer the reader to Hale & Reiss (2008) for basic arguments that there is such a thing as UG, but I take it to be a theoretical desideratum that the content of UG should be minimal. In searching for a "theory of phonology", we are simply asking, what is it about the nature of phonological grammars that is inherent in the language faculty. If the reader doesn't believe that there is such a thing as the language faculty, then the enterprise of constructing a "theory of phonology" is senseless.

UG, they result from something in the specific language. The universal-substance theory has a specific inventory of types (as well as a set of predetermined dominance relations).

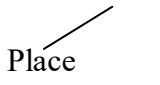
The precedence relationship holds of entities of the same type: any two instances of [voice] have a precedence relation (either $voice_i \prec voice_j$ or $voice_j \prec voice_i$), likewise any two instances of Place, two instances of root, or two instances of σ. In using the term "precede", that does not imply that the alternative notion "succede" is either theoretically inferior, or is also justified. At this point in the articulation of the theory, we can by convention focus on left-to-right ordering rather than right-to-left ordering, without implying that, specifically, preceding versus succeeding is correct. In phonology, same-type nodes have precedence relations.

The dominance relation is applicable to things not in a precedence relation, namely nodes of different types. Place dominates Coronal, and Coronal dominates anterior and distributed. The entity Root ("segment") dominates all segmental features. It is not clear what Root immediately dominates,[24] but probably it immediately dominates [continuant], [nasal] and Place, but not [voice] (which is immediately dominated by Laryngeal). The same notion of dominance applies to rule-derived suprasegmental grouping entities such as syllable and foot. Dominance is not a total relationship, and it is quite possible for type-distinct nodes to have no relationship. For example, [distributed] and [nasal] typically (possibly always) are not at all in a dominance relationship, that is, [distributed] does not dominate [nasal] and [nasal] does not dominate [distributed]. The notion of "floating" specification means that in a specific instance, a token of a typically "lower" node α which might be dominated by β is in fact not dominated by anything: the dominance relation between α and some "typically higher" node β is a potentially distinctive representational property for a given segment. (5a) is minimally distinct from (5b), in that Laryngeal is dominated by Root in (5a) but not in (5b).

(5) a.



b.



The minimal theory of relations is that the kind of relation between two nodes depends on whether the nodes are of the same type (one or more instances of [nasal]) or of different types (an instance of [voice] and an instance of [round]). Whether or not we require other relations is a matter for investigation, for instance the notion "head" may also exist, but clearly "precedence" and "dominance" are existing relationships.[25]

We have the following theoretical identifications regarding representational relations.

[24] These are exemplifying / explanatory statements, not definitional, and they are not claims about essential properties of the concept. This leaves open the possibility that what root immediately dominates is determined on a language-by-language basis. In UFT, for example, it would be a fundamental property stated in UG that [voice] is dominated by Laryngeal, and Laryngeal is dominated by Root.

[25] A consequence of this minimal theory of representation, and the mutual-exclusivity of dominance and precedence, is that recursion would be impossible, where A dominates A, insofar as the exact relationship that exists between two entities in a relationship is said to depend on whether the entities are of the same type or different types. Later, we consider the consequences of establishing structures where A dominates A.

(6) *Order: nodes all of the same type* α *are ordered*
Dominance: nodes of type α *dominate nodes of type* β.

The linguistic justification for the ordering proposition is ubiquitous, and can be observed in the fact that "ask" and "ax" are different words in English, composed of the same segments but in different orders. It is attested in the computations of myriad languages where velars palatalize before front vowels but not after front vowels. It is virtually unimaginable that a theory of phonology could dispense with the proposition that nodes of the same type are ordered.[26]

The dominance proposition is justified by the numerous computational facts that motivate autosegmental phonology, which need not be recited here. Certain alternative statements of the proposition ought to be considered, nevertheless. First, to clarify what the claim is, it is that a dominance relationship *can* exist between some two nodes of different types. Place can dominate labial (for a particular instance of Place and labial), Root can dominate Place (for a particular instance of Root and Place), σ can dominate Root (for a particular instance of σ and Root). It is also possible that some instance of Place does *not* dominate some instance of labial. In other words, dominance is an allowed but not mandatory relationship between nodes. Since it is not presently clear whether the theory should include recursive representations, I leave open the possibility of a condition that $\alpha \neq \beta$. The plan is to not prematurely hard-wire the answer into the foundational concepts of the theory.

A major question remains regarding the theory of representations, namely how specific node-types and their properties are registered in a particular language. The answers in universal-substance theory and substance-free theory are roughly the same, except that the mechanism is expressed in the grammar of the particular language in the substance-free account, rather than in the universal theory of representations. We can partially answer the question by stating the rule which states that there exists a node Place which may dominate Coronal, Labial or Dorsal.

(7) Place → Coronal, Labial, Dorsal

This rule states the node-types which Place can immediately dominate. We discuss this below under the theory of computation.

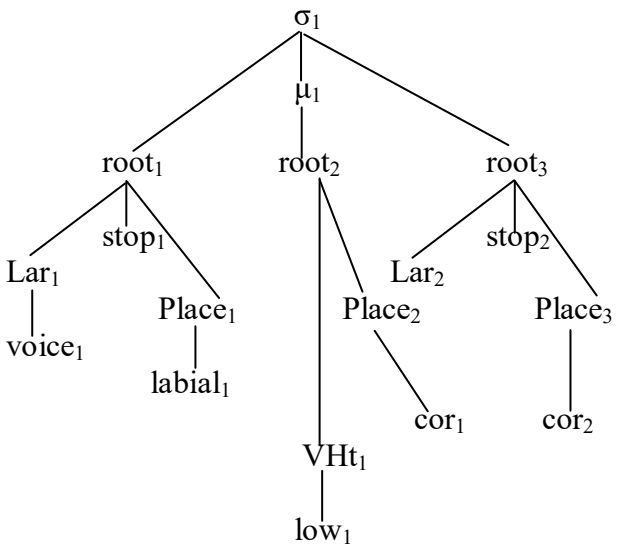
This constitutes the entire proposed minimal UG theory of phonological representations. There are nodes, and two relational concepts – precedence and dominance. Nodes have types; they enter into ordering and dominance relations as stated in (6). The challenge, now, is to see whether such a minimal theory of UG is sufficient (a question whose answer requires a theory of computation). As a brief review, here is the (full) representation of [bæt].[27]

---

[26] The CD model of Fujimura, which is a model of phonetic computation, does appear to dispense with ordering within the syllable, though not at higher levels. However (a) this is not a model of phonological computation and (b) it eliminates ordering by a significant expansion of the theory of representational units, where syllables have many more components and there are such things as "pre-spirantized stops". On the other hand, substance-free models of representation do not claim that there is a predetermined inventory of classificatory features, so tradeoff between more node types and fewer relations is a valid possibility. We will explore this tradeoff in part 2.

[27] This is [bæt] in an arbitrary language – the features required for a specific language must be determined on the basis of the phonological facts of that language. The features specified here are somewhat arbitrary. Integer subscripts on nodes are a means of stating what precedes what, and have no theoretical significance.

(8)



(8) is the graphic representation of the following conjunction of relations, which is the minimum necessarily-stored content of [bæt]. D is "dominates", P is "precedes".

(9) $D(\sigma_1,root_1)$ & $D(\sigma_1,root_3)$ & $D(\sigma_1,\mu_1)$ & $D(\mu_1,root_2)$ & $D(root_1,stop_1)$ & $D(root_1,Lar_1)$ & $D(root_1,Place_1)$ & $D(Lar_1,voice_1)$ & $D(Place_1,labial_1)$ & $D(root_2,VHt_1)$ & $D(root_2,Place_2)$ & $D(Place_2,cor_1)$ & $D(VHt_1,low_1)$ & $D(root_3,stop_2)$ & $D(root_3,Lar_2)$ & $D(root_3,Place_3)$ & $D(Place_3,cor_2)$ & $P(root_1,root_2)$ & $P(root_2,root_3)$

The reader will notice that this is not an exhastive list of the dominance relations implied by (8), for example σ_1 dominates low_1, a fact which is not listed in (9). Likewise, $Place_1$ precedes $Place_2$, Place2 precedes $Place_3$, and $Place_1$ precedes $Place_3$, but these facts are not registered in (9). Because dominance is a transitive relationship (if D(a,b) and D(b,c) then D(a,c)), "σ_1 dominates low_1" follows from the above set of relations. The same goes for the ordering relation $P(Place_1,Place_3)$. Some additional theoretical statement is required so that we can infer that $Place_1$ precedes $Place_2$ from the fact that $root_1$ precedes $root_2$ and the respective root nodes dominate the Place nodes. This points to the need for refinements in the concepts of precedence and dominance. One might consider the alternative of replacing "dominate" with "immediately dominate", but this would complicate the theory of computations, as discussed below.

3.2.3. THEORY COMPARISON: REPRESENTATIONS

One of the main reasons for explicitly articulating the things that a theory claims exist is to allow theories to be evaluated in terms of simplicity. The above theory of representation in RSFP can be contrasted with Unified Features Theory as promulgated in Clements & Hume (1995). That theory adds other propositions to the theory.[28] For example, UFT has a fixed set of representational entities in UG – coronal, dorsal, radical, open, Vocalic and so on. The difference between RSFP and UFT is that RSFP only assigns the general concept "node" to UG, where different node-types are acquired based on phonological patterning (not mandated by UG), and are named

[28] Historically speaking, RSFP which is proposed later, subtracts (implicit or explicit) propositions from UFT.

arbitrarily. The number of learnable nodes is not set by theory in RSFP. In UFT, dominance relations have numerous properties specified by the theory, which are propositions about UG: "Coronal dominates anterior", "Laryngeal dominates [c.g.]", "the root node is composed of the unordered features *sonorant, approximant, vocoid*", "C-place immediately dominates vocalic", "C-place **may** immediately dominate Coronal", "V-place **may** immediately dominate Coronal" – that is, features generally have a unique immediately-dominating node, but in the case of Coronal (idem Radical, Dorsal, Labial), those features may be immediately dominated by either V-place or C-place. To the extent that such statements are grammatically useful, the grammar of a particular language in RSFP could also contain such statements, i.e. rules like (7). The primary difference between RSFP and UFT in terms of how representational objects relate to each other is that UFT assigns much more to genetic endowment (being in UG), and RSFP assigns much more to learning.

The basic existential concepts of node, precedence and dominance are shared between RSFP and UFT. The theories differ in the auxilary propositions attributed to these properties as part of UG, for example the specific list of nodes as UG statement, and the required or optional dominance relationships between entities attributed by UFT to UG. Both theories have ordering as an essential property of same-type nodes. In other words, the representational statements characterising RSFP are a proper subset of those of UFT. From this, we conclude that RSFP is formally simpler than UFT. Of course, the proper comparison is between two whole theories, not just the sub-theories of representation. However, RSFP also does not require a more complex theory of computation.[29]

As a second example of theory comparison, the preceding discussion has taken from RSFP the simple object "node", but UFT, possibly, and other representational theories, certainly, impute to UG a richer inventory of primitives, especially nodes classically called features. Is voicing represented with just [voice] which either exists or does not exist? Or is voicing a complex of an attribute [voice] and a value, + or – (or possibly other values)? In the latter case, it is even possible (especially in post-SPE autosegmental theories with underspecification) that a feature either exists or does not exist for a given segment, **and** if it exists it has either the value plus or the value minus. Reducing representations to informational minimality, we might have the following distinct representations in the value-attribute (binary) theory of features.

(10) A A A
| |
[+B] [–B]

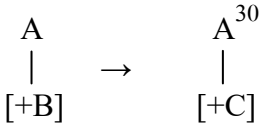
The privative theory admits the following relations:

(11) A A
|
B

[29] This is an instance of a more general difference between FP and RSFP within FP, versus earlier UG-dependent practices where it was considered desirable to encode as much of our knowledge of language to UG as is possible. for discussion of the proper burden to be placed on UG.

The set of possible specifications in value-attribute theory is larger than that of privative theory, but this fact has no direct bearing on theory-evaluation. What matters is which theory is simpler – which theory posits the fewest theoretical entities or identifying propositions. The valued-feature theory posits all of the theoretical concepts of the privative theory, plus some. First, *feature* has to be distinguished from *node*, since in valued-feature theory, there are still nodes without values (Place or σ, for example). There are two representational things, *node* and *feature* in this theory. A feature is not a primitive thing, it is a pairing of two logically independent things, a *value* and an *attribute*. It is typically held in valued-feature theory that nodes can dominate an entity, but *features* do not dominate anything (this is an identifying proposition of the theory regarding the nature of *feature*). Morover, valued-feature theory states a proposition to the effect that a value cannot exist free of an attribute (minus by itself does not exist), and an attribute cannot exist free of a value (there are no bare features). Finally, only the value of a feature can be changed by a computation. The operations (12a) are possible, and (12b) is not.

(12) a.

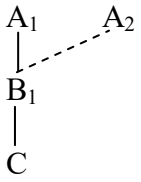

b.


In other words, valued-feature theory posits all of the things of privative theory, plus 3 additional concepts (in addition to node, it has feature, value, attribute) and 3 propositions (only nodes dominate features; only values can be changed in a computation; a value exists iff it is paired with an attribute). A common 4th inventory proposition is that "values" can only be either "+" or "–". On grounds of theoretical simplicity, privative theory has the logical advantage and is to be adopted, until it can be shown that the additional claims of valued-feature theory are justified.

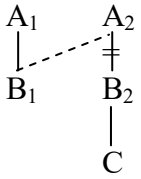
An empirical issue sometimes raised in discussions of privative vs. binary theories of features is a putative prediction of privative theory, that only one value of a feature can spread or "be active", compared with the prediction of binary feature theory that either value can spread / "be active". If we assume, as is often done, that vowel height is privatively represented as [high] and the vowels [e, o, æ, ɑ] lack that property, no rule of assimilation can spread "–high" from non-high vowels where /i/ → [e] after /e,o/. The reason is that "–high" doesn't exist in privative theory and mid vowels don't have a "–high" which can in some way be copied from one vowel to the next. But many languages have such a vowel harmony process, including most Bantu languages. How is this not a counterexample to the claim of privative theory?

Spreading of node B_1 from A_1 to A_2 means that A_2 will gain whatever structure – or lack – exists under B_1. Suppose B_1 dominates C, then such spreading assigns the property C to A_2, as a consequence of spreading B_1.

[30] This rule says "replace the attribute [B] with the attribute [C] when paired with the value [–] and immediately dominated by A". The same string of symbols in a rule would actually be interpreted as meaning "insert the value-attribute pair [+C] under a node A which also imediately dominates [+B]".

(13) 

Suppose instead that B_1 dominates nothing, and futhermore that B_2 is already associated with A_2 dominates C. If B_1 spreads to A_2, A_2 then loses any specification for C.

(14) 

The appearance of anomalously "spreading a lack" can result from spreading node B which does not dominate C, and not from "spreading lack of C". The phenomenon of "spreading a lack" is accounted for formally by spreading a higher node that lacks the property in question (thus supplanting any pre-existing specification).

While simplicity suffices to choose privative feature theory over binary feature theory, such simplicity arguments may not always exist. It is entirely possible for two theories to be equally complex. Is it then valid to appeal to the greater representational power of binary features in deciding between the theories, in case the theories are equally simple? It is, via the requirement for justification. Assume two distinct theories, which differ in containing sets of claims A versus B. The choice of A versus B is based on comparing their justifications, J(A) versus J(B). The gratuitous representational power of A is where A predicts a class of facts P_A which is not predicted by B – A makes a false prediction. If P_A is shown to exist, not only does P_A not count against A (predicting something now known to exist), it counts against B (which predicts that P_A will not exist).

As a point of clarification regarding the approach advocated here, notice that enumeration of theoretical entities was separated from discussion of the properties of those entities. The point of doing so is to emphasize the logical independence of questions of existence versus questions of identity. Everything that exists has some identity. It might therefore seem to be advantageous to state all of the properties of an entity when stating that the entity exists. This would be a self-defeating method of gaining knowledge, because we do not know all of the properties of nodes, order, or dominance *ab initio*. We do know some – the essential property which distinguishes one concept from the others (the definition) is given. Thereafter, other properties can be given.

3.2.4. THEORY COMPARISON: ORDERING

It is claimed above that nodes of the same type have the relation of precedence. This leaves open some room for variation: does this mean that such nodes are strictly (totally, linearly) ordered, or are they partially ordered? Is this result encoded in the theory of representation, or does it result from something outside of phonological theory? If same-type nodes are totally ordered, then a 4-item list of nodes has the following precedence relations.

(15) a. $n_1 < n_2$

b. $n_2 < n_3$ $n_3 < n_4$

c. $n_1 < n_3$ $n_1 < n_4$ $n_2 < n_4$

If not all pairs stand in a precedence relation, the ordering is a partial ordering. If we take segments, that is root nodes, is there any case where we would say that two segments are phonologically unordered? For example, [a{k,p}a], where [k] does not precede [p] and [p] does not precede [k]. Or, [a{d,t}a] where [d] does not precede [t] and [t] does not precede [d]. There are certainly cases where different articulator features overlap within a single segment, e.g. [akpa] with a labiovelarized stop, but this is understood phonologically to be distinct features overlapping within one segment, and distinct features have no precedence relations. No evidence has been uncovered showing the need in some language for two same-type nodes being unordered.

Does this mean that the representation of [spɪt] requires the whole set of precedence relations of (15), or does it need just those in (15a-b)? There is no sensible physical interpretation of phonological representations where *less* than the relations in (15a-b) still represents |spɪt|, but it is conceivable that the order of {s,p} is the product of a phonological rule – there is no *evidence* for such an analysis, but this is a conceptual possibility. It is most reasonable to assume that the representation of [spɪt] includes at least (15a-b). While it is true that *s* precedes *t*, this is not a necessary fact of the representation (as opposed to the relation $n_2 < n_3$, which is a necessary representational relation that distinguishes [spɪt] from [sɪpt]). A precedence relation between the first and last segments can be inferred from the transitive nature of precedence, but even that fact is irrelevant until we establish the need to compute the precedence relation between the first and last segment of the utterance. Adding (15c) to the representation of [spɪt], and the myriad analogous precedence relations that would be needed to impose strict ordering on representations, complicates representations without reason: Occam's Razor prohibits unnecessary complications.

The theory of representation advanced here says that (15a-b) is a possible representational fact, but it does not impose conditions of obligatory precedence on representations. This means that a sequence composed of {s,p,ɪ,t} with the vowel having no ordering relative to the other segments is a grammatically possible representation. Should we *allow* that degree of freedom, or should be ban it? Why not add to the theory a codicile to the effect that *all* same-type nodes must enter into two precedence relations – one stating which thing the node immediately precedes, and one stating which thing is immediately preceded by the thing? The reason is simple: such a statement is unnecessary, as part of grammatical theory. The result, that the segments of [spɪt] do in fact have such precedence relations, follows from simple learning, and need not be re-stated in grammatical theory. The physical inputs that give rise to (s,p,ɪ,t) will always give rise to the ordering relations (15a-b), and it is not necessary to burden the theory of phonological representation with an additional stipulation mandating the encoding of certain precedence relations in a grammar. The explanatory burden for the lack of instances of the type [a{d,t}a] is borne by theories of speech production, phonetic representation, and phonological learning from phonetic stimuli. The simplest grammatical theory is one without such a condition in phonology (either as a language-specific condition or as an aspect of UG). Thus it remains conceptually possible for there to be a phonological representation [a{d,t}a], if such a representation follows from facts of the language. Perhaps if there *is* no theory of learning that renders [a{d,t}a] unlearnable, we may need some phonological account of the lack of such representations. Or, we may find in pursuing rule theory that there is independent reason to include or preclude such representations. Lacking

any evidence to support adding other ordering conditions to the theory of representations, we don't add anything to the theory.

3.3. Concepts of phonological computation

The input to phonology (coming from syntax – we need not care if there is a morphological component in the middle) is the *underlying form,* and the output of the phonology is the *surface form*. Whatever relates the input to the output in phonology is the system of phonological computation. In the most uninteresting case, there is no computation: any input directly becomes the output. Usually, one or more rules produces a representation that is not the same as the input. The justification for claiming that there are phonological computations ought to be well-known to anyone in the field, except that there has been significant dispute as to what counts as evidence for phonological computations (and therefore what a computation is). See Odden (2013) and Hale & Reiss (2008) for some discussion. Pursuant to the narrower scope of phonological theory assumed in such works, the computational component is only assumed to contain whatever devices are required to get from the underlying form to the surface form.

3.3.1. RULES

The basic primitive concepts of the theory of phonological computation are *representation*, *rule*, and a relationship of *ordering* between rules. In applying a rule to a representation, the input representation becomes the output in a specific way. The total grammar maps the underlying form to the output in a specific way, as determined by the *ordering* of the rules. Rules may be in a precedence relation – the simplest theory is that representational precedence and computational precedence are the same concept, applied to different objects. Rules can be numbered 1 through *n*, and the computation starts by applying rule 1 to representation R_0 (the initial representation, the underlying form) which maps R_0 to R_1, then rule 2 applies to R_1 converting it to R_2, and so on until rule *n* applies to give R_n which is the surface form, the output of the grammar. I set aside the complication of cyclic rule application, where a rule applies to a sub-representation R_i and where rules apply order …R_i, R_{i-j}, R_{i-j+1}, $R_{i-j+2,\ldots}R_{i-(j-(j-1))}$, R_i, R_{i+1}… ($R_{i-j\ldots}R_{i-(j-(j-1))}$ being the reapplication of a cyclic sequence of rules composed of *j* rules, the last of which is rule R_i).[31]

The concept of *rule* is somewhat complicated (because it has been so well studied). First, there is "what the rule does", that is, the *structural change* (*s.c.*). Examples are "spread [nasal] leftward", "delete a vowel", "insert [ʔ]". Every rule applies when specific representational conditions are met – this is the *structural description* (*s.d.*) of the rule. A rule "spread [nasal] leftward from a sonorant to a sonorant" has the structural description "$root_i$ dominates [sonorant], $root_j$ dominates [sonorant] and $root_j$ dominates $[nasal]_k$". When this is true, the structural change of the rule is implemented: a dominance relation is added between $[nasal]_k$ and $root_i$.

All we have said about grammar, so far, can be graphically charted as follows:

[31] A downstream problem regards the possibility that some computations may state global properties of relations in a grammar, for example legitimate node-domination relations such as the requirement that [coronal] be dominated by [Place]. Such computations are requirements that must always be true, not just true at a single point in the derivation.

(16) Phonological component
Is a set of rules
The set of rules is ordered
A rule maps a representation to a representation
A rule is a representation
A rule has a structural description and a structural change

We may need to say more about the phonological component – if there are other essential properties of the phonological component, they can be added as further propositions about the phonological component. We can also add or modify propositions about what a rule is or does. At some point, we will say how a representation is scanned for application of this rule (does the rule start at the right end of the representation and work leftward; does it start at the left end and work rightward; does the rule simultanously apply at all possible points where the conditions described by the rule are satisfied at the initial input to the rule?). Finally, a rule might be obligatory, meaning that the operation described by the rule is always performed when the described conditions are satisfied; or a rule might be optional, meaning that the operation may be performed *or* the substring may be passed through unmodified: the rule has two outputs for an input.

If we can develop a sufficiently precise theory of "rule", we can reduce rules to symbolic strings – equations of a sort – which state what any rule does. This approach to phonology was followed reasonably closely in SPE-theoretic accounts of rules, but formalization practices became looser in the autosegmental era where it was assumed that UG provides a rich arsenal of automatic operations that "flesh out" the actions of less-specific rules of particular languages, where rules are only a small part of phonological computation. Still, it remains possible to give a reasonably precise formalization of rules under the theory of autosegmental representations with minimalist computations. Although it is possible to go straight to the formalism and develop rule concepts based on the notation, it would be wrong to do so, because that misconstrues the nature of notation. A system of notation encapsulates the concepts that are implied by the concept "rule", so first we must understand what the concepts behind "rule" are, and then we devise a notation for expressing those concepts.

The two rule concepts that are most in need of elaboration are *structural description* and *structural change*. We start with structural description. Whatever it is that happens when a rule applies, it happens in some context, the *s.d.* This is a specification of the class of representations that must be present, in order for the *s.c.* to be implemented. Is there a significant difference between the notion of *representation* as used in the preceding section and the notion of *representation* for phonological rules? Our previous discussion of *representation* was a bit underspecified, since we just took for granted that we understand what the general term "representation" means. Let's see if that is true. Here is the Wikipedia page on "mental representations", for what it is worth: https://en.wikipedia.org/wiki/Mental_representation (I disagree with it to some extent).

A "rule-representation" refers to a class of "representation-representations", i.e. it refers to, picks out, or describes, a certain class of stored-or-computed non-rule "representation-representations" (henceforth termed, for purposes of clarity, a *phonological datum*, or simply *datum*, which is not to be confused with the scientific concept of "data" where [bɨβ̰at] is a piece of phonological data in that sense). A representation-representation or "datum" is a thing outside of the rule system, it is what a rule applies to and produces as output. (17) as a structural description *refers to* a class of data: any datum that has a root node which dominates a Place node that

dominates a Dorsal node. It does not distinguish between voiced vs. voiceless segments, stops vs. fricatives, or rounded vs. non-rounded versions of this class.

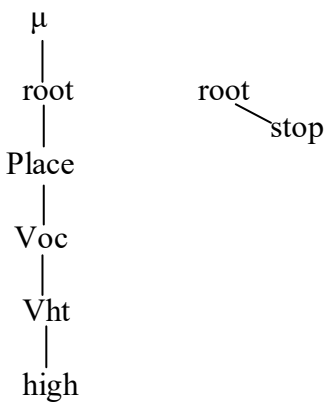
(17)


A datum such as [frɔg] refers to something very specific (an aspect of phonetics dealing in parsing acoustic signals into phonological representations, which produces a pre-phonological representation of atomic segments, e.g. [frɔg]). The place where rule-representation and phonological data differ is in what kind of things they refer to. A rule-representation *refers to* classes of data existing within the phonology – those data which are picked out by the rule. A phonological datum may refer to something outside of phonology, though I am not clear on exactly what it refers to – something in phonetics. It is not important to the theory that we say what a phonological representation represents – from the perspective of grammar, a representation simply "is". The interpretation of "representation" has some significance in terms of justification, since we don't just arbitrarily posit grammatical concepts simply because we can, we do so because it serves some purpose. The theory is that representations are real mental things, which are caused by something – encountering a physical utterance like |frɔg|, for example. Surface-phonological [frɔg] may thus "represent" a class of physical utterances that a hearer needs to subsume under some kind of cognitive symbol.

Do the two notions of representation differ in an important theoretical way? Are there things that can be rule-representations but not phonological data, or vice versa? For instance, can (17) be a datum? It satisfies the requirements of being a data-type representation: there are nodes, and dominance and precedence relationships between them. It does not obviously correspond to a complete segment, though it might be the representation of [k]. Even if it didn't refer to a complete segment, being a complete segment is not a condition that is imposed on a datum in phonology – "floating" subsegmental representations are phonologically possible. At this point (pending the discovery that the two kinds of representation have to be formally distinguished), we conclude that we do *not* have to say anything other than the fact that a structural description is a representation.

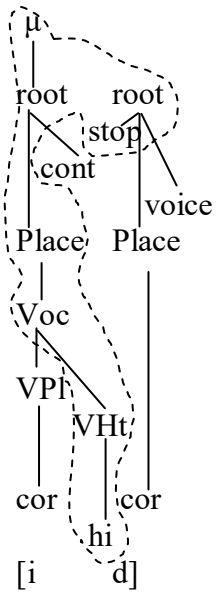
We do have to do more than just say that such-and-such representation is the *s.d.* of a rule, we have to say how the representations are related when a rule applies, but what we have to say about s.d. type representations is not filed under the concept "representation", it is filed under "computation". If a rule-representation "matches" a datum, one thing happens, if it doesn't, something else happens. Grammar has a rule-to-string matching function. Let us consider a specific *s.d.* and a datum. The rule below looks for a stop which comes right after a high vowel, as encapsulated in representation (18). The reader will hopefully detect standard representational premises which say "vowel".

(18)



Ordering of segments ("after a vowel") is established by mentioning root nodes, which are ordered. The dominance relations (association lines) of the first segment encode "vowel" via dominance by μ, and vowel height is specified by reference to a series of linkages from [high] to Root, likewise "stop" is specified by the linkage of [stop] to $Root_2$. As a string that might be picked out by (18), consider [id]. How is "match" characterized so that [id] is an example of what the s.d. of a rule refers to?

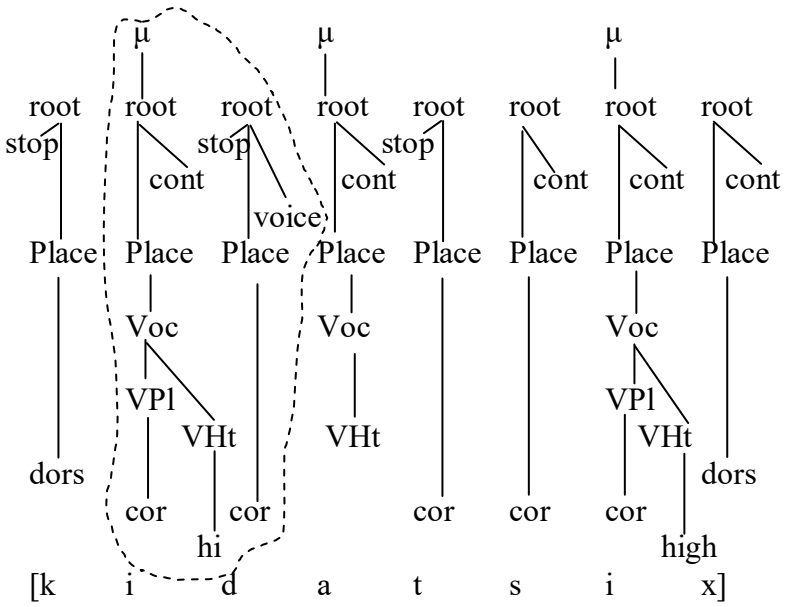
(19)



The representation that constitutes the s.d. of the rule, (18), is included in (19): everything in (18) is also in (19). The subpart of (19) that corresponds to (18) has been circled for convenience. There are things about representation (19) that are not in (18), such as the place specification of the vowel and the voicing and place specification of the stop, but that does not matter – there can be other things in the representation. Because everything in (18) is also in (19), the string "matches" the s.d. of the rule at this location – therefore the rule will apply.

What about (20), [k**id**atsix], with other stuff in the representation? Only the sequence [id] matches the rule-representation (18), the structure corresponding to [id] being encircled.

(20)



Again, only one specific subpart of the representation "matches" the *s.d.* representation – but there is a match in that circled part. There is *not* a match between the final VC ([ix]) and the s.d. in (18), since the datum contains [cont] within that subrepresentation whereas the s.d. requires presence of [stop]. Assume a third representation, one like (20) but with three sub-sequences matching (18) – then there would be three sub-representations with matches, and no more. So far, we have seen no hints of any problem with the concept *structural description*, which is simply "a representation in a rule". Paired with the concept *match*, we have a good start on the project of saying how a rule applies to a string.[32]

The *structural change* (s.c.) of a rule is a statement of what the rule does.[33] A rule either adds or subtracts something to a representation, or conceivably "changes" something. After all, we commonly say "/g/ changes to [γ] (between vowels)". The difference between [g] and [γ] is that [g] has (or lacks) a property that [γ] lacks (or has) – such a change involves an addition or a subtraction. Given the assumed theory of representation, "change" as a primitive operation is a conceptual anomaly. A change in identity (distinct from insertion or deletion) would be if the actual representational node changed its identity (type) from being one thing to being another, e.g. a node of the type [nasal] converts into a node of the type [voice]. It is hard to see what could justify the concept "change" along with "insert" and "delete". Any apparent "change" might simply be the result of a combined addition and deletion ("delete [nasal] and insert [voice]"). The simplest theory has only "insert" and "delete" as components of s.c., and does not admit the third option "change of node type". Then the question is whether the s.c. of one rule can have both insert and delete operations.

What can be inserted? A common kind of insertion is insertion of a dominance relation – things spread. Deletion applies to dominance relations, as well (delete an association between H

[32] In §4.2, we will see that this good start needs some correction, that the concept of "match" is a bit more subtle that simply computing a subset relation on representations.

[33] It would be a mistake to say that the s.c. is what the rule does. The reason for this distinction is discussed in section 6.5.1.

and μ). Can a node be deleted? Examples of node-deletion are devoicing (delete [voice]), spirantization (delete [stop]), delete a whole segment (delete a root node), deletion of a Place node ("debuccalization"). Less common is deletion of a syllable (attested as "haplology" under conditions regarding identity of the elements of the syllable). In other words, it looks like the theory of *s.c.* may be that you insert or delete a representational entity (node or relation). There is, at present, no reason to say anything more than that a rule can insert or delete a node or relation.

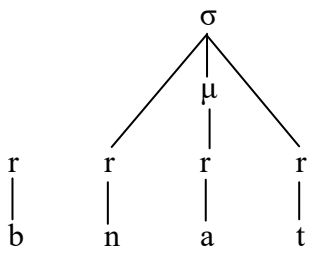
The case of insertion still requires a bit more analysis. First, insertion of a relation requires specifying which entities are in that relation. When a rule says "insert a domance relation", it says which two things gain that relationship. By our theory of relations between type-distinct nodes, if A and B have a relationship, either A dominates B or B dominates A.[34] A rule of spreading says "insert the relation D(A,B)". This seems to suffice, for the moment, to deal with spreading.

Second, when a node is inserted, is it just inserted at random within the representation on the appropriate tier, or is it inserted along with some relation between the node and the rest of the representation? Although a dominance relation D(A,B) entails existence of A and B, existence of A and B in a potential dominance relation does not entail that there *exists* such a relation. Can we just insert a node, or do we always insert a node plus a relationship? To answer the question, we should ask whether we ever need to insert a node *without* a dominance relation – to what extent is node-insertion factually independent of node-and-dominance insertion? The most common examples of node-insertion are prosodic structure-building operations: morification (rules making certain segments become moraic), initial syllabification (inserting a syllable in connection with a moraic segment), footing (creating a foot that contains some syllable). Other examples include the insertion of root nodes (a syllable is provided with an onset consonant, perhaps to be later specified as [ʔ], [j] or [k]; a mora gains some content that wasn't already present). Perhaps some feature is inserted into a segment in some condition (i.e. "default specification", though that is a theory primarily applicable to binary feature theory with universal surface specification).

A typical line of phonological reasoning regarding prosodically-conditioned insertion has been that a "defect" is identified, usually the presence of an unsyllabified consonant, but perhaps a foot without enough stuff in it is created (or, there is an attempt to create a foot which results in a defective foot that has to be repaired), and this defect is ameliorated by inserting a mora, which then forces construction of a new syllable, which in turn provides a place for the unsyllabified consonant to be associated. This UG-heavy functionally-driven account is the antithesis of the FP approach to phonological analysis, and unfortunately the vast majority of work in prosody-construction has been carried out in the UG-laden metatheory, so it is not clear what an explicit and minimalist account of prosody should looks like, though see Baal et al. (2013) for some proposals. Let us nevertheless forge ahead with a simple (constructed but representative) case: the language only allows a single C at the beginning or end of a syllable, but underlying forms contain representations like /bnat/. Ordinary syllabification will yield the following representation (specifics of segmental structure is suppressed here).

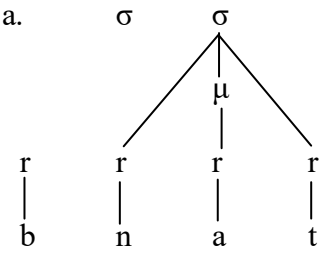
[34] Traditionally, when we talk about relations in terms of "association", we simply say that A and B are associated, and dominance is talked about as a separate property. However it us not clear that we need both relations: the way we have talked about things is not sufficient evidence for what things actually exist.

(21)

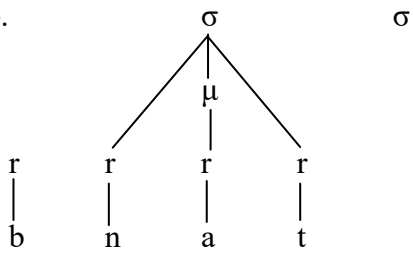


To this structure, we apply a rule that inserts a syllable,[35] stated (initially) as applying in the presence of a root node not dominated by a syllable. The s.c. of the rule is "insert syllable". The representation already has a syllable – does the rule produce (22a) or (22b)?

(22) a.



b.



Representation (22a) and not (22b) should be the output of the rule. Now suppose the unsyllabified consonant is towards the right end of the representation, then given /tanb/, /b/ would be unsyllabified, and we expect the inserted σ to be to the right in the representation, not to the left: a vowel is not inserted word-initially in case there is a final consonant cluster. In other words, the locus of the action relates to the place in the representation where there is a match between the s.d. and the entire representation.

The traditional rule-concept *focus* conveys location-of-action information in s.d., where one part of the s.d. is its "focus". The action to be carried out, as described in the structural change, involves specifically-identified entities in the matched subpart of the representation. In the case of the insertion of a mora in the context of an unsyllabified consonant, the s.d. is the description of an unsyllabified consonant, and the focus is also the unsyllabified consonant. The change is the addition of μ, and this takes place "near" the unsyllabified consonant. In a rule of intervocalic voicing, the s.d. is the sequence VCV, and the *focus* within the s.c. is the consonant.

I have advocated not letting notational practices predetermine the concepts underlying rules, but it is worth pointing out that to the extent that rules have ever been explicitly written to

[35] Maybe what happens first is that a mora is inserted, then a syllable is erected over the mora. The goal here is simply to see what we would have to say when we say "insert a prosodic entity".

implement bare insertion of prosodic entities (leaving their association to "other principles"), a rule of insertion could be written as (23).

(23) $\emptyset \rightarrow \sigma$

root′

This says in notation what I said above, that addition of μ takes place "near" an unsyllabified segment. There isn't yet an answer to the question raised by (22) regarding the position of other syllable nodes in the representation, but we now know that the answer depends in part on refining the concept of structural description.

Because a dominance relation is ultimately created in connection with insertion of a prosodic node, the correct theory of s.c. may be that node-insertion *always* comes with insertion of a dominance relationship between the focus and inserted nodes. In autosegmental rule theory, we would say that a vowel is "assigned" a mora or that a vowel is "assigned" a syllable node (or feature), meaning that the node is inserted along with the relationship between the node and the dominating / dominated node. (23) might be explicitly be written as (24).

(24) $\emptyset \rightarrow \sigma$
⋮
root′

Is there evidence that phonological rules insert purely floating material, which becomes linked to particular positions by one or more later rules? No clear examples come to mind that show the need to separate entity-addition from relation-addition, making it credible to conclude that "insert" always includes a relation (spreading/dominance-insertion), and may include a node. While this is an important matter, we do not need to know this, in order to "define" the concept of rule.[36] We do need to resolve this, in order to have a complete theory of rules.

3.3.2. STRUCTURAL CONSTRAINTS

This could have been the end of discussion for elementary phonological computation, but that would leave a gap in our ability to compute the things that we have to be able to compute. This subsection is speculative and inconclusive, and is included to show how the program of constructing an ontology for phonology is relevant to hypothesis formation. Providing an ontology of phonology is not just a post hoc accounting measure for well-established theories.

[36] I would be remiss if I failed to point out that in various earlier works, I myself have proposed rules which insert floating tones in some morphologically-defined context, followed by one or more association rules to link that tone to specific vowels. For example, I proposed that all verbs in Kuria receive a floating H (though actually all words have a H, not just verbs) and then different rules link that H to the fourth stem mora, or the second mora, etc. The problematic aspect is the initial insertion rule, which insert a floating H. There are no phonological conditions on this class of insertions, instead insertion of H is a characteristic of all verbs in Kuria or Matumbi (to take two example languages). In neither case is there compelling evidence that the H in question is inserted by phonological rule. Marlo, Mwita & Paster (2015) provide evidence for Kuria that H is present only in certain tenses, thus is arguably an affix.

We are missing any account of apparent global properties of phonology. "Global properties" are grammatical behaviors regarding aspects of phonological representation and computation that do not seem to be attributable to properties of a specific rule's operation. One example is the segmental inventory of a language. English (most dialects) do not have [ħ ɗ ɬ | ʉ̃ œ] as segments in any stage of a derivation. Another is syllable-structure conditions: syllables in some language may be limited to CV(C), or $C^3V(C)$, or CV(V). Syllable-initial consonants may preclude [ŋ], or sequences like [lk]. In some languages, only sonorants can be dominated by a mora. Some languages require feet to have exactly two moras. In the substance-free theory of representations, the grammar of a language may need to contain some statement of what node-types exist in the language, and what regularities there are regarding dominance (e.g. does Place dominate Coronal?; can a token of Place dominate more than one node?).[37]

The approach to syllable construction advocated in Kahn (1976) places the entire burden of stating what is a "possible syllable" on the specific rules that construct syllables, which are, in theory, ordinary autosegmental rules of insertion and association. Limits on consonant sequences in onsets are supposed to be directly expressed via specific statement of the onset-association rule; likewise coda sequences are controlled by rules that explicitly encode the extant patterns. The problem with this program is that it failed *ab initio* – it was never possible to express those patterns in simple formal rule statements, even given the broad latitude of SPE rule theory.[38]

These phenomena suggest that certain nodes have a nature which is stated at the level of the whole phonological grammar, and are not just isolated to one rule. This implies a kind of statement governing the node types of a language, and some of their properties – what they (immediately) dominate. Suppose for example that some language has higher level prosodic nodes U, φ, Pwd, F and σ. The typical dominance relations found between these nodes is stated in the following rule-like statements, where "=>" means "immediately dominates".

(25) U => φ
φ => Pwd
Pwd => F
F => σ

Rather than doing what rules do (mapping from string to string, applying in a specified order), these computational objects state node admissibility conditions on representations, and they hold

[37] The necessity of node lists is not at all assured. Assume a language with no phonological evidence for voicing (e,g, all obstruents are phonetically voiceless, all sonorants are phonetically voiced). Nothing in the computational apparatus requires a statement that voicing does not exist in the grammar, just as nothing in the computational apparatus requires a statement that bananas do not exist in the grammar. In a language which contrasts /p/ and /b/, the existence of [voice] in lexical entries suffices to establish that [voice] exists. In a language which has a derived-only distinction of voicing, coming from some rule of intervocalic voicing, the existence of [voice] in the statement of a rule suffices to establish that [voice] exists. There is no reason to construct a separate statement to the effect that a language does exploit [voice], therefore the claim that computation theory must include lists of nodes lacks support, and simplicity dictates that there whould be no such statements. The question of what nodes dominate what other nodes is a matter of computational gravity, bearing on the theory of structural change representations. This point is discussed further in section 4. A list of extant nodes in a grammar can be computed from a list of possible dominance relations.

[38] As an exercise, we may at some point consider an explicit pure rule-based autosegmental ananlysis of English syllabification, *sans* global constraints.

through the computation.[39] These statement say, for example, that φ (immediately) dominates Pwd, and because there is no statement Pwd => φ, a structure where D(Pwd,φ) does not observe this condition is impossible. A rule "attempting" to insert such a dominance relationship cannot apply (it contradicts the condition expressed in Pwd => φ), and the rule would have no effect (therefore could not be learned).[40]

To avoid confusion over different types of "rule" that would result if we called statements like (25) "rules", I refer to such statements as *structural constraints* (they *are* constraints, in the original sense, which has gotten inundated by the OT concept of "constraint". See Odden (2011) for discussion of the different senses of "constraint"). Structural constraints do not map representations to representations, they regulate possible immediate dominance relations in a language. They implement the notion "structure preservation", as originally proposed for syntax in Emonds (1970), that is, they impose a requirement that rule outputs conform to the set of phrase-structure rules of a language.[41] They differ from other conceptions of constraint which have a wider scope, whereby there could be a "constraint" prohibiting voiceless stops between vowels. This is outside the scope of structural constraints, which only state allowed relations between mother and daughter nodes.

We can continue stating typical structural relations that exist in most languages:

(26) σ => X
X => Root
Root => Place Laryngeal [nasal] [continuant]…
Place => Cor Lab Dorsal Radical
Cor => [ant] [distrib]

The above statement for syllables is not the only constraint that pertains to what things syllables immediately dominate. Given the well-motivated claims of moraic theory, a more accurate statement is that a syllable branches to a skeletal position or a mora.

(27) σ => X μ

The theory of structural constraints mostly places supposed universal representational properties of theories such as UFT within the reach of language-specific statement. Where UG-heavy theories assume (25)-(27) as part of the immutable language faculty, FP/RSFP attributes such relations to the grammars of particular languages, to the extent that there is evidence for such statements. The impact of (25)-(27) in both interpretations is that these are relationships that repre-

[39] See McCawley (1968), where following a proposal of Stanley, such rules are termed node admissibility conditions. This is, essentially, the base component for phonology. Sampson (1970) postulates a phonological rule subsystem, the "phonological base" which is somewhat similar to the present structural constraint, but operates in the strongly non-minimalist universal-substance exploiting SPE framework, and the arguments come down to the invalid premise that grammar must express all apparently relevant factual generalizations about the data.

[40] Alternatively, the structural constraint would not be learned. The theory does not favor one mechanism over the other, it simply says that contradictions are not learnable.

[41] Emonds does not make structure-preservation a requirement of all transformations, so there are certain classes of rules that are not subject to those limits. No exceptional rule-types are proposed here – the structural constraints of a language are conditions on all representations.

sentations must obey, so computations must somehow be made to obey. The difference between the two interpretations is that in the substance-free approach, being part of the grammar of specific languages, the exact content of such statements can vary between languages.

The underlying concept behind structural constraint as given above is that the dominance relation in representations can be subject to limitations. It does not require much change in perspective to also express generalizations about the other relationship that holds between nodes, namely ordering. Where (27) says what things σ can immediately dominate, analogous grammatical statements like (28) could say what order those things must have. English syllables may begin with [s] followed by a stop or nasal, but may not begin with a nasal or stop followed by [s]. They may begin with an obstruent followed by a glide or liquid, but they may not begin with a glide or liquid followed by an obstruent.[42]

(28) a. [$_{\sigma}$ [s], Stop b. [$_{\sigma}$ Obstruent, [sonorant, continuant]

Another example of nodes whose subordinate elements are ordered is the foot. In some languages, the foot may dominate syllables in the order light-heavy but not heavy-light, and in some languages the opposite order holds. Of course, we must consider the possibility that such patterning is simply the consequence of the statement of the foot-construction rule in a particular language. We will not here attempt to establish that the notion of structural constraint must include both dominance and order, we will simply point out that this is a possibility that needs to be evaluated. What we have proposed is the addition of a second kind of computation, *structural constraint*, which regulate derivations as well as stored relations (saying what kinds of dominance relations can exist in phonological inputs).

(29) *Structural constraints*
State admissible relations between nodes in an immediate dominance relation
Include specification of allowed dominance relations between mother and daugher
May include ordering requirements on sisters

Again, this is not the complete theory of structural constraints, and further elaboration may be needed to say *how* they state admissible relations. It should be clear where in this hierarchy of concepts such additional claims fits: it is somewhere in the theory of computation, and not a subcase of "rule".

3.3.3. WHAT IS THE THEORY OF COMPUTATION (VERSION 1)?

We now need to take stock of our theory of computation. The epistemologically most basic computational concept is *grammar*, which embodies what the phonological component does. The theory is that a phonological grammar is a set of *rules* and *structural constraints*. A rule is composed of a *structural description* and a *structural change*, which maps a phonological representation to a phonological representation. The set of rules in a grammar apply in an order. The struc-

[42] We have inadvertently stumbled onto a notational problem, that we will need some way to specify "position withing a higher constituent", e.g. expressing the distinction foot-initial versus foot-final. This matter will be attended to in the next section on notation.

tural description of a rule is a representation which identifies those data-representation that are subject to the actions of the structural change. A data-representation *matches* a structural description iff the s.d. representation of the rule is included in the string. One aspect of the s.d. of a rule is a specification of the two entities which constitute the *focus* of the rule (a specification which indicates where the rule action takes place). The structural change of a rule is a representation plus the operations "insert" or "delete". The thing that is inserted or deleted includes a relationship between the focal entities, and may include an entity.

Structural constraints are a kind of representation, one which specifies an immediately dominating node α and an immediately dominated node β. The product of any rule which inserts a dominance relationship between α, β must be consistent with some structural constraint regulating dominance between α and β.[43]

This does not constitute the entire theory of the concept "rule" in phonology. Here are some not-yet answered questions about rules.

(30) What happens to representations that do not match the *s.d.* of a rule?
Rules can be optional: how does that affect the concept *s.c.*?
Rules can apply from left-to-right or from right-to-left: how does this happen?
Rules apply in a linear order: what exactly does that mean? How does the cycle fit with this requirement
Do rules (either the *s.d.* or the *s.c.*) involve auxilary concepts not part of data-representations – is there anything remotely like the SPE concept of rule schemata?
Does *s.d.* include the concept of non-presence (blocking) where presence of some entity means that the string does **not** match the *s.d.*? (A specific instance of the preceding question)
Can association lines cross; if not, why not?
What exactly do "structural constraints" do, if anything?

But perhaps this constitutes the entirety of the "basics" of phonological theory.

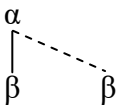
[43] This allows the possibility of seemingly contradictory conditions such as both α => β and β => α. It is not immediately clear what would motivate having both relations, nor is it immediately obvious that we need to care right now. The identification made here can be easily modified to change "some" to "all", if necessary.

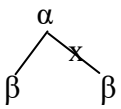
4. Formalization

If concepts are explicit, a notation can be developed where a concept is represented in a fixed way. Outside of linguistics, this is well-known in arithmetic where the concept "divide" is abbreviated as "÷", and the distinction "divisor" versus "dividend" is denoted by distinctive order of operands relative to the operator "÷". There have long been standards in phonology for notating data representations, which is unsurprising, since the core representational concepts are very simple. *Order* of entities is indicated by writing them on the same line, one before another.[44] *Dominance* is indicated by drawing a line between nodes. Conventions for representing data in SPE theory do not concern us, since we have not explored the SPE theory of representations.[45]

Rule concepts are harder to formalize, in part because autosegmental practice inherited an antipathy towards formalism arising from reaction against SPE theory, thus rule-writing practice became more informal. The notation expresses the theory, and the theory does not "interpret" the notation *qua* existential fundamental. When theorists seem to need to interpret notation, that is evidence that the notation has been reified – taken to be primary – and we need to "discover" what the notational symbols mean. In other words, we would have evidence that the underlying theory is not sufficiently clear and coherent. To repeat, FP and the ontological approach to theory puts concepts in first place and notation in last place. What exists in the language faculty is homologs of theoretical concepts, not notation for expressing concepts.

Notational practices in autosegmental theory are a blend of SPE notation and some specific autosegmental practices. In both approaches, insertion or deletion of a representational object is indicated via the standard "Ø→α" scheme for insertion and "α→Ø" for deletion – in SPE theory, entire segments are deleted and inserted, and in autosegmental phonology nodes are inserted or deleted. Autosegmental phonology also provides for insertion and deletion of dominance relations – adding or deleting association lines, where dashed lines (31a) denote an added dominance relation and crossed-out lines (31b) denote a deleted dominance relation.

(31) a. 

b. 

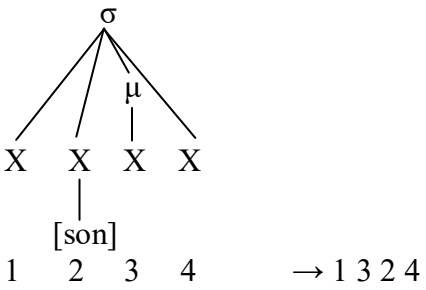
[44] Occasionally, elements which are *not* ordered are written on the same line. Generally this is because a representation with lots of node-types are difficult to draw – graphs like (20) are a nuisance to draw. Unfortunately, on occasion, it is unclear what an author intends by writing different things on the same line.

[45] There are two versions of SPE theory. Best known is "received" SPE, i.e. principles inferred in the literature based on SPE, including statements in SPE. The second is "technical" SPE, the formal model set forth in SPE's appendix on formalism. The two theories are quite different, indeed Received SPE is the conceptual statemenbt of the theory and Technical SPE is a mathematical interpretation of Received SPE. However, there are errors in translation.

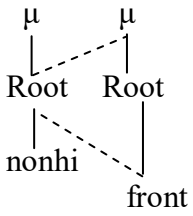
The notation is suboptimal in that it does not graphically recognize the unity of the insertion and deletion operations. The theory of rules proffered in Archangeli & Pulleyblank (1993) unifies both subtypes of insertion and deletion (taking "path" and "node" to be a parametric choice for a given rule, same as this theory), but does not use a special symbol to denote "insert" or "delete". Notation like (31a-b) can be criticized for obscuring the underlying concepts, but my intent is not to reform notational practices when there are other more important issues needing attention.

Another basic representational property that can be manipulated is precedence relations – languages may have rules of metathesis. Neither autosegmental theory nor SPE theory was particularly explicit about how such rules should be notated. Likewise, notational practices for fusions (e.g. /kj/ → [tʃ], /ai/ → [e:]) are not well-developed. A metathesis rule (CRVC → CVRC within a syllable) could be stated as (32a), and fusion of /ai/ into [e:] as (32b).

(32) a.



b.



A problem with either of these formalizations is that they contradict the premise that a rule performs a single action. It is possible that (32a) is a single action (re-numbering), but it is hard to see how (32b) is not at least two operations – and indeed, it may be that there are two rules that implement fusion. But since the conceptual underpinnings of fusion and metathesis are not entirely clear, we will leave the matter as it stands, as a problem in need of resolution.

4.1. The order of nodes in a rule

Interpretive difficulties may arise in notating the structural description of a rule, which refers to a class of data, and is not a complete specification of all of the segments that exist in forms described by the rule. Suppose we have an s.d. which refers to "a nasal followed by a high vowel". Does (33) correctly describe that class of representations?

(33)

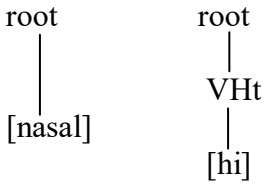
Since only same-type entities are ordered and (33) has no two same-type entities, this does not encode any notion of "following" segment. (33) is not different from (34a), just as (34a) is not different from (34b) (the difference between the representations can be measured with a ruler).

(34) a.

b.

(35) does encode ordering, via reference to things that are ordered – two root nodes. The potential problem here is that the strings described by (35) are not obviously encountered in any data-representation, since VHt is (probably) not immediately dominated by root, it is immediately dominated by Vocalic (which is probably immediately dominated by Place).

(35)



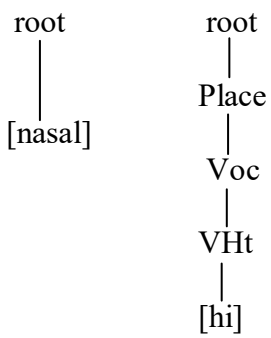
One solution to this matter is to fill in the full sequence of nodes that intervene between root and the terminal features that must be referenced by the rule. Another possibility – the position adopted here – is that (35) is literally correct, because (35) does not say that a root node *immediately* dominates VHt. This is the question considered in the next sub-section.

4.2. Dominance

FP and RSFP do not assume that feature relationships are universally set; but it is likely that languages have similar organizations, so it is likely that some language with the standard sequence root-place-Vocalic-Vheight-hi has a rule refering to "a nasal followed by a high vowel". Literally interpreted, (35) does not exist as a possible string in the language. The root does not immediately dominate VHt, so (35) seems to fail as an expression of the intended class of data strings.

One solution is to explicitly list all intervening nodes in the representation, as in (36).

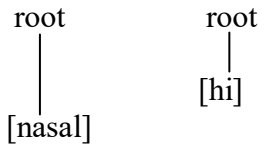
(36)



The proposition of (36) is "(if) $root_i$ precedes $root_j$ and $root_i$ immediately dominates [nasal][46] and $root_j$ immediately dominates $Place_j$ and $Place_j$ immediately dominates Voc_j and Voc_j immediately dominates VHt_j and VHt_j immediately dominates [hi] (then…)". This complexity and the underlying premise that the mind operates with simple concepts and propositions forces us to ask whether the complexity is an artifact of the theory. Have we identified the correct relationship, in saying that a line denotes immediate dominance? The artifact is clear, it is the premise that the non-precedence relation is *immediate dominance*. We have no need for immediate dominance as a grammatically significant concept.

Understanding association lines as refering to dominance, (36) simplifies to (37).

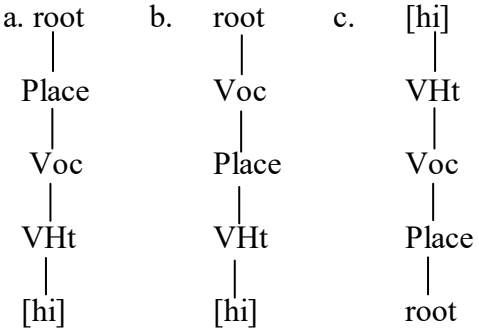
[46] Indices clarify when instances of a node type are the same or different tokens, and are not part of the theory.

(37)


This rule (fragment) embodies the proposition (fragment) "(if) $root_i$ precedes $root_j$ and $root_i$ dominates [nasal][47] and $root_j$ dominates [hi] (then…)" – we don't care if there is some node that $root_j$ dominates which stands between that root node and [hi]. The only reason to include the intermediate structure in (36) was the thought that association lines encode immediate dominance.[48] The reason we thought that, is that we only draw lines between a node and what it immediately dominates.

What *should* the theory say about immediate dominance and dominance? The absolutely weakest[49] claim would be that both *dominance* and *immediate dominance* are completely independent primitive relations in phonology. We can immediately reject that possibility without even considering facts of language, since *i.d.* is definitionally a subtype of dominance – one where no other node stands between higher node α and lower node β. Clearly, immediate dominance is not primitive, but it still might be a convenient fact that needs to be referred to in grammar. Should we admit the possibility of having, as an option for a representation, a condition like "and no node γ is dominated by α and dominates β", that is, can we include bare dominance plus an expasion that identifies immediate dominance in the arsenal of rule relations?

Any such expansion of the theory must be justified. What would be the utility of *i.d.* distinct from dominance in phonological theory? Suppose we had contrasting structures similar to the specification of a high vowel in (36). If association lines only encode dominance, what guarantees that (38a) is distinct from (38b) or (38c). Clearly, the representations differ and the basis is what immediately dominates what.

(38)


These graphic structures translate trivially to a minimal set of unique dominance relations.

---

[47] Arbitrary indices are only included when it is necessary to state that two instances of a given node type are either the same of different tokens.

[48] This same conclusion is encoded in most exemplars of autosegmental rule-writing practice such as Clements (1985), Hayes (1986) and is explicitly connected to the concept "dominance" by Clements. Since I knew in advance that the solution was that lines encode dominance, I did not earlier press the analysis that they encode immediate dominance, so this was a bit of a setup.

[49] By "weakest" I mean the claim that most expands what a phonological grammar can do.

(39) a. D(root,Place) & D(Place,Voc) & D(Voc,VHt) & D(VHt,hi)
b. D(root,Voc) & D(Voc,Place) & D(Voc,VHt) & D(VHt,hi)
c. D(hi,Vht) & D(VHt,Voc) & D(Voc,Place) & D(Place,root)

In both (38a) and (38b), root dominates Voc, but it is only in (38b) that root *immediately* dominates Voc. It is obvious from (39) that the two representations are distinct, and there is no need to include "immediate dominance" as a device for encoding distinctness in the arsenal of representational machinery.

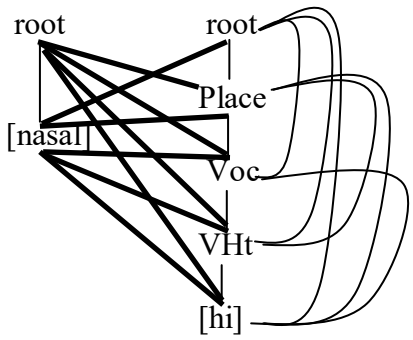
Might the theory of computation be the source of justification for adding this concept? If we encounter a phonological rule treating (38a) and (38b) differently, where the rule only applies to Voc *immediately* dominated by root, we would have evidence for such an expansion of the theory. Until we find clear evidence for such a rule, we do not add that concept to the theory. At the end of this section, we will briefly consider what such evidence might look like.

In arguing against adding "immediate dominance" to the arsenal of phonology, we have now uncovered a lurking need for *some* revision to the theory of phonology. Remember that in §3 we adopted a very simple theory of representation and rule, where a rule is a resentation, one which performs an action, and the algorithm for applying a rule to a datum is simply that if the rule "matches" the string, the action of the rule is performed. We now have to re-scrutinize "matching". §3.3.1 posits that a match occurs when the nodes and relations in a rule are included in the nodes and relations present in the datum – if the rule requires the relation "root dominates hi", the relation "root dominates hi" must exist in the datum. Inspection of (39a) reveals that this relation is lacking in the assumed datum. Therefore the proposals for representations (both data and rules) and the concept "match" fail. How do we resolve this problem?

Abandoning simplicity is not a solution, it is a repudiation of the framework. But simplicity is a comparative measure, so maybe one complication is required, to achieve greater simplicity elsewhere. Let us expand (39) so that all dominance relations are included, including those which can be computed from the minimal set of dominance relations (which are, in effect, the immediate dominance relations). A theoretical compulsion to include all dominance relations in a representation is one complication. This is a problematic complication, since it is not clear exactly what principle is to be added to the theory. Superficially, it looks like there might be a connexity requirement on nodes of different types, where every node of a different type must have a dominance relation with every other node of a different type – this is simply "total ordering", imposed on dominance instead of precedence.

The problem is that a connexity requirement is completely incompatible with autosegmental relations. This is graphically clear if we chart (36) with a dominance relation between *all* un-ordered nodes.

(40)



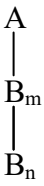
The curved thin lines are relations in a segment compelled by a connexity requirement *within* a segment. The thick straight lines are compelled by that requirement *between* nodes of two segments (this is one sample of such dominance relations). Since what impelled the search for a new theoretical proposition was the need to say that root dominates [hi] within a segment, we might attempt an alternative proposition restricted to the segment: all nodes dominated by the same root node must be in a total dominance relation, so only the thin curved and straight relations in (40) are compelled. Informally, we are trying to require that "all nodes within a segment have some dominance relation with all other nodes within that segment", but how do we define "in a segment"? It is not simply by reference to "is dominated by the same root node" – this is exactly the problem that we are trying to solve, since as we see in the list of (39), [hi] is *not* dominated by root, so it is technically not "in the same segment", and therefore is not compelled to enter into a dominance relation with Place, and so on. If we add a requirement that all nodes must be associated to *some* root node, plus the total dominance requirement for nodes "in a segment", then we have a means to enforce an exhaustive specification of dominance relations. The only problem is that this is empirically wrong (or counter-indicated): it fails in the face of floating features, ones that are not linked to any root node. Repudiating autosegmental representations as well as the desideratum of simplicity is an alternative, but one that is outside the scope of this work.

Rather than pursue some set of further modifications to the theory which would compel presence of superfluous domination relations, an alternative solution is to focus on the concept of "match", and to recognize that we were hasty in characterizing when a datum "matches" a rule representation. In 3.3.1 we observed that a "rule-representation" refers to a class of "representation-representations" and we attempted to explicate what that means via a concept of formal matching. The flaw is that the attempt to characterize "refer" in terms of a subset relation between two representations was wrong. As a rule representation, the expression "root dominating hi" refers to numerous representations including (38a), (38b) and many other representations. The concept "refer" is not a peculiar construct of phonological theory, it is a broad epistemological one. Remember that generative phonology is a theory of the mind. Whatever "refer" means in the broader cognitive sense, that is what "refer" means in phonological theory.

Without giving a full philosophical account of "reference", we know that symbolic expressions refer to things, and things do not refer to things (except insofar as a symbolic expression is a thing). The essence of "referring" is "picking out", where "dog" refers to certain things and picks out dogs as distinct from other things, and "fry" refers to certain situations. The expression "D(root,hi)" refers to or picks out (38a) because "D(root,Place) & D(Place,Voc) & D(Voc,VHt) & D(VHt,hi)" further logically *implies* the dominance relations D(root,Voc), D(root,VHt) and D(root,hi), owing to the transitive nature of the relation "dominate". The "match" between a rule expression and the class of stored representations which are *referred to* by the rule is a superset of the nodes and relations in the stored representation, including as well those nodes and relations which are logically implied by those relations. Of course many questions are raised in invoking broader cognitive notions – what exactly does it mean to "refer", what does it mean to "infer"? There is no reason to believe that the answer to these questions differs essentially for phonology, syntax, visual identification, auditory identification or any other form of cognition. The solution to the problem of matching as "inclusion of relations" is that the matching computation includes both stored relations and those that are logically inferrable from stored representations. The simplest account of (35) is to better articulate "match" in rule theory.

We have not directly established whether immediate dominance is also a phonologically significant concept. We should still sketch the kinds of considerations that could yield an answer.

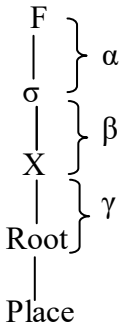
A theoretical instance of a rule distinguishing A **immediately** dominating B from A non-immediately dominating B would be structures like (41a-b). A dominates both token *m* of B and token *n* of B, but it only immediately dominates token *m*.

(41) a. 

b. 

Does "A dominating B" refers to B_m or B_n, or both? This is the A-over-A principle (Chomsky 1964). The issue comes up only if phonology has recursive structures. Such structures have been proposed in early metrical phonology with recursive binary tree structures having an in principle unbounded number of levels. Recursive arboreal structure has been replaced with a small number of distinct constituents such as syllable, foot, perhaps cola, and word, with an orthogonal indication of "prominence", so that recursion is not necessary to account for stress facts. There are ongoing controversies over recursion when it comes to accounting for syntax-phonology interaction – at this point the evidence for phonological recursion in that domain is not compelling. Finally, there are proposals for sub-segmental recursion in UFT, where feature A may dominate A, e.g. in the account of vowel height (Parkinson 1996), and proposals for recursive structures in Element Theory (which is not easily reconcilable with the structures assumed in feature-geometric models of representation). Even if recursive structures exist in data representation, it is not obvious that rules treat the relationship between A and B_m vs B_n differently in (41a-b) – this is an empirical area that needs to be researched. In other words, it seems that with a correct understanding of *dominance*, there are no interpretive difficulties pertaining to the conventional graphics for data-representations and the *s.d.* in a rule – they are interpreted the same way.

What about the fact that, in standard notions of prosody, F immediately dominates σ, and this fact is encoded in structural constraints of the type proposed in (25)-(27)? In that discussion, the applicable constraints were stated in terms of immediate domination, so would it be possible for F to immediately dominate [Place] in this scheme, in lieu of a special rule F => Place? The idea behind structural constraints is that the dominance relations expressed in the set of constraints must always be satisfied, and other dominance relations are not possible. When constraints α, β, γ respectively allow linkage of F to σ, σ to X and X to root, what permits F to (non-immediately) dominate X, Root or Place?

(42) 

Given transitivity of dominance, if D(F,σ) and D(σ,X), it follows that D(F,X): and D(F,Root) and D(F,Place) follow via transitivity from X => Root and Root => Place. In the case where the structural constraint subset of the grammar contains only a listing of immediate dominance relations, non-immediate dominance relations follow by transitivity. Since F can dominate σ (F => σ), sub-representation α is allowed, and since σ can dominate X (σ => X), sub-representation β is allowed. D(F,X) is possible by a chain of admissible node relations as in (42). But F immediately dominating X is not allowed in (43) because there is no structural constraint F => X, nor is it admissible via transitivity from one of the existing constraints. X immediately dominating σ likewise is not an existing constraint, and does not follow by transitivity from existing constraints.

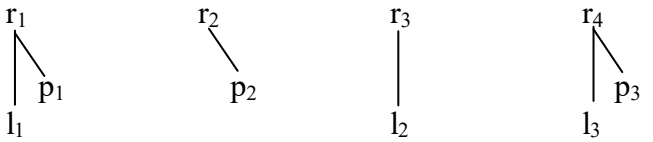
(43) * F
|
X
|
σ
|
Root
|
Place

It thus seems that only a single notion of dominance is needed in phonological theory. A rule statement like (37) only says "root node$_i$ precedes root node$_j$ and root node$_j$ dominates [hi]", which is true and matches a data representation like (36).

4.3. Precedence and matching

The preceding analysis of dominance, and the discovery of a need to modify the notion of match, compels us to ask what analogous things might be said about precedence. In particular, is there anything about the notion "precede" in the theory of rules that might flow from the fact that rule-representations refer to broad classes of data?

A refinement of the concept 'precede' does follow from the discussion of 4.2. Consider the following schematic datum.

(44)


This encodes the minimal set of precedence relations in (45).

(45) $P(r_1,r_2)$ $P(r_2,r_3)$ $P(r_3,r_4)$
$P(p_1,p_2)$ $P(p_2,p_3)$
$P(l_1,l_2)$ $P(l_2,l_3)$

The following precedence relations are inferable from (45).

(46) $P(r_1,r_3)$ $P(r_1,r_4)$ $P(r_2,r_4)$
$P(p_1,p_3)$ $P(l_1,p_3)$

As far as we know, no language maintains a representational contrast between a version of (44) stated as (45) versus a version of (44) that is the union of (45) and (46). For the needs of stored representations, nothing tells us whether the applicable ordering concept if precedence of immediate precedence.

Turning to the needs of rule-representations, assume a rule which has the following structural description.

(47)


What data representations does this refer to? The answer depends at least in part whether "p p" refers to "p preceding p" or "p *immediately* preceding p". The answer, standardly assumed in autosegmental theory and discussed in Odden (2020), is that p_i must immediately precede p_{i+1}. (47) will apply to r_1,r_2; p_1,p_2 in (44) but it will not apply to r_1,r_4; p_1,p_3, because r_2,r_3,p_2 are "in the way". What does this mean for "match" in the theory of rules? One possibility is that inapplicability of the rule to r_1,r_4; p_1,p_3 comes from the fact that $P(r_1,r_4)$ and $P(p_1,p_3)$ are lacking in (45): but furthermore, we must not *infer* such ordering relations in computing the match. That is, dominance relations may be inferred for matching but precedence relations are not. But there is a much simpler solution: "P" is the relationship "immediately precedes", not "precedes". If α dominates β and β dominates γ, α dominates γ, because dominance is a transitive relation. If α immediately dominates β and β immediately dominates γ, α does not automatically immediately dominate γ: immediate dominance is not a transitive relationship. Likewise with precedence and immediate precedence. We may now conclude that the concepts of hierarchy and ordering invoked by grammar are in fact *dominance* and *immediate precedence*. We will reconsider precedence and immediate precedence in §5.4.2, but will reconfirm that there is no computational need to additionally encode general precedence.

4.4. Floating nodes and denial

Here is another way in which rules are more than general representations. One indication that we need something else is the existence of rules where a node must be identified in a rule as "empty; unassociated; free; floating; unspecified". An unsyllabified segment is one which is not dominated by any σ; a floating tone is one not dominated by any μ; a place-unspecified consonant is one not dominating a Place node. The literature contains ample examples of rules operating only on empty / unassociated structures, so there is a fact to be accounted for.[50] A related concept is the notion of "feature-changing" versus "feature-filling", see Reiss (2003) and ?Reiss & Volenc?. Let us see why "floating" elements are a problem for rule theory. Consider the following data representations, involving relationships between moras and tones.

[50] I discuss the matter empirically below, and do not require the reader to have faith that the assertion is true. For the moment, we can consider this a logical exploration of the consequences of such a condition on rules.

(48) a. H b. μ | H c. μ

As a data representation (48a) is a floating H, (48b) is a H linked to a mora, and (48c) is a toneless mora. But as a rule representation, (48a) refers to both (48a) and (48b) *qua* data representations; (48c) refers to both (48b) and (48c) *qua* data representations. (48b) is the only rule representation that refers to just one class of data representations – H linked to a mora. In rule representations, specified elements are not interpreted as exhausting the data representations identified by the rule representation. Specifying the s.d. of a rule as "H" does not refer to "A H which is the only node in the entire representation". What we require, in order to identify "H not linked to μ", is a way to *exclusively* identify a H not linked to μ; or, to say that the H is not a match if it is also linked to μ.

The difference between a syllabified segment and an unsyllabified segment, or a floating tone and an associated tone, is that in the former cases, D(σ,root) or D(μ,T) as propositions about the data representation are true. The relationships D(σ,root) or D(μ,T) exist within the set of nodes and relationships that constitute the actual data. But in the latter cases, when referring to unsyllabified segments and floating tones, we are asserting that D(σ,root) or D(μ,T) is false. The propositions say "¬D(σ,root)", "¬D(μ,T)", i.e. "it is not the case that there is a dominance relation between the root and a syllable". There is no data-representational equivalent of "¬D(σ,root)" within the set of nodes and relationships – there is no "thing" in any datum that corresponds to that rule expression. The previously-noted differences in ontological status of "representation" in rules and data is thus not just a matter of a less-concrete interpretation of "representation" as applied to rules and data, it translates into formal differences between rules and data. The s.d. and s.c. of a rule are propositions that refer to classes of representations, such as "D($root_i$,hi) & D($root_i$,coronal) & D($root_{i+1}$,nasal)". This immediately raises the question, what other mechanisms do propositions bring to bear, in describing classes of data representations?

Formal logic provides a rich set of tools for expressing propositions. The fact that a certain logical tool may have utility in a domain like mathematics does not mean that the same tool is valid as a theory of phonological rules. Whatever machinery is imputed to phonological theory, that machinery has to be justified linguistically. We have admitted ordering and dominance relations since they are essential to describing virtually every phonological facts: we have admitted conjunction of relations for the same reason (a segment is not just a single dominance relation, it is a conjunction of a range of dominance relations; the s.d. of a rule refers to some conjunction of properties). We have provisionally accepted the notion of propositional negation. Perhaps other relations will be shown to be necessary.

Returning to the question of notation – the form of rules – the issue which set off this proposed modification of the theory is, how do we notate negation in rule-formalism? Autosegmental practice did not adopt a uniform notation for referring to lack of association. Practices in the era when rules were formalized were variable, and often relied on intuition about intent to correctly interpret. C′ would generally mean "C which is not syllabified", H′ would mean "floating H", and V′ would mean "Vowel lacking a tone". See Steriade (1982), Pulleyblank (1983), Archangeli (1984) for examples of notation for expressing "lack" (of association) in rules.

We could start very simply by conventionalizing negation with the symbol "~", a symbol that has been used in rule statements to assert "not". Before proceding too far (even one step further, in fact), we need at least one specific instance of "not associated" in rule statements. Draw-

ing on discussion in section Q^{tbw}, there is rule in Logoori which applies between words, linking an unassociated tone to the first mora of the word to the right. The H tone in bold is a syntactic "linker" present between words within NPs in certain constructions, underlyingly a floating H. That tone links to the penultimate syllable of a preceding head of the NP if the noun is toneless as in (49c), and it deletes when the preceding word ends with a H-toned vowel as in (49d). Otherwise, a floating H between words associates to the first syllable of the word to the right, as in (49a-b).

(49) a. H **H**
avadóto vǎra
'those infants'

b. H **H**
avíísɔkɔrɔ vǎra
'those grandchildren'

c. **H**
avǎándɔ vara
'those people'

d. H H **H→Ø**
vǎáguugǎ vara
'those grandfathers'

The tone-association rule must specifically identify floating tones as the target. In this language, a linked H does not shift or associate to any vowel of the following word, for example *avǎándɔ vara* or, in a context lacking the linker H, *vǎáguugǎ vɔza* 'only grandfathers', *avadóto vɔza* 'only infants'. The traditional way to write this rule in autosegmental phonology is (50).

(50) H′
[ω μ

In this context, the second Hs in (49a-b) are subject to (50) because the H is present, but there does not exist a dominance relationship between that H and any mora. The problem is that (50) doesn't mention any mora that the H lacks a dominance relation with. We can try to remedy this by restating the rule:

(51) H
~/
μ [ω μ

This is an attempt to say "there is no dominance relation". But this does not correctly express the concept of "floating tone", since it requires there to be a preceding mora to which the tone is *not* associated. As it happens, this works empirically in Logoori, since the floating tone that is docked is always precede by some word with a mora. The point is, this attempt to develop a notation ran aground, because we do not yet have a proper understanding of what concepts of negation are to be encoded in rules. Do we also need to refer to "lack of nodes" (perhaps as a way of stating blocking effects – X spreads "unless blocked by Y")?

The primary point of this subsection is to give some support to the idea that rules are slightly different from data, in that some propositional concepts such as "negation" may be relevant to rule statements, because rules describe classes of data, and some properties of proposi-

tions do not make sense as properties of data. Secondarily, the point has been to indicate why it is premature to make a strong claim about how negation is manifested in rule statements. The next section briefly considers another relationship that needs to be expressed in rules, but which plays no role in data.

4.5. Identity

It is well-known that a variety of rules involve computing the identify of two segments with respect to some subset of properties – this is frequently labeled "the OCP". See Odden (1988) and Reiss (2003b) for examples, and Odden (2013) for theoretical discussion in the current framework. The fact which a theory of rules must account for is that there are computations where it matters either that two segments are the same in terms of a set of features, or they are different. For example, some syncope rules are blocked when the consonants preceding and following a vowel are identical for a subset of place features, termed "antigemination" following McCarthy (1986). Some syncope rules only apply when the surrounding consonants are identical for a set of place features, termed "anti-antigemination" following Odden (1988). The proposal set forth in Odden (2013) is, simply, that "identity" is a relationship that can be part of a rule. As a first attempt to include the Identity relationship, we can say that two nodes are identical if they have the same type, and all of the the nodes that they dominate are also identical. Thus [p] and [m] have identical Place nodes, but not identical Root nodes.

Defining identity is not trivial, as the OCP literature shows. Postvocalic spirantization in Biblical Hebrew does not apply to the first of a sequence of identical consonants, e.g. *karat-ti* (**karaθti*) 'I cut', cf. *lamað-ti* ← /lamad-ti/ 'I learned' – a voicing difference makes the sequence *dt* non-identical. In this instance, identity is computed with respect to the entire set of nodes dominated by Root. Modern Hebrew, on the other hand, appears to insert [e] between /d,t/ and following /t/, that is, identical consonants except for voicing (cf. *kafar-ti* 'I tied', *hitlabaf-ti* 'I got dressed', *mazag-ti* 'I poured' but *lamad-eti* 'I learned', *fafat-eti* 'I judged'. It is clearly beyond the scope of this work to undertake a systematic search of all putative OCP effects, but it is well within the scope of this work to raise the relevant question regarding identity effects. Namely: what range of identity relations, or non-identity relations, are manifested in rules? Do rules ever compute whether two vowels are identical in height, or nasality; do rules ever compute whether the entire feature structure under Place for two consonants is exactly the same?

A simple claim regarding "identity" as a relation in phonological rules is that a rule may, as an option for rule statement, require two nodes of the same type to be "identical". In the case of a terminal privative feature, requiring α_i and α_j to be "identical" is the same as saying that α_i and α_j exist. If a feature is a value-attribute pair or if a node is non-terminal, requiring two segments to be identical w.r.t. [voice] means that the segments either both lack voice, or both have it and if they both have it their values are both + or both – (if those are the values assignable to [voice]). Requiring two segments to be identical for Place likewise means that both lack a Place specification, or they both have one and then the content that they dominate must also be identical. Having a clear understanding of the facts regarding identity relations would give us a basis for reasoning about the details of the relationship "identical" as an aspect of computation, for example is a placeless segment 'identical' with every configuration of Place-dominated features; is a placeless segment 'identical' with another placeless segment? While we do not have answers to these questions at the moment, there can be no serious doubt that the relationship "identical" is

also an aspect of rules – it constitutes a proposition that describes classes of representations that a rule applies to. We revisit identity relations in evaluating theories of assimilation in §5.4.1.

4.6. Edges

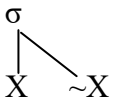
If some node α can dominate more than one token of a node of type β, those instances of β are ordered, and some β_i is the first β within alpha, or the last β within α. This is a property that rules may refer to, for example the last consonant in a syllable may be devoiced, the last vowel of a word may be deleted, or the first consonant of a syllable may be somehow modified. How do we refer to the fact of being the first or last elements within a constituent? It is not enough to express "last segment of the syllable" as (52): this simply says "a segment which is dominated by a syllable", and does not distinguish first, last, or medial.

(52)


One approach pursued in early autosegmental studies was to place "x" or "=" on or near the line with the node in question, and draw a line between the higher node and that symbol.

(53)


This simply means "X_i is dominated by σ, and there does not exist X_{i+1} which is dominated by σ". Given the discussion in 4.4, this could be stated as follows.

(54)


Another way of expressing this is with labeled directional brackets:

(55) $[_\sigma$ X

Again, this refers to "X_i is dominated by σ, and there does not exist X_{i+1} which is dominated by σ", but calls on a special notation standing for "there does not exist__ which is dominated by __". It is not too important which notation we use, as long as the underlying concepts are justified. In this case, we appear to need reference to the lack of a specific node or relation, a notion that we have had some reason to consider already. Perhaps "edge" is just an application of the concept of denial in a rule.

4.7. The calculus of the metatheory

While we are most concerned with how to write particular rules in a formal notational system, consistent with our adopted theory of grammar, there is also a desideratum of reducing the con-

tent of the theory itself to a notational system. This is particularly important because FP holds simplicity to be a prime adjudicator between competing theories. In order to measure the complexity of two theories, you need comparable units to count.

Take the following two competing theories as an illustration. Theory (56a) is simpler than (56b) – just count up the things in the two theories.

(56) a. $\forall x(A(x) \supset B(x))$
b. $\forall x((A(x) \wedge \neg C(x)) \supset B(x))$

A problem with this approach is that it does not correspond well to the cognitive thing that the formulas are intended to recognise. The first is supposed to encode the idea that having the property A causes a thing to have the property B, but it does so by saying "for all x, if A is true of x then B is true of x". Conventions of mathematical formalization don't necessarily reflect cognitive processing. Or, to put it another way, some mathematical formalization must surely be optimal for the expression of theoretical concepts, but not necessarily *this* one. In an extensionally-focused metatheory, where one is only concerned with the sets of strings that are admitted by a theory, it is not important for there to be a close analogy between the propositions of the theory and the mathematical interpretation of the theory, but in an intensionally-focused metatheory, it is essential that the mathematical model of the theory faithfully translates the concepts of the theory. For example, the concept of "directed acyclic graph" is a reasonable basic model of domination relations. Importing *all* of graph theory is not a good model of phonology, because graph theory contains concepts that are superfluous to phonology, for example "order" (of a graph: the number of nodes), "loops" and "adjacent" (immediate dominance).

The second and much larger problem is that the content of A and B (etc.) must ultimately reduce to some minimalist lexical calculus. For instance, phonological theory needs a concept of "precedence" (immediate or otherwise), because [it] and [ti] are distinct and differ in precedence relations. What does it formally mean to say that [i] precedes [t] in [it]? We could say that the segments in a representation are in correspondence with the set of natural numbers, so that node n_j precedes n_k iff $j<k$. But that implies that "the set of natural numbers" is a valid phonological concept, which is surely false (in addition, the determination that $j<k$ at least in some theories of numbers is based on a successor function, raising the spectre of circularity). From a phonological perspective, the solution is simple: precedence is an axiomatic relation in grammar, just as dominance is. Those with an interest in translating phonological theory into non-phonological mathematical concepts can pursue that enterprise, as well.

Axiomatic concepts tend to be very theory specific. For example, rule theory tends to take "rule" to be an axiomatic concept, and contraint-based theories tend to take the concept of "constraint" to be axiomatic, therefore a conceptual comparison of rule theory and constraint theory is impossible because the fundamental units of these theories are in a black box. In classical philosophy (thus epistemology), an axiom is not an arbitrary declarations stipulated to be true, it is a *necessary truth*, one worthy of general acceptance. A *very* partial solution to this problem could be (as advocated by the present ontological-decomposition approach) to locate that concept which OT and RBP have in common, namely "computation" – the theories differ in the propositions they posit about a "computation", but they agree that input-output relations are

the product of such a thing.[51] Or, consider the concept "dominance" adopted in the present theory. Within generative phonology, this cannot be treated as axiomatic, insofar as SPE theory has no need for such a concept. Autosegmental phonology introduced this notion into the theory. Given that one of the driving motivations for this paper is to methodologically facilitate theory comparison, invoking axiomatic concepts can be dysfunctional with respect to the purpose of setting forth the ontology of a theory. The remedy for this problem is simply to recognise when a result specific to one theory is stipulated as being foundational for all theories within the science.

4.8. Summary of formalization

The point emphasized here is that formalization concepts of phonological theory requires a clear understanding of what concepts are being formalized. In the attempt to formalize at least an elementary theory of phonology, we have uncovered conceptual unclarities: but we have also uncovered and eliminated some unclarities and error in theoretical concepts.

[51] An additional problem is that the thing that a single rule does and the thing that a single constraint does can be called a "computation", but so can the entire structured set of things done by a collection of rules or constraints, and there is no terminology distinguishing the two senses of computation.

5. Justification and theory evaluation

To summarize the preceding results, the program of rational theorizing depends on rational theory evaluation. The primary principle of theory evaluation is correctness, which entails that a false claim is rejected in favor of a true claim. Within the realm of competing true (non-false) claims, the consideration of simplicity dictates that a more complex theory is to be rejected in favor of a simpler theory. What is the role of justification in theory-selection? It lies in the first desideratum, the partitioning of claims into true claims and false claims. In this section, we scrutinize the hierarchical nature of knowledge and the method of inducing high-level theoretical propositions based on simple sense data, where "justification" provides the logical transition between observation and theory.

We must first be clear about what "truth" is. A proposition is true if it correctly describes a fact. As framed by Aristotle (*Metaphysics* 1011b25): "To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, and of what is not that it is not, is true". This is a cognitive view of "truth", meaning that the truth of a statement derives not from an intrinsic property of a sequence of symbols on paper, it derives from a mind comprehending a relationship between a proposition and a fact – some aspect of the universe. When that happens, we traditionally say that the proposition "corresponds" to, or describes, reality. The proper method of establishing such a relation in science is to rely on logic, applied to facts. Other methods have been employed to identify "truth", for example one may intuit how one feels about a proposition and use an emotional reaction to tell you what is "true", but emotions are not a source of knowledge about facts (other than questions about one's emotional state). You can use hallucinogens or animal entrails as a tool to tell you what is "true", but these are highly unreliable tools. The judgment that a claim is possibly true, or probably true, is based on evidence, and we need a theory of what kind of evidence is needed to justify the claim.

Evidence is the logical relationship between a proposition which constitutes the conclusion under scrutiny, and some set of propositions that are firmly established conclusions about the universe. For instance, the statement "all of the doors to this room are locked" could be supported by statements such as "I turned the knob on door 1 and it would not turn" or "I just locked door 2". If these facts are all of the evidence that supports the claim, and you know that there are 30 doors to the room, you would not say that the claim is probably true. The level of support for the claim is way too weak, and it is not reasonable to leap from "2 doors are locked" to "30 doors are locked". Superficially (so superficially that this point may be hard to follow), if there are only 2 doors to the room and we have these same true statements as evidence, it might seem that the claim is more than probable, perhaps it is certain. The flaw in assigning certainty to the conclusion here is that the available evidence fails to disprove alternatives.

One alternative response to the door 1 fact is "Maybe there was some other reason that the knob didn't turn". Another response is "There was a magical spell on the knob". A third hypothesis is that the cylinder was temporarily jammed by an obstruction. Magical claims have no potency in science, and deserve no effort at disproof – such claims are outside of the realm of reason. The same is true of the less blatantly irrational claim "maybe there is some other reason". It is impossible to test the truth of such a claim, and saying "maybe it's not true" cannot convert an arbitrary claim into a probable claim which creates a permanent shield against knowledge. Such responses should be disregarded. That is why the first two responses were dismissed as mere "responses". The third hypothesis, which is worthy of consideration, is that there exists some other obstruction on the mechanism. This is a verifiable claim grounded in ordinary experience. This constitutes a legitimate alternative hypothesis, although it is not yet supported by direct observation. But it is a claim that can easily be tested. The reader can, as an exercise, construct other non-fanciful alternative hypotheses addressing the status of the assertion "I just locked the door". Alternative explanations play a fundamental role in evaluating knowledge claims. It should be apparent that we need a theory of evidence in its relation to justification.

The first question about evidence is, what is the logical relationship between a conclusion and its factual support, when devising a scientific theory? We have some number of observations about releasing bricks and observing that they fall down (not up, or sideways). We arrive at a theory of a concept "force" whereby the earth exerts a force on objects that makes them fall. There is some connection between the observations, and our theoretical abstraction. What is it that the theoretical abstraction conveys? It is something about a property of the particular objects, and different objects have different properties. The properties of rules are not the same as the properties of features or the properties of the immediate precedence relation. Properties are the facts about a thing which determine what the thing does, which means that the logical relation underlying scientific evidence is *causation*. Certain facts (causal properties) cause other facts (observed events or states).[52] We are justified in arriving at the separate concepts of segment, rule or immediate precedence, because they "do" different things – they enter into different causal relations. Justification is the system that proves the connection between real-world things and a causal claim. The standard of evaluation is the viability of alternative causal claims.

An exhaustive reduction of the claims of theoretical phonology to undeniable experiences would be impractically complicated, but it is in principle possible, and it is important for phonological theorists to understand these foundations and the transitions to the higher-level knowledge that are scientific theories. To illustrate this method of reasoning, we consider some examples of building towards high-level knowledge, a theory of phonology, from simple experience. In the attempt to identify very simple knowledge, we will be reminded that even simple behavioral observation presupposes other knowledge. The point is not to inundate the reader with an infinite regress of unanswered questions, it is to identify some of the alternative causes which the reader may not have considered, when evaluating a theory.

---

[52] See Falcon 2019, citing Aristotle *Posterior analytics* 71 b 9–11 as "we think we have knowledge of a thing only when we have grasped its cause", or the fuller Mure 1925 rendition "We suppose ourselves to possess unqualified scientific knowledge of a thing, as opposed to knowing it in the accidental way in which the sophist knows, when we think that we know the cause on which the fact depends, as the cause of that fact and of no other, and, further, that the fact could not be other than it is. Now that scientific knowing is something of this sort is evident – witness both those who falsely claim it and those who actually possess it, since the former merely imagine themselves to be, while the latter are also actually, in the condition described. Consequently the proper object of unqualified scientific knowledge is something which cannot be other than it is".

5.1. Establishing the nature of human language phonology

The theory of human language phonology is a body of claims about rules and representations, as briefly sketched in §§3-4. The empirical fodder for that theory are the rules and representations in particular languages. Where do we get the rules and representations of particular languages?

5.1.1. PROPOSITIONS ABOUT INDIVIDUAL LINGUISTIC DATA

This is where we start to question "data", which is the foundation of linguistic theory. Claims of phonological theory rest on logic which organizes out knowledge of the real world, and we deal with the real world via "data". This sub-section provides a basic walk-through of the nature of data, and the following sub-section subjects the concept to closer scrutiny.

Linguists take data for granted as self-evident truths, which is not an entirely wrong view. It is a datum that the Arabic word for 'vertebra' is [fiqra].[53] What is the source of that datum? We start with something closer to direct experience rather than inference: the proposition "Smith said [fiqra] when I asked him what the Arabic word for 'vertebra' is", which is a report of an actual experience. As the experiencer in this scenario, I do not entertain outlandish alternative hypotheses that Martians transmitted brain waves into my head and deluded me into thinking that I had a perceptual experience, nor do I fantasize that it was really an actor wearing a Smith mask. These are *unreasonable* alternatives, outside the scope of scientific methodology. Almost all of my statement falls within the realm of that which cannot reasonably be denied (note that I didn't assert that Smith is a speaker of Arabic), *except* the linguistically-pertinent datum [fiqra]. That is a conclusion, a perceptual judgment of physical sound: what justifies that judgment?

Every science requires some theory of its basic instruments. Chemists and physicists don't reach conclusions about temperature associated with a reaction because they "feel" that the temperature must be so-and-so. They have a complex system of internally-consistent propositions allowing them to objectively measure a temperature. There is,[54] likewise, a complex system of principles that linguists call on to validate the conclusion that Smith said [fiqra]. The inference from physical experience to discrete categorization depends primarily on the ability of people to auditorily judge acoustic similarities, according to an objective standard. The details of this process are developed within the specific linguistic sub-sciences of field work and phonetics. A simple model of science of that sub-domain of linguistics is that we assemble examples of [f], [ɪ], [q] and so on. This may involve observing the behavior of speakers of particular languages, where we establish a convention that [q] stands for that sound appearing in various words of Arabic, these words of Lushootseed, such-and-such words in Berber, other words of Cuzco Quechua, certain words of Somali, and so on. Ideally, there are recordings available for training and calibration purposes. In writing "q" for Arabic [fiqra], we appeal to that standard, concluding that the utterance was not [fikra], [fiχra], [fiɢra], [fiɢ̊ra] or [fiqʰra].

There are numerous reasonable alternative propositions regarding this situation, for example "Smith said [fiʔra] which I interpreted as [fiqra]" – that is, the transcribed datum is an error of judgment (a measurement error). To decide between these propositions, we look at the

[53] Assertions about "Arabic" generally have to be sociolinguistically hedged since, like English, there are many dialects of Arabic.

[54] Is, or should be, a point which holds throughout this work.

"machine" which performs the measurement, just as we check the thermometer to see if it is functioning properly. Of course, we don't check out measurements if we are unaware of the possibility of error, so I assume that the scientist making these recordings knows of the possibility of error. Plausible discrepancies between physical utterance and measurement can be reduced to one-time errors or systematic errors – a nonce lapse of judgment vs. being systematically misguided in perception of [q] versus [ʔ]. These matters are well-known to field workers, and are guarded against in the case of occasional errors, using the standard practice of repeated elicitation. Systematic inaccuracies are a more significant problem, in that self-correcting procedures do not rapidly remedy a problem with a structurally-faulty machine (a field-worker who does not hear a distinction correctly). Independent verification may be necessary.

If we can rule out measurement error, we have a single datum, which by itself has little scientific value. A collection of similar data could justify a higher-order conclusion which is not just a singular statement of experience. Parallel observations could lead to higher-order conclusions, for example given repeated observation of the same type, we might induce the open-ended conclusion "Smith says [fiqra] when asked what the Arabic word for 'vertebra' is". The difference between these two statements (obscured somewhat by the tense form in English) is that the first statement is a singular observation, and the second is an universal claim, applicable not only on June 9th 2006 in Chicago, but on any other date and in any other place. Again, the science of fieldwork addresses the specific methods of validating the universal claim, the general form of which reduces to considering and rejecting any reasonable alternatives. There is more to it than just accumulating two or three additional tokens, but not much more.[55] If we are also armed with a few more examples like the fact that Smith says [qaraʔa] for 'he read', [waraqah] for 'paper' and [fiqh] for 'jurisprudence', we might draw a simple conclusion "Smith's grammar has the consonant [q]". An alternative is that the behavior is not the result of a grammar, in the standard generative sense.[56]

5.1.2. What exactly is that method of creating sequences of symbols?

The end product of the previous subsection is a set of linguistic forms, or at least it seems that is what we have gained. Now we need to start over, looking closer at the method of transcribing, which underlies all of phonology, to ascertain the relationship between these things and linguistic theory. Again returning to fundamental questions, what is it that phonological theory is responsible for, and what is the relationship between "data" and phonological theory? As I and others have said many times, it is responsible for theorizing the nature of phonological representations and computations, which map from representation to representation.[57] We reason to claims about

---

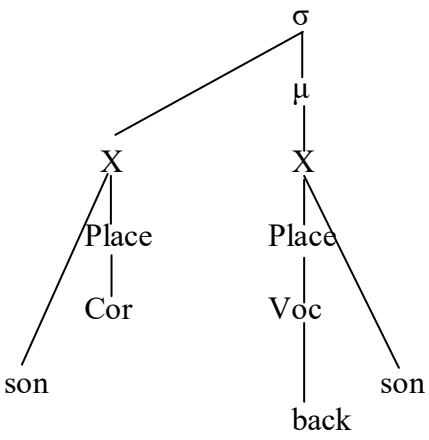
[55] Humans being volitional creatures, we can always decide to speak randomly in response to a request to utter a particular word in a particular language. The claim that a person will always say such-and-such is therefore too strong, and in general any claim about gramar that is predicated on universal quantification over human behavior is doomed. Recall that a generative gramar is a model of a cognitive faculty, and not a model of observed behavior.

[56] I can pronounce these words and if you ask me what the word for paper is in Arabic, I will say [waraqah], but I don't have a grammar that generates this form, it is simply a thing that I know about Arabic.

[57] This subsection substantially echoes the sentiments of Hammarberg 1976, though I have a few disagreements with him. I disagree with his Kuhnian, rationalist stance and the treatment of prior knowledge as 'presuppositions', but I agree with his mentalist stance as to the nature of grammar. Insofar as he appears to be operating in a classical SPE phonological framework, I disagree with that framework. I also disagree with his apparent stance that "phonet-

representations and computations based on the nature of "the $\text{data}_{\text{phonological}}$" of human language, the most fundamental kind of which is "the outputs of the phonological grammars of human languages".[58] What then is the output of the phonological grammar when I say the word "raw"? I claim that it is (57).

(57)



A similar-looking display would be the object of interest for Arabic [fiqra], namely some feature-geometric display. In other words, [fiqra] has a very remote relationship to what phonological theory is about, but (57) and similar objects have a very close relationship to what phonological theory is about. But for some reason, we take objects like [fiqra] to be more "fundamental" and beyond question than we do (57). Why do we act this way?

There are not many grammars or other data sources for English that give displays like (57) for "raw", in fact this may be the only source where you will see such a datum. Another problem is that some people may disagree with my featural analysis of "raw" (leaving aside the tangential question of whether I pronounce the word the same way that they do), and someone else might propose (58) as the output of the phonological component.

ics" is automatic and non-cognitive, a view inherited from SPE. Still, we fundamentally agree that there is nothing for phonetics to study until a basic phonological analysis – a transcription – has been created.

[58] It is the closest thing in phonology to what can be observed, and it is the basis from which a child learning a language reasons to more abstract underlying forms.

(58)

+vocalic	+vocalic
+sonorant	+sonorant
+coronal	–coronal
–anterior	–anterior
+voice	+voice
+cont.	+cont.
–lateral	–lateral
–nasal	–nasal
–low	+low
+round	–round
+back	+back
+high	–high
.	.
.	.
.	.

Two cooperative but disagreeing scholars probably can develop a translation program for English that converts one kind of output to the other. But the problem with saying that "the theory of phonology is responsible for (57), or for (58)" is that the "facts" so represented are the product of extended theoretical analysis, so we have a seeming circularity problem, that we have to know the nature of phonological representations and computations, in order to know what "the facts are" that phonological theory explains, with properties of "the facts" being the logical foundation of that theory.

To avoid circularity, we need a theory-independent mode of representing the object of interest (the output of the phonological computation). Since that representation is theory-neutral, it cannot itself *be* the thing that a phonology produces – it *stands in for*, or *represents* whatever the phonology produces. There are very many imaginable candidates for such a stand-in representation, such as a digitized speech file; a spectrogram; a matrix of formants, amplitudes and fundamental frequencies; a set of listener reaction times to some test involving listening to an utterance. None of these are useful as a model of what the phonology produces: they all suffer from the problem of indiscriminately providing too much information. You will surely recall a fundamental distinction between physical speech versus phonological representations, that speech does not come divided into segments, but phonological representation does. What is useful for phonology is a segmental transcriptions and, occasionally, some kind of prosodic notation that records comparative information about syllables, such as that syllable 2 is in some way "more prominent" than syllable 1. The phonology of Arabic does not produce [fiqra], it produces a feature representation analogous to (57) or (58), which we can conventionally and compactly present as [fiqra], given certain conventions about sound and a theory of how phonological forms relate to physical sounds. The transcription [fiqra] is an evidentiary object that represents what was said and experienced, and was cause by Smith's grammar.

A transcription of an utterance is a sequence of informed category judgments, involving at least three things: an utterance, a standard, and foils – alternative standards. When you encounter [fiqra], you need a set of measurement standards, thus a standard for judging purported [q] *qua* category. But furthermore, an instance of suspected [q] will not be *exactly* the same as the standard value, it will be *more* like that value, and *less* like some other standard value, such as that for [k] or for [ʔ]. The most-generally accepted orthographic standard for the sound of the

second consonant in our Arabic word is the letter "q", which is available in the International Phonetic Alphabet. The letter "q" is defined as a "voiceless uvular plosive" in the IPA. The IPA itself does not give technical or operational definitions of these terms, but the purported articulatory correlates of "voiceless", "uvular" and "plosive" are not so severely in doubt that such a label is meaninglessly arbitrary. The judgment about this word does not typically rest on actual inspection of articulators, it is based on an inference from sound to probable articulatory state. We claim that [fiqra] has a voiceless uvular plosive (or similar terminology) because it sounds like a standard example of [q] which has that articulatory description, which indeed may have been independently subjected to a relatively sophisticated articulatory study. Theoretically, there *is* a standard pronunciation whereby one can consult a performance of [aqa] and compare it to [aka] to judge whether an Arabic speaker said [fiqra] or [fikra].

This is sort of true, and sort of false. It is true that there are recordings on the internet purporting to illustrate particular IPA letters. Since "what's on the internet" is a constantly-evolving "standard", one should be concerned about those purported standards – who decides what the standard pronunciation of [q] is? Who decides what the standard distance "meter" is? That was first decided by the French National Assembly in 1790, based on the length of a pendulum with a half-period of 1 second, then by The French Academy of Sciences, in 1793, which concluded that the meter is one ten-millionth of the distance from the equator to the North Pole along a great circle; it was redefined by an international commission in terms of a prototype meter bar in 1799 (the bar was changed in 1889), then a wavelength definition was adopted first by the International Union for Cooperation in Solar Research, and more recently in 1983, the Conférence Générale des Poids et Mesures defined the meter in terms of the speed of light. Linguistic transcriptions are analogous to the platinum-iridium bar standard for the meter, that there are agreed-on performances, accepted by phoneticians, as exemplars of [q] or [ɪ].

The foundation of this system is the cardinal vowel system developed by Daniel Jones. Jones developed auditory and (introspective) articulatory standards for the so-called cardinal vowels. Via repeated practice, Jones' students gained expertise in the production of these vowel standards, wherein they were adopted as the received interpretation of the IPA letters. Thus, the foundation of IPA transcription is a learned system of performances, passed down from teacher to student over generations. See Ladefoged (1967) for discussion and experimental efforts at validating the system.

The main problem with this system of standards is that it is not entirely fixed, in the way that the meter-bar is a fixed object. There exists a low-quality recording of Jones cardinal vowel performances, which however is a subset of the vowels sanctioned by the IPA – there appear to be no standard recordings of consonants by Jones. A teaching tradition was developed in the UK which resulted in the training of a number of illuminaries in phonetics such as David Abercombie, A. C. Gimson, Dennis Fry, Eugénie Henderson, Ian Catford, Peter Ladefoged, John Laver, John Wells, John Esling and others. Unfortunately, there are at present negligible publicly-available reference recordings of the standard value of IPA letters as performed by such expert phoneticians, in fact there are only two generally-available "charts", one (8-bit) set of performances by Ladefoged and one (higher quality) set by Esling. It is fair to say that the tradition of auditory expertise developed by Jones is endangered.[59]

[59] There exists another much more extensive set of recordings by William Smalley to accompany his manual of articulatory phonetics, which uses model performances by Smalley as well as real-language performances by speakers of particular languages, however these recordings are outside the UK IPA-based tradition. There are also plainly incompetent assemblages of IPA performances available on the internet.

An alternative approach was pursued in earlier field-working traditions, the "language comparison" method (alluded to in the previous subsection). That is, a certain sound of a target language is compared to productions in other languages, thus "similar to *r* in Parisian French", "like Norwegian æ but further back". The primary drawback of this method is that we do not all speak Parisian French or Norwegian fluently (or at all), so while the prospects for complete loss of historical knowledge are lessened with the language-reference method, there is still a significant problem with accessing reference performances. While there certainly are recorded speech samples of a number of languages available on the internet, it is still not standard or even common that language descriptions provide adequate associated recordings, so that references like "similar to Lushootseed [ɬ]" can, on practical ground, be challenging or uninformative.

The main "problem" of transcriptional judgments, from the creation and consumption perspective, is that linguists often do not give sufficient thought to what a transcription is, or should be taken to be, or what it is intended to be in a particular instance. A transcription is not a highly precise measurement of a known acoustic property, it is a relatively coarse judgment "more like X than like Y", which facilitates the processing of category judgments for utterances in a language. This is evident to any field worker who consults existing expert performances of IPA letters, and compares what the experts say with what one hears in the particular language. Take the question of what the values of [e] and [ɛ] should be. Many linguists would point to examples like "take" and "tech" in English as exemplifying the vowels [e] versus [ɛ]. And yet, in reviewing expert IPA performances, I have found no performances of those letters that closely match how the vowels of "take" and "tech" are pronounced in General American English. The reason for this is simple: it is a mistake to say that *phonetically*, "take" is absolutely [tek] and "tech" is absolutely [tɛk], buy IPA standards. I would proffer [tɛɪk] and [tɛk] as the closest appropriate IPA transcriptions, but even then, the standard (as performed by Jones, Wells, House, Esling and Ladefoged) for [ɛ] isn't the same as how the vowel is pronounced in GAE. It is, however, *closest*. Somewhat unfortunately, there is also variation *within* this set as to what canonical [e] and [ɛ] sound like: or, fortunately, if we want to emphasize the point that auditory standards are a range, and not a precise point.

The IPA (as manifested in expert performances) has a certain degree of coarseness, and while it would be very inaccurate to transcribe the vowel of "sit" as [æ], it is unconvention but not wrong to consider the vowel of English "sit" to be IPA [e] rather than [ɪ]. The low vowel "a" is notoriously variable, in that there is a strong tendency to use "a" for the low vowel, and to call on letters like "æ", "ɑ" "ɐ", "ɒ" only when there is a two-way (or more) contrast within a language. A similar quasi-problem is that the letters [ʕ, ħ] as performed by Ladefoged and Esling sound rather different from what those consonants sound like in Arabic. Still, Arabic [ʕ, ħ] sound more like IPA-expert [ʕ, ħ] than they do like any other IPA sounds. In fact, [ʕ] in Iraqi Arabic doesn't sound the same as it does in Moroccan Arabic, and Somali [ʕ,ħ] also don't sound the same as Arabic or the same as the expert performances (likewise, Tigrinya [ʕ] doesn't sound the same as any of the above – it is closest to [ʔ] at least in some contexts). The problem with pharyngeals is perhaps more acute (see Esling et al. 2019) and could indicate a need for expansion of transcriptional standards, but even without any expansion due to pharyngeals being understudied, the demand for an exact match between language sound and transcriptional letter is a fundamentally mistaken goal. An IPA (or other) transcription simply provides a means of objectively talking about linguistic utterances in a way that most efficiently mediates between physical utterances (which may be studied using the tools of phonetics) and grammar outputs such as (57) or (58).

It is generally understood that there are two kinds of transcription, a "broad" and a "narrow" transcription, but again it is not well understood what these kinds of transcription are, in part because the concepts are used differently, by different practicioners. The main concept of "broad transcription" is a phonemic transcription, for example English [set] "sate", [sɛt] "set", [sæt] "sat". This kind of transcription is the result of a phonological analysis, where it is decided that the vowel of "sate" can be treated as manifesting the English phoneme /e/. Where does such a decision about phonemes come from? Typically, as taught in Phonology 101, it comes from a study of the distribution of actual sounds in English, using the test of "complementary distribution". We elect to not indicate glottalization in the final stops of these words because that is a phonetic property that can be provided by rule; we elect to indicate a "phonemic" difference between [set] and [sɛt] because we cannot predict that difference by any rule of English. A broad (phonemic) transcription in this sense presupposes a narrow transcription: it is the product of a phonological analysis of something more basic, something that is directly perceived.

However, another kind of "broad transcription" (a practice actually employed in the profession though not raised to the level of a terminological standard) is a tenative or *approximate*, less-precise transcription. The letter [e] may simply be a ballpark estimate of vowel quality, an indication that the vowel was "e-like", but not necessarily IPA standard [e]. A field-working phonologist transcribing a language does not have instant access to infinitely-precise categorization judgements, instead they work towards that level of accuracy through repeated exposure and token comparisons. It is a waste of a transcriber's time to obsess about a single token to nail down the exact vowels and consonants of an utterance [bɨ̆ɣ̊a̰:t̪ʰ] 'ability' in Tigrinya. It is more likely that the transcriber will initially record something like [bɨʁat] and then refine the transcription based on successive exposures and the collection of additional tokens. The letter "ʁ" might be subdivided into [χ, ʁ̆, ɣ̊] and so on when the range of similar sounds expands to indicate that it is *possible* (not even probable) that there could be [χ] and [ʁ], or [ʁ̆] as well, in the output of the phonology of the language.

Our main conclusion about the nature of transcriptions as data for phonology is,

> A transcription is a transduction from continuous physical speech to a sequence of symbols.

Transcriptional symbols are perceptual categories that unify a range of physical events, just as "dog" is a category that unifies a range of physical animals, and distinguishes "dog" from "coyote". A transcription thus "represents" the physical event, but gives it a form that is more compatible with the nature of phonological grammars. But still, phonology deals in features and not letters, so a transcription is not an object of linguistic theory, it is an evidentiary object that relates to an object of linguistic theory. The datum [fiqra] is a tool that mediates between classes of waveforms (actual or potential utterances), and the actual output of the phonology of Arabic.

More specifically, a field-worker's transcription is a shorthand notation of a hypothesis as to what the output of the phonology is – the hypothesis that the grammar produces the featural analog of [fiqra]. How does a phonological field-worker arrive at such a hypothesis? First, the field-worker has to perceive distinctions between sounds, so if you cannot perceive a difference between [t] and [t̪], you cannot *judge* that there is a difference. Bear in mind that perceiving a difference between two sounds is not the same task as knowing how to categorize a sound. It is very likely that any field-phonologist (sans hearing impediments) can perceive a difference be-

tween two actually occurring utterances.[60] Being able to localise the difference to specific segments is another matter, and being able to perform an *n*-way comparison ("that consonant is more like Esling's [χ] than it is like his [ħ] or his [x]") is an advanced, acquired linguistic skill. Second, and this is crucial, the linguist has to have a theory of what a phonological output is. This point is made numerous times in the works of Hale & Reiss, who distinguish between body outputs versus phonological outputs. The body does not output a sequence of segments nor does it output a feature structure, it outputs an acoustic waveform, which has some relationship to a transcription. Therefore, one has to have a theory of how a waveform relates to a transcription.

There are two main theories for addressing that relationship. One is that there is a phonetic component, which acts on symbolic phonological outputs and creates immediate precursor representations for the motor system. The other is there is no phonetic component, therefore the output of the phonology is that immediate precursor to the motor system. Under the single-component theory of language sound, the output of phonology is closer to being a symbolic re-creation of the acoustic waveform, therefore it contains much more information and many more distinctions than would likely arise in a two-component theory. As an example, the two-component analysis of English would most likely hold that in "exclude", the phonological output contains a round vowel (u), and does not have a string of variably-rounded consonants. The fact that lip-rounding commences at some early point in the string of consonants, probably during initial [k], and procedes continuously to the round vowel, would be attributed to some operation in the phonetic component. The alternative, in the one-component model, has many categories of rounding and there are rules assigning different values of rounding based on what immediately follows.[61] From the data-producer's perspective, a transcription is a measurement of auditory experience, calibrated by reference to a standard, where the choice of symbol reflects comparative auditory similarity, and at the same time reflects a theory of what information is produced by the phonological component.

Things are a bit more complicated, though, because a field phonologist (who is tasked with eliciting hundred of thousands of paradigmatic examples, and transcribing them all) is highly unlikely to provide maximally-differentiated narrow transcriptions capturing all of the possible sound categories for the entire corpus of hundreds of thousands of utterances. The job of exhaustive re-listening and comparing becomes intractable at a certain point, and often apparently pointless (and error-prone: tedious, repetitive tasks fall victim to shortcuts designed to alleviate boredom). Therefore, a field phonologist is likely to retreat to a different standard for transcription. Rather than judging what *could* be, theoretically, a phonological output (given the existence of optional phonological rules and a perhaps slightly enhanced set of transcriptional symbols), the phonologist is likely to make assumptions as to what the "phonemes" of a language are, and will only provide phoneme-level transcriptions, not paying attention beyond a certain level of accuracy. After all, if you can apply a set of rules to phoneme-level transcriptions to derive an accurate narrow transcription, it's so much easier to just create phoneme-level transcrip-

[60] It is not difficult to digitally manipulate a recording to make a modified version of an utterance that is technically different but imperceptibly so. The pertinent test of perceptual discriminiation is to provide actual speech examples including repetitions and imperceptible modification, and test whether the listen can identify that A B is composed of distinct recordings, versus the same recording played twice. This is different from the categorization task "is [ta] the same thing as [da], or is it the same as [tʰa]".

[61] In the SPE phonetic tradition, which is a single-component theory, features have scalar values, so this process is modeled by a numeric equation [+cons] → [*n*-*1*round]/__[*n*round], where [u] is [1round]

tions and let others compute the pronunciation. I do not argue strongly against that viewpoint, but simply point out that (a) this works – for you personally – only if you have a good-enough narrowly-transcribed basis for positing those rules creating the purported allophones, (b) you retain enough commitment to the perceptible, as opposed to your analysis, that you can recognize a counterexample to an allophonic claim when you encounter one (*milit**a**ristic* vs. *capit**a**listic*), (c) this only works for us (the consumers of your data) if you provide the evidence and analysis that underlies the previous two considerations. It is a bit of a disservice to the data-consumer to present [bɨk'at'] for (apparent) [bɨʁ̆ã̰:t̪ʰ] in Tigrinya, on the ground that the word might be *underlyingly* /bɨk'at'/, while omitting whatever evidence there is to support that theory. However, it may not matter to the consumer how the word is pronounced (not matter at all, or not matter beyond a certain very coarse level of accuracy).

The caveat to the data-consumer is, on what basis should you believe that a particular data claim is valid? In 5.1.1 we took it to be a straightforward and nearly trivial matter of deeming that Arabic 'vertebra' is [fiqra], but now we see that this is a non-trivial conclusion. The units in a linguistic data collection cannot be identified without reference to higher theory and standards of measurement. Do we know those standards, do we all follow those standards, do we know the implicit theories mediating between the object in theoretical phonology and the acoustic waveform? What are the reasonable alternatives, and why do we accept the final conclusion over the alternatives? The transcription [fiqra] is acceptable under "standard" (Esling- or Ladefoged-performance) IPA conventions as the closest symbolization, setting aside use of IPA micro-adjustment diacritics. The entire science of measuring speech and producing data depends on a rich set of higher-level propositions. Language-particular reasoning will almost certainly be crucial to admitting as a datum of Arabic a transcription [fiqh] ('jurispridence') as opposed to [fiqʰ], [fiqh] or [fiqʰ].

Refering only to a single pronunciation, [fiqʰ] is probably the most accurate representation of what was said. In the context of a larger collection of pronunciations and a partial phonological analysis of the language, [fiqh] is the most *phonologically*-justified transcription.[62] It superficially seems that there is circularity in rejecting [fiqʰ] in favor of [fiqh] on phonological grounds based on related forms of the root such as [faqihtu] 'I understood', [faqiha] 'he understood', [ʔafqahu] 'I understand', [mafqu:h] (passive participle). But, given the accumulated corpus of transcriptions of Arabic, [fiqh] and [fiqʰ] are comparable as faithful recordings of the physical event, [fiqʰ] being perhaps slightly more faithful to the physical event. Judged from the perspective of phonological correctness, [fiqʰ] in no way simplifies the analysis of the language, and it would indeed make analysis of related forms more complex. As a claim about what might be the output of the phonology – one made with knowledge of [faqihtu, faqiha, ʔafqahu, mafqu:h] – [fiqʰ] as phonological output requires a phonological rule which not otherwise justified (any voiceless stop plus *h* merges into an aspirated stop).[63]

An important caveat is that phonetic implementation should not become a garbage pail for disposing of epistemological problems. Simplicity – in phonology – tells us that it is simpler to have the phonological output be [fiqh] (or [fiqh]) rather that [fiqʰ] because segment fusion is not necessary, but what are the phonetic consequences of that decision? What burden does that

---

[62] The vowel /i/ is somewhat lax in closed syllables, there is no aspiration contrast in the language, and *q* and *h* are unquestionably separate segments in the root.

[63] The reader can presumably see why it is not a viable hypothesis that the underlying form contains /qʰ/, since that would compel the addition of a number of segment-splitting rules.

choice shift the phonetic component? From a phonologist's perspective, it doesn't seem any harder for the phonetics to output physical |fiqʰ| given phonological [fiqh] that to get the same output from phonological [fiqʰ]. The phonologist's (this phonologist's) uninformed perspective on the matter is pretty much irrelevant, what matters is, what theory of speech production and phonetic grammar is correct, and what is the effect of the competing phonological hypotheses on the phonetic grammar – are the hypotheses neutral, or does one complicate the grammar? Phoneticians are especially invited to weigh in on this topic: I will simply say that if you plan to leave a fact to phonetic implementation, you better have a theory of phonetic implementation.

In lieu of a theory of phonetics, a phonologist can at least do some preliminary vetting of the competing hypotheses (if you have some direct awareness of the facts of other languages). For the sake of discussion, I simplify the choices to just [fiqh] and [fiqʰ], focusing on competing analyses of the consonants. Human languages can have consonant clusters, and they can have aspirated consonants. Human languages can also have words that end with [m]. The null-grammar hypothesis – where there are no relevant phonological or phonetic rules – is that in all languages, final phonological [m] is pronounced the same in all languages.[64] If in some language underlying /m/ in word-final position is pronounced as [β̃], either there is a phonological rule that does this, or there is a phonetic rule. Deviation from the "universal pattern" is proof of a language-specific rule. Likewise, languages can have words that end in the cluster /lm/, and can have words that end with /ləm/. If there are no relevant rules, /lm/ will be pronounced the same in all languages, /ləm/ will be pronounced the same in all languages, and /lm/ will be pronounced differently from /ləm/. If /lm/ is pronounced differently in some language, or /ləm/ is pronounced differently, or if /lm/ and /ləm/ are pronounced the same in a language, then we have proof of a language-specific rule. While we do not know if the rule so detected is in the phonology or in the phonetics, we can use the uniform-output assumption to pin down the logic of choosing between [fiqʰ] and [fiqh] as phonological output. Given the facts of Arabic, if there is no phonological rule in the language that creates aspirated consonants from stop-plus-h sequences, then [fiqh] will be pronounced the same in Arabic as it is in all languages with such a phonetic input; and it will be pronounced different from how aspirated consonants are pronounced.[65] Of course, this is an extremely, indeed unreasonably optimistic view of phonological and phonetic knowledge, and I see no realistic prospects for proving that the phonological output [fiqh] for Arabic results in a net-simpler grammar that does the output [fiqʰ]. While in principle it could be detectable whether the phonology produces [fiqʰ] versus [fiqh], the matter is at presently unanswerable. This then leaves us with the "garbage can" alternative, that there is no compelling reason to *not* assume the simplest phonological grammar and assume that whatever happens in the phonetics is automatic and cross-linguistically invariant.

Now to summarize this subsection, to the extent that there is "a problem" with reputable transcriptions as data, the main problem is that we can't always be sure what a transcription represents.[66] One may not be certain as to the claim embodied in a transcription (especially if

---

[64] This is not so in RSFP, since there is no phonetically-defined universal feature object corresponding to [m].

[65] Just as a theoretician can over-burden the theory of phonology by adding auxiliary hypotheses requiring /qh/ to always phonologically neutralize with /qʰ/, a theoretician can over-burden the theory of phonetics by imposing some additional mechanism forcing phonological [qʰ] and [qh] to always be pronounced the same way.

[66] I have nothing to say here about disreputable transcriptions. I have encountered transcriptions that omitted important facts of a language, especially contrastive tone and vowel length, or certain vowel distinctions, and I have encountered a few transcriptions that made what turned out to be incorrect simplifying assumptions – I have *made*

you don't know the language). It is possible that a transcription is a more-faithful rendering of speech events; or that it omits certain hard-to-decide differences, especially ones that do not have IPA-category equivalents (where micro-adjustment diacritics may be called on). It may represent a particular kind of analysis, such as a taxonomic phonemic analysis, or the output of the phonology up to a certain point (e.g. "the output of the lexical phonology", "abstract away from the effects of palatalization"). Since transcription involves categorization according to a standard and a given language sound may fall between category-exemplars, the choice of [x] versus [χ] may be a result of a quantization problem. It is also quite possible that transcriptional practices are significantly influenced by orthographic practices rather than properties of utterances – this is certainly the case with most sources of Modern Mongolian or North Saami and apparently (based on the discussion of Christdas (1988)) Tamil. There are, additionally, forms of "noise" in interpreting data, since not every field linguist benefitted from a two-year training course in practical field-phonetics at UCL or Edinburgh, and often (and for good reason) linguists make their data available in a form that is at least somewhat intelligible to speakers of the language under study (example: "y" is a palatal glide, not a front round vowel, in most roman-based orthographic traditions.

What we can (should) say about a transcription and therefore phonological data, and omitting factors that constitute noise, is that:

There is a causal relation between an utterance and a transcription
There is a causal relation between an utterance and a phonological goal
There is a causal relation between an utterance and a phonological representation
A transcription is the only valid evidence about what phonological theory models

5.1.3. General data propositions about a language

With suitable valid linguistic data, we can draw inferences about how a particular language works, for example we can posit rules and representations that explain the nature of the data (causal principles): we can analyze phonology problems. As a starting point on that journey, we can gather together a number of words in Damascene Arabic, which justify certain factual claims about the language. Reasoning on the basis of relationships between word (especially words which have something in common – the same root, for example), we can eventually make justified claims about underlying forms and phonological rules. This subsection is only concerned with the logical foundation for an analysis of data in terms of a system of rules and representations. An analysis functions to explain the facts of a language, which means that fact-patterns must be identified. Here are some data, organized with words sharing a particular lexical meaning put together in a column, and words with a shared grammatical meaning together in a row.

(59)					
	nə́zel	ʕə́ref	ləbes	ʔə́der	3sm
	nə́zlet	ʕə́rfet	ləbset	ʔə́dret	3sf
	nə́zlu	ʕə́rfu	ləbsu	ʔə́dru	3p
	nzə́lət	ʕrə́fət	lbə́sət	ʔdə́rət	2sm
	nzə́lti	ʕrə́fti	lbə́sti	ʔdə́rti	2sf
	nzə́ltu	ʕrə́ftu	lbə́stu	ʔdə́rtu	2p

some such errors myself, but have worked to obliterate all such errors. Most often, I find that apparent transcriptional discrpancies reduce either to real dialect differences, or comprehensible level-of-analysis differences.

nzə́lət	ʕrə́fət	lbə́sət	ʔdə́rət	1s
nzə́lna	ʕrə́fna	lbə́sna	ʔdə́rna	1p
'descend'	'know'	'dress'	'be able'	

In these data, the suffixes /Ø, et, u, ət, ti, tu, ət, na/ are evident (other evidence shows that [ət] is /t/). Each column has common subparts because each column has a single root plus a suffix. The question of interest is, what are those roots, and what explains the variation in the realization of the roots. The answer depends on abstracting fact patterns from the data. Using analogical reasoning, we can generalize from these particulars to a general pattern, abstracting from the particular consonant in roots.

(60)	Cə́CeC	3sm	CəCCet	3sf
	Cə́CCu	3p	CCə́Cət	2sm
	CCə́Cti	2sf	CCə́Ctu	2p
	CCə́Cət	1s	CCə́Cna	1p

Each particular consonant as first, second or third C of a root constitutes a degree of freedom in the correlation between specific word and abstract fact pattern. It turns out that there is no significance to the choice of specific consonant in any position and the particular pattern that a verb follows.[67] There are numerous other examples such as those in (61) which follow this pattern – I only give the 3sm form, and assure the reader that the remaining forms follow this same pattern.

(61)	mə́sek	'he held'	ħə́sen	'he was able'
	ʕə́mel	'he made'	fə́hem	'he understood'
	ʔə́tel	'he killed'	lə́ʕeb	'he played'
	kə́ber	'he grew up'	wə́sʕel	'he arrived'
	wə́ʔeʕ	'he fell'	jə́bes	'he dried up'
	xə́reb	'he was ruined'	tə́ʕeb	'he got tired'
	zʕə́ʁer	'he became small'	sə́hel	'he became easy'

Abstract symbolization in the form of (60) is not advancement of a proposition, but it is an important aid to developing propositions. One might claim "All verbs have the form Cə́CeC in the 3rd singular masculine, Cə́CCet in the 3rd singular feminine…" – this is a testable claim (which is false, as we shall now see: other conditions have to be imposed on the descriptive proposition).

There are other inflectional patterns, seen in (62).

(62)	a.	kátab	dáras	tʕábax	ʔáxad	3sm
		kátbet	dárset	tʕábxet	ʔáxdet	3sf
		kátabu	dárasu	tʕábaxu	ʔáxadu	3p
		katábət	darásət	tʕabáxət	ʔaxádət	2sm
		katábti	darásti	tʕabáxti	ʔaxátti	2sf
		katábtu	darástu	tʕabáxtu	ʔaxáttu	2p

[67] This is an accidental property of this dialect. In some dialects of Arabic, choice of certain consonants on position 1, 2 or 3 determines which analytic pattern a verb follows.

	katábət	darásət	tˤabáxət	ʔaxádət	1s
	katábna	darásna	tˤabáxna	ʔaxádna	1p
	‘write’	‘study’	‘cook’	‘take’	
b.	sákkar	ʒárrab	ʔáʕlan	tárʒam	3sm
	sákkaret	ʒárrabet	ʔáʕlanet	tárʒamet	3sf
	sákkaru	ʒárrabu	ʔáʕlanu	tárʒamu	3p
	sakkárət	ʒarrábət	ʔaʕlánət	tarʒámət	2sm
	sakkárti	ʒarrábti	ʔaʕlánti	tarʒámti	2sf
	sakkártu	ʒarrábtu	ʔaʕlántu	tarʒámtu	2p
	sakkárət	ʒarrábət	ʔaʕlánət	tarʒámət	1s
	sakkárna	ʒarrábna	ʔaʕlánna	tarʒámna	1p
	‘close’	‘try’	‘announce’	‘translate’	

Lexical subset (a) is populated by many stems having three consonants and the vowel [a] in the two stem syllables. Roots in subset (b) have more than three consonants (which may be two adjacent identical consonants in the middle of the root), including certain derivational prefixes, or else a long first vowel. Further lexical probing reveals an ample supply of words of subtypes (a) and (b), from which we can abstract the following patterns.

(63)	a.		b.		
	CáCaC		CáCCaC		3sm
	CáCCet		CáCCaCet		3sf
	CáCaCu		CáCCaCu		3p
	CaCáCət		CaCCáCət		2sm
	CaCáCti		CaCCáCti		2sf
	CaCáCtu		CaCCáCtu		2p
	CaCáCət		CaCCáCət		1s
	CaCáCna		CaCCáCna		1p

We temporarily set aside (63b) which correlates (very strongly) with the number of consonants. The descriptive proposition which we are working towards is that the inflectional pattern is partially determined by the number of consonants in the root, and partially is an unpredictable lexical fact. We can now state a valid generalization: if the root contains three consonants, the inflectional pattern is either (60) or (63a), regardless of the specific consonant on the root.

The last pattern of interest arises in roots having two consonants, examples given below (roots with final geminate consonants are “counted” as having two consonants, not three).

(64)					
	sá:ʔ	bá:ʕ	bána	wáfa	3sm
	sá:ʔet	bá:ʕet	bánet	wáfet	3sf
	sá:ʔu	bá:ʕu	bánu	wáfu	3p
	sáʔət	báʕət	bané:t	wafé:t	2sm
	sáʔti	báʕti	bané:ti	wafé:ti	2sf
	sáʔtu	báʕtu	bané:tu	wafé:tu	2p
	sáʔət	báʕət	bané:t	wafé:t	1s
	sáʔna	báʕna	bané:na	wafé:na	1p
	‘drive’	‘cook’	‘build’	‘fulfill’	

máʃi	nə́si	mádd	támm	3sm
máʃjet	nə́sjet	máddet	támmet	3sf
máʃju	nə́sju	máddu	támmu	3p
mʃí:t	nsí:t	maddé:t	tammé:t	2sm
mʃí:ti	nsí:ti	maddé:ti	tammé:ti	2sf
mʃí:tu	nsí:tu	maddé:tu	tammé:tu	2p
mʃí:t	nsí:t	maddé:t	tammé:t	1s
mʃí:na	nsí:na	maddé:na	tammé:na	1p
'walk'	'forget'	'extend'	'remain'	

These exemplify the 4 extant subtypes of biconsonantal patterns in the language, and expanding the data coverage to more examples yields no new sub-types. The following patterns can be abstracted from the data, generalizing across all consonant distinctions.

(65)

Cá:C	CáCa	Cə́Ci	$CáC_iC_i$	3sm
Cá:Cet	CáCet	Cə́Cjet	$CáC_iC_iet$	3sf
Cá:Cu	CáCu	Cə́Cju	$CáC_iC_iu$	3p
Cə́Cət	CaCé:t	CCí:t	$CaC_iC_ié:t$	2sm
Cə́Cti	CaCé:ti	CCí:ti	$CaC_iC_ié:ti$	2sf
Cə́Ctu	CaCé:tu	CCí:tu	$CaC_iC_ié:tu$	2p
Cə́Cət	CaCé:t	CCí:t	$CaC_iC_ié:t$	1s
Cə́Cna	CaCé:na	CCí:na	$CaC_iC_ié:na$	1p

Patterns (60), (63) and (65) exhaust the possibilities, at least for the perfective form of verbs.

What we have done here is induce a set of higher-level generalization about data, where we have organized the data of the language into analogous sets. Every word that we have considered (verb, in the perfective) can be subsumed under one of these generalizations. The particular consonants that occupy C slots in these patterns are specified in the lexicon, and there is no correlation between specific lexical C and the paradigmatic subclass induced above – there is no rule to the effect that C_1 must be [b] or cannot be [l]. There is also a two-way contrast in vowel patterns for those roots with three consonants, namely (60) versus (63a) which have alternating [ə~e~Ø] versus relatively stable [a…a]. Two-syllable biconsonantal vowel-final roots have a similar lexically-determined pattern [ə~i~Ø~i:] versus [a…a~e:]. These generalizations then form the basis for a set of rules and representations.

5.1.4. ANALYTIC PROPOSITIONS ABOUT A LANGUAGE

The patterns induced above are, in generative phonology, taken to indicate something about a system of underlying forms and rules: we don't just say "here are the patterns, everything can be classified in one of these ways". We are reaching for first-order causal principles. Now we need to understand the rules and representations that account for triconsonantal roots, exemplified by [kátab] 'he wrote' and [nə́zel] 'he descended', whose forms are repeated in (66).

(66)

kátab	nə́zel	3sm
kátbet	nə́zlet	3sf
kátabu	nə́zlu	3p
katábət	nzə́lət	2sm
katábti	nzə́lti	2sf
katábtu	nzə́ltu	2p
katábət	nzə́lət	1s
katábna	nzə́lna	1p
'write'	'descend'	

As a hypothesis regarding the underlying form of the verb, the data most directly supports three alternatives for the second class of words: /nə́zel/, /nə́zl/ or /nzə́l/. Since the data pattern of the verbs [lə́bes] 'dress' and [ʔə́der] 'be able' is identical to that of [nə́zel], we have no reason to assume that other verbs in that behavioral class have different underlying forms, apart from choice of particular consonant. Therefore, we are evaluating the three hypotheses (a-c) regarding underlying forms, plus some set of rules which derive (d).

(67)

a.	Cə́CeC	Cə́CeC-et	Cə́CeC-u	Cə́CeC-ti	Cə́CeC-na
b.	Cə́CC	Cə́CC-et	Cə́CC-u	Cə́CC-ti	Cə́CC-na
c.	CCə́C	CCə́C-et	CCə́C-u	CCə́C-ti	CCə́C-na
d.	[Cə́CeC]	[Cə́CCet]	[Cə́CCu]	[CCə́Cti]	[CCə́Cna]

One further possibility, motivated by the fact that different vowels within the word are stressed, is that underlying form have no stress indications, which instead result from a rule: this results in hypotheses (a′-c′) without underlying stress marks.

We can compare the entire set of hypotheses lacking underlying stress, versus the hypotheses with underlying stress on the root (the first root vowel; the vowel schwa). A stress assignment rule can indeed put stress on the penult under all analyses at least for words like [nə́zel], but the hypotheses with underlying stresses also require rules that eliminate lexical stresses, since the location of stress is not fixed within the root. For example, under the /Cə́CeC, CáCaC/ and /Cə́CC/ hypotheses, initial stress must be eliminated in [CCə́Cti, CaCáCti]; under the /CCə́C/ hypothesis, final stress must be eliminated in [Cə́CeC] and [Cə́CCet]. No auxiliary rules are necessary under the theory that there are no underlying stresses, and no complications result from the hypothesis that underlying forms have no stresses. Although we have not yet proffered a stress rule and certainly haven't done so in a formal theory where the simplicity of two analyses can be explicitly compared, we don't need to in this case, because the rules necessary for the no-stress analyses are a proper subset of the rules needed under the underlying-stress hypothesis: the no-stress analysis is necessarily simpler.

The data represented in (60), (63) and (65) reveals a simple generalization, which is that stress is on the rightmost heavy syllable, with the proviso that a word-final single coda consonant does not render a syllable heavy. This problem would be extremely difficult to solve and the generalizations would be extremely difficult to state, if Damascene Arabic were the only language that had been analyzed. Perhaps these generalizations would only emerge after we attempted similar analyses of stress in Ancient Greek, Latin, and Classical Arabic. If all syllables are light, stress is on the antepenult. The one superficial exception is forms like [katábət] 'I wrote; 2sm. wrote'. There is much evidence that these suffixes are /t/ and not /ət/, and thus the

underlying form is /katabt/. As seen in (64), the 1s and 2sm suffixes which surface as [ət] behave like the other C-initial suffixes /-tu, -ti, -na/ in other respects: they trigger the lengthening and fronting of final vowels ([bana, məʃi]) like the other consonant-initial suffixes. They trigger shortening of a preceding long vowel to schwa (sa:ʔ). They trigger insertion of [e:] after geminates (madd). The seemingly exceptional stress pattern and the fact of functioning like a consonant-initial suffix in other respects justifies the claim that the suffixes are /-t/, and schwa is inserted by general rule. The alternative, positing /ət/, results in significant complications involving disjunctions of the type "a consonant, or suffix-initial schwa plus consonant".

Referring to the three hypotheses in (67) and understanding the pattern of stress assignment, under hypothesis (a) with /CəCeC/, we appear to need a rule deleting /e/ before CV (Cə́**e**Cet → Cə́CCet, Cə́**e**Cu → Cə́CCu) and a rule deleting unstressed /ə/ before CV (C**ə**CéCti → CCéCti). Because *ə́* does not delete in [nə́zel], where schwa is stressed, deletion of schwa must be restricted to unstressed schwa. The lack of examples with stressed [e] in an open syllable means that we cannot directly test whether e-deletion requires a stress condition on it. But one other generalization (and rule) must be considered, which impinges on deletion of *e*. We also observe the root-internal vowel alternation CəC**e**C ~ CC**ə́**C-CV, where V_2 *e* becomes [ə] before CC. It is possible (most likely) that no rule directly deletes /e/ in /nə́z**e**l-u/ → [nə́zlu], instead, /e/ becomes schwa, and then schwa deletes before CV, by a rule that we already know exists. Since direct e-deletion is not clearly necessary, we can shift our focus to the alternation between [e] and [ə].

Schwa and [e] do not occur in the same contexts in the language. The short vowels [e,o] only appear in a final syllable before one consonant (an exception is the noun suffix [-a]~[-e] depending on whether the root contains pharyngealized or guttural consonant), and schwa appears elsewhere.[68] Because of vowel-deletion before CV, schwa is always followed by CC or is stressed. Distributional complementarity does not prove that the vowels have the same underlying form, but at least makes such an analysis possible. Other alternations between *e,o* and *ə* governed by position give support to such a rule, for example [ʃá:f-et] 'she saw', [ʃa:f-ə́t-na] 'she saw us'; [kto:b] 'write!', [ktə́bi] 'write f.!'; [ħme:l] 'carry!', [ħmə́li] 'carry f.!' where monosyllabic *ktob, ħmel* are subject to vowel lengthening. These facts suffice to justify a rule reducing mid vowels to schwa except before a single final C. For the moment, we will leave this and all rules formally unstated, because the proper statement of the rule is a function of the general theory of rules constituting UG, which must itself be validated against observation of language specific rules. We are a few steps from inducing a theory of rules: we are still evaluating descriptions of the languages facts, to be plugged into a theory.

Hypothesis (b), where the root is /CəCC/, requires a rule of *e*-insertion applying to word-final clusters, so that /CəCC/ → [Cə́CeC]. While final consonant clusters are broken up by an epenthetic vowel, the inserted vowel is [ə] and not [e], e.g. /katab-t/ → [katábət] 'I wrote'. It is possible to postulate a separate rule which inserts [e] specifically between two root consonants, but that rule has no independent utility, unlike mid-to-schwa reduction. When followed by a CV suffix, purported /CəCC/ becomes [CCəC]. That might be implemented via a direct metathesis rule, or by a sequence of schwa insertion and deletion – insert [ə] in the context C_CC (this is a context for epenthesis, for example imperfective /b-tə-ktob-u/ → b-tə́-ktob-u → b-tə́-ktb-u → [b-

[68] Underlying /o/ does not appear in perfectives, but does appear in imperfectives, for example [b-jə́-ktob] 'he writes', [b-jə́-ħmel] 'he carries'. Schwa does appear in a final syllable before CC in imperfectives, for example [b-i-ħə́ss] 'he feels'. Schwa can appear before a single final consonant, but only when it is epenthetic, e.g. /katab-t/ → [katábət] 'I wrote'.

tə́-kətb-u] ‘you pl. write’). Schwa so inserted is *not* assigned stress, and we would expect /nəzl-ti/ → nə́zl-ti → *[nə́zəlti], analogous to /b-tə-ktob-u/ → [b-tə́-kətb-u]. Therefore, such schwa-insertion would have to be via a separate rule for which there is no justification. The alternative of direct metathesis, where /CəCC-C/ → CCəC-C prior to stress assignment, avoids this problem. Comparing (a) and (b), in particular the rules that crucially distinguish the hypotheses, (a) only requires a rule deleting unstressed schwa before CV, but (b) requires a rule of final e-epenthesis which only applies to stem-internal final CC clusters, plus a rule of metathesis.

Hypothesis (c) which posits /CCəC/ requires some sort of metathesis rule so that CCə́Cet → [Cə́CCet], and requires some account of the fact that unsuffixed /CCəC/ becomes [Cə́CeC], e.g. metathesis plus final e-insertion. Hypothesis (c) is similar to hypothesis (b) and suffers from the same disadvantages of complexity compared to hypothesis (a).

We have now arrived at a system of rules, albeit not familiar formalized generative rules. We have considered and ruled out the reasonable alternatives. We thus have a valid understanding of relations between data. Our next step is to treat particular rules of languages like data in a language, and arrive at even higher-level propositions regarding the theory of “rule”.

5.2. A theory of rules

An extensive review of how the contemporary theory of rules developed in generative phonology would take us very far from the central topic of this work, but a brief overview of the fundamental developments is in order, in order to concretize the claims of the present approach to theory construction. I claim that there have been errors in theory-construction in the course of generative phonology: what is the source of those errors? I believe that the root cause is a little-recognized change in goals of theory, without a corresponding change in the associated theory. Generative phonology no longer models sets of data produced by speakers, it models the *causes* of such data, the system of rules. This is a definite improvement in goals, but calls for re-evaluation of the tools of the theory.

The first theory of rules is embodied in Pāṇini’s Aṣṭādhāyī, who set up a formal system of about 4000 rules describing Sanskrit, covering all aspects of the language. As adjuncts to this system, there are two lexical lists and a structured phoneme inventory of 14 lines known as the Śivasūtras. There was a concept corresponding closely to “underlying form”, and there are specific rules which often correspond to the rules that a contemporary phonologist would postulate. For example, rule 6.1.77 [iko yaṇ aci] says that ‘(in the place) of [i u ṛ ḷ] (there is, respectively) [y w r l] before vowels (in connected speech)’, i.e. a rule of glide formation. The expressions [ik], [yaṇ] and [ac] refer to classes of segments, interpreted by reference to the Śivasūtras, where “i” is second segment in the list and “k” is a dummy letter (called *anubandha*) identifying the end of the second line (sutra) – “ik” thus means “everything from [i] to the end of the second line” (i u ṛ ḷ). “yaṇ” begins with “ya” (the consonant *y*) which is the second entry in line 5, and “ṇ” is the anubandha ending the 6th line. “ac” refers to “vowels” because “a” is the first item in the Śivasūtras and “c” is the anubandha terminating the fourth line. (68) gives the Śivasūtras.

(68)

1. a i u Ṇ	2. ṛ ḷ K	3. e o Ṅ
4. ai au C	5. ha ya va ra Ṭ	6. la Ṇ
7. ña ma ṅa ṇa na M	8. jha bha Ñ	9. gha ḍha dha Ṣ
10. ja ba ga ḍa da Ś	11. kha pha chha ṭha tha ca ṭa ta V	
12. ka pa Y	13. śa ṣa sa R	14. ha L

The expressions which can be so constructed constituted the Paninian theory of "natural class". The resulting system of grammar is extremely economical, economy being measured in syllable count. However, it is only designed to be a theory of Sanskrit, designed to describe the strings that constitute that language.

The primary wellspring of ideas about rules leading to generative phonology is the theory of sound laws in historical-comparative linguistics, and the development of the synchrony-diachrony distinction. Rask's famous 1818 essay *Undersøgelse om det gamle Nordiske eller Islandske Sprogs Oprindelse* introduces elements of a theory of rules through his account of correspondence relations between words in different languages such as 'transposition of letters', 'interchange of mute letters', 'increment of letters added at the beginning of words', and posits particular general rules such as what we know as gradation in Finnish: "in Finnish *p* changes to *v*, *t* is dropped or becomes *d*, *k* is dropped or becomes *j*; *pp, tt, kk* become *p, t, k*, and so forth". The general practice of plain-language verbal description of linguistic relationships prevailed, frequently augmented by non-systematic symbolic expressions such as correspondence tables or formulae ("ʔ+ʔ=ʔ", Osborn & Smalley 1949; Harris (1951) "[t^h] …occurs in [#__V]", "/d/ will not occur in /ˈV__V/"). Up to the advent of generative phonology, there was little consideration that there should even *be* a theory of rules. There was, instead, a theory of how one should construct an analysis, with no notion that rules have a definite nature.

Exploiting developments in the recently developed theory of computation, mathematical modelling of language properties by Harris introduced abstract notions of substitutability and rule-notational practices into the field, for example (Harris 1946) "BC = A" means that a morpheme of class B followed by a morpheme of class C can be substituted for a single morpheme of class A, and "AC = BC" adds to that "only if C follows A". Chomsky (1951) presents such a formulaic account of the transformational rules in Modern Hebrew phonology using a notational system which is language-neutral, fixed in advance, and structured so as to favor simplicity, specifically shortness of description. The system of notation should be reasonably familiar to linguists familiar with rule-based grammars, or standard string-manipulation systems in mathematics. Like the theory in SPE, this theory exploits notions of "basic" transformations changing one thing to another, versus compound expressions that expand stepwise into sequences of basic transformations – the simple rule / schema distinction of SPE. For example p. 46 MR34, looks like a number of rules of SPE, though using lists of letters rather than features.

(69) [B,P,K] →{[b,p,k] in env.{__:W,{#,{V,N}C_j}__} j=0,1;C_j≠y,[v,f,x]}

Despite being inauspiciously catapulted into a hyper-mathematical theory of rules, phonology at MIT persisted in using ordinary English to state rules, until a more familiar theory of rules was developed in various papers in the MIT RLE Quarterly Progress Reports, which culminated in *The sound pattern of English* as the essential statement of generative rule theory for a period.

There are two problems with ordinary English language descriptions of phonological rules. One is the potential for ambiguity, if a rule statement is poorly drafted. Based on observation of actual verbal rules in linguistics, I don't think this is a significant concern, at least for reasonably simply rules (something like (69) probably cannot be clearly expressed in plain English). Concern number two, which is more significant, is that plain English treats as comparable some ideas that are not comparable. For example, "and" and "or" are comparable as language units, but they are not comparable as grammatical operators. Any actual rule is replete with instances

of "and" when translated into predicate calculus; few if any well-motivated rules involve "or". The word "unless" thrown into a linguistic formula sets off, or should set off, alarms. The seemingly simple rule "a stop devoices before *h*, unless preceded by hV̆" implies a very complex set of plain-language statements can formulate as standard rules, following the usual practice of stating what things are true of input strings, and not what things are false of such strings:

(70) a stop devoices when preceded by C and followed by *h* or
a stop devoices when preceded by V and followed by *h* or
a stop devoices when preceded by pV̆ and followed by *h* or
a stop devoices when preceded by tV̆ and followed by *h* or
a stop devoices when preceded by kV̆ and followed by *h* or
a stop devoices when preceded by sV̆ and followed by *h* or…

Plain-language rule-writing obscures the conceptual complexity of certain imaginable rule types. Another example is the word "respectively", which could easily enter into plain English rules.

(71) /p/ becomes [b] when preceded by a liquid, nasal or fricative and followed by a vowel, glide or stop, respectively
/p/ becomes voiced, fricative or ejective when preceded by a stressed vowel, nasal or liquid respectively

If we agree that plain-language rule-writing is not a satisfactory model of phonological rules, then what alternative must we embrace? Clearly, we ought to embrace a formal theory: but *which* formal theory. Why *not* embrace the theory of Chomsky (1951)?

The theory of Chomsky (1951), and to a large extent Chomsky (1955: *Logical structure of linguistic theory*) and even *Syntactic Structures* has a different goal from that of contemporary phonology. Those earliest works were carried out in a metatheoretical framework whose goal was modeling a corpus of data, characterizing a language as a set of sentences constructed out of a finite set of elements, and seeing the aim of linguistic analysis as being partitioning the grammatical sequences which are in the language from the ungrammatical ones which are not in the language. The main desiderata of that approach are, as articulated in *LSLT* I-15, whether the theory is rigorous, and does it apply to other languages? Importing the entire arsenal of mathematical tools serves the purpose of rigor. The highly expansive theory of phonology that emerged as SPE theory follows from the premise that our sole concern is with abstract properties of data sets. Even though there had been a shift in metatheoretical function of grammar in favor of a cognitive view which led to the emergence of generative phonology (as distinct from Chomsky 1951 as pre-generative phonology), there was no repudiation of the assumption that the arsenal of mathematics is part of the arsenal of linguistic theory. Metatheoretical concepts of theoretical minimalism only gained popularity in linguistics in the mid 90's, and acceptance of those ideas has a substantially less well established position in phonology, even today.

5.2.1. BUT FIRST, WE NEED A THEORY OF GRAMMAR

While there are good methodological reasons to not adopt the full package of theoretical claims advanced in Chomsky (1951), that work moves phonological theory in the correct general direction, by setting forth a tenable theory of grammar as computation. Under a different set of re-

search goals which define what phonology investigates (ones prioritizing simplicity and the cognitive role of grammatical theory), a very different theory of phonology would, no doubt, have emerged. Rather than try to construct a model of phonology which simultaneously presupposes the now-standard (?) view of grammar as computation and reaches for a minimalist cognitive system, let us look simply at the question of the influence of adopting the "grammar as computation" approach. That approach is not entirely the product of generative grammarians, but it is an issue brought sharply into focus by Chomsky.

We can try to discern what view of grammar existed in ordinary phonological analyses set forth in pre-generative literature through the mid-50's. Proposed rules often have a decent resemblance to verbal statements of rules as would be posited in a generative analysis.[69]

(72) Bloomfield (*The Menomini language*): Nonsyllabic sounds at the beginning of a word are palatalized before front syllabics (*ɛ e i*) and labialized before back syllabics (*a, o, u*);

C → C^{y}/#__front V

C → C^{w}/#__back V

Nonsyllabic sounds after a front vowel are palatalized, after a back vowel labialized.

C → C^{y}/front V __

C → C^{w}/back V __

A short vowel before or after a long vowel is dropped

$\breve{V}$ → Ø % __$\bar{V}$

Sapir (*The Takelma language of southwestern Oregon*): tx always simplifies to s, t'x to ʔs.

tx → s

t'x → ʔs

A diphthong in *u* tends, by an easily understood dissimilatory process, to drop the *u* before a labial suffix (*-g^{w}-*, *-p^{h}*, *-baʔ*).

u → Ø / V__+ [labial]

In attempting to interpret pre-generative works in phonology, the main challenge is determining whether a particular author has in mind any theory of phonology whatsoever. The way an author talks about phonological rules may simply reflect a higher interest in saying what the facts of the language are, or perhaps in explaining the history of the language. For example, language descriptions often contain distributional statements, such as a list of attested word-initial consonant clusters, or a statement (Bloomfield *Menomini*) that "All consonants occur, singly, in final position". I know from my own practice that in writing a description of a language, the factual statements that I give about a language are usually surface-oriented statements directed at saying what characteristics the data have: they do not directly reflect my view of the mechanisms which generate the data. A particularly common mode of conveying facts about a language is a schematic table, for example listing all possible consonants as C_1 vs C_2 and at the intersection of the row and column for two sounds listing the surface outcome. This does not necessarily mean that the author believes that it is in the nature of the grammar that there are such tables. We can even, on some occasions, discern evidence for a competing non–tabular theory of grammar, when an

[69] Phonetic symbols employed in the original sources are modified to conform with standard transcriptional practices of the 50's and 60's, for example ejectives <k!> are re-transcribed as <k'> and <ˀ> is re-transcribed as <ʔ>.

author offers both tables and rules. For example, Bogoras (*Chukchee*) gives a table of "admissible" medial consonant clusters, a table of medial consonant changes, but also a set of general verbal rules which resemble ordinary verbal statements of generative phonological rules

(73) *k, γ* and *y* before dentals become *γ*.
{k,γ,y} → γ / __ dental
ŋ before labials becomes *m*; before dentals, *n*.
ŋ → {m,n}$_i$ / __ {labial,dental}$_i$
k and γ before labials become *w*; with *v* and *w*, they form a labialized *k*
{k,γ} → w / __ labial
{k,γ} {v,w} → k^w

The rules regarding *k* and *γ* before labials pose an mild computational problem, that the rule deriving [w] is, apparently, "overridden" by the rule deriving [k^w]. They are presented as two separate rules, but since *v,w* are labials, the question is, how do we determine that /kv/ → [k^w] and not *[wv]? The logical connection is that {k,γ}{v,w} is, extensionally, a proper subset of {k,γ}{labial}, thus /kv/ → [k^w] must take priority. Otherwise, there could be no rule /kv/ → [k^w].

Here are some other examples of rules whose statement involve some kind of disjunctive mechanism.

(74) Sapir (Takelma): When an element ending in a nonsyllabic is followed by an element with an initial consonant, a connective *-e-* appears between them; Connective *-e-* is not used before semivowels
Ø → e / C __ C[not semivowel]

That is, a general rule is stated, then it is modified by a more specific limiting statement.

(75) (*Takelma*) All *k-* sounds (*k^h, g, k', k^{hw}, g^w, k'^{w}*) simply disappear before *x* without leaving any trace of their former existence, except in so far as *k'* and *k'^{w}* remain as *ʔ*.

Deletion and replacement are subsumed under a single generalization, the choice being governed by presence of the glottalized subset of velars.

(76) (*Takelma*) … the two a- vowels would not coalesce into one long vowel, but would be separated by an inorganic … catch. The same thing happens when two verbal prefixes, the first ending in and the second beginning with a vowel, come together.
If a diphthong in i or u precedes a catch followed by a vowel, the i or u often appears as y or w after the catch: k'wãʔya` 'just grass' (= k'wāĩ+-a`)

Insertion of glottal stop is 'within' a diphthong, not before the second vowel.

The following statement of vowel epenthesis reflect a quasi-cyclic word-to-word comparison view of phonology, where rules derive the form of inner morphemes such as roots, and if suffixes are added that take away the phonological condition triggering epenthesis, the rule is "undone".

(77) Frachtenberg (*Coos: an illustrative sketch*): Terminal consonantic clusters are avoided by inserting a weak vowel between two consonants standing in final position. But as soon as a suffix is added to a stem thus expanded, changing the cluster from a terminal to medial position, the inserted vowel is dropped, and the consonants are combined into a cluster

Another well-known property of the underlying metatheory which affected how people viewed rules is the infamous separation of levels problem, the premise that so-called allophonic rules had a different ontological status from so-called morphophonemic rules, whereby rules which neutralize distinctions in sounds have one status, and "reversible" rules that select from a list of surface realizations according to physical context have a completely different status. Voicing of /t/ in Russian has one status because /t/ and /d/ are linguistically distinct sounds where one cannot always be predicted from the other based on the physical context of their appearance. But [x] and [γ] have a different status: [γ] only appears before voiced obstruents, where [x] does not appear. To the extent that a linguist adopts the "separation of levels" premise, the theory of rules possible under that theory of grammar will be radically different from the theory of rules adopted without the separation of levels premise.

Before we can sensibly talk about what a minimalist generative theory of phonology ca. 1960 might have looked like, we need a basic theory of grammar, specifically the grammar-as-computation framework. The most important claim in that theory is that a grammar operates in a regular way to map from an input representation to an output representation. A grammar has some number of independent computations: this means, each computation identifies a class of substrings that are subject to the computation.[70] Computation c_i applies solely on the basis of string properties specified in c_i, and what is present in the immediate input string in_i. When the conditions set forth in that rule are satisfied, the operation characterizing the rule applies.[71]

This theory of grammar makes three essential claims:

(78) Computations map representations to representations
Each computation defines a "level": each computation operates in terms of properties visible at its level
Computations are of the form "If X is true, Y happens"

This theory does not say what the technical nature of a computation is. That is the next topic. Instead, it says, whatever the exact nature of a computation is, these are the properties of the *system* of computations that constitute a grammar. The validation of this theory of computation is not difficult, and apart from the specific quasi-mathematical language used to state the theory, this corresponds to how many linguists have thought about phonological facts. The main conceptual difficulty is the serial view of computation, where an alternative view would be to state surface properties of utterances. We can thus factor out one issue under serious contention: do grammatical computations operate in a stepwise fashion, or are they performed all at once? To chose between these alternatives, we have to have a theory of the computations themselves.

[70] Or, more accurately, computations are performed but often the result is that there is no change.

[71] Here I am vacillating between "rule" and "computation" because what I mean is "rule, as envisioned in this computational theory of grammar"

5.2.2. What would be the initial theory of rules?

As an illustrative exercise in theory-formation, I intend to sketch below what I believe to be a plausible first attempt at a formal mimalist theory of the basic units of phonological computation, the rules, given general knowledge about phonological systems available up to the early 60's. This chronological restriction means that the wealth of knowledge of autosegmental exotica uncovered and flowing from the autosegmental research program in the 70's (phenomena regarding tone, compensatory actions and so on) is simply not part of the discourse. The reason for excluding these phenomena is that there are two classes of arguments supporting autosegmental phonology: the empirical, and the conceptual. We are immediately interested in the conceptual arguments that might have led to autosegmental phonology. As the reader has no doubt discerned, I am extremely skeptical about a lot of the SPE theory of rules, but it is important to know that certain elements of the theory had reasonable empirical motivation.

Propositions about the theory of rules are formed by induction from specific instances of rules. But you can't instantly validate a theory of rules by pointing to examples of rules and observing what they have in common. The apparent circle in reasoning is that you seem to have to know what a rule is in general, in order to identify an instance of a rule, but you cannot validate your concept of rule without referring to specific instances of the concept. This is not just a problem with phonological rules, it is a general puzzle of epistemology, which leads many people to believe that you have to arbitrarily stipulate a claim about the universe and see what things might be deduced from that claim. Arbitrary stipulation is the antithesis of the approach advocated here. Sometimes (but generally not in science), these stipulations are thought to come from "insight", a special ability to see the Platonic Forms.

The problem alluded to here is specifically about equivocation between evidentiary concepts and ontological claims – treating "facts of languages" as the same as "rules of grammar". A secondary problem, which I believe to be important but which is not the main focus of this (or the next) subsection, is that theory development is an iterative process. The growth of knowledge initially starts with weak knowledge, mere possibilities worth checking out; patterns of confirmation that lead to probabilities, refinements in the theoretical claims, and eventually to certainties. Conceptual plausibility may lead to empirical possibility, and if you have a coherent idea of how language works in a particular domain, you may discover an example of that idea. The justification for the theoretical claim can then procede by showing that the characteristic facts which identify the theory as being superior to alternatives are indeed a general property of language, and not an error of observation.

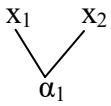
5.2.3. Phenomenological concepts

In §5.1, I outlined the logical progression from direct perception to general statement about language facts. The starting point is singular statements – reports of having experienced a linguistic form – then building up to replicable patterns first about specific words, then drawing factual conclusions about systems of words, and (phonological) relations between words. Conclusions about the patterns of fact *can* be valid, independent of a theory of the underlying cause of the facts, indeed a theory of causation depends on prior knowledge of the effects whose cause we seek.

Some fact patterns are of limited utility, not extending beyond being a report of the facts of a particular language. Above, we considered some specific fact patterns in Damascene Arabic,

but the patterns discussed are just facts about one language. Some fact patterns (including elements of the latent rule system of Damascene Arabic) are of broader utility, appearing in the analysis of many languages, suggesting that they have some relation to the theoretical concept of "phonological grammar". In theorizing, linguists also form phenomenological concepts, which are a precursor to theoretical concepts. As an example, "inalterability" is a proposition about what languages do or do not ever do, as evidenced by a class of phenomena found in numerous languages. It has been repeatedly observed that geminates and other homorganic consonant sequences fail to undergo phonological processes which other types of consonant sequences do undergo (it is a separate matter that, individually, we may not be fully aware of which languages have been shown in the literature to have this property, or what the alternative analyses are). A specific instantiation of inalterability[72] is that in Palestinian Arabic, syllable-final C_iC_i sequences do not undergo *i*-epenthesis where other comparable non-homorganic syllable-final CC clusters are separated by epenthetic *i*. Epenthesis is observed in /ʔakl-na/ → [ʔakilna] 'our food', and failure of epenthesis into the middle of a geminate is seen in [ʔimm-na] 'our mother'. This fact-pattern is of considerable generality across languages, common to the point that it became a desideratum to find a theoretical explanation for why the fact exists. Certain concepts and propositions have been advanced to explain this phenomenon, and these concepts and propositions constitute the theory of inalterability. This systematic fact pattern stands as partial justification for making such claims within the theory. The phenomenological concept "inalterability" then becomes part of the justification for a more general theoretical claim.

The connection to theory is made when the evidentiary concept (the explicandum) "inalterability" becomes the explicatum, because a causal claim is advanced about the representation of geminates as being the following:

(79) 

where α represents a single segment, "x" is a timing unit of some kind, geminates and consonant clusters have two such units, and lines indicate "association" relations. Certain generally-valid principles of phonological theory have the consequence that a vowel segment (x_3 linked to α_2) cannot be inserted between x_1 and x_2). Coupled with these explanatory principles, the concept *inalterability*$_P$[73] stands as evidence for representation (79) of geminates.

Another example of an phenomenological concept is "overcopy", which refers to a complication in reduplication. The main generalization about reduplication is that in some morphologically-defined context (to mark some meaning), a certain structural unit is copied to the right or left of the basic unit. "Overcopy" refers to the fact that in some languages, material from out-

[72] Historically "inalterability" and "integrity" were separate concepts, both of which pointed to a special property of geminates not undergoing rules. Inalterability referred to the fact that geminates resist featural changes such as becoming [+continuant] in a context, and integrity referred to the fact that rules do not separate geminates. These two subcases were treated as reflecting the same basic fact, and are terminologically unified here.

[73] The P-subscript notation identifies this concept as a phenomenological concept, that is, a recurring pattern of facts which appear to require an explanation. It is not an existential concept, that is, "inalterability" is not a thing in the theory like feature, rule or constraint are. OCP and Markedness enjoy, or suffer from, a mixed status in this respect: they may refer to an actual mechanism of grammar, but they may also refer to a fact pattern that may need to be explained by an existential concept.

side that basic unit is copied contrary to the general pattern, just in case certain rules apply which segmentally merge or resyllabify morphemic content that stands outside of that base unit with the base for reduplication. A concrete example is the fact that in Hehe, the entire stem of a verb is copied to form the attenuative, and this copying excludes the infinitive prefix or object prefixes (80a). However, an infinitive or object prefix is copied just in case it merges syllabically with a vowel-initial stem as in (80b). In addition, the regular rule nasalizing voiceless stops after a nasal applies to the initial stop of both copies of a reduplicated stem as in (80c).

(80)	a.	kú-[lim-a]	'to cultivate'
		kú-[lim-a]-[lím-a]	'to cultivate a bit'
		kú-lu-[lím-a]	'to cultivate it'
		kú-lu-[lim-a]-[lím-a]	'to cultivate it'
		kú-[telek-el-án-a]	'to cook for each other'
		kú-[telek-el-an-a]-[telek-el-án-a]	'to cook for each other a bit'
	b.	kw-[íit-a]	'to spill'
		kw-[íit-a]-**kw**-[iít-a]	'to spill a bit'
		kú-lw-[iít-a]	'to spill it'
		kú-lw-[iit-a]-**lw**-[iít-a]	'to spill it a bit'
	c.	/kú-n̩-[telek-él-a]/ → kúu-[nelek-él-a]	'to cook for me'
		/kú-n̩-[telek-él-a]/ → kúu-[nelek-el-a]-[**n**elek-él-a]	'to cook for me'

Overcopy$_P$ has been taken to be evidence for a number of theoretical conclusions, for example that phonological rules can be ordered before reduplicative copying, or the related conclusion that the constituent string copied under reduplication may be defined in part with respect to derived phonological properties. Many phenomenological concepts are available in the literature, for example natural class$_P$, heavy syllable$_P$, weakening$_P$, neutralization$_P$.

Phenomenological concepts are the observational currency with which we purchase theoretical causal concepts. They are very valuable, when they retain their essential nature of being a systematic integration of observations about language facts. We conclude that there are rules because there are recurring facts of language, which we integrate into various phenomenological concepts that stand for general classes of observed fact (paradigms of Damescene Arabic), and the theoretical concept "rule" provides a causal explanation for why the myriad facts are what they are. In arriving at a theory of the nature of the computational concept "rule", we directly organize the facts of language – the things that are to be explained – with one set of concepts which are part of the justification for the theoretical concepts. Then we add to that an account of what *causes* these facts, in the form of a theory of the nature of rules. We do not start by listing all and only the facts that we want to be explained by such-and-such theoretical notion, we start by being aware of all the facts that we can grasp, and we organize them into a conceptual system according to some perceivable similarity.

It is common practice in phonology to seek phenomenological concepts as evidence for a claim, or to hold theories responsible for explaining the facts represented by these concepts. When we claim that rules are formulated with conjunctions of features, we often refer at a phenomenological level to a very large set of cases across languages where certain segments function together as terms in a rule, and there is a pattern to this partitioning of segments into rule-term subsets: we reify the fact patterns into the concept of "natural classes". That phenemenon is, in some theories, the product of a claim about how rules are stated (as representations, which contains conjunctions of features), and the product of some other claims about the relationship

between segments as acquisitional phenomena, and features. Phenomenological concepts also have utility for theory comparison – how does OT handle "rule ordering"? How does rule-based phonology handle "conspiracies"? Phenomenological concepts may also be a curse, if we equivocate over what facts are referred to by "marked", "natural class" or "harmony".

In other words, with the aid of phenomenological concepts, we are able to comprehend numerous facts relevant to discovering the nature of rules, without having to arbitrarily define "rule" in some technical fashion. We have valid observational concepts, which support valid theoretical concepts.

5.2.4. AND FINALLY, RULE THEORY 1.0

Here is what the previous parts of §5 have done. First, we have explicated the relationship between justification and theory, reducing claims to sensory experiences (but usually via a chain of inferences). Claims are validated in a bottom-up fashion by ultimately being based on specific observations of linguistic behavior abstracted as token data, then type data, then elementary propositions about what the facts of the language are. At all stages, justification goes beyond simply pointing to a relationship between observations and claim, it also involves an evaluation of alternatives. We accept claims when alternative claims are not reasonably viable.[74]

Broad propositions about the structure of data in a specific language and in languages in general lead to the claim that there is some system of computation that relates underlying forms to surface forms – this is part of the even broader program of understanding the system behind form-meaning pairing. As part of the metatheory of investigation – the notions that define what thing we are investigating – we always bear in mind that we are concerned with the cognitive mechanisms that cause the fact patterns in linguistic data. We are not just concerned with abstract statements about the data.

Part of the theory of that relation which defines "phonology" is the concept of "rule", so going beyond the simple generalization that "a grammar is a set of ordered rules", we need to develop propositions about the nature of "rule", also (less evidently so, from a historical perspective) about "representation". A fundamental fact about a rule is that it changes a representation in a specific way, when certain conditions are satisfied. This leads us to focusing on specific questions about rules: what does a rule say about the conditions for application, and what do we say about the change that constitutes the action of a rule? We induce the properties of rules by observation of specific properties of rules, exploiting various phenomenological concepts to organise the specific observations made about Damascene Arabic, Takelma, Menomini, Chukchi, Logoori and so on. We will not have progressed to the advanced stage of theorizing that characterizes Autosegmental Phonology, Feature Geometry, and Substance-Free phonology; we will not have been enlightened by the works of Hale & Reiss, but we will also not have automatically adopted the practice of prohibiting at a theoretical level all fact patterns that are not yet established as existing. Obviously, what follows is my conjecture about what the theory of rules would have been, had the universe and especially the metatheoretical assumptions made been somewhat different.

---

[74] There is a large grey area where certain claims seem very likely but alternatives could be reasonably entertained. Depending on the comparative strength of the competing theories, we might say that we tend to favor A over B but do not claim to know that B is unlikely to be true; or we may say that we strongly favor A over B and seriously doubt that B is true. The force of feeling that A is correct is not what is important, what is important is the ability to say why B might be true.

One characteristic of phonology was largely established well before the theory of rules emerged in generative phonology, namely the theory of distinctive features. Although I have argued for and assume privative features, that argument depends on a sophisticated theory of structured representations. I take a binary or possibly ternary (+,–,0) theory to be the most self-evident account of representations at the assumed level of knowledge. A segment is a conjunction of a set of valued features. The exact set of features that most clearly "fits the facts" is not obvious, but for the most part there has not been a massive difference of opinion as to those features. The most problematic representation question would be how specific places of articulation are encoded, one possibility being that the standard IPA columns are either individual mutually-exclusive features, or that they are a scale of a dozen values on a POA continuum. Because groupings of segments that span multiple IPA columns are not rare, the question of how to account for "back consonants" and "coronals" as a class in rules would have arisen, but I doubt that the feature "anterior" applicable to both labials and front coronals would have ever emerged from a study of language patterns.

Inspection of numerous exemplars of what rules do would lead to the conclusion that rules describe sequences of segments which constitute the "where" of rules, for example "before X", "after Y", "between W and Q", where these letters represent some class of segments. This leads to a theory that rules are of the form "X Y Z" where Y is the class of segments being changed and X and Y are class of segments that trigger the change. Input class Y changes in a fashion which is definable in terms of features, thus we can write rules as mappings of the form XYZ → XQZ, or to use more familiar notation, Y → Q /X__Z. A question that would naturally arise is whether X, Y, Z are necessarily single segments, or might they be sequences of segments? While rules most commonly refer to one segment changing either before or after one segment, there are (in the literature assumed to be available) sufficient numbers of examples of operations requiring two-element sequences in the context, and examples of segment merger or segment-splitting, so that X, Y and Z are empirically one or two segments. Instances of three-segment contexts such as Ganda Law (l→Ø/__CV[+nasal]) are not common, but awareness of such cases would or should incline the theoretician to not hard-code into the theory a specific upper limit on how many segments X, Y and Z stand for.

One other lacuna that would be addressed is the well-known fact that segments can be inserted or deleted. We notate this with the special symbol Ø, so for the material before the slash, Y and Q may be Ø. Does this necessitate a special statement in the theory of rules, where we say e.g. "X and Z may be any number of segments, X and Z may be present or absent, and Y and Q but not both may be Ø"? Such a statement is unnecessarily complex: simplying saying "X, Y and Z may be any number of segments" includes the possibility that any of these rule terms are null. In simplifying the statement of what a rule is – "Y → Q /X__Z: X,Y,Z,Q are any number of segments" – we must also have clear propositions regarding the interpretation of rules.

5.3. Damascene Arabic rules 1.0

The reader may easily have lost sight of the fact that the whole purpose of subsection 5.2 is to enable us to posit and evaluate rules in Damascene Arabic, as part of the enterprise of incrementally inducing a theory of phonological grammar. We cannot discuss the import of Damascene Arabic for phonological theory until we *have* an initial theory of rules which the grammar of Damascene can be relevant to. We now have a primitive theoretical framework, so we can attempt to state rules for the language. The reader is referred to (66) for a reminder of the basic

paradigmatic facts, §5.1.3 for most of the relevant data, and §5.1.4 for the evidentiary propositions (proto-rules) that justify the specific rules that we posit. We have concluded that the relevant underlying forms of roots are /katab/ and /nəzel/.[75] There is a stress rule, which we will need to formalize, but its output is subject to two rules. A relatively easy rule to state is schwa-deletion.

(81) [+syl,–stress,–low] → Ø / __ [–syl] [+syl]

The facts alluded to above indicate that unstressed /a/ does not delete in this context, and that stressed schwa does not delete. Nothing else needs to be said about that rule. There is also a rule deleting unstressed /a/ before [Ce] (but not [Cu]).

(82) [+syl,+low] → Ø / __ [–syl] [+syl,–hi]

Given the paucity of contrasting suffixal vowels, the fact that /et/ triggers the rule and /u/ does not can be expressed in a number of ways: the choice of referring to [hi] as opposed to [round] or [back] is arbitrary, and outside the scope of the computational theory.

Another rule for which we have seen evidence is epenthesis, which inserts [ə] between two word-final consonants, where /katab-t/ → [katábət], a rule which can be stated as (83).[76]

(83) Ø → [+syl,–hi,–rd,–lo] / [–syl] _ [–syl] [+WB]

Since we are emulating the effect of increments in factual knowledge on theoretical knowledge, we are unaware that epenthesis also applies word-internally, for example in imperfective /b-yə́-ktəb-u/ → b-yə́-ktb-u → [b-yə́-kətb-u] – this issue does not arise in the perfective, which is the domain that we are aware of.[77]

The alternations between [e,o] and [ə] led us to positing a rule changing short /e,o/ to [ə] everywhere except before a single word-final consonant. Because there are no vowel sequences in the language, the extant contexts always involve __C, and the input vowel is always [–long]. Differentiation comes from considering what follows __C.

(84)

ə	e,o
__CC#	__C#
__CV	

[75] It has occurred to me that this might be /nezel/ since the rules seem to give the correct output from that input, but I don't at this point see any advantage to that particular abstraction.

[76] This rule is posited in a universe where we have considered the fact that word boundaries can be part of a rule, and where they have a featureal analysis as non-segments, marked as [+WB] – the SPE solution. In an alternate universe where we have never considered the role of word boundaries, this rule of Damascene causes a disturbance in the theory, which either complicates the theory of rules by including a disjunction "or #" in our proposition, or leads to the featural solution which we presupposed.

[77] The theoretical problem is that the rightmost element in strings that undergo epenthesis can be # or C, but not V. One solution is that boundaries are [–syllabic], or more generally, they are [–F] for all non-boundary features. Another is that there are two rules of insertion, one triggered by # in the right and another triggered by C on the right. A small systematic adjustment of the theory of features may be justified to the extent that it makes grammars simpler and more coherent.

We can now see what the common context is for reduction: on the right there is C plus another segment. The underlying vowel is retained when no segment follows, which means that the consonant is word-final. Under the assumption (already made) that there is a feature [segment] where segments are [+segment] and boundaries are [–segment], the reduction rule is (85).

(85) [+syl,–hi,–lo,–long] → [–round,+back] /__[–syl] [+segment]

It is now time to address that problem of stress assignment. The factual generalization is that stress appears in one of three positions. To aid the reader's comprehension of the patterns, the formal rule expressions [+syl], [+WB] and [–syl] are respectively replaced with V, # and C.

(86) a. V → [+stress] / __ C C #
V[+long] → [+stress] / __ C #
b. V → [+stress] / __ C C V #
V → [+stress] / __ C C V C #
V[+long] → [+stress] / __ C V #
V[+long] → [+stress] / __ C V C #
c. V → [+stress] / __ C V C V #
V → [+stress] / __ C V C V C #
V → [+stress] / __ C C V C V #
V → [+stress] / __ C C V C V C #

The complexity of this rule is a clear indication that something is missing in the theory, a suspicion confirmed by finding numerous similar rules in other languages. Rather than taking Theory 1.0 to be axiomatically the final theory, we entertain a few alternatives. One is to include parentheses, a device allowing the disjunction {X,Ø} within a rule, that is (to take a simpler case):

(87) X → Y / __ (Z) Q =
X → Y / __ {Z,Ø} Q

That is, X becomes Y whether or not Z follows but Q must follow – Z can be anything. If this is correct, we would try to build an expression that allows certain sequences to stand between the word boundary and the vowel which is stressed, and by carefully regulating the pattern which can precede the word boundary, we get exactly the result in (86). That leads to rule (88).

(88) V → [+stress] / __ C (C) (V[–long]) (C) (V[–long]) (C) #

Remember that the generalization is that consonant clusters delimit where stress is, and long vowels do too, so we simply write an expression that allows skipping over only short vowels and single consonants adjacent to them, and we limit the number of short vowels that can be ignored so that stress is at most on the antepenult. By adding this one device, we have greatly simplified (86).

In positing this device, we are obligated to consider the reasonable alternatives. One such alternative is explicitly stated in (87), that we instead use the disjunctive expression "{Ø,X}". We might try to decide between brace theory and parenthesis theory based on conceptual sim-

plicity, that is, carefully scrutinizing the set of concepts and propositions that have to be added to the theory under each of these hypotheses, to decide which is the simplest theory. To do this properly, we would need a more explicit version of Theory 1.0, so that we can put the competing propositions side by side, and since Theory 1.0 is inadequate, such a fleshing out would be mostly a waste of time except for pedagogical purposes, which is not a waste of time. We may find such a comparison in Part II. We might discover that the brace theory is actually simpler if the parenthesis theory includes everything in the brace theory plus the condition "where Z is Ø". Another avenue to consider is whether one theory makes a false prediction that the other theory does not make, that is, has one theory made a false identification of the nature of a concept? If, for instance, all cases of {Y,Z} refer to {Y,Ø}, then simple brace theory has not identified the correct concept, compared to what parenthesis theory identifies.

We are (no doubt) empirically aware of another alternative because of a wealth of already existing descriptive observations about "heavy syllables". The well-known solution is to have a mechanism (like parentheses) which allows skipping, but in addition there is some device that allows further compression of the strings in (88) – we can skip up to two light syllables.

(89) V → [+stress] / __ (σ̌) (σ̌) #

This adds all of the machinery of parenthesis theory, plus it introduces a variable device that defines σ̌ somehow. This is the dispreferred theory on grounds of simplicity (adding more propositions), though it does do a better job of identifying the correct concept. We will leave this as an interesting quandry, and we will simply note that when faced with a challenge to the theory, one must not just come up with a solution, one must compare that solution to other possible solutions, therefore one must be able to come up with two or three solutions and chose between them, on some rational grounds.

5.4. Theory 2.0

It would not take long for problems to arise with Theory 1.0 (some of which we touched on in the immediately preceding subsection). One issue which has been prominent in our knowledge of phonology for a century and remains not satisfactorily resolved is how to limit rules to particular grammatical contexts – since this is still unsolved, we will just ignore the problem for present purposes. Another is the fact that a number of rules take word-beginning or word-ends as part of their conditioning, thus some expression "at the beginning / end of a word" is needed. We would then add a symbol "#" to indicate word boundaries, and expand the characterization of what X, Y, Z might be, that is (one or more) sets of features or # – except that "or #" is only possible for X, Z but not for the segment undergoing the rule. We already touched on that issue in the preceding discussion.

5.4.1. MULTIPLE-FEATURE ASSIMILATION

One very prominent theoretical problems would have arisen in light of numerous common phonological processes, namely general assimilations.[78] It is very difficult in Theory 1.0 to account

[78] As a historical note, Halle (1962) proposes alpha notation in MIT RLE QPR 66, as the first published proposal for generative rule theory.

for general assimilations, where e.g. nasals become homorganic with a following consonant, or /l/ assimilates completely to a following consonant or a following coronal consonant. Well-known facts, widely available to phonologists at the time, would have impelled a theoretician to come up with a mechanism for saying "becomes the same as", w.r.t. some specific set of features.

The fact-pattern of Arabic l-assimilation is:[79]

(90) l → {t,d,s,z,θ,ð,ʃ,ʒ,tˤ,dˤ,sˤ,ðˤ,r,n} / __{t,d,s,z,θ,ð,ʃ,ʒ,tˤ,dˤ,sˤ,ðˤ,r,n}

One approach is to posit a series of seven specific rules assimilating the individual features *voice, son, cont, strid, ant, low* and *nas*, when those features appear in a coronal consonant. A problem with the 7-rule approach is that this assimilation only applies to the /l/ of the definite article (examples from Damascene Arabic).

(91)	γáltˤa	'mistake
	nzə́lna	'we descended'
	twaldan	'he acted childishly'
	sallti	'my basket'
	lsa:n	'tongue'
	xəlžá:n	'gulfs'

It is highly implausible that seven independent rules would be restricted to applying to just /l/ of the definite article.

One possibility for a single rule calls on the hypothesized "(X)" notation. The rule, so conceived, would be:

(92) [+lat,+DEF] → [(–voice) (–son) (–cont) (+strid) (–ant) (+low) (+nas)] / __ [(–voice) (–son) (–cont) (+strid) (–ant) (+low) (+nas) +cor]

This rule relies on a version of parenthesis theory, where (X) stands for the set {X,Ø}, and as in SPE theory, a rule containing (X) is a schema which in interpreted as two rules, one including X and one not including X. In the specification of a segment, [+A,(–Z)] expands to [+A,–Z] and [+A]. What that resolves to in (92) is that /l/ changes to acquire any feature of a following coronal which is different from that of /l/, without requiring that the following segment actually has the opposite value for all of the features. In ordinary words, this means "becomes [–voice] if the following segment is [–voice] but otherwise does not change voicing (since /l/ is voiced); becomes [–son] if the following segment is [–son] but otherwise does not change sonorancy…".

[79] Depending on dialect, θ,ð,ðˤ may not exist and ʒ may not exist or may not trigger assimilation.

Unfortunately, there is an important detail (problem) to be scrutinized (eliminated) in this theory. Looking only at the structural change matrix, and only considering the first three features in the set, the rule above represents eight subcases.

(93) {[–voice,–son,–cont], [–voice,–son], [–voice,–cont], [–son,–cont], [–voice], [–son], [–cont], Ø}

Likewise, the trigger matrix would represent the same set of subcases.

The problem is that as we have conceptualized segments and rule statements, ordering of features in a matrix is not significant – a segment is a *set* of features, not an ordered *n*-tuple. That means that nothing guarantees that [–voice,–cont] in the SC is selected only when [–voice,–cont] in the trigger expression is selected – nothing prevents the SC expression from being interpreted as [–son,–cont] when the trigger expression is just [–son]. Thus we expect it to be possible that /l/ becomes [t] before any obstruent, but this does not happen. Obviously, we need a mechanism for connecting the specific expansion of the expression in the SC to a corresponding expression in the trigger matrix. We could attach a coindexing mechanism to the parenthetic expressions:

(94) $$[+\text{lat},+\text{DEF}] \rightarrow \begin{bmatrix} (-\text{voice})_i \\ (-\text{son})_j \\ (-\text{cont})_k \\ (+\text{strid})_l \\ (-\text{ant})_m \\ (+\text{low})_n \\ (+\text{nas})_o \end{bmatrix} / \_\_ \begin{bmatrix} (-\text{voice})_i \\ (-\text{son})_j \\ (-\text{cont})_k \\ (+\text{strid})_l \\ (-\text{ant})_m \\ (+\text{low})_n \\ (+\text{nas})_o \\ +\text{cor} \end{bmatrix}$$

But this obviously complicates the theory in undesirable ways. We could, alternatively, expand the theory of rule expressions so that [A,B] denotes an unordered set of features and |A,B| denotes an ordered *n*-tuple, so that the order in which the rules are listed in the rule becomes significant.

(95) $$[+\text{lat},+\text{DEF}] \rightarrow \left| \begin{matrix} (-\text{voice}) \\ (-\text{son}) \\ (-\text{cont}) \\ (+\text{strid}) \\ (-\text{ant}) \\ (+\text{low}) \\ (+\text{nas}) \end{matrix} \right| / \_\_ \left| \begin{matrix} (-\text{voice}) \\ (-\text{son}) \\ (-\text{cont}) \\ (+\text{strid}) \\ (-\text{ant}) \\ (+\text{low}) \\ (+\text{nas}) \\ +\text{cor} \end{matrix} \right|$$

In the Arabic case, only a single segment undergoes the rule, so the input segment has a unique set of features. But nasal assimilations especially may apply to multiple places of articulation (though some are limited to one or two of the extant places). How would we accommodate a language where /m n ɲ ŋ/ all assimilate place from following /p t tʃ k/, where all possible place features are present in the input and all possible place features are assimilated? Restricting outselves

to just the features *ant, cor, high* to express consonant place of articulation, the canonical rule of nasal place of assimilation applicable to all nasals may be expressed as:

(96) $[+nas] \rightarrow \left|\begin{matrix}(+ant)\\(–ant)\\(+cor)\\(–cor)\\(+high)\\(–high)\end{matrix}\right| \; / \; \_\_ \left|\begin{matrix}–syl\\(+ant)\\(–ant)\\(+cor)\\(–cor)\\(+high)\\(–high)\end{matrix}\right|$

Alternatively within the set-only account, the individual features would be co-indexed as they are in (94), so that SC $(+ant)_i$ is selected only when trigger $(+ant)_i$ is selected.

Solution (96) is very similar to the familiar mechanism adopted in SPE theory, which expands the class of feature values, but only in rule expressions, to include general variables notated α, β, γ… If we reduce (96) to just one feature, (97a) expands to the 4 subcases (b.i–iv).

(97) a. $[+nas] \rightarrow \left|\begin{matrix}(+ant)\\(–ant)\end{matrix}\right| \; / \; \_\_ \left|\begin{matrix}–syl\\(+ant)\\(–ant)\end{matrix}\right|$

b.i [+nas] → [+ant,–ant]/ __[–syl,+ant,–ant]
ii [+nas] → [+ant]/ __[–syl,+ant]
iii [+nas] → [–ant]/ __[–syl,–ant]
iv [+nas] → [] / __[–syl]

Case (b.i) never applies since no following segment is ever [+ant,–ant]. Cases (b.ii-iii) are exactly what "→[αant]/__[–syl,αant]" says. Case (b.iv) requires some interpretive convention to the effect that "[]" is "no change", whereas Ø is "replacement with null". Interestingly, by switching the order in which values are presented in one matrix in (97a), but not both, we get the following result.

(98) a. $[+nas] \rightarrow \left|\begin{matrix}(–ant)\\(+ant)\end{matrix}\right| \; / \; \_\_ \left|\begin{matrix}–syl\\(+ant)\\(–ant)\end{matrix}\right|$

b.i [+nas] → [–ant,+ant]/ __[–syl,+ant,–ant]
ii [+nas] → [–ant]/ __[–syl,+ant]
iii [+nas] → [+ant]/ __[–syl,–ant]
iv [+nas] → [] / __[–syl]

That is, we get "→[–αant]/__[–syl,αant]".

Other technical details need to be specified to render these theories technically tenable as a treatment of assimilation, but since this is (in the author's opinion) the wrong direction to be going, I shift attention to the question of whether we have identified the correct concept, technical implementation notwithstanding. The fact that the theories which we have so far identified

potentially (in the case of the parenthesis account) or actually (in the case of alpha notation) allow feature switchings, e.g. αF….–αF, indicates that a broader concept is being identified, than what is involved in the data justifying the expansion of rule theory – assimilation, the case where segments become *the same as* another segment, w.r.t. some set of features. The foregoing mechanisms clearly don't just convey the notion "becomes the same", that notion is just one of a range of notions expressable. The (contrived) theory of parentheses doesn't convey any intelligible notion, instead it is a notation with a range of interpretations having something to do with "disjunction involving something with the empty string". The theory of alpha notation at least has the advantage of expressing the general concept "variable" ranging over the set of possible featrure specifications: for that reason alone, we will only consider alpha notation, and its next competitor.

The best alternative, at least given the assumed theory of representations, is, simply, to include "is the same" as a basic function in phonology. Borrowing a proposal from McCawley (1970), we may posit that $[=F_{ij}]\ldots[=F_{ik}]$ is true iff whatever value exists for F_i in position *j* is the same as the value for F_i in position *k*, which is mathematese for "is the same". Generic nasal place assimilation is thus expressed as (99).

(99) $$[+\text{nas}] \rightarrow \begin{bmatrix} =\text{ant} \\ =\text{cor} \\ =\text{hi} \end{bmatrix} / \_\_ \begin{bmatrix} -\text{syl} \\ =\text{ant} \\ =\text{cor} \\ =\text{hi} \end{bmatrix}$$

The interpretation of "=" is driven by the specific feature that bears it, thus a rule can say "takes the same value of *anterior* as the value of *anterior* that exists in the following segment", but a rule cannot say "takes a value attached to *anterior* that is the same as the value attached to *coronal* that exists in the following segment". Since such phenomena are not encountered in language, this approach, "identical value" theory, is conceptual superior – it better matches the phenomenological concept that is justified by the data.

Of course, the discussion ought to continue to explore the possibility that there actually *do* exist cases that are beyond the reach of identical-value theory, but the goal here is not to definitively resolve this technical question, it is to scrutinize the logic of answering the question. The alert reader may wonder, what about dissimilations? The answer, which depends on an empirical conclusion, is that there are no problematic cases requiring alpha notation. That is, no persuasive cases of the type "[αvoice] → [–αvoice] / X__" exist. The two best-known such examples are length-polarity in Dinka and voicing polarity in Luo, where vowels putatively swap length to form the plural in Dinka, and stops swap voicing in Luo to form the plural. The Dinka case derives from Gleason's workbook, where examples are cherry-picked to lead the student to a particular analysis. Plural formation in Dinka is remarkably complex and lexeme-specific, where plurals are less predictable than they are in Arabic. Luo voicing-switch is more robust, but still ultimately not true insofar as the fact pattern is that some nouns voice in the plural and some nouns devoice in the plural, and some nouns (having either voiced or voiceless stops) do neither, so it is an interesting question how the facts should be accounted for. As for general dissimilation, such as continuant dissimilation in Chukchi where /ɬɬ/ → [tɬ], this is a unidirectional change where homorganic continuent clusters change the first segment to a stop, but there is no analogous change of stops to fricatives.

Two considerations have driven the comparative evaluation of theoretical alternatives. One is economy: select the simplest theory over a more complex theory. The second is conceptual correctness – not just "can the theory grind out the data", but does the theory correctly state the causal principle that explains the data. When two theories differ substantially in what they "say", and one of these theories makes arbitrary claims (ones for which there is no supporting evidence), the theory that doesn't make arbitrary claims is superior to the theory that makes unsupported claims.[80] If we have clearly identified the concepts and propositions about concepts that comprise our theory of phonology, and if we have clearly identified the justification for these claims, we can evaluate their correctness.

5.4.2. STRING VARIABLES

In the case of vowel harmony, the problem for rule theory is that the segments relevant for the rule are vowels, and not consonants, and it is well known that (with a few exceptions) vowel harmony just ignores consonants. In the non-generative literature, the treatment of vowel harmony, and the irrelevance of consonants, received a number of treatments. One approach is to express the generalization as a property of words. Menomini has a long distance vowel raising rule which applies to /e:, o:/ conditioned by high vowels following anywhere in the word. Bloomfield says "One of the conditioning factors for the raising is the occurrence of *i* later in the word... Whenever the high vowels *i, i:, u, u:*, or the semivowels *y, w* after a nonsyllabic, occur later in the word, the vowel *e:* is raised to *i:*". There is a blocking factor, the vowel [ɛ:] – "An intervening *ɛ* or *ɛ:* normally prevents this raising". The long-distance nature of the rule is conveyed by referring to the trigger being "later in the word".

Whitney's statement of the Sanskrit ṇati rule, whereby n→ɳ if preceded by a retroflex continuant anywhere in the word, with dental, retroflex and palatal consonants blocking the process, is that

(100) The dental nasal *n*, when immediately followed by a vowel or by *n* or *m* or *y* or *v*, is turned into the lingual *ṇ* if preceded in the same word by the lingual sibilant or semivowel or vowels – that is to say, by *ṣ, r* or *ṛ or ṝ* – and this, not only if the altering letter stands immediately before the nasal, but at whatever distance from the latter it may be found: unless, indeed, there intervenes (a consonant moving the front of the tongue: namely) a palatal (except *y*), a lingual, or a dental.

Bogoras likewise treats Chukchi vowel harmony as being a property of word: they may only contain a particular combination of vowels:

(101) The vowels have been classified in three groups, – weak, strong, and neutral. … A word, simple or compound, must contain only strong vowels and neutral, or onle werak vowels and neutral, or vowels of one of the three classes. When, in composition, weak vowels and strong vowels come together in the same word, the former are changed by the ablaut into strong vowels

[80] In Odden (2013), these two principles appear to be ordered in such a way that simplicity takes precedence over conceptual correctness: that was my thinking in that work. I now understand that it is a mistake to presuppose that there has to be a ranking of these principles – the theory to be selected is both the simplest and the most correct among alternatives.

Sapir & Hoijer handle long-distance sibilant harmony in Navaho by saying "The consonants *s, z, j…* followed by a stem which contains one or more of the consonants *š, ž, ǰ, č* or *č'*, alter, respectively, to *š, ž, and ǰ*", thus the matter is handled by referring to a stem and the fact of containing a certain sound. It should be noted that the stem containing the triggering consonant does not have to be immediately after the harmonizing prefix.

A very common treatment, in describing vowel harmonies which are not of the strict "all vowels of the word must be…" type, is to refer the issue of consonant irrelevance to the fact that the process refers to certain syllables. Sapir describes Takelma *i*-umlaut as conditioned by *i* in a following syllable (also noting the iterative nature of the process).

(102) the process is a regressive assimilation of a non-radical *-a-* to an *-i-*, caused by an *-i-* (*-ii-*) in an immediately following suffixed syllable, whether the *-i-* causing the umlaut is an original *-i-*, or itself umlauted from an original *-a-*;

Bloomfield likewise appeals to syllables to account for another rule of Menomini, a vowel laxing rule which only applies between adjacents syllables, where "when another long vowel follows in the next syllable, they are often relaxed and shortened". Likewise, "If the preceding syllable contains a back vowel, the beginning of the *e:* is labialized". Lamb (*Mono grammar*) similarly states that "/u/ does not occur after syllables whose vocalic nucleus is /o/ or /oo/".

Pitkin (*Wintu Grammar*) deals with the problem of ignoring consonants by ignoring consonants, stating the rule as though consonants aren't part of the equation.

(103) Vocalic ablaut involves a variation in vowel quality conditioned by the preceding or following vowel…The morphophoneme |I| is phonemically /i/ preceding /a/, and /e/ preceding /i/, /u/, /e/, or /o/. The morphophoneme |U| is phonemically /u/ preceding /a/, and /o/ preceding /i/, /u/, /e/, or /o/.

Examples are |lIla| → /lila/, |lIlu| → /lelu/ – no mention is made of the fact that a consonant intervenes. It is interesting to note a different wording in a similar rule, a progressive total harmony rule: "The morphophoneme |V| is phonemically a short vowel of the same quality as the immediately preceding vowel". However, the morphophoneme |V| is actually never *immediately* preceded by a vowel, as seen in the examples /c'ot**o**, c'o:r**o**, c'e:w**e**, c'uq**u**/ which exemplify the pattern CV(:)C|V|.

Pitkin's wording seems to be accidental, but it suggests a possible account in the theory – sometimes a rule specifies that X *immediately* precedes Y, and sometimes X simply precedes Y. A segmentally-local assimilation might then be formulated as (104a) and a vowel harmony rule that can skip segments is formulated as (104b) – the rules differ in whether the segments merely precede, versus immediately precede.

(104) a. $[+x]_a \rightarrow [+y]/ \_\_ [+y]_b$ where *a* immediately precedes *b*
b. $[+x]_a \rightarrow [+y]/ \_\_ [+y]_b$ (where a precedes b)

An alternative notation would be:

(105) a. [+x] → [+y]/ __ [+y]
b. [+x] → [+y]/ __ X [+y]

where "X" means "precedes, not necessarily immediately".

The main problem for a theory attributing segmentally long-distance relations to a difference in precedence types is that there is no obvious way to talk about blocking conditions, such as the fact that Menomini harmony operates across vowels except for /ɛ/, or that the Sanskrit rule ṇati is blocked by lingual consonants. Such cases are very rare, and might not have caught the attention of phonologists at the beginning. Such examples are a problem for this simple proposal, but the counter-example has no affect if we are unaware of the facts.[81]

Another possibility is to incorporate some notion of recursive string variable, that is to write harmony as (106).

(106) $[+x] \rightarrow [+y] / \_\_ \ X^* \ [+y]$

Here, "X*" simply means "any number of segments". As such, this is not better at a conceptual level than the (immediate) precedence solution, and it is worse, because it introduces a much larger set of novel concepts into phonological theory. This does, however, have the advantage that it can more easily generalize to include complex conditions on material between target and trigger. It is not hard to see that statements analogous to (106) could encompass lingual blocking in Sanskrit, blockage by [ɛ] in Menomini, by allowing "X" to have various feature values.

The SPE approach involved a lot more apparatus (conceptual expansion), and even on its own terms is more expansive than necessary for long-distance rule application.[82] It would suffice to introduce the notation "[…]*", that is, a segmental matrix plus the Kleene-star operator – any number of segments which are […]. The actual SPE theory of rules involves operators for grouping linearly ordered expressions of arbitrary content, disjunctive collections of such expressions, and repeat-operators specifying upper and lower bounds. Thus rules might involve C_0, in the most common case of vowel harmony, $(C_0V_0)_0$ or equivalently $[+seg]_0$ for a rule applying at any distance, or $[–coronal]_0$ for a rule applying across any number of non-coronals. The full SPE account countenances patterns such as "a→æ when preceded by [i], separated by any multiple of 3 occurrences of the string [p][l][a]".

(107) $[a] \rightarrow [–back] / [+hi,–back] (([p][l][a])_3^3)_0 \ \_\_$

Since nothing approaching this pattern is found in language, we don't have empirical justification for including the range of concepts implied by the system of SPE abbreviatory devices. There *is* good evidence for some device conveying the notion "identical" for a feature, and some account is needed of the fact that rules sometimes apply with variable-length strings of segments stading between target and trigger though that may (does) depend on a refined theory of representations.

To wind up this section, the point made here is that justification plays a essential role in theory construction. Justification specifically means, showing that the evidence indicates the

---

[81] Many phonological facts have been established in the literature, and then forgotten. Overcopy in Tagolog reduplication was forgotten for almost a decade, but seems to be part of the "everybody knows" corpus, unless it has resubmerged.

[82] A smidgen of support for the full scheme comes from unbounded quantity-sensitive stress assignment rules, which were not considered in SPE.

truth of the proposition, and the falsity of competing propositions. This system of evidence requires building up knowledge from the lowest-level of the senses, proceding incrementally to higher level abstractions about the senses and the concepts integrating what we have sensed.

6. Conceptual inheritance and hierarchy

The logical framework advocated here relies on concepts being hierarchically arranged. That is, gaining conceptual knowledge is not just the memorization of a huge bag of concepts, it is the acquisition of a hierarchical system of concepts and propositions about concepts. What does that mean?

6.1. Kinds of hierarchies

A simple example of a hierarchy is biological taxonomy, which is an arrangement of the myriad concepts pertaining to the evolution of living things. A concept is a mental integration of some group of things under one label, where the grouping is based on the perceptible similarities of the individuals subsumed under the concept, compared to other individuals which are not included. At the lowest level, the basic individuals are particular dogs. We group individual dogs into various first-order concepts like husky, malemute, samoyed etc. Those breed concepts are then integrated into a second-order concept "dog", which then is integrated with other similar non-dog lower-order concepts like "wolf", "jackal" and "coyote", resulting in a third-order concept *Canis*. We continue the process to arrive at *Canina* (integrating groups of dog-like animals such as "short-eared zorro", "bush dog", "maned wolf"), then *Canidae* which includes foxes, continuing upwards to *Mammalia* and ultimately the "kingdom" *Animalia*.

Hierarchy follows from the nature of concepts. *Qua* concept, *Canis* is not made up of all of the individual wolves, coyotes, jackals and dogs, it is made up of (integrates) the concepts wolf, coyote, jackal, dog. Higher-level concepts are not bags of increasing size containing all of the individuals, they are small folders of folders: the order of folders is the hierarchy.

Taxonomic biological concepts generally constitute a claim about the historical development of species. By the nature of evolution and species differentiation, the individuals making up the immediately subordinate concepts have all of the properties characteristic of the superordinate concept, with the addition (replacement) of some properties that characterize the distinctive properties of the various immediately subordinate concepts.[83] Dogs can be classified according to functional criteria as well, such as sporting, working, herding, toy, hound and terrier. Again,

[83] I am mixing pre-contemporary systems of classification to simplify the exposition. Speciation probably represents successive binary splits, and not 4-way simultaneous splits into wolves, coyotes, dogs and jackals – the Linnean system. More recently, the unity of dogs and wolves is reflected in modified terminology, when *canis familiaris* became *canis lupus familiaris*. I also set aside the non-canonical "wolf" species of Africa, *canis anthus* and *canis simensis*, which are less closely related to coyotes, dogs and wolves = *canis lupus*.

just as husky, malemute, samoyed and so on are kinds of dog, so are sporting, working, herding, toy, hound and terrier.[84] The difference is that similarity and differentiation is based on different criteria: function, not evolutionary history.

Another hierarchical organization of concepts is "metal", a group of substances with common properties of conductivity and malleability or ductility. Subgroups of metals are functionally or compositionally defined, such as ferrous metals (containing appreciable amounts of iron), refractory metals (extraordinary resistance to heat and wear), white metals (white-colored with low melting points), base metals (easily oxidized or corroded) or noble metals (resistant to oxidation and corrosion). Ferrous metals subdivide into numerous other types such as cast iron, wrought iron, steel, staineless steel, based on other properties. Cast iron, wrought iron, steel and staineless steel are all types of ferrous metals; ferrous metals are a type of metal.

Geometry provides myriad examples of conceptual hierarchies. "Polygon" (in the ordinary sense) is a connected finite series of line segments on a plane where the first vertex coincides with the last one, has no holes, and does not intersect itself. There are very many (indeed infinitely many) sub-types of polygon, such as triangle, quadrilateral, pentagon, hexagon and so on, each concept being similar in having the properties of a polygon, and differing from each other in terms of the number of line segments that compose the sub-types. Quadrilaterals are subclassified as rectangles, parallelograms, kites and so on based on properties of their angles; squares are a subcase of rectangle. Triangles are classified by two orthogonal properties, length of side (equilateral, isosceles, scalene) and internal angles (right, oblique, acute, obtuse). An isosceles right triangle is a kind of right triangle, and is also a kind of isosceles triangle. Right triangles are kinds of triangles, isosceles triangles are kinds of triangles. The idea of multiple orthogonal classifications is trivial to any phonologist who know a theory of distinctive features.

The concept and word "hierarchy" itself arises from medieval Christian theology,[85] and refers to an ordering of nine types of celestial beings set forth by Pseudo-Dionysius the Areopagite in "On the celestial hierarchy" (Περὶ τῆς οὐρανίου **ἱεραρχίας**) ch. 5-10. This theory of celestial beings primarily distinguishes according to sphere[86] of influence using a concept termed "order" or "choir", and within each order there is a ranking (Seraphim are above Cherubim, which are above Thrones – it is unclear to me what the order-internal ranking reflects). Unlike conceptual systems of biology, metallurgy or geometry, reduction of these nine being-concepts into units bypasses the ultimate step of actual concept formation where the unity is recognized by special lexicalization (that is, the first, second and third orders don't have special names, they are simply labeled compositionally with an ordinal number and the concept "Order"). Seraphim, Cherubim and Thrones are not types of Order, they are the units that *constitute* an Order. This exemplifies another kind of hierarchy, a compositional hierarchy, where a concept is said to be composed of other units – they are not the integration of distinct units into more general "is a" concepts.

Military rank illustrates the partial independence of "concept" and "hierarchy". Taking the structure of the US Army as an example, members of the military can be divided into two

---

[84] As it happens, these distinctions are interposed between "dog" and individual breeds – they organize breeds, but not necessarily based on evolutionary facts.

[85] I have not submitted this summary to any theological scholars, so there is a high probability that I've missed or incorrectly stated details. I am sure that such a hierarchy was posited.

[86] Literally, a sphere on which they reside, where the first sphere is closest to the divine center.

types: officers and enlisted personnel.[87] Enlisted personnel are subclassified into other types, e.g. private, private first class, corporal, sergeant, staff sergeant all the way up to sergeant major of the army. This is an bare ordering of types: a sergeant is "above" a corporal, a corporal is above a private. This is not an "is a" relationship: a private is not a kind of corporal and a corporal is not a kind of sergeant. But the officer / enlisted distinction is an "is a" classification.

Many hierarchies are compositional in nature. For example, a molecule is an electrically neutral unit composed of two or more atoms. An atom is composed of a nucleus and a number of electrons. A nucleus is composed of protons and neutrons. A proton is a particle with positive charge composed of 3 quarks mediated by gluons, and a neutron is a particle with no charge composed of 3 quarks mediated by gluons. Quarks are not kinds of molecules or atoms, atoms are not types of molecules. Any proposition of the kind "W is composed of X, Y and Z" creates a hierarchy.

A third kind of hierarchy is a presuppositional hierarchy. This is the situation where the definition or essential propositions about a concept logically depend on some other concept. For example, the definition of "polygon" depends on the concept "plane". A polygon is not a kind of a plane, and a plane is not composed of a polygon and something else. Instead, simply, you cannot understand what a polygon is unless you understand what a plane is, and you cannot understand what a plane is until you understand a dimension. Point, plane and polygon *are* concepts within the domain of geometry, making the distance between these concepts relatively short.

Concepts which we call on to understand other concepts are frequently of such generality or orthogonality that they are simply "things we know". A "drawer" is a box-shaped container closed at the bottom and open at the top, which fits into a piece of furniture so that it can be pulled out horizontally. "Drawer" has a hierarchical relationship to "furniture", because it is a part of some particular piece of furniture. "Furniture" is a movable object used to support actions such as seating, eating, sleeping, or to hold objects at a convenient height for work, viewing or storage, thus "drawer" may be an aspect of a desk, a cabinet or a table, all of which are kinds of furniture. The hierarchical relationship between furniture and drawer is compositional in nature.

But the essential properties of a drawer include being box-shaped, being a container, being closed at the bottom and open at the top, and being designed to be pulled out horizontally. To distinguish drawers from other things in the furniture kingdom such as cutting boards and table legs, we use concepts like box, container, closed, top, bottom, pulling and horizonal, but these are not at all concepts introduced under the domain of the theory of furniture, they are autonomous concepts, from the perspective of a theory of man-made artifacts. Presuppositional hierarchy, where understanding some concept requires prior understanding of another concept, is ubiquitous and the most challenging for theory construction.

Hierarchies order concepts based on the logic of a concept. Why do we care about such an ordering? First, we have to understand "*prior* knowledge" in order to properly structure a particular idea. We have to know, when we advance a theory of "phonological rule", what we are positing at the moment and thus what needs to be proven – as opposed to that which has already been proven. Second, we need to understand the consequences of error for our theory. In a well-structured theory, a speculative concept such as "structural constraint" can be contemplated on the basis of weak evidence, but it should not the the foundation on which all other phonological

[87] This is the position of the US Army. Other classifications make a primary division between officers and non-commissioned officers, where the latter are enlisted personnel who can command others below them. I omit warrant officers because I don't understand them.

ideas are grounded. This is a restatement of the premise made throughout science that theories should be built in a solid foundation. If you understand the hierarchical relation between theoretical ideas, you can say what ideas are foundational. Awareness of the applicable hierarchy of ideas in a theory also allows one to avoid logical circularity – the situation where X is defined by whether a thing is Y, but Y is defined by whether a thing is X. A "segment" cannot be defined as a unit having a collection of "features" if "feature" is defined as a "phonetic property of a segment". The typical generative approach to this problem is to take "feature" to be existentially primitive, and to define "segment" based on "feature".

6.2. Hierarchies and inheritance

An optimally-structured conceptual system is "local", in the sense that relations are stated between a superordinate concept and its immediate daughters. This follows from the nature of conceptual structuring, where one concept integrates other concepts, or one concept "is composed of" other concept. Even though it is true that a husky is a type of animal, we would not say such a thing in a theory. We would also not seek to directly distinguish huskies from bears, we would seek to distinguish *Canidae* from *Arctoidea* (which integrates bears, racoons and pinnipeds), within the concept *Caniformae*. Instead, the conclusion about huskies follows from the fact that a husky is a kind of dog (etc), and from the hierarchy of concepts we can *infer* that a husky is a kind of animal. We do not distinguish huskies from rocks, or from the act of tree-felling, or from radio waves. We only distinguish huskies from their conceptual sisters which taken together form an immediately higher concept – huskies are a type of dog.

A proper epistemology does not state all of the imaginable true propositions about a thing at all levels of knowledge, it states the essential ones. With the aid of the *hierarchy* of the concepts standing between "husky" and "animal" or other concepts, we derive the conclusion that a husky is an animal, and not a rock. Without a hierarchical system organizing concepts, we would not know the proper treatment of the husky-rock distinction. Huskies *inherit* the property of being animals, from this hierarchical organization of concepts.

As a linguistic example of conceptual inheritance, the concept "rule" is not exclusive to phonology, so phonology does not have to completely reinvent or redefine the wheel in invoking a notion of "rule", instead, we inherit the concept from a higher domain. A phonological rule is indirectly a kind of cognitive rule, but it is directly a kind of linguistic rule. In a properly modular and local theory epistemology, we would *not* say (stipulate as part of the theory) that a phonological rule is a type of rule or a type of cognitive rule, we would say that it is a kind of grammatical rule, referring to the theory of grammar. We are only concerned with the characteristics that distinguishing it from other kinds of grammatical rules. Therefore, linguists in general are interested in just the properties that distinguish "linguistic (grammatical) rule" from other kinds of cognitive rule, and phonologists in general are interested in just the things that distinguish phonological rules from semantic, phonetic and other kinds of linguistic rule. The properties that distinguish linguistic rule from other kinds of cognitive rule are thus inherited by phonological rules (idem semantic, morphological, syntactic and phonetic rule) from the superordinate concept of linguistic or grammatical rule.

Similarly, we do not need to rediscover or redefine the concept "representation" in phonology, we only need to distinguish phonological representations from other kinds of representation – being local, from other kinds of *linguistic* representations. We derive the fundamentals of the phonological concept from the higher-order concept of "linguistic representation", and that

concept in turn derives its fundamentals from an even-higher level concept of representation. Phonological theory only specifies the special properties of phonological representations, and linguistic theory only specifies the special properties of linguistic representations.

Inheritance is a one-way street, in that discrimination of a class of existents into lower-level concepts only adds information to the properties of the higher-level unit. An egg-laying dog would contradict the presumed properties of *Carnivora*,[88] *Caniformia, Canida* and so on, at least under the evolutionary interpretation of taxonomic categories. The invention by Hale & Reiss of *homo collitumens* reminds us that we must focus on the "historical sequence of events" aspect of biological taxonomy, thus a dog or human who subsequently developed egg-laying abilities does not contradict any conceptual properties – it just adds "and later started to lay eggs". Conceptual inheritance means that if a property of a higher order concept is "they lost the ability to lay eggs", the lower order concept cannot have the property "they never lost the ability to lay eggs". Reversal of neutralization in historical linguistics (the effect of final devoicing being eradicated in the history of Yiddish, the loss of the change *dʒ → ʒ in Ukrainian) is a classic linguistic analog, that the historical change imputed to the languages is not magical reversal of the learned representations – underlying forms – is it loss of a rule that affected the learned representation.

Automatic inheritance is, however, only a property of lower-level concepts with respect to higher-level concepts that integrate them. The other major hierarchical relation between concepts,[89] the compositional hierarchy, does not entail inheritance. That is, the objects that compose atoms are not themselves atoms, and the essential properties of atoms are not (necessarily) inherited by its constituents; the parts of a car are not themselves cars. The structural description and structural change of a rule are not themselves rules, however, all three *are* representations. Thus it is possible that the components of some concept will share a type with the higher concept that they are constituents of. In structuring a theory of these concepts, we would therefore start by saying that "a structural description is a representation which…", "a structural change is a representation which…", thereby passing the general properties of phonological representations down to SD and SC.

In a compositional hierarchy, the objects comprising the higher-order concept are each classified as being of some essential type. The immediate components of an atom are a nucleus and a kind of particle. "Nucleus" is, in general, a central mass (originally, the nucleus of a comet, also the umbra of a sunspot, the core of a galaxy, a particle leading to crystal formation, the nucleus of a cell, a small bee colony including a queen, and now including the center of an atom).

6.3. Prelude to the hierarchy of phonological concepts and propositions

We will attempt to refocus on phonology, but will get briefly distracted. The theory of phonology is an aspect of the theory of grammar, which is an aspect of the theory of mind. We can try

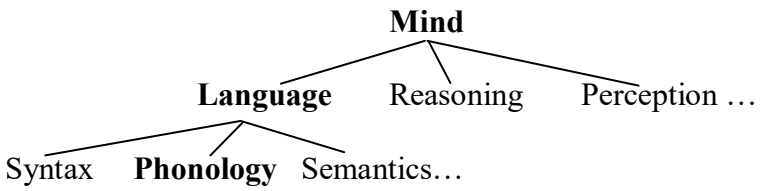
---

[88] Inconveniently, there are egg-laying mammals, the monotremes. There is controvery over the details, but it seems that *Marsupiala, Monotremata* and *Placentalia* are intermediate concepts, where *Australosphenida* (which includes monotremes) was a first branch from what are known as "crown mammals", then numerous other branches leading to *Metatheria* (marsupials) and *Eutheria* (placentalia). A number of extinct intermediate animals appeal to be non-placental, such as *Allotheria*, but evidence regarding egg-laying is sparse to non-existent. Thus "not egg-laying" is a property of some higher classification of mammals, but below the level of "crown mammal".

[89] It appears that the union of these studies is known as "mereology", but a lot of what I have seen about that area seems to be tangential.

to convey this by a graph such as (108). The point of this excursus is to underline the risk of assembling words into a structured graph, without clearly saying thing what the graph represents.

(108)



The relationship between these terms is not the "type of" relations that we see in a genus-species analysis of entities. Language is not a kind of mind, reasoning is not a kind of mind; syntax is not a kind of language, phonology is not a kind of language… The highest-order relations are compositional relations, that is, the mind is composed of these things.[90] What type of thing is it that makes up the mind? At least w.r.t. these aspects of the mind, these lower-order concepts are "faculties": language is a mental faculty, as are reasoning and perception. "Faculty" means "ability", a potential and not an actual act. When actualized, the perceptual faculty acting on sense data results in the percept of an apple or of a stick in water. The faculty of reason acting on other knowledge and sense data may yield a proof that water bends light, or give reason to doubt that the sky is falling. What is the Language Faculty (LF – *not* "Logical Form")?[91]

As a concept, "LF" integrates all of the LFs existing in all of the minds out there, which kicks the can down the road only a short distance. In linguistic theory, there has been some unclarity or equivocation about specific grammars, LF, and UG, and it behooves us to make explicit some of the issues. Under certain standard generative assumptions (see Odden in press for some discussion of views of UG), LF is a fixed cognitive ability, and it is not the same as the grammar of a particular language. The LF of a specific individual *enables* the learning of the grammars of English, Logoori, Tamazight… so the LFs of all individuals, as a whole, *enable* the learning of the grammars of English, Logoori, Tamazight… I leave unanswered for the moment the question whether LF is anything more than this ability, but obviously the simplest theory is that that is all that there is to LF.

So next we ask, how does LF enable this learning? The most important thing LF does is say what "a language" is, in the mental sense: what is learned? A "language" from the I-language perspective is composed of a grammar, and an associated data structure generally called "the lexicon". LF then states what the structural properties of any grammar must be (e.g. describes the components of a grammar: embodies the theory known as UG). When a person learns a (particular) language, that means they learn the grammar and its lexicon (the rest of the "how" is primarily a specification of what a grammar can be). The grammar of Logoori, for example, is itself a specific data-structure, containing structured sets of rules or other things which constitute the phonological, morphological, phonetic… components of the language. The phonological component of Logoori is composed (at least) of the phonological rules of the language, and the ordering of the rules of the language. A mind can contain multiple languages, meaning that mind can contain multiple datasets composed of grammars and lexicons.

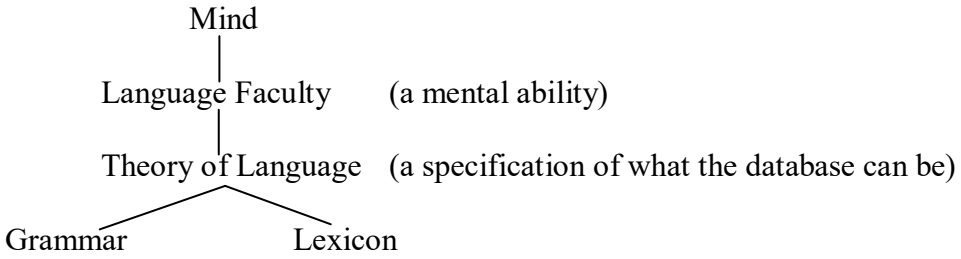
---

[90] There is ample confusion about the dichotomy *mind* versus *brain*. What the mind does is produced in part by the brain. I have no intention of touching the third rail of the mind-body problem, and will only talk about the mind.

[91] To mildly reduce the speculative content of this work, I will not argue that the faculty of language is an essential component of the faculty of reason.

The problem with a picture like (108) is that it equivocates over what "language" is, even from the mentalist perspective. When we distinguish "Language" from other aspects of the mind, we are talking about an ability, embodied for language as a statement of what a language "is" – a particular kind of database. Thus LF provides a (homolog of a) general scientific theory of that database, and what is in it. That guides the acquisition of specific languages in some way (it limits the nature of the learned database that constitutes "a specific language").

(109)



For the moment we will just assume that the Language Faculty is only composed of the Theory of Language, that is, the architectural specification of language in the form of a theory of grammar and a theory of stored representations. In the last subsection of this section we discuss the concept of The Engine, which "does" what the grammar and lexicon tell it to do.

In other words, the result of successful generative theoretical investigation in phonology is not a theory of phonological grammars, it is a theory of the theory of phonological grammars. The theory of phonological grammars is a mental object, which is about the instruction-template available during language acquisition controlling what a phonological instruction (rule) or computation can be. This distinction is important in part because it introduces additional causal principles that mediate between direct perception à la §5.1.1 and the final conclusion about what the Language Faculty is. More important, from the perspective of this section, is that phonological theorizing is a model of a specific kind of epistemological system.

We now return to more concrete questions regarding the structure of phonological theory.

6.4. The hierarchy of phonological concepts and propositions (for real)

Above, I have set forth the following essential concepts for a theory of phonology, some being undeniable[IMO] and some being plausible. These are the concepts of the present theory, not the historical intermediates, and they are not the evidentiary concepts formed in the course of arriving at this system. However, these are not necessarily concepts strictly *of* phonology – that is the question which we are now investigating.

(110)

node	precede	dominate
type	rule	structural description
structural change	insert	delete
focus	structural constraint	denial
identity	edge	

6.4.1. CONCEPTS NOT IN PHONOLOGY

The reader may be shocked at the omission of concepts actually used in phonology, such as:

(111)	phoneme	contrast	underlying form
	segment	counterfeeding	dissimilation
	floating	cycle	

On the assumption that the reasons for the omission of (111) are not self-evident, we quickly consider why these concepts are off the table. In case of doubt, this epistemological framework is not all-inclusive, so excluding concepts from the domain of a theory is a virtue, but of course nihilistic rejection of all concepts from theory is not a virtue. Only *necessary* concepts are part of phonological theory: the above are not necessary concepts. The fact of people using a concept does not confer scientific validity on the concept.

The concept "phoneme" is about particular kinds of representations. It refers to a specific kind of segment (which we will discuss), and asserts that segments with a certain derivational property ("biuniqueness") have other special properties – the rule assigning aspiration in English has that property, the rules of epenthesis of devoicing which apply to plurals (*bushes, cats*) do not. Why is "phoneme" not included here? Because there is no empirical justification for including the concept in the theory – there are no *special* phonological properties attributable to the segments that have the distributional relation of biuniqueness. The concept and specific details was historically mandated in part because of a particular epistemological framework (positivism), that is, the justification is purely conceptual – and the conceptual foundation is invalid. Outside that framework, the concept "phoneme" (as taxonomically conceived) does nothing for phonological theory. It does relate to the concept of "allophone", but that concept is also not of use to phonological theory. But to be clear: the concept phoneme, also allophone, *was* historically useful, as an approximation of the distinction between categorial units of sound as found in feature structures – phonological representations – and the physical manifestation of phonological representations, coming from the phonetic component. To the extent that past phonologists used these terms to distinguish between phonological computations and phonetic realization, these *were* useful ideas, like "ether" was in the history of physics.

The concept "contrast" is a relationship between phonemes – two segments "contrast" iff they are distinct phonemes.[92] This is a slightly more useful concept, but still no fact about language justifies adding "contrast" to phonological theory. Instead, "contrast" was a tool in making the phoneme / allophone distinction. We might seek to identify the [ɪ] of [fiʃ, dɪʃ, dɪg] by reference to the fact that those segments do *not* contrast within these words, but even then, the identify of [ɪ] is expressed in the present theory by the fact that they have the same representation, that is, the same sequence of nodes and relations – they are the same type of thing.

Surely there can be no quarrel with the concept of an underlying form. Indeed, it does refer to something meaningful, but it is not a significant concept for *phonology*. The underlying form in phonology is simply the input to the phonological component: it has no special properties to which phonology must refer. In any reasonably modular theory of language, each component has an input and an output, and *that* is a property of Language, inherited by each component of a grammar. It follows from the hypothesis that a (whole) grammar is a series of computations where the output of one component forms the input to another, and phonology has an input. No

[92] An alternative and narrower view of contrast is that two segments X, Y contrast iff there are two linguistically distinct forms which are identical except for the choice of X versus Y, i.e. if there exists a minimal pair.

property of phonological rules, representations or computations is stated in terms of whether something is an underlying representation.[93]

The concept of segment is, again, a historically and pedagogically useful concept, being a nearly self-evident basic which is combined to form syllables and words, but there is nothing technically special about "segment". A segment is simply whatever sequence of nodes are dominated by a root node. For any non-terminal node, one could reify "the set of nodes dominates by node J". The issue for theory is, does the collection of nodes dominated by root have any special and essential property, e.g. is that collection treated distinctly from the collection characterizing Place, Laryngeal or Tone? In lieu of evidence for segment as a significant grammatical concept, we would not posit such a theoretical concept.

The concept of counterfeeding functionally identifies a kind of rule-ordering relationship, specifically where a later-ordered rule creates a string that could have undergone an earlier rule, but because of the ordering of rules the earlier rule does not apply – therefore the later rule creates what look like counterexamples to the earlier rules. The concept was postulated in the SPE era in an attempt to explain some patterns of historical change, but its relevance to the theory of grammar has been mostly typological, being a means of referring to certain classes of rule ordering which have posed problems for the Unordered Rule Hypothesis of Koustsoudas, Sanders & Noll or for Optimality Theory. As such, it is a phenomenological concept relevant to the theory of ordering in grammar, but is not a theoretical concept itself. The concept of dissimilation, a sub-type of phonological operation, is the change of one of two identical values for some node. Dissimilation is also a phenomenological concept. Its utility is only as a source of information about what the theory must account for, but it is not a distinct part of the theory itself.

The concept of floating has an uncertain status in grammatical theory. A floating node is a node not dominated by the "usual" dominating node, for example a Tone node not dominated by a mora or a mora not dominated by a syllable. A specific mechanism has been proposed in §4.4 to account for the class of facts known as floating specifications, via the "denial" subproposition, whereby rules can apply in a context where "~P" it true. "Floating" is a phenomenological concept that leads to this broader concept – it is one kind of evidence for the concept.

The concept of the cycle, which we have only alluded to, is omitted from the privileged list in (110) in part because we have not considered the evidence for the cycle, so it has not been justified, at least w.r.t. the present discussion. However, "cycle" is also not an obviously necessary and useful grammatical concept, because again it seems to describe the results of some aspect of grammar, and is not itself an aspect of grammar. The specific causal mechanism is something filed under the theory of phonological grammar, where under conditions related to the grammatical structure of a representation, the output to phonology becomes the input to phonology. So while some additional theoretical concept is necessary, it isn't "cycle".

6.4.2. CONCEPTS THAT **ARE** IN PHONOLOGY

Having denied phonological-theoretical significance to a number of popular terms used by phonologists, we should then review the set of concepts that were not rudely dismissed, those in (110), to verify that they are properly concepts of phonological theory (they are, with some reservations). The difficulty in making that judgment is the problem of dismissing the alternative

[93] In standard rule-based phonology. There have been rule-based theories – global rules – positing that computations can consult the input to phonology, and OT posits a class of constraints which refer to the underlying form.

that these are higher level concepts pertaining to Language or Mind. These concepts all play a role in the theory of phonology, but which of them are exclusively phonological, and which are modifications or applications of concepts from other domains – how much of the content of phonology is inherited from other domains?

We start with the four concepts *node, precede, dominate* and *type*. Each of these concepts has an analog in syntactic theory. What *Merge* produces in syntax is a node, one that dominates two other nodes. There is some question of where in syntax the concept "precede" becomes relevant (word order is traditionally syntactic), and there are thirty million theories of syntax. The concept "type", exemplified in syntax by N, V, A etc. and bar levels, has fallen into disfavor in some theories. But actually, these are not purely linguistic concepts, they are general cognitive concepts. That is, these are essentials of cognition, and while they may have some phonology-specific properties, the concepts are not just invented for the purposes of phonology. Consider the general cognitive task of organizing information, such as the relationship of dogs to everything else in the universe. Every concept is a "node", and the idea of hierarchy captures what dominance captures. Concepts are not just pure organizing points, they are labeled – they have types. Grammar and phonology exploit the pre-existing concepts of node, dominate and type (though they are not necessarily given these names in other theories of mind).

The concept of precedence or ordering is likewise a very general cognitive concept, which conveys position relative to a reference point. The chair is closer to me than the table, the white dog is first in a sequence of sitting dogs, looking left-to-right, physical causes precede their effects in time. By *some* method, the syntax determines that in the English sentence "The dog is on the log", the word "the" precedes "dog" and "dog" precedes "is".

All evidence says that the four concepts *node, precede, dominate* and *type* are the same in all domains of grammatical theory and there is no evidence for a phonologically-special sense of "node" etc, but there are some ways in which we use these concepts specially in phonology. Thus we would simply say that these concepts are directly inherited in phonological theory from linguistic theory. In phonology, representational ordering is defined on nodes of the same type, but that is not the case in syntax (and is certainly not a property of ordering in general). Thus the concept "precede" is involved in a special phonological way, meaning that there is a special proposition in phonology. What does that proposition modify? It is not "precede", it is "representation" – the claim is that in a *phonological* representation, only nodes of the same type have a precedence relation. Or, only nodes of a different type have a dominance relation.[94] The concepts *node, precede* and *dominate* remain completely uniform in grammatical theory, and the one thing that changes is a property attributed to *representation* in phonology. It is not clear how monolithic the concept "type" is in grammatical theory. In some theories of phonology, there is a specific inventory of possible types said to be available, where those theories add lists of the type "*Type* is one of {a,b,c…}", and in some theories of syntax there are analogous inventories imputed to the theory. But, some versions of syntax (apparently) and phonology (for certain, see Odden in press) attribute no lists of types to the theory of grammar.

The concept of "rule" in the broadest sense is also a general cognitive one, but it is significantly less specified in general epistemology, compared to what it means in the theory of grammar and phonology. Note, incidentally, that "rule" is only specialized as a concept of gram-

[94] Given the calculus of the metatheory problem discussed in §4.7, I will not explore exactly how this connection is expressed. The simplest account is to reduce the matter to a biconditional relation "dominate iff different type", rather than as two unrelated propositions.

mar and not of language. Why do we conclude this? Because unlike *node, precede* and *dominate* which figure both in rules and stored representations, *rule* only pertains to the grammatical aspect of language. A rule, in general, is an "if A then B" cause-effect relation. Historically speaking, the concept of "rule" was part of the common currency of linguistic theory, refering to a "template" for mapping one representation to another. In that account, the notions of structural description and structural change are not unique to phonology. Such concepts have been largely deprecated in syntactic theory, so it is hard to say what aspects of "rule" are inherited from the theory of grammar and what aspects are intrinsic to phonology. In the interests of concreteness, I will impute to phonological theory all technical propositions about "rule", but some of this might be "linguistic" i.e. grammatical in nature.

As spelled out in §3, a "rule" is one of two kinds of representation, the other being a datum (alternatively, string). The subset of concepts of (110) taken here to be specifically phonological are repeated below (also somewhat reorganized).

(112)

rule	structural constraint	
SD	SC	focus
denial	identity	edge
insert	delete	

How are the concepts of (112) organized into a hierarchy?

First, ignore "structural constraint" for a moment, because this is a less well established concept, and we are interested in seeing how adding that concept might affect other aspects of the theory. The remaining concepts are, in general, "about" rules. These are not kinds of rules, but they are concepts relevant to the identity of the concept "rule". As already stated, a rule is composed of two sub-parts, a structural description and a structural change, both of which are themselves representations. There are a number of other propositions about rules in phonology that need to be specified, including their organization into a grammar and propositions about how rules "apply". This relationship between rules and SD, SC introduces one kind of hierarchical relation between concepts: SD and SC have no status in phonology except as properties of rules. SD and SC logically depend on, and flesh out the composition of, "rule".

The concept "focus" is an aspect of the SD, being a (required) sub-part of the SD, which relates to the SC – it says which part of the SD the action of the SC is performed on. Denial, identity and edge are concepts that pertain to the SD as well,[95] expanding the expressive range of "representation" beyond simple conjunctions of nodes and their relations. Recall that "denial" means that a particular sub-representation of a SD is specified as being obligatorily missing. How exactly this happens is within the purview of the sub-theory of "applying" a rule, a matter needing more discussion. In other words, "denial" is a sub-part of SD, one which has special properties. Likewise, "edge" is a sub-part of an SD stating a specific dominance and precedence relationship between two nodes – subordinate node β is at the right (left) edge of superordinate node α just in case nothing following (preceding) β is dominated by α; again, like "denial", an "edge" sub-part to a SD determines whether a string "satisfies" the conditions described by the SD, and crucially impinges on the sub-theory of "applying". "Identity" then states a condition on instances of nodes of a type in the SD, and refers to conditions on a rule "applying".

[95] The possibility of "identity" being involved in SC is considered later.

SC is at least in part a representation, specifically a node and/or a relation (e.g. a node with a dominance relation is inserted, or a dominance relation is inserted – a node spreads). But the rest of the change, the action of inserting or deleting, is outside the scope of what a representation is: these are actions. The problem that is that a rule apparently cannot be *just* a triggering representation and a representation listing representational primitives that result from the rule. There also has to be an indication of the action, either inserting or deleting. To recapitulate the concepts in (112), we conclude that a rule is composed of SD and SC, that denial, identity and edge are (probable) concepts regarding SD, that focus relates an element of SD to the action of the rule, SC says what things are acted on. What do we say about SC that provides a means of conveying the action which a rule performs?

Re-examining what an SC is, it is one or two representational objects (a relation and optionally a node) and an operation on those objects. The four[96] configurations countenanced are as follows, with examples of what that refers to.

(113)	Delete dominance	debuccalization
	Delete node	dissimilation
	Insert dominance	spread
	Insert node (and dominance)	default specification; prosodification

The simplest things that we can say is that SC is composed of an operation (insert, delete) and operands (the representational objects that are operated on). But how can an action be in a representation?

6.5. The engine

The main problem with the theory of grammar developed so far is that we have skirted the question of how the rules of a grammar are executed. Especially given the proposal of §6.3 that a language is a data structure composed of a grammar and a lexicon, and a grammar is a set of rules with certain properties, how do we actually map underlying forms to surface forms? To say that rules are self-executing, we would have to expand the theory of rules quite substantially. While we do need to say more, what we have to say doesn't have to be directly about the concept "rule". To draw an analogy to computers, a program (structured set of instructions) can operate on a database and produce an output, but a program requires something other than the instructions, it requires an engine that implements the instructions. A program is thus a *potential* action. This subsection provides a bit more interpretation of what the previous concepts might "mean", by making conjectures about the primary phonological concepts proposed here. A mental theory of grammar doesn't care intrinsically about a mathematical interpretation, it cares about the properties of the mind that we are talking about. However, we have pretty much reached the end of the road as far as empirically-supportable propositions about the mind are concerned. The proposal here reaches[IMO] the minimum acceptable level of support required for a scientific theory, namely possibility. The reason why the theory is possible, at the conceptual level, is that it is structurally analogous to actually-existing computational devices, namely computers.

One thing we demand from the engine, on conceptual grounds, is that it be invariant across languages. Language-specific variation is the function of the system of rules comprising

[96] A lacuna that needs to be filled is the lack of account of movement, which is attended to below in part II.

the learned grammar of a particular language. It would stretch credibility to posit that there exists a separate learned "engine" for every language that one knows – instead, one should pursue the alternative of making rules self-executing (which would largely reduce to cannibalizing The Engine and distributing its parts into the theory of rules). If "grammar" is just a data structure, where is the engine? The obvious place to assign it is as an aspect of the Language Faculty. As previously proposed, LF states what a possible grammar is, thus it is an embodied theory – a description of the architecture of language. The present proposal is that this "ability to produce language" also includes a device that *does* so – The Engine. The computer analog is that The Engine is a specific System 370 device, and the theory of rules is the instruction manual for that device. We want to know, how does the engine operate on the set of rules and the lexical database?

Just as we do not "care"[97] in computer architecture how data is moved from location to location, how bit shifts, addition and XOR operations are carried out, we are not asking about the lowest level of mental operations involved in saying that a rule applies to an input and produces a certain output. The data structure that is a grammar is a set of rules, an ordering of the rules, and perhaps a set of structural constraints. Rules are analogous to lines of code in a higher-order language like Pascal, C or Perl, and the engine is analogous to the Perl interpreter.[98] A theory of The Engine has to say what happens when a rule is executed. Structural constraints are not the same as rules, but they impinge on how a rule is executed. Ordering is a control property, describing the next state of the machine.

Just as we do not care (in the theory of computer operation) how data ends up in a memory location or how pressing a key causes some datum to come within the scope of a program, we do not care how we get underlying forms, we only care what the initial state is, and what is the result of the series of instructions operating on the initial datum. In phonology, there is (at a particular moment) some single string that is the target of computation. Computation is relative to that string, and is performed without reference to other items in storage or that might be processed by other aspects of a grammar. When a phonological computation starts (in the simplest case), the control component initially identifies "1" as the next instruction state for the engine to be in. The Engine "does what it does" (details below) then moves to the next instruction state, "2". In terms of this "fetch and execute" cycle, there are a few other scenarios worth considering, one as a possibility and one as a certainty.

The first, which is speculative, is that the engine may have the capacity to do computations in parallel, just as some computers can execute multiple instructions simultaneously (current Intel processors can, 8086 cannot). Very often, the nature of some subset of rules does not require a particular order of application, which means that the order could be strictly linear, one rule at a time, with the order determined at random, or that subset of rules could all be applied

---

[97] By "care", I mean that these are aspects of computer operation which go below the directly-observable, the execution of an instruction. The System 370 manual tells you what XOR and ADD do, and the controllable aspects of how the device knows what to do next, but it says nothing about the wiring of a control unit or the registers, or how a comparison is physically calculated. Since computers are built to order, where are no mysteries to solve: somebody actually knows, though *I* don't know. A brain scientist would of course be interested in the wiring of the control unit, but the theoretical linguist operating as theoretical linguist is only interested in what it means to say that the rules are ordered. A theoretical linguist can, at will, pursue brain science.

[98] The interpreter is an essentially invisible behind-the-scenes program which reads a program and somehow converts it into a sequence of actually-executable instructions that the computer can implement. I assume that "The Engine" for phonology is more like an interpreter that creates some cognitive thing which is less phonology-specific. It does not directly send voltages to nerves, but it does something that eventually leads to that outcome.

simultaneously. It is totally beyond the scope of phonological theory to say anything about the language faculty at this level of implementation-specificity. The mind either does operate in a maximally parallel fashion, or it does not. One of these accounts (maybe both) is wrong.

The second, which refers to something known to exist, is "the cycle". This refers to the situation where some subsequence of the rule ordering $r_i \ldots r_k$ is subject to "extraordinary" control, where instead of simply "add 1" to give the next rule, if the current rule is r_k, the next rule is (returns to) r_i. Under standard assumptions about the cycle, this is paired with a change in what constitutes "the datum", so that additional material is added to the string (newly-visible stuff contributed by the morphology). At some point, rather than infinitely cycling between r_k and r_i, namely when you come to the "outermost cyclic domain", the state which follows r_k is r_{k+1}. This is all that needs to be said about the flow-of-execution aspect of the engine for now, leaving out complicated questions about domains of application.

As for the operation of a single rule to a datum, we need to be a bit clearer on the specifics of scanning a string and applying the rule. There are four sub-topics to consider in this domain: the decoding of the operation, the comparator, iteration, and the effect of structural constraints.

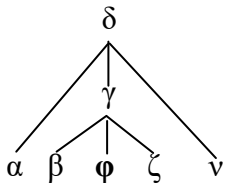
6.5.1. DECODING THE OPERATION

Continuing the computer instruction analogy, like a computer instruction a phonological rule has an operator (opcode) and operands. Where a computer opcode may be add, subtract, shift, move, bit-or and so on, the operators proposed for phonology are simply insert and delete. In a computer program, an instruction is just a representation (string of bits), so the operation XOR in 8086 architecture is x'30 which usually looks like "0", if interpreted as letter data. This helps us to resolve the previous quandry in characterizing the "action" part of a rule, which otherwise has the characteristics of a representation. The "action" aspect of a rule is the result of The Engine *executing* a representation – a rule, and a specific aspect of that representation, namely the "operation" which is what tells the machine which operation to perform. A rule therefore does not contain an actual action, which would be a conceptually incoherent theory, a rule contains information that causes an action: it *represents* an action. Similarly, the feature "voice" in substance-based theories of representation is not actual vocal fold vibration, it *represents* such substance (in substance-free theories, it heads a causal path that eventual results in such vibration).

In order to reduce "operation" to a representation, we do need to explicitly expand the theory of "representation". While "node" and the two relations defined on nodes are general aspects of language (phonology), "operation" is a expansion of the concept "representation", specifically within the domain of structural change. The structural change is composed of operation and operators specifications. The operations are delete and insert, coming in two subtypes, node and relation, and the operands are the specific type and token operated on.

We set aside iteration and structural constraints for a moment, and presuppose that we have identified a substring γ in the datum δ where γ which satisfies the conditions in the SD, and the focus within the SD (therefore a substring of γ) is φ. In other words, we want to apply the SC within γ to φ.

(114)



In the case of node spreading, the focus is the pair of nodes ν^i_j (the j^{th} node of type *i*) and ν^k_l. What is added to δ is the relation $D(\nu^i_j,\nu^k_l)$. In the case of delinking, the relation $D(\nu^i_j,\nu^k_l)$ is removed from δ; and a node can also be deleted. This raises the question raised in previous sections regarding bare insertion of a node without an association relation. Is default specification always a two-step process of inserting a floating node followed by insertion of a dominance relation? As observed above, gratuitous node insertion does not seem to exist in phonology, so there is no gratuitous insertion of a node with no consideration of some triggering node, but that could be because such a process is unlearnable. It may be that when The Engine is instructed to insert, it is supplied with two nodes as operands, and detects that one of the nodes is not present in φ, so the operation *supplies* that node, automatically including a dominance relation. One can reasonably object that this is just a technical way of avoiding two simpler rules. In the same mode of reasoning, it has been postulated in syntax that there is no movement, there is just copying followed by deletion under doom-marking.

This is an area needing focused thinking, and perhaps simplicity should rule out the idea of assigning dominance as a consequence of node insertion. However, it is also a mistake to micro-manage the micro-code. Nothing within our grasp allows us to say what the cellular composition of linguistic theory is (what kind of nerve fibers correspond to "dominance" or "cycle"). We have no factual basis for deciding between random linear orders vs. partial ordering as the underlying reality regarding the organization of rules. It does matter what the actual case is, therefore one or more of those competing theories of order must be wrong. If our primary logical tool for resolving controversies is – as advocated here – simplicity, we need a full account of the entire theory and the competitors. That level of evaluation is beyond the scope of this sub-subsection, but might sensibly be taken up in part II. At a purely conceptual level, an enriched string computation is a reality, as we can see from many computational machines that have atomic instructions which execute in a single computer operation what would otherwise be complicated programs (the string reverse instruction in 370, floating point math instructions in 8087, specialized instructions for digital signal processing).

6.5.2. THE COMPARATOR

The "comparator" is the name for the aspect of rule-execution that determines whether a particular substring of the datum matches the requirements of the SD. It is fairly simple to say what this device does. The SD of a rule is a set of node relations which are either satisfied or not satisfied. If the conditions contained in the SD are satisfied, the instruction constituting the SC is executed, and in either case, we "move along" (see the immediately following and related topic of iteration). The job of the comparator is to compare the relevant substring of the datum to see if it matches the requirements constituting the SD.

To illustrate the comparator, we will take the concrete rule Ganda Law of Logoori, selected because it has non-trivial structural requirements. In the data below, y=IPA j, j=IPA dʒ; ɲ varies between dental, palatalized dental and palatal, sometime neutralizing with the distinct se-

quence /ny/ which is IPA [nj], with alveolar [n]. Ganda Law deletes a root-initial sonorant consonant /y,r/ when it is preceded by a nasal in the onset and the following syllable begins with a nasal. Syllables in Logoori have up to three consonants in the onset, in the order nasal-consonant-glide, and by separate rule the nasal assimilates in place to the following consonant. There are no codas.[99] The trigger of this process in these examples is the 1s object prefix /N/, a nasal whose underlying place of articulation is indeterminate (it is always followed by some consonant which it assimilates to in place of articulation). There is also a hardening rule which changes /r/ into [d] after a nasal, and this rule applies to any post-nasal instance of /r/ not deleted by Ganda Law. Any vowel is lengthened before NC, and there is regressive vowel height harmony lowing /ɪ,ʊ/ to [e,o] before [e,o]. Any /Ns/ cluster simplifies by deletion of the nasal.

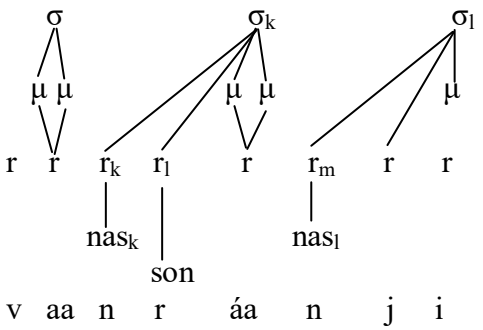
(115)	vakʊ**r**ááŋji	'they called us'	vaanááŋji	'they called me'
	vakʊ**r**ómi	'they bit us'	vaanǿmi	'they bit me'
	vakʊ**r**wáani	'they fought us'	vaanwáani	'they fought me'
	vakʊ**y**áánzi	'they loved us'	vaaŋáánzi	'they loved me'
	vakʊ**y**ááminyi	'they made us swell'	vaaŋááminyi	'they made me swell'
	vakʊ**y**áári	'they sued us'	vaan**z**áári	'they sued me'
	vako**r**ééti	'they brought us'	vaan**d**ééti	'they brought me'
	vako**r**ékanizi	'they separated us'	vaan**d**ékanizi	'they separated me'
	vako**d**ónyi	'they chopped us up'	vaan**d**ónyi	'they chopped me up'
	vakʊ**d**íginyi	'they ticked us'	vaan**d**íginyi	'they tickled me'
	vakʊ**z**úuni	'they pricked us'	vaan**z**úuni	'they pricked me'
	vako**n**óji	'they plucked us'	vaa**n**óji	'they plucked me'
	vakʊ**s**yéeni	'they stepped on us'	vaa**s**yéeni	'they stepped on me'
	vakʊ**d**yéemi	'they emulated us'	vaan**d**yéemi	'they emulated me'

The data show the following. The forms with the 1pl object prefix /kʊ/ all show that a nasal on the left is crucial to applying the rule since there is no deletion. The examples [vaanzáári, vaandééti] show that a nasal on the right is also crucial. The examples [vaandónyi, vaandíginyi, vaanzúuni] show that only /r,y/ undergo deletion, and [vaandékanizi] shows that the triggering nasal on the right must be in the immediately following syllable. The example [vaanóji] shows that would-be geminate nasals simplify to a single nasal, which opens up the possibility that the rule nasalizes /r,y/ which then degeminates. Finally, [vaansyéeni, vaandyéemi] shows that /y/ in an onset cluster not preceded by a nasal but followed by a nasal does not change.

The SD of this rule must therefore be consistent with the representation (116), irrelevant properties omitted (the reader can presumably construct alternative representations, in case you adhere to a different theory of representations). The pattern that we seek to express is an onset with a nasal followed by a sonorant, which owing to lack of w-initial roots is limited to /r,y/. Indices *k,l,m* are also assigned to certain elements in this datum, to facilitate later discussion about the representation.

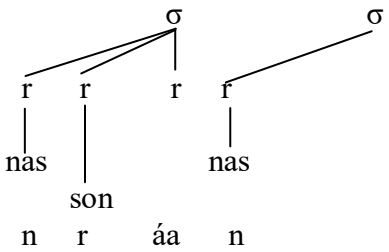
[99] In a tiny number of borrowed nouns, there may be coda consonants which are optionally made into onsets by inserting a vowel.

(116)



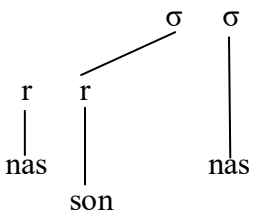
To recapitulate the essential properties of targets, they involve a certain sequence of consonants in one syllable which is immediately followed by a syllable with certain consonants (a nasal). Moraic structure is irrelevant, the preceding syllable is irrelevant, the vowel of the following syllable is irrelevant, as are any additional consonants in that syllable. We thus reduce the structure that needs to be considered to (117). Although the intervening vowel is "irrelevant" and root nodes are not clearly necessary, they are included because they have the potential (not realized) to cause problems for the rule.

(117)



Because the order of segments in the syllable is completely predictable, it does not matter for the rule what order the onset nasal and approximant stand in. It is almost sufficient to say that the syllable has these two properties, except that as we see in [vaandyéemi], the nasal must stand immediately before the consonant that is deleted (under the deletion analysis) or nasalized (under the nasalization analysis). We therefore reduce the structure in (117) to the essentials in (118).

(118)



We seek two adjacent syllables, where a nasal specification "precedes" (via root-node precedence) a sonorant specification, and the following syllable contains a nasal. The representational propositions that constitute the SD of this rule are (119). Recall that the "precede" relation is always "immediately precedes", but dominance is "dominates" and some further calculation may be necessary to conclude that A dominates B when dominance is non-immediate. Token indices are node-type relative, so there is no significance to using *i,j* for both the types *r* and σ.

(119) $D(r_i,nas)$ $D(r_j,app)$
$D(\sigma_i,r_j)$ $D(\sigma_j,nas)$
$IP(\sigma_i,\sigma_j)$ $IP(r_i,r_j)$
$IP(nas_i,nas_j)$

The precedence relations on root nodes plus their dominance conditions encode the condition that there is a sequence of adjacent nasal plus approximant. The requirement for a nasal in the following syllable is the product of the precedence relation between syllables (*following* syllable) and the requirement that the syllable dominate nasal. In fact it dominates nas_j: we should consider whether it is necessary to say that.

This is the set of propositions that must be satisfied, next we need to come up with a subset of the full representation of (116) to see if there is a match. There is a match within some substring γ if each proposition of (119) is "satisfied" in γ. In the case of the immediate precedence relations in the datum and the SD, this reduces to seeing whether the set of IP relations in (119) exist in γ. We can equate the IP propositions by stating the equivalence between the token indices of the datum and the token indices of the SD: *k=i*, *l=j*. The dominance relations are not evident from simple inspection, but they are trivially computable from the listed dominance relations (the association lines) via the recursive rule $(D(\alpha,\beta)\&D(\beta,\gamma))\supset D(\alpha,\gamma)$. This is relevant for deriving the data proposition $D(\sigma_j,nas)$ (from the dominance relations given in the datum, $D(\sigma_j,r_m)$ and $D(r_m,nas)$). We now have a simple match: the rule will apply.

The three devices of identity, edge and denial require some additional machinery, to be integrated with the comparator apparatus. A typical application of an unambiguous identity condition is (120), deletion of a vowel between homorganic consonants.

(120) mupaka → mpaka (mutaka n.c.)

r	r → Ø	r
|		|
=Place		=Place

The difference between (120) and an analogous rule without the identity indicators is that in the latter case, the simple presence of a Place node in the first and last segments in the sequence suffices to condition the rule. An identity condition is satisfied iff for every node type N^i [100] of γ corresponding to $=N^i$ in the SD, the set of nodes dominated by that node in γ is identical.

The "edge" conjecture of rule theory is that there is a special status to being "last in…" or "first in…": a node N^i is at the left edge of a word iff no node of type N^i precedes that node and is dominated by that word node. In general, we compute that node *j* of type *i* is leftmost in superordinate node *s*, i.e. it is true that $Edge(left,n^i_j,s)$, iff $(\neg\exists\, n^i_k|D(s, n^i_k)$ & $IP(n^i_k,n^i_j))$, and $Edge(right,n^i_j,s)$ is true iff $(\neg\exists\, n^i_k|D(s,n^i_k)$ & $(IP(n^i_j,n^i_k))$. Or, if $IP(n^i_j,n^i_k)$, then $\neg D(s,n^i_k)$…

"Denial" consists of bracketing a sub-expression with the SD and negating the expression. In a simple case, the expression "X ¬(Z)" is true if a substring matches "X" on the left, and does not match "Z" on the right. X is some conjunction of node relations, and X is some other conjunction of node relations. Ordinarily, a requirement that there be "X Y" is simply the conjunction of the requirement that there be "X" and that there be "Y", i.e. "(propositions constituting X) & (propositions constituting Y)" is true. When Y is denied, that means "(propositions

[100] Here, the superscript refers to the node type – identity is defined only between nodes of the same type.

constituting X) & ¬ (propositions constituting Y)" is true – therefore, at least one of the propositions making up Y is false.

The alert reader may have detected that "edge" can be handled via a specific scheme of "denial", namely "and is not followed by a sister x within y". In that case, if "denial" is a valid concept in rule theory, "edge" becomes superfluous, and is deleted from further discussion.

6.5.3. ITERATION

On the off-chance that the reader is not entirely familiar with the theory of rule iteration, I start with an overview of that concept, drawing primarily on Howard (1973) and Vago & Battistella (1982). The competing theory of rule application was that a rule applies simultaneously to a string at all points where the structural description is satisfied. If a rule changes /b/ to [β] between vowels, every instance of intervocalic *b* is identified and changed to *β* in /ba**b**a**b**amba**b**ab/ in a single rule operation. This theory faced a number of technical problems which proved unresolvable in the SPE period, and the alternative was accepted that rules apply by starting at one end of the string, working its way through the string to the other end. In a typical vowel harmony case, an input like /kukitipute/ → *kukitipote* → *kukitepote* → *kuketepote* → [koketepote], where the rule applies to successive partially-overlapping substrings *ute*, *ipo, ite* and *uke*, each substring accessing a change that is provided by the immediately preceding application of the rule. Howard discusses in detail the iterative (directional) algorithm for rule application.

There are many details in Howard's theory that pertain to the peculiarities of the SPE theories of schemata and representation, and a few questionable choices were made such as allowing simultaneous application as a "marked" option, a choice motivated by the (ultimately untenable) desire to not specify rule direction in a rule statement. A widely-debated issue in the literature is the possibility of predicting the direction in which rules apply, indeed that is the focus of Howard (1973). Benefitting from a large dose of hindsight, we can reduce the theory to the following.

(121) A rule is specified as applying from the left, or from the right
The string is scanned starting at the specified initial point, looking for a segment satisfying the internal requirements of the focus.
The string is scanned to the left and right of the focus for a match between the SD of the rule and the string.
If a match is found, the rule applies.
The scan for a focus re-commences one segment further away from the initial point, until the string is exhausted.

The algorithm became more complicated when rule expressions can contain variable-length expressions like "$(C_0V[\text{-back}]_0C_0)_0$", or expressions like "[αX] → [–αX]". It might appear to be easier to spell out the technology of iteration, since segments in that theory are fixed-length ordered n-tuples of value-attribute pair, so "moving on" appears to be trivial (look to the next segment, it is obvious what that is).

The implementation of iteration in the current theory does not require any major changes from this scheme. Rather than seeing the matter of matching as being a comparison of a segment and a matrix in a rule, the present theory operates in terms of the set of relations charactering the two relevant kinds of representation (datum or SD). The initial scan looks for that set of relations

defining the focus – not for a segment of a particular type. For example, if the rule is a leftward rule of vowel harmony and the focus is a pair of nodes in a relation which is "mora dominating back", the initial scan looks for that relation in the set, starting with the rightmost token of that relation. Notice that this involves the "can be inferred from" computation attributed in the preceding section to the comparator: that computation is general within phonological theory of "rule". Then the comparator matches the required elements of the SD centered around the focal segment, the rule applies or not, then the focal scan looks for the next (preceding or following) example of such a relation pair.

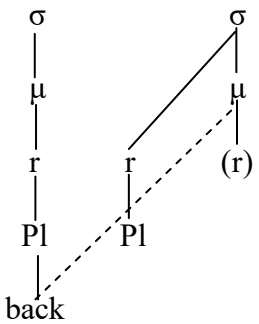
The one complication that might be introduced by the present theory is implicit in the requirements that we start with the rightmost token, and then move to the next preceding token (or the opposite for a rightward rule). But "ordering" is an epiphenomenal logical fact in this theory, not an irreducible and self-evident physical fact. The ordering of elements in a string /kukitipute/ is a graphic representation of an unordered chain of precedence relations IP(k_1,u_2), IP(u_2,k_3), IP(k_3,i_4), IP(i_4,t_5), IP(t_5,i_6), IP(i_6,p_7), IP(p_7,u_8), IP(u_8,t_9), IP(t_9,e_{10}). It would seem that we ought to address the question of how The Engine efficiently computes where the rightmost focus is (or, leftmost). But we won't do that at least now, since that is more about general computation, and we really cannot confidently say how the brain implements the task of finding the rightmost or leftmost something. We have reached the limits of linguistic reasoning.

6.5.4. Effect of structural constraints

Since it is not entirely clear what the scope of structural constraints is (because they are highly speculative), it is hard to say exactly what they correspond to in The Engine. The point of this pointedly conjectural analysis is to see how the question should be treated, provided that it *is* to be treated. To recapitulate the discussion of §3.3.2, a vague notion of "structural constraint" has been contemplated primarily to encode the dominance relations that may be secondarily inserted, when inserting a dominance relation between a higher and lower node, such as the path from a mora to a place feature, which is one way to describe vowel harmony. The theoretical motivation for this device is to simplify the theory of rules, allowing SC to be reduced to "insertion/deletion of a node/relation".

Suppose that we have a representation like (122) as the input to a vowel harmony rule spreading [back] to a mora on the right.

(122)



In this case, the following syllable has an underspecified vowel, perhaps lacking a segmental root node – perhaps the vowel arises from epenthesis, only inserting a mora. We single out vowel as opposed to consonants by referring to moras, thus a simpler rule results if the target is character-

ised just as "μ". If the rule adds an (immediate) dominance relation between [back] and μ_2, we end up with an anomalous potentially-contrastive structure – a mora immediately dominating [back] (as suggested by the dashed line). On the other hand, insertion of a series of nodes and dominance relations as the SC of the rule completely changes the nature of the theory of rules. The third alternative is that a series of independent rules must first apply to supply root and place nodes, prior to vowel harmony.

The latter solutions is the simplest to implement without any change to the existing theory, thus this stands as the obvious alternative to the changes to the theory implied by structural constraints. The outcome – an inserted series of nodes and dominance relations – is the same under both theories, the only question is whether this happens as an automatic consequence of a rule inserting a dominance relation, with the additional assistance (at the level of The Engine) of the automatization of this process. In order to compare the two theories, we have to state the competing theories. The "specific rule" account would (probably) posit rules such as (123).

(123)

That is, when a mora does not dominate a root node, one is associated to the mora; when a root node does not dominate a place node, one is provided. As specific rules of the grammar, they have to be ordered somewhere – before vowel harmony, after epenthesis. Ordering problems might arise in case the language has multiple harmonies and multiple rules of epenthesis.

Imputing this sequence of associations to The Engine has two consequences: it makes the insertion mandatory rather than being a matter of grammatical discretion, and it makes the insertion ubiquitous, being a side-effect of applying any rule and not specifically orderable to not apply before or after after a certain point. At a non-technical level, what the procedure does is compare the operands of the rule e.g. μ and [back], then consults the set of structural constraints in the grammar, to determine if μ=>[back] is included. If not, and if there is a constraint μ=>X (and no other) for some X, then a dominance relation between μ and X is inserted. If there are multiple constraints, others introducing Y, Z etc. then each of these insertions must apply virtually, and we wait a bit for resolution of the choice. Finally, if there is no constraint at all allowing something to be dominated by μ (inconceivable, but the computation does have to handle this possibility), the derivation "crashes". Then the procedure re-applies to the output of this first level of node insertion. Since μ=>r exists in the constraint set, a dominance relation and root node can be supplied. On reapplication, The Engine encounters the set of constraints r=>Lar, r=>Place, r=>nas etc. Each of these will apply as a "virtual option", but r=>Lar is ill-fated, because there is no constraint admitting [back] under Lar. That path eventually crashes; however, the path introducing Place is viable, and since Pl=>back exists, that specific set of nodes is what will be inserted.

In choosing between alternatives, we would primarily be concerned with the empirical questions of whether some choice has to be made – is there an actual problem that needs to be solved? If there is, then we again look primarily to the facts – does one theory make a false claim that the other does not? We also appeal to secondary logical questions of simplicity: which theory is simplest? If the theories are empirically comparable, simplicity says "do not add this mechanism to the theory of human language" unless the alternative of language-specificity is unconscionably complex. This is, of course, contrary to the widespread assumption of UG free

ride – if you can put it in UG rather than the learned grammar of a language, you should do so. The minimalist epistemological framework is generally opposed to the UG free ride premise. On the third hand, the computational rigamarole proposed above for implementing node fill-in is simply based on how I would write a computer program to do this, and we can speculate that the mind is slightly different from a computer program that I might write. While asserting without evidence that "the mind can do wonderous things" is an irrational response to the challenge "*how* does the mind do this?", the fact that it's not unreasonable to think that the mind has mechanisms for interpolation is enough of a reason to say that the proposal is not dead on arrival. Obviously, we have not undertaken a thorough empirical investigation of this topic, what we have done is lay bare the logic of deciding which move is warranted. In other words, we don't know.

7. Are theories comparable?

The alert reader will have noticed that I earlier invoked theory comparison as a desideratum, and have engaged in acts of comparison, but I have not presented a theory of theory comparison. Theories are not self-comparing, so exactly how do or should we compare one theory with another? I focus on two primary questions in theory comparison: when are theories logically comparable, and what are the standards for evaluating theories? The former question is harder to answer, and the main point here is to be aware of problems of incommensurability.

A prerequisite to comparison is having theories to compare. Not every assembly of words constitutes a theory. Linguists should take a hard line and only compare actual theories (a system of concepts and propositions). The string of words "A segment the context after the rule applies, delete any" does not correspond to a proposition, and it has no truth value. A theory which is based on such a string has no truth value, and is not a theory. It is not even *possibly* true. Or, there might be an expressive ambiguity in an apparent proposition, for instance "an X is Y" might mean "for all X, that X is Y", or "there exists an X which is Y". When X_i is not Y (but X_j is Y), the statement is both true and false – a contradiction. Again, we rule out such statements as not even being theories, because they *cannot* be true. You might counter-claim that it is possible to assign a truth value to the theory, if you just change word salad or an ambiguity into a proposition with a truth value. This is true but irrelevant to a comparison of theories A and B, when you change B into C: you are comparing A and C, not A and B. These problems are largely ameliorated if theoretical propositions are expressed in some formal calculus, or at least plain English that has an obvious translation to such a calculus. However, the lexical semantics of such a system is a huge problem, which we will not discuss.

7.1. Scope

It is sometimes invalid to compare apples and oranges, but not always. We will talk more about evaluation in the next section, but in order to understand the problem of incommensurability, we should reference the most elementary criterion of comparison: truth. There is usually little dispute in science that we should accept a theory which is true and reject a theory which is false. Remember that a theory is a system of concepts that states "what objectively is". We reject the theory "The Earth is flat" in favor of the theory "The Earth is a sphere" for empirical reasons which I assume the reader partially knows. The flat-Earth theory does not describe reality, only the spherical-Earth theory does. Scientists follow the Law of Excluded Middle (the Principle of Non-Contradiction) as a fundamental principle of reason, and science is all about reason. We do not purport that no person can think that the Earth is a sphere and also that the Earth is flat, we just say that that is an unreasonable position. We compare theories in order to maintain rationality – we won't adhere to a theory that is false, and we won't adhere to one theory that is false and another that is true. We only adhere to theories that are true.

The logically first thing to do in theory comparison is to determine that the theories are distinct. If two theory are identical, one cannot be true while the other is false (Principle of Non-Contradiction). I've presented a theory of phonology here which employs the concepts "precede", "node" and "dominate". You may take that theory and replace these terms with "come before", "point" and "associate". Those two theories are not different except for the specific word assigned to three concepts – the theories are notational variants. Any fact that supports or refutes one of these theories has the same effect on the other theory. Therefore, we have to first determine that we have two *distinct* theories. Two theories might be extensionally identical, meaning that they make the same predictions about nature; or they might be conceptually identical, meaning that the very conceptual systems of the theories are identical. In my (invented) example, the propositions composing the theories are exactly identical except for the names assigned to three concepts, that is, they are conceptually identical if not linguistically identical. Even when two theories make identical empirical predictions and are extremely similar in propositional content, if theory I posits proposition "(A⊃B)" while theory II positions the proposition "(A⊃C)" plus "(C⊃B)", the theories are not conceptually identical. The theories are distinct, and can be compared. The difference is not one of fact, it is a difference in logic.[101] My claim (hypothesis, hope, delusion) is that it will eventually be possible to convey the conceptual content of a theory in an objective calculus.

There are non-rational ways to block theory comparison. One can deny that there is any valid metric of comparison. One can take the subjective stance that they personally like a particular theory and they don't like a competing theory, and that ends the discussion. If a person's criterion for theory selection is emotional introspection, you will get a certain kind of (bad) result. Science rejects appeal to subjective criteria. We seek objective criteria that can be understood and followed by anyone. If we apply logic and knowledge of the facts according to the criteria, we should all agree on the status of theories. But it is still logically invalid to compare apples and oranges, at least some of the time. A theory of the physical nature of apples cannot be compared to a theory of the economic value of oranges, because the things being described are different things. We can only logically compare distinct theories of the same thing.

7.1.1. SCOPE AS A PERSONAL VALUE

I have focused on developing a theory of phonological representation and computation, because my goal is to gain knowledge of the nature of phonological representation and computation. What if I change my goal? What would be the consequence of adopting one of the following goals?

a. Gaining knowledge of the nature of phonetic implementation
b. Gaining knowledge of human cognition
c. Learning what phonological patterns exist in languages
d. Creating a descriptive grammar of Logoori
e. Developing a computer program that emulates the SPE theory of phonology
f. Developing a profitable text-to-speech program between Arabic and English
g. Securing a job in an endowed chair in Indo-European historical linguistics

[101] It would be correct for the reader to surmise that I reject that view that theory comparison can only consider factual differences between theories.

h. Gaining knowledge of how mammals evolved
i. Determining if neutrinos have mass

There is, or should be, a relationship between goals and the actions taken to reach them. If one has an entirely different goal, entirely different actions are called for.

Looking down this list, we see a reasonable but decreasing relatedness between my goal and the particular alternative goal, diminishing to the point that the theory of phonology has no realistic relationship to the neutrino-mass question. While the lowest ends of this list have nothing to do with phonology, it's not at all unreasonable to think that some knowledge of the theory of phonology could be useful as a tool in text-to-speech programming for Arabic, because of the nature of the Arabic language (see the discussion in §5.3). On the other hand, such knowledge would be of significantly less utility for an analogous project on Vietnamese, since that language apparently doesn't have phonological alternations, or at any rate has very few of them. But there is still no harm in knowing about phonology if you are a physicist, or vice versa.

It is pointless to argue about whether one *should* have the goal of acquiring knowledge of the nature of phonological representation and computation, rather than determining whether neutrinos have mass, or the goal of developing a descriptive grammar of Logoori. Personal goals are the unarguable foundation for choosing actions. The entire framework of reasoning about theory-construction is appropriate to the goal that characterizes science: gaining conceptual knowledge[102] of the nature of the universe. At the most concrete level, one may want a theory that answers the question "How do I feed and shelter myself", and answering the question "What is the grammatical structure of Logoori" may answer the survival question. Answering the question "What is phonology" may aid in answering the Logoori grammar question. Or it may satisfy a desideratum of intellectual curiosity, making life interesting. Scientifically speaking, it doesn't matter why I want to investigate the theory of grammar, it isn't even relevant that I *want* to and that I do so voluntarily and happily. Hopefully this point is not in serious doubt.

The reason why I am not certain that there is no doubt about the irrelevance of personal wishes is that meta-scientific discussions are sometimes known to consider the "utility" of a theory, as though that has some bearing on a scientific question. Utility or usefulness may be obscure social code for "correct" or "comprehensible", allowing one to indirectly say that some theory is wrong or false, or that it is incomprehensible or incoherent, without being so direct about the matter. A theory may have utility in (more or less) correctly expressing a fact in a new and better way, allowing the creation of a hybrid theory, but that does not make the theory a better theory than the competitors. Utility is a social and individual criterion ("I like it", "Our people like it"), not an objective scientific criterion where rejection implies irrationality. It is not invalid to like, use or advocate a theory for non-scientific reasons, it might even be required by law (especially if a person has a contractual obligation to produce a technical product like a software package). It is just important to not equate practical utility with scientific correctness. And it is also important to recognise that scientists can be influenced by non-scientific considerations.

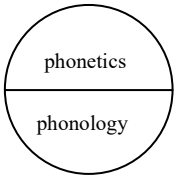
[102] That is, general principles *qua* propositions which account for the facts, not just knowing a list of facts about languages.

7.1.2. COMPARING THE INCOMPARABLE

What is the impact of the fact that there can be very many purposes behind theory-construction, in terms of theory-comparison? Do we have to agree on goals before we compare theories? That would be the wrong conclusion to draw. We *set aside* our subjective goals, at most relegating them to a gate-keeping function that says "It's not relevant to *my* interests to look at computer programming research on grammar emulation", or "It's not relevant to *my* interests to model how the larynx moves in producing pharyngeals". For the purposes of science, the only purpose is "gaining conceptual knowledge". Now we must say "of what?".

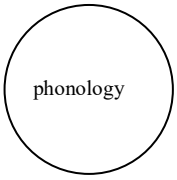
A theory of sub-atomic particles has no palpable non-scifi relevance to a theory of phonological grammar, and vice versa. A comparison of particle theory and phonological theory would be meaningless, because they are about completely different things. Linguists don't engage in such obviously fallacious comparisons, but we have a tendency to disregard subtle and not-so-subtle incomparabilities in subject matter. A classic incomparability is the SPE theory that "grammar" contains syntax and phonology where phonology performs all linguistic operations pertaining to sound,[103] versus the post-SPE view that there is a phonetic component distinct from the phonological component. We can graph this relation between theories thusly:

(124) a.



Language sound grammar

b.



Language sound grammar

Both theories (apparently) agree that there is a coherent empirical domain (language sound grammar, which is a thing claimed to exist) and they disagree on whether two separate theories are required, each addressing a specific aspect of that object, thus defining a subdomain of inquiry. Theory comparison is meaningful when one compares two theories of language sound grammar, one with two components and one with a single component, but comparison of theories called "phonology" in this case is not meaningful, since they are about different things. A comparison of a two-component theory of phonology with a one-component theory of phonology depends on resolving a prior question: is Language Sound Grammar composed of one object or two? If we don't first address this question, comparison will not be useful in answering the grand unifying scientific question "what is the nature of reality?". We come back to this point when we discuss the importance of an *integrated* theory of reality in §7.1.4.

There are analogous domain-delimitation problems regarding phonology versus phonetics – assuming that there is a phonetic module, and in phonology versus morphology. Facts of English aspiration cannot refute a phonological theory if those facts are computed in the phonetics, but they can if they are not. Is it valid to consign aspiration to the phonetics in order to refute a phonological theory, or should we have an independent principle that decides what is in phonetics vs. phonology – and what are those principles? Is it valid to claim that a phonological alterna-

[103] Or perhaps, Articulatory Phonology if it only admits a phonetic component without any phonological computations.

tion which applies in one morphologically-characterizable phonological context but not when the phonological context is the same and the morphological context is different is *therefore* not part of phonology? The consequences of such a strategy for theory evaluation should be familiar in the history of generative phonology – phonological alternations have often been consigned to the morphological component (e.g. various things associated with reduplication; umlaut in German; suffix variation in Korean and Turkish), and the reader is often left to wonder whether the strategy of data-exiling is a facile technique for theory-prophylaxis.

You should not think that I am arguing that we must each have precise theories of morphology, phonology and phonetics which uniquely characterizes how any set of imaginable facts will be disposed of, and that discussion is impossible until we accomplish that. Even developing a coherent theory of phonology is a monumental task. What is necessary is recognizing that when we arbitrarily expel a fact from phonology, we take on an obligation to find a home for the fact elsewhere. The word "arbitrarily" is crucial here, because we also should not import random facts *ad libitum* into the domain of phonology: both decisions must be principled, meaning we require principles, and as I mentioned earlier in §§1-6, theoretical principles require justification.

A conspicuous example of divergence in goals which has led to distinct domains of inquiry and has contributed to theoretical confusion is the competence / performance distinction. The standard generative position is that an individual's linguistic behavior is determined in part by linguistic competence, and other factors. The "grammar" of a language is standardly understood to be the device which embodies competence – it *generates* the language. Some larger cognitive system enables humans to think and communicate, and that system (which calls on the grammar) is what *produces* sentences and other communicative and non-communicative behaviors. Some people may have a broader personal goal of accounting for communicative behavior, and can be said to be most interested in explaining the nature of language data. Others, such as myself, focus on discovering the nature of the mechanisms that underly the observed data. As I mentioned, personal desires don't validate theoretical claims. But it can explain why some disagreements exist – names (such as "phonology", "markedness") are held constant while the domain of investigation varies.

This has consequences for theories of phonology which satisfy those competing desiderata, because there is a difference in the data domains which the approaches are responsible for. Performance-oriented theories of phonology take abstractions about language data to be the very thing that the theory of grammar is to account for. Segment coocurrence patterns are a classical example, for example insisting on grammatically encoding the fact that the sequence [wu] is never attested in a language where it is never attested. That might be a fact in some language *and* could be paired with a truly phonological fact, such as the mapping /w+uta/ → [uta] contrasted with /w+ata/ → [wata]. The very justification for positing such a thing as a "phonological grammar" is that the output of morphological concatenation cannot (usually) be directly fed into phonetic implementation. No direct argument exists that grammars must encode the possible or impossible segment sequences of a language, rather, that is a consequence of an assumption about the thing being investigated. If you change the object of investigation to "what the speaker 'knows' in some sense", then it might be the case that coocurrence patterns are necessarily part of grammar. Speech errors and loanword adaptations are sometimes thought to be within the domain of what grammar is responsible for, and thus a pattern of inserting a vowel after a final consonant may be seen as evidence that the grammar must contain a latent rule of final vowel epenthesis, since these are facts about performance data. This kind of disagreement over the

scope of phonology tends to make apparently-competing theories be actually non-competing, since they are about different things.

"Substance abuse" as identified by Hale and Reiss *is* the unjustified inclusion into the domain of investigation theoretical objects from a different domain of inquiry.[104] For example, there is a well-known physical explanation for a significant tendency that consonants after a nasal should be voiced, and this depends on facts and theories of vocal tract physiology and aerodynamics. The role of these non-phonological propositions in phonology is explaining the asymmetrical distribution of post-nasal voicing versus post-nasal devoicing as "phenomena" in the phonological data from human languages (post-nasal voicing is much more common). The broader explanation recruits a superset of conceptual principles from human anatomy and physics, whereas a strict competence-based theory excludes such principles from grammar (strict grammar has no access of the range of posible volumes of an individual's pharynx, air pressure above or below the glottis, the mass, elasticity or stiffness of the vocal folds, or the aerodynamic equations that allow one to compute whether the vocal folds will vibrate or not).[105] Since the performance-oriented perspective is primarily about language data, and since principles of physical and anatomy have a causal relation to those data, the goal (accounting for language productions) affects the theory in an unsurprising way: it is desirable to model the factors that determine production. The competence-oriented perspective has a different relation to language productions, since it sees productions as being the result of multiple factors, including grammar, physics and anatomy. To state the obvious, only those things relevant to grammar are relevant to a theory of grammar.[106]

Remember that science does not just produce theories with certain names, it produces theories *of* a specific thing. There is a thing which exists (if indeed it exists) that is the explanatory domain of the theory. Comparison of "non-competing theories" is as pointless as comparing apples and prime numbers. This does not mean that a comparison of the single-component theory of phonology and the two-component theory of phonology (as in (124)) is nonsensical – it is simply much more complicated than comparing SPE and autosegmental theories of phonology, which agree on the subject matter. The scope of comparison needs to be where there is agreement about subject matter, over the nature of "language sound grammar", where the theories are talking about "the same thing".

---

[104] The "substance" being appealed to in their view is facts of physical realization, such as acoustics, articulation or perception. It is not clear whether they would also include appeal to physical measurements of brain states: I shall do so. I will also include references to non-linguistic constructs such as "state-transition table" or "recursively enummerable set" in the class of "substance" – substance is anything outside of phonological grammar. As far as I know, nobody claims that a phonological grammar *is* an actual state-transition table, and I'm not even certain that anybody believe that there exists an "actual state-transition table". This may be a topic for later discussion.

[105] There may be some confusion over the point that collections of non-linguistic explanations could be reified into a strictly grammatical declaration. If contrary to fact there were *no* cases of post-nasal devoicing, a grammatical principle could be posited that "the structural change of a rule cannot include [–voice] if the triggering context includes [+nasal] to the left of the target". It would be arbitrary that this putative principle of grammar says "[–voice]" and not "[+voice]": and then, the supposed additional principle of grammar lacks explanatory value, it merely functions to better fit the data to the theory. We discuss data-fitting below.

[106] Generative linguists do need some understanding of peripheral domains in order to give reasonable interpretations to data, that is, to say whether a certain instance of behavior should be seen as grammatical data versus noise.

7.1.3. ARE WE EVER REALLY TALKING ABOUT THE SAME THING?

In the previous section I pointed to some big differences between theories that probably result in unsurmountable barriers to theory comparison. The difference between the "data-modeling" and "device-modeling" approaches is so profound that I don't see how to get past the "your facts, my facts" divide. But there are other, smaller theoretical divides that can make theories (or versions of theories) non-comparable, because of quirky individual differences. In this part, I consider data-related differences in assumptions that can affect the scope of theories.

"Marginal data" exist in many languages: apparently anomalous words, such as the language name Xhosa which is pronounced [||ʰosa] by a number of English speakers (most of them living in South Africa). English otherwise does not have lateral clicks. For those speakers who say [||osa] rather than [kosa] or [hosa], should we say that their grammars countenance such a click, or should we exclude the datum as marginal, despite being used? Should we exclude [x] as in *Bach, Loch, chutzpah* because it is marginal and is only found in a small number of loanwords? I don't propose to answer that question, I simply point out that this is a kind of case where a particular goal will shape the scope of the theory that one is developing, and may thereby make theories incomprehensible. If we can agree to adhere to the view that "phonology" is about a particular kind of computational and representational system, we still don't have a solid reason for stipulating that those speakers who say [||osa] do or do not have a segment [||] in their system. Assumptions about marginality can lead to adding methodological principles that effectively redefine the thing being investigated, for example if a segment is attested in only one or two words, it may not be included in the data being accounted for (meaning that it isn't generated by the grammar, it is a "performance" factor of speakers). The exclusion of low-frequency segments which only appear in words historically borrowed from some other language is a very common premise of data-marginalizing, and the same goes for low-frequency sequences (such as *sf* onsets in English), perhaps again on the premise that the sequence is "not native" with the attendant implication that only "native" structures are part of a person's competence in a language.

One type of non-marginal marginal data (in some social contexts) is code switching, whereby linguistic unit from one language can be often transported on the spot into a "matrix language" by a bilingual speaker. The phonological behavior of the other-language form therefore becomes a matter of theoretical interest. If the word follows the rules of the matrix language, can we say that this is additional support for those rules, and if it does not follow the rules, must we say that this refutes the rules? Or do we adopt auxiliary readjustments, positing that outside of the grammar, the form must first be somehow filtered to be consistent with the requirements of the matrix language. Again, ad hoc methodological decisions especially based on whether the decisions favor vs. disfavor a theory are not good practice.

I have rhetorically *presumed* that we are interested in the same object and have the same goal, but that is manifestly not so in all phonological discussions. One must bear in mind the possibility that two theories cannot be compared because they are not about the same thing.

7.1.4. INTEGRATED KNOWLEDGE

My solution to the apparent chaos of different scopes and domains of investigation is to posit that science seeks integrated knowledge of the universe. That means that science does not care if you get a simpler theory of phonology by exiling a fact and leaving it without explanation, or forcing some other area of general science to devise an explanation: there has to be conservation of ex-

planation and the whole system of knowledge is evaluated. Whether you call aspiration in English phonological, there must be an account of aspiration.

As I have emphasized in previous chapters, all claims require justification. Expelling a datum like [||osa] from the domain of English grammar requires justification, and not just a tad bit of justification but *sufficient* justification, to overcome the arguments that support retaining the word. Even including the word in the first place requires justification. For a person who pronounces the word [kosa], the matter ought to be as simple to resolve as it is to determine whether the word for "cat" is [kæt] (presuming that this is not a dialect where it is pronounced [kʲæt]). The null hypothesis regarding the question "what the word is" is, "it's however it is pronounced". We've already covered the question of data-creation and the requirement of stability (the phonological form of the word abstracts away from errors and random implementational variants). To reject the null hypothesis, there has to be a compelling principle demanding expulsion of the datum. It is irrelevant that some people say [kosa] – we are asking a question of individual psychology. It is not a justified principle of phonological analysis that phonemes must be universal within an E-language. If it were, English would have no [a, ɔ] contrast because I don't have such a contrast, in fact English would not have any vowels other than [i e a o u] because those are the vowels of Kenyan English, and Kenyan English is a dialect of English (one of two official languages and widely spoken).

Saying that it simplifies the theory of grammar to remove the OCP from the list of principles of grammar is not, by itself, sufficient. But also, just stipulating that the OCP is a principle of grammar is not, by itself, sufficient. The addition of such a principle requires justification, for example it might explain why certain phenomena exist in human languages and why other phenomena are lacking. If the OCP is to be eliminated from grammatical theory, an alternative explanation for those facts is necessary – for example, an alternative device like variables. Or, the empirical foundation must be disproven, by showing that the strong claims embodied in the OCP *qua* universal fact of human language are false. In which case, alternative explanations for the observed correlations are necessary (for instance, reducing the matter to a learning theorem, a perceptual tendency, or some articulatory challenge – all of which are outside the domain of grammar).

7.2. Comparative theory evaluation

Having pointed out the necessity for comparison to be between competing theories (distinct theories of the same fact), we move to evaluation of theories. The most obvious requirement of a theory is that it must be true, it must describe what is, and not describe what is not. Another criterion of theory selection is that we select a theory that is simple and reject a theory that is complex. A third criterion is that we should accept a theory that is enlightening, and reject a theory that is obfuscating. Clearly, we should accept a true theory that is simple and enlightening and reject a theory that is false, complex and obfuscating, but what about a theory that is true, complex and obfuscating, or false, simple and enlightening? Can one theory be more true than another, and likewise does falsehood admit of degrees?

Since truth is the central desideratum behind "knowledge", I start with degrees of truth. There are exactly two degrees of truth, "true" and "false". A theory is also either true or false, because a theory is a conjunction of propositions (and if one conjunct is false, the conjunction is false). It is also true that a complex proposition can be false because of a single false element, which can be replaced: that resembles "being mostly true". What does come in degrees is *evi-*

dence. There may be a smattering of evidence for a claim, or a fair amount of evidence, or overwhelming evidence. It is important to not be confused over quality of evidence, and what fact the evidence might diagnose. As our first evaluative principle of a theory, we simply require it to be true.

7.2.1. SIMPLICITY

There might exist two theories which are comparable w.r.t. empirical support. If truth is the only rational criterion for theory-selection, we face a quandry that we cannot rationally decide between the theories. But truth is not the only criterion, even though it is the primary one. We also evaluate theories in terms of their cognitive function. Since the purpose of a theory is to distill knowledge into a particular form which the mind can efficiently process, a theory that doesn't have the required cognitive properties should be rejected as dysfunctional.

Another possible problem would be incomprehensibility: it may be practically impossible to understand the theory, even if the propositions are indeed meaningful noncontradictory propositions. Incomprehensibility is a product of something else: unnecessary complexity. Examples of unnecessary complexity resulting in incomprehensibility can be found in the appendix on formalism of SPE. There is a simple detector for unnecessary complexity: Occam's Razor, but more precisely, a version of the SPE evaluation metric, applied to the formalized statements that constitute a theory.

The evaluation procedure of the present approach is to list the things that constitute the technical content of the theory, and count the number of objects. This assumes that we have a neutral descriptive calculus which we do not have, but even in lieu of such a tool, we can grasp the main implications of the procedure. As a first step, we can compare two theories of representation, the one proposed by the end of §4 (RSFP), and one related to SPE, which I term neo-SPE. This is based on SPE of the commonly-used version of the 60's and 70's, not the mathematical version presented in the appendix on formalism (which played no detectable role in phonological theory). SPE primarily presents a user-friendly account with segmental matrices and well-known features, and then in a brief appendix formally reinterprets representations as being an ordered sequence of values, where every 40th value marks the beginning of a new segment or boundary, and "coronal" is the user-friendly name for the 9th (49th, 89th…) instance of + or –. Since nobody ever wrote rules or did phonology that way, I ignore that approach. I only consider a subset of the theory of representation with binary features, excluding integers, *u* or *m*, and considering only segments and phonetically-defined features. The SPE interpretation that I present is my best effort to present a minimal interpretation of the theory. It is an attempt to say what SPE theory *would have* looked like, had its creators applied the evaluation metric to their theory.

The "counting rules" of evaluation are as follows. For every proposition, the terms of the proposition are counted: "A dominates B" counts as 3. Furthermore, every concept introduced by the theory counts as 1, though in the present discussion, all phonology-specific representational concepts can be introduced as part of a list of types that exist for nodes. A concept that already exists and is inherited from a higher theoretical domain is not introduced by the theory of phonology, therefore it does not count against the theory. Any modification to the properties of a concept within a phonological theory are accomplished by one or more propositions, which incur a cost to the theory. The sub-theory of representation set forth here excludes word, morpheme and syntactic boundaries.

RSFP has four representational concepts: node, type, dominate, precede. All four are inherited from the theory of grammar,[107] thus incur no cost to RSFP. The grammar of a *specific* language will have a provision for what node-types exist, what dominance relations are allowed, and limitations on precedence, but those are not details of the theory of phonology (phonological UG). This is the entire theory of phonological representation according to RSFP: the cost is zero. There is a non-zero cost in the theory of rules, which we could also consider.

SPE theory does not use the general grammatical concept "dominate", but it is not necessarily required to stipulate a proposition "canceling" the concept of domination (this is a question needing further discussion: should SPE's UG say "dominance is not available"?). Otherwise, this theory too has the freely-available concepts "node",[108] "precede" and "type". A phonology-specific concept of the theory is "segment", which is a specific type of node. While RSFP does not posit that the segment is a special kind of node, in neo-SPE it is, and theoretical principles and rules of specific grammar refer to the fact of being a segment. The plain-English proposition that "*Segment* is a type of node" can be expressed as:

(125) type(node) → *segment*

This proposition will expand to include other types.

A segment is a "collection" of some other sort of nodes, which have some properties. We posit that the sort of collection that a segment is, is like a set in that its members are unordered, and we will call it a set. We disregard the mathematical concept of "set" because the language faculty is not a subordinate of mathematics, it is a subordinate of cognition. I speculate that there does exist a higher cognitive concept of "collection" or "set" in a non-technical sense, thus non-mathematical *set* is freely available to any linguistic theory, and at least historically underlies mathematical theories of set. The essential property of set is just that it is a collection of things (nodes). This is not a concept of neo-SPE, it is a general linguistic and cognitive concept, so it does not count against neo-SPE (but imputing a specific mathematical theory of set into phonology would count against the theory, unless one can show that it is a ubiquitous aspect of human cognition).

In phonology, what is collected together into that set is another type of node, a *feature*, therefore the ontological rule above expands to

(126) type(node) → {*segment, feature*}

The relationship between nodes of the type segment and nodes of the type feature is that a segment is a set whose members are all of type *feature*, which we can abbreviate as

(127) segment = set{feature}

This does not say which or how many features make up a segment: that is a matter to be addressed below.

---

[107] I hold that these are the representational essential concepts unifying all senses of "grammar".

[108] It is not entirely clear to what extent this is a useful concept in SPE theory, but I will use it as I do in RSFP to refer to an entity in a representation; the name could be changed.

Every feature has the same structure, being a set of two kinds of things: an attribute (specific "feature" such as *coronal*), and a value. This gives us all the needed types of node.

(128) type(node) → {*segment, feature, value, attribute*}

Beyond adding "value" and "attribute" to the collection of types, we require a rule governing composition of feature: it is a set whose members are of type value and attribute.

(129) feature = set{value, attribute}

In this case, any feature has one value and one attribute: again, this characteristic also has to addressed with appropriate propositions.

It is not clear how much of the concept of "feature" is inherited from general grammatical theory. Since the concept "feature" comes from phonology, it is not surprising that the *Aspects* theory of features is not formally distinct from the neo-SPE theory, or from actual SPE (except for the specific values and attributes posited). The nature of features in syntactic theory is also unclear (see discussion in Svenonius 2019): they may be privative, binary, or multi-valued; they may have relations to each other apart from being members of the set that characterizes a node (that is, features have structure, as is true in all theories of phonological feature geometry and in HPSG; and is implicit in the syntactic notion of φ-features). Since we cannot say with confidence that value-attribute structure derives from UG, we will count this as a property of neo-SPE.

Clearly, the list of attributes assumed in neo-SPE is specific to that theory. The range of values likewise is specific to the theory, and is even broader in actual SPE theory. The following rules state the actual "terminal" values that instantiate the types "value" and "attribute". In these rules, "actual" refers to the fact of being an ontological primitive, which is not a specific construct of the theory ("actual" is a syntactic frame for saying in any thory of anything what the irreducibles are).

(130) actual(value) → {+,–}
(131) actual(attribute) → {cons, sonor, syl, voice, s.g., c.g., cont, nas, ant, cor, hi, back, lo, rd, tense…}

Neo-SPE also asserts some counting properties of values in a feature, and of features in a segment. In the case of values, each attribute is paired with exactly one value, so bare [coronal] does not exist, only [+coronal] and [–coronal] do (that is, {+,coronal}). Secondly, in neo-SPE, in any segment, $feature_x$ is present for all x features, and $feature_x$ is present exactly once. This means that a segment cannot be [+coronal,+coronal…] or [+coronal,–coronal], and also that a segment cannot be totally devoid of a specification for *coronal* (likewise any other attribute). Although I think the meaning of these three propositions is clear, it might be worth explicit formalization, since this raises some questions about usual mathematical interpretation and plain-language descriptions.

The first challenge is stating the relationship between attributes and corresponding values. In typical low-order mathematese, this means stating a relationship between any feature, and the values and attributes that it contains (which conveys "corresponding"). For every feature x, x contains a value y and an attribute z, and x does not contains w unless w is the same object as y or z. Consider a collection of integers {1,2,4,1,3,4}. The collection contains two instances of *1* and

two instances of *4* – the values of these instances are the same, but they are different instances. In writing *w=y*, we mean that the *instance* is the same, not that the value is the same (the value *will* always be the same when the instance is the same). We then write the first proposition as

(132) $\forall x \exists y \exists z$ ((feature(x)) & (attribute(y)) & (value(z)) & (y∈x) & (z∈x)) ⊃ (¬ $\exists$ w ((w∈x) & ¬ ((w=y)∨(w=z))))

An alternative exploits a special symbol $\exists !$ symbolizing "unique existence", that is, there exists exactly one, which shortens the proposition.

(133) $\forall x \exists !y \exists !z$ ((feature(x)) & (attribute(y)) & (y∈x) & (value(z)) & (z∈x))

That is, all features contain exactly one value and one attribute. This simplifies the statement, but also raises the question of what a linguistic theory is a model of. As mentioned above, there has been a widespread practice of freely adopting the devices of formal mathematics in linguistic theory, which is not problematic if the theory is a Platonic theory of the nature of the data. But generative grammar is a theory of an aspect of the mind, and a phonological theory is a model of that aspect of the mind. Therefore, we need reasonable evidence that "$\exists !$" is a general cognitive or linguistic operation. I am not contending that unique existence is an arbitrary mathematical stipulation, I am pointing out that it requires linguistic justification to be included in the theory, or compelling general cognitive justification if it is inherited from general cognition. Compare, for example, the alternative symbolization "$\exists \underline{\varrho}$" defined as "there exist exactly three", which one would be hard-pressed to consider to be a "fundamental concept of cognition or linguistics".

The condition that every "feature" is specified exactly once in a segment is really a statement about the attributes contained in a feature, since it is only at the level of attributes that "features" are distinct. An attribute is not, in one view, an element of a collection, it is an element of an element of a collection. At considerable peril, I simply stipulate that "∈" (as used in linguistic formalism) is computed recursively, so "coronal∈seg_3" is true in [ræt] (also [ræk]), and will take this to be a general linguistic truism and not a specific stipulation of neo-SPE (that is, I will not force the theory to write an even more complicated formula). Paraphrasing the intended proposition, for all segments and for all of the attribute types that exist, that attribute type is (ultimately) present in the segments, exactly once.

(134) $\forall x \forall y$ ((segment(x)) & (type(attribute(y))) ⊃ ((y∈x) & $\exists !$x))[109]

Having built up the list of representational statements constituting neo-SPE step by step, here is a restatement of the final versions, along with the counts.

(135)	type(node) → {*segment, feature, value, attribute*}	7
	segment = set{feature}	3
	feature = set{value, attribute}	4
	actual(value) → {+,–}	4
	actual(attribute) → {cons, sonor, syl, voice, s.g., c.g., cont, nas, ant, cor, hi,	42

[109] I don't really know if this is a legitimate formula, but we can also write a function "unique in".

back, lo, rd, tense…}

$\forall x \exists !y \exists !z ((\text{feature}(x)) \& (\text{attribute}(y)) \& (y \in x) \& (\text{value}(z)) \& (z \in x))$ 22

$\forall x \forall y ((\text{segment}(x)) \& (\text{type}(\text{attribute}(y))) \supset ((y \in x) \& \exists !x))$ 17

Every symbol other than brackets and commas "counts" ($\exists !$ is treated as one symbol though it uses two letters: each named concept like "segment", "cons(onantal)" is one symbol). The list of attributes assumed to be in UG by neo-SPE is about 40, which is about the number of features posited in actual SPE.[110] RSFP added no propositions or concepts that are specific to phonological representations, neo-SPE added 99. This is the concrete sense in which RSFP is a simpler theory of phonological representations than neo-SPE. Of course, complexity considerations are secondary to the empirical adequacy of the theory, therefore if RSFP fails on some account where neo-SPE succeeds, we handle the theory-evaluation question earlier in the logical cycle.

The main methodological difference between neo-SPE and RSFP is that the former imputes to UG as much as possible, in order to simplify individual grammars. RSFP pre-specifies much less as part of the theory, and correspondingly leaves as much as possible to learned rules of specific languages. However, it also does not purport that it is learned for each language whether features are binary or privative, for instance. It claims that the necessary set of attribute types is learned, and that dependency rules relating to dominance relations are learned ("Place dominates coronal"). Needless to say, the competing goals of simplifying UG versus designing a version of UG that seems to yield "simpler language specific rules" have to be adjudicated first. Chomsky & Halle (1965) argue, correctly, that it is invalid to compare the simplicity (symbol-count) of individual linguistic descriptions across metatheories, thus it is irrelevant that RSFP might require more rules or more learning compared to neo-SPE (although on other grounds, it seems that individual neo-SPE grammars turn out to be longer that RSFP grammars). They also comment that there is another kind of simplicity, "in which one talks of the simplicity, elegance, etc., of theories (relativistic physics is more elegant – simpler – than Newtonian physics, Copernican astronomy is simpler than Ptolmaic, etc.)". They acknowledge the "interest" of this notion but have nothing to say about it. The present approach takes up the centrality of theoretical minimalism, since the field has now progressed to the point that we can talk about competing versions of generative grammar, and can realistically expect substantial simplifications both of UG and of individual grammars.

I leave as an (extensive) exercise for the reader the task of undertaking a similar comparison of the simplicity of neo-SPE rule theory versus the minimalist rule theory associated with RSFP. The phonology-specific content of rule theory under RSFP is not completely null, but it is still rather simple. There are rules mapping from string to string. The basic organization of a phonology is essentially the same in the two theories (a set of ordered rules that compute the input-output relation, rules iterate through the string in some manner as discussed in the literature of the 70's, there is a domain of rule-reapplication involving "the cycle"). When only considering neo-SPE "simple rules", those is that do not involve any abbreviatory conventions, the theories do not make substantually different claims, except that since neo-SPE has no notion of dominance between nodes (features) and since an expression "[(+)coronal] → [(+)voice]" is meaningless in RSFP, a certain amount of work will be necessary to restate what a rule does in neo-SPE. The main difference in the rule theories is that neo-SPE will at least require devices to express

[110] SPE does not give a definitive list, and completely abdicates on tone features, so one can only estimate.

value variables (αcoronal), unbounded segment variables ([-voice]*) and probably finite optional strings (C(V̆C)…).

7.2.2. ACCEPTANCE

Suppose we have an established theory for which we have a lot of evidence. We would probably adopt the theory. Then suppose a new theory is proposed, and we are presented with a fair amount of evidence in support of it.[111] The new theory seems to be *almost* as good as the old theory. Would we not then accept the new theory? Based on the history of phonology, it is hard to deny that there seems to be an additional selection criterion at work in linguistics, one that favor new theories which seem to be promising, even if they are not yet proven to be empirically equal or better. How can it be rational to reject one theory that is less-proven? Is "being new" an intrinsic advantage for a theory. It would be plainly irrational to deem a newer, less-well supported theory to be inferior as a theory just because it is older. Neither the truth nor the simplicity of a theory is affected by its age.

It is possible that an individual phonologist never accepts a novel theory until it is shown to be empirically almost as good, *and* it is shown to be simpler. If one were to adhere to the principle that almost-proven and simpler theories are to be accepted, one might be taking a rational gamble that the novel theory *will* maintain its empirical success, so acceptance could be based on an estimate of future truth evaluation and a calculation of simplicity, the two logical criteria that this work holds out as being correct for theory evaluation. However, such a pattern doesn't accurately describe the broad social history of Optimality Theory or Autosegmental Phonology

There is more to theories than their logical structure.[112] There is also *acceptance*, which is a relation between a theory and the belief systems of scientists in a field. I don't intend to delve deeply into the psychology of theory acceptance, I am simply compelled to point out that concluding that a theory seems "promising" or "enlightening" does not change the logical status of a theory. Because one *should* be confused over the status of such peripheral considerations, I will spend a little time on theory-aesthetics. It is very difficult to determine what additional factors, other than truth and simplicity, actually influence theory-acceptance in phonology. I believe that "predictive power" and "constrainedness" are two (related) criteria which strongly influence acceptance.

Every theory makes predictions. A theory's propositions allow for the existence of certain facts, and preclude others. What does it mean for a theory to "predict" something? A "prediction" is in part a relationship between phenomena and how the phenomena are treated by the theory. For example, various theories of gravity predict that if you release a brick when it is 4 feet above the ground, the brick will fall to the ground. However, no theory of physics talks about bricks, 4 feet, or releasing things, rather, there are a number of abstract equations that constitute the theory, and practicioners in physics have devised an art for relating those equations to experimental setups and the interpretation of those experiments. In physics, a prediction will be a

---

[111] I avoid invented numbers like "76.6% probability", because I think that is an incorrect view of evidentiary support. Instead, I use ordinary words that evoke comparative strength, for example "a lot of evidence" is stronger than "a fair amount". This is because in our present epistemological state, we cannot objectively assign evidence to a continuum, and when evidence bearing on competing propositions is similar in strength, we may not be able to distinguish which of two is "better supported" (and why).

[112] This is an observation, not an endoresement or condemnation of that fact.

relationship between the laws constituting a theory and a verification procedure. A prediction may be negative, stating what cannot happen (e.g. various conservation laws in physics), even when the specific event is not deterministically computable (various decay modes for a nucleus). At the most elementary level, if a theory allows an account of the phenomena that arise in the domain of the theory or those that we believe will arise in the future, and can distinguish between the possible and the impossible, then we can say that it makes correct predictions.[113] For instance, Newton's general theory of gravity allows one to calculate the future position of Mars at an arbitrary time in the future (given knowledge of the sun and other planets, or other big gravitational sources), and that prediction has been verified innumerable times over the centuries. The position of Uranus can likewise be calculated. When it turned out that the prediction was false, there were two choices, either the theory is wrong,[114] or an auxiliary hypothesis is posited (a planet now known as "Neptune"). Because of the precision of Newtonian mechanics, it was possible to predict where said planet must exist, and Neptune was then observed in that location – the prediction was verified. Analogously, anomalies in the orbit of the planet Mercury led to the prediction of another planet "Vulcan" inside the orbit of Mercury. The predictions were only "transiently verified", that is, numerous observations were made that seemed to count as another planet, but the observations themselves could not be confirmed in over a half-century of searching.[115] Failure to observationally match prediction may be excused when the conditions of observation are more ambiguous.

The content of a specific grammar is highly variable and grammatical theory cannot deterministically predict what system of rules and representations exist in a language (apart from stating a deterministic relationship between "the data" of a language and the grammar that generates it), so we must be content with conditional relations of the type "if there is such-and-such pattern of data, the analysis must be X". This is generally how phonologists relate observation to theory. We organize language phenomena into sets of similar patterns, for example hundreds of Bantu languages have a vowel harmony rule where /i/ becomes [e] when the vowel of the preceding syllable is [e,o]. Every theory allows this to happen, and it is not a significant discovery that it exists in Luganda when it is known to exist in Kerewe, Zinza, Shona and so on. Eventually we may abstract a general pattern of facts saying that languages can have a phenomenon of "high-to-mid lowering harmony", where we predict the possibility of hi-to-mid lowering processes, abstracting away from direction or domain of application, as well as possible limitations on the members of the target class. From a diverse range of concrete instances, we can drop non-essential differences in the specific rules of languages and we are left with a core of high-to-mid lowering facts. Again, every theory of rules and representations (apparently) can handle this fact – it is "predicted" by theories, specifically, predicted to be possible. Two theories that predict these data are the SPE model and the autosegmental model. In the SPE model, this follows from the fact that the language-specific rule has a common formal core [+hi] → [–hi] / [–hi], and in

---

[113] The reader is advised that by talking about "what happens", I am not endorsing a data-fitting theory of science that blurs the distinction between competence and performance. What "happens" in grammar is the system of rules and representations, not the behavior that the rules and representations contribute to.

[114] This happened hundreds of years ago, and I have no idea if rejecting Newtonian mechanics was ever contemplated.

[115] This failure of prediction did not overthrow Newtonian mechanics, instead, the observational anomaly was explained once Newtonian mechanics was replaced with the theory of relativity, which allowed for a special gravitational effect of close proximity to a really enormous body.

autosegmental phonology the core is spreading of [–hi] to a segment that is [+hi]. More generally, any SPE rule of the form [αX] → [–αX] / [–αX] has an autosegmental equivalent which spreads [–αX] to a segment that is [αX]. Predictive power does not distinguish the approaches.

The feature-geometric notion of constituency changes the predictions of autosegmental theory, in predicting certain additional phenomena while precluding others. Constituency (along with universality assumptions as to the structure of representations) gives rise to the possibility of general "height harmony", which generalizes over different sub-classes of vowel harmony. The constituency hypothesis (in one version) says that [hi,low,ATR] form a representational constituent (VHeight), therefore a rule can individually spread [hi], [low] or [ATR], or the entire set [hi,low,ATR] as a group. These predictions are (more or less)[116] verified.

It is not that SPE theory was incapable of also stating these facts, since →[αhi,βlow,γATR]… is a possible rule in SPE theory. The predictive power of the constituency argument stems from how it partitions the universe into "possible" vs "impossible" patterns. One prediction (exploited in early autosegmental phonology) was that there could be autosegmental rules corresponding to those of the SPE-type →[αX] / [αX], but not those of the type →[αX] / [αY] (autosegmental phonology eliminates feature coefficients, and handles the phenomenon of assimilation as feature spreading). Autosegmental phonology partitions the imaginable world of phonological phenomena into the possible (spreading a feature) versus the impossible (one feature taking on the value born by a different feature), and the partitioning is different from that of SPE theory. Insofar as the latter class of rules is unattested, a prediction of the autosegmental theory is validated, and the autosegmental theory leads to new knowledge about language. Since theories are a tool for creating and retaining knowledge, acceptance of the theory is an appropriate response. Universal feature constitency makes a further prediction, stating what are possible multi-feature assimilations. It allows any individual feature spreading, and also spreading of any complete set such as {hi,low,ATR}, but not groups that do not constitute a defined set, such as {hi,round}. Again, compared to SPE theory, feature geometry predicts a fact pattern which is not predicted by SPE theory.

Generalizing these observations about prediction from a specific area of phonology, we can say that the "predictive power" of a theory is high when the partitioning of phenomena into "possible" versus "impossible" closely matches observation of attested phenomena. But furthermore, as an aspect of theory comparison, this partitioning should constitute substantial, perhaps surprising new knowledge. It's not just that we don't observe constituency of the type {hi,round}, it's that it is (somewhat) surprising, in the context of the history of feature geometry, that we don't observe that pattern, or myriad other kinds of multiple-feature spreadings.

The related concept of "constrainedness" also played a major role in shaping the development of linguistic theory, but is more problematic. Early generative grammar was "very powerful" in the sense that you could do just about anything with it. In syntax, there was a proof in Peters & Ritchie (1973) that existing syntactic theory had the ability to generate the same class of languages as a Turing machine, which is the "most powerful" computational device in existence – it can compute anything than can be called computable (though some things cannot be computed). In syntax the concern was with characterizing language using devices that are lower on the Chomsky hierarchy, an abstract mathematical characterization of what different kinds of computations can do. The specific concern was about sets of strings, which is how "language"

[116] There is a conspicuous lack of evidence for spreading of [low], which may then motivate a search for a better theory of vocalic features.

was viewed, and a theory of machine-type that could produce those sets. These concerns filtered into phonology, and contributed to a feeling that there should be limits on what phonological theory can do. We viscerally feel that we should not allow rules to refer to whether or not the index of an element in a string is a prime number, because we have never seen such a thing in phonology (or syntax). Yet the quasi-formal approach of earliest generative phonology allowed statements such as "where *j* is not prime". Chomsky-hierarchy concerns generally had little influence in phonology,[117] but there still was concern was over the non-existence of imaginable yet (apparently) unattested data patterns which might motivate rules like ŋ→p/n__, p→s/__r or $\{b,t\}_i \rightarrow \{m,ʒ\}_i / \_\_ \{a,f\}$. The approach is well summarized in McCawley (1973), that "One who takes 'excessive power' arguments seriously has as his goal characterizing 'phonological rule' so as to include all and only the phonological rules that the phenomena of a natural language could demand".

This concern with matching theory to observed data patterns gave rise to various added principles of grammar which prevented otherwise describable phenomena from being produced by a grammar on the grounds that they had never been seen, for example the Crossover Constraint, the Relevancy Condition, the Revised Alternation Condition, the Obligatory Contour Principle, the Well-Formedness Conditions, the Twin Sister Convention, the Sonority Sequencing Principle, the Iambic-Trochaic Law. To some extent, such additions to theory were exploited to modify the statement of language-specific rules, so that rules of syllable-structure construction in a language might be simplified because the rule itself would simply be "syllabify everything", and the specifics of how that happens is then left to Sonority Sequencing from some set of "parameters". It has thus been considered desirable for a theoretical modification to impose novel limits on the range of allowed phenomena.

A very simple but unconstrained theory of phonological computation is "anything can change to anything anywhere". But this is not a theory of computation, it is a meta-theory that attempts to summarize an actual theory of computation. If any language had a computation "anything can change to anything anywhere", then every input would nondeterministically map to every linguistically-possible string, the underlying form /dɔg/ would probably never actually map to [dɔg], and nobody would understand or learn this language. Instead, the supposedly unconstrained theory would require individual rules that do specific things, just as are posited in SPE, and specific rules could be written to do strange things, such as round a vowel which has the same value of nasal as it has for high and the opposite value as it has for back, in every syllable that is a multiple of 7, counting from the left of the utterance. Whether this is a formally possible rule depends on the theory of rules under consideration. It is not possible in the theory advanced here, but it is stateable in SPE theory:

$$(136)\quad \begin{bmatrix} +\text{syl} \\ \alpha\text{nasal} \\ \alpha\text{high} \\ -\alpha\text{back} \end{bmatrix} \rightarrow [+\text{round}] / \#\#\#(([-\text{syl}]_0 \#_0 [+\text{syl}] \#_0))^6_6)([-\text{syl}]_0 \#_0 [+\text{syl}] \#_0))^7_7) * [-\text{syl}]_0 \#_0 \_\_$$

[117] However, the use of transformational technology, i.e. unrestricted rewrite rules, to simultaneously modify multiple parts of a representation was a concern, and eliminating such devices by exploiting other means to get metathesis and fusion was part of the motivation for autosegmental phonology.

The present approach to theory construction eschews independent complications of grammar whose purpose is to state that certain things are "not possible". This is because the concern of theory is not with stating what phenomena will be observed, it is with stating the causal principles that underly phenomena. The use of algebraic variables, value-polarity (indeed, values) and numeric repeat-operators are prime examples of overgeneration that can be blamed on accepting insufficiently-justified claims in the theory. Historically speaking, it seemed at the time that these computational concepts were necessary, but in the light of alternative concepts of representation and rule statement (primarily via the exploitation of "dominance" and the lack of precedence relation between elements on separate tiers), we see that there are superior alternatives. Setting aside the fact that autosegmental representations allow in a formally simpler theory of rules, the fundamental representational concepts exploited by autosegmental theory is "more true", by the Aristotelian correspondence theory of truth – "To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, and of what is not that it is not, is true". That is, the account not only predicts that which is observed, it says correctly that which will not be observed. To the extent that "constrainedness" is a restatement of "that which is true" for causal principles, this is clearly a valid criterion for theory comparison. To the extent that "constainedness" is appealed to in order to shrink the size of the dataset potentially covered by the theory and to give the impression that we already know everything there is to know about the domain, it is not a valid criterion for theory comparison.

8. Recapitulation

To sum up the conclusions of this first part, what have we learned? First, as I (and myriad others before me, usually not linguists) have repeatedly argued, scientific theories require an epistemological framework, in order for sound theories to be constructed. The essential epistemological principles advocated here are that knowledge results from the evaluation and integration of sense data. That knowledge is organized hierarchically so that there is a meaning to the idea of "building up" a theory, starting at the bottom. The cycle of constructing a theoretical system proceeds by first building a system of pre-theoretical concepts and propositions that state what the observable facts are. Theory is built on this foundation, by adding causal principles that explain the nature of the universe, as we have observed it.

A claim to having knowledge requires justification, which relates a conclusion to the undeniable – an actual observation, and not just a belief about what an observation indicates. Concepts and propositions are evaluated in the face of alternatives, therefore one must *consider* reasonable alternatives, from which it follows that one must *seek* alternatives. The primary metric for evaluating claims is reality, meaning that a theory must correctly describe what is, and not mistakenly say that that which is not, is. In case we can reasonably consider two ideas as competent to describe reality, we appeal to simplicity – the theory which says fewer things is correct. One point which I have *not* repeatedly hammered away at is that it can be reasonable to say "I don't know if it is A or B that is true". When massive evidence supports A and no evidence supports B, it is unreasonable to maintain that B is possible – that is just nihilism. When 55% of the evidence supports A and 45% supports B, it is reasonable to consider B a viable theory – we just don't know, yet. It is important to be able to both say "I know" and "I don't know".

EXERCISES FOR THE READER

I think enough of the framework as been set forth that the reader can answer some un-asked questions.

1. Rules may be optional, e.g. /ata/ → [ata,ada] via a rule of intervocalic voicing. What do we have to add to the theory to get this? Deliberate ignore question 2.

2. In the case of an optional rule that can apply to *n* points in a string where applicability of the rule to node *j* is not affected by application to node *i*, 2^n surface forms are generated. How does this affect the answer to Q1?

3. Rules are ordered. What exactly does that mean, in terms of the theory?

4. Autosegmental phonology has maintained that association lines do not cross. What thing(s) in the theory does (would) that claim correspond to?

5. A number of languages have processes of "merger", such as /ai,au/ → [e,o]; /kj,gj/ → [tʃ,dʒ]; /kw/ → [p]. Assume that this is accomplished by a single rule: what change in the theory is necessary to express merger rules?

6. Find two actually-formalized real rules[118] in the literature that can't be stated in the present theory: explain why they can't be stated.

7. Find one fact pattern that cannot be handled by this framework; what do you need to add to the theory to handle the pattern?

PART II

TBW

[118] That is, exclude fake rules which are constructed for the purposes of saying that X is not a possible rule.