The OCP and NoBanana

OR

Philosophical and Empirical Reasons to Ban Constraints from Linguistic Theory*†

Charles Reiss
Concordia University, Montréal
reiss@alcor.concordia.ca

February 14, 2002

That which is wanting cannot be numbered.
Ecclesiastes 1:15

1 Introduction

This paper has two goals. One goal is to argue that constraints are inappropriate computational devices for modeling grammar. The second goal is to make concrete proposals concerning the nature of phonological computation and representation. While I hope to have integrated these two aspects of the paper, they are logically independent. Accepting or rejecting the philosophical arguments against constraints is not necessarily tied to one's views of the validity of the discussion of identity reference in phonological processes later in the paper. The discussion of constraints, at the very least, provides a springboard into the subject of identity reference, and thus we start there.

Let's begin with a preposterous example. Suppose we are seeking a constrained theory of *UG* for syntax and we are trying to choose between a theory with the components in (1a)

*This document is available as Carleton University Cognitive Science Technical Report 2002-03. URL http://www.carleton.ca/iis/TechReports.

†Thanks to audiences at McGill, UQAM, MIT, Utrecht and Michigan and to Andrea Gormley, David Odden, Jean-Philippe Marcotte, Daniela Isac, Geoffrey Pullum, Marshall Wong, Brendan Gillon, Patrick Davidson, Bill Idsardi, Madelyn Kissock, Jonathan Bobaljik, Glyne Piggott, Eric Raimy, Ida Toivonen, Ash Asudeh, Morris Halle and Bert Vaux for useful comments and criticism, not all of which have been addressed. Some of these people remain outraged by the contents of this paper. Many of these ideas were developed in conversations with Mark Hale. It should be clear that many of my proposals draw on various ideas in the literature. I do not pretend to have approached exhaustive acknowledgement of these sources.

and another with the components in (1b)[1]:

(1) Which model of *UG* is better?

a. *Merge*
Lexicon

b. *Merge*
Lexicon
NoBanana

Model (a) contains the rule *Merge* which operates on elements of the *Lexicon*. Model (b) contains these components as well as the additional constraint NoBanana which marks as ungrammatical any representation of a sentence containing a banana—an actual banana, not the lexical item *banana*. Is it useful to claim that (b) is a more constrained model than (a) is, since (a) has no way of ruling out sentences that contain bananas? Obviously it is not useful or necessary to do this—(a) does not generate sentences that contain bananas since bananas are not contained in the set of items (the *Lexicon*) over which *Merge* operates. The more constrained model is thus (a), since it is characterized by a subset of the elements needed to characterize (b), and the two models have the same extension.

Consider another preposterous example in (2).

(2) Which model of *UG* is better?

a. *Merge*
Lexicon

b. *Merge*
Lexicon
NoLexicalItemsSentToCleveland

In (1) we considered the effect of enriching a model of grammar by adding a constraint referring to entities not found in the set over which *Merge* applies. In (2b), we have added a constraint referring to an operation that is not present in the model of the grammar in (2a). Again, it should be clear that since *Merge* does not have the effect of sending lexical items to Cleveland, and since the grammars characterized in (2) contain no other operations, it is not necessary to rule out representations in which lexical items have been sent to Cleveland.

What makes the preceding examples preposterous is that constraints are supposed to be formulated in terms of a (typically implicit) *universe of discourse*. Note that the claim intended by the constraint NoBanana, that no representation of a sentence contains bananas, is probably true for all human languages. However, there are an infinite number of true claims of this type. No language requires speakers to dance a jig to express iterativity; no language has pizza as an element of syntactic trees; *etc.* Bananas, pizza, dancing of jigs, sending and Cleveland are not elements of grammatical models. In other words, we do not want our model of grammar to express every true statement about what structures do not occur, since there are an infinite number of such statements and the grammar must be statable in finite terms if it is to be instantiated in human brains.

The conclusion suggested by the preceding discussion is that the search for *UG* should be conceived of as the attempt to characterize the universe of discourse, the entities and

[1] I am obviously making simplifying assumptions here. The point is just that one model has a set of entities and the second has all those plus an additional constraint.

operations that constitute the representations computed by the language faculty. *UG* is thus to be characterized by a list of categories and rules that take these categories as arguments—and nothing else.

A coherent conception of the 'perfection' of the language faculty, one that does not cave into the temptation of functionalism, is that the formal system that defines *UG*, as well as every particular grammar, is exhaustively definable: there is a finite list of categories and rules that uniquely determines all and only possible linguistic structures.[2] Again, *UG* should not be conceived of as a set of constraints defining directly what is *not* a possible human language, because this set has an infinite number of elements. The notion of what is not a possible language will follow from an appropriate characterization of the properties of possible languages, but this notion need not be independently formulated in the grammar.

We can further motivate the elimination of constraints by considering how positing of constraints forces us into an overly rich view of *UG*. Linguists in general adopt the assumption that *UG* contains constraints, *qua* prohibitions on possible structures. However, such prohibitions *cannot* be learned by positive evidence (an infinite number of well-formed structures are absent from the PLD—we may find a supposed ill-formed structure in the next sentence we encounter). Therefore, these prohibitions could only be learned *via* negative evidence. However, it is generally accepted that negative evidence is neither supplied to the child with sufficient regularity, nor attended to by the child when supplied, to play a significant role in language learning. Therefore, since the prohibitions cannot be learned *via* positive evidence (for reasons of logic), nor through negative evidence (according to the emprical data), they must be innate.

This conclusion appears to follow from the premises, but I believe it is false. The fault lies with the assumption that *UG* consists of constraints. In this paper, I justify rejection of this premise; in another paper (Reiss, in prep) I demonstrate how the need for constraints can be circumvented in learnability theory.

2 Overview

This paper not only develops this argument concerning what *UG* should not be, but also makes concrete suggestions concerning how the study of *UG* should be approached. In section 3, I define constraints in opposition to rules, then I return to the issues raised in the Introduction in order to point out two slightly different ways in which inviolable constraints have been used. I then turn to a discussion of violable constraints, as used in Optimality Theory (Prince & Smolensky 1993). I conclude on philosophical grounds that linguistic theory should be rule-based rather than constraint-based: grammars contain rules (as defined below), not constraints (as defined below).

In section 4, I briefly show that the ideas presented here converge with some recent work in syntax. I then discuss, in section 5, the use of constraints in conjunction with rule-based phonology, concentrating on the Obligatory Contour Principle (OCP) for illustration.

[2]In other words, the definition of UG, and of particular grammars, can be understood as including a final, exclusion clause of the type used in recursive definitions in logic. I address below the problem of overgeneration—the fact that the set of possible linguistics structures is a superset of attested structures.

Following Odden (1988) I argue that there is no good theoretical or empirical motivation for positing the OCP. The argument extends readily to other constraints that have been posited in the literature.

In addition to this somewhat negative conclusion, I show in sections 6 through 12 that an investigation of the formalization of the types of data for which the OCP was invoked leads to interesting discoveries concerning the nature of phonological formalism. Specifically, Autosegmental Representation and Feature Geometry are found to be insufficient for the representation of certain well known processes. An alternative formalism, Feature Algebra, in combination with quantificational statements, is introduced in section 9. This formalism not only allows us to state the range of conditions on rules that are relevant (as examples of or counterexamples) to the OCP, but it also allows us to discover, in section 10, two types of rule condition which apparently are unattested. I sketch a tentative explanation for why these pattterns are unattested. Section 11 briefly discusses Rose's (2000) treatment OCP effects in OT. Section 12 suggests that, at a certain level of abstraction, the computations that are unattested in phonology are used by the syntax. Section 13 compares rule- and constraint-based approaches to phonology. I argue for a revival of rule-based phonology, but not a return to the mixing of rules and constraints, and I offer a contribution to the understanding of *formal* aspects of Universal Grammar. The results presented here demonstrate that progress in our understanding of UG does not depend upon the characterization of substantive tendencies subsumed under the notion of markedness. Some conclusions and open questions are discussed in section 14.

3 On Constraints

This section discusses in general terms various uses of the notion of constraint in linguistic theory. First I discuss constraints on grammars, that is, constraints on what is a possible language. Then I turn to inviolable constraints within grammars. Next, I discuss violable constraints as the basis of grammatical computation, as in Optimality Theory. I argue that each of these approaches to defining UG suffers from a combination of a lack of elegance and a mistreatment of the problem of inductive uncertainty.

3.1 What is a rule? What is a constraint?

Mohanan 2000:146 argues that, due to basic logical equivalences, the constraint/rule distinction is incoherent once we adopt the view that both rules and constraints express propositions. However, in the following definitions I distinguish rules and constraints in terms of their role in a computational system (a grammar) as a whole.

I will now provide a description of what I mean by a linguistic rule and what I mean by a linguistic constraint. Various formal statements found in the literature may go by names that are at odds with the descriptions given here. This purely terminological issue does not bear on the validity of the dichotomy proposed. So, for example, we may find formal statements that are called 'constraints' in the context of a given theoretical framework, but which are in fact examples of what is here called a 'rule'.

A RULE R can be viewed as a function that maps an input representation I defined in terms of a set of representational primitives (features and relations) to an output representation O which is defined in terms of the same set of primitives. The application of a rule depends upon a potential input representation matching the structural description of the rule. This representational matching procedure (RMP) outputs two possible results: YES, I satisfies the structural description of R; or NO, I does not satisfy the structural description of R. If the output of the RMP is YES, R applies and relevant parts of I are rewritten as O. If the output of the RMP is NO, I is not affected.

In a constraint-based theory, constraints also contain RMPs that serve to map an input I to one of the two possible results YES or NO, as above. However, for each constraint, one of the two values, YES or NO, maps to a further evaluation called VIOLATION and the other to NOVIOLATION. The use to which this evaluation is put rests with another part of the computational system. In (all?) constraint-based linguistic theories a crucial aspect of constraint evaluation leading to the output value VIOLATION is the notion of illformedness. Depending on the formulation of a given constraint, either matching or failing to match the structural description of the constraint signals illformedness. For example, a constraint formulated as 'Don't have a coda' leads to an evaluation of illformedness for a syllable which *has* a coda. A constraint formulated as 'Have an onset' leads to an evaluation of illformedness for a syllable which *does not have* an onset. The distinction between such negatively stated and positively stated constraints will not be relevant to the remainder of this paper. A further aspect of constraint-based theories that should be noted is that violation of a constraint (that is, incurring the evaluation which is assumed to denote illformedness) is passed on to other parts of the computational system. In theories incorporating inviolable constraints, constraint violation prevents a representation from being evaluated as grammatical. In Optimality Theory (Prince & Smolensky 1993) the violations are used by EVAL, the evaluation procedure which interprets violation with respect to the relative ranking of the constraints.

To reiterate: a rule is defined as a function from representations to representations; a constraint is defined as a function from representations to the set {VIOLATION, NOVIOLATION}.

3.2 Constraints on Grammars

It is a commonplace in the linguistic literature to find statements suggesting that a goal of linguistic research is to define UG by formulating the constraints on what is a possible language. This enterprise is typically seen as integral to explaining the paradox of language acquisition, in the following way. If the child is endowed with innate knowledge of the constraints delimiting the set of humanly attainable languages, then the child's hypothesis space is limited. Instead of choosing from the infinite set of (not even necessarily attainable) grammars, the learner need only select from a predetermined subset of those. Of course, we might make this idea more palatable to some by referring to constraints on the learner's ability to make hypotheses, rather than to knowledge of these constraints, but this is just a matter of terminology. I wish to argue that a characterization of UG in terms of such constraints can be at best merely a derivative notion.

It is necessary to stress that I am concerned in this subsection with constraints *on* gram-

mars, not constraints *in* grammars. I am not concerned, for the moment, with evaluating the merits of constraint-based computational systems such as Optimality Theory (Prince & Smolensky, 1993) *vis-à-vis* rule-based grammars, for example, although I turn to this topic below.

Instead of the preposterous examples in (1) and (2) above, consider the question of hierarchical structure in syntax. Let's imagine that we want to express the claim that all structure is hierachically organized as a trait of UG. How should this proposal be formulated? If one seeks to characterize UG by listing constraints on the set of possible languages, then one might say something like "Flat structure is not possible" or "All structure is hierarchical". Again, since UG is instantiated in real brains, it must consist of a finite set of characteristic properties. Note again, that we would actually need an infinite set of constraining statements to characterize UG—those referring to bananas, jigs, *etc.* Again, there are an infinite number of such constraints on the set of possible languages.

In order to avoid having an infinitely long list of constraints, constraint-based theories need a *list* or positive statements of entities (distinctive features, primitive operations like *Merge*, *etc*). This list will define the universe of discourse in which we interpret a constraint like "Flat structure is not possible". We see, then, that a theory which formulates linguistic universals in terms of constraints must *also* contain a vocabulary of elements and operations in which those constraints are expressed, or to which they refer. This vocabulary of items and processes is presumably based on empirical observations and inferences. Consider a simpler alternative.

If our current hypothesis concerning UG is stated only in *positive* terms, as statements of what grammars have access to or consist of, without prohibitions or constraints, we can achieve a more economical model. The positive terms are just those entities and operations (features, deletions, insertions, *Merge*, *Move*, *etc.*) which have been observed empirically or inferred in the course of model construction. When faced with a phenomenon which is not immediately amenable to modeling using existing elements of the vocabulary, scientific methodology (basically Occam's Razor) guides us. We must first try to reduce the new phenomenon to a description in terms of the vocabulary we already have. If this can be shown to be impossible, only then can we justify expanding the vocabulary.

Thus, a "constraining approach" to UG, stated in terms of what is disallowed, requires a set of constraints, as well as a vocabulary which defines the universe of discourse in which the constraints are valid. The alternative proposed here requires only the vocabulary of possible entities and operations, along with the metatheoretic principle of Occam's Razor. The alternative is thus more elegant and should be preferred.

In more concrete terms this means that our theory of UG should consist of the minimum number of primitives that we need to describe the grammars we have seen.[3] Note that we should not be influenced in our search by preconceived notions of simplicity. For example, if we know that we need hierarchical structure for some phenomena, but there exist other phenomena which are ambiguous as to whether they require flat or hierarchical structure, then we should assume that the ambiguous cases also have hierarchical structure. If our current theory of UG contains an operation to generate hierarchical structure from primitive

[3]According to Rennison (2000:138) this principle has, in practice, been more vigorously upheld by proponents of Government Phonology (GP), than by members of other schools of phonology.

elements, constraints against flat structure will be superfluous. In fact, positive statements like 'Structures are organized hierarchically' and 'All branching is binary' (assuming they are correct) are also superfluous within the grammar itself, even though they are descriptively accurate, since they are just a reflection of how structure building operations work (see section 4).

The approach advocated here seems to be consistent with that used in science in general. If a physicist observes a 'constraint' on the behavior of a particle, say, then s/he posits a set of properties for that particle from which the observed behavior emerges. The constraint thus has the status of a derivative and not primitive aspect of the theory.

3.3 Inviolable Constraints in Grammars

It was suggested above that the issues raised thus far are irrelevant to the choice between rule-based and constraint-based computational systems. In a sense this was an overstatement and the discussion above is in fact clearly relevant to a certain class of constraints invoked in versions of Optimality Theory, as well as other models of phonology: constraints that are never violated, either universally or within individual grammars.

For the sake of concreteness let's adopt a version of Optimality Theory which assumes that it is never the case that the winning candidate in a derivation, in any language, has crossing association lines.[4] There are several ways to deal with this. One possibility is to claim that there exists a constraint, NoCross, that is part of the OT constraint hierarchy which incurs a mark when a candidate contains crossing association lines. This constraint can be posited to be universally undominated, or rather, universally undominated by a 'competing' constraint. A competing constraint which dominated NoCross would be one whose satisfaction could 'force' a violation of NoCross in the winning candidate. This possibility can be construed as allowing simplicity in the theory—allow GEN to generate candidates freely, and leave it to universally undominated constraints like NoCross to rule out candidates with no chance of surfacing. However, the simplicity achieved is somewhat illusory.

This approach introduces a complication into the core idea of Optimality Theory, the idea that grammars are defined by constraint hierarchies. If one adopts the view that constraints are universal and innate, then certain constraints, the undominatable ones like NoCross will have to be kept in a separate stratum of the constraint hierarchy, one whose members are not subject to reranking. Equivalently, they can be marked as not susceptible to reranking.

Yet another approach is to claim that these constraints are high-ranked at the initial state of the grammar. According to the claim of Smolensky (1996) and most other scholars, they would therefore start out at the top of the block of initially high-ranked Wellformedness constraints. If one is willing to accept such a scenario,[5] then the undominatable constraints

[4] This is a particularly well-known and easily discussed constraint. However, Local & Coleman (1994) have demonstrated that it is basically contentless.

[5] But see Hale & Reiss (1998) for arguments that it is untenable. They argue that acquisition under such an initial ranking, with Wellformedness constraints outranking Faithfulness constraints, is impossible. They claim that the (normal, rerankable) Wellformedness constraints must start out ranked below the Faithfulness constraints in order to allow the acquisition of a lexicon. If one adopts this assumption, then, the undominated Wellformedness constraints like NoCross would have to be initially ranked in a block separated from all

need not be marked as unrerankable, since, by hypothesis, no language ever has evidence that they are dominated. However, the generalization that OT grammars consist of freely rerankable constraints becomes empty, if in fact, some of the constraints are never reranked in any language.

We see then that each of the versions of undominatable constraints proposed here leads to complications in the theory of grammar. An obvious alternative is to state the constraints as limitations on GEN. In other words, assume that GEN freely generates—except that it does not generate forms that violate NOCROSS and other undominatable constraints. But this still fails to solve the need to define the universe of discourse for GEN. We would need constraints on GEN to keep it from generating representations that violate NOCROSS, but not ones that violate NOBANANA, presumably. But GEN has certain properties, it does certain things with inputs, and we should try to characterize those properties. Therefore, it seems preferable to model GEN in such a way that it does not have the capacity to output forms with crossed association lines and other impossible traits (including bananas). In other words, the arguments against constraints on grammars and undominatable or inviolable constraints in grammars are the same—we always need a positive characterization of the formal system we are modeling.

3.4 Free generation and constraints as filters

The dominatable, or violable, constraints of both standard OT, which assumes universal, innate constraints, and other theories which allow language specific constraints, do not immediately appear to pose the problems discussed thus far. Such constraints are formal devices for evaluating candidates, but they do not, each on its own, define what is a possible linguistic representation. However, I will argue in this subsection that even a constraint-based grammar which contains violable constraints is to be avoided. In section 7, we will see that the original motivation for such constraints may have been empirically and methodologically misguided.

Various theories of grammar, including Optimality Theory and some versions of Minimalism and its predecessors posit a mechanism that allows unconstrained generation of linguistic representations. In OT this device is GEN which, given an input, generates the universal candidate set of possible outputs. In various syntactic theories, an analog to GEN is the 'free' concatenation of morphemes, or the 'free' application of operations such as *Move* α. A derivation which is thus generated will either satisfy certain conditions at PF and LF, the grammar's interface levels, and thus *converge*; or it will not satisfy those conditions and it will *crash*. Both the OT approach and the free-generation-with-interface-conditions approach in syntax are flawed in the following (related) ways.

First, it is easy to proclaim something like 'GEN generates any possible linguistic representation' or 'The syntactic component allows *Move* α to apply freely'. However, it is not clear what such statements mean. One could argue that the theory of grammar need not be computationally tractable, since grammar models knowledge and does not necessarily map directly to an algorithm for generating grammatical output. However, it does not follow from this that we should immediately aim for a model that we cannot imagine being implemented

the rerankable Wellformedness constraints, or somehow marked as not rerankable.

in the mind. It seems that any implementation of GEN or the syntactic component that incorporates *Move* α will have to be very explicit about what it does. One way to achieve this is to be explicit about what the abstract grammar generates.

Second, the *free generation-cum-filters* model stinks somewhat of antimentalism. It basically says 'We don't care how the candidate forms are generated, as long as they are generated. One way is as good as the next, as long as they are *extensionally* (empirically) equivalent.' This is parallel to the position taken by Quine (1972, discussed by Chomsky 1986) in arguing that it is incoherent to talk about the 'correct' grammar among a class of extensionally equivalent ones. In defining I-language, a matter of 'individual psychology' as the domain of inquiry for linguistics, Chomsky (1986) argued convincingly that the fact that knowledge of language is instantiated in individual minds/brains means that there is necessarily a 'correct' characterization of a speaker's grammar (or grammars). It is worth pointing out that such anti-mentalism is a real part of much current theorizing. Consider, for example, McCarthy's (1999:6) discussion of the OT 'principle' of RICHNESS OF THE BASE, the idea that there are no restrictions on inputs: "with faithfulness bottom-ranked, the choice of input [among three alternatives] doesn't matter, since all map to [the same surface form]. So there is no need to restrict the inputs." McCarthy is confusing the issue of the linguist designing a grammar, *qua* computational system, with the problem of discovering which *mental* grammar the learner acquires. There is no question of 'restricting' the inputs, but rather a question of figuring out which inputs the learner constructs given the observed data. It is something of a perversion of terms to label our hypothesis about what the LAD does a 'restriction', when in fact we mean 'selection of a uniquely defined choice.' See Reiss (2000) for a fuller discussion.

Once one accepts that modules/processes, like GEN and *Move* α, must have a certain set of properties; and that these properties ultimately must be derived from a set of positive statements (a vocabulary); and that these properties can be incorporated into the structural descriptions of rules; it appears to be the case that a procedural, or rule-based approach to grammar that generates a sequence of representations constituting a derivation is to be preferred to a constraint-based, non-derivational theory. In other words, grammars can be understood as complex functions mapping inputs to outputs. A rule-based model just breaks the complex function into simpler components, in order to understand the whole. A theory that incorporates GEN or *Move* α avoids the problem of characterizing the function that is the grammar.

Thus a rule-based derivational model of grammar is better, since it can be stated in purely positive terms, without prohibitions.

3.5 The fallacy of imperfection

In phonology at least, it appears that the obstacle to developing such a theory has been an *a priori* belief in the relative well-formedness of abstract representations based on the never formalized notion of markedness. In other words, even the rule-based phonological literature is rife with constraints which are meant to 'motivate' the application of rules that repair structure. In syntax, the tradition of appealing to markedness is more subtle, but it has basically been adapted in that the grammar, or perhaps the processor, is characterized with

respect to derivations which 'crash', as well as with respect to ones that 'converge'.

Consider for comparison the visual system. Given an input, the visual system is assumed to have certain biases, probably manipulable via the little understood mechanism of *attention*, but no visual input leads to a failure to assign a representation. It is also not clear what it would mean to say that a given representation generated by the visual system was less well-formed, or more marked than another representation. Presumably the visual system generates representations based on the input it is given, and these representations are unique—they are the best and the worst (or rather, neither best nor worst) that the system generates. Outputs are generated which depend on the input and the state of the system processing the inputs—hardly a controversial view. The same holds true of phonological representations—they are not perfect or imperfect, THEY JUST ARE.

Since the violable OT constraints are posited on the basis of cross-linguistic typology, data from child speech and the informal intuition of linguists, it is worth evaluating these criteria. I do so here only briefly. Defining markedness based on cross-linguistic *tendencies* of absolute and implicational patterns of attestation (*e.g.*, If a language has voiced stops, it also has voiceless ones) raises many difficult issues, not least of which is 'How do we count?'. Do we count tokens? E-languages like 'English' or 'Chinese'? Grammars?[6] Without an explicit theory of what gets counted, generalizations based on intuitive 'statistical' patterns are worthless. Furthermore, at least some of the reported statistical tendencies, such as the more common absence of [p] from voiceless stop inventories, in comparison with [t] and [k], are highly reflective of areal biases in the sampling procedure (see Engstrand 1997 and Hale & Reiss 2000ab for discussion).

Hale & Reiss (1998) have argued in detail that the use of child speech data to determine markedness status is flawed since this data is rendered opaque by the effects of children's performance systems. I will not repeat these arguments here. Linguists' intuitions concerning 'better' (unmarked) and 'worse' (marked) structures reflect a confusion of levels of analysis, as well as other conceptual problems. A problem addressed in detail by Hale (2000) is that discussion of the evaluation of 'output' forms often fails to distinguish between the output of the grammar (a feature-based representation) and, say, the output of the speaker (an acoustic or articulatory event). As demonstrated most clearly by our ability to construct 3D representations based on a black and white pattern on a printed page, there is a vast gap between physical stimuli and outputs and the representations that relate to them. Therefore, even if phonologists had a metric of the complexity or difficulty inherent in interpreting or creating certain physical stimuli or outputs (which they do not), it is apparent that there is no reason to believe that such a scale would translate straightforwardly to a markedness scale for representations. There is no reason to believe, for example, that the representation of the act of pushing a boulder is more difficult or complex or marked than the representation

[6]I am collapsing Chomsky's discussion of a sociopolitical conception of 'language', common in everyday parlance, with the E-language conception which he includes among the scientific approaches to the study of language. The E-language approach treats a language as an external artifact, say a text or corpus of texts, rather than as a knowledge state. This collapse is, I believe, justified and consistent with Chomsky's views, since the decision to include various texts or utterances within a single E-language corpus is typically made on the basis of the everyday sociopolitical notion of language—how else can an E-linguist decide that a set of texts constitutes a single corpus, except by appealing to the pretheoretical notion that they are all French or English or Swahili?

of the act of pushing a feather.

3.6 OT constraints as fallible intuitions

> "We should know that one intrinsic characteristic of a heuristic is that it is *fallible*, and that it may be unjustified." *Inevitable Illusions*, Massimo Piatelli-Palmarini (1994:22)

The preceding discussion suggests an explanation of why the constraints of OT are violable. These constraints are for the most part derived from so-called 'principles of well-formedness' or 'markedness' found in other phonological theories. I propose that these 'principles' are actually just the heuristic devices that constitute our intuitions as experienced linguists. For example, we may assume that a sequence like [akra] will more likely have a syllable boundary before the stop-liquid cluster than between the two consonants. This is because we seem to believe, rightly or wrongly (it is hard to imagine how to collect the appropriate statistics under the I-language approach) that the majority of languages 'maximize onsets' in such cases and leave the first syllable without a coda.

However, both syllabifications are found, for example, in the Ancient Greek dialects. Lacking information to the contrary, it may be useful to assume that the more common syllabification is present in a new, unfamiliar language. This will allow the formulation of hypotheses that may then be tested, and the guess will turn out to be correct more often than not, if our intuitions have any basis. However, we must take care not to confuse our intuitions concerning what happens often with the actual nature of the system under study. Based on our experiences and expectations, we apply our intuitions, in attempting to solve the problems involved with analyzing data, but there is no reason to expect that these intuitions directly reflect the nature of the actual mental grammar constructed by a learner. The intuition that heavy things fall faster than light things is very useful when someone drops something from a window, but the intuition needs to be transcended to understand the workings of gravity. Heuristics are used by the analyst to make useful guesses about data, and guesses can be wrong. This is why OT constraints need to be violable—they reflect the fallibility of our guesses.

It may be useful to refer to the error under discussion as a confusion of epistemological issues (concerning the nature of our knowledge) with ontological ones (concerning the nature of phonological systems). One explanation for the pervasiveness of such errors may lie with our terminology. A term like *physics* or *phonology* is used in a systematically ambiguous fashion. *Physics* means both 'the study of the properties of the physical world, including gravitational attraction, *etc.*' and 'the properties of the physical world, including gravitational attraction, *etc.*'. When I fall down the stairs, I do so, not because there is a field of study that concerns itself with gravity, but because of the nature of the physical world, because of gravity itself. I would fall down the stairs even if all the physicists and physics books disappeared—I assume people fell down the stairs before Newton. By failing to make this crucial distinction we can be misled into believing that the *tools* (intuitions) we use in phonology *qua* field of study of the nature of sound systems are constitutive of phonology *qua* the nature of sound systems.

I think the use of violable wellformedness or markedness constraints in OT that are based upon putative statistical tendencies has exactly the status of this kind of error. Reiss (2000) discusses another such case in the OT literature.

3.7 Overgeneration

Pylyshyn (1984:205ff) describes a box emitting certain recurrent patterns of signals. He then asks what we can conclude about the nature of the computational mechanism inside the box, based on the observed pattern of output. The answer is that we can conclude nothing, since the observed patterns may reflect the nature of what is being computed (in his example, the output is a Morse Code rendering of English text, and the observed regularity is the 'i before e, except after c' rule), not the nature of the computer. In Pylyshyn's words "the observed constraint on [the system's] behavior is due not to its intrinsic capability but to what its states represent." (The observed 'constraint' on output vanishes if we input German text, instead of English, since German texts will have words that violate the English spelling rule.) Pylyshyn's example suggests that we should expect our models to overgenerate with respect to the corpus of attested data, since this data is 'sifted' by language change, for example. The language faculty may be able to perform computations that we have not observed because of the forms that language data just happen to take. The solution to this situation is clear: posit the minimal theoretical apparatus needed to generate attested patterns, and don't worry too much about overgeneration. We must assume, as a matter of scientific practice, that newly encountered phenomena will be amenable to modeling using current theories. We may be proven wrong, this is in the nature of inductive reasoning. When we are proven wrong we change the assumptions. This is not a bad situation—it just reflects the eternal incompleteness of scientific knowledge.

The fact that we predict the computational possibility of unattested forms is not only possible, but highly likely, given the fact that the language faculty is embedded in a complex system of other cognitive and physiological modules with which it interfaces. Consider the following example. Suppose that the rule $\mathcal{R}$ of a formal system combines the primitive categories of the system $\{a, b, c, d, e\}$ into ordered pairs such as $< a, b >$, $< e, c >$, $< b, d >$, *etc.*. Suppose that after collecting a sample of data we notice that all ordered pairs have occurred except for $< a, d >$. If we then supplement our characterization of the formal system by adding a constraint $*< a, d >$, what have we gained? We have merely built the descriptive generalization into the grammar. Two preferable alternatives come to mind.

The alternative suggested by Pylyshyn's example is to look outside of the formal system itself. In phonology, for example, the shape of phoneme inventories reflects the nature of sound change and physiological constraints on articulation, not just the cognitive capacity of humans. Not only is it misleading and uninsightful to posit constraints on the formal system that do no more than recapitulate observation, but it also discourages us from looking for a real explanation in a domain other than the characterization of the formal system. (See Hale & Reiss 2000ab for discussion.) This approach is adopted later in this paper (section 10.2) to account for an apparent gap in rules types.

A second alternative to explore is to examine whether $\mathcal{R}$ has been correctly formulated. Many constraint based linguistic analyses are built by positing a spurious generalization,

then adding constraints to the model to account for the cases which do not match the generalization. It seems more elegant to posit our generalizations more carefully. This approach is taken in sections 7-10 of this paper, which deal with so-called OCP effects.

Has the preceding dismissal of concerns of overgeneration made the proposals here vacuous? For example, does the position reduce to the following: 'posit a rule that generates all the attested data, and assume that unattested data is the result of accidental gaps in the corpus'? Fortunately, the answer is that this is not the position I am advocating, and this is because of a simple claim that is in direct conflict with general practice, at least in the phonology literature. The claim is that rules are formulated in the *least* general form that is compatible with the data.[7] Generality of application results from lack of specification in structural descriptions; lack of generality, that is, restrictiveness of application results from richly specified structural descriptions. In the view of acquisition developed in Reiss (1995, 1999) and Hale & Reiss (1999), it is claimed that representations that are more highly specified than necessary for the purposes of generating target output, are a logical necessity in early grammars. Rules are only made more general, that is with less specified structural descriptions, upon exposure to positive evidence. Therefore, a rule of a particular grammar will generate all and only the data whose representations are subsumed by that encountered during the acquisition process. This model of 'restriction without prohibition' is the topic of Reiss (2001).

4 A right-minded approach to syntax

The conclusion to be drawn from the discussion above is that it is in fact best to state our theory of *UG* in terms of a positive list of what can occur. This approach actually does delimit the set of possible languages as well as a theory that states constraints on possible linguistic structures, because the normal interpretation of a formal system defined by a set of properties (a vocabulary) is that the system is exhaustively defined by those properties. (See Rennison (2000) for an explicit discussion along these lines.) One can add or subtract one of Euclid's Postulates and explore the consequences of such a move, but any set of postulates is assumed to be exhaustive once stated. Similarly, in physics new elementary particles are posited only when a phenomenon cannot be accounted for by appeal to those currently identified, or when their existence is predicted on other grounds. Since linguistics posits formal models of (indirectly) observable systems, our current theory is open to revision when forced by new discoveries, but Occam's razor serves as a check on the current version at any particular time. A model characterized by prohibitions in the form of constraints constraints must implicitly be itself constrained by a vocabulary defining the universe of discourse in which the constraints hold. Therefore, such a model contains a certain amount of unnecessary redundancy.

The derivational approach to syntactic relations developed in Epstein, Groat, Kawashima & Kitahara (1998) adopts a viewpoint consistent with the 'rules only' approach to modeling

[7] For example, a palatalization rule that applies before the vowels [i,e] in a language with only the vowels [i, e, a, u, o] should be formulated with the conditioning environment as 'before [-back, -round, +tense, -low] vowels', and *not* as 'before [-back] vowels'.

grammar advocated here. These authors claim (pp. 13-14) that their theory has five innovative properties. The first and the last are most clearly relevant to the discussion in this paper and can be summarized as follows:

(3) Epstein, Groat, Kawashima & Kitahara (1998)

- The syntactic computational system consists only of syntactic rules. There are no relations (like *Government*) that are not derivable from the nature of the rules;
- There are no filters or constraints (on non-existent levels of representation such as DS and SS), but only lexical items and operations on these items.

These authors are able, for the most part, to do away with independently stipulated constraints on movement such as Greed and Shortest Move and instead build their effects into the nature of the rule/process *Merge* itself. I understand the goal of this model to be to formulate a rule/process *Merge* which applies in such a way that its outputs are well-formed, as long as it is possible to generate a well-formed output from the current input. Perhaps a better way to describe the model is to say that outputs are 'formed', or 'not formed', and that the notion 'well-formed' is undefined—and unnecessary.

In the rest of this paper, I explore a parallel approach to phonological derivation. First, I provide some background on the use of constraints within primarily rule-based phonologies. Then I demonstrate the insight that can be gained by building the effects of constraints into the statements of the rules themselves.

5 Constraints in rule-based phonology

Despite the fact that phonologists tend to characterize current debate concerning OT as a question of 'rules vs. constraints', this is misleading (see Archangeli 1997). Many rule-based analyses make use of constraints such as the Obligatory Contour Principle (OCP). Constraints in otherwise rule-based phonologies serve two main purposes. Either they define certain structures as disfavored or illformed, and thus subject to modification by rule; or they are used to block the application of a rule just in case the rules output would be disfavored or illformed. Work by Paradis (1988) and Calabrese (1988) are typical of the use of constraints as diagnostics for repair of certain structures: if a string satisfies the structural description of a constraint, that is, if it violates the constraint, it must be repaired by a rule. The rule-based account of stress systems presented by Halle & Idsardi (1995) appeals to 'Avoidance Constraints' (422ff.) which prevent the application of rules in cases where the rules' output would be a 'disfavored' structure. The OCP has been invoked for both of these purposes in a number of papers, most notably McCarthy (1987) and Yip (1988).

Given the problems with markedness theory alluded to above, note that in the absence of a theory of disfavoredness, this approach is circular: the only real evidence for the disfavored status is that the posited rule appears to be blocked; and the posited reason for the blocking is that the resultant structure would be disfavored. Halle & Idsardi point out that certain advantages derive from mixing rules with constraints in the analysis of individual languages.

In general, the use of constraints allows us to formulate simpler rules. However, they note that a fully rule-based analyis is in principle always possible—Halle & Vergnaud (1987) is an example they cite:

> In Halle & Vergnaud (1987), the full metrical constituency was constructed, and at the end disfavored configurations [like stress clash] were eliminated by the application of a rule.

I propose that considerations of elegance for a theory of UG take precedence over elegance in the analysis of individual languages, and thus the Halle & Idsardi system, for example, should be adapted in a way that preserves its mathematical explicitness, while doing away with constraints on unattested structures. A possibility which Halle & Idsardi do not consider[8] is to make the structural descriptions of their rules more complex. As these authors point out, some languages do tolerate stress clash and thus their avoidance constraint is specific to those languages which do not tolerate clash. The rewards of allowing for more complex rules are considerable: constraints become unnecessary and the effects of earlier rules need not be undone.

In brief, Halle & Idsardi need the avoidance constraint AVOID(x(to prevent the generation of Line 0 metrical structures such as (x (x x (x x in a language like Garawa that (1) inserts the leftmost left parenthesis on the basis of an Edgemarking rule, and (2) inserts left parentheses interatively from the right edge after every second syllable. In a word with an even number of syllables, steps (1) and (2) give, *e.g.*, *(watjim(paŋu*. However, in a word with an odd number of syllables the rules outlined above would generate a 'disfavored' (x(structure like *(na(řiŋin(muku(njinam(iřa* where the leftmost syllable has a left parenthesis on both its right and its left. The avoidance constraint blocks the insertion of a parenthesis to the left of the second syllable from the left, and the actually generated Line 0 form is *(nařiŋin(muku(njinam(iřa* with a trisyllabic leftmost constituent.

Instead of appealing to an avoidance constraint, the so-called Iterative Constituent Construction rule can be specified to insert a left parenthesis only in the environment x x _ x x. By the normal conventions of interpretation, the structural description is not satisfied by the following structure: x (x _ x x. Thus, the stress clash configuration is not generated.[9] Again, we cannot rule out such complications to rules *a priori*, without considering that the use of the simpler rule requires adding an additional rule to the grammar (in the Halle & Vergnaud formulation) or else enriching grammatical theory by the use of avoidance constraints (in the Halle & Idsardi formulation).[10]

[8]Idsardi (1992), however, does have a useful discussion of rule-, constraint-, and rule-and-constraint-based approaches to stress.

[9]Because it is not relevant to the discussion, I ignore here the further steps in the derivation, those which follow the construction of the Line 0 structure.

[10]There are, in fact, other plausible rule-based analyses. Morris Halle (p.c.) points out that by first building a single binary foot from the *left* edge of the word, then building binary feet iteratively from the right, the third syllable from the left will remain unfooted in words with an odd number of syllables, but not in those with an even number.

Even number of syllables: x x) (x x (x x
Odd number of syllables: x x) x (x x (x x

By projecting the leftmost syllable of each foot, the correct Line 1 configuration is generated for all words.

I thus propose that a goal of future phonological research should be to take the idea of rule-based phonology seriously—by avoiding constraints altogether. Such an approach will offer a principled alternative to Optimality Theory and other constraint-based models. In other words, rather than stating simple, but empirically inadequate rules, reinforced by an arsenal of language particular or universal constraints, we should attempt to understand what kind of rules we actually need if we are to do without any constraints.

Part of the groundwork for this approach was done over ten years ago in a pair of underappreciated papers by David Odden (1986, 1988). Odden demonstrated that the OCP is demonstrably *not* a universal constraint on either underlying representations or on the workings of the phonological component. Odden also points out that work appealing to the OCP is unacceptably vague in defining how, for example, identity of representations is computed. These arguments need not be repeated here, since my goal is to reject the use of all constraints on more general grounds.

6 Identity references—a positive contribution

In the next few sections I attempt to develop the "adequate formal account of identity references" that Odden (1988) demonstrates is necessary for phonological theory. As Odden points out

> ...languages differ in what constitutes 'identical' segments. Biblical Hebrew identical consonant fusion requires reference to complete identity (including voicing). Syrian Arabic allows identity to ignore pharyngealization and voicing, Koya allows identity to ignore retroflexion, and Telugu Syncope requires only rough identity computed at the place of articulation, which ignores voicing and narrow place distinctions such as alveolar/retroflex/palatal [461].

We will look at some of this data below in section 7, which discusses the status of the Obligatory Contour Principle (OCP) from both an empirical and methodological perspective.

Again following Odden, I assume that "It is misguided to attribute every accidentally true statement about human language [or particular human languages—cr] to UG, for doing so trivializes the theory of UG itself"(461). Thus, linguistic theory should attempt to unify diverse phenomena by analysing them at an appropriately abstract level, instead of merely cataloging observations. In this spirit, I follow up on Odden's groundbreaking work, and related observations by Archangeli & Pulleyblank (1994), to propose that phonological theory needs the power provided by the existential quantifier and the universal quantifier to express identity references as conditions within the strutural description of rules. Section 8 develops the notion of identity and nonidentity conditions in rules, and in section 9 I propose and justify a new notation for expressing such conditions. My development of Odden's work serves as a small step towards making explicit what formal apparatus phonology must have access to.

Building on these results, I show that the need for quantificational statements entails the rejection of feature geometry in phonological representation. Feature geometric representation is insufficiently powerful, and must be replaced by an algebraic form of representation that allows the use of variables and indices for the purposes of identity checking.

An immediate benefit of approaching phonology in this minimalist fashion is that it helps us to discover that, in section 10, that certain *a priori* plausible rule- (or constraint-) types are actually unattested. In 10.2, I discuss this kind of 'overgeneration' and propose an explanation for such gaps in attestation.

7 The dubious status of the OCP

McCarthy (1986) discusses data from several languages in which a vowel which is expected for independent reasons to be deleted, is instead preserved if its deletion would cause identical consonants to be adjacent: Biblical Hebrew /ka:tab-u:/ → [ka:θvu:] but /sa:bab-u:/ → [sa:vavu:] because deletion would bring together the two underlying [b]'s (both of which are spirantized by an unrelated process).[11] The 'failure' of the deletion rule to apply is dubbed *antigemination* by McCarthy, since the rule is 'blocked' if its application would produce a geminate. McCarthy invokes the Obligatory Contour Principal (OCP) as the constraint which blocks the rule from applying. This phenomenon involves the failure of deletion rules just in cases where the rule would result in a string of identical adjacent consonants.

Yip (1988) provides a very useful summary, elaboration and discussion of McCarthy's treatment of the OCP as a blocker of rules. Consider the following argument:

> If a language has a general phonological rule that is blocked just when the output would contain a sequence of identical feature matrices, we can conclude that the OCP is operating to constrain derivations ... The alternative is an ad hoc condition on such rules, as in [4]:
>
> (4) $A \rightarrow \emptyset / B\_C$
> Condition: $B \neq C$
>
> Such a condition not only incurs an additional cost (whereas the OCP is taken to be universal) but also lacks explanatory power, particularly if contexts B and C are necessary only to state the ad hoc condition.

In other words, Yip argues that a theory with language specific rules and a universal OCP is a better theory than one with language specific rules that correctly encode where the rule applies, because adding the necessary conditions to the statement of such rules makes them more complex.

Note that the examples that Yip mentions conform to the first (a) of the following three types of conditions on rule application, but Odden (1988) points out that in fact vowel syncope rules are found with all three of the following types of conditioning:

(5) Some conditions on vowel deletion rules (Odden 1988:462)

a. Delete a vowel unless flanking Cs are identical.

[11] It has been brought to my attention that vowel length in the Hebrew is actually difficult to determine. However, this issue is irrelevant to the point under discussion—any example of 'antigemination' will do and additional ones are provided below.

b. Delete a vowel blindly [whatever the flanking Cs are].

c. Delete a vowel only if flanking Cs are identical.

Condition (a) can be restated as 'Delete a vowel if flanking Cs are *not* identical'. This is the condition described but rejected by Yip in (4) above: $B \neq C$. But note that Odden's type (c) condition would be written as follows:

(6) Odden's Condition (c) in the notation Yip rejects: $B = C$

In other words (a) demands nonidentity and (c) demands identity of segments in the structural description of a rule. Thus, there is no reason to propose, as McCarthy and Yip do, that rules that conform to condition (a) illustrate a universal principle of markedness—condition (c) is also a possible rule condition. A rule like (5c) *only* applies when it creates OCP violations—Odden refers to this phenomenon as *antiantigemination*. So a theory of *UG* must allow for both types. There is thus no good reason to claim that a universal principle, the OCP, *blocks* deletion in the (a) cases, since deletion can also be *required* in cases that lead to apparent OCP violations when a rule with conditions (b) or (c) applies. Stated in McCarthy's terms (although he does not mention such cases), deletion can be blocked (in case (c)) if the rule will *not* generate an OCP violation. This point was clearly made by Odden, though it seems to have been ignored in most of the subsequent literature.[12]

Note that the logic of attributing cases that fit the profile of (a) to a universal principle and ignoring cases that fit (c), is incoherent. Suppose we examine some data concerning a certain phenomenon and find that all cases fall into two categories, x or y. If we present only cases of x and proclaim that we have found that x is always true, then our claim is not valid, *no matter how many positive examples of* x *we adduce*. The existence of (c) cases, makes the existence of (a) cases uninteresting on their own. Odden's observations taken together *are* interesting, as we will see below. Simply put, case (c) is a counterexample to the claim that (a) is universal.[13]

7.1 Treating phonological pathology: The OCP as a rule trigger

The main point of Yip's paper is that the OCP not only *blocks* rule application as in McCarthy's antigemination cases, but also *triggers* it—it may be the case that a rule applies only to an input that violates the OCP. Instead of an argument based on formal simplicity in rule statements, as discussed above, Yip's discussion of the OCP as a rule trigger illustrates particularly well the assumption that the phonology repairs structures that are somehow pathological—illformed or marked or disfavored: "The main contribution of the OCP is that it allows us to separate out condition and cure. The OCP is a trigger, a pressure for change" (74).

[12]For example, Keer's (1999) recent OT thesis on the OCP, lists Odden's papers in the bibliography, but makes no reference to them in the text, even in sections discussing antigemination.

[13]Providing a principled response to the reader who finds this discussion to constitute an argument for the violable constraints of Optimality Theory is beyond the scope of this paper, or perhaps even impossible, reducing to a question of faith, but see section 3.6.

In Yip's model the 'cure' is effected by language specific rules. In OT models that make use of similar constraints the 'cure' emerges from the constraint ranking. Because of the violability of OT constraints, the winning candidate in an OT derivation is typically not fully 'cured'—certain marked structures may be present in the output form.[14] One goal of this paper is to work towards removing the notion of ill-formedness from the generative component of the phonology. There are representations that are generated, or formed, by grammars; there are representations that are not generated—that is, not formed; but there is no reason to believe that anything a grammar actually generates is ill-formed.

Yip provides a range of examples that show how different solutions can be applied to OCP violations. They include deletion, dissimilation and assimilation rules (where assimilation represents multiple linking of a single node, and not identical adjacent nodes). One example of repair by deletion comes from Seri (Marlett and Stemberger 1983). This language has a rule that deletes a coda glottal stop in a syllable with a glottal stop in the onset:

(7) *Seri Glottal Stops*

a. ʔa-a:ʔ-sanx → ʔ-a:-sanx 'who was carried'

b. ʔi-ʔ-a:ʔ-kašni → ʔi-ʔ-a:-kašni 'my being bitten'

c. koʔpanšx 'run like him!'

The rule only applies to tautosyllabic glottal stops so the second glottal stop in (7b) is not affected. In general, coda glottal stops can surface, as shown by (7c).

Yip's account of this process is the following:

> [We can] assume that the Laryngeal node is absent except for /ʔ/, and the entries for glottalization in [7ab] are thus adjacent and identical and violate the OCP. This violation triggers a rule that operates in the domain of the syllable, and the language chooses [one of the possibilities for repairing OCP violations,] deletion of one matrix (either [+constricted] or [Laryngeal]). The actual rule has four parts, as shown in (8):
>
> (8) *Glottal Degemination*
> Domain: Syllable
> Tier: Laryngeal
> Trigger:
> Change: Delete second
>
> The environment is not stated, so the rule is unable to operate unless triggered "from the outside". The outside trigger is, of course, the OCP, a universal principle and thus free of charge.

In another example, Yip proposes that English uses epenthesis to 'cure' OCP violations of adjacent coronal stridents, thus accounting, for example, for the form of the plural morpheme

[14] We might refer to this idea as OT's Fallacy of Imperfection. Imperfection, or markedness, seems to be as irrelevant to linguistic theory as the notion of perfection.

after coronal stridents: *judges, couches, bushes, cases, etc.* In other words, if epenthesis did not apply, the adjacent coronal stridents would constitute an OCP violation. As Odden (1988) points out, the OCP is invoked rather opportunistically—note that it appears to be irrelevant to identity of adjacent [+voiced] specifications in words like *bins, rugs, hills, cars.* More seriously, Odden points out that there are rules that insert vowels only when doing so will specifically *not* repair an OCP violation. This is case (d) below. There are also rules that insert vowels regardless of the nature of the flanking consonants—case (e). And of course, there are rules that, like English epenthesis, depend on the total or partial identity of flanking segments—case (f).

(9) More conditions on vowel insertion rules (Odden 1988:462)

d. Insert a vowel unless flanking Cs are identical.

e. Insert a vowel blindly [whatever the flanking Cs are].

f. Insert a vowel only if flanking Cs are identical.

Parallel to (a), condition (d) can be restated as 'Insert a vowel if flanking Cs are *not* identical.' Thus there is no reason to see (f) as reflecting the OCP as a trigger when (d) shows that rules may be triggered if and only if they *fail* to fix OCP violations. The existence of rules with conditions (c) and (d) make it unlikely that appealing to the OCP as either a trigger or blocker of rules is a fruitful endeavor.

8 The Identity and Nonidentity Conditions

More of Odden's data will be presented below. For now, note that it is equally possible for a rule to generate OCP violations (c) as it is to repair them (f). And it is equally possible for a rule to be 'blocked' from generating OCP violations (a) as to be blocked from fixing them (d).[15] Since the goal of phonological theory should be to define the set of computationally possible human languages, Odden's observations provide an excellent opportunity to study the purely formal nature of linguistic rules. In the following discussion, we will concentrate on syncope rules as a matter of expository convenience. Again, for expository convenience, we will refer to a schematic representation C_1VC_2. Odden's conditions (a) and (c) can be restated the following:

(10) The Nonidentity Condition on syncope rules (Version 1)

Delete a vowel if flanking Cs are *not* identical ($C_1 \neq C_2$).

(11) The Identity Condition on syncope rules (Version 1)

Delete a vowel if flanking Cs are identical ($C_1 = C_2$).

[15] Of course, (b) also potentially generates OCP violations, and (e) potentially repairs OCP violations.

The apparatus of phonological representation must be at least powerful enough to express the NONIDENTITY CONDITION and the IDENTITY CONDITION. This issue has implications for Feature Geometry as a model of phonological representation.

There is an insightful discussion of the need for Identity Conditions in Archangeli & Pulleyblank (1994:368-373). These authors point out that "linked structures themselves are simply one type of configuration involving identity" (369). Archangeli & Pulleyblank present the 'Identity Predicate', a relation holding between two arguments, which "is important in a wide variety of phonological contexts" (369). In addition to the OCP cases, they cite the case of Tiv where [+round] spreads between vowels, if and only if they agree in height. Arguments against a linked structure analysis of identity conditions include cases where identity holds across a morpheme boundary—since the identical features belong to different lexical items, they cannot be stored as linked.

In the next section, I will formalize the identity condition and offer further arguments for the inadequacy of a 'linked structure' analysis of these conditions. Archangeli & Pulleyblank mention identity conditions holding of whole segments, as well as of individual features. We will see that it is also necessary to allow identity conditions over arbitrary subsets of the feature set. I will also show that linking is inadequate for the expression of non-identity conditions.

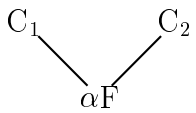
9 Feature algebra and conditions on rules

This section demonstrates that the notation of standard Autosegmental Representation (AR) is insufficiently powerful to represent the NONIDENTITY CONDITION and the IDENTITY CONDITION. I propose a solution to this problem by incorporating basic quantificational logic into an algebraic system of phonological representation called feature algebra (FA). I will argue that this system of representation has all the expressive power of Feature Geometry, as well as additional power that allows us to state hitherto unformulated aspects of phonological UG. The concern of this section will be with stating the structural descriptions (SDs) of rules. In other words, we will concentrate on describing the representations to which a rule applies, and not on the statement of the structural changes effected by the rule.

9.1 Shared feature values

One of the advantages of autosegmental theory, including Feature Geometry, is that it provides us with a visual representation of a situation in which two segments share a single specification for a given feature F_n. Such a situation can be a condition on the application of a rule (or the relevance of a constraint):

(12) Two segments linked to the same valued feature



Suppose that instead of representing this situation geometrically we did it algebraically with indices. For example, let C_1 and C_2 be understood as abbreviations for feature matrices such as the following:

(13) Segments as feature matrices

$$C_1 = \begin{bmatrix} (\alpha F_1)_1 \\ (\beta F_2)_1 \\ (\gamma F_3)_1 \\ \vdots \end{bmatrix} \qquad C_2 = \begin{bmatrix} (\delta F_1)_2 \\ (\epsilon F_2)_2 \\ (\zeta F_3)_2 \\ \vdots \end{bmatrix}$$

F_i denotes a feature, such as [nasal] and Greek letter variables denote the value ($\pm$) that feature F_i has for a given segment.[16] The subscript outside of a pair of parentheses containing αF_i denotes the segment in question; thus, these subscripts are always 1 for C_1 and 2 for C_2.

If we want to express a state of affairs in which C_1 has the same value for some feature F_n as another segment C_2, we can express this as follows:

(14) Identical values for F_n using FA: $[(\alpha F_n)_1] = [(\beta F_n)_2]$

We thus express the fact that C_1 and C_2 have the same value for the feature mentioned. Perhaps we lose the visual metaphor of shared nodes, but the required *identity* condition on values is expressed by the equation. Obviously, this system can be extended to an arbitrary subset of the total set of features, even to the set of all features. We will do so below to formalize the IDENTITY CONDITION, corresponding to Odden's condition (c).[17]

9.2 Indifferent feature values

An Autosegmental Representation may show two segments which are not linked with respect to a given feature. In such a case the standard interpretation is that such linking or lack thereof, is irrelevant to the application of the rule in question. The two segments *may* have identical values for a given feature, but this issue does not bear on the rule's applicability. This corresponds to Odden's condition (b). An example is a rule of schwa syncope in Hindi (Bhatia & Kenstowicz 1972). The form *daanəw+i* surfaces as *daanwi*, and *kaanən+i* surfaces as *kaanni*, showing that syncope is indifferent to identity or nonidentity of flanking consonants. The absence of association lines in the AR model is equivalent to the absence of an explicit algebraic statement of a relationship in our algebraic model. Since we constantly write rules which apply to *classes* of sounds in *classes* of environments, it is obvious that

[16] I continue to refer to segments for expository convenience, however, the valued features belonging to a given segment are more accurately characterized as the valued features sharing an index. These indices, in turn are best understood as denoting association to elements of an X-slot or CV timing tier—valued features with identical indices are linked to identical elements of the timing tier.

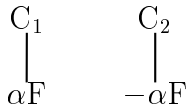
[17] A reviewer of an abstract of this paper complained that the formalization developed here merely restates Odden's original observation. I refer to Halle (1975:532) for discussion of the importance of explicit, careful formalization: "[D]etailed concern for the formal machinery of phonology has led to significant insights into the relationship between superficially disparate facts ... [I]t has paid off in terms of a deeper grasp of the significance of certain empirical facts." My debt to Odden is, I assume, obvious.

some feature values are irrelevant to the application of certain rules. For example, a rule that voices all stops between vowels does not refer to the various place of articulation features of the potential rule targets.

9.3 Obligatorily different feature values

Consider a rule which is only applied if two segments *disagree* with respect to some features. That is consider rules conforming to Odden's conditions (a) and (d). For now let's consider a simple case where the two segments must disagree with respect to a single feature F_n. This can be represented trivially by overtly specifying the two segments, one as αF_n and the other as $-\alpha F_n$:

(15) Autosegmental representation of two segments with distinct values for a feature F



In algebraic terms this is easy to represent:

(16) Distinct values for F_n using FA: $[(\alpha F_n)_1] \neq [(\beta F_n)_2]$

The algebraic formulation (16) expresses the fact that C_1 and C_2 have the opposite value for the feature mentioned. Again, it is obvious that this system can be extended to an arbitrary subset of the total set of features. (It can be extended to the whole set, but appears not to be—see below.)

9.4 The extra power of feature algebra with quantifiers

Note, however that there are examples of rule conditions (a) and (d), those that require that two segments be distinct, that cannot be expressed using just feature geometric association lines or feature algebra as sketched thus far. For example, imagine a requirement that C_1 and C_2 be different with respect to some arbitrary feature, that is any feature, or any feature out of a predefined subset of all the features. In other words, the two segments must *not be identical*, but it doesn't matter how they differ.

Let **F** be the set of all features. In order to express such a NONIDENTITY CONDITION we can make use of the existential quantifier:

(17) The NONIDENTITY CONDITION in FA (defined over all features)

$\exists\ F_i \in \mathbf{F}$ such that $[(\alpha F_i)_1] \neq [(\beta F_i)_2]$
There is at least one feature for which segment$_1$ and segment$_2$ have different values.

That is, there is some feature for which the two segments have a different value. Note that there is no way to represent NONIDENTITY using just Autosegmental Representation. This is because nonidentity can be due to a disagreement with respect to *any* arbitrary feature. The essence of autosegmental notation is the way in which it provides a geometric model

of phonological structure. Being geometric, autosegmental represetation does not make use of variables. Therefore, autosegmental notation is not sufficiently powerful to express non-identity conditions.

The requirement of difference can also be restricted to a subset of the features, $\mathbf{G} \subseteq \mathbf{F}$, for example, to the place of articulation features. Then, the more general version of the NONIDENTITY condition is the following:

(18) The NONIDENTITY CONDITION in FA (Final version)

$\exists$ $\mathrm{F}_i \in \mathbf{G}$ such that $[(\alpha \mathrm{F}_i)_1] \neq [(\beta \mathrm{F}_i)_2]$
For some specified subset of the features, there is at least one feature for which $\mathrm{segment}_1$ and $\mathrm{segment}_2$ have different values.

As we will see below, such a condition is necessary for the formulation of some well known phonological processes.

Once we admit the necessity of quantificational statements in our phonology we can see that conditions of identity can be also be expressed in such a fashion. Total identity can be expressed as follows:

(19) The IDENTITY CONDITION in FA (defined over all features)

$\forall$ $\mathrm{F}_i \in \mathbf{F}$ $[(\alpha \mathrm{F}_i)_1] = [(\beta \mathrm{F}_i)_2]$
For all features, $\mathrm{segment}_1$ and $\mathrm{segment}_2$ have the same value.

whereas partial identity can be expressed by defining a subset of features, $\mathbf{G} \subseteq \mathbf{F}$ over which identity must hold. Total identity is just a special case of partial identity, where $\mathbf{G} = \mathbf{F}$.

(20) The IDENTITY CONDITION in FA (Final version)

$\forall$ $\mathrm{F}_i \in \mathbf{G}$ $[(\alpha \mathrm{F}_i)_1] = [(\beta \mathrm{F}_i)_2]$
For some specified subset of the features, $\mathrm{segment}_1$ and $\mathrm{segment}_2$ have the same values.

Again, we shall see that such conditions are part of phonology.

9.5 Examples of conditions on identity and nonidentity

Note that McCarthy's account of antigemination, which uses the OCP to block rule application, involves a 'lookahead' effect: the phonology must see what the outcome of the rule *would be* and then 'decide' whether or not the rule is to be applied. In effect the rule must be done and undone if the outcome is not satisfactory. An alternative to the rules-and-constraints lookahead solution is to build into the rule the conditions on its application. Note that this condition is just a part of the rule's Structural Description (SD), and a SD is needed in any event. In the case of antigemination, if the OCP is dispensed with, there is no lookahead, but instead a NONIDENTITY CONDITION is built into the rule. McCarthy's rule deletes a vowel in the environment $\#\mathrm{CVC}_1\_\mathrm{C}_2\mathrm{V}$, unless it is blocked by the OCP. I propose replacing McCarthy's rule with one that deletes the vowel in the environment $\#\mathrm{CVC}_1\_\mathrm{C}_2\mathrm{V}$ if $\exists$ $\mathrm{F}_i \in \mathbf{F}$ such that $[(\alpha \mathrm{F}_i)_1] \neq [(\beta \mathrm{F}_i)_2]$.

9.5.1 The Nonidentity Condition

All the examples of OCP blocking cited by McCarthy, including the Biblical Hebrew case illustrated on page 18, can be restated as rules with a condition 'apply unless two segments are identical'. Again this is equivalent to 'apply only if two segments are different, that is, non identical'. So these rules all exemplify the Nonidentity Condition. Odden and McCarthy also provide an example of anti-gemination from Iraqi Arabic:

(21) Anti-gemination in Iraqi Arabic (from Odden 1988:452)

a.	xaabar	'he telephoned'	xaabr-at	'she telephoned'
	ħaajaj	'he argued'	ħaajij-at	'she argued'

b. *Syncope*

$$V \rightarrow \emptyset \ / \ V(C)C\_\_\_CV$$

The syncope rule applies normally in the form *xaabrat*, but is blocked, according to McCarthy, in *ħaajijat* to avoid generating an OCP violation.

Under the theory developed here, rule application is not blocked by the OCP, but rather, the rule's SD includes a nonidentity condition. Rule 22 shows the necessary indexing of the consonants in the structural description:

(22) Revised Iraqi rule

$$V \rightarrow \emptyset \ / \ V(C)C_1\_\_\_C_2V$$
$$\text{if } \exists\ F_i \in \mathbf{F} \text{ such that } [(\alpha F_i)_1] \neq [(\beta F_i)_2]$$

To reiterate, the syncope rule is not written in an overgeneral form and *blocked* by the OCP, but instead the rule contains a condition (part of the structural description) which determines that the rule only applies when the consonants on either side of the vowel differ with respect to at least one feature. Since structural desciptions are obviously needed to determine where rules apply, there is no reason to decide arbitrarily that, say, the presence of flanking consonants should be part of the structural description, but that their identity or non-identity should not be.

The Cushitic language Afar, dicussed by McCarthy (1986) and Yip (1988), based on data found in Bliese (1981) provides a similar case of antigemination which is sensitive to a nonidentity condition. Vowels which appear under stress in the left hand column of (23b) are deleted in the related forms to the right, since the stress has shifted. The unstressed vowels in (23c) do not delete. The rule is stated in (23a) with the relevant nonidentity condition, and the notation [-stress].

(23) *Syncope in Afar*

a. $V \rightarrow \emptyset \ / \ \#CVC_1\_\_\_C_2V$ if $\exists\ F_i \in \mathbf{F}$ such that $[(\alpha F_i)_1] \neq [(\beta F_i)_2]$
[-stress]

b.	xamíla	xamlí	'swampgrass' (acc/nom-gen)
	ʕagára	ʕagrí	'scabies' (acc/nom-gen)
	digibté	digbé	'she/I married'
	wagerné	wagré	'we/he reconciled'
c.	mi~~d~~á~~d~~u	mi~~d~~a~~d~~í	'fruit' (acc/nom-gen)
	xararé		'I, he burned'
	danané		'I, he hurt'

The alternations in the (b) forms shows the deletion of unstressed vowels in open syllables. The first two lines show the relevance of stress— only unstressed vowels are deleted. The next two lines show that deletion does not occur when the syllable is closed. The (c) forms show that the rules does not apply when the flanking consonants are identical. In other words, the rule applies only between non-identical consonants.

As noted above, there is no way to refer to non-identity with respect to some arbitrary feature without making use of the existential quantifier. In the next subsection, I will show that the use of the universal quantifier is not only as good as a geometric representation to express identity conditions, but that in some cases, a geometric representation will be insufficient, so the quantificational formulation is the only one with sufficient power.

9.5.2 The Identity Condition

In this section, I repeat three of Odden's examples of deletion processes requiring identity between flanking segments. According to Odden, Sherwood (1983) motivates a rule in Maliseet-Passamaquoddy which deletes the short vowels /ə/ and /ă/ in doubly open syllables when flanking consonants are identical. Obviously, this identity condition must be stated so that it applies to the whole feature set (or at least to those relevant to consonants). The Hebrew syncope rule we began with deletes vowels unless the flanking consonants are identical. We have encoded the condition 'unless identical' as 'necessarily nonidentical', using the existential quantifier and negation of identity. In the Maliseet-Passamaquoddy rule, a vowel deletes only when the flanking consonants are identical. This requirement can be encoded using the universal quantifier and the identity relation.[18]

Other rules mentioned by Odden, such as one posited by Jensen (1977) in Yapese, demand only homorganicity between flanking consonants, and not identity of Laryngeal or manner features. The Yapese rule deletes a vowel flanked by homorganic consonants if the first consonant is postvocalic or word-initial:[19]

(24) Yapese: syncope between homorganic consonants

a. $\mathrm{V} \rightarrow \emptyset$ /{V, #} C_1__# C_2
if $\forall \mathrm{F}_i \in$ {[coronal], [labial], [dorsal]} $[(\alpha \mathrm{F}_i)_1] = [(\beta \mathrm{F}_i)_2]$

[18] We can remind the reader here that, as is always the case, only one of the two quantifiers is necessary, since they can each be derived from the other *via* negation. For example, '$\forall x, x = y$' is equivalent to '$\neg\exists x$ such that $x \neq y$'; and '$\exists x$ such that $x \neq y$' is equivalent to '$\neg\forall x, x = y$'. I continue to make use of both quantifiers for ease of exposiion.

[19] Presumably the correct generalization is that the first consonant is in an onset.

	Underlying	Surface	Gloss
b.	ba puw	bpuw	'it's a bamboo'
	ni te:l	nte:l	'take it'
	rada:n	rda:n	'its width'

The Yapese data shows that e can use the universal quantifer to express identity. We now demonstrate that we should use the FA system with quantification, since Feature Geometry is insufficiently powerful.

9.5.3 Another failing of Feature Geometry

Odden cites data from Koya (Taylor 1969:38) in which word final vowels are deleted if flanking consonants are identical, except that retroflexion is not used in the computation of identity. In other words, retroflex consonants group with plain coronals for the purposes of computing identity. The data and a FA formulation of the rule are given in (25).

(25) Koya: syncope between identical consonants—ignoring retroflexion

a. $V \rightarrow \emptyset$ /C_1__# C_2
if $\forall$
$F_i \in$ {[coronal], [labial], [dorsal], [spread glottis], [sonorant], [nasal], [lateral]}
$[(\alpha F_i)_1] = [(\beta F_i)_2]$

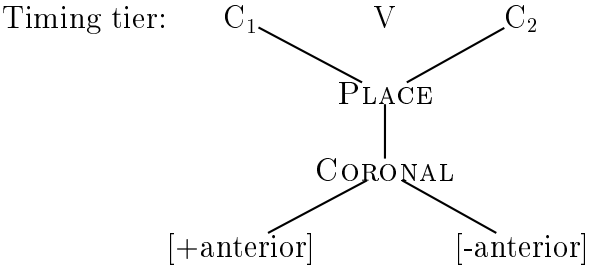
	Underlying	Surface	Gloss
b.	na:ki ka:va:li	na:kka:va:li	'to me it is necessary'
	a:ru ru:pa:yku	a:rru:pa:yku	'6 rupees'
	verka:ḍi digte	verka:ḍdigte	'the cat got down'

Rule (25a) is not particularly pretty, but it is correct (insofar as the set of features listed as relevant to consonant identity, retroflexion excepted, is correct) as a formalization of Odden's discussion.

It is extremely important to note that we have here further evidence for the inadequacy of Feature Geometry. As discussed above, Autosegmental Representation and Feature Geometry can be used to denote shared feature values, or even shared classes of features by a many to one mapping of segments (say, Root nodes) to feature structures. There are two implicit assumptions involved in using such representations. First, there is the assumption of locality: linked structures are adjacent in some sense, so that there are no intervening association lines to cross. In order to treat consonant identity across an intervening vowel as due to autosegmental linking, it is necessary to assume that the intervening vowel have no features which are shared with the two consonants. This assumption forces us to posit a different set of features for vowels and consonants, thus leading to the question of how the two types of segment can influence each other, or else it requires ad hoc segregation of vowels and consonants onto seperate tiers. See Archangeli & Pulleyblank (1994:368-70) for discussion.

Second, if nodes X and Y on a given tier both dominate a node A on a lower tier, then X and Y are assumed to be identically specified for node A (obviously) *and* any nodes that A dominates (by the transitivity of dominance). To illustrate, a structure like (26) is not well-formed since the [+anterior] and [-anterior] nodes are not ordered—temporal relations can only be determined with reference to the timing tier or Root tier.[20] In order to encode the ordering between these two specifications, one could mark them with an index to link them with C_1 and C_2 respectively. But in doing so, we have reverted to an algebraic notation.

(26) An illicit representation



But this view is incompatible with the situation in Koya where the retroflexion features are irrelevant to the compuation of identity. Let's simplify matters and assume that the Koya vowels are not specified for Place features and thus flanking consonants can be considered adjacent. The syncope rule might then be assumed to apply when the flanking consonants share a Place node. However, the consonants should then be assumed to share all the dependents of the Place node. The Place node dominates the Coronal node and its dependents [anterior] and [distributed] in most models of Feature Geometry. However, the Koya rule applies even when the flanking consonants are not identically specified for these features. So, the rule cannot be stated using multiple linking of PLACE nodes or CORONAL nodes.

If we take this as evidence that [anterior] and [distributed] are not dominated by Coronal, but are perhaps directly dominated by the Root node, we are just claiming that Feature Geometry contains less structure, that is we are arguing against its usefulness as a representation of the organization of features, and we are heading back to a model of unorganized feature matrices, as in (13). FA, in contrast to Feature Geometry, allows us to list all and only those features which are relevant to the Koya identity condition. We thus see that even in certain cases of the identity condition, FG representation is insufficiently powerful. It does not allow us to *exclude* from consideration of identity computations those nodes whose dominating nodes are *included* in the computation.

In order to show that this is not an isolated example, consider an additional case concerning place assimilation in modern Irish (pointed out to me by Morris Halle, p.c., discussed by Ni Chiosán and Padgett (1993) and Halle, Vaux & Wolfe(2000). In this language, "word-final coronal nasals assimilate the primary place of articulation of a following stop, but crucially do not assimilate the secondary articulation of the stop. Palatalized [n^j] assimilates the Dorsal

[20]Such representations have been proposed for affricates–with a single Root node dominating both a [+continuant] and a [-continuant] specification. In such cases, one might propose that a default ordering principle ensures that the [-continuant] specification be ordered first, since that is how they are ordered in affricates. Such a strategy will not work in the present case, since the order of the features is not fixed in Koya.

articulation of a following [g], but not its contrastive [+back] specification. Nonpalatalized [n] assimilates the Dorsal articulation of following [g^j], but not its [-back] articulation. In other words, the secondary articulation feature [back] does not spread whenever its supposedly dominating node spreads. The authors cited offer different solutions to this problem, while still maintaining the basic FG model. I take the data to be consistent with a rejection of FG, as argued for above.

Odden also mentions several cases of insertion rules that rely on an identity condition. For example, in Lenakel (Lynch 1978) schwa is inserted between identical consonants. Yip's examples of rules that break up (partially) identical clusters of consonants, such as the English epenthesis between coronal sibilants, can all be restated in terms of rules constrained by identity conditions.

10 Unattested Conditions

The previous section was really no more than a slight elaboration on Odden's important work on defining the range of possible conditions on rule application. In this section I suggest that the use of FA actually helps us discover that two types of rule condition are unattested in phonology.

Phonological theory needs at least the power of the NONIDENTITY CONDITION and the IDENTITY CONDITION. Interestingly, it appears not to need the power of the two conditions which are made by switching = and $\neq$ in the conditions already established. That is, no phonological rule appears to require what can be called the COMPLETE NONIDENTITY CONDITION (two segments must have opposite feature values for all of a given subset of features)[21] or the VARIABLE PARTIAL IDENTITY CONDITION (two segments must be identical with respect to one member of a given subset of features, but it doesn't matter which particular member it is).

(27) Unattested: COMPLETE NONIDENTITY CONDITION

$\forall\ F_i \in \mathbf{G}\ [(\alpha F_i)_1] \neq [(\beta F_i)_2]$

For some specified subset of the features, segment$_1$ and segment$_2$ have different values.

(28) Unattested: VARIABLE PARTIAL IDENTITY CONDITION

$\exists\ F_i \in \mathbf{G}$ such that $[(\alpha F_i)_1] = [(\beta F_i)_2]$

For some specified subset of the features, there is at least one feature for which segment$_1$ and segment$_2$ have the same value.

An example of the COMPLETE NONIDENTITY CONDITION would require that two segments have opposite values for, say, all place features, or even for *all* features. I know of no such case. For example, 'delete a vowel in the environment #CVC$_1$__C$_2$V if C$_1$ is [-anterior, -labial, +dorsal] and C$_2$ is [+anterior, +labial, -dorsal], or C$_1$ is [+anterior, -labial, +dorsal] and C$_2$ is [-anterior, +labial, -dorsal], *etc*'.

[21] That is 'complete' refers to all members of the given subset, not in general, to the whole feature set.

An example of the VARIABLE PARTIAL IDENTITY CONDITION would require that two segments have the same value for some feature in a given subset: 'delete a vowel in the environment #CVC$_1$__C$_2$V if C$_1$ and C$_2$ are both [αanterior], or [αlabial], or [αdorsal], *etc*'.

In other words, we can trivially construct a condition using the universal quantifier and a nonidentity relation, or the existential quantifier and an identity relation, but it seems that such conditions never are needed by the phonology. In section 10.2, I make a proposal concerning how this generalization should be treated by the theory.

10.1 Ambiguous cases

Obviously, there are rules for which the conditions on application *could* be expressed as an example of case (27) or (28), such as a rule requiring nonidentity for a single specific feature value. For example, Yiddish (Perlmutter 1988, Sapir 1915) has a rule that deletes the vowel in the plural suffix *-en* as long as the stem final consonant is not a nasal.[22] In terms of McCarthy's view of the OCP as a rule blocker, the rule fails to apply if it will bring two nasals into contact. Data appears in (29).

(29) Yiddish vowel deletion

	Sg	Pl
'language'	šprax	špraxn
'ear'	oyer	oyern
'magazine'	žurnál	žurnáln
'sea'	yam	yamen

Let's assume that the only relevant feature in conditioning the rule is [nasal], and that, for example, [sonorant] and [voice] are not part of the rule's structural description. Then, this rule *could* be stated in terms of a condition like (30), which is a highly restricted example of the COMPLETE NONIDENTITY CONDITION: the set of relevant features contains the single feature [nasal], and the rule has the added conditionn that the second consonant be [+nasal].

(30) Delete the vowel between an onset consonant C$_1$ and a nasal C$_2$
if $\forall$ F$_i$ $\in$ {[nasal]}[(αF$_i$)$_1$] $\neq$ [(βF$_i$)$_2$]

However, this condition can be expressed as well as an example of the NONIDENTITY CONDITION, as follows:

(31) Delete the vowel between an onset consonant C$_1$ and a nasal C$_2$
if $\exists$ F$_i$ $\in$ {[nasal]} such that [(αF$_i$)$_1$] $\neq$ [(βF$_i$)$_2$]

[22] The rule is also sensitive to syllable structure in a way that is irrelevant to the point under consideration, so I have given only examples of stems that end in a single consonant. Other data makes it clear that this is not a rule of epenthesis: *tate* 'father', *(dem) tatn* DATIVE.

So it is not *necessary* to combine the the universal quantifier with a nonidentity requirement to express the correct condition. Thus the claim that the COMPLETE NONIDENTITY CONDITION and the VARIABLE PARTIAL IDENTITY CONDITION are unattested in phonological rules as necessary kinds of conditions remains valid.

Other potential cases of the COMPLETE IDENTITY CONDITION arise when segments have redundant values for certain features, values that can be predicted on the basis of other features. Suppose a language $\mathcal{L}$ has in its consonant inventory only voiceless obstruents and voiced sonorants. Then a rule in $\mathcal{L}$ which apparently demanded nonidentity of voicing between two consonants could have the condition shown in (32):

(32) $\exists$ $F_i \in$ {[voice]} such that $[(\alpha F_i)_1] \neq [(\beta F_i)_2]$
that is, $[(\alpha \text{voice})_1] \neq [(\beta \text{voice})_2]$

Alternatively, the correct (though extensionally equivalent) form of the condition might be one that referred to the feature [sonorant], as in (33)

(33) $\exists$ $F_i \in$ {[sonorant]} such that $[(\alpha F_i)_1] \neq [(\beta F_i)_2]$
that is, $[(\alpha \text{sonorant})_1] \neq [(\beta \text{sonorant})_2]$

Finally, the COMPLETE IDENTITY CONDITION in (34) is also extensionally equivalent to the two versions of the NONIDENTITY CONDITION just listed.

(34) $\forall$ $F_i \in$ {[sonorant], [voice]}$[(\alpha F_i)_1] \neq [(\beta F_i)_2]$
that is, $[(\gamma \text{sonorant})_1] \neq [(\delta \text{sonorant})_2]$ AND $[(\epsilon \text{voice})_1] \neq [(\zeta \text{voice})_2]$

Again, since we do not need to use the formulation in (34), and since we have reasons to believe that we never need to use such conditions, we have a tool for choosing among extensionally equivalent grammars. The phonology of $\mathcal{L}$ potentially has the condition in (32) or the one in (33), but definitely not the one in (34). This is the kind of argument suggested by Chomsky (1986) in refuting the claim of Quine (1972) that it is futile to attempt to choose among extensionally equivalent grammars: "Because evidence from Japanese can evidently bear on the correctness of a theory of S_0, it can have indirect—but very powerful– bearing on the choice of the grammar that attempts to characterize the I-language attained by a speaker of English" (Chomsky 1986:38).

10.2 Towards an explanation for gaps in attestation

Assuming that the COMPLETE NONIDENTITY CONDITION and the VARIABLE PARTIAL IDENTITY CONDITION are really absent from human languages, how are we to treat this fact? One way to do so is to build the fact into UG as an explicit constraint against quantificational statements conforming to certain formats. This strikes me as the wrong way to approach the issue, if it merely consists of restating the descriptive observation as a principle of grammar and not being open to explanations outside of the realm of grammar. In this particular case, it may be possible to derive the gap, from the relationship between language change and phonetics. Note that this approach is in no way incompatible with a

nativist perspective—the nativist position is just that some (not necessarily *all*) non-trivial aspects of the language faculty are innate.

Following work on the nature of sound change (Ohala 1990, Hale, forthcoming) and theoretical work in cognitive science (Pylyshyn 1984), Hale & Reiss (2000ab) argue that it is to be expected that attested patterns in the phonological systems of the world's languages reflect only a subset of what is computationally possible for the human phonological capacity.[23] In other words, all attested patterns must be generatable by the UG-given phonological capacity, but not all generatable patterns will arise, due to the nature of sound change and language acquisition. This point of view may be helpful in explaining why the COMPLETE NONIDENTITY CONDITION and the VARIABLE PARTIAL IDENTITY CONDITION are unattested.

In general, phonological processes arise diachronically from the reanalysis of sublinguistic (gradient) phenomena as grammatical (categorical, feature-based) phenomena. Now note that there is at least a partial correspondence between phonetics and phonology—for example, features referring to place of articulation tend to correspond to the nature of formant transitions between vowels and consonants. Work on feature detectors and the like, though far from complete, reflects the belief that we can study the nature of the tranduction processes between phonetics (gradient phenomena) and feature-based phonology.[24] An identity condition defined over a subset of phonological features, therefore, will tend to be to some extent related to a 'natural class' of phonetic properties.

Similarly, a non-identity condition on, say, segments in the environment of a rule implies identity of the segments in those environments in which the rule does *not* apply. In other words, nonidentity entails the existence of identity in the complement set of environments. From the phonetic/diachronic perspective, then, these two conditions are the same. These two types of condition depend, at least at the point when they are phonologized by a learner, on clusters of phonetic properties. However, from a synchronic, phonological perspective, they are computationally distinct—one requires the equivalent of universal quantification and identity; the other, existential quantification and non-identity.

In contrast to these two cases, it is hard to imagine how either the COMPLETE NONIDENTITY CONDITION or the VARIABLE PARTIAL IDENTITY CONDITION could be derived from definable clusters of shared phonetic properties. For example, there is no phonetic unity to be found between segment transitions involving [± voiced] agreement and those involving [±coronal] agreement. But this is the kind of phonetic phenomena required to give rise to the VARIABLE PARTIAL IDENTITY CONDITION which requires agreement for an arbitrary feature among segments in a structural description.

It is also hard to imagine how having opposite values for a given set of features, as

[23]This point is too obvious to be credited to Hale & Reiss or anyone else, for that matter. However, it seems to be ignored in OT arguments which suppose that the factorial typology, the set of possible ranking of a constraint set, should reflect *attested* languages. Obviously, the factorial typology must generate all attested patters, but it is clear that some may not be attested, for reasons that have nothing to do with phonology.

[24]See Harnad 1987 for critical discussion of feature detector theory. Current work that attempts to incorporate phonetic description (acoustic parameters, trajectory of articulators, *etc.*) into the phonology represents an overly naive approach—one that essentially equates physical partameters with representational constants.

required by the COMPLETE NONIDENTITY CONDITION, could lead to a phonetically stable pattern. Recall that the relevant case would allow *every* pairing (of members of the relevant set **G**) of opposite feature values. Restricting the relevant set to the members A and B, each line of (35) would instantiate an environment for rule application under a COMPLETE NONIDENTITY CONDITION:

(35) Permutations of feature specification in the COMPLETE NONIDENTITY CONDITION

[+A +B] *vs.* [−A −B]
[−A +B] *vs.* [+A −B]
[+A −B] *vs.* [−A +B]
[−A +B] *vs.* [+A −B]

It is hard to imagine how such sets of representations could correspond to a phonetically natural grouping.

To summarize, the NONIDENTITY CONDITION and the IDENTITY CONDITION provide us with a lower limit on the computational resources of UG, whereas patterns of attestation reflect extragrammatical factors. I am not claiming that these conditions are in principle uncomputable by the phonological component of the mind, but rather that the nature of language transmission makes it unlikely, or perhaps even impossible, that they will arise.

It is beyond the scope of this paper to provide a diachronic phonetic analysis of how, say, antigemination and antiantigemination arise, but we can indicate some directions for future research. Odden (1988:470) makes some suggestions related to the fact that the consonant closure gestures for non-identical places of articulation can overlap in time, since they involve different articulators, whereas repeated, identical consonants cannot have overlapping gestures. We can imagine that vowels between identical consonants will be somewhat longer, and thus less susceptible to deletion diachronically than vowels between nonidentical consonants. If only the phonetically shorter vowels are deleted, the resulting grammar will manifest antigemination.

Antiantigemination could perhaps result from a sequence of reanalyses. First, consonant-to-consonant place transitions are reanalyzed as epenthetic vowels. The resulting grammar thus has epenthetic vowels between nonidentical consonants only. Next, *epenthesis* between nonidentical consonants is reanalyzed as *syncope* between identical consonants.

The first stage of this scenario results in a type (9d) condition, insertion between (partially or fully) non-identical consonants. Such rules appear to have arisen in various Oceanic languages such as Marshallese (Mark Hale, p.c.), Mokilese (Harrison 1976), Ulithian (Sohn & Bender 1973).[25]

[25]It is beyond the scope of this paper to provide full analyses of these processes. It is clear, however, that they involve complex chains of diachronic events. Here are some typical descriptions:

- "Consonant clusters are often broken up by the insertion of a vowel ... An excrescent vowel is never inserted betweeen identical consonants, as in *kodda* above. Clusters consisting of a [nasal and a stop, a liquid and a stop, or a fricative and a stop] are not broken up by an execrescent vowel when the two consonants have the same place of articulation. ... [W]here the two consonants have different points of articulation ... excrescent vowel insertion appears to be optional" (Harrison 1976:42-3)

Once we accept the existence of such epenthesis rules and the possibility of rule inversion we see that it is not surprising that antiantigemination (5a), deletion between (partially or fully) identical consonants, can arise diachronically.

11 Rose's (2000) account of antigemination

It may be a matter of faith whether the data cited by Odden demonstrates that the OCP is ill-founded as a principle of grammar or if, instead, it shows that OT constraint violability is central to an understanding of grammar. I have adopted the first alternative. Rose (2000) implicitly adopts the second, but I will now argue that her model is actually more unwieldy than indicated by the examples she discusses.

Rose achieves an analysis of antigemination by positing a constraint against gemination, No-Gem, and a constraint called OCP, which is violated by 'adjacent' (see below) identical segments. Actually, she posits a segmental OCP constraint (p.102) and a family of OCP constraints for different features, such as OCP/Coronal (p.97). She also makes the following crucial, though non-standard, assumptions:

(36) Assumptions of Rose (2000)

a. Consonant adjacency: Two consonants in sequence are adjacent irrespective of intervening vowels (p.95).

b. A surface sequence C_iVC_i violates the OCP under consonant adjacency (p.101).

c. Any surface C_iC_i sequence in a given domain is a geminate and does not violate the OCP (p.101).

Antigemination, the failure to delete a vowel which otherwise should not surface between identical segments, derives from a ranking in which No-Gem outranks the relevant OCP constraint. The opposite ranking leads to deletion, since not deleting maintains an OCP violation—by (36a) identical consonants separated by vowels are adjacent.

Rose does not provide much of a model of phonological representation. For example, she does not tell us whether she is using binary or privative features, or if she is assuming some kind of feature geometric representation. This vagueness makes it somewhat difficult to state exactly what her model predicts in cases of partial identity conditions. However, it appears that we would have to endow her model with a much greater number of constraints than she discusses.

- "Clusters are allowed in medial position ... in which case an excrescent vowel optionally intervenes if the members of a cluster are not in the same position of articulation and if the first consonant is not one of **l**, **n** and **g** [the velar nasal—cr]" Sohn & Bender (1973:38).

- "In order to maintain the phonetic and structural equilibrium, such forces as compensatory lengthening, excrescent vowel insertion, vowel reduction, etc. are constantly in operation. The above rule deals with vowel reduction ... When the single vowel to be reduced is preceded by nasal or **l**, the reduction seems almost complete. In the case of non-nasal and non-**l**, the reduction, which is incomplete [to the high central vowel [ɨ]—cr] is applicable only where the neighbouring vowels are dissimilar" Sohn & Bender (1973:67).

Consider the case of epenthetic vowels that appear between English coronal stridents, mentioned above. Note that this epenthesis cannot be driven in an OT model merely by adjacent coronal features. Following Rose's approach, the epenthetic vowel in, say, *bushes* would be due to the avoidance of a *partial* geminate consisting of string adjacent segments with linked nodes for the features which define coronal stridents—[+continuant, -sonorant, +coronal, +strident]. So we need a version of No-Gem that is violated by such partial feature sharing.

This cannot be accomplished merely by positing individual constraints for each feature, since epenthesis does not occur to avoid partial gemination with respect to certain subsets of these features. A form like *cliffs* has (in Rose's model) [+continuant] shared between the [f] and the [s]. A form like *bins* has a shared [+coronal] specification on the [n] and the [s], and so on. In other words, we cannot appeal to a constraint No-Gem-Coronal, banning linked coronal nodes, because this feature is relevant in driving epenthesis *only in the context of the other features that define coronal stridents.*

Therefore, Rose's model requires a separate No-Gem-type constraint for each combination of features that can group together in the structural description of a phonological pattern. As she acknowledges, she also needs separate OCP constraints, such as OCP/Coronal, for each feature.

In OT, such constraints are part of UG. My proposal instead endows UG with the necessary apparatus to construct rules (or constraints) in accordance with input from the target language. Again, I cite Odden (1988:461): "It is misguided to attribute every accidentally true statement about human language to UG" (461).

The simplicity of the theoretical claims I am making can perhaps be appreciated with the following paraphrase: Phonological processes/alternations are context sensitive. One kind of information that can be used to define context is feature identity.

12 A difference between phonology and syntax

It seems clear that syntactic/semantic interpretation makes use of the existential and universal quantifiers—these are standard primitives of LF represesentational apparatus. In the following, I point out that that the apparent gap of type Complete Nonidentity (27) and Variable Partial Identity (28) conditions in phonological rules is not paralleled in syntax.

The nonidentity condition, requiring that a set of features contain some differences, might be abbreviated thus (where k refers to a particular attribute or feature and x_k refers to the value that k has on segment x):

(37) $\exists k\ x_k \neq y_k$

But, of course, this is equivalent to the following

(38) $\neg\forall k\ x_k = y_k$

Similarly, the identity condition can be abbreviated as follows:

(39) $\forall k\ x_k = y_k$

Of course this is equivalent to the following:

(40) $\neg\exists k\ x_k \neq y_k$

The two apparently unattested conditions can thus, also be given in two forms each. First is the COMPLETE NONIDENTITY CONDITION:

(41) i. $\forall k\ x_k \neq y_k$
ii. $\neg\exists k\ x_k = y_k$

And the VARIABLE PARTIAL IDENTITY CONDITION can be represented thus:

(42) i. $\exists k\ x_k = y_k$
ii. $\neg\forall k\ x_k \neq y_k$

Since each condition has two logically equivalent formulations, I will simplify the exposition by not using the negations of the quantifiers. So, this leaves us with the two quantifiers, used to quantify over the set of features, and the relation "=" and the negation of this relation. Now, let us rewrite the four conditions, replacing "=" with "R", to stand for an arbitrary relation:

(43) Four types of condition

A. $\exists k\ x_k \neg R y_k$
'Under some condition k, x is not in a given relationship with y.'

B. $\forall k\ x_k R y_k$
'Under all conditions k, x is in a given relationship with y.'

C. $\forall k\ x_k \neg R y_k$
'Under all conditions k, x is not in a given relationship with y.'

D. $\exists k\ x_k R y_k$
'Under some condition k, x is in a given relationship with y.'

A question to ask is whether conditions C (which encompasses the COMPLETE NONIDENTITY CONDITION) and D (which encompasses the VARIABLE PARTIAL IDENTITY CONDITION) are *ever* used by the language faculty, since they appear not to be used by the phonology. The answer seems to be that they, or their parallels, are used in the interpretation of binding relations.

Suppose we let the relation R be the *c-command* relation. So xRy means 'x c-commands y' and $x\neg Ry$ means 'x does not c-command y'. Epstein, et al. (1998:62) assume the following interpretive procedure in their derivational theory of syntax:

> The application of "disjoint" interpretive procedures occurs at every point in the derivation, whereas the application of "anaphoric" interpretive procedures occurs at any single point in the derivation.

Abstracting away from locality conditions, we can rephrase this generalization as follows.[26]

- A pronoun P is disjoint in reference from a category X if for all points in the derivation, X does not c-command P. (If X_i is category X at point i, then P is disjoint from X if $\forall i\ X_i \neg RP$—compare the COMPLETE NONIDENTITY CONDITION.)
- An anaphor A is anaphoric with a category X (whose features are compatible with A) if there is a point in the derivation where X c-commands A. (If X_i is category X at point i, then A is coreferential with X if $\exists i\ X_i RA$—compare the VARIABLE PARTIAL IDENTITY CONDITION.)

In other words, if we quantify over steps in a derivation (replacing the k in our phonological examples with the i in the binding examples, and replacing phonological identity with binding, we see that the interpretation of binding relations uses exactly those combinations of the quantifiers and a negated and un-negated relation that the phonology does not use. Syntactic operations like feature checking clearly involve the evaluation of identity relations—the features of a functional head only check *identical* features of lexical categories— so then there are types of conditions that are shared by syntax and phonology. Thus there is no simple complementarity between the types of conditions the two components make use of, nor is there any reason to expect there should be.[27]

13 Constraints alone *vs.* Rules & Constraints *vs.* Rules alone

A reader may have been convinced to accept the necessity for the additional power granted to the representational component argued for here—the necessity of quantification—without accepting rejection of constraints. The formulation of constraints that can evaluate identity and nonidentity would also require the use of quantification. Therefore, constraints on their own, or constraints in conjunction with rules do not vitiate the need for quantificational statements in grammars.

Consider, however, what we gain by adopting a minimalist approach to characterizing the phonological component in terms of rules: we have a rule component which allows the use of quantificational statements; we have no notion of wellformedness or illformedness—the phonology maps inputs to outputs. In the following table I compare three approaches to building a phonology, under the assumption that they are all empirically nondistinct, that

[26] It must be noted that this discussion slightly simplifies that of Epstein, *et al.*, since for them c-command is also derivable from the nature of *Merge*.

[27] At least one other explanation is available, namely that the binding conditions suggested are not, in fact, those of natural language, and that their dependence on the kinds of conditions that are unattested in the phonology is further evidence of this fact.

is, that they can generate the same sets of output. The Just Rules (JR) approach outlined in this paper is compared to 'standard' OT and a generic Rules & Constraints (RC) model.

(44) Comparison of various approaches to phonology

	OT	RC	JR
a. List of Primitive Entities	yes	yes	yes
b. List of Possible Operations/Functions	yes	yes	yes
c. List of Constraints	yes	yes	no
d. Notion of Illformedness	yes	yes	no
e. Notion of Repair	no	yes	no
f. Quantifiers in SDs	yes	yes	yes
g. Representational Matching Procedure	yes	yes	yes

A complete formal theory of phonology must specify what it can generate, so it is necessary to define the universe of discourse by listing the entities (a) and operations (b) that the computations have access to. In OT there are no rules, but as discussed above, a fully explicit version of OT will have to provide a finite characterization of what GEN actually does—a list of possible operations on representations is in fact a necessary part of the model. In addition, OT contains other functions, such as EVAL, so all three theories contain functions. The three models cannot be distinguished on these grounds.

Obviously, there are constraints (c) in OT and RC models, and there are none in JR. As Yip explains, the use of constraints presupposes a notion of illformedness (d), which I have argued is circular at best, and incoherent at worst, as an explanation of phonological alternation. The constraints are posited on the basis of this intuited sense of wellformedness *vs.* illformedness or markedness. This notion does not exist in the JR model, in which a set of rules maps phonological inputs to outputs.

OT does not prescribe a specific repair (e) for individual markedness violations, but conceives of the grammar as finding an optimal solution across all outputs, which emerges from the ranking. In RC, rules are applied to repair illformed structures or to block rule application, thus also appealing to markedness theory. Repair is not part of JR theory.

In all three theories, quantifiers (f) are necessary to evaluate the SDs of rules or constraints which refer to identity and nonidentity. Similarly, all three theories need some kind of Representational Mapping Procedure to determine which representations satisfy the structural description of its rules or constraints.

Recall that we are assuming that we can compare extensionally equivalent grammars. While straightforward theory comparison is difficult, the 'rules only' approach appears to be the most elegant. The list of possible operations is stated in positive terms and thus characterizes the universe of discourse with no additional apparatus. There is no notion of markedness, and thus no reason to conceive of rules as repairing representations. The theory requires rules with a sufficiently rich representational apparatus to define their condition of application. However, as exemplified by the discussion of quantification, this apparatus may be needed by any empirically adequate theory.

13.1 Structural descriptions are constraints on application

Let's look back to the type of rule discussed by McCarthy to motivate the restriction of rule application by the OCP. Notice that blocking of a rule R can be achieved in one of two ways—either by applying R and undoing its effects if they are 'undesirable', or by 'looking ahead' to see what the output would be before applying R, and not applying R if the projected output is undesirable. There is, however, a simpler way of avoiding rule outputs that result in ungrammatical surface forms: reformulate the rule as R', so as to apply only when it should. We have said this much already, however, it is important to realize that the structural description of a rule, the representation that determines whether the rule applies via the representational matching procedure discussed in section 3.1, is nothing other than a constraint on application. McCarthy's rule of vowel syncope in Hebrew applies to vowels between consonants, not to any segment that is between any other two segments. The rule applies only under certain metrical conditions, not under others. The condition that the flanking consonants be non-identical, is thus of the same type as the other constraints on application, the other components of the rule's structural description. In other words, there is no motivation in a rule-based grammar that uses a RMP to also have constraints that are not just part of the structural description of rules.

Analogies may again be useful. There is no reason to assume that a law of Newtonian physics, $f = ma$, that refers to entities like *force*, *mass* and *acceleration* is actually better seen as a relation between variables $x = yz$, which is constrained by a constraint system that rules out any possible instantiation of $x = yz$ other than $f = ma$. Similarly, a rule or law includes a specification of when it is applicable. Writing highly general rules that lack appropriate structural descriptions to restrict sufficiently when the rules actually apply, and then positing constraints that limit the applicability of a rule seems unproductive.

13.2 A historical irony

There is a certain historical irony to be noted here. It has been claimed that the demise of rule-based phonology was due to a failure to formulate a theory of possible rules. Recall that Yip claimed that OCP 'effects' appear over and over again in phonological rules. She concludes that such conditions should therefore *not* be stated in structural descriptions. It should now be apparent that the opposite conclusion was warranted: by observing what type of conditions (for example, the Identity and Non-Identity Conditions) appear in structural descriptions, we approach a theory of what is a possible rule! In other words, appeal to constraints not only complicates the theory of grammar unnecessarily, since the RMP used in the notion of structural description already provides the computational power that additional constraints were meant to supply, but also undermines the investigation of the most pressing question in rule-based phonological theory: 'What is a possible rule?'.

13.3 Conspiracies

Before proceeding we must dispense with the oft-cited claim that the existence of 'conspiracies' motivates the use of constraints in phonological theory: for example, the OT literature

is rife with claims of OT's superiority at accounting for conspiracies: "One of the principal reasons that rule-based theory has come under attack is that it offers no satisfactory explanation for conspiracies" (Kager 1997:463). However, Kiparsky (1973) has shown, that generative phonology does not need the notion of conspiracy. Here is my interpretation of Kiparsky's argument.

(45) The epiphenomenality of conspiracies (based on Kiparsky 1973b:75ff)

i. A conspiracy is a set of rules that are "functionally related", that is they lead to the same kinds of output configurations, such as 'all syllables are open'.

ii. If a language has such a set of rules, then the rules of the language will tend to be surface true (transparent).

iii. Non-transparent (opaque) rules are not surface true.

iv. Rules that are not surface true are hard for a learner to learn.

v. Things that are hard to learn are more likely *not* to be learned than things which are easy to learn.

vi. Failure to learn aspects of the ambient language constitutes a diachronic change.

vii. Therefore, (E-)languages are more likely to lose opacity than gain opacity.

viii. Therefore, grammars are likely to look like they have conspiracies.

In other words, the existence of conspiracies is an epiphenomenon due to the fact that languages tend to have transparent rules. This in turn is an epiphenomenon derived from the undeniable fact that individual languages must be learned.

Kiparsky's explanation of conspiracies depends on the fact that acquisition can be *unsuccessful*, resulting in so-called language change (Hale, to appear). In other words, tendencies such as 'conspiracies' are to be explained by reference to diachronic linguistics where the goal is to define possible changes and to explain why certain changes are more or less likely to occur.

14 Conclusions

I have argued that an algebraic formulation of phonological representation facilitates the incorporation of quantificational logic into structural descriptions. I have shown that well known cases of NONIDENTITY CONDITIONS demonstrate the necessity of the existential quantifier (or its equivalent) for phonology. The cases of IDENTITY CONDITIONS that refer to arbitrary sets of features, sets that are not members of a class according to Feature Geometric models, support the claim that the universal quantifier, or its equivalent is also necessary. Since the Feature Algebraic notation is more powerful than Geometric notation, and since the Algebraic notation seems necessary, the Geometric notation can be dispensed

with. This is not surprising since the original motivation for Feature Geometry now seems to have been somewhat misguided (see below). I have also argued that certain kinds of logically possible conditions, the VARIABLE PARTIAL IDENTITY CONDITION and the COMPLETE NON-IDENTITY CONDITION, though potentially paralleled in other domains, including syntax, appear not to be needed in the phonology.

If correct, the generalization about phonology using only conditions of type (43A & B) may, as suggested, be derivable from principles outside the realm of grammar, and not reflect any real constraint on the nature of phonological computation. In any case, we can posit the IDENTITY CONDITION and the NONIDENTITY CONDITION as *necessarily formulatable* by phonological *UG*.

The negative claim, that the COMPLETE NONIDENTITY CONDITION and the VARIABLE PARTIAL IDENTITY CONDITION are not possible conditions on phonological rules, need not be made if the set of possible conditions is positively specified. This might seem unsatisfying, since it requires positively characterizing the partially overlapping condition types independently for phonology and syntax. However, the necessity of doing so (if we reject the phonetic account sketched in section 10.2) just reflects their status as separate modules of grammar—the set of possible conditions on phonological rule application and the set of possible conditions of pronouns and anaphor interpretation must be specified separately, just as the set of primitive elements used by the phonology and the syntax must be specified separately.

It is useful again to make an analogy to see that characterizing *UG* in terms of constraints on possible grammars, instead of in positive terms, is potentially misguided. This will help to relate the second half of this paper to the first half. When a physicist claims that there are, say, five types of fundamental particle, s/he is not explicitly claiming that no others exist—it is impossible to know everything that exists (inductive uncertainty again). What is being claimed is that all known phenomena (within the relevant domain) can be explained using these five particle types, and so there is no reason to posit any others. Similarly, we can now propose the hypothesis that all conditions on phonological rule application (of the narrow type considered here) are cases of the IDENTITY CONDITION or the NONIDENTITY CONDITION, but we need not posit a constraint that prohibits the other types of condition which were discussed, but claimed to be unattested. First of all, we do not know that they are, in fact, impossible conditions. Second, they are members of an infinite set of unattested conditions. It should be satisfying enough to get a handle on what we know *UG* can do, what its minimally necessary formal properties are, without worrying about what it can't. We need to posit such constraints only to the extent that we need to posit NOBANANA. Obviously, that is no extent at all.

This philosophical argument is bolstered by the empirical arguments given in the paper. These can be summarized as follows. The invocation of universal constraints depends upon a notion of relative ill-formedness or markedness. Such a notion cannot be justified empirically. There are rules that seem to be blocked if their output would violate the OCP, as well as those that seems to be blocked only if their output would *not* violate the OCP, so there is no reason to grant primacy to one type over the other. So without markedness, universal constraints are unjustified. Language specific constraints are unnecessary, since their effects can be captured by a more precise formulation of rules.

It is worthwhile to compare the approach proposed here to that presented in an influential pre-OT paper by a phonologist who is one of the most important contributors to the success of OT. McCarthy (1988:84), in an exposition of Feature Geometry, states that "The goal of phonology is the construction of a theory in which cross-linguistically common and well-established processes emerge from very simple combinations of the descriptive parameters of the model". For example, "Assimilation is a common process because it is accomplished by an elementary operation of the theory—addition of an association line" (86). After attempting to motivate two operations and two constraints on well-formedness, McCarthy declares that "each operation and constraint is predicted to operate on each class node of the feature geometry in some reasonably well-attested linguistic phenomenon" (90).[28] The vagueness of terms like *common, well-established* and *reasonably well-attested* should alert us to the lack of rigor inherent in such an approach. A simpler, more explicit approach is to figure out what is the minimum amount of representational and computational machinery needed to generate attested patterns. Rather than seeing this as an original suggestion, it strikes me as "the natural approach: to abstract from the welter of descriptive complexity certain general principles governing computation that would allow the rules of a particular language to be given in very simple forms, with restricted variety" (Chomsky 2000:122).

With this goal in mind, phonology should not return to the rules-and-constraints models that predate Optimality Theory, but to a pure rule-based formalism. The nature of the types of rules needed by phonological theory thus becomes an empirical question that promises to yield answers if not prejudiced by preconceived notions of what rules 'should' look like.

Instead of the taxonomic generalizations offered by spurious markedness-based theories like OT, the approach advocated here will offer deeper insight into the nature of phonological computation. Such insight is the goal of cognitive science in general:

> "[I]f we confine ourselves to the scientific and intellectual goals of understanding psychological phenomena [as opposed to predicting observed behavior–cr] one could certainly make a good case for the claim that there is a need to direct our attention away from superficial "data fitting" models toward deeper structural theories" [Pylyshyn 1973:48].

As discussed in section 10.2, explaining the actual corpus of attested data may require an understanding of the interaction of phonology, phonetics and language change.

A final philosophical issue arises from the results of this paper. Given that the use of quantification and identity are not specific to the language faculty, are we justified in labeling them as part of the language faculty, or as general cognitive mechanisms that the language faculty has access to?[29] Perhaps, the correct answer is that the use of quantification in logic and mathematics are somehow secondary manifestations, relatively recent in human history, of its use by the language faculty. Chomsky (2000: 3-4, most recently) has suggested that

[28]The following sentence is much closer to a coherent proposal: "In other words, we should be able to freely combine the predicates of our theory of representations and our theory of operations and constraints and, in each case, come up with some real rule that languages have." See, however, Hale & Reiss (2000ab) for arguments that the set of actually attested languages is expected to be only a subset of the set of computationally possible human languages allowed by UG.

[29]See Kempson (1986) for related discussion.

another property of the language faculty, the property of discrete infinity, is another such case. The answer to this question is probably not relevant for the purposes of phonological theorizing and modeling, but, consideration of such issues is necessary for an understanding of phonology in the context of cognitive science.

References

Archangeli, Diana. 1997. Optimality Theory: An introduction to linguistics in the 1990s. In Diana Archangeli and D. Terence Langendoen (eds.), *Optimality Theory: An Overview*. Oxford: Blackwell: 1-32.

Archangeli, Diana & Douglas Pulleyblank. 1994. *Grounded Phonology*. Cambridge, MA: MIT Press.

Bhatia, Tej & Michael Kenstowicz. 1972. Nasalization in Hindhi: A reconsideration. *Papers in Linguistics* 5:202-212.

Bliese, L.F. 1981. *A generative grammar of Afar.* Summer Institute of Linguistics, University of Texas, Arlington.

Chomsky, Noam. 2000. *New Horizons in the Study of Language and Mind.* Cambridge Universty Press.

Chomsky, Noam. 1986. *Knowledge of Language*. Westport, CT: Praeger.

Engstrand, Olle. 1997. Areal biases in stop paradigms. Papers from *Fonetik 97, The Ninth Swedish Phonetics Conference*, held in Umeå, May 28-30, 1997. Reports from the Department of Phonetics, Ume University (PHONUM), 4, 187-190.

Epstein, S., E. Groat, R. Kawashima & H. Kitahara.1998. *A Derivational Approach to Syntactic Relations.* Oxford: Oxford University Press.

Hale, Mark. Forthcoming. *Theory and Method in Historical Linguistics.* Cambridge: Blackwell.

Hale, M. 2000. Phonetics, Phonology and Marshallese Vowels. *The Linguistic Review 17.*

Hale, M. & C. Reiss. 2000a. Substance abuse and dysfunctionalism: Current trends in phonology. *Linguistic Inquiry* 31:157-169.

Hale, M. & C. Reiss. 2000b. Phonology as cognition. To appear in N. Burton-Roberts, Philip Carr & Gerry Docherty (eds.), *Phonological Knowledge*. OUP: Oxford.

Hale, M. and C. Reiss. 1999. The subset principle in phonology: Why the *tabula* can't be *rasa*. Submitted to *Journal of Linguistics*.

Hale, M. & C. Reiss. 1998. Formal and empirical arguments concerning phonological acquisition. *Linguistic Inquiry* 29: 656-683.

Halle, Morris. 1975. Confessio Grammatici. *Language* 51:525-35.

Halle, M. & W. Idsardi. 1995. Stress and metrical structure. In J. Goldsmith (ed.), *Handbook of Phonological Theory*. Oxford: Blackwell.

Halle, M. & J.R. Vergnaud. *An essay on stress.* Cambridge, MA: MIT Press.

Halle, M., Bert Vaux, & Andrew Wolfe. 2000. On feature spreading and the representation of place of articulation. *Linguistic Inquiry* 31:387-444.

Harnad, Stevan (ed.) 1987. *Categorical perception : the groundwork of cognition.* Cambridge: Cambridge University Press.

Harrison, Sheldon, with the assistance of Salich Albert. 1976. *Mokilese Reference Grammar.* Honolulu: University Press of Hawaii.

Jensen, J.T. 1977. *Yapese Reference Grammar.* Honolulu: University Press of Hawaii.

Kager, R. 1997. Rhythmic vowel deletion in OT. In Iggy Roca (ed.), *Derivations and Cconstraints in Phonology.* Oxford: Clarendon.

Keer, E. 1999. Geminates, the OCP and the Nature of CON. PhD thesis. Rutgers University. New Brunswick, New Jersey.

Kempson, R. 1988. On the grammar-cognition interface. In R. Kempson (ed.), *Mental representations : the interface between language and reality* Cambridge ; New York : Cambridge University Press, 1988.

Kiparsky, P. 1973. Phonological Representations? In O. Fujimura (ed.), *Three Dimensions of Linguistic Theory.* Tokyo: The TEC Corporation.

Marlett, S. & J. Stemberger. 1983. Empty consonants in Seri. *Linguistic Inquiry* 14: 617-639.

Local, J. & J. Coleman. 1994. The line crossing constraint. *Linguistics and Philosophy*
Lynch, J. 1978. *A gramar of Lenakel.* Pacific Linguistics Series B #55. Australian National University, Canberra.

McCarthy, J. 1988. Feature geometry and dependency: a review, *Phonetica* 45:84-108.

McCarthy, J. 1986. OCP effects: Gemination and antigemination. *Linguistic Inquiry* 17:207-263.

Mohanan, K. P. 2000. The theoretical substance of optimality formalism. *The Linguistic Review* 17: 143-166.

Ni Chiosáin, M. & Jaye Padgett. 1993. Inherent V-Place. UC Santa Cruz Working Papers. Linguistics Research Center, University of California, Santa Cruz.

Odden, D. 1988. Antiantigemination and the OCP. *Linguistic Inquiry* 19:451-475.

Odden, D. 1986. On the Obligatory Contour Principle. *Language* 62:353-383.

Ohala, J. 1990. The phonetics and phonology of aspects of assimilation. In J. Kingston and M. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech.* Cambridge: Cambridge University Press.

Paradis, Carole. 1988. On constraints and repair strategies. *LR* 6: 71-97.

Perlmutter, David. 1988. The split morphology hypothesis: evidence from Yiddish. In Michael Hammond and Michael Noonan, (eds.), *Theoretical Morphology.* San Diego: Academic Press.

Prince, A. & P. Smolensky. 1993. *Optimality Theory: Constraint Interaction in Generative Grammar.* Technical Report RUCCS, Rutgers University, New Brunswick, N.J.

Pylyshyn, Z. 1984. *Computation and cognition: toward a foundation for cognitive science.* MIT Press, Cambridge, MA.

Pylyshyn, Z. 1973. On the role of competence theories in cognitive psychology. *Journal of Psycholinguistic Research.*

Quine, Willard V. 1972. Methodological reflections on current linguistic theory. In G. Harman and D. Davidson (eds.). *Semantics of Natural Language.* New York: Humanities Press.

Reiss, C. 2000. Optimality Theory from a Cognitive Science Perspective. *The Linguistic Review 17*:291-301 (2000).

Reiss, C. 1999. Acquisition and Post-OT Phonology. To appear in *Phonology 2000.* B. Vaux & M. Halle, eds.

Reiss, C. 1995. *A theory of assimilation, with special reference to Old Icelandic phonology*, Ph.D. Dissertation, Department of Linguistics, Harvard University, June, 1995.

Rennison, J. 2000. OT and TO. *The Linguistic Review* 17: 135-141.

Rose, Sharon. 2000. Rethinking Geminates, Long-Distance Geminates, and the OCP. *Linguistic Inquiry* 31:85-122.

Sapir, E. (1915) Notes on Judeo-German phonology. *The Jewish Quarterly Review*, n.s., 6, 231-266. Reprinted in *Selected Writings of Edward Sapir in Language, Culture, and Personality.* David G. Mandelbaum (ed.). Berkeley: University of California Press.

Sherwood, David. 1986. *Maliseet-Passamaquoddy Verb Morphology.* Ottawa, Ontario: National Museums of Canada.

Smolensky, P. On the comprehension/production dilemma in child language. *Linguistic Inquiry* 27.

Sohn, Ho-min & Byron W. Bender. 1973. *A Ulithian Grammar.* Pacific Linguistics, Series C—No. 27. Linguistic Circle of Canberra.

Taylor, S. 1969. *Koya: An outline grammar, Gommu dialect.* University of Caliornia Publications in linguistics 54. University of California Press: Berkeley and Los Angeles.

Yip, M. 1988. The Obligatory Contour Principle and Phonological Rules: A Loss of Identity. *Linguistic Inquiry* 19: 65-100.