Comparing Two Optimality-Theoretic Learning Algorithms for Latin Stress

Diana Apoussidou and Paul Boersma
University of Amsterdam

1. Introduction

This paper compares the performance of two formal Optimality-Theoretic learning algorithms in modelling the acquisition of Latin stress from overt language data: Error-Driven Constraint Demotion (EDCD) and the Gradual Learning Algorithm (GLA). We present computer simulations of learners who are trained on several kinds of overt Latin stress patterns: a case with main stress only, three cases with overtly available secondary stress, and a case in which the learners are free to invent their own secondary stress patterns. Several of these cases turn out to be learnable with the GLA, none with EDCD.

2. Latin stress: overt forms

As we see in (1), Latin stress assignment follows a clear-cut rule: stress the penultimate syllable if it is heavy, else the antepenultimate syllable.

(1) Examples

misericordia	[mi.se.ri.kór.di.a]	[L L L H1 L L]
domesticus	[do.més.ti.kus]	[L H1 L H]
rapiditas	[ra.pí.di.ta:s]	[L L1 L H]
perfectus	[per.fék.tus]	[H H1 H]
incipio	[iŋ.kí.pi.o:]	[H L1 L H]
amice	[a.mí:.ke]	[L H1 L]
homo	[hó.mo:]	[L1 H]

The weight of a syllable is determined by its segments: a syllable ending in a short vowel is light (“L”), a syllable ending in a long vowel or a consonant is heavy (“H”). For each word, the table shows the written form, the syllabification (with main stress indicated by an accent), and the overt pattern of heavy and light syllables (with main stress indicated by “1”). Further on in this paper, these word-sized sequences of light and heavy syllables, sometimes with additional markings for secondary stress, will form the language data that we will feed to our simulated learners.

 WCCFL 23 Proceedings, ed. B. Schmeiser, V. Chand, A. Kelleher and A. Rodriguez, pp. 101-114. Somerville, MA: Cascadilla Press.

3. Abstract analyses

When trying to make sense of the overt data in (1), linguists have proposed analyses in terms of hidden prosodic structures, both smaller and larger than the syllable. Jakobson (1937/1962) proposed the subsyllabic structure of the *mora*, of which light syllables contain one, heavy syllables two; in his view, Latin stress falls on the syllable that contains the second mora before the final syllable. Hayes (1981, 1995), Mester (1994), Prince & Smolensky (1993) and Jacobs (2000) invoke the suprasyllabic structure of the *foot*; in their view, the final syllable is discarded as *extrametrical*, after which a bimoraic foot with a trochaic rhythm is assigned on the right. Overt forms are depicted in square brackets, surface forms in slashes.

(2) Foot structure
a) [ra.pí.di.ta:s] /L **(L1 L)** <H>/ b) [per.fék.tus] /H **(H1)** <H>/

However, there are forms in which bimoraicity and right alignment cannot both be satisfied at the same time. For instance, how would you assign a foot to *misericordia*?

(3) [mi.se.ri.kór.di.a]
a) /L L L **(H1)** L <L>/ or b) /L L L **(H1 L)** <L>/ ?

In (3a), bimoraicity is obtained by parsing the main-stressed heavy syllable into a foot, but a light syllable is skipped in the metrical construction, leaving in effect two syllables extrametrical. In (3b), right alignment is achieved, but bimoraicity is violated by the construction of a trimoraic foot. Both (3a) and (3b) have been defended in the literature.

Foot structure is a metrical construct, not traceable in the phonetic signal. That is, while our learners can directly spot the main stress placement, they cannot directly read the foot structure from the overt language data. In this paper we follow Tesar & Smolensky's (1998, 2000) proposal that learners use their OT grammar to construct the foot structure by themselves from the overt stress patterns. The question, then, is: will our Latin learners have problems assigning a foot structure compatible with the overt forms? And what foot structure will they assign to *misericordia*?

4. The grammar

4.1. Twelve structural constraints

We will equip our virtual learners with the same 12 metrical constraints that Tesar & Smolensky (2000) used in their simulations of the stress systems of 124 languages. Eleven of these constraints are listed in (4).

(4) All-Feet-Left/Right (AFL, AFR): feet are left/right aligned with the left/right word edge.
Main-Left/Right (Main-L, Main-R): the foot with main stress is left/right aligned with the left/right word edge.
Word-Foot-Left/Right (WFL/WFR): the left/right word edge has to be aligned with a foot.
Nonfinal: the final syllable is not included in a foot.
Parse: every syllable is parsed into a foot.
Weight-to-Stress Principle (WSP): heavy syllables are stressed.
FootBin: feet are binary, i.e. consist of two syllables or two morae.
Iambic: the stressed syllable is final in its foot.

4.2. The 12th constraint: trochaicity

For a trochaic rhythm pattern, two possibilities can be found in the literature: FootNonfinal (Tesar & Smolensky 2000), or Trochaic (e.g. Jacobs 2000 for Latin).

(5) FootNonfinal: the stressed syllable is not final in its foot.

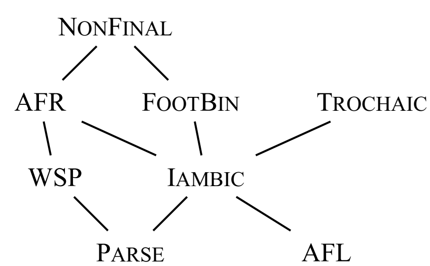
(6) Trochaic: the stressed syllable is initial in its foot.

FootNonfinal punishes candidates with monosyllabic feet, while Trochaic is satisfied in monosyllabic feet (as is Iambic). We will perform simulations both with Trochaic and with FootNonfinal.

4.3. A possible crucial ranking for Latin

A possible crucial ranking for Latin main stress, explained in detail by Apoussidou & Boersma (2003), is given in (7).

(7) A crucial ranking



This ranking generates the main-stress forms in (1), with foot structures as in (2) and the uneven trochees of (3b).

5. Acquisition

In OT approaches of language acquisition the task for a learner is to find an appropriate constraint ranking for the target language, i.e. a ranking that reproduces the same patterns in her outputs that she encounters in her language environment. Spoken in terms of stress, she has to come up with a ranking of the metrical constraints that produces the same foot structures that (the learner thinks) the adults produce. However, the information to rank the constraints can only be taken by the child from the impoverished overt language data. That brings the notion of *perception* into play.

5.1. Perception

In Tesar & Smolensky's (1998, 2000) proposal of *robust interpretive parsing*, the information of the overt language data is processed as follows. The learner hears an *overt form* (OF), which in general could be a phonetic speech signal (Boersma 1998), but in Tesar & Smolensky's simplified case (which we follow in this paper) it is actually the overt stress pattern, i.e., the learner is assumed to already have divided up the string into syllables and words and to know what a light or heavy syllable is. By use of her current grammar (the same constraint ranking that she uses for production) she maps the signal onto a *surface form* (SF) that contains foot structure. As a result, the interpreted foot structure is determined partly by the stress information in the overt input, partly by the grammar. With Boersma (1998: 269), we will simply call the OF-to-SF mapping *perception*, in order to make it explicit that low-level processes like phonetic categorization and high-level processes like foot assignment are parts of one and the same mapping from raw auditory data to abstract phonological structures. In (8) we show a schematized picture of the perception process.

(8) Perception

$$\boxed{\begin{array}{c}\text{OF}\\ \text{[H1 L]}\end{array}} \xrightarrow{\text{perception}} \boxed{\begin{array}{c}\text{SF}\\ \text{/(H1 L)/}\end{array}}$$

What does this schematized process look like in an OT tableau? Imagine a child that at some point during acquisition happens to have a grammar with the ranking in (9). What surface structure will she impose on an overt form like, say, [H1 L]?

Perception in an OT tableau differs in a few respects from the traditional OT production tableaus. The input to perception is not the underlying form, as it is in production, but the overt form. Here, the input has information about syllable weight and stress. The output candidates are the possible perceived surface structures. In (9) there are only two candidates. GEN cannot generate more candidates, because (by Tesar & Smolensky's assumption, which we follow) every stressed syllable has to be licensed by a foot and every foot has to contain a stressed syllable. A candidate like /H1 (L)/, therefore, is not generated. Another assumption here is that all stress information of the overt input has to appear in the surface candidates, so there are no candidates like /H L/ or /(H1) (L2)/.[1]

The evaluation procedure for a perception tableau is the same as for a production tableau. The second candidate in (9) is ruled out by MAIN-R, since the foot with main stress is not aligned with the right word edge.

(9) Perceiving a disyllabic trochee[2]

overt form [H1 L]	MAIN-L	MAIN-R	PARSE	AFL	AFR	FOOTBIN	IAMBIC	TROCHAIC	WSP	NONFINAL
☞ /(H1 L)/							*			*
/(H1) L/		*!	*		*					

The comprehension process is not restricted to perception alone. The listener will also infer an *underlying form* (UF), i.e. the form as it is stored in the lexicon. In Tesar & Smolensky's model of metrical phonology, the underlying form can be derived from the surface form in a trivial way, namely by erasing foot boundaries and stress marks. This is possible for languages that have grammatically assigned stress only, i.e. languages without lexically assigned stress. Since Latin is such a language, we can follow Tesar & Smolensky in this simplified view of the SF-to-UF mapping (which Boersma 1998 calls *recognition*). The surface form /(H1 L)/ will thus simply be recognized as the lexical entry |H L|, as shown in (10).

(10) Comprehension

OF [H1 L] —perception→ SF /(H1 L)/ —recognition→ UF |H L|

1. We leave out issues of OF-SF faithfulness, i.e., we ignore SF candidates that have a stress pattern different from the OF. See §6.3 for a relaxation of this restriction, by which /(H1) (L2)/ becomes a third candidate.
2. The two constraints WFL and WFR are left out of all our tableaus for the sake of readability.

5.2. Production

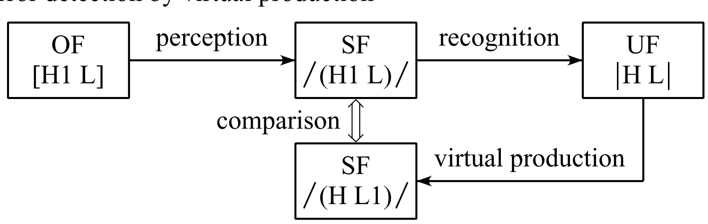
Comprehension is not enough to learn from. A child needs to compare the perceived form to something else in order to change her grammar. This something else is the learner's produced surface structure, as Tesar & Smolensky (1998, 2000) and Boersma (1998) propose: every time the learner interprets an overt form, she computes the corresponding *virtual production*, i.e. the form she herself would say given the underlying form she has recognized. For our example of the overt form [H1 L], which led to underlying |H L|, the computation of the virtual production is shown in (11), a conventional production tableau where the underlying form is the input to the evaluation. Applying the same ranking as in perception, an iambic form /(H L1)/ is chosen as the winner.

(11) The virtual production of an iamb

underlying form |H L|	MAIN-L	MAIN-R	PARSE	AFL	AFR	FOOTBIN	IAMBIC	TROCHAIC	WSP	NONFINAL
/(H1 L)/							*!			*
/(H1) L/		*!	*		*					
/(H1)(L2)/		*!		*	*	*				*
/H (L1)/	*!		*	*		*			*	*
☞ /(H L1)/								*	*	*
/(H2)(L1)/	*!			*	*	*				*

Comparing this output to the perceived form, we see a discrepancy between the perceived surface structure (the uneven trochee /(H1 L)/) and the produced surface structure (the iamb /(H L1)/). And that is exactly what the learner does in Tesar & Smolensky's (1998, 2000) and Boersma's (1998) proposals: comparing these two forms and labelling the discrepancy as an *error*. The whole process is summarized in (12).

(12) Error detection by virtual production



The general task of the language learner, now, is to reproduce the same surface structure that she thinks she heard the adult produce. Note that this interpreted adult structure may be different from real adult structures; real adults may have /(H1) L/, with final-syllable extrametricality, but Tesar & Smolensky's point is that this does not matter: the child may learn even from partially correct inferred structures.

The learner, then, will strive to bring the two surface forms into conformity by adjusting the produced form she deems incorrect to the perceived form she deems correct. This is done by constraint reranking. In the next two sections we discuss two implementations of this idea.

5.3. Grammar adjustment: Error-Driven Constraint Demotion

In EDCD (Tesar 1995), constraints can only be *demoted*, i.e. moved downwards in the hierarchy. This is triggered by error detection: if the learner observes a mismatch between her perceived form and her produced form, she will rerank the constraints. All the constraints that favour the virtually produced ('incorrect') form are demoted just below the constraints that favour the perceived ('correct') target form. In this way the learner will make it more likely that at the next evaluation of the underlying form the target form will win. In our test case IAMBIC is demoted below TROCHAIC, onto the same stratum as WSP. In (13), '√' depicts the winner of the perception tableau, while '☞' depicts the winner of the production tableau.

(13) Error-Driven Constraint Demotion

underlying form |H L|	MAIN-L	MAIN-R	PARSE	AFL	AFR	FOOTBIN	IAMBIC	TROCHAIC	WSP	NONFINAL
√ /(H1 L)/							*!			*
/(H1) L/		*!	*		*					
/(H1) (L2)/		*!		*	*	*				*
/H (L1)/	*!		*	*		*			*	*
☞ /(H L1)/								*	*	*
/(H2) (L1)/	*!			*	*	*				*

How has this grammar adjustment changed the learner's behaviour? The perceived form for [H1 L] will still be /(H1 L)/, as we can see when demoting IAMBIC below TROCHAIC in (9). The produced form for |H L| will be /(H1 L)/, as we can see by demoting IAMBIC below TROCHAIC in (11).

In EDCD there is no possibility to promote a constraint again, once it has been demoted. Constraint demotion goes on until a grammar is reached that generates the same surface forms in perception and production.

5.4. Grammar adjustment: The Gradual Learning Algorithm

Another strategy of reranking constraints in the learning process is the GLA (Boersma 1997): to gradually shift the constraints up or down the hierarchy, depending on the evidence in the input. The ranking steps in this model are much smaller than in EDCD. Moreover, the constraints may overlap. This has the consequence that the borders between strata fade, and different outputs become possible, with a probability depending on the extent of overlap. Again, the reranking of constraints is triggered by error detection. Applied to our test case, the error leads to a demotion of IAMBIC down in the hierarchy, and at the same time to a promotion of the constraints punishing the produced form, as shown in (14). But since this algorithm is gradual, the constraint shifts do not lead instantly to a full reversal of the constraints. Instead, when the child evaluates this UF in the future, it will become possible that she may utter the target form /(H1 L)/ in some instances, while still using her old form most of the time.

(14) The Gradual Learning Algorithm

underlying form |H L|	MAIN-L	MAIN-R	PARSE	AFL	AFR	FOOTBIN	IAMBIC	TROCHAIC	WSP	NONFINAL
√ /(H1 L)/							*!			*
/(H1) L/		*!	*		*					
/(H1) (L2)/		*!		*	*	*				*
/H (L1)/	*!		*	*		*			*	*
☞ /(H L1)/								*	*	*
/(H2) (L1)/	*!			*	*	*				*

With a new encounter of the target form (or other forms) the constraints will move further along the ranking scale, leading ultimately to a full reversal of the constraints. Constraints that were demoted in one learning instance can be promoted again in another, depending on the evidence in the input.

6. Simulations

So much for the learning process. Up to now we have explained two different strategies OT learners can employ. Now we turn to the metrical pattern of a natural (although dead) language, Latin. Our first simulation deals with main stress only. Since it is controversial whether Latin had secondary stress, and if so, what it exactly looked like, we felt free to design several different data sets with secondary stress. We then tested whether our virtual learners were able to learn from these data sets, provided with the basic metrical constraint sets. For each simulation, we created a number of virtual EDCD and GLA learners, with constraint sets that contained either TROCHAIC or FTNONFINAL. The initial ranking was all-constraints-equal (100.000). The training data were two to four syllables long and drawn randomly with equal probability from 28 possible overt forms. The EDCD learners were fed with 1,000 data pieces, while the GLA learners were fed with 10,000 up to 40,000 data pieces.[3] All of the simulations were carried out with the PRAAT program (www.praat.org). The first simulation was carried out with the main-stress-only data set in (15), and with the constraint set containing TROCHAIC.

(15) Simulation 1: 28 main-stress-only overt forms

Disyllables	Trisyllables	Quadrisyllables	
[L1 L]	[L1 L L]	[L L1 L L]	[H L1 L L]
[L1 H]	[L1 L L]	[L L1 L H]	[H L1 L L]
[H1 L]	[L H1 L]	[L L H1 L]	[H L H1 L]
[H1 H]	[L H1 H]	[L L H1 H]	[H L H1 H]
	[H1 L L]	[L H1 L L]	[H H1 L L]
	[H1 L H]	[L H1 L H]	[H H1 L H]
	[H H1 L]	[L H H1 L]	[H H H1 L]
	[H H1 H]	[L H H1 H]	[H H H1 H]

The results are shown in (16). All of the EDCD learners failed to converge upon a correct grammar. The crucial problem here was the ranking of IAMBIC, WSP and TROCHAIC: the data led the three constraints to tumble down below each other over and over again. After some learning the incorrect forms /(L L1) X/[4] appeared, which brought about a reranking of IAMBIC below TROCHAIC. After that, forms like /(H1 H) X/ appeared, leading to a demotion of TROCHAIC below IAMBIC and WSP again. By contrast, all GLA learners were able to create a correct constraint ranking from the data. The analysis that the learners came up with was: all final syllables were extrametrical; feet were assigned from right to left; feet were binary at some level of analysis (thus allowing uneven trochees of the form (H1 L)); heavy syllables attracted stress; and feet were trochaic.

3. GLA learners need more data pieces because their reranking step (*plasticity*) of 0.1 is much smaller than the *evaluation noise* of 2.0. It turned out that EDCD learners were able to learn either from the 1,000 data pieces or not at all, while GLA learners were sometimes late: most of them converged onto a correct grammar after 10,000 data pieces, but some only after 40,000.

4. "X" stands for either a light or a heavy syllable in final position.

(16)

An EDCD learner		A GLA learner	
Constraints	Ranking values	Constraints	Ranking values
FOOTBIN	100.000	NONFINAL	115.651
NONFINAL		AFR	110.638
AFR	99.000	FOOTBIN	106.635
MAIN-R		WSP	105.462
PARSE		TROCHAIC	105.460
WFR		IAMBIC	101.955
AFL	98.000	MAIN-R	100.200
MAIN-L		AFL	100.193
WFL		MAIN-L	93.301
IAMBIC	–157.000	WFR	84.349
WSP		WFL	80.189
TROCHAIC	–158.000	PARSE	78.679

Examples for the produced forms under the constraint ranking of the GLA learner in (16) are /L H (H1) L/ as in *vo.lup(tá:)te:s*, /H (L1 L) H/ and /H (H1 L) L/. No GLA learner would generate strictly bimoraic forms like /(H1) L L/ (as proposed by e.g. Mester 1994): this analysis is simply not possible with the implemented constraints.

6.1. Simulation 2: very weight-sensitive secondary stress

One possibility of secondary placement would be to stress every heavy syllable before the main-stressed one, as in (17) (see Apoussidou & Boersma 2003 for a more thorough discussion of this simulation). Disyllables were also used in the simulations, but they are suppressed in the table since they do not differ from those in (15).

(17) Very weight-sensitive secondary stress

Trisyllables		Quadrisyllables		
[L1 L L]	[H2 H1 L]	[L L1 L L]	[L H2 H1 L]	[H2 H1 L L]
[L1 L L]	[H2 H1 H]	[L L1 L H]	[L H2 H1 H]	[H2 H1 L H]
[L H1 L]		[L2 L H1 L]	[H2 L1 L L]	[H2 H2 H1 L]
[L H1 H]		[L2 L H1 H]	[H2 L1 L L]	[H2 H2 H1 H]
[H1 L L]		[L H1 L L]	[H2 L H1 L]	
[H1 L H]		[L H1 L H]	[H2 L H1 H]	

EDCD learners training with these 28 overt forms failed again: with TROCHAIC they produce initially stressed forms like */(L1 L) L L/, and with FTNONFINAL they produce forms like */(H2) (L1 H) L/. Again, the GLA learners training with the same data succeeded, independently from the constraint set they were using. GLA learners with TROCHAIC produced forms like /(L H1) L/, and /(L H2) (H1) H/ as in *(vo.lùp)(tá:)te:s*. GLA learners with FTNONFINAL produced forms like /L (H2) (H1) H/ as for *vo(lùp)(tá:)te:s*, /L (H2) (L1 L) L/ as for *a(mì:)(kí.ti)a*, and /(H2) (H2) (H2) (H1) H/ as for *(dè:)(fì:)(nì:)(tì:)vus*.

6.2. Simulations 3 to 5

Another possibility for secondary stress is to build a weight-insensitive disyllabic foot at the left edge of the word, given the overt forms in (18) in which the 20 forms without secondary stress are suppressed. All learners fail, simply because there *is* no ranking that can describe the data.

(18) Left-aligned binary weight-insensitive secondary stress

Quadrisyllables			
[L2 L H1 L]	[L2 H H1 L]	[H2 L H1 L]	[H2 H H1 L]
[L2 L H1 H]	[L2 H H1 H]	[H2 L H1 H]	[H2 H H1 H]

The same happens with a training set that has left-aligned, binary weight-sensitive secondary stress, as in (19): there is no OT analysis with our constraint sets that could describe this pattern, so again, all learners fail (data without secondary stress, like [L H1 L H], are again suppressed, although they would make the weight-sensitivity more explicit).

(19) Simulation 4: left-aligned binary weight-sensitive secondary stress

Trisyllables	Quadrisyllables			
[H2 H1 L]	[L2 L H1 L]	[L2 H H1 H]	[H2 L H1 L]	[H2 H1 L H]
[H2 H1 H]	[L2 L H1 H]	[H2 L1 L L]	[H2 L H1 H]	[H2 H H1 L]
	[L2 H H1 L]	[H2 L1 L L]	[H2 H1 L L]	[H2 H H1 H]

The same happens with data that contain left-aligned weight-insensitive secondary stress that is not binary, as in (20): all learners fail.

(20) Simulation 5: left-aligned weight-insensitive secondary stress

Trisyllables	Quadrisyllables			
[L2 H1 L]	[L2 L1 L L]	[L2 H1 L L]	[H2 L1 L L]	[H2 H1 L L]
[L2 H1 H]	[L2 L1 L H]	[L2 H1 L H]	[H2 L1 L L]	[H2 H1 L H]
[H2 H1 L]	[L2 L H1 L]	[L2 H H1 L]	[H2 L H1 L]	[H2 H H1 L]
[H2 H1 H]	[L2 L H1 H]	[L2 H H1 H]	[H2 L H1 H]	[H2 H H1 H]

6.3. Simulation 6: freely assignable secondary stress

A further possibility to assign secondary stress is to let the learners invent it. This is done in the sixth simulation. The idea is that even if there appears only one audible stress in a word, the surface structure could be made up with several feet that are simply not articulated (see Halle & Vergnaud's 1987 *conflation*, and Hayes' 1995 reformulation of it). The consequence is that although children hear only main stress, they could construct more than one foot in a word. Given this, GEN would then generate an additional candidate for a form like [H1 L]: /(H1) (L2)/. Alternatively, secondary stress in the input could be ignored by a learner so that [H1 L2] could be perceived as /(H1) L/. Both strategies constitute a violation of OF-SF faithfulness for secondary stress. As usual, the choice between the candidates is determined by the ranking.

The input to simulation 6 was therefore the same as in simulation 1: main-stress only overt forms. But this time the learners were allowed to invent secondary stress. The resulting constraint rankings are listed in (21):

(21) An EDCD learner

Constraints	Ranking values
FootBin	100.000
NonFinal	99.000
AFR	98.000
Main-R	
AFL	97.000
Main-L	
WFL	
Parse	96.000
WFR	
WSP	
Iambic	-104.000
Trochaic	-105.000

A GLA learner

Constraints	Ranking values
NonFinal	116.962
Main-R	110.198
WSP	106.139
Parse	105.612
AFL	104.276
Main-L	100.623
Trochaic	99.743
WFL	99.185
Iambic	97.045
FootBin	87.208
WFR	83.038
AFR	80.461

The EDCD learners failed again: they were not able to produce main stress correctly. The GLA learners assigned main stress correctly, and furthermore created secondary stress in some forms: e.g. /(L2) (L1 L) X/ as for *(fà)(kí.li)ter*, and /(L2) (H1 L) X/ as for *(sù)(pér.bi)ter*, as well as /(H2) (H1) X/, /(L2) (L H1) X/, /(L H2) (H1) X/, /(H2) (L1 L) X/, /(H2) (H1 L) X/, /(H2) (H2) (H1) X/, and /(H2) (L H1) X/.

Their generalizations to longer forms were correct but weird, though: /(L2 L) (L H1) X/ as for *(rà.pi)(di.tá:)tem*, and /(L2) (L2 L) (H1 L) X/ as for *(rà)(pì.di)(tá:.ti)bus*. They are weird because the secondary stress assigned to the left of the main stress is influenced by what happens to the right of the main stressed syllable: if it is heavy and penultimate, an iambic foot is built as in /(L2 L) (L H1) X/; if it is heavy and antepenultimate, a trochaic foot is built, as in /(L2) (L2 L) (H1 L) X/. Has something like this ever been observed in the languages of the world?

7. Summary

7.1. Summary of successes

The successful simulations included three very different stress patterns, which could be learned by GLA learners only. A word like *voluptates* was analysed in simulation 1 as *vo.lup(tá:)te:s*, in simulation 2 as *vo(lùp)(tá:)te:s* (with FTNONFINAL), and as *(vo.lùp)(tá:)te:s* (with TROCHAIC; the same in simulation 6). EDCD learners always failed to converge upon a correct grammar. The immediate cause for this lies in the behaviour of the most uncertain constraints (those for trochaicity and iambicity), since the EDCD strategy moves them to the bottom of the hierarchy early, while the GLA keeps them ranked in the middle.

7.2. Summary of failures

What is missing in our simulations are analyses with strictly bimoraic feet such as /(H1) L <L>/, as proposed by Mester (1994) on the basis of segmental changes such as iambic and cretic shortening. Our constraint sets are not capable of producing this pattern, regardless of the input. Furthermore, although forms like [L2 H H1 H] were given in simulations 3 to 5 and allowed in simulation 6, no learner came up with the analysis of secondary stress actually proposed in the literature (Allen 1973a, b), which contains wretched trochees such as /(L2 H) (H1) H/ as for *(vò.lup)(tá:)te:s*.

8. Therefore, future work

Looking at the results of our simulations, we feel that something has to be done about FOOTBIN. As explained in Apoussidou & Boersma (2003) its makeshift formulation (referring to syllables and morae) is the cause of the failure of our learners to come up with strictly bimoraic analyses like (3a). Rather than filling this gap with ad hoc constraints like RHYTHMIC HARMONY (Prince & Smolensky 1993) we should find a way to have this constraint refer to morae in weight-sensitive languages, to syllables in weight-insensitive languages.

Also, if perception precedes lexical access, as in (12), foot structure has to be assigned before word boundaries are. This order is problematic because some of the constraints we used imply a dependence of foot assignment on word boundaries. Consider the overt form [á:.bra.ka.dá:.bra], to which the learner has to assign two feet and a word boundary. Under an analysis with uneven trochees the following problem emerges: if the word boundary is as in *a:.bra#ka.da:.bra*, the footing would have to be *(á:)bra#ka(dá:)bra*. If it is *a:.bra.ka#da:.bra*, footing would have to be *(á:.bra)ka#(dá:)bra*. This makes the strictly bimoraic analysis more likely, since this bimoraic analysis would predict identical footing in *(á:)bra#ka(dá:)bra* and *(á:)bra.ka#(dá:)bra*, so that feet can be assigned independently from (e.g. before) word boundaries.

A last point is that we gave the learners too much information about syllable weight. Real children have to learn the heaviness of CVC syllables by themselves. In some languages, CVC is light (e.g. final, monomoraic CVC-feet in Chuukese, as described in Kennedy 2003), while in others it is heavy (e.g. in Latin).

In sum, it all smells like we need a more emergentist modelling of representations and constraints, meaning that much less is given to the learner than is assumed in Tesar & Smolensky's (2000) and our simulations.

References

Allen, W. Sidney. 1973a. *Accent and rhythm*. Cambridge University Press.

- 1973b. *Vox Latina. The pronunciation of Classical Latin*. Second edition. Cambridge University Press.

Apoussidou, Diana and Paul Boersma. 2003. The learnability of Latin stress. In *Proceedings of the Institute of Phonetic Sciences* 25: 101-148.

Boersma, Paul. 1997. How we learn variation, optionality, and probability. In *Proc. of the Institute of Phonetic Sciences* 21: 43-58. University of Amsterdam.

- 1998. *Functional phonology: formalizing the interactions between articulatory and perceptual drives*. Doctoral thesis, University of Amsterdam.

Halle, Morris and Jean-Roger Vergnaud. 1987. *An essay on stress*. Cambridge: MIT Press.

Hayes, Bruce. 1981. *A metrical theory of stress rules*. Doctoral thesis (1980), Cambridge: MIT Press.

- 1995. *Metrical stress theory*. University of Chicago Press.

Jacobs, Haike. 2000. The revenge of the uneven trochee: Latin main stress, metrical constituency, stress-related phenomena and OT. In A. Lahiri (ed.), *Analogy, levelling, markedness*, 333-352. Berlin: Mouton de Gruyter.

Jakobson, Roman. 1937/1962. On ancient Greek prosody. In R. Jakobson 1962, *Selected writings I. Phonological studies*, 262-271. The Hague: Mouton & Co (from *Z zagadnień poetyki: prace ofiarowane K. Wóycickiemu*. Wilno).

Kennedy, Robert. 2003. *Confluence in phonology: evidence from Micronesian reduplication*. Doctoral thesis, University of Arizona.

Mester, R. Armin. 1994. The quantitative trochee in Latin. In *Natural Language and Linguistic Theory* 12: 1-61.

Prince, Alan and Paul Smolensky. 1993. *Optimality Theory: constraint interaction in generative grammar*. Ms., Rutgers University, New Jersey.

Tesar, Bruce. 1995. *Computational Optimality Theory*. Doctoral thesis, University of Colorado.

Tesar, Bruce and Paul Smolensky. 1998. Learnability in Optimality Theory. In *Linguistic Inquiry* 29: 229-268.

- 2000. *Learnability in Optimality Theory*. Cambridge: MIT Press.