University of Pennsylvania Working Papers in Linguistics

Volume 16
Issue 1 *Proceedings of the 33rd Annual Penn Linguistics Colloquium*

Article 10

1-1-2010

Comparing Incomparable Frameworks: A Model Theoretic Approach to Phonology

Thomas Graf
University of California, Los Angeles, tgraf@ucla.edu

Comparing Incomparable Frameworks: A Model Theoretic Approach to Phonology

Abstract
In previous work, we used techniques from mathematical logic and model theory to study and compare two phonological theories, SPE and Government Phonology. The surprising result was that Government Phonology corresponds to a very weak fragment of SPE, yet it can attain the full expressivity of the latter through more powerful mechanisms of feature spreading. An issue that we didn't elaborate on, however, is the question of what this increase in expressivity buys us in terms of empirical coverage, which we pick up in this paper. Again making good use of our model theoretic techniques, we investigate two phonological phenomena --- Sanskrit n-retroflexion and primary stress assignment in Creek and Cairene Arabic --- and show how much power feature spreading has to be granted in any descriptively adequate account which does not invoke additional technical machinery. These technical results are accompanied by reflections on the relation between empirically minded theory comparisons and the model theoretic approach.

Comparing Incomparable Frameworks: A Model Theoretic Approach to Phonology

Thomas Graf

1 Introduction

Theory comparisons have always enjoyed a prominent role in modern linguistics. Surprisingly, though, almost all comparative studies base their results exclusively on the meticulous analysis of empirical data, which is a laborious process. So-called model theoretic approaches have recently emerged as an intriguing alternative for specific problems (Rogers 1998, Potts and Pullum 2002, Kracht 2003). They involve less analytical toiling and prove results that could not be derived from empirically grounded work alone. Moreover, they unearth the implicit cognitive claims of the theory under consideration, thus strengthening the bonds between competence-focused theorizing and psycholinguistic research.

In Graf (2009), I used these mathematical techniques to study and compare two phonological theories, SPE (Chomsky and Halle 1968) and Government Phonology (Kaye et al. 1985, 1990). The unexpected result was that Government Phonology corresponds to a very weak fragment of SPE, yet it can attain the full expressivity of SPE through more powerful mechanisms of feature spreading. An issue that I did not elaborate on, however, is the question of what the increase in expressivity buys us in terms of empirical coverage, one which I take up in this paper. Again making good use of model theoretic techniques, I investigate two phonological phenomena, Sanskrit n-retroflexion and primary stress assignment in Creek and Cairene Arabic, and I show how much power feature spreading has to be granted in any descriptively adequate account which does not invoke additional technical machinery.

Since I derive my expressivity results by specialized mathematical means, which most linguists presumably are not familiar with, I preface the empirical part of this paper (section 3) with a high-level yet in-depth discussion of the formal machinery underlying my investigation (section 2). First, I explain and motivate the model theoretic approach to phonology in general, and I am careful to point out both its virtues and its limitations as well as how it complements the well-established tradition of empirically grounded theory comparison and evaluation. This is followed by a brief and accessible overview of the axiomatization of Government Phonology developed in Graf (2009). In particular, I justify specific parts of the formalization that might at first sight look like crucial deviations from the original theory. Beyond its more tenable expressivity results, then, this paper also offers reflections on why linguists should be at all interested in the model theoretic approach and how it supplements empirical theory comparisons to the benefit of the entire field.

2 Model Theoretic Phonology

2.1 The Basic Idea

Linguistic theories aren't monolithic entities; they usually come in different flavors that differ to a varying degree from the original proposals. Variants of Optimality Theory (OT), for example, can be built from a vast array of components such as correspondence theory, output-output correspondence and sympathy constraints. Clearly, these modifications aren't *ad hoc* inventions but are motivated by empirical concerns, so we should expect them to a have a noticeable impact on the inner workings of the theory. Unfortunately, it is often difficult to see how exactly these changes affect the original theory and the predictions it makes. This creates a big problem for theory comparisons: instead of comparing, say, OT to SPE, different incarnations of OT have to be compared to different incarnations of SPE. A quick survey of the development of phonology over the last 40 years shows that any well-developed phonological theory has at least three such optional modifications which can be mixed and matched, thereby giving rise to eight variants. Even under optimal conditions, then, a thorough comparison would have to consider at least sixteen theories, truly a herculean task.

An efficient way to reduce the complexity of the comparisons is to group theories into classes from which they inherit certain properties. If these properties are our only concern, it is sufficient to consider entire classes instead of each individual theory they contain. But what measure should be used as a classification scheme? Ideally, it will be general enough to allow for an easy and reliable classification of specific theories, yet at the same time detailled enough to capture properties of genuine linguistic interest. Model-theoretic phonology (MTP) is based on the realization that tools from mathematical logic provide us with a scheme that fulfills both requirements.

Let me illustrate the connection between linguistic theories and mathematical logic with an example first. Consider the formula $L \to \neg H$ of *propositional logic*.[1] It states that the presence of a low tone implies the absence of a high tone. We could just as well phrase the logical formula as a linguistic constraint: "No segment associated with L may be associated with H". Now take a look at the structure in Figure 1. It is easy to see that only the leftmost structure obeys the constraint, or as logicians would say, only the leftmost structure *satisfies* $L \to \neg H$ and is thus a *model* for it. We may use additional formulas to impose further well-formedness conditions, just as we can add further rules and constraints to phonological theories. But when we try to write a formula which imposes the requirement that every syllable with a high tone is both preceded and followed by syllables with a low tone, we run into a problem. Propositional logic cannot do this, since it considers only isolated nodes and fails to take context information into account. But this is easy to fix by adding two *operators* $\lhd$ and $\rhd$ which talk about the nodes immediately to the left and immediately to the right, respectively. Our propositional logic has now become a *modal logic*. The formula $H \to \lhd\lhd L \land \rhd\rhd L$ then enforces that two steps to the left and two steps to the right of a high tone there is a low tone. By now the reader should be able to check that only the structure in the middle is a model of this formula. Crucially, this implies that none of the structures is a model for both formulas.

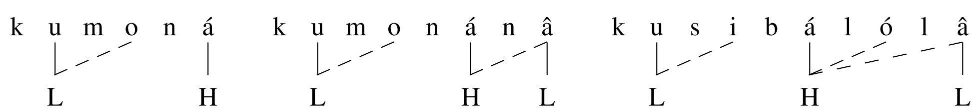

Figure 1: The formulas $L \to \neg H$ and $H \to \lhd\lhd L \land \rhd\rhd L$ are satisfied only by the leftmost structure and the one in the middle, respectively. None of the three structures satisfies both formulas.

Using more and more formulas as illustrated above, we restrict the set of well-formed structures in the same way a phonological theory does, although certain parts of a theory might require further operators or other modifications to propositional logic. The mathematical literature offers a broad range of logics which can be obtained in this way and whose properties are well-known. MTP is about establishing connections between theories and these logics, or putting it differently, MTP uses logics to classify linguistic theories. In particular, a theory can be assumed to inherit some of the properties of the weakest logic that is still sufficiently powerful to formalize it.

While the idea seems rather abstract, it is natural and efficient in practice. In particular, the classification of multiple variants requires hardly any additional work. After formalizing the original proposal, such as OT with correspondence theory, one is left with the easy task of formalizing the modifications, for instance sympathy constraints. If it turns out that the logic used for the original theory is too weak for the altered version, then the latter is more powerful than the former. This is exactly what Potts and Pullum (2002) did in their investigation of OT to show that sympathy constraints and output-output correspondence are proper extensions of standard OT. Potts and Pullum's case study also demonstrates that MTP produces new insights that are of immediate linguistic relevance and would be very difficult to obtain with traditional methods based on empirical comparisons. For instance, MTP allows us to derive *universal insufficiency* results by proving that specific phonological phenomena are beyond the reach of certain classes. That is to say, every phonological theory belonging to class C will fail to account for phenomenon P if P cannot be described in

[1] I use the standard notation for the logical connectives: $\neg$ for "not", $\land$ for "and", $\lor$ for "or", and $\to$ for "implies".

the logic corresponding to C. The traditional approach, on the other hand, derives *theory-specific sufficiency* results by devising an account of a specific phonological phenomenon in the theory under scrutiny. Rather than a replacement, then, the MTP approach is a useful complement to the traditional approach.

That MTP is no substitute for thorough empirical comparisons is also witnessed by the fact that the gains from a deliberate restriction to classes of theories come at the cost of reduced granularity. Hence MTP fairs better than alternative approaches in getting the big picture right and deriving general, broadly applicable results pertaining to generative capacity, computational complexity, memory requirements and parsing. On the other hand it has less to say about technical minutiae that do not correspond to class distinctions. This does not mean that we cannot use MTP for the investigation of such details, but it would be just as laborious a task as in any other approach.

2.2 Formalization of Government Phonology

In Graf (2009), I sought to demonstrate the usefulness of MTP by comparing SPE to the superficially very different Government Phonology (GP) and concluded that the class of SPE theories was identical to the class of GP theories when the latter is enhanced by a very powerful spreading mechanism.

GP as defined in Kaye et al. (1985, 1990) and Kaye (2000) differs from SPE in that it uses *privative* features (features without values) rather than binary ones, assembles these features in operator-head pairs instead of feature matrices, builds its structures according to an elaborate syllable template, employs empty categories and allows all features to spread (just like tone features in autosegmental phonology). Figure 2 gives an elaborate example of a GP structure. As can be seen, GP structures are built up from strings of skeleton nodes (designated by x), which are connected to constituents at the top and phonological expressions at the bottom. The dashed lines in the structure denote that the features **U** and **I** spread from the fifth into the second skeleton node.

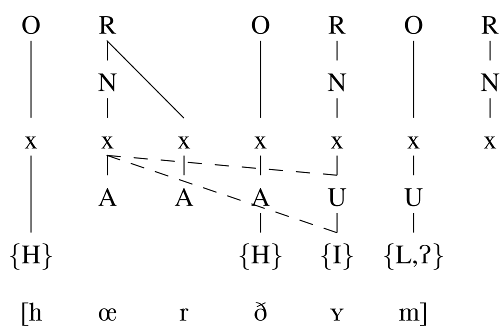

Figure 2: A complex example of a phonological structure in Government Phonology

The first step of the formalization in Graf (2009) is to accommodate GP's feature system. As indicated in Table 1 on the following page, GP replaces feature matrices by a pair consisting of a set of privative features, called *operators*, and a single privative feature, which functions as the *head*. It is usually stipulated that no feature may occupy both the head and an operator position. In my formalization, head and operator features are distinct features that are generated from a unique set of base features. Given a GP theory with three features **A**, **I**, **U**, one would use three head features $\mathbf{A}_h$, $\mathbf{I}_h$, $\mathbf{U}_h$ and three operator features $\mathbf{A}_o$, $\mathbf{I}_o$, $\mathbf{U}_o$. This makes it possible to regulate the entire feature calculus using only propositional logic, the weak logic encountered in our example in the previous section. Given the six features just listed, the pair for the sound ε, for instance, is represented by the logical formula $\mathbf{A}_h \wedge \neg\mathbf{I}_h \wedge \neg\mathbf{U}_h \wedge \neg\mathbf{A}_o \wedge \mathbf{I}_o \wedge \neg\mathbf{U}_o$. It is also straightforward to enforce the uniqueness of the head feature by formulas such as $\mathbf{A}_h \rightarrow \neg\mathbf{I}_h \wedge \neg\mathbf{U}_h$ for every head feature. Similarly, formulas such as $\mathbf{A}_h \rightarrow \neg\mathbf{A}_o$ ensure that a feature, in this case **A**, does not occupy both a head and an operator position.

Let us now turn to GP's syllable template, which establishes a distinction between vowels (in

r	({_},A)	a	({_},A)
j	({_},I)	i	({_},I)
w	({_},U)	u	({_},U)
g	({ʔ},_)	ə	({_},_)
s	({A,H},_)	e	({A},I)
n	({L,ʔ},A)	ɛ	({I},A)

Table 1: Some common phonological expressions

the *nucleus* position N) and consonants that do/do not need special licensing (*onsets* O/*codas*, which are dominated by the rhyme R). The six basic building blocks of the syllable template are listed in Figure 3 on the next page. They can be combined into bigger structures according to the rules in (1).

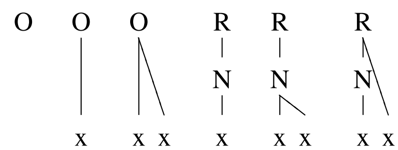

Figure 3: The six basic building blocks of phonological structure

(1) *How to combine the building blocks*
  a. Every structure consists of at least one rhyme.
  b. Every rhyme is immediately preceded by exactly one onset.
  c. Every onset immediately precedes exactly one rhyme.
  d. Every branching rhyme immediately precedes a unary branching onset.

As in the example from the previous section, propositional logic is too weak to express such structural information. But we can use the modal logic from the previous section, which allows us to take the neighborhood of a node into account thanks to its operators $\lhd$ and $\rhd$. With dedicated symbols for R, N and O, this logic is capable of expressing all constraints in (1). Condition (1d), for example, can be rendered as the formula $(R \wedge \lhd R) \rightarrow \rhd (O \wedge \rhd \neg O)$.

There are two special cases that need to be taken care of, though. The first one concerns the leftmost building block in Figure 3, a single O that is not associated to any skeleton node. This rather odd device is used to explain certain phenomena pertaining to word initial h in French. For mathematical reasons, I model this as a normal unary branching O — i.e. an O associated to a single skeleton node — which in turn hosts a special feature `fake`. The feature `fake` tells us that the onset in question represents an unassociated onset. Therefore, restricting the distribution of unassociated onsets is tantatmount to restricting the distribution of the feature `fake`, a simple task. The second minor complication is due to binary branching constituents, which I encode as two adjacent unary branching constituents of the same type. This does not introduce any conceptual confusions since such configurations cannot normally arise in GP, whence it is safe to assume that they represent binary branching constituents. Note that I am driven to this move by considerations of mathematical simplicity and elegance, but none of my results hinge on these minor alterations.

Given these minor modifications and the additional feature C, which is introduced for the sake of convenience to explicitly mark codas, we get bare GP syllable structures that look like the ones in Figure 4 on the next page.[2] It might be at this point that government phonologists start to take issue with my formalization and the slight simplifications it embodies, and the subsequent treatment of *empty categories* will in all likelyhood raise even greater concerns. By empty categories, I refer to

[2]GP practitioners might wonder how I accommodate short diphthongs, which are represented by two distinct phonological expressions, such as one for ɑ and one for ɛ, associated to the same skeleton node. I choose to handle this within the feature calculus by introducing another feature parameter (the first one being the distinction between heads and operators) that tells us whether a feature belongs to the first or the second expression.

GP's unique trait of allowing nuclei to remain unpronounced under specific conditions, as illustrated in Figure 5 on the facing page for Hebrew. The definitions and constraints involved in the distribu-

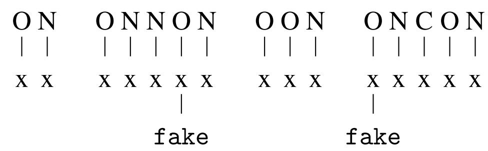

Figure 4: Examples of structures in simplified notation

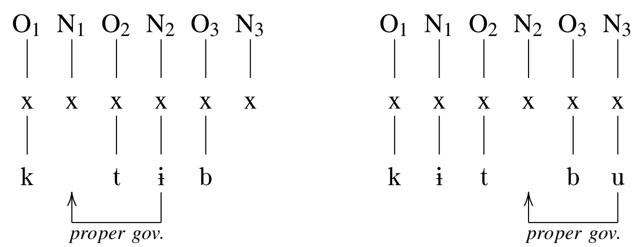

Figure 5: Proper government in Hebrew paradigms

tion of empty categories are rather complex, but if we ignore for a moment domain-final nuclei and so-called magic licensing configurations, the underlying intuition is easy to express in the simplified template: If a nucleus is pronounced, the preceding nucleus may remain unpronounced. If a nucleus is not pronounced, the preceding nucleus has to be pronounced. However, if the two nuclei are separated by two or more skeleton nodes associated to O or C, both have to be pronounced under all circumstances. Disbelieving readers may want to check themselves that this formulation yields the same results as the original definition of the Proper Government condition:

(2) *The phonological ECP*
A p-licensed empty category receives no phonetic interpretation.

(3) *p-licensing*
  a. Final Empty Nuclei Parameter (FEN)
     Domain-final empty categories are/aren't p-licensed.
  b. Magic Licensing
     s+consonant sequences license a preceding empty nucleus.
  c. Proper Government
     Properly governed (empty) nuclei are p-licensed.

(4) *Proper Government*
a properly governs *b* iff
  a. *a* and *b* are adjacent on the relevant projection level, and
  b. *a* is not itself licensed, and
  c. Neither *a* nor *b* are government licensors.

Admittedly, though, a single result that shows how these conditions can be simplified given a different encoding of the syllable template is insufficient to dispel doubts about the faithfulness of the formalization. Even though it is foremost an issue of philosophy of science to determine which parts of a theory need to be represented explicitly in its technical machinery, I believe my modelling decisions can be sufficiently supported on purely pragmatic grounds alone. For one has to keep in mind that the goal is to use as weak a logic as possible; but the weakest logic that might be expressive enough to allow for a direct translation of the conditions above, the *two variable fragment of first-order logic* (FO^2), is significantly more powerful than the modal logic I propose to use. One could of course try to put further restrictions on FO^2 to push it down to the level of a modal logic, but there is nothing to be gained from such a cumbersome move, because the two logics would then be identical from the perspective of MTP. In sum, the underlying issue here is that a formal approach always has to reconcile linguistic faithfulness with mathematical desiderata if it wants to be useful; the changes to GP I adopt above are, in my opinion, the best compromise between those two poles. Even if some of the readers may still take issue with the slight deviance of our formalization, they can rest assured that it is immaterial for the claims made in this paper, thanks to the granularity of the properties an MTP analysis investigates.[3]

The last module to be formalized is spreading. The technical details are rather involved (see Graf 2009:77ff.), but the general upshot is that our simple modal logic with the operators $\lhd$ and $\rhd$ captures only bounded spreading and unbounded spreading that arises from the iteration of bounded spreading steps. The latter is often assumed to underly processes such as vowel harmony, where a feature spreads from one vowel into the next one to the right and then uses this vowel as the new starting point for another spreading step, and so on. But truly unbounded spreading, i.e. unmediated spreading from one node into another one that can be arbitrarily far away, cannot be modeled with $\lhd$ and $\rhd$ alone. In Graf (2009), I showed how the use of further operators increases the power of the spreading mechanism and hence GP. The operators $\lhd^+$ and $\rhd^+$ from restricted temporal logic (RTL) enables us to spread into nodes that are arbitrarily far away, but we cannot tell how far a spreading step will take us. Metaphorically speaking, we have an unlimited field of vision but no depth perception. This problem does not arise with the operator U ("until") from linear temporal logic (LTL). A formula such as $O \rightarrow U(\mathbf{A}_h, \neg(\mathbf{I}_h \vee \mathbf{I}_o))$ asserts that for every onset o, there is a node n somewhere to the left of o with an $\mathbf{A}$ in head position and no $\mathbf{I}$ occurring between o and n. The counterpart of U for spreading in the other direction is S ("since"). While U and S seem to offer all the power one could ever need for a phonological analysis, some elaborate spreading patterns require the so-called least fixed point operator ν. It has been shown before by Vardi (1988) that ν-LTL is equivalent to monadic second-order logic (MSO), which in turn is the logic corresponding to the class of SPE theories (Kaplan and Kay 1994). Hence, the class of GP theories with elaborate spreading patterns is identical to the class of SPE theories. Table 2 gives a short overview of the spectrum of GP classes ordered by expressivity.

	$GP^{\lhd}$	$GP^{\lhd^+}$	GP^U	GP^{ν}/SPE
Modal logic	$\lhd$	RTL	LTL	ν-LTL
Predicate logic	—	FO^2	FO	MSO
Formal language	—	—	star-free	regular

Table 2: Hierarchy of classes of phonological theories

Note that spreading is the only decisive factor in establishing class membership. The feature system and the syllable template are in general ignored. In fact, I showed in Graf (2009) that the distinction between privative, binary and finitely valued feature systems is immaterial in so far as for every theory using finitely valued features, there is an equivalent one using binary features, and for every theory using binary features, there is an equivalent one using privative features. However, for individual theories that use the same set of features, it might indeed matter whether these features are allowed to take values or not. But since this is a question situated at a level between theories and theory classes, conclusive answers are difficult to obtain for both MTP and traditional comparisons.

3 Expressivity Results for Natural Language Phonology

3.1 The Power of Features

Before the empirical part of this paper can finally commence, it needs to be prefaced by a short disclaimer on how to interpret the results. Using a comparatively little-known mathematical theorem

[3] Moreover, GP actually benefits from these slight re-encodings, as we would otherwise be pressed to postulate that a very expressive logic is necessary for modelling GP when, in fact, it is not.

(Thatcher 1967), one can show that all variants of GP have the same power as SPE if non-local dependencies may be encoded by diacritic features (roughly, features that never have a visible effect on the surface string). That is to say, for any SPE grammar over a set F of features, there is an equivalent GP grammar over a set of features that is an extension of F. But fortunately, linguists do not generally use features in this way, and if we assume that the set of features is fixed across all theories, the expressivity hierarchy in Table 2 holds unchanged. Nevertheless, the power of feature coding forces us to explicitly relativize the results to specific feature sets, which makes them somewhat cumbersome to read.

3.2 Beyond $GP^{\lhd^+}$: Sanskrit n-Retroflexion

My first case study revolves around a well-known long-distance phenomenon, n-retroflexion in Vedic Sanskrit, also known as *nati*. As discussed in Schein and Steriade (1986) and Hansson (2001) (building on data given in Whitney 1889 and Macdonell 1910), the process turns the first postvocalic /n/ following a continuant retroflex consonant (called /r/ and /s/ here rather than /ɻ/, /ʂ/ for typographic ease) into a retroflex /ɳ/ iff the following conditions are fulfilled:

(5) a. No coronal consonant intervenes between trigger and target.
    b. The nasal is immediately followed by a (nonliquid) sonorant.
    c. No retroflex continuant occurs in the string after the nasal.

Two claims are true of nati. The first one asserts that there is a GP-variant that is expressive enough to faithfully model this process, whereas the second revolves around the insufficiency of $GP^{\lhd}$ and $GP^{\lhd^+}$ to this end.

Theorem 3.1. *There is a feature set* Σ *such that there is a theory over* Σ *that belongs to the class* GP^U *and accounts for nati.*

Proof. The conditions in (5) translate into the three GP^U axioms below. Note that the feature μ is used to denote unpronounced nuclei. Furthermore, for any sound or class of sounds i, I use $\ulcorner i \urcorner$ to denote the propositional formula over Σ that uniquely represents i. In the special case of $\ulcorner$*derived* ɳ$\urcorner$, this corresponds to a formula that is obtained from the formula $\ulcorner$ɳ$\urcorner$ by replacing the feature(s) that distinguish /ɳ/ from /n/ by their spread analogue. It is easy to see that there are feature systems where /ɳ/ has more features than /n/ (so that /ɳ/ has some features that can be replaced) and where all expressions in the axioms can be uniquely represented. Hence there is at least one suitable Σ. Together with the three GP^U axioms this proves the theorem.

N1 $\ulcorner \textit{derived}\ ɳ \urcorner \wedge \neg N \wedge \lhd (N \wedge \neg\mu) \rightarrow U(\ulcorner \mathrm{r} \urcorner \vee \ulcorner \mathrm{s} \urcorner, \neg \ulcorner \textit{coronal} \urcorner \wedge \neg \ulcorner \mathrm{n} \urcorner)$
"If a non-vocalic node X is postvocalic and contains a ɳ derived from an /n/, then there is a node X' labeled r or s to the left of X and neither a coronal nor an /n/ occurs between X and X'."

N2 $\ulcorner \textit{derived}\ ɳ \urcorner \wedge \neg N \wedge \lhd (N \wedge \neg\mu) \rightarrow \rhd \ulcorner \textit{sonorant} \urcorner$
"If a non-vocalic node X is postvocalic and contains a ɳ derived from an /n/, then there is a sonorant immediately to the right of X."

N3 $\ulcorner \textit{derived}\ ɳ \urcorner \wedge \neg N \wedge \lhd (N \wedge \neg\mu) \rightarrow \neg \rhd^+ \ulcorner \textit{retroflex continuant} \urcorner$
"If a non-vocalic node X is postvocalic and contains a ɳ derived from an /n/, then there is no retroflex continuant to the right of X." □

Theorem 3.2. *There is a feature set* Σ *such that at least one theory over* Σ *which belongs to the class* GP^U *accounts for nati but no theory over* Σ *belonging to a weaker class does.*

Proof. Since the formulas for (5b) and (5c) are in fact $GP^{\lhd^+}$ formulas, the culprit has to be (5a). As seen above, $GP^{\lhd^+}$ is a special variant of restricted temporal logic over strings, which can be formalized in FO^2. This fragment contains all and only those formulas of full first-order logic that use at most two variables (e.g., not $\exists x \exists y \exists z[give(x,y,z)]$, but $\forall x \exists y[loves(x,y)]$ and also $\forall x[\exists y[loves(x,y)] \rightarrow$

$\neg\forall y[hates(x,y)]]$). Hence it suffices to show that (5a) cannot be stated in FO^2. This can be demonstrated using standard techniques from finite model theory (in particular pebble proofs), which would take us too far here. But even on an intuitive level it should be clear that one needs three variables for (5a): two in order to mark the edges of the interval defined by trigger and target, and a third one to restrict the nodes within said interval. □

Presumably, the astute reader has already realized that our proof establishes a stronger result: given such a Σ, there is no theory in the class $GP^{\triangleleft^+}$ that accounts for *any* process which involves checking nodes within an interval of unbounded size. Under the proviso that there are empirical phenomena besides nati that exhibit this property, Theorem 3.2 presents a catalog of phenomena that $GP^{\triangleleft^+}$-theories cannot account for without careful tweaking of their feature system.

But the discovery of further "unbounded interval" processes is likely to prove difficult, because it is a notoriously hard task to establish conclusively that the unboundedness of the size of the interval does not arise from the iteration of bounded spreading steps. In the case of nati, for instance, it is also conceivable (see Hansson 2001:242f) that what we are actually dealing with is a sequence of local retroflexion-steps, only the last of which is marked in the Sanskrit writing system. Then nati could easily be accommodated in a $GP^{\triangleleft}$-theory. Well-formedness conditions not involving any spreading to begin with (e.g., "no s between p and t") would thus constitute a better place to look for such unbounded intervals, but they all seem to be restricted to small phonological constituents such as syllables or onsets.

3.3 Beyond GP^U: Primary Stress Assignment in Creek and Cairene Arabic

I now turn to primary stress assignment in Creek and Cairene Arabic (Mitchell 1960, Haas 1977). I only list the stress rules of Cairene Arabic, as the general reasoning applies to Creek as well.

(6) *Stress assignment in Cairene Arabic*
  a. Stress the final syllable, if it is superheavy (CV:C or CVCC).
  b. Else stress the penult, if it is heavy (CV: or CVC).
  c. Else stress the penult or the antepenult, whichever is separated by an even number of syllables from the closest preceding heavy syllable (or, if there is no such syllable, from the beginning of the word).
  d. There is no overt marking of secondary stress.

Ignoring for a moment (6d), we note that conditions (6a–c) are fairly unremarkable from a typological perspective and can be found in hundreds of languages, often in conjunction with trochaic or iambic secondary stress assignment. It is also easy to see how these constraints could be captured in GP^U.[4] Clearly, (6a) and (6b) are rather simple implicational statements of the form $a \rightarrow b$, where a and b are logical descriptions of the respective structural configurations. Constraint (6c) is just as straightforward to accommodate if we make use of the distribution of secondary stress, thereby simplifying it to (6c′) below.

(6c′) *Modified stress assignment for a system with trochaic secondary stress*
  Else stress the penult or the antepenult, whichever
  a. immediately follows a heavy syllable, or else
  b. if the closest preceding heavy syllable has secondary stress or if there is no preceding heavy syllable, is two syllables to the right of a syllable with secondary stress, or else
  c. if the closest preceding heavy syllable is stressless, is one syllable to the right of a syllable with secondary stress.

The revised rule replaces the unbounded counting of syllables between the closest heavy syllable and the potential targets of primary stress by a strictly local iterative process of distributing secondary

[4]The technical details are a little intricate due to stress rules operating on syllables, which do not exist as discrete entities in GP. This makes it necessary to distribute stress features over multiple segments, wherefore the formulas would look more complicated than they actually are.

stresses in a trochaic rhythm and calculating primary stress based on this secondary stress pattern, much in the spirit of a metric analysis using binary branching feet (Hayes 1995).

This approach, however, faces a severe problem: (6d) makes it clear that there is no overt secondary stress in Cairene Arabic. As a consequence, the secondary stress feature is degraded to the status of a coding feature, and as such it cannot be assumed to be an integral part of all feature systems.[5] Thus, we find a bifurcation between GP^{v} and GP^{U} as stated below.

Theorem 3.3. *There is a feature set Σ such that there is a theory over Σ that belongs to GP^{v} and accounts for primary stress assignment in Cairene Arabic.*

Proof. From our discussion above it follows that the crucial factor in assigning primary stress is counting the number of syllables. In particular, we need to be able to distinguish even from odd syllables. This corresponds to counting *modulo 2*.[6] It is known that GP^{v} is equivalent to MSO (Vardi 1988), which in turn provides a logical characterization of the regular stringsets (Büchi 1960), which in turn are equivalent to the string yield of finite state automata. Finally, there is a simple and well-known algorithm for constructing a finite state automaton that counts *modulo n*, *n* finite, whence any GP^{v} theory can count *modulo 2*. Since the only feature needed for assigning stress is a primary stress feature, almost every choice of Σ is sufficient. □

Theorem 3.4. *There is a feature set Σ such that at least one theory over Σ which belongs to the class GP^{v} accounts for primary stress assignment in Cairene Arabic but no theory over Σ belonging to a weaker class does.*

Proof. It is well known that full first-order logic cannot count *modulo n*. As was proven by McNaughton and Pappert (1971) and Thomas (1979), the stringsets definable in first-order logic are the star-free stringsets, which in turn are also the stringsets definable in LTL (Cohen et al. 1993), in which I formalized GP^{U}. □

I feel obliged to point out that the alleged absence of secondary stress marking is highly contentious. Therefore, both instances of "non-$GP^{\lhd}$ phenomena" involve a considerable amount of empirical uncertainty. The apparent scarcity of conclusive evidence for mechanisms in natural language phonology that go beyond $GP^{\lhd}$ is somewhat unexpected considering that both SPE and OT are significantly more powerful (see Kaplan and Kay 1994 and Frank and Satta 1998, respectively).

4 Conclusion

The findings in this paper complement some earlier results of mine (Graf 2009) by tying the expressivity hierarchy of GP variants to concrete empirical phenomena and explicitly comparing the model-theoretic perspective with other approaches to theory comparison. On a more general level, I wanted to show how the application of model-theoretic methods to phonology initiated in Potts and Pullum (2002), despite its somewhat unusual focus on entire classes of theories, supplements traditional methods of theory comparison and unearths interesting new results with broad applicability. I hope that linguists will conceive of it as a welcome addition to their analytic toolbox.

References

Büchi, J. Richard. 1960. Weak second-order arithmetic and finite automata. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik* 6:66–92.

Chomsky, Noam, and Morris Halle. 1968. *The Sound Pattern of English*. New York: Evanston.

[5] Keep in mind that this isn't a claim about the validity of metric approaches such as Hayes's but rather a methodological requirement that prevents the phonological hierarchy from collapsing and ensures the general validity of our theorems.

[6] *n* the value of *n modulo 2* is the remainder of dividing *n* by 2, so n *modulo 2* $= 0$ for *n* even and 1 otherwise.

Cohen, Joelle, Dominique Perrin, and Jean-Eric Pin. 1993. On the expressive power of temporal logic. *Journal of Computer and System Sciences* 46:271–294.

Frank, Robert, and Giorgio Satta. 1998. Optimality theory and the generative complexity of constraint violability. *Computational Linguistics* 24:307–315.

Graf, Thomas. 2009. Towards a factorization of string-based phonology. In *Proceedings of the* 14*th* *Student Session of the European Summerschool for Logic, Language and Information*, ed. Thomas Icard, 72–84.

Haas, Mary R. 1977. Tonal accent in Creek. In *Southern California Occasional Papers in Linguistics*, ed. Larry M. Hyman, volume 4, 195–208. Los Angeles: University of Southern California.

Hansson, Gunnar Ólafur. 2001. Theoretical and Typological Issues in Consonant Harmony. Doctoral Dissertation, University of California, Berkeley.

Hayes, Bruce. 1995. *Metrical Stress Theory*. Chicago: Chicago University Press.

Kaplan, Ronald M., and Martin Kay. 1994. Regular models of phonological rule systems. *Computational Linguistics* 20:331–378.

Kaye, Jonathan. 2000. A User's Guide to Government Phonology. Ms., University of Ulster.

Kaye, Jonathan, Jean Lowenstamm, and Jean-Roger Vergnaud. 1985. The internal structure of phonological representations: a theory of charm and government. *Phonology Yearbook* 2:305–328.

Kaye, Jonathan, Jean Lowenstamm, and Jean-Roger Vergnaud. 1990. Constituent structure and government in phonology. *Phonology Yearbook* 7:193–231.

Kracht, Marcus. 2003. Features in phonological theory. In *Foundations of the Formal Sciences II, Applications of Mathematical Logic in Philosophy and Linguistics*, ed. Benedikt Löwe, Wolfgang Malzkorn, and Thomas Räsch, number 17 in Trends in Logic, 123–149. Dordrecht: Kluwer.

Macdonell, Arthur. 1910. *Vedic Grammar*. Strassburg: Trübner.

McNaughton, Robert, and Seymour Pappert. 1971. *Counter-Free Automata*. Cambridge, Mass.: MIT Press.

Mitchell, Terence F. 1960. Prominence and syllabification in Arabic. *Bulletin of the School of Oriental and African Studies* 23:369–389.

Potts, Christopher, and Geoffrey K. Pullum. 2002. Model theory and the content of OT constraints. *Phonology* 19:361–393.

Rogers, James. 1998. *A Descriptive Approach to Language-Theoretic Complexity*. Stanford: CSLI.

Schein, Barry, and Donca Steriade. 1986. On geminates. *Linguitic Inquiry* 17:691–744.

Thatcher, James W. 1967. Characterizing derivation trees for context-free grammars through a generalization of finite automata theory. *Journal of Computer and System Sciences* 1:317–322.

Thomas, Wolfgang. 1979. Star-free regular sets of ω-sequences. *Information and Control* 42:148–156.

Vardi, Moshe Y. 1988. A temporal fixpoint calculus. In *Proceedings of the 15th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*, 250–259.

Whitney, William D. 1889. *Sanskrit Grammar*. London: Oxford University Press.

Department of Linguistics
University of California, Los Angeles
3125 Campbell Hall
Box 951543
Los Angeles, CA 90095-1543
tgraf@ucla.edu