

An Iterative Strategy for Learning Metrical Stress in Optimality Theory

Rutgers University has made this article freely available. Please share how this access benefits you.
Your story matters. <https://rucore.libraries.rutgers.edu/rutgers-lib/41981/story/>

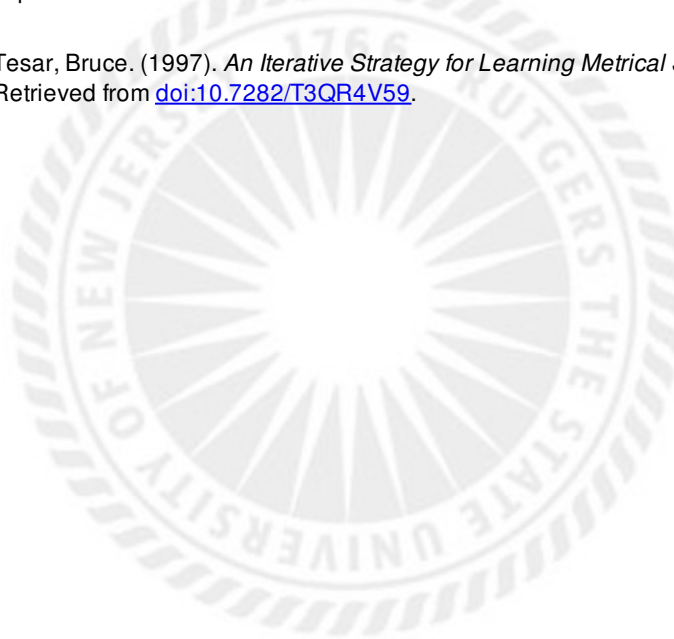
This work is an **ACCEPTED MANUSCRIPT (AM)**

This is the author's manuscript for a work that has been accepted for publication. Changes resulting from the publishing process, such as copyediting, final layout, and pagination, may not be reflected in this document. The publisher takes permanent responsibility for the work. Content and layout follow publisher's submission requirements.

Citation for this version and the definitive version are shown below.

Citation to Publisher Tesar, Bruce. (1997). *An Iterative Strategy for Learning Metrical Stress in Optimality Theory*.
Version: <http://www.cascadilla.com/buclid21toc.html>.

Citation to *this* Version: Tesar, Bruce. (1997). *An Iterative Strategy for Learning Metrical Stress in Optimality Theory*.
Retrieved from [doi:10.7282/T3QR4V59](https://doi.org/10.7282/T3QR4V59).



Terms of Use: Copyright for scholarly resources published in RUcore is retained by the copyright holder. By virtue of its appearance in this open access medium, you are free to use this resource, with proper attribution, in educational and other non-commercial settings. Other uses, such as reproduction or republication, may require the permission of the copyright holder.

Article begins on next page

To appear in *The Proceedings of the 21st Annual Boston University Conference on Language Development*, November 1996.

An Iterative Strategy for Learning Metrical Stress in Optimality Theory

Bruce Tesar
Rutgers University

1. Introduction

Optimality Theory (Prince & Smolensky, 1993) accounts for linguistic competence in terms of optimization over a ranked, universal set of violable constraints. For any given underlying form, there is a general space of candidate structural descriptions of that input. The candidate which best satisfies the constraints is the grammatical description. Thus, a linguistic structure need not satisfy all of the constraints completely in order to be grammatical; it is required only to better satisfy the constraints than all competing candidates (the other candidates for the same input). Conflicts between constraints are mediated by the ranking imposed upon the constraints. If one constraint dominates another, the dominating constraint gets priority, so that candidates which better satisfy the dominating constraint are preferred, even if other candidates much better satisfy the dominated constraint.

One of the central claims of Optimality Theory is that cross-linguistic variation is accounted for by different rankings of the same universal constraints. The space of possible linguistic inputs, knowledge of the candidate descriptions for each input, and the content of the universal constraints are all innate; only the lexicon and the ranking of the constraints are language-specific. The possible rankings of the universal constraints gives rise to a typology of human linguistic grammars. Thus, to learn the core grammar of a language is to learn the language-specific ranking of the universal constraints. In previous work, Tesar and Smolensky (to appear) developed an algorithm, the constraint demotion algorithm, that correctly and efficiently deduces the constraint ranking for any language, given a representative set of grammatical structural descriptions from the language. Thus, if the learner is able to recover the full structural descriptions for the positive data they observe, the constraint ranking is learnable.

However, one of the major challenges of language acquisition is the fact that the auditory signal received by the learner underdetermines the structural description of the utterance. This “gap” between the overt, phonetic form and the full grammatical description must be overcome during the process of learning. This gap can appear to put the language learner in a paradoxical situation: the correct grammar is needed to

ensure that the overt forms are correctly interpreted (assigned the correct descriptions), but the correct descriptions are needed to ensure that the correct grammar (here, the ranking of the constraints) is learned.

Tesar and Smolensky (to appear) outlined a proposal for a larger theory of language learning, of which the constraint demotion algorithm is a part. The first results from examining this proposal examined a system of quantity insensitive-only stress systems (Tesar, to appear). This paper extends that investigation to a more complex system, including many grammars with quantity sensitivity. Section 2 briefly describes the optimality theoretic system of metrical stress grammars used here. Section 3 very briefly describes the constraint demotion algorithm. Section 4 describes the strategy used to attack the problematic gap between overt forms and full descriptions. Section 5 presents the results of some computer simulations used to test and investigate the learning strategy.

2. Metrical Stress in Optimality Theory

The following optimality theoretic system is loosely based upon analyses proposed by John McCarthy and Alan Prince (Prince, 1990; McCarthy & Prince, 1993). While it is not a complete account of all attested metrical phenomena, it does capture many of the central cross-linguistic stress patterns, making it a worthy test bed for the investigation of learning. Metrical stress was selected because it permits the issue of input/output faithfulness to be set aside: underlying forms are strings of syllables, and structural descriptions assign stresses to the syllables, but no insertion/deletion of syllables is considered (see (Tesar & Smolensky, to appear) for discussion of the learning of lexical underlying forms). Thus, the underlying form for an utterance can be directly (and correctly) inferred from the overt form; the underlying form is simply the syllables of the overt form (without the stresses). Each structural description is of a single prosodic word, and so all overt forms are of single prosodic words.

The underlying forms are strings of syllables marked for weight (light or heavy); L denotes a light syllable, H a heavy syllable. The inputs range from 2 to 5 syllables in length, with all possible combinations of heavy and light syllables, giving a total of 60 distinct possible inputs. The overt forms are strings of syllables with marked stress levels: 1 denotes main stress, 2 denotes secondary stress, and 0 denotes unstressed. A structural description is a grouping of the syllables (with their stress levels) into feet. Feet are strictly bi-syllabic (i.e., heavy syllables cannot form monosyllabic feet). Every word is required to have precisely one head foot, and the head syllable of the head foot receives main stress. If other feet are present, their head syllables receive secondary stress.

As an illustration, some candidate structural descriptions and corresponding overt forms for the underlying form [L H L L L] are

[(L0 H1) (L0 L2) L0]	[L0 H1 L0 L2 L0]
[L0 (H1 L0) (L2 L0)]	[L0 H1 L0 L2 L0]
[(L1 H0) L0 L0 L0]	[L1 H0 L0 L0 L0]

The first description has left-aligned iambic feet, with main stress on the leftmost foot. What is missing from the overt forms is the foot structure; it is the foot structure that must be recovered when the learner interprets an overt form. The overt forms can be ambiguous; observe that the first two descriptions above have identical overt forms.

There are 12 universal constraints, all freely rankable:

WSP	heavy syllables must be stressed
PARSE-SYLLABLE	a syllable must be footed
MAIN-RIGHT	align the head-foot with the word, on the right edge
MAIN-LEFT	align the head-foot with the word, on the left edge
ALL-FEET-RIGHT	align each foot with the word, on the right edge
ALL-FEET-LEFT	align each foot with the word, on the left edge
WORD-FOOT-RIGHT	align the right edge of the word with a foot
WORD-FOOT-LEFT	align the left edge of the word with a foot
IAMBIC	align the head-syllable with its foot, on the right edge
TROCHAIC	align the head-syllable with its foot, on the left edge
NONFINAL	the final syllable should not be footed
NONINITIAL	the initial syllable should not be footed

WSP (the weight-to-stress principle) produces quantity sensitivity effects when suitably ranked. The ranking of PARSE-SYLLABLE determines when and if secondary stresses will occur. The ALL-FEET-R/L constraints capture what has previously been analyzed as directional iterativity, while NONINITIAL and NONFINAL have effects often analyzed in terms of extrametricality.

Four of the constraints, MAIN-R/L and ALL-FEET-R/L, are gradient alignment constraints between edges of feet and the prosodic word. Each such constraint assesses a constraint violation for every prosodic word constituent intervening between the relevant foot-edge and the relevant word-edge.

3. Constraint Demotion

The error-driven constraint demotion algorithm (Tesar and Smolensky, 1995; Tesar, 1995) learns constraint rankings on the basis of grammatical full structural descriptions (including both overt and hidden structure). The procedure works by pairing a grammatical structural description, called the *winner*, with a competing description (a different description of the same underlying form as the winner), called the *loser*. The winner is presumed to be more harmonic (better satisfies the ranked

constraints) than any competitor, including the loser, with respect to the correct constraint ranking. If the winner is not more harmonic than the loser with respect to the learner's ranking, constraint demotion modifies the ranking, producing a new ranking with respect to which the winner is more harmonic than the loser.

The basic principle behind constraint demotion is rather simple. All constraints violated more by the winner than by the loser are demoted down in the hierarchy. Each is demoted just far enough so that it is dominated by a constraint violated more by the loser. In the new hierarchy, the loser is less harmonic than the winner, because it fares worse on a constraint which now dominates all the constraints on which the winner fares worse.

Further discussion of constraint demotion, including formal proofs of the algorithm's correctness and of its data requirements, as well as the interpretation of non-total rankings presumed by the algorithm, can be found in (Tesar, 1995) and (Tesar & Smolensky, to appear).

4. The Iterative Learning Strategy

The iterative strategy is based in part upon proposals for linguistic processing in Optimality Theory. The core function defined by an optimality theoretic grammar maps each underlying form to a structural description, the description which best satisfies the ranked constraints, out of all the candidate descriptions for that underlying form. This mapping has natural interpretation in terms of language generation. The language user, given an underlying form (e.g., the syllables of the word), computes the optimal description (e.g., metrical structure). The process of computing the optimal description of an underlying form is called *production-directed parsing*.

A corresponding mapping for language comprehension would be one which maps from an overt form of the language to the grammatical structural description. A definition for such a mapping was suggested by Tesar and Smolensky (to appear); their definition also uses optimization with respect to the user's constraint ranking, but this time the candidate structural descriptions are those descriptions whose overt portion matches the observed overt form exactly. A structural description assigned to an overt form is called an interpretation of that overt form. The act of interpreting an overt form is that of computing the optimal structural description with an overt portion matching the observed overt form. The process of computing the optimal interpretation of an overt form is called *interpretive parsing*.

In the case of metrical stress, the underlying form is contained in the overt form (the underlying form is the overt form with the stress levels removed), so the set of descriptions matching a particular overt form are a subset of the set of candidate descriptions for the underlying form. For a language user with a full and correct competence (i.e., the correct constraint ranking), applying interpretive parsing to an overt form will result in a description identical to the one produced by applying production-directed parsing to the corresponding underlying form. This fact is central

to learning, because it will be precisely the mismatch between the interpretation of the surface form and the learner's own stressing of the same underlying form that triggers learning. That is, when the learner hears a word (with stresses), and the learner's own current grammar stresses the same word differently, the learner must actively attempt to change their grammar (the constraint ranking) so that the learner's new grammar stresses the word more like what the learner heard.

The familiar challenge is, of course, *how* to change the learner's grammar to one which more closely corresponds to what is heard. This problem is non-trivial because (a) the overt form often does not uniquely determine the corresponding full structural description; and (b) the principles of the grammar available for manipulation crucially evaluate full descriptions.

Consider the following simple example, for the underlying form consisting of three light syllables, [L L L]. Suppose the overt form [L0 L1 L0] is observed. There are two structural descriptions consistent with the overt form: a description with a left-aligned iambic foot, [(L0 L1) L0], and a description with a right-aligned trochaic foot, [L0 (L1 L0)]. If the learner's current grammar assigns to the underlying form the description [(L1 L0) L0], then it is clear that the learner needs to change their grammar, but they have a choice between changing the foot alignment from left to right, and changing the foot form from trochaic to iambic.

The problem of interest, then, is the ambiguity of some overt forms between multiple structural descriptions. The solution advocated here relies heavily on the optimization-based processing account described at the beginning of this section. In particular, the claim is that the language learner is not merely using an overt form for comparison, but is actively trying to interpret it, which requires assigning it a structural description. For metrical stress, this means assigning a foot form supporting the observed stressing. The interpretation function is that function computed by interpretive parsing, as previously defined: the learner uses their current grammar to compute the most harmonic description matching the observed overt form.

Table 1 shows an example where production-directed parsing and interpretive parsing arrive at different descriptions. Consider the subset of the constraints displayed in the table, ranked in the order that they appear, from left to right. The overt form observed is a word of five light syllables with penultimate main stress only. The form labeled 'winner' is the result of interpretive parsing applied to the overt form; the sole stressed syllable must be the head of the only foot, and the ALL-FEET-LEFT constraint forces the foot to be iambic in order for the left edge of the foot to be as close to the left edge of the word as possible. The form labeled 'loser' is the result of production-directed parsing applied to the underlying form; the stress levels of the overt form are ignored, and the optimal description of five light syllables assigns two feet (PARSE) which are left-aligned (ALL-FEET-LEFT) and iambic (IAMBIC).

		Parse	A-F-L	M-R	Iamb	A-F-R	Troch
loser	[(0 2)(0 1)0]	*	*	*		* * *	* *
winner	[0 0 (0 1) 0]	* * *	* *	*		*	*

Table 1: Underlying form: [L L L L L] Overt form: [L0 L0 L0 L1 L0]

The winner is less harmonic because it has more violations of PARSE, the highest-ranked constraint. Constraint demotion will demote PARSE and ALL-FEET-LEFT to below ALL-FEET-RIGHT, so that, in the resulting hierarchy, the highest constraint on which the loser and winner differ will be one violated more by the loser, making it less harmonic. Once that demotion has been made, interpretive parsing may be re-applied with the new hierarchy to the overt form to get the new winner, and production-directed parsing re-applied to the underlying form, with the result shown in table 2.

		M-R	Iamb	A-F-R	Troch	Parse	A-F-L
loser	[0 0 0 (0 1)]				*	* * *	* * *
winner	[0 0 0 (1 0)]		*			* * *	* * *

Table 2: Underlying form: [L L L L L] Overt form: [L0 L0 L0 L1 L0]

As a result of the demotion, ALL-FEET-RIGHT \gg PARSE, so the new loser only has one foot, and MAIN-RIGHT \gg IAMBIC, so both descriptions have their head foot right-aligned. The only difference between loser and winner is foot-form; the next demotion will demote IAMBIC to below TROCHAIC, at which point both production-directed and interpretive parsing will produce the same description, namely the winner of table 2.

More precisely, the learning procedure is as follows:

Starting with ranking H_0 and overt form F:

1. Apply interpretive parsing, using H_x , to F, getting interpretation D_i
2. Apply production-directed parsing, using H_x to the underlying form of D_i , getting D_p
3. If $D_i = D_p$:
learning is done for overt form F, returning H_x
4. If $D_i \neq D_p$:
(a) apply constraint demotion to H_x using D_i and D_p , getting new ranking H_{x+1}
(b) repeat the procedure from step 1, using the new ranking H_{x+1}

If interpretive parsing always produces the correct description for each overt form, then this procedure is guaranteed to converge upon a correct hierarchy. However, the ambiguity of some of the overt forms means that there are situations where interpretive parsing will produce the incorrect description. The question is raised, then, as to how the procedure will perform in general on the overt stress patterns of a variety of grammars. Of particular interest are cases where interpretive parsing does not produce the correct description.

One cause for optimism lies in earlier work investigating this learning strategy. Tesar (to appear) applied the same strategy to a simpler system of stress grammars. In that simpler system, the algorithm reached a correct hierarchy for a significant majority of the tested stress patterns. Significantly, cases were discovered in which the algorithm converged to a correct hierarchy despite the fact that interpretive parsing produced the wrong description at some point, demonstrating the possibility of such recovery (an example is the illustration given above in tables 1 and 2).

A more extensive discussion of this strategy, including a discussion of some of the ideas from learning theory that originally motivated it, can be found in (Tesar, to appear).

5. Simulation Results

The stress system previously investigated (Tesar, to appear) was significantly simpler than the one presented in this paper. Specifically, no analysis of syllable weight was used: there was no WSP constraint, and syllables were not distinguished as light or heavy. Thus, there was only one input for each word length. There were 6 different underlying forms of length 2 through 7 syllables.

The system investigated here does distinguish between light and heavy syllables, and the WSP constraint is sensitive to the distinction. This makes the system of grammars more complicated. In particular, the number of distinct underlying forms expands greatly; the number of possible underlying forms grows exponentially with the maximum length of form considered. Because of this expansion, the experiments described here only used forms of from 2 to 5 syllables in length, giving a total of 60 distinct underlying forms.

A simple implementation of the iterative strategy was run on the overt stress patterns of 50 different grammars. Each stress pattern consisted of the overt forms generated by the corresponding grammar for all 60 possible words. The stress patterns were obtained by selecting 50 different total rankings of the constraints, and generating the 60 overt forms for each total ranking. In 30 of the 50 tested grammars, WSP was ranked high, resulting in quantity sensitivity (QS); the other 20 were quantity insensitive (QI).

One potentially influential factor in the application of the learning algorithm is the choice of a starting hierarchy. Any exhaustive investigation of all possible initial

hierarchies would be a huge task indeed, and the effects of the initial hierarchy certainly interact with a number of other design factors. In the current work, two starting hierarchies were investigated. The first starting hierarchy, the monostratal hierarchy, is maximally non-committal: all of the constraints are lumped into a single stratum of a hierarchy. This means that at the beginning of learning, no constraint dominates any other constraint. It is worth pointing out that this hierarchy does not correspond to any adult grammar.

When the starting hierarchy had all the constraints in one stratum, the following results were obtained:

	Number of patterns	Number that converged	Median number of demotions	Range of number of demotions
QS	30	24	13	5 - 16
QI	20	16	4	3 - 5
Total	50	40		

Table 3: Results with the Monostratal Starting Hierarchy

Most of the cases which failed to converge did so as a consequence of a particular circumstance, in which both of the foot-form constraints, TROCHAIC and IAMBIC, were demoted low in the hierarchy. One version of this difficulty was responsible for all of the convergence failures for quantity insensitive target patterns, and is identical to the failure discussed in (Tesar, to appear). The problem occurs when an overt form has, as its best interpretation, a structural description with inconsistent foot form (trochaic and iambic feet in the same description). In this grammar system, inconsistent foot form can only occur optimally as a consequence of quantity sensitivity effects; it cannot occur optimally in forms not subject to quantity sensitivity, such as a word of all light syllables. The selected winner is not a possible optimal description, and the algorithm performs an endless cycle of demotions, demoting TROCHAIC below IAMBIC and then the reverse, in a futile effort to make the inconsistent form optimal.

The other version of the difficulty was responsible for most of the convergence failures for quantity sensitive patterns. This problem has inconsistency of footing, but across two different forms. The result is that, for every pass through the overt forms, one overt form demotes TROCHAIC below IAMBIC, and then another form does the reverse.

An example of such a situation is shown in table 4. To keep the table to a manageable size, only four constraints are shown; assume the other constraints to be ranked at least below ALL-FEET-LEFT. In the target grammar, WSP \gg TROCHAIC \gg ALL-FEET-LEFT. However, the relative ranking of TROCHAIC and ALL-FEET-LEFT only

matters in a few forms containing heavy syllables. The overt form in table 4 is such a form. The loser, optimal under the current ranking, has inconsistent foot form, due to the high ranking of WSP and ALL-FEET-LEFT. The winner, the best interpretation of the overt form, satisfies ALL-FEET-LEFT as much as possible, at the expense of two TROCHAIC violations. The correct description from the target language, shown in the bottom row, achieves equal satisfaction of WSP, but with two trochaic feet. The learner does not know this, however.

[L0 H1 H0 H2 L0]		WSP	ALL-Ft-L	TROCH	IAMB
loser	[(L0 H1)(H2 H0) L0]	*	*	*	**
winner	[(L0 H1)(H0 H2) L0]	*	*	**	*
Correct	[L0 (H1 H0)(H2 L0)]	*	***		**

Table 4: ALL-Ft-L \gg TROCHAIC, causing the wrong interpretation (winner).

The only constraints on which the winner and loser differ are the foot-form constraints. Constraint demotion will demote TROCHAIC, violated more by the winner, to below IAMBIC, violated more by the loser, to make the winner more harmonic. The resulting hierarchy will only last until a form is reached in which TROCHAIC has effects that do not interact with ALL-FEET-LEFT, as shown in table 5.

[H1 L0 H2 H0 L0]		WSP	ALL-Ft-L	IAMB	TROCH
loser	[(H1 L0)(H0 H2) L0]	*	*	*	*
winner (Correct)	[(H1 L0)(H2 H0) L0]	*	*	**	

Table 5: With ALL-Ft-L indifferent, IAMBIC will be demoted below TROCHAIC.

The winner here is the correct form, having consistent foot-form (trochaic feet) because WSP is indifferent as to which heavy syllable in the second foot receives stress. Here, constraint demotion (correctly) demotes IAMBIC to below TROCHAIC, making the winner optimal. However, the problematic domination of TROCHAIC by ALL-FEET-LEFT remains, leading to a repeat of this problem on all subsequent passes through the overt forms.

Out of the 50 grammars whose results are shown in table 3, 40 of the target grammars (20 QI, 20 QS) had the dominant foot-form constraint (TROCHAIC or IAMBIC) ranked high, at the top for the QI cases, and dominated only by WSP in the QS cases. This was done intentionally, in an effort to find the kinds of failures just described. To test the hypothesis that having the foot-form constraints high in the target grammars was the problem, the overt forms for 10 more quantity sensitive target grammars were generated. The generating rankings for these 10 were identical to 10

of the earlier QS languages except that the dominant foot-form constraint was ranked low. When the learning algorithm was applied to these 10 with the same initial hierarchy, all 10 cases rapidly converged to a correct constraint hierarchy.

After an analysis of these results, a new starting hierarchy, shown below, was hand-constructed to attempt to avoid the problems appearing for the languages on which the algorithm failed to converge:

WSP » PARSE » IAMBIC » TROCHAIC » {all other constraints}

This starting hierarchy was motivated by the following observations. Failure to converge resulted most frequently from both of the foot-form constraints being low in the hierarchy. Ranking them both above most of the other constraints permitted the dominant one of them in the language to remain high, with the other one being demoted if necessary. The only two other constraints that the foot-form constraints directly interact with are WSP and PARSE. If the dominant (higher-ranked) foot-form constraint is ranked high, its demotion will usually only be triggered if the target language requires that it be dominated by either WSP or PARSE. By having an initial hierarchy with those two constraints already above both foot-form constraints, the dominant foot-form constraint should never get demoted: if it must be dominated by either WSP or PARSE, it already is, and if it must dominate one or both of WSP and PARSE, then it is the latter constraints that will be demoted, leaving the dominant foot-form constraint at the top. The goal is to arrange the system so that the dominant foot-form constraint never has to be demoted, thus avoiding the pathologies sometimes caused by having both foot-form constraints too low.

The same algorithm was applied to the same stress patterns using this starting hierarchy, and the following results were obtained:

	Number of patterns	Number that converged	Median number of demotions	Range of number of demotions
QS	30	28	4	3 - 6
QI	20	20	4	2 - 6
Total	50	48		

Table 6: Results with the Designed Starting Hierarchy

Using this starting hierarchy resulted in rapid convergence for 48 out of the 50 languages. For all 48 languages which converged, a correct constraint hierarchy was found after at most 6 instances of demotion (six mismatch errors) total out of all the 60 forms of the language. Interestingly, the two failures were on QS grammars where the foot-form constraints were required to be ranked low, but remained too high during the course of learning.

It is worth noting for purposes of comparison that an exhaustive search of the space of the possible total rankings of the constraints would have to test 479,001,600 total rankings. The number of actual distinct grammars is clearly far less than that; the higher-ranked constraints usually carry the day, and the lower-ranked constraints, while active, do not normally interact with each other a great deal, so that the relative ranking among lower-ranked constraints makes little, if any, difference. The fact that some grammars can be learned from the monostratal initial hierarchy with only four demotion events shows that some grammars can be fully determined by hierarchies with the 12 constraints spread over 5 strata. However, the actual number of distinct grammars realizable in an optimality theoretic system, short of the factorial of the number of constraints, is entirely system-specific; there is no trivial way of counting them.

6. Discussion

The primary motivation for the larger learnability project, of which this work is a part, is to see how much the formal structure of Optimality Theory can contribute to language learnability. The iterative strategy employed is general to optimality theoretic systems. The approach is not defined in terms of the substantive details of particular constraints or candidates; it only requires that it can be supplied with candidates, along with their constraint violations, which are optimal with respect to hypothesized constraint hierarchies. Rankable constraints and evaluatable candidates are necessary components of any optimality theoretic system (by definition), and the particular optimization functions, production-directed parsing and interpretive parsing, are already motivated as being the basis for an optimality theoretic account of language production and comprehension.

One element of this work not general to all optimality theoretic systems is the assumption that the underlying form is fully apparent in the overt form. Metrical stress was selected as a domain in part to delay having to deal with the acquisition of underlying forms which are not faithfully realized in optimal descriptions. The current work focuses on the issue of underdetermination of structural descriptions by overt forms. There is, however, a proposal for extending the iterative strategy to deal with the learning of lexical underlying forms, even in the presence of alternations. The proposal is presented in (Tesar & Smolensky, to appear); an investigation of the proposal remains the topic of future work.

Another system-specific element is the specially-designed initial hierarchy used in the second set of simulations discussed in section 5. This hierarchy provided impressive learning performance on the stress grammars, but its construction was based entirely upon an analysis of the specific constraints and candidates used. The issue of the initial hierarchy, however, is not restricted to the substantive details of particular systems; it is the subject of more general proposals for both formal properties of language learning and in the optimality theoretic analysis of empirical

child language acquisition data. Further discussion of the formal issues can be found in (Smolensky, 1996b) and (Hale & Reiss, 1996). Discussion and presentation of the issues as related to empirical acquisition data can be found in (Gnanadesikan, 1995), (Demuth, 1995), (Demuth, 1996), and (Smolensky, 1996a).

While the OT system employed here captures a significant range of stress phenomena, it is not an exhaustive account of the attested stress systems of the world. the consequences for learning of a more ambitious system of grammars cannot be easily predicted in advance. However, in the overall picture of acquiring stress in particular, there are at least two other possible sources of information not addressed in this work, but which could be of great benefit to a learner. First, stress is known to interact with other phonological phenomena, such as vowel quality (Hayes, 1995). An optimality theoretic analysis would make it possible for the constraints determining stress to interact in limited ways with the constraints determining other phonological processes, allowing the learner to learn from observed consequences of those interactions via the kind of learning strategy advocated in this paper. Second, there may be useful phonetic correlates indicating aspects of metrical structure beyond the location of stresses (Prince, 1996). While such information is not contained in the more abstract overt forms used here, they would certainly be overtly available to child learners.

For language learning in general, these results suggest that there is significant benefit for learnability in the structure of Optimality Theory. The optimizing structure provides the learner with a way to not only determine that their current grammar is wrong, but to give strong indications of what changes need to be made to find the correct grammar, via the principle of constraint demotion. In particular, the speed of convergence shows the power of the iterative learning strategy, making it a quite plausible basis for an overall account of language acquisition, even if domain-specific enhancements, such as those discussed above for stress, ultimately need to be added to provide a complete account.

Endnotes

* The author would like to thank Paul Smolensky, Alan Prince, John McCarthy, Joe Pater, and B. Elan Dresher for useful discussions of this work.

References

- Demuth, K. (1995) "Markedness and the Development of Prosodic Structure". In J. Beckman, ed., *Proceedings of the North East Linguistic Society* 25, 13-25.
- Demuth, K. (1996) "Alignment, Stress, and Parsing in Early Phonological Words". In B. Bernhardt, J. Gilbert, and D. Ingram, ed., *Proceedings of the UBC International Conference on Phonological Acquisition*, 113-125.
- Gnanadesikan, A. (1995) "Markedness and Faithfulness Constraints in Child Phonology". Ms., the University of Massachusetts, Amherst.
- Hale, M., and C. Reiss. (1996) "The Initial Ranking of Faithfulness Constraints in UG". Ms., Concordia University, Montreal.
- Hayes, B. (1995) *Metrical Stress Theory*, University of Chicago Press, Chicago.
- McCarthy, J., and A. Prince. (1993) "Generalized Alignment". In G. Booij and J. Van Marle, ed., *Yearbook of Morphology*, 79-154.
- Prince, A. (1990) "Quantitative Consequences of Rhythmic Organization". In K. Deaton, M. Noske, and M. Ziolkowski, ed., *CLS26-II: Papers from the Parasession on the Syllable in Phonetics and Phonology*, 355-398.
- Prince, A. (1996) Personal Communication.
- Prince, A., and P. Smolensky. (1993) *Optimality Theory: Constraint Interaction in Generative Grammar*, Technical Report TR-2, Rutgers University Cognitive Science Center, and CU-CS-696-93, Department of Computer Science, University of Colorado at Boulder.
- Smolensky, P. (1996a) "On the Comprehension/Production Dilemma in Child Language". *Linguistic Inquiry* 27.
- Smolensky, P. (1996b) "The Initial State and 'Richness of the Base'". Ms., The Johns Hopkins University, July.
- Tesar, B. (1995) *Computational Optimality Theory*. Ph.D. thesis, University of Colorado at Boulder, June.
- Tesar B. (to appear) "An Iterative Strategy for Language Learning". *Lingua*.
- Tesar, B., and P. Smolensky (1995) "The Learnability of Optimality Theory". In *Proceedings of the Thirteenth West Coast Conference on Formal Linguistics*, 122-137.
- Tesar, B., and P. Smolensky (to appear) "The Learnability of Optimality Theory: An Algorithm and Some Basic Complexity Results", *Linguistic Inquiry*.