Modeling metrical stress acquisition through alignment constraint induction

Jeroen Breteler
Utrecht University

The paper models the acquisition of quantity insensitive metrical stress through constraint induction. A single constraint format is specified that regulates the alignment of prosodic categories. A binary and ternary foot-based prosodic hierarchy are compared in their conduciveness to learning a range of stress patterns, with clear advantages for the latter. The paper also points out the interaction between grammatical modeling and acquisition modeling with regards to the typological predictions of the grammar formalization.

Keywords: metrical stress, acquisition modeling, alignment, prosodic hierarchy, layered feet

1. Introduction

Previous approaches to acquisition in Optimality Theory often frame the learning task as a re-ranking task. Furthermore, such approaches usually stipulate an initial set of constraints or an initial ranking. The present article demonstrates an approach where the learner's grammar starts out empty and constraints can be induced (see also Hayes & Wilson 2008; Adriaans & Kager 2010). The learning task in question is the acquisition of quantity insensitive metrical stress. The model induces alignment constraints that follow a specific format, varying only in the prosodic categories they make reference to. Furthermore, two different conceptions of the prosodic hierarchy are tested for their conduciveness to learning in the present framework.

Section 2 describes the two conceptions of the prosodic hierarchy, as well as the format of the alignment constraints. Section 3 describes the learning process, including parsing and induction. Sections 4 and 5 discuss the data and testing method, respectively. The results are then presented in Section 6, with a closer consideration of the typological predictions in Section 7. The paper closes off with

a discussion of two related issues, and a conclusion. Finally, two appendices contain the forms for all tested stress patterns.

2. Prosodic structure and alignment constraints

2.1 Defining a prosodic hierarchy

Since metrical stress languages assign stress according to a prosodic structure, and constraints driving this assignment process make reference to the edges of prosodic categories, it is imperative for any model of OT-based metrical stress acquisition to give a concrete definition of a prosodic hierarchy for the learning model to work with.

A definition of a prosodic hierarchy describes a number of prosodic categories and describes how lower-order categories can be contained in higher-order ones (Selkirk 1980; Nespor & Vogel 1982). For example, a common definition of the prosodic hierarchy, up to the word level, is shown in Figure 1. Here, the Prosodic Word (PrWd) category is the top level category. It can contain several types of feet, or an 'unparsed', stressless syllable. For each of those prosodic categories, the accepted parsings are shown one level further down in the tree. A 0 stands for unstressed syllable, 2 for stressed syllable, and 1 for head-stressed syllable. Since the metrical learner is concerned with word-level stress, prosodic categories on phrase level will not be considered.

For the purposes of the present study, prosodic hierarchies also define a specificity ordering. That is, a prosodic category may be a specific, derived version of a general category. The specificity relations for the prosodic hierarchy shown in Figure 1 have been made explicit in Figure 2.

For example, for the hierarchy described in Figure 2, we define the Foot as the generic category for all types of feet, and the UnaryFoot category as a more specific version of the Foot. In the definition of Figure 2, there is no way for alignment constraints to make reference exclusively to binary feet, since a reference to the Foot
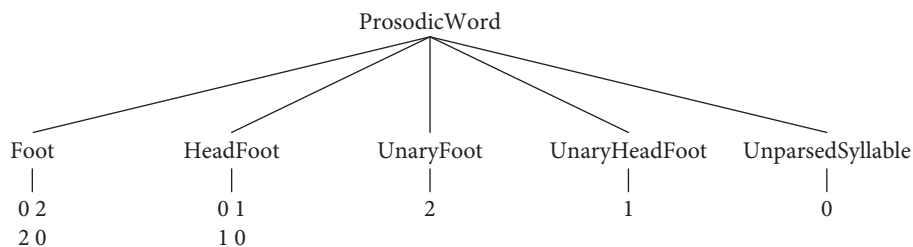

Figure 1. Possible containment relations for a traditional prosodic hierarchy
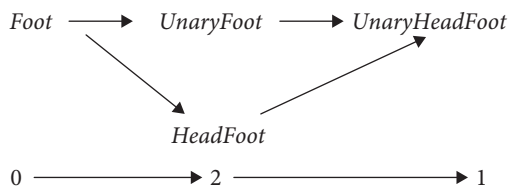
Figure 2. Specificity relations for the categories described in Figure 1

category will also include more specific derivations of Foot, such as UnaryFoot. The format of alignment constraints will be discussed in detail in Section 2.2.

The concept of specifity holds some relation to the idea of markedness; unary feet are considered more marked than binary feet, based on the typological prevalence of binary feet. However, the comparison does not hold for prosodic hierarchies with ternary feet. Hence, the specificity relation is to some extent stipulative, but has been found to be an essential part of the model in practice, as it is used in both the formulation of constraints and the parsing of overt forms, which will be described in Section 3.1.

The present article will compare the effects of two different conceptions of the prosodic hierarchy on learnability. The binary foot-based hierarchy described above was compared with a ternary foot-based hierarchy, argued for in Kager (2012), among others. The prosodic hierarchy and specificity relations for the ternary foot-based hierarchy are shown in Figures 3 and 4.
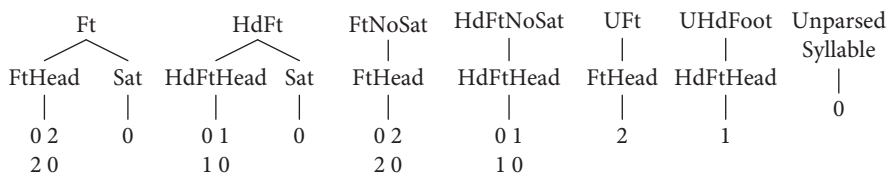

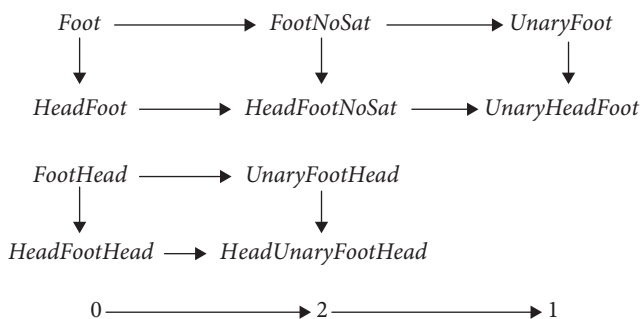**Figure 3.** Containment relations for a ternary foot-based prosodic hierarchy.

Figure 4. Specificity relations for a ternary-foot based hierarchy

2.2 Alignment constraints

The previous section stated that a prosodic hierarchy defines which elements can be contained in which others. The alignment constraints employed by the metric learner make reference to these containment relations. Concretely, for two elements A and B that have a containment relation, there are three possible scenarios:

1. *Subordinate Relation:* A is contained in B
2. *Sibling Relation:* A and B are both contained in a third element, C
3. *Domination Relation:* A contains B.

Inspired by Generalized Alignment (McCarthy & Prince 1993), an alignment constraint has three arguments: Two prosodic categories, CAT1, CAT2, and a direction, left (L) or right (R). An example of a constraint name is ALIGN-L(FOOT, PRWD). The order of the categories is important for the interpretation of the constraint. The first category determines which elements the constraint applies to; in the example, this is a constraint that states something about elements of the Foot category. The relation between the first and second category, either subordination, siblinghood, or domination, determines the interpretation of the constraint, taking into account the direction. For expository purposes, the direction is L in the description below:

– *Subordination:* For every element A of CAT1, if it is contained in an element of CAT2, it should be LEFTMOST in that element.
– *Siblinghood:* For every element A of CAT1, if there is an element to its LEFT contained in the same constituent as A, this element should be of category CAT2.
– *Domination:* For every element A of CAT1, its LEFTMOST dominated element should be of category CAT2.

In the example, the relation between Foot and PrWd elements is one of subordination (from the perspective of the Foot). Hence, the example constraint is interpreted as:

> ALIGN-L(FOOT, PRWD): For every Foot, if it is contained in a PrWd, it should be leftmost in that PrWd.

Constraints can also punish alignment behavior. For every alignment constraint as described above, there is also such a 'misalignment' version, that assigns violations in exactly the inverse way of the usual alignment constraint. Misalignment constraints are denoted by an asterisk in front of the constraint name. The misalignment version for the example constraint is *ALIGN-L(FOOT, PRWD). This constraint assigns a violation mark for every time that a Foot is leftmost within a PrWd.

An important reason for defining the alignment constraints in the above way has to do with computational tractability. An argument that started with Ellison (1994) and was expanded upon in Eisner (1997) for general OT constraints, and Buckley (2009) specifically for metrical stress, says that constraints should be formulated in a 'local' manner if the proposed phonological grammar aims to be cognitively plausible.

2.2.1   *Constraint sets*

The set of constraints that can be induced by the model depends on the prosodic hierarchy definition. For a given hierarchy, the constraint set contains constraints pertaining to every possible direct domination, siblinghood, and subordination relation between two prosodic categories. There are in fact four versions for every such relation; the constraint can be either left or right-oriented, and it can be either an alignment or misalignment constraint.

The size of the constraint set grows with the number of structural relations that a prosodic hierarchy allows. For the present experiments, the size of the set was around 170 and 300 constraints, for a binary feet and ternary feet-based hierarchy, respectively. This number was counted after removing superfluous constraints such as those regulating siblinghood for pairs of head feet, which are never applicable since no word can contain more than one head foot, or siblinghood with a regular syllable, in which case the constraint is always satisfied.

Alignment constraints that targeted head-stressed syllables specifically were excluded from these constraint families. This represents the idea that these syllables are actually not considered as separate prosodic categories from normal stressed syllables, but that their status (and phonetic realization) is simply due to the nature of the foot that they are contained in, which is the head foot of the word. However, this prevents the model from dealing with languages where the head foot form deviates from the language's default foot form. This may be troublesome when modeling quantity sensitive languages, where such behavior is attested, for example in Finnish, Tübatulabal, and Huariapano (Pruitt 2012).

3.   The learning process

3.1   Input format and parsing

The present model assumes that a learner has mastered the task of word segmentation, so that a stream of input can be processed word by word. It is also assumed that a learner can reliably distinguish the presence or absence of stress on a syllable, and the difference between first and secondary stress.

Apart from the two assumptions above, there is a parsing bias for the structure of the input forms. Recalling the definition of specificity relations in Section 2, the model will prefer structures that involve less specific prosodic categories.

An example is a three-syllable word with main stress on the second syllable. This can be schematically represented by the string '010', where 0 stands for an unstressed syllable, and 1 for a stressed syllable. The possible structural parsings of this input are shown in (1). Note that there is no parsing where the stressed syllable is in a foot by itself, since there is always room to form a binary foot, so forming a unary foot would go against the parsing bias in this case.

(1)

a. [(01)0]

ProsodicWord

HeadFoot   0

0   1

b. [0(10)]

ProsodicWord

0   HeadFoot

1   0
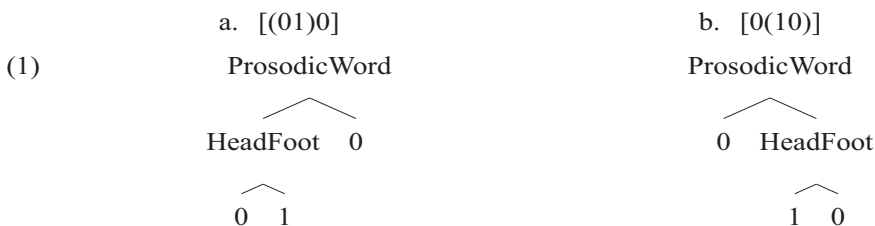
Note that for prosodic hierarchies with a distinction between more than two types of feet, this specificity bias can be implemented in two ways; it can be either specificity-avoiding, which entails minimizing the occurrence of forms that are the most specific, or generality-preferring, which entails maximizing the occurrence of forms that are the most general. These two options may give rise to different predictions, for example in the case of a prosodic hierarchy with unary, binary, and ternary feet, the latter being the least specific; a form 1020 could be parsed as either (10)(20), minimizing the specific unary foot, or (1)(020), maximizing the number of ternary feet. In the present ternary feet-based hierarchy, the bias is specificity-avoiding, i.e. it prefers structures like (10)(20).

3.2 Processing and induction

This section will describe the way that the learner processes its input, and the way that constraints are induced. The outline of the process is as follows. The model processes word tokens one at a time. Each token is parsed, and for each parsing, statistics are updated concerning the alignment behavior of all prosodic categories. After a number of tokens, the learner induces the constraints which show the most consistent alignment behavior. The induced constraints become part of the grammar, and guide the parsing process for subsequent input tokens. The grammar for the target language thus grows from a number of induction steps. The model terminates after a high number of induction steps, allowing the learner to converge on the target grammar. In the rest of this section, the process will be

described in more detail, and the final part of the section will go through a detailed example case.

The learner starts out with an empty grammar. At every time step, the learner is presented with an input token, i.e. the overt form of a word. Input tokens are randomly selected from a uniform distribution of words of length two to eight from the target language. The learner processes the token by considering all the different parsings of the overt form, taking into account the parsing bias. For each parsing, the alignment behavior is observed, and the statistics regarding this behavior are updated. Specifically, two statistics are tracked for each constraint in the constraint family: *frequency* and *obedience*. Frequency refers to the number of times that a constraint *could* possibly be violated. Sometimes, constraints are vacuously obeyed, since there is simply no target for the constraint, for example when an alignment constraint targeting unary feet is applied to a candidate form with only binary feet. Obedience refers to the number of times that a constraint actually *has* been obeyed.

Some input tokens may have more possible parsings than others. Since each input token should have the same impact on the grammar, the values for both frequency and obedience are divided by the number of different parsings for the token. In other words, if an input token has a high degree of structural ambiguity, each of its individual parsings will only play a small role in the learning process.

After a sufficient number of input tokens, the learner selects the constraint that has been the most consistently obeyed and induces it at the bottom of the ranking. The consistency score is found by dividing the obedience value of each constraint in the constraint family by its frequency value, which gives the ratio of obedience to the total number of instances. If the frequency count is zero, it means there is no context where the constraint plays a role. In such cases, the constraint is removed from the constraint set. The number of tokens after which an induction step occurs was set at 2000, which is relatively high in the sense that the impact of the random selection of input tokens is limited.

Once a constraint has been induced, it is employed in parsing the input, similarly to robust interpretive parsing (Tesar 1998). Only parsings that are optimal according to the grammar that has been induced so far will be weighed for the alignment observations. Furthermore, after an induction step, the observations are *reset*. This is because the old observations are outdated since they are based on parsings that may be excluded by the newly induced constraints. Straightforwardly resetting the observations is a simplification in the model; a way to adapt the observations to the new grammar in a more biologically plausible way might be to have the strength of an observation decay over time. This way, old, incorrect observations will simply decay to zero whereas observations that are still correct will remain high.

In practice, there may be many ties when determining the most consistent alignment constraint. In such cases, the model induces all the constraints in one single constraint stratum. The model makes use of what Boersma (2003) refers to as 'crucial ties'; the violation marks of constraints in the same stratum are added together and the constraints are evaluated as one constraint. Using constraint strata allows the learner to postpone decisions on frequent but ambiguous phenomena until more information is available from other observations. For example, consider a language with the two forms 10 and 010. This translates to the parsings (10), 0(10) and (01)0, where the parsing of the two-syllable form weighs twice as heavily as either of the three-syllable forms' parsings. Since languages prefer a single foot type, this language should be analyzed as trochaic, with feet aligning to the right edge of the word. However, the short length of the 2-syllable form also allows left-alignment with the word, clouding the alignment statistics of the learner. The learner finds a 75% obedience ratio for ALIGN-L(FOOT, PRWD), ALIGN-R(FOOT, PRWD), and ALIGN-L(STRESSEDSYLLABLE, FOOT), and a 25% obedience ratio for the iambic footform constraint, ALIGN-R(STRESSEDSYLLABLE, FOOT). If a constraint were randomly induced, there is a one in three chance of selecting ALIGN-L(FOOT, PRWD), a constraint that goes against the analysis provided above. By inducing all three constraints, any malicious influence of the word-alignment constraints is neutralized, and the impact of the trochaic foot-form constraint ALIGN-L(STRESSEDSYLLABLE, FOOT) will cause the learner to prefer 0(10) over (01)0, as desired.

The model is terminated after 95000 tokens have been processed. This number of tokens was found to be enough for every language tested to converge on a grammar that produces only one stress placement for every word length, whether correct or incorrect. Many languages converged much faster, but since induced constraints are always added to the bottom of the ranking, the predictions in such cases do not change by processing more tokens.

3.3 Example case

To demonstrate the details of the process described above, this section will consider an example with a toy language and a toy constraint set in a binary foot-based hierarchy. The toy language will consist of the single form 10202. The example will assume an empty starting grammar. Furthermore, the example will consider the induction of three constraints: ALIGN-L(STRESSEDSYL, FT) which prefers stressed syllables at the left foot edge, *ALIGN-L(PRWD, UNARYHDFT) which is violated when a unary head foot is at the left word edge, and *ALIGN-L(PRWD, HDFT), which is violated when any head foot is at the left word edge.

At every learning instance, the model is presented with a word. Since there is only one form in the toy language, the model is presented with the form 10202.

Under the parsing bias, forms should be specificity-avoiding. There are only two unstressed syllables for three stressed ones, so the form will contain two binary feet and one unary foot. This leaves room for three possible parsings: (10)(20)(2), (10)(2)(02), and (1)(02)(02). Since the grammar is empty, all parsings are equally harmonic, so none are discarded by the robust interpretive parsing step.

At the end of the parsing process, three optimal parsings remain. These are the parsings that the model will learn from. For each of the constraints, a frequency and obedience score are calculated. Table 1 shows the calculated scores. Notice that for the ALIGN-L(STRESSEDSYL, FT) constraint, there are three instances in which the constraint can possibly be violated in each parsing, since there are three stressed syllables. Furthermore, the *ALIGN-L(PRWD, UNARYHEADFT) constraint is a more specific version of the *ALIGN-L(PRWD, HEADFT). The former manages to escape the violations of the latter because in two of the forms, the head foot is not unary.

Table 1. Obedience/Frequency ratios for each of the parsings and constraints. Constraint names are abbreviated

	(10)(20)(2)	(1)(02)(02)	(10)(2)(02)	Average
L(StressSyl, Ft)	3/3	1/3	2/3	2/3
*L(Wd, UnaryHdFt)	1/1	0/1	1/1	2/3
*L(Wd, HdFt)	0/1	0/1	0/1	0

For each constraint, the score is averaged over the different parsings and stored. At this point, the model might process more input tokens before inducing more constraints, but for the purpose of the demonstration we will immediately move to an induction step.

For all three constraints, the obedience ratio is calculated by dividing the obedience score by the frequency. In the example, this means that the scores are equal to the values in the 'Average' column of Table 1.

Two constraints are tied for best obedience ratio. This means that both constraints will be induced into the grammar as a single stratum. The induced constraints go at the bottom of the constraint ranking, but so far, the grammar was empty. The induced constraints are removed from the constraint set, which now only consists of *ALIGN-L(PRWD, HDFT). This constraint might be induced at a later stage. After the induction step, the frequency and obedience scores are reset.

The effect of the newly induced constraints can be seen by parsing the 10202 form again; there is now only a single form left, as shown in Table 2. This means that in future parsing steps, the model will only consider the parsing (10)(20)(2) for learning.

Table 2. The newly induced stratum rules out two parsings; only one optimal parsing remains

10202	*Align-L(UnaryHdFt,Wd)	Align-L(Stress,Ft)
☞ a. (10)(20)(2)		
b. (10)(2)(02)	*!	
c. (1)(02)(02)	*!	**

4. Testing

After the learning phase is terminated, the model's state can be assessed to see if a correct grammar for the target language has been acquired. The model runs a typical Optimality Theory production process for each word length from two to eight. For each length, the model considers all possible parsings of all possible forms, even those with stress placements that did not occur in the input data. This reflects OT's Richness of the Base and Freedom of Analysis assumptions (McCarthy & Prince 1993). Using the acquired grammar, the optimal form or forms for each word length are derived.

For the present experiments, a learning trial was considered successful if for all the optimal forms, stress placements correspond with the overt forms of the target language. That is, no presumption is made about the structure that the model 'should' learn, and even structural ambiguities in the output are acceptable as long as the model comes to produce the same stress placements as speakers of the target language would.

Since there is some randomness in the distribution of the input tokens, it is possible that the model gives different results for the same language in different trials. To make sure that the outcome of the learning process is not too sensitive to chance, testing should be run with multiple trials for each language. In the present experiments, each language was tested with 5 trials.

5. Data

Language data is taken mainly from Gordon (2002) and Elenbaas & Kager (1999), with the addition of Estonian data reported in Prince (1980). Crucially, a number of patterns with ternary rhythm have been included, since these are expected to benefit from the ternary foot-based hierarchy the most. All tested patterns are quantity insensitive; they do not distinguish between heavy and light syllables.

Many of the included languages have 'mirror cases': A pattern that is symmetrical or almost symmetrical (e.g. only a difference in the edge orientation of the head foot). Consequently, although only 14 patterns are listed, the result can be generalized to at least a similar number of other patterns. It is not uncommon for several tens of languages to share the same pattern, so the coverage in terms of number of languages runs into the hundreds.

6.   Results

The success ratios out of five trials for both binary and ternary foot-based hierarchies for all languages are shown in Table 3. For all languages, a ternary foot-based hierarchy either matches or outperforms a binary foot-based one. Instances where the ternary foot-based hierarchy shows an improvement over the binary foot-based hierarchy are indicated in bold.

There is one case of partial success in the results: Araucanian in a binary foot-based hierarchy. This language is traditionally analysed as having iambs from the left. In words of an odd-numbered length, these forms are highly structurally ambiguous, and this perhaps leads the model to incorrect conclusions in some cases. A similar problem occurs with Cayuvava, which builds long (ternary) trochees from the right, causing a similar ambiguity, even in ternary foot-based hierarchies.

Both models also perform quite poorly on bidirectional patterns. This may in part be due to structural ambiguity again, especially in the case of Piro. Although the Piro pattern is highly similar to that of Garawa, the latter is more easily acquired because the structural ambiguity is centered around the head foot, which can be more easily targeted by head foot-specific constraints. The Armenian case is problematic because the grammar set is too restricted; for the present experiments, the model could not form constraints directly relating the prosodic word category to syllable-level categories, which would have enabled the model to align a stressed syllable with both edges. As discussed earlier, the model was also not allowed to model foot forms separately for head feet, which might alternatively have allowed the grammar to capture the pattern.

7.   Typological predictions

The typological predictions of a grammatical formalization in Optimality Theory are often interpreted by considering a factorial typology, i.e. the set of outputs predicted through all different constraint rankings. However, the induction learning mechanism described here puts additional restrictions on the acquisition process.

Table 3. Success rates for the learning model for binary and ternary feet-based hierarchies on unbounded, binary, ternary, and bi-directional stress patterns

Language	Success Rate	Success Rate	Comment
Yakut	100%	100%	Final stress, mirror case: Afrikaans
Arapesh	100%	100%	Prefinal stress, m.c.: Lakota
Macedonian	**0%**	**100%**	Antepenultimate stress, m.c.: Kashaya
Pintupi	100%	100%	Binary rhythm
Araucanian	**40%**	**100%**	Binary rhythm, iambs from the left
Passamaquoddy	100%	100%	Binary rhythm, allows unary feet
Murinbata	100%	100%	Binary rhythm, allows unary feet
Cayuvava	0%	0%	Ternary rhythm, `trochees' from right
Estonian	100%	100%	Ternary Rhythm, allows binary feet
Ioway-Oto	**0%**	**100%**	Ternary Rhythm, allows binary feet
Tripura Bangla	**0%**	**100%**	Ternary Rhythm, allows binary feet
Piro	0%	0%	Head trochee right, trochees from left
Garawa	100%	100%	Head trochee left, trochees from right
Armenian	0%	0%	Stress at both edges

Some languages can be represented by a ranking of alignment constraints, but cannot be learned. Consequently, the learning model explains typological gaps in the grammar formalization, and the typological predictions of the grammatical formalization should be analysed in conjunction with the restrictions placed by the learning model (Boersma 2003; Heinz 2009).

For lack of a formal way of determining the limits of the induction model, Table 4 presents 'manual' tests for a number of patterns that could be modeled using alignment constraints, but that are unattested and thought to be unnatural. Each of the languages is described in prose below.

– Unaries Interleaved: Alternates binary and unary feet.
– Quaternary Stress: Stress every four syllables.
– Odd-Even Stress Flop: Stress on the left of the word if the leftmost foot is binary, else on the right.
– Unary with Binary Right-Edge: All unary feet, except for a binary foot on the right edge of the word.

Most patterns, while grammatically possible, cannot be learned by the model. The exception is the final pattern; this indicates that the model or grammar formalization may be in need of extra provisions to rule out the rampant use of unary feet.

Table 4. Success rates for the learning model for binary and ternary feet-based hierarchies for a set of non-natural languages

Language	Success Rate	Success Rate	Example Form
Unaries Interleaved	0%	0%	1022022
Quaternary Stress	0%	0%	1000200
Odd-Even Stress Flop	0%	0%	102020 vs. 2002010
Unary with Binary Right-Edge	100%	100%	2222210

8.   Discussion

8.1 The need for locality restrictions

Recently, Kornai (2009) questioned the line of argument pursued since Ellison (1994) on locality. Kornai argues that the mathematical technique of amortized analysis, used to argue in favor of locality restrictions, does not hold for the limited domains present in phonology. If this reasoning is correct, then the grammatical formalizations employed in this study have been unnecessarily restricted. It would be interesting to use the alignment observations for a different induction mechanism that can employ more powerful alignment constraints, which paint a tighter typological picture.

8.2 Learnability definition

The term 'unlearnable' has been used here to refer to stress systems whose surface forms do not lead the learning model to a grammar that will output the same surface forms again. However, that does not guarantee that such a grammar is unlearnable in the strictest sense of the word; it is possible that some grammars cannot be learned by the model through surface forms of an identical grammar, but can be learned by forms from a different grammar. Consequently, rather than 'unlearnable', we might refer to such grammars as 'unstable' or 'intransferable'. However, if there exists a grammar A whose surface forms lead the model to learning a different, unstable grammar B, then A is itself also unstable. Without some form of linguistic mutation then, it is impossible for the learning model to arrive at an unstable grammar. Since such mutations are not part of the present modeling approach, we can consider unstable grammars to be unlearnable for all practical purposes.

9. Conclusion

In conclusion, the paper has demonstrated an approach to learning in the Optimality Theory framework that uses alignment constraint induction instead of reranking. In the presented framework, a ternary foot-based hierarchy was found to be beneficial to learning despite its more complicated prosodic hierarchy and constraint family. The model has proven to be able to learn most stress patterns presented to it, while excluding most of the unattested patterns. Crucially, this exclusion was not due only to the grammatical formalization, but was for a large part determined by the nature of the learning model. Further research will have to show if the present methods can be applied to a wider range of phenomena, the clearest next objective being the acquisition of quantity sensitive stress.

References

Adriaans, Frans & René Kager. 2010. Adding generalization to statistical learning: The induction of phonotactics from continuous speech. *Journal of Memory and Language* 62 (3). 311–331.

Boersma, Paul. 2003. Bruce Tesar and Paul Smolensky (2000). Learnability in optimality theory. Cambridge, Mass.: MIT press. *Phonology* 20 (3). 436–446.

Buckley, Eugene. 2009. Locality in metrical typology. *Phonology* 26 (3). 389–435.

Eisner, Jason. 1997. Efficient generation in primitive optimality theory. In *Proceedings of the Eighth Conference on European chapter of the Association for Computational Linguistics,* 313–320. Association for Computational Linguistics.

Elenbaas, Nine & René Kager. 1999. Ternary rhythm and the lapse constraint. *Phonology* 16 (3). 273–329.

Ellison, T. Mark. 1994. Phonological derivation in optimality theory. In *Proceedings of the 15th Conference on Computational Linguistics. Volume 2,* 1007–1013. Association for Computational Linguistics.

Gordon, Matthew. 2002. A factorial typology of quantity-insensitive stress. *Natural Language & Linguistic Theory* 20 (3). 491–552.

Hayes, Bruce & Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39 (3). 379–440.

Heinz, Jeffrey. 2009. On the role of locality in learning stress patterns. *Phonology* 26 (2). 303–351.

Kager, René. 2012. Stress in windows: Language typology and factorial typology. *Lingua* 122 (13). 1454–1493.

Kornai, András. 2009. The complexity of phonology. *Linguistic Inquiry* 40 (4). 701–712.

McCarthy, John J. & Alan Prince. 1993. Generalized alignment. *Yearbook of Morphology* (1993). 79–153.

Nespor, Marina & Irene Vogel. 1982. Prosodic domains of external sandhi rules. *The Structure of Phonological Representations* 1. 225–255.

Prince, Alan. 1980. A metrical theory for Estonian quantity. *Linguistic Inquiry* 11 (3). 511–562.

Pruitt, Kathryn R. 2012. *Stress in Harmonic Serialism*. Amherst, MA: University of Massachusetts Amherst dissertation.

Selkirk, Elisabeth O. 1980. The role of prosodic categories in English word stress. *Linguistic Inquiry* 11 (3). 563–605.

Tesar, Bruce. 1998. An iterative strategy for language learning. *Lingua* 104 (1). 131–145.

Appendix. Natural language input tokens

Afrikaans

# of Syllables	Form
2	10
3	100
4	1000
5	10000
6	100000

Lakota

# of Syllables	Form
2	01
3	010
4	0100
5	01000
6	010000

Macedonian

# of Syllables	Form
2	10
3	100
4	0100
5	00100
6	000100
7	0000100

Pintupi

# of Syllables	Form
2	10
3	100
4	1020
5	10200
6	102020
7	1020200
8	10202020
9	102020200

Araucanian

# of Syllables	Form
2	01
3	010
4	0102
5	01020
6	010202

Passamaquoddy

# of Syllables	Form
2	10
3	210
4	2010
5	22010
6	202010
7	2202010

Murinbata

# of Syllables	Form
2	10
3	102
4	1020
5	10202
6	102020

Cayuvava

# of Syllables	Form
2	10
3	100
4	0100
5	00100
6	200100
7	0200100
8	00200100

Tripura Bangla

# of Syllables	Form
2	10
3	100
4	1000
5	10020
6	100200
7	1002000
8	10020020

Estonian

# of Syllables	Form
2	10
3	100
4	1020
5	10020
6	100200
7	1002020
8	10020020
9	100200200

Ioway-Oto

# of Syllables	Form
2	01
3	010
4	0100
5	01002
6	010020
7	0100200
8	01002002
9	010020020

Piro

# of Syllables	Form
2	10
3	010
4	2010
5	20010
6	202010
7	2020010
8	20202010

Garawa

# of Syllables	Form
2	10
3	100
4	1020
5	10020
6	102020
7	1002020
8	10202020

Armenian

# of Syllables	Form
2	21
3	201
4	2001
5	20001
6	200001
7	2000001
8	20000001

Appendix. Natural language input tokens

Binary-Unary Alternation

# of Syllables	Form
2	10
3	102
4	1022
5	10220
6	102202
7	1022022
8	10220220

Odd-Even Stress Flop

# of Syllables	Form
2	10
3	100
4	1020
5	20010
6	102020
7	2002010
8	10202020

Trochaic Quaternary Stress	
# of Syllables	Form
2	10
3	100
4	1000
5	10002
6	100020
7	1000200
8	10002000

Unaries with Binary Right-Edge	
# of Syllables	Form
2	10
3	210
4	2210
5	22210
6	222210
7	2222210
8	22222210

Author's address

Jeroen Breteler
Utrechtseweg 87
3702 AA Zeist
The Netherlands

jmwbreteler@gmail.com