The Learnability of Metrical Phonology

Published by
LOT
Janskerkhof 13
3512 BL Utrecht
The Netherlands

phone: +31 30 253 6006
fax: +31 30 253 6406
e-mail: lot@let.uu.nl
http://wwwlot.let.uu.nl/

Cover illustration by Diana Apoussidou

ISBN 978-90-78328-18-6
NUR 632

THE LEARNABILITY OF METRICAL PHONOLOGY

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof.mr. P.F. van der Heijden
ten overstaan van een door het college voor promoties ingestelde
commissie, in het openbaar te verdedigen in de Aula der Universiteit
op dinsdag 9 januari 2007, te 14.00 uur
door Diana Apoussidou
geboren te Mönchengladbach, Duitsland

Promotiecommissie

Promotor: prof. dr. P.P.G. Boersma

Overige commissieleden: prof. dr. H.M.G.M. Jacobs
prof. dr. R.W.J. Kager
dr. W. Kehrein
dr. N.S.H. Smith
prof. dr. P. Smolensky

Faculteit der Geesteswetenschappen

Acknowledgements

I consider myself very lucky to have had the support when writing this thesis, whether directly or indirectly, from the following people:

Essential to starting and finishing this work is Paul Boersma. I thank him for his support as much as for challenging me, and for enabling me to push the limits.

I thank my reading committee Haike Jacobs, René Kager, Wolfgang Kehrein, Norval Smith, and Paul Smolensky for their valuable comments; especially Wolfgang and Norval for enduring tedious questions about phonology.

My paranymphs Maren Pannemann and Petra Jongmans never failed in supporting me; they prevented me from going nuts in quite some moments of panic. I wouldn't have known what to do without them, and my Dutch would be much worse if it weren't for them.

Furthermore, I was very happy to be part of the phonological reading groups in Utrecht and Leiden. It was always fun and helpful to discuss anything phono with Dafna Graf, Paula Fikkert, Nancy Kula, Claartje Levelt, Jeroen van der Weijer, Marc van Oostendorp, and all the other people who attended. Being part of the Sound Circle organizing committee with Ivana Brasileiro, Ania Kijak, and Shakuntala Mahanta was, and still is, great.

The core of the IFA as I got to know it made me always feel welcome. Florien van Beinum, Jan van Dijk, Cecilia Odé, Louis Pols, Jeannette van der Stelt, Rob van Zon, David Weenink, Ton Wempe en Dirk Jaasma took more than good care of me, not to forget my fellow gratiën Irene Jacobi and Wieneke Wesseling. Being a phonologist among phoneticians is not so bad at all. In the bigger picture, I thank everyone at the ACLC for contributing to a stimulating work environment, and all AiOs for being an enjoyable group open for discussion.

Being responsible for my beginning interest in science, I like to thank Janet Grijzenhout and Martin Krämer, who have always been there with good advise for me beyond Düsseldorf.

There are many non-phonologist linguists I need to thank for a lot inside and outside linguistics: Martin Salzmann for having the first fun-nights out in

Amsterdam; Berit Gehrke for many more fun-nights out and in; Mika Poss for being there in Boston; Silke Hamann for being a role model, and Robert Cloutier for his patience and help with my English. My fellow mappreviators need to be mentioned, too: Roland Pfau for being my other role model; Rachel Selbach and Karina Hof for brightening my day with their eloquence and quick cheer-ups; and Rafael Fischer for being a fun office mate and utterly cheek.

Completely outside of linguistics, there are many more people who contributed in mastering my time as a PhD student. I thank my homies from Germany Yvonne Salcewics, Anja Ammenhäuser, Marie Reuter, and Wolfram Bernhardt for keeping the faith. Last but definitely not least, I thank Georgios, Annelie and Aliki for being there.

The list is not exhaustive; many others have contributed in one way or the other to my work. However, I would like to leave it at that and let the interested reader go on to the main part of this book.

2 Grammatical vs. lexical stress

3 The learnability of hidden structure and the grammar

4 The learnability of grammatical stress in Latin

5 The learnability of grammatical stress and weight in Pintupi

6 The learnability of grammatical and lexical stress in Modern Greek

7 Conclusions

1 Introduction

Linguistic theory needs to consider the fact that a language must be learnable in order to stay alive. *Learnability theory* is the part in linguistics that searches for formalized explanations for language acquisition. The study of learnability can provide insights into the mechanisms that underlie language acquisition and how languages can change. These insights have repercussions on the shaping of linguistic theory.

The general goal of this dissertation is to provide a comprehensive proposal for the learning of phonology. This is exemplified by the case of metrical phonology. I argue for a bidirectional[1] approach to phonology and its acquisition (Boersma 1997, 1998, 1999, 2000, 2001, 2005, 2006ab). The adopted grammar model uses the framework of Optimality Theory (Prince & Smolensky 1993) and formalizes two interfaces: the interface between phonology and phonetics, and the interface between phonology and semantics. For each of the three language modules there are probably multiple levels of representation; an all-embracing learnability approach would have to account for this. The present dissertation, however, restricts itself to the four representations shown in figure (1): one semantic representation (*meaning*, represented as morphemes in the lexicon), two phonological representations (the lexical *underlying form* and the abstract phonological *surface form*), and one phonetic representation (the concrete auditory *overt form*).

Throughout the book, I use the following symbols to mark the various representations, as in figure (1): meaning is represented between single quotes (‘ ’), underlying forms between pipes (| |), surface forms between slashes (/ /), and overt forms between square brackets ([ ]). For instance, meaning can be a single morpheme such as ‘day’ or a morphemic structure such as ‘day-Nom.Pl’.[2] In German, the underlying form that corresponds to this morphemic structure is |taːg+ə|, where |taːg| corresponds to the morpheme ‘day’, and |ə| corresponds to the morpheme ‘Nom.Pl’.

---

[1] The term ‘bidirectional’ stems from OT semanticist Blutner (2000).

[2] ‘Nom.’ stands for nominative case, ‘Pl’ for plural. This means that the term *meaning* as used here can refer to *syntactic* functions as well.

(1) The grammar model

semantic representation	\`Meaning'
phonological representations	|Underlying Form|
	/Surface Form/
phonetic representation	[Overt Form]

The corresponding surface form is /(tá$_{\mu\mu}$.gə$_{\mu}$)/, which contains hidden structures such as feet (indicated by parentheses), moras ($_{\mu}$), stress (), and syllable boundaries (.). The corresponding *overt form* could be the detailed phonetic transcription [ˈtʰaːgə], but in this book I abstract away from segmental detail and limit myself to representing stress as in [táː.gə] (I do, however, include syllable boundaries in the overt form and therefore assume that they are known to learners from the start). Feet and moras are not contained in the overt form.

The grammar model in figure (1) is bidirectional in the sense that it can handle the two main linguistic processing tasks of *production* and *comprehension*. In production, the figure in (1) can be read top-down: a speaker is given meaning, from which she computes form: the underlying form, the corresponding surface form, and the corresponding overt form. In comprehension, figure (1) can be read bottom-up: a listener is given the overt form, from which she computes surface form, underlying form, and meaning. Within OT, this computation of representations is achieved through the constraint ranking of a language.

The child's learning task is to become a good speaker and a good listener. Her task is therefore *learning to compute* surface form, underlying form and meaning in comprehension, and underlying form, surface form and overt form in production. Within OT, the learning task is to learn an adequate constraint ranking. **The specific goal of this dissertation is to show that the child can create the whole phonology of the language, if given a sufficient number of informative pairs of overt forms and meaning in the course of acquisition. This means, in fact, that the child bootstraps into phonology by using phonetic and semantic information.** The dissertation will pass through several steps before arriving at the ultimate goal.

The first step is to model the acquisition of the mapping between surface form and underlying form, the two middle forms in figure (1), when both underlying forms and surface forms are given. This has already been done in former OT approaches on learnability: if the learner is given a sufficient number of informative pairs of underlying and surface forms, the learner can learn how to compute a surface form for any given underlying form (Tesar & Smolensky 1993, 1996, 1998, 2000; Tesar 1995; Boersma & Hayes 2001; Apoussidou and Boersma 2003[3]).[4] Likewise, it has been shown that if the learner is given a sufficient number of informative pairs of overt and surface forms (the two bottom representations in figure (1)), the learner can learn to compute a surface form for any given overt form (Boersma 1997, Escudero & Boersma 2001, 2004). There are various proposals of how this learning can proceed: Tesar & Smolensky (1993) and Tesar (1995) formalized this learning as *Constraint Demotion* (henceforth CD), while Boersma (1997) and Boersma & Hayes (2001) formalized this learning as the *Gradual Learning Algorithm* (henceforth GLA). I explain the mapping between surface form and underlying form and the formalized learning approaches in chapter 3, and apply them in chapter 4 on Latin stress.

The second step is to model the acquisition of the mapping between overt form, surface form and underlying form, the three bottom representations in figure (1). It has already been shown that if a learner is given a sufficient number of informative pairs of overt and underlying forms, the learner can learn how to compute surface and overt forms for any given underlying form (in a CD manner shown by Tesar 1997, 1998ab, 1999; and Tesar & Smolensky 1996, 1998, 2000; in a GLA manner shown by Boersma 1998, 2003; Apoussidou & Boersma 2003, 2004ab[5]; and Apoussidou 2006a[6]). In this learning situation, surface forms are hidden; they have to be created by the learner. This requires a translation mechanism for the comprehension process. In the OT formalizations the child is enabled

---

[3] This paper has been incorporated into this dissertation in chapter 4.

[4] The reader familiar with the literature might object that in the listed learnability approaches, the learners were not fed underlying forms; however, the underlying forms were known to the learner. The learner had no other option than mapping a surface form like /(σ σ́) σ/ onto an underlying form |σ σ σ|, or a surface form like /σ (σ́ σ) σ/ onto |σ σ σ σ|. This boils down to saying that the learner was provided with both surface form and underlying form in the learning process.

[5] These papers have been incorporated into this dissertation in chapter 4.

[6] This paper has been incorporated into this dissertation in chapter 5.

to interpret the speech signal by using her current grammar to assign hidden structure, introduced as *Robust Interpretive Parsing* by Tesar & Smolensky (henceforth RIP; 1996, 1998, 2000). While the interpretation is only a first guess and might be wrong, the mechanism itself is robust, in that it never fails to give an interpretation. This mechanism enables the child to detect what might be wrong with her grammar and needs to be changed by comparing her interpretation of a heard form to what she would produce herself. In OT, the grammar is the language-specific ranking of constraints. Children's grammars differ from adult ones, and although children have no idea about what the adult constraint ranking looks like, they have access to the constraint violations of the forms they process, and thereby the means to change their grammar in response to the language data they encounter. I pick up on the ideas of RIP, CD and the GLA, and explain them in chapter 3. I apply them in chapter 4 on Latin stress and in chapter 5 on Pintupi stress.

The third step is to model the acquisition of the mapping between surface form, underlying form and meaning, the three top representations in figure (1). I propose an *on-line* learning approach, where the learner does not have to wait and gather data before she can make a learning move, but where lexicon and grammar are learned concurrently. In fact, underlying forms are learned through the grammar. I show how a learner can learn to compute underlying forms and surface forms if provided with a sufficient and informative number of pairs of surface forms and meaning by applying the mechanism of RIP (see also Apoussidou 2006b[7]). The situation of learning underlying form and grammar, given surface form, has indeed been tackled by e.g. Tesar et al. (2003) with a rather complicated learning algorithm that switches back and forth between the modification of the underlying forms in the lexicon and modification of the grammar, and by Tesar (2004, 2006). The relation between meaning and underlying form was not handled by the grammar. Crucial for these approaches is the ability of a learner to compare paradigms, implying *off-line* learning: the learner gathers data and stores them for later processing. I argue that this is not a realistic learning situation and therefore propose the on-line alternative, outlined in chapter 3 and applied in combination with the GLA in chapter 6.

As a fourth and final step I model the acquisition of the mapping between meaning, underlying form, surface form and overt form that are shown in figure (1). I demonstrate how a learner, if given a sufficient

---

[7] This paper has been incorporated into this dissertation in chapter 6.

number of informative pairs of overt forms and meaning, can learn to compute underlying forms, surface forms, and overt forms for any given meaning. This approach to the learning of underlying forms relaxes the demarcation between grammar and lexicon: the connection between underlying forms and surface forms, and between underlying forms and meaning, is handled by the grammar. Again I make use of RIP; this time to model the two levels of hidden structure, surface and underlying form. Step 4, the goal of this dissertation, is achieved in chapter 6 on Modern Greek.

With the OT learnability approaches, and especially when multiple representations are considered, many of the learning problems stated in generative linguistics can be solved. One of the main learning problems is the 'poverty of the stimulus' (Chomsky 1986:7): the fact that during the learning period a child never gets to hear every possible sentence that the adult language can construct (the child is exposed to an 'impoverished input'), yet is able to learn the language. It has therefore been argued that a child does not learn language by heart in memorizing every sentence she ever gets to hear, but that she can learn the language by inferring hidden structure from the speech samples that she encounters. By abstraction, the learner can adapt her grammar and create new, meaningful utterances. This abstraction can be handled in phonology by the mentioned mechanism Robust Interpretive Parsing. Another problem in the literature on learnability is the problem that children can learn only from positive evidence in the data. A child cannot deduce from the fact that she does *not* hear a certain structure that this structure is *not* permitted. She can only deduce from a certain structure she *does* hear that it *is* permitted. OT deals with this problem by providing the learner with implicit negative evidence: any form that is regarded as optimal by the child's grammar constitutes the positive evidence, and all other candidates that are discarded as not optimal constitute the negative evidence (Tesar & Smolensky 2000:33). The mechanisms of interpretation and constraint reranking are outlined in chapter 3, and applied in chapters 4, 5 and 6.

I exemplify the proposed approach to learning with word stress of different languages. The dissertation at hand models three different kinds of stress systems, outlined in chapter 2. I distinguish between *grammatically assigned stress* (modelled in chapters 4 and 5) and *lexically assigned stress* (modelled in chapter 6). Within grammatical stress systems I distinguish between *weight-sensitive* stress and *weight-insensitive* stress. The first language is Classical Latin in chapter 4, a language with weight-sensitive stress. Although a dead language, the prosodic system of Latin is well-studied in phonology. The study of Latin provided insights into cross-linguistic principles of phonology, and the study of the learnability of Latin stress can provide insights into cross-linguistic principles of learnability. Stress in Latin is largely determined by *heavy* syllables, and therefore provides an example for a weight-sensitive language. With the modelling of Latin stress it will be shown that under the learning model argued for, learners are not only capable of inferring hidden structures from what they hear, but can create structure on their own, if permitted (e.g. create secondary stress from data with only primary stress). They come up with slight variations in their grammar. This chapter furthermore provides a comparison between different analyses that have been proposed for Latin stress, and their learnability. It turns out that some of the analyses are better learnable than others, and that some are not learnable at all. The chapter also provides a comparison between the two reranking strategies CD and GLA, and a comparison of different constraint sets that the learners are equipped with. Three representational levels are involved: one phonetic level and the two phonological levels of surface form and underlying form (the three bottom representations of figure (1)).

The second language is Pintupi, a Pama-Nyungan language spoken in Western Australia, modelled in chapter 5. Pintupi has weight-insensitive stress. The modelling of Pintupi stress makes clear that the overt forms of the learners are uniform, yet the grammars of the different learners can vary. Communication among each other is still guaranteed. Again, CD and the GLA are tested with respect to their performance in learning, as well as different constraint sets. As in the Latin simulations, three levels of representation are modelled: phonetic form, surface form, and underlying form.

The third language that is modelled is Modern Greek in chapter 6, where stress is largely determined by the lexicon. Lexically assigned stress is interesting to model, because it is not predictable form the grammar and makes it necessary to account for a learning of underlying forms. In this chapter, all four levels of representation that are shown in figure (1) are modelled.

I conclude in chapter 7 that within the proposed learnability approach:

1. learners are able to create structure (shown in chapter 4),
2. the grammars of speakers can differ while they are still speaking the same language: they can have different surface and underlying structures, but as long as their overt forms are the same, communication is guaranteed (shown in chapters 5 and 6), and
3. parts of the lexicon can be learned by acquiring the phonological grammar (shown in chapter 6).

2 Grammatical vs. lexical stress

2.1 Introduction

In the pre-linguistic stage (if there is any) infants probably experience the language they are exposed to as an impenetrable speech stream. Stress functions as a help to break up the speech stream into smaller, processable units. But different languages make different use of stress. In languages where stress serves a word boundary marker, it would come in handy for the child to know that a stressed syllable signals the beginning or end of a word. However, the child does not know that stress is a boundary marker unless she knows the word boundaries. In languages where stress carries information about the morphological components of the word, the child does not know this until she knows these components and can link them to stress. In short, the language-acquiring child no more knows the function of stress in her mother tongue in advance than she knows how to assign the correct stress pattern. Learning the stress pattern of a language is complicated by the fact that stress is only the overt manifestation of covert, not directly observable structure of a word, like metrical *feet*. Stress is the clue for the child to construct feet, but languages make use of different patterns of feet, and stress is sometimes ambiguous with respect to foot structure: whereas one language might interpret a trisyllabic form with medial stress, [σ σ́ σ], as having a trochaic rhythm /σ (σ́ σ)/, other languages might interpret it as having an iambic rhythm /(σ σ́) σ/. This complication is tackled in the present approach on the learning of stress patterns and hidden structures. The three languages modelled in this book stand for three different kinds of stress systems: Latin as a weight-sensitive language in chapter 4, Pintupi as a weight-insensitive language with rhythmical stress in chapter 5, and Modern Greek as a weight-insensitive language with lexical stress in chapter 6.

This chapter provides the background for the metrical analyses of the different languages that will be modelled in chapters 4 to 6. Throughout the dissertation, the expression "stress" refers to word stress.[8] I give a short

---

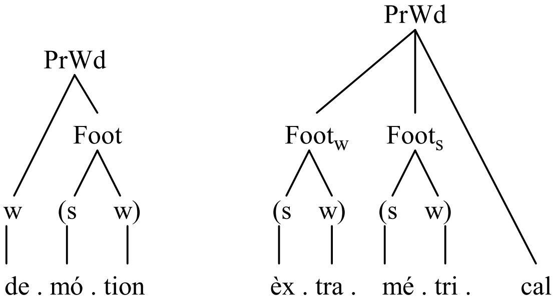
[8] Stress can be classified in different ways. The concrete phonetic manifestation of stress and its learnability is not a subject in this dissertation; I assume that the languages modelled in this book realize stress in at least one of the correlates

classification of word stress in sections 2.2-2.4 and a basic approach in section 2.5 to how stress is handled in Optimality Theory (Prince & Smolensky 1993). Section 2.6 discusses the constraint families used, while section 2.7 briefly discusses the shape that overt forms can take in this book. Section 2.8 summarizes.

2.2 Typological grass roots

I basically assume that stress patterns are rhythmically organized (Liberman & Prince 1977) by feet (e.g. Prince 1976a; Halle & Vergnaud 1978; McCarthy 1979ab, Selkirk 1980b). A foot is the metrical constituent that groups smaller units within a word, such as syllables or moras, into bigger units. Each foot has exactly one head syllable (marked with 's' for 'strong'; 'w' stands for 'weak'), and each prosodic (i.e. content) word has exactly one head foot, no matter how many feet it contains, as illustrated in figure (2).[9]

(2) Metrical constituents



The foot inventory adopted here consists of maximally binary feet (as in Hayes 1991, 1995).[10] Feet with a strong-weak pattern are called *trochees*, and feet with a weak-strong pattern are called *iambs*:

---

duration, intensity, loudness, pitch or clarity. I therefore regard a stressed syllable as a syllable that jumps out in one way or the other compared to unstressed syllables in a word.

[9] I assume that the syllable is the intermediate level between moras and feet; otherwise, it would be predicted that foot boundaries can fall within a syllable.

[10] Although intriguing, ternary stress systems and systems with level stress are not discussed here.

(3) Foot inventory

Trochee: (σ́ σ), (σ́) Iamb: (σ σ́), (σ́)

The following two subsections outline the distinction between *grammatically* assigned stress and *lexically* assigned stress, which I use throughout the book. I refer to grammatically assigned stress when talking about stress that is predictable from the grammar (i.e. when there is a “rule” or a ranking of structural constraints that determines the stress of a word). I talk about lexically assigned stress when stress is not (fully) predictable by the grammar, but when the grammar interacts with the lexicon in order to determine stress.

2.3 Grammatical stress systems

Grammatically assigned stress can serve as a word boundary marker by stressing the first or last syllable of a word. Grammatically assigned stress can also serve as a marker of phonetically salient syllables, as in weight- or quantity-sensitive languages.

A language is called quantity-sensitive if stress is assigned depending on the structure of a syllable, i.e. depending on the *weight* of a syllable. The weight of a syllable is determined by the number of *moras* it contains.[11] A syllable with a short vowel has one mora and is light, and a syllable with a long vowel has two moras and is heavy. Coda consonants can count a mora, but this is not universal. In some languages coda consonant are moraic, making the syllable heavy (e.g. Yana; Sapir & Swadesh 1960), but in other cases the coda does not contribute to the weight of a syllable (e.g. Khalkha Mongolian; Walker 1997; or Ancient Greek). In quantity-sensitive languages with trochaic rhythm, feet are ideally bimoraic: they should contain exactly two moras. Feet with a strong-weak pattern are then called *moraic trochees* (Hayes 1991). Moraic trochees can either consist of two light syllables (‘σ’) containing one mora (‘μ’) each, or of one heavy syllable containing two moras, as illustrated in (4).

---

[11] Usually it is assumed that only the rime of a syllable can contain moraic elements. A famous exception is Pirahã (Everett & Everett 1984, Everett 1988), where the onset contributes to the weight of a syllable. Gordon (2005; see references therein) discusses twelve other languages.

(4) Moraic trochees

$$\begin{array}{ccccc} (* & .) & & & (*) \\ | & | & \text{or} & & | \\ \sigma_{\mu} & \sigma_{\mu} & & & \sigma_{\mu\mu} \end{array}$$

An example of this kind is Latin, modelled in chapter 4. Latin has phonemic vowel length and stresses the pre-final (penultimate) syllable if heavy, otherwise the antepenultimate syllable.

Iambs are weight-sensitive if they consist either of a light syllable followed by a stressed light or heavy syllable, or of one heavy syllable (Hayes 1995):

(5) Iambs

$$\begin{array}{cccccccc} (. & *) & & (. & *) & & (*) \\ | & | & \text{or} & | & | & \text{or} & | \\ \sigma_{\mu} & \sigma_{\mu} & & \sigma_{\mu} & \sigma_{\mu\mu} & & \sigma_{\mu\mu} \end{array}$$

If a language does not employ any weight distinctions (e.g. it does not have distinctive vowel length or moraic codas) stress cannot be sensitive to weight. The language is *trivially quantity-insensitive* (as Kager 1992 calls it), and stress is assigned by e.g. the alignment to word edges. Trochees are in this case syllabic trochees (Hayes 1991) and contain two syllables, regardless of the syllable structure. This is illustrated in (6):

(6) Syllabic trochees

$$\begin{array}{cc} (* & .) \\ | & | \\ \sigma & \sigma \end{array}$$

Next to this kind of stress system are languages that do have weight distinctions, but assign stress independently of the weight of a syllable. Kager (1992) classifies these languages as *truly quantity-insensitive,* stating at the same time that these languages assign stress not completely independently from weight. If these languages have trochaic feet, the feet take the form of *generalized trochees* (Hayes 1991, 1995): feet are preferably disyllabic, else bimoraic, as in (7).

(7) Generalized trochees

$$\begin{array}{ccc} (\,*\;\;.\,) & & (\,*\,) \\ |\;\;| & \text{else} & | \\ \sigma\;\;\sigma & & \sigma_{\mu\mu} \end{array}$$

These truly quantity-insensitive languages reveal some sort of weight-sensitivity in that e.g. the assignment of secondary stress is weight-sensitive, or in that the language has a bimoraic word minimum. For instance in Estonian, main stress as well as secondary stress is assigned by linking together syllabic trochees (Prince 1980, Kager 1992). However, in words with an odd number of syllables the last syllable is footed (and thereby stressed) only if heavy. Chapter 5 will deal with another quantity-insensitive language, Pintupi (Western Australia; Hansen & Hansen 1969). It is traditionally analyzed as having syllabic trochees assigned from left to right, leaving final syllables unfooted in words with an odd number of syllables.

2.4 Lexical stress systems

Lexical stress systems assign stress on the basis of marks in the underlying form of a word. While there are some languages which seem to have straightforward marks for stress (in the sense "stress the syllable in the output that is stressed in the underlying form") there are others that involve more than that, supposedly foot structure (e.g. Inkelas 1998 for exceptional stress in Turkish). Revithiadou (1999) proposed for Modern Greek (the language that will be modelled in chapter 6) that morphemes can be underlyingly marked for "strong" or "weak" accents. Stress in Modern Greek is limited to the last three syllables of a word, but within this range it can occur on any syllable. Morphemes can be stressed, unstressed, pre- or post-stressing. Stress is largely determined by the lexical specifications of the morphemes of a word, and their interaction with each other. The challenge that lexical stress systems pose to a theory of learnability is that it involves the modification of the lexicon, a far more complex task for the learner than solely adjusting the grammar. This is shown in chapter 6.

The next section gives a brief introduction of how stress is handled in OT, and discusses the constraint families that are used in the computer simulations.

2.5 Stress in Optimality Theory

According to OT (Prince & Smolensky 1993), all languages of the world share the same set of violable constraints, and the languages differ only in the *ranking* of these constraints, i.e. their relative degree of importance. The hierarchical ranking of all the constraints constitutes the *grammar* of a language. In this section, I explain how a ranking of constraints determines the stress pattern in a language.

Standard OT is production-directed. A speaker of a given language chooses an underlying form (the *input* to the grammar) from a set of lexical items. On the basis of this underlying form she then chooses a surface form (the *output* of the grammar) from among a set of possible *output candidate*s. These candidates are provided by the function GEN, which can generate by *Freedom of Analysis* (Prince & Smolensky 1993:6, McCarthy & Prince 1993b:21) any possible linguistic structure. The candidate chosen as the *optimal* one from all these candidates is the one that satisfies the highest-ranked constraints best. At the same time it might abundantly violate lower-ranked constraints. This process of *evaluation* is portrayed in a *tableau*, where the candidates are compared with respect to their fulfilment or violation of the constraints.

This can be applied to stress assignment. Phonologists generally agree that while in some languages stress can be assigned by referring to word edges (e.g. “always stress the first syllable in a word”), the analysis of other languages requires one to assume that syllables are grouped into hidden structures called feet. Every foot has one stressed syllable. For the purposes of this section and the next, I only consider disyllabic feet. Imagine a simplified foot inventory, where feet are disyllabic and are either trochaic (σ́ σ) or iambic (σ σ́). Consider now the small universal constraint set in (8), where two constraints (IAMBIC and TROCHAIC) are responsible for the placement of the stressed syllable within the foot, and two constraints ALIGNFT-R and ALIGNFT-L are responsible for the placement of the foot within the word. These four constraints are among the many that have been proposed in the literature to account for generalizations on the phenomena of metrical phonology.

(8) Constraints on metrical constituents

IAMBIC: The rightmost syllable in a foot is the head syllable.
TROCHAIC: The leftmost syllable in a foot is the head syllable.
ALIGNFT-R (AFR): Align the right edge of the foot with the right edge of the word.
ALIGNFT-L (AFL): Align the left edge of the foot with the left edge of the word.

The constraints IAMBIC and TROCHAIC (short for Prince & Smolensky's 1993 RHYTHMTYPE=IAMBIC/TROCHAIC) stem from the observation that languages tend to have either iambic or trochaic feet, rather than a mix of them (McCarthy & Prince 1986, Kager 1996, Van de Vijver 1998). AFR and AFL stem from the observation that languages tend to have feet that are either close to the beginning or to the end of a word, or tend to assign feet iteratively starting either near the beginning or near the end of the word.

Consider now an underlying form with three syllables, represented in pipes |σ σ σ|. If we assume that stress is assigned purely by the grammar (i.e. the language at hand does not have lexical stress), then we have at least the four different candidates shown in tableau (9). The four candidates have different main stresses, denoted as "σ́", and different foot structures, denoted by parentheses. Suppose now that in a specific language the highest ranked (i.e. most important) constraint is IAMBIC and the lowest ranked constraint is AFR. This ranking is denoted in the tableau by sorting the constraints from left to right. The asterisks *(violation mark*s) in the tableau depict which candidates violate which constraints. The candidates /(σ́ σ) σ/ and /σ (σ́ σ)/ both violate the highest ranked constraint IAMBIC, since they contain a trochaic foot. These violations are marked with a "!" because they are the crucial violations that rule out these two candidates from further consideration. The choice between the remaining two candidates /(σ σ́) σ/ and /σ (σ σ́)/ cannot be made by the two highest ranked constraints IAMBIC and TROCHAIC, since these two constraints have an equal number of violations for these two candidates. The matter is decided by AFL, which prefers the candidate /(σ σ́) σ/, since this form has a left-aligned foot, unlike /σ (σ́ σ)/. The grey cells in the tableau are those that do not contribute to the determination of the winning form. The winning candidate itself is denoted by '☞'.

(9) An iambic left-aligning language

Underlying: |σ σ σ|	IAMBIC	TROCHAIC	AFL	AFR
/(σ́ σ) σ/	∗!			∗
☞ /(σ σ́) σ/		∗		∗
/σ (σ́ σ)/	∗!		∗	
/σ (σ σ́)/		∗	∗!	

The language in tableau (9) is an iambic left-aligning language: the foot in the winning candidate is iambic and a result of the ranking IAMBIC >> TROCHAIC. The left-alignment within the word is the result of the ranking AFL >> AFR.

An early assumption in OT is that any ranking of the constraints should correspond to a possible language. OT thus makes typological predictions. If we assume, for instance, a universal grammar with the four constraints AFL, AFR, TROCHAIC and IAMBIC, these constraints can be ranked in 24 different ways, and in the extreme case this could lead to 24 different types of languages. We get a different type of language if TROCHAIC dominates IAMBIC, as in tableau (10): stress is on the first syllable, due to a trochaic foot that is aligned at the left edge of the word.

(10) A trochaic left-aligning language

Underlying: |σ σ σ|	TROCHAIC	IAMBIC	AFL	AFR
☞ /(σ́ σ) σ/		∗		∗
/(σ σ́) σ/	∗!			∗
/σ (σ́ σ)/		∗	∗!	
/σ (σ σ́)/	∗!		∗	

If iambic foot structure is preferred, and AFR outranks AFL, then stress will be on the last syllable in the output, as in tableau (11).

(11) An iambic right-aligning language

Underlying: |σ σ σ|	IAMBIC	TROCHAIC	AFR	AFL
/(σ́ σ) σ/	∗!		∗	
/(σ σ́) σ/		∗	∗!	
/σ (σ́ σ)/	∗!			∗
☞ /σ (σ σ́)/		∗		∗

If, however, the trochaic foot form is preferred, and AFR outranks AFL, stress will be on the second syllable, as in tableau (12).

(12) A trochaic right-aligning language

Underlying: |σ σ σ|	TROCHAIC	IAMBIC	AFR	AFL
/(σ́ σ) σ/		∗	∗!	
/(σ σ́) σ/	∗!		∗	
☞ /σ (σ́ σ)/		∗		∗
/σ (σ σ́)/	∗!			∗

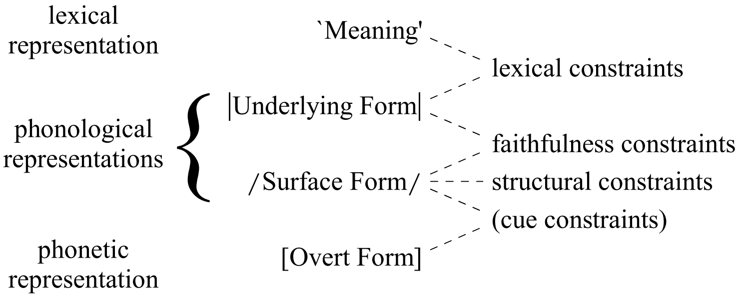
The typological possibilities of these four constraints is now exhausted. The 24 possible rankings lead to only four different types of languages, since changing the ranking of the alignment constraints with respect to the foot form constraints does not lead to any new types of languages. Within this simple set of constraints, IAMBIC only competes with TROCHAIC, and AFR with AFL. The rankings predict that in the language in (9), feet will be ranked at the left edge in every word. And since every foot in this language is iambic, stress will always be on the second syllable in a word. In the language shown in (10), stress will always be on the first syllable in a word, since here the foot form is trochaic, but feet are still ranked at the left edge of a word. The language in (11) will always have final stress, since feet are iambic and are aligned at the right edge of the word. Finally, the language in (12) will always have stress on the pre-final (*penultimate*) syllable. So whenever a linguistic principle is translated into an OT constraint, it should make typological predictions about languages. As we have seen, though, not all constraints have to be in competition with each other.

What we have also seen (and that brings us closer to the learnability problem) is that there are actually two grammars here, shown in (9) and (12), that show the same *overt* stress pattern in trisyllabic words, namely stress on the second syllable, as demonstrated by Tesar & Smolensky (1996, 1998, 2000). In other words, the two surface forms /(σ σ́) σ/ and /σ (σ́ σ)/ share the same overt form [σ σ́ σ] (at least if there are no other phonetic cues such as iambic lengthening or trochaic shortening etc.). A child can only observe stress in the overt form, but not foot structure, this is something she has to assign herself. In the case of [σ σ́ σ], the learner cannot learn the ranking of constraints of her language from trisyllabic words alone. She will crucially depend on the presence of other, either shorter or longer forms, to figure out the exact ranking. Luckily, most languages do employ words with more and less than three syllables. The next section outlines the constraint families that are used in the simulations.

2.6 Constraints for stress

The different levels of representation in figure (1) of chapter 1 are connected through different families of constraints, illustrated in figure (13). Underlying forms are connected to surface forms by *faithfulness constraints*, and to meaning by *lexical constraints*. Surface forms are restricted by *structural constraints*. Between surface forms and overt forms operate cue constraints (Boersma 1998; Escudero & Boersma 2003), which I do not discuss here.

(13) The levels of representations and their constraints



Section 2.6.1 discusses structural constraints for grammatical stress systems. Section 2.6.2 discusses the family of faithfulness constraints on stress, and section 2.6.3 discusses the family of lexical constraints for the learning of underlying forms.

2.6.1 Structural constraints for metrical phonology

Grammatical stress systems are determined by the ranking of structural constraints. The structural constraints applied in the simulations of chapter 4 and 5 are based on the constraint set of Tesar & Smolensky (2000) that contains twelve widely accepted structural constraints on metrical phenomena, listed in (14):

(14) Tesar & Smolensky's (2000) constraint set

AFL: The left edge of a foot is aligned with the left edge of a word.
AFR: The right edge of a foot is aligned with the right edge of a word.
FOOTBINARITY (FTBIN): Feet are binary on the mora or syllable level.
FOOTNONFINAL (FTNONFIN): The head syllable of a foot is not final in the foot.
IAMBIC: The rightmost syllable in a foot is the head syllable.
MAIN-LEFT (MAIN-L): The leftmost foot in a word is the head foot.
MAIN-RIGHT (MAIN-R): The rightmost foot in a word is the head foot.
NONFINALITY (NONFIN): The final syllable is not included in a foot.
PARSE: Every syllable is included in a foot.
WORD-FOOT-LEFT (WFL): The left word edge is aligned with a foot.
WORD-FOOT-RIGHT (WFR): The right word edge is aligned with a foot.
WEIGHT-TO-STRESS PRINCIPLE (WSP): Heavy syllables are stressed.

The alignment constraints AFL and AFR (McCarthy & Prince 1993a) make sure that a foot is aligned with one of the edges of a word. Their violation is *gradient*: AFL is assigned one violation mark for every syllable between the left edge of the word and the left edge of every foot. In a candidate with five light syllables such as /σ (σ̀ σ)(σ́ σ)/, AFL is violated four times: once for the first foot, three times for the second foot. In a grammar where AFL is high-ranking, the number of feet in a word is kept to a minimum, since any additional foot would incur at least one constraint violation. AFL and AFR were introduced as constraints, and therefore as a part of Universal Grammar, because languages tend to start footing at the periphery of words.

The constraints MAIN-L and MAIN-R (termed as such in Tesar & Smolensky 2000; based on EDGEMOST by Prince & Smolensky 1993, and ALIGN(PrWd, Edge, H(PrWd), Edge) by McCarthy & Prince 1993b) are similar to AFL and AFR, but apply only to the foot that contains the main stress. Thus, a candidate like /σ (σ̀ σ)(σ́ σ)/ violates MAIN-L three times, while it violates MAIN-R not at all. MAIN-L and -R have their raison d'être as part of the universal constraint set because languages might differ whether they assign main stress to the first or the last foot within a word, rather than assigning main stress to the *middle* foot of a word with e.g. three feet, */(σ̀ σ)(σ́ σ)(σ̀ σ)/.

The two WORDFOOT-alignment constraints WFL and WFR (McCarthy & Prince 1993b) favour candidates in which at least one foot is aligned with the word edge. This is different from AFL/AFR and MAIN-L/MAIN-R. WFL and WFR are not gradient, but *binary*: they are assigned a single violation regardless of how many syllables are between the word and the foot edge. Thus, a candidate like /σ σ (σ́)(σ̀ σ)/ violates WFL only once, while it would violate e.g. AFL twice. The same candidate violates AFR twice, but not WFR.

The constraint NONFINAL (inferred from Liberman & Prince 1977, Hayes 1980; introduced as OT constraint by Prince & Smolensky 1993) expresses extrametricality: it is violated if the last syllable is parsed (included) in a foot. This constraint prefers /(σ́) σ/ over /(σ́ σ)/. Note that WFR and NONFINAL have *complementary violations* on the word level: a word that violates WFR does not violate NONFINAL, and a word that violates NONFINAL does not violate WFR. Syllable extrametricality (i.e. when the final syllable does not count in the assignment of metrical structure) is a well-known phenomenon in metrical processes, but extrametricality as such can apply to other constituents such as segments or feet. The rationale behind such a constraint might be that the least salient part of the word is ignored, or that it leaves more room for boundary tones, if you exclude the final element from the metrical analysis (Norval Smith, p.c.).

The constraint PARSE (Prince & Smolensky 1993, McCarthy & Prince 1993) favours candidates in which all syllables are parsed into feet. It is assigned one violation mark for each unfooted syllable. Thus, the candidate /σ (σ́ σ) σ σ/ violates PARSE three times. This constraint, if high-ranking, maximizes the number of feet (or the size of a foot): if it outranks both AFL and AFR, the language tends to have secondary stress. If AFL or AFR outranks PARSE, words tend to contain a single foot with main stress. It is practically inevitable to assume a constraint like PARSE if one wants to explain stress systems that have secondary stress.

The constraint FTNONFIN (Tesar 1998) favours candidates with trochaic feet such as /(σ́ σ)(σ̀ σ)/. However, *degenerate feet* consisting of only one syllable, like /(σ́)/, violate this constraint. The constraint IAMBIC favours candidates with iambic feet like /(σ σ́)/, and this constraint is *not* violated in degenerate feet like /(σ́)/. This asymmetry in the formulation of the two foot form constraints (other than in TROCHAIC/IAMBIC we have seen in section 2.5) leads to complementary violations on the foot level: FTNONFIN is assigned one violation mark for each occurrence of /(σ σ́)/ *or* /(σ́)/, and IAMBIC for each occurrence of /(σ́ σ)/.

The WSP (Prince 1990; OT-version in Prince & Smolensky 1993) favours candidates that have stress on a heavy syllable (represented as 'H'; light syllables are represented as 'L'; primary stress is represented with '1', and secondary stress is represented with '2'). Every heavy syllable that is not stressed in a form causes a violation of this constraint. Thus, /(L2 H) H (H1) L/ violates WSP twice (once for the unfooted H, once for the H in the first foot's weak position). A candidate like /(L H2)(H2)(H1) L/ does not violate WSP, because each heavy syllable carries either primary or secondary stress. Like PARSE, this constraint tends to maximize the number of feet (though less strongly), and it prefers sequences of heavy monosyllabic feet like (H1)(H2) to superheavy feet like (H1 H). This constraint can also override the foot form constraints: even in a basically trochaic language (FTNONFINAL/TROCHAIC >> IAMBIC), a high-ranked WSP can force the occurrence of an iambic foot (L H1). The rationale behind this is the earlier mentioned phenomenon of making salient syllables even more prominent.

FTBIN (Prince & Smolensky 1993) is the constraint on foot size: feet should be binary with respect to either syllables or moras. This constraint is assigned a violation mark for every monosyllabic light foot, i.e. (L1) and (L2), whereas feet like (L1 H) and (H H2) do not violate this constraint (for the purpose of this dissertation, I do not admit feet in candidates with more than two syllables in GEN).

The list of structural constraints for stress is not exhaustive, but captures a great deal of metrical phenomena. Some add-ons will become necessary for modelling Latin and Pintupi stress, and will be introduced in sections 4.2.2.3, 5.2.2 and 5.2.3. The constraints used in the simulations of lexical stress systems are discussed next.

2.6.2 Faithfulness constraints for lexical stress

The metrical constraints presented in the section above are structural constraints and hold on surface forms. In languages without lexical stress, structural constraints is all you need: their ranking will tell you where to stress a word. Stress is less (or not at all) predictable from the grammar if lexical specifications come into play. In languages with lexical stress, morphemes can be underlyingly stressed. In interaction with structural constraints, these underlying specifications determine the stress in the surface form. This is ensured by faithfulness constraints. The faithfulness constraints I employ in chapter 6 are based on Correspondence Theory (McCarthy & Prince 1995:14):

(15) Correspondence:

> Given two strings S_1 and S_2, **correspondence** is an relation $\Re$ from the elements of S_1 to those of S_2. Elements $\alpha \in S_1$ and $\beta \in S_2$ are referred to as **correspondents** of one another when $\alpha \Re \beta$.

Following Revithiadou (1999), I assume faithfulness constraints that relate underlying stress to stress in surface forms. Underlying stress has a correspondent in the surface form (ensured by MAX), and stress in the surface form has a correspondent in the underlying form (ensured by DEP). Contrary to Revithiadou (1999), who paints a bigger picture of morphologically determined stress, I content myself for the purpose of this dissertation to the four simplified constraints in (16), which make a difference between faithfulness to lexically stressed roots, and faithfulness to lexically stressed affixes.

(16) Faithfulness to stress:

MAX(Root): A root that is stressed in the underlying form is also stressed in the surface form.

DEP(Root): A root that is stressed in the surface form is also stressed in the underlying form.

MAX(Affix): An affix that is stressed in the underlying form is also stressed in the surface form.

DEP(Affix): An affix that is stressed in the surface form is also stressed in the underlying form.

The MAX constraints are violated whenever a stress of the underlying form is not realized in the surface form. The DEP constraints are violated whenever there is a stress inserted on the surface form although the corresponding underlying form does not have stress.

In the next chapter I introduce the concept of the lexical constraints that enable the learning of underlying forms. I argue that faithfulness constraints are not enough for the learning of underlying forms, because with faithfulness and structural constraints alone a new underlying form would be created for alternating surface forms. A form /(tá$_{\mu\mu}$k)/ 'day' would faithfully be mapped onto an underlying form |taːk|, and a form /(tá$_{\mu\mu}$.gə$_{\mu}$)/ 'day-Nom.Pl' would faithfully be mapped onto an underlying form |taːgə|. A learner would not necessarily realize that the two forms share the same morpheme, but create two allomorphs |taːk| and |taːgə| for it without even knowing that they are allomorphs. I propose that the learner needs to be able to make a connection between these forms, which is achieved by linking the form to the meaning 'day', as outlined in the next section.

2.6.3 Lexical constraints for underlying stress

As will become evident in chapter 6, we need more and other constraints than metrical and faithfulness constraints to model lexically assigned stress, and underlying forms in general. These *lexical* constraints link *meaning* to underlying forms and are based on Boersma (2001) and Escudero (2005); unlike Boersma's and Escudero's proposed constraints, the lexical constraints I assume are formulated in a *do not connect the meaning 'XY' to an underlying form |xy|*-manner. For instance, there are constraints as *don't connect the meaning 'dog' to the form |kæt|*, which militate against the connection of a meaning to a form that is already occupied with another meaning, in this case 'cat'. But there are also constraints as *don't connect the meaning 'cat' to the form |kæt|*. In the comprehension process (outlined in §3.2 and 3.3), these constraints suppress, and thereby guide, lexical *access*. In the production process (outlined in §3.4), these constraints suppress, and thereby guide, lexical *retrieval*. Depending on the ranking of these constraints, there will only be one optimal underlying form for a meaning. The learner needs to acquire a ranking of these constraints that will enable her to compute the appropriate forms for communication. These lexical constraints need not be innate but could rather be very language-specific. The learner needs to be equipped with the *ability* to create these constraints, and that these constraints only emerge on the moment the learner encounters a new meaning along with a new form. In this sense, these constraints are hypotheses about how an underlying does *not* look like. Since these constraints have a very language-specific character, they will be discussed in more detail in section 6.3.1, where they are applied to Modern Greek.

2.7 The form of overt forms

We have encountered the grammatical restrictions in this chapter that are necessary for modelling stress. A note on the shape of overt forms is advisable. For the purposes of chapter 3, overt forms are represented with ‘σ’-symbols like [σ σ́ σ]. In the computer simulations in chapters 4 and 5, the input to the learners is always labelled for syllable and word boundaries. In the chapter on Latin, the input consists of word-like forms with syllables labelled as heavy (‘H’) or light (‘L’). In the chapter on Pintupi, the input consists of word-like forms with syllables labelled as consonantal (‘C’) and vocalic (‘V’) structure. In the chapter on Modern Greek, the input consists of Modern Greek words where the C- and V-slots are actually filled with segmental information.

All inputs have in common that they are overt in the sense that with respect to metrical information, they are only labelled for main and secondary stress, but not for foot structure. The virtual learners have to choose the appropriate foot structure (and sometimes the appropriate moraic structure) themselves.

2.8 Summary

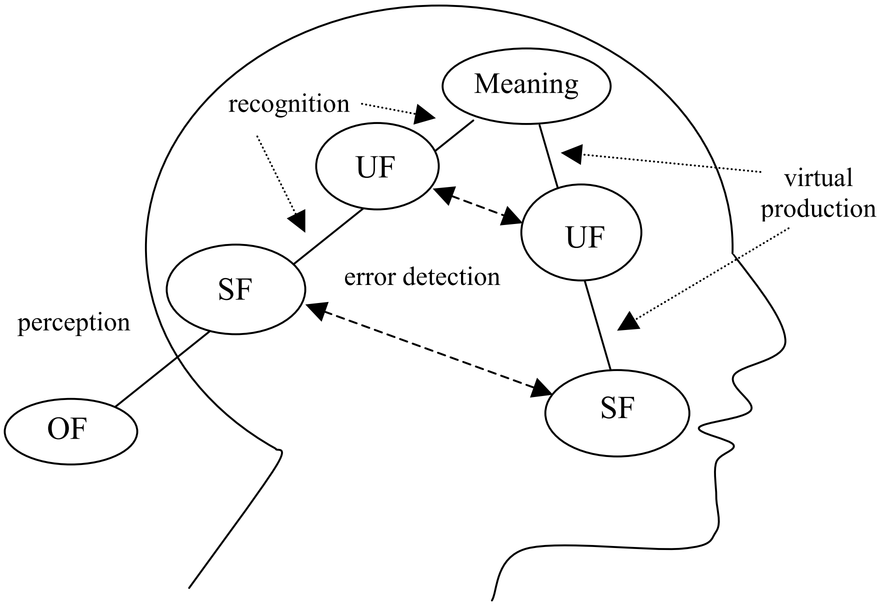
This chapter provided the linguistic ingredients to the computer simulations of chapters 4, 5 and 6. I outlined the distinction between grammatical and lexical stress systems, and between weight-sensitive and weight-insensitive stress systems. To model grammatical stress, structural constraints are needed, while for the modelling of lexical stress, faithfulness and lexical constraints play a crucial role. The following chapter discusses the ingredients for learnability. Section 3.1 provides an introduction to learnability in OT. The different modelling processes perception, recognition, virtual production and error detection are discussed in sections 3.2 to 3.5. The two different reranking strategies I apply and compare are discussed in section 3.6. The case of learning from pairs of surface and underlying forms is outlined in section 3.7. I summarize in section 3.8.

3 The learnability of hidden structure and the grammar

3.1 Introduction

A language-acquiring child needs to learn the grammar and the lexicon of her parent's language in order to become a proper communicator. Within OT, the grammar is defined as the constraint ranking of a language. The lexicon is usually seen as pairs of form and meaning. In OT, the learning task is usually regarded as learning the constraint ranking, abstracting away from the problem of learning the lexicon. I argue that at least the form-part of the lexicon is acquired by means of constraints, i.e. the grammar, and that underlying forms have to be computed, much in the same way as surface forms have to be computed. I build upon the assumption that the child uses the same grammar for comprehension as for production (Smolensky 1996). Tesar & Smolensky (1996, 1998, 2000) distinguished between overt forms with surface structure and underlying forms that play a role in learning. These forms are connected to each other by the grammar. Tesar & Smolensky decomposed the learning process into three components: a device that enables the learner to interpret an incoming form and assign it a structure (*Robust Interpretive Parsing*, Tesar & Smolensky 1996, 1998, 2000; Tesar 1997), a device for changing the grammar (*Constraint Demotion*, Tesar 1995), and a device for deriving forms for the lexicon (*Lexicon Optimization*, Prince & Smolensky 1993; Itô et al. 1995). I argue in this chapter that the function of RIP can be furthermore applied to the learning of underlying forms, and that the grammar is changed in a *Gradual-Learning-Algorithm* manner. What follows is an outline of the learning process in the terminology of functional phonology (Boersma 1998 et seq.). I distinguish between the three representational levels of overt forms, surface forms and underlying forms, and a fourth level of 'meaning' that I deem necessary for the learning of underlying forms, as illustrated in figure (17):

(17) The learning process in a scheme[12]



In the case of adults, who already know the grammar of the language, there will be no mismatch between what they hear and what they produce (in an idealized situation). In the case of the language learner, who wants to become a good listener and a good speaker, the grammar will be under-developed and will yield mismatches in comprehension and production. By comparing what she herself would say for a given form to what she hears, she is able to find a possible mistake in her grammar (or in OT terms: to find a mistake in her constraint ranking). This is indicated in figure (17), where the dashed lines indicate the comparison between the underlying form in comprehension and the underlying form in production, and the comparison between the surface form in comprehension with the surface form in production.

To model and understand the comprehension process (the listener's perspective) we need to distinguish between *perception* as outlined in §3.2,

---

[12] Due to spatial reasons, I left out an additional bubble in the production process: the surface form in production should be accompanied by an overt form. For the purpose of this dissertation, though, the overt form in production will have stress on the same syllable as the corresponding surface form, so it can be derived in a trivial way from the surface form in production. It feels therefore licit to exclude this representation.

recognition as outlined in §3.3, and *virtual production* as outlined in §3.4. The trigger to change the grammar, *error detection*, is outlined in §3.5. The means to change the grammar is *constraint reranking*. I will outline two strategies for the reranking of constraints in §3.6: Constraint Demotion (CD; Tesar 1995) and the Gradual Learning Algorithm (GLA; Boersma 1997). I will furthermore discuss the case of learning from pairs of surface forms and underlying forms in §3.7. Section 3.8 provides a short summary.

3.2 Perception

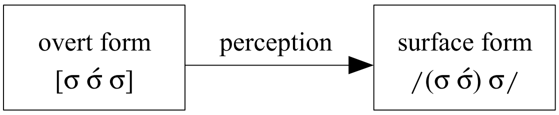
The input to perception is the auditory signal, the *overt form*. For my purposes of learning stress only, the overt form does not contain phonetic detail, but phonemic or syllabic representations, as well as syllable boundaries and stress marks. For instance, in an overt form like [σ σ́ σ], the symbol 'σ' stands for a syllable, and the symbol 'σ́' stands for a stressed syllable with primary stress. Stress is directly observable in the form of pitch, intensity, duration, or vowel quality. Languages make use of some or even of all of these phonetic cues.[13] The overt form lacks 'hidden' structure like feet and moras. Hidden structure has to be assigned by the listener. Having established in chapter 2.2 how syllables and moras are grouped into higher constituents, namely feet, it has to be stressed that feet and moras are not directly observable in a speech stream.[14] Hidden structure is represented in the *surface form*, e.g. /(σ σ́) σ/: this surface form has the syllabic and stress information that the overt form has, but additionally it has foot structure. This mapping from overt form onto surface form is what Tesar & Smolensky (1996) and Tesar (1997) call Robust Interpretive Parsing (henceforth RIP) and what Boersma (1998:269) calls *perception*. This is illustrated in figure (18):

---

[13] Languages do not only differ in their stress patterns; they also differ in their use of phonetic cues for encoding stress. The acquisition of the language-specific use of stress cues could be modelled as well, but is beyond the scope of this dissertation.

[14] Nor can syllables be observed as such for that matter, but a full account of the learnability of syllables is beyond the scope of this dissertation. A partial account for the learnability of syllable structure, namely the acquisition of coda-moraicity, is given in chapter 5 on Pintupi stress.

(18) Perception



RIP or perception is an interpretation of the overt form: the listener hears a speech signal, for instance a trisyllabic word [σ σ́ σ], with primary stress on the second syllable. This form is ambiguous in terms of foot structure: it could have an iambic foot structure /(σ σ́) σ/ in languages such as the one we have seen in (9) of chapter 2. It could also have a trochaic foot structure /σ (σ́ σ)/ like the language in (12) of chapter 2. Or it could have a monosyllabic foot /σ (σ́) σ/ in yet another language. The grammar of a listener will tell her how to interpret the overt form. In RIP (or perception), the listener applies the constraint ranking she uses in production also in perception. GEN will give her three possible candidates for an overt form [σ σ́ σ], listed in tableau (19). She will interpret this form by applying her current constraint ranking. Imagine that her grammar consists of four constraints AFL, AFR, Iambic, and Trochaic, ranked as in (19). In this case high-ranked AFL rules out the candidate with the monosyllabic foot in (19a) and the candidate with a trochaic foot in (19c). The candidate with an iambic foot /(σ σ́) σ/ wins (indicated with ‘ 👂 ’ in the tableau): this is the form that the listener *thinks she hears*. The candidate list consists of pairs of overt forms and surface forms. A candidate overt form has always the same stressed syllable than the corresponding surface form; there will never be a candidate like [σ́ σ σ] /σ (σ́) σ/, where stress is on e.g. the initial syllable in the overt form, but on e.g. the second syllable in the surface form.

(19) Perception (RIP) in OT

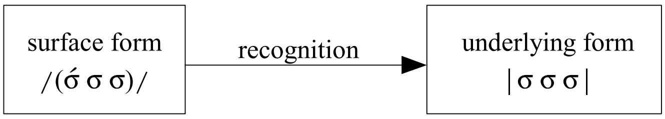
overt form: [σ σ́ σ]	AFL	AFR	Trochaic	Iambic
a. /σ (σ́) σ/ [σ σ́ σ]	*!	*		
👂 b. /(σ σ́) σ/ [σ σ́ σ]		*	*	
c. /σ (σ́ σ)/ [σ σ́ σ]	*!			*

For readability reasons I will often refrain from including the overt forms in the candidate lists, since it can be trivially computed from the surface form (by taking away the foot structure). The next section shows the recognition process.

3.3 Recognition and comprehension

In the recognition step, a listener looks up the just perceived surface form in the lexicon. Tableau (19) showed the form that the learner perceived. In order to *recognize* the word, she needs to look it up in her lexicon by mapping the perceived surface form onto an underlying form:

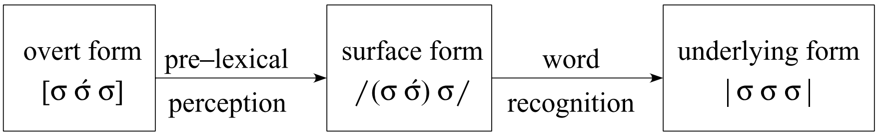
(20) Recognition



In the approaches to the learning of metrical surface forms so far (Tesar & Smolensky 1996, 1998, 2000; Apoussidou & Boersma 2003, 2004ab, also in chapters 4 and 5 of this book), the mapping from the surface structure to the underlying form is trivial: the syllabic and segmental information is kept, and surface structure such as stress marks and feet are stripped off.

The more complete mapping from overt form to surface form to underlying form is called *comprehension*, illustrated in figure (21). The process of perception is comparable to the notion of pre-lexical perception in psycholinguistics, while the recognition step is comparable to the notion of word recognition (e.g. McQueen & Cutler 1997).

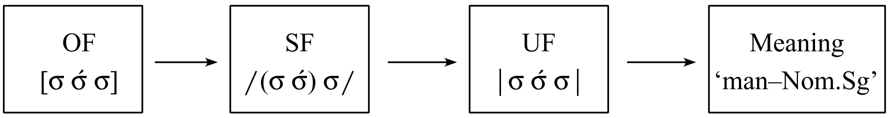
(21) Comprehension including the underlying form



For the purposes of chapters 4 and 5, the mapping from surface form to the underlying form is not evaluated by the grammar, in opposition to the mapping from overt form onto surface form that is evaluated by the grammar, as shown in tableau (19).

For the purposes of chapter 6, the mapping from surface form to the underlying form is also evaluated by the grammar. I argue that we need to take the interpretation function one level higher, and include *meaning* in the comprehension process, illustrated in figure (22), to get even closer to a realistic model of comprehension and language acquisition. The term ‘meaning’ as used here covers both the semantic content and the syntactic function of a word.

(22) Comprehension including meaning



The input to the comprehension process is again the overt form, as in perception alone, but the candidates of the evaluation are now quadruplets consisting of the meaning of the form (represented as e.g. ‘man-Nom.Sg’) with combinations of different underlying forms, surface forms, and overt forms. Thereby, the determination of underlying forms becomes the task of GEN and Freedom of Analysis (Prince & Smolensky 1993:6, McCarthy & Prince 1993b:21). I exclude the possibility of homonymy and assume that in the comprehension process, the meaning is given to the learner. Therefore, the meaning will always be the same in the candidates of a comprehension tableau. The overt form will also be the same in all the candidates of a comprehension tableau, because there is only one overt form for each word, as we have seen in the perception evaluation in (19). This is illustrated in the comprehension tableau in (23) with the input [σ σ́ σ]. There are two possible surface forms and two possible underlying forms. For each possible underlying form, there is a lexical constraint militating against it, in this case *|σ σ σ| ‘man-Nom.Sg’ *don’t connect the meaning ‘man-Nom.Sg’ to the underlyingly unstressed form* |σ σ σ| and *|σ σ́ σ| ‘man-Nom.Sg’ *don’t connect the meaning ‘man-Nom.Sg’ to an underlyingly stressed form* |σ σ́ σ|. For each surface form in the candidates, there is a structural constraint militating against it (TROCHAIC/IAMBIC, AFR/AFL). The correspondence between a surface form and an underlying form is given by faithfulness (MAX/DEP(Stress)). Given the ranking in tableau (23), the candidate with an underlyingly stressed form and a right-aligned trochaic surface form |σ σ́ σ| /σ (σ́ σ)/ is chosen (indicated by ‘ ☞ ’). The first and third candidates are ruled out, because their surface forms violate high-ranked TROCHAIC. The last candidate is ruled out because its underlying form violates the next-ranked lexical constraint *|σ σ σ| ‘man-Nom.Sg’, which militates against underlyingly unstressed forms.

(23) The evaluation of comprehension

[σ σ́ σ]	TROCHAIC	*|σ σ σ| 'man-Nom.Sg'	*|σ σ́ σ| 'man-Nom.Sg'	DEP(Stress)	IAMBIC	AFL	MAX(Stress)	AFR
'man-Nom.Sg' |σ σ σ| /(σ σ́) σ/ [σ σ́ σ]	*!	*		*				*
'man-Nom.Sg' |σ σ σ| /σ (σ́ σ)/ [σ σ́ σ]		*!		*	*	*		
'man-Nom.Sg' |σ σ́ σ| /(σ σ́) σ/ [σ σ́ σ]	*!		*					*
👂 'man-Nom.Sg' |σ σ́ σ| /σ (σ́ σ)/ [σ σ́ σ]			*		*	*		

For an adult language user, the forms she computes in the production process are the same as in the comprehension process. For the language learning child with an under-developed grammar, this might not be the case. The child applies the same mechanisms in comprehension as the adult listener, but she might compute different forms in the comprehension process than in the production process. As soon as a learner has the perceived form and the underlying form of the comprehension process under her belt, she will compute what she would produce herself for this lexical item: she will *virtually produce* it. This is demonstrated next.

3.4 Virtual production

Virtual production is the key for the learner to find out whether her grammar needs to be adjusted, and also how the grammar needs to be adjusted. From listening alone the learner does not know whether her grammar is correct. She needs to compute what she herself would have said given the recognized form or meaning. In this way she can compare her own produced form to what she perceived and recognized. She deems her perceived form correct, and her produced form incorrect, and will strive to bring her production closer to her perception.

For the purposes of the computer simulations in chapters 4 and 5, the input to virtual production is traditionally the underlying form, as became clear from figure (21): the comprehension process in (21) ends in the underlying form. Virtual production from the underlying form is outlined in section 3.4.1, and applied in section 4.5 for Latin stress and in section 5.3.2 for Pintupi stress. For the purposes of chapter 6, the input to virtual production will be meaning, as became clear from figure (22): the comprehension process in (22) ends in the meaning of the form. Virtual production from meaning is outlined in section 3.4.2.

3.4.1 Virtual production from the underlying form

Virtual production from the underlying form connects to the comprehension process shown in figure (21) and is illustrated in tableau (24). In this virtual production step, we see that more candidates become available to the evaluation than there are in the comprehension step in tableau (19), simply because there are more surface structures competing for the underlying form |σ σ σ| than there are surface structures competing for the overt form [σ σ́ σ] in the comprehension step.[15] In tableau (24) we see a virtual production tableau, where the optimal surface form for the underlying form |σ σ σ| is evaluated. The candidates consist of triplets of underlying forms/surface forms/overt forms. The underlying forms are the same for each candidate. The overt forms can straightforwardly be computed form the surface forms by stripping off the foot structure.

(24) Virtually producing |σ σ σ|:

underlying form: |σ σ σ|	AFL	AFR	TROCHAIC	IAMBIC
a. |σ σ σ| /σ (σ́) σ/ [σ σ́ σ]	*!	*		
b. |σ σ σ| /(σ σ́) σ/ [σ σ́ σ]		*	*!	
c. |σ σ σ| /σ (σ́ σ)/ [σ σ́ σ]	*!			*
☞ d. |σ σ σ| /(σ́ σ) σ/ [σ́ σ σ]		*		*
e. |σ σ σ| /σ (σ σ́)/ [σ σ σ́]	*!		*	
f. |σ σ σ| /(σ́) σ σ/ [σ́ σ σ]		**!		
g. |σ σ σ| /σ σ (σ́)/ [σ σ σ́]	*!*			

---

[15] Note that the candidate set in perception is a subset of the one in production, at least for the cases under discussion in this book.

Candidate (24d), the triplet of |σ σ σ| /(σ́ σ) σ/ [σ́ σ σ], is the winner because it satisfies high-ranked AFL. The competing candidate (24b), the triplet |σ σ σ| /(σ σ́) σ/ [σ σ́ σ], (which contains the surface form/overt form pair that was chosen as optimal in perception) is less harmonic because it violates lower-ranked TROCHAIC. The learner will proceed to compare the surface form of the virtual production step, i.e. (24d) to the surface form that she perceived, i.e. (24b). If they match, i.e. if they are identical, she will not change her grammar. If they do not match, she will adjust her grammar. This is demonstrated in sections 3.5 and 3.6. But first, the virtual production step from meaning is discussed.

3.4.2 Virtual production from meaning

Virtual production from meaning connects to the comprehension process shown in figure (22) and is illustrated in tableau (25). In this virtual production process, overt form, surface form *and* underlying form have to be computed by the grammar, given meaning. All candidates with iambic feet are ruled out by high-ranked TROCHAIC. All candidates with underlyingly unstressed forms are ruled out by high-ranked *|σ σ σ| ‘man-Nom.Sg’.

(25) Evaluation of production, given meaning

‘man-Nom.Sg	TROCHAIC	*|σ σ σ| ‘man-Nom.Sg’	*|σ σ́ σ| ‘man-Nom.Sg’	DEP(Stress)	IAMBIC	AFL	MAX(Stress)	AFR
‘man-Nom.Sg’ |σ σ σ| /(σ σ́) σ/ [σ σ́ σ]	*!	*		*				*
‘man-Nom.Sg’ |σ σ σ| /σ (σ́ σ)/ [σ σ́ σ]		*!		*	*	*		
‘man-Nom.Sg’ |σ σ́ σ| /(σ σ́) σ/ [σ σ́ σ]	*!		*					*
‘man-Nom.Sg’ |σ σ́ σ| /σ (σ́ σ)/ [σ σ́ σ]			*		*	*!		
‘man-Nom.Sg’ |σ σ σ| /(σ́ σ) σ/ [σ́ σ σ]		*!		*	*			*
‘man-Nom.Sg’ |σ σ σ| /σ (σ σ́)/ [σ σ σ́]	*!	*		*		*		
👄 ‘man-Nom.Sg’ |σ σ́ σ| /(σ́ σ) σ/ [σ́ σ σ]			*		*		*	*
‘man-Nom.Sg’ |σ σ́ σ| /σ (σ σ́)/ [σ σ σ́]	*!		*			*	*	

This leaves two candidates for competition, one with a left-aligned trochee, and one with a right-aligned trochee. Both candidates have underlyingly unstressed forms. Low-ranked AFL decides in favour of the candidate with the left-aligned trochee, ‘man-Nom.Sg’ |σ σ́ σ| /(σ́ σ) σ/ [σ́ σ σ].

We established what the learner perceived (§3.2) and what the learner recognized (this section). What is missing now is the comparison between what the learner perceived or recognized, and what she produces. As will be outlined in sections 3.5 and 3.6, the learner will adjust her grammar in response to the detection of an error.

3.5 Error detection

To detect whether her grammar matches the target language the learner wants to acquire, she needs to compare her production to what she perceived and recognized. If there is a mismatch between the comprehended form and the produced form, the learner detects an *error* (Wexler & Culicover 1980:127, Tesar 1995). Error detection in virtual production from underlying form compares the overt and surface form of perception with overt and surface form in production, as outlined in section 3.5.1. Error detection in virtual production from meaning compares the overt, surface, and underlying form of the comprehension step with the overt, surface, and underlying form in the production step, outlined in section 3.5.2.

3.5.1 Error detection in virtual production from underlying form

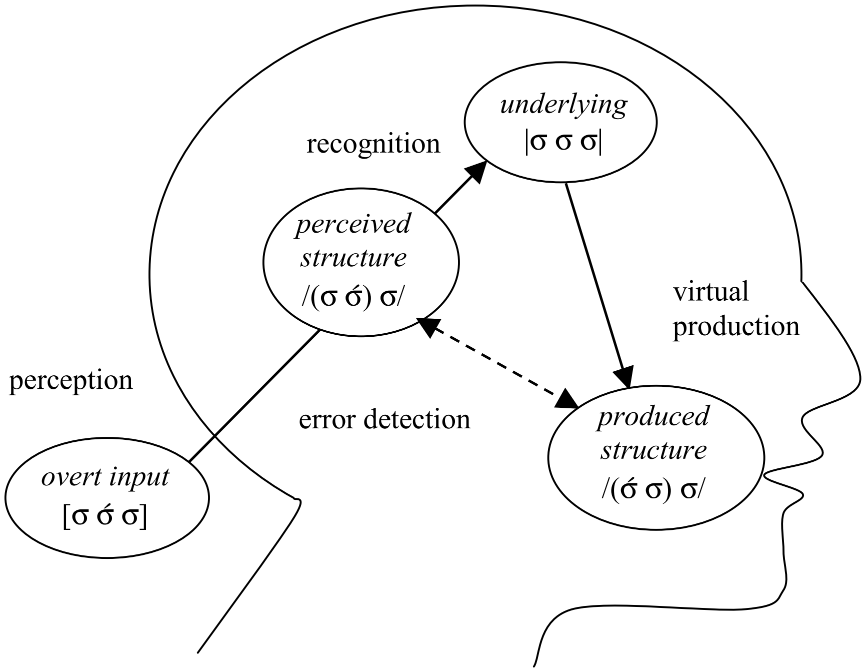
In virtual production from underlying form, the underlying form is the input to the production evaluation in tableau (24), which I repeat as tableau (26). In this virtual production step, the grammar of the learner gives out a triplet with a trochaic, left-aligned foot /(σ́ σ) σ/, indicated by ‘👄’. The form that the learner perceived, though, is a different one, as we have seen in the perception tableau in (19): the perceived form has a left-aligned iamb, /(σ σ́) σ/. Perceived/ recognized forms are henceforth indicated by ‘👂’. There is a mismatch between the perceived and the produced form, which means that the learner detected an error.

(26) Error detection, given underlying form

underlying form: |σ σ σ|	AFL	AFR	TROCHAIC	IAMBIC
a. |σ σ σ| /σ (σ́) σ/ [σ σ́ σ]	*!	*		
👂 b. |σ σ σ| /(σ σ́) σ/ [σ σ́ σ]		*	*!	
c. |σ σ σ| /σ (σ́ σ)/ [σ σ́ σ]	*!			*
👄 d. |σ σ σ| /(σ́ σ) σ/ [σ́ σ σ]		*		*
e. |σ σ σ| /σ (σ σ́)/ [σ σ σ́]	*!		*	
f. |σ σ σ| /(σ́) σ σ/ [σ́ σ σ]		**!		
g. |σ σ σ| /σ σ (σ́)/ [σ σ σ́]	*!*			

The error detection elicits an adjustment of the grammar, which means in OT that the constraints are reranked. The learning process with overt forms, surface forms and underlying forms is illustrated in figure (27):

(27) The “small” comprehension/production loop



I discuss two possibilities of grammar adjustment that can follow on error detection. The two reranking strategies are outlined in section 3.6; but first we will have a look at error detection in virtual production form meaning.

3.5.2 Error detection in virtual production from meaning

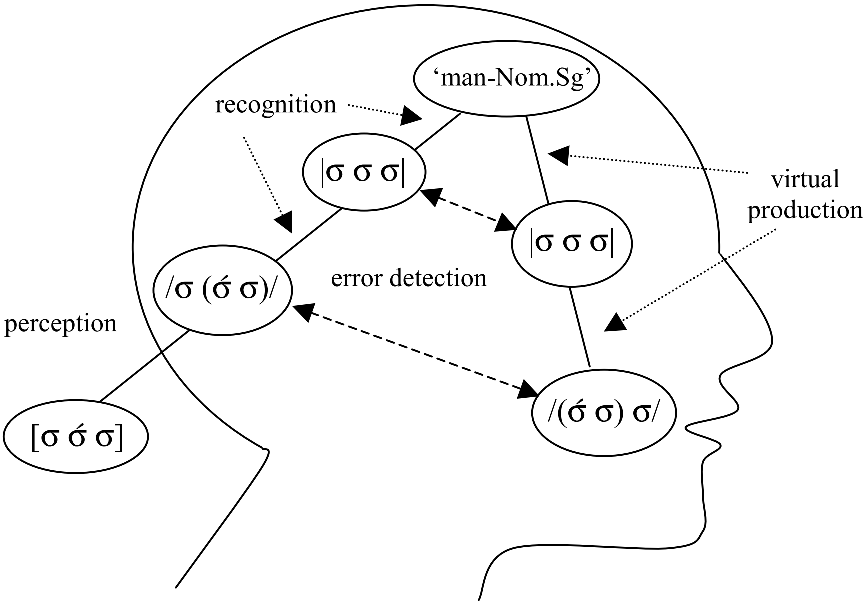
In (28) we see the same tableau as in (25), except that this time both the winning candidate in production (marked with ‘👄’) as well as the candidate of the comprehension process (marked with ‘👂’) are marked. The winning candidate in comprehension differs from the one in production only in the surface forms (and necessarily in their overt forms as well, because I do not allow a mismatch between a surface form and its overt form). In general, the competing candidates can differ in either underlying form or surface form or in both. Any deviation will elicit error detection, and therefore constraint reranking.

(28) Error detection, given meaning

‘man-Nom.Sg	TROCHAIC	*|σ σ σ| ‘man-Nom.Sg’	*|σ σ́ σ| ‘man-Nom.Sg’	DEP	IAMBIC	AFL	MAX	AFR
‘man-Nom.Sg’ |σ σ σ| /(σ σ́) σ/ [σ σ́ σ]	*!	*		*				*
‘man-Nom.Sg’ |σ σ σ| /σ (σ́ σ)/ [σ σ́ σ]		*!		*	*	*		
‘man-Nom.Sg’ |σ σ́ σ| /(σ σ́) σ/ [σ σ́ σ]	*!		*					*
👂 ‘man-Nom.Sg’ |σ σ́ σ| /σ (σ́ σ)/ [σ σ́ σ]			*		*	*!		
‘man-Nom.Sg’ |σ σ σ| /(σ́ σ) σ/ [σ́ σ σ]		*!		*	*			*
‘man-Nom.Sg’ |σ σ σ| /σ (σ σ́)/ [σ σ σ́]	*!	*		*		*		
👄 ‘man-Nom.Sg’ |σ σ́ σ| /(σ́ σ) σ/ [σ́ σ σ]			*		*		*	*
‘man-Nom.Sg’ |σ σ́ σ| /σ (σ σ́)/ [σ σ σ́]	*!		*			*	*	

The whole process of comprehension and virtual production including meaning is illustrated in figure (29): the underlying form in perception is compared to the underlying form in production, and the surface form in perception is compared to the surface form in production. If the underlying forms differ from another, or if the surface forms differ from another, the learner detects an error that elicits constraint reranking.

(29) The "big" comprehension/production loop



Once the learner detects an error, she will proceed to adjust her grammar in order to make the produced form more likely to match the perceived form the next time she encounters it. In the following I describe two different ways to rerank constraints in learning that have been proposed in the OT literature.

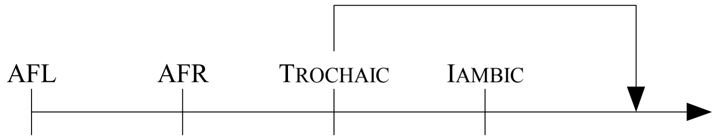
3.6 The reranking strategies

The two reranking strategies that I discuss here are Constraint Demotion (CD; Tesar 1995) and the Gradual Learning Algorithm (henceforth GLA; Boersma 1997). Both reranking strategies can be combined with the error detection procedure as described above, but where CD only allows constraint demotion, the GLA also allows constraint promotion.

3.6.1 Constraint Demotion

The reranking strategy of Constraint Demotion makes use of an ordinal ranking scale and as the name already indicates it only moves constraints *down* the hierarchy, not up. Constraints in standard OT occupy *strata*. This is indicated in figure (30): the four constraints occupy four distinct strata on the ranking hierarchy (where the left edge of the arrow marks the top of the hierarchy). In the learning process, constraints are demoted along the ranking scale to lower strata, in response to error detection:

(30) Constraint demotion



How does the detection of an error tell the language learner which constraints she has to demote? She will focus on the two crucial forms: the perceived and the produced form in (28), both repeated in tableau (31). The present ranking is AFL >> AFR >> TROCHAIC >> IAMBIC (which is not the initial ranking, but a ranking after some learning already took place). The learner wants to make the perceived form (the target form) more harmonic than her own form (the produced form). As soon as the learner detects an error (as seen in the production-comprehension mismatch shown in tableau (31)), she looks up the highest-ranked constraint in her grammar that prefers the perceived form. Here, this constraint is IAMBIC (indicated by a ‘√’ in the column of IAMBIC). She will also look up all the constraints that prefer the produced form, and are ranked at least as high as Iambic. This is TROCHAIC in this case (indicated by a ‘√’ in the column of TROCHAIC). She demotes all constraints that prefer the produced form (here: TROCHAIC) directly below the stratum of the constraint that prefers the perceived form (here: IAMBIC; this stratum can be occupied by a lower ranked constraint).

(31) A constraint-demotion tableau

underlying: |σ σ σ|	AFL	AFR	TROCHAIC	IAMBIC	
👂 |σ σ σ| /(σ σ́) σ/ [σ σ́ σ]		*	*!	√	
👄 |σ σ σ| /(σ́ σ) σ/ [σ́ σ σ]		*	√	*	

In this way the learner will make it more likely that the perceived form matches with the produced form next time she encounters this lexical item. With CD, constraints are demoted minimally, i.e. not further down the hierarchy than absolutely necessary. We can check whether the two forms match in a new perception/virtual production loop. In a new perception evaluation with the new ranking (AFL >> AFR >> Iambic >> Trochaic), the learner still perceives [σ σ́ σ] as having iambic foot structure /(σ σ́) σ/, just as she did in tableau (19).[16]

(32) New perception of [σ σ́ σ]

overt: [σ σ́ σ]	AFL	AFR	Iambic	Trochaic
a. /σ (σ́) σ/	*!	*		
👂 b. /(σ σ́) σ/		*		*
c. /σ (σ́ σ)/	*!		*	

The surface form /(σ σ́) σ/ will again be (trivially) recognized as |σ σ σ|, as in §3.2 . If she virtually produces this form with the adjusted ranking, as we can see in tableau (33), the winner is the same form as in perception: a left-aligned iamb /(σ σ́) σ/.[17] The algorithm brought perception and production into agreement.

(33) New production of |σ σ σ|

underlying: |σ σ σ|	AFL	AFR	Iambic	Trochaic
a. /σ (σ́) σ/	*!	*		
👂 👄 b. /(σ σ́) σ/		*		*
c. /σ (σ́ σ)/	*!		*	
d. /(σ́ σ) σ/		*	*!	
e. /σ (σ σ́)/	*!			*
f. /(σ́) σ σ/		**!		
g. /σ σ (σ́)/	*!*			

---

[16] According to tableau (19), the candidates should include overt forms; they are excluded from the tableau because they are equal to the overt form in the input cell.

[17] According to tableau (23), the candidates should include underlying forms and overt forms because the underlying forms are equal to the one in the input cell and the overt forms look like the surface forms without foot structure.

This grammar adjustment proceeds until the constraints converge to a ranking with which all produced forms match the perceived forms. In our case with a reduced grammar, the convergence to such a grammar turned out to be possible in one go, at least for this trisyllabic form. In general, it could happen that the adjustment of the grammar leads to a different perception of the same item, which in turn leads to a form in production that differs again from the perceived form and so on. This can also happen when the learner encounters forms with more or less syllables. Usually more than one go is needed for convergence. There is also the possibility that constraints do not converge at all: the forms that the learner encounters are contradictory, leading the constraints to keep tumbling down the hierarchy without converging to a stable grammar (Tesar & Smolensky 2000:67f.). In such a case the data that the learner encounters are either not informative enough, or there simply exists no ranking (i.e. no grammar) that could describe the data.

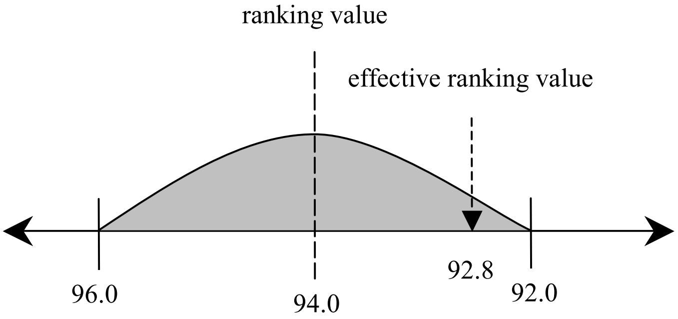
In real life, there is a considerable amount of optionality, both in what children produce and in what they encounter in the data. In this respect CD is not a realistic model of acquisition, since it cannot handle this phenomenon. A development towards solving this problem is the GLA, which is described in the next section.

3.6.2 The Gradual Learning Algorithm

With the GLA, constraints can shift in both directions of the ranking scale, contrary to CD. The process pictured in (18) throughout (26) is the same, i.e. the GLA also makes use of error detection. Once the GLA learner detects an error, she will look up *all* the constraints that prefer the perceived form over the produced form and look up *all* the constraints that prefer the produced form over the perceived form. As with CD, the GLA learner deems the perceived form the correct form. She subsequently moves the constraints down the hierarchy that prefer the produced form and moves the constraints up the hierarchy that prefer the perceived form. The GLA makes use of what became known as Stochastic OT (Boersma 1998). This involves two things: a) the constraints are ranked on a continuous scale rather than on an ordinal scale as in Constraint Demotion, b) the constraints are evaluated with a little bit of noise. On a continuous scale, the constraints cover a range in the form of a Gaussian distribution, rather than occupying strata. This is shown in (34). The mean of the distribution marks the actual *ranking value* (here:

94.0) of the constraint on the scale. However, the distribution has a standard deviation, the *evaluation noise* (say, 1.0), and given that, another point than the ranking value can be chosen in a given evaluation of a form.

(34) A constraint on a continuous ranking scale



For the constraint pictured in (34), this means that any point within the grey area can be chosen, with less and less *probability* the more the point (the *effective ranking value* or *disharmony*) departs from the actual ranking value.[18] Therefore, the most probable effective ranking value that is chosen for the constraint in (34) is around 94.0. A less probable, but nevertheless possible effective ranking value, would be e.g. 92.8.[19]

This property leads to situations where constraints can overlap, if they are close enough to each other. Adding a bit of random noise to each evaluation of a form has the consequence that overlapping constraints might swap places in the hierarchy for the time being, i.e. in a particular evaluation. This is demonstrated in figure (35), where constraint C_1 has a ranking value of 94.0, and constraint C_2 has a ranking value of 90.2. In a particular evaluation, C_1 gets an effective ranking value of 92.0, and C_2 gets an

---

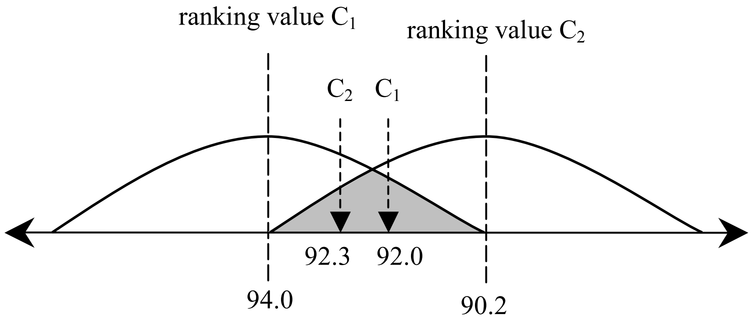
[18] In fact, a Gaussian or *normal* distribution only approaches zero, but never touches it, contrary to what the figure in (34) indicates.

[19] The formula for the mathematically versed reader to compute the exact probabilities, with which these two constraints can swap their ranking in a given evaluation, is provided here (taken from Boersma 1998:331):

$$P(disharmony_1 > disharmony_2) = \tfrac{1}{2} \cdot \left[1 - \mathrm{erf} \left[\tfrac{1}{2}\sqrt{2} \cdot \frac{r_1 - r_2}{rankingSpreading \cdot \sqrt{2}} \right] \right]$$

effective ranking value of 92.2, thereby outranking C_1 in this particular evaluation.

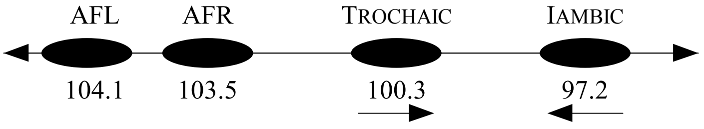
(35) Two constraints on a continuous ranking scale with overlap



The bigger the overlap of the two constraints, the bigger the possibility that the constraints swap places in a particular evaluation.

With gradual learning, a nearly-absolute ranking of constraints is achieved when the constraints have moved far apart so that there is no overlap anymore (indicated in figure (36)). The probability that two constraints swap places in this case becomes very small.

(36) A continuous ranking scale



During reranking, constraints are not immediately demoted below another constraint or promoted above another. When the learner encounters an informative form (i.e. when she detects an error), the constraints move stepwise. In the case of (37), TROCHAIC moves a tiny bit further down the ranking scale, and IAMBIC a tiny bit up the ranking scale, depending on the amount of *plasticity*. The plasticity is the amount that a constraints shifts up or down the ranking scale, i.e. the learning speed. Throughout the simulations conducted for this book, the GLA will shift constraints by 1.0, with a plasticity *decrement* of 0.1, and 4 plasticities in total. This means that in the beginning of learning, the constraints will be shifted by 1.0 on the ranking scale (e.g. from 100 to 99 or 101), later by 0.1 (e.g. from 93.0 to 92.9 or 93.1), even later by 0.01 (e.g. from 104.60 to 104.59 or 104.61), and finally by 0.001 (e.g. from 108.630 to 108.629 or 108.631). The effect of this is that the learning speed decelerates in the course of time and the learning steps become small and smaller.

(37) Grammar adjustment with the GLA

underlying form: |σ σ σ|	AFL	AFR	Trochaic	Iambic
a. /σ (σ́) σ/	*!	*		
👂 b. /(σ σ́) σ/		*	*!→	
c. /σ (σ́ σ)/	*!			*
👄 d. /(σ́ σ) σ/		*		←*
e. /σ (σ σ́)/	*!		*	
f. /(σ́) σ σ/		**!		
g. /σ σ (σ́)/	*!*			

As soon as Trochaic and Iambic overlap, this gradual reranking will make it more likely that in the next production evaluation of a trisyllabic form, the iambic form will become the winner. The learning steps that the GLA learner takes are much smaller than in CD, meaning that she will need a lot of data to converge to a constraint ranking where the overlap between constraints is kept to a minimum and all forms are brought to agreement. The power of the GLA lies in the fact that it can handle noisy data as well as free variation in the data. It can also explain the intermediate stages in the acquisition process (Boersma & Levelt 1999, Curtin & Zuraw 2001): a child starts to produce a modified form or even the correct adult form while still using her old form at some time or other.

CD and the GLA are both on-line learning algorithms, i.e. they modify the learner's grammar directly on the basis of incoming language data. The idea is that the learner considers incoming adult forms as 'correct', and her own forms (i.e. forms that she will produce virtually) as 'incorrect'.

My main concern is the learning of hidden structures alongside with the grammar; in the simulations of Latin stress I will include the modelling of the grammar alone. This is outlined in the following section.

3.7 Learning from full information

The modelling processes outlined in sections 3.3 to 3.6 involved the learning of hidden structure. However, learnability in OT started out from a different position, mentioned in step 1 of chapter 1: the learning from full information, i.e. learning from pairs of surface form and underlying form (Tesar 1995). This entails that the learner is explicitly provided with the correct surface structure, that is with surface forms that already contain adult foot structure. Tableau (38) shows the comprehension process that is in this case trivial: the surface form is given and is the input to the evaluation. Only one candidate is available for comprehension, therefore it does not matter which constraints are violated, because the only candidate is the optimal candidate. The learner is also provided with the corresponding underlying forms. For the surface form /(σ σ́) σ/, this is |σ σ σ|.

(38) Trivial comprehension with given representations

/(σ σ́) σ/ |σ σ σ|	AFL	AFR	TROCHAIC	IAMBIC
☞ /(σ σ́) σ/ |σ σ σ|		*	*	

Production is not trivial: several candidates with all kinds of stress and foot structure are available, and the constraint ranking becomes important. With the ranking in tableau (39) the candidate with a right-aligned trochee, /(σ́ σ) σ/, is chosen in virtual production. Because this form differs from the given surface form in comprehension, this will trigger a reranking of TROCHAIC and IAMBIC in e.g. a CD fashion or a GLA fashion.

(39) Error detection

underlying: |σ σ σ|	AFL	AFR	TROCHAIC	IAMBIC
a. /(σ́) σ σ/		**!		
👄 b. /(σ́ σ) σ/		*		*
c. /σ (σ́) σ/	*!	*		
d. /σ (σ́ σ)/	*!			*
👂 e. /(σ σ́) σ/		*	*!	
f. /σ σ (σ́)/	*!*			
g. /σ (σ σ́)/	*!		*	

When learning from given surface structure and given underlying forms in combination with error detection, CD is called *Error Driven Constraint Demotion* (EDCD; Tesar & Smolensky 2000:50ff.). It is guaranteed that EDCD will converge onto the correct grammar if fed with a sufficient number of fully specified pairs of underlying and surface forms, at least if the target ranking is a language with a total ranking (Tesar & Smolensky 1998, 2000). As a comparison, it is not guaranteed that the GLA will converge onto a correct grammar when provided with pairs of surface and underlying forms (Pater 2005); tested on a set of possible languages it converged in 98.5% of the cases (Paul Boersma, p.c.) onto a correct grammar. This has sometimes been taken as criticism against the GLA as a learning algorithm. One should keep in mind, though, that learning from phonological surface forms is an unnatural learning situation: surface forms are not directly present in the speech signal, and the learner has to construct these surface forms herself from overt forms. The guaranteed convergence of the EDCD when learning from surface and underlying forms is therefore a handy tool to see whether there exists a constraint ranking for the provided data, but it is not a proof for the adequacy of the EDCD as a (natural) learning algorithm. As we have seen in the previous sections, the more realistic task for a learner is to assign surface forms herself. In the case of learning hidden structure, i.e. learning from pairs of *overt* and underlying forms, there is no guarantee that either CD or GLA will converge onto a correct grammar. Tesar & Smolensky (2000) conducted a simulation where the metrical patterns of 124 language types had to be learned by virtual CD learners from pairs of overt and underlying forms. This led to a success rate of only 60%; this means that 40% of those 124 languages could not be learned by CD. The GLA fares a bit better, but is still unable to learn 30% of those languages (as was tested by Boersma 2003). Moreover, chapters 4, 5 and 6 will show that the GLA fares better in certain ways than CD does.

The failure of CD and GLA on some language types is not necessarily bad. A learning algorithm should work for all existing languages, and by failing on some language types it should be able to predict what kinds of languages are impossible to learn. OT learning algorithms, for instance, could predict holes in the factorial typology, i.e. they could predict what permutations of the constraints are impossible to learn; such languages would be allowed by the framework of OT itself, yet would not exist, because there is no path by which children can acquire them.

I test the learnability of pairs of surface and underlying forms only in the simulations of Latin stress, and refer to this kind of learning as “informed learning”. For the rest, I am only concerned with the learnability of hidden structures.

3.8 Summary

We have encountered the phonological ingredients to the modelling of stress in chapter 2, and the modelling processes in learning in this chapter. The modelling of perception is needed to enable the learner the mapping of overt forms onto surface forms. The modelling of recognition is needed to enable the learner to derive underlying forms and meaning from surface forms. With the virtual production step, the learner is enabled to detect an error in her grammar, which will lead her to a systematical adjustment of her grammar. Two possible reranking strategies came up: CD and the GLA. The following chapters mix these ingredients together to simulate stress acquisition in three different languages: Latin (chapter 4), Pintupi (chapter 5), and Modern Greek (chapter 6).

4 The learnability of grammatical stress in Latin[20]

4.1 Introduction

This section compares the performance of the two reranking strategies Constraint Demotion (CD) and the Gradual Learning Algorithm (GLA) for the metrical stress system of Classical Latin. It turns out that the GLA has a higher success rate than CD when learning from overt forms. This suggests that the GLA may be a better model of acquisition than CD. The results also provide evidence for the discussion in the literature about what the correct linguistic analysis of Latin stress is: if overt forms contain main stress only, the GLA makes the child posit an analysis that makes use of uneven trochees (like the analysis by Jacobs 2000) rather than strictly bimoraic trochees (as suggested by Prince & Smolensky, Mester 1994 and Hayes 1995).

To linguists, learnability theory is about creating formal models of language acquisition, i.e. it investigates what precisely is known by the beginning learner and how precisely the learner proceeds from this initial state to an adult state on the basis of language input. While the universality of constraints could be questioned in general, it is assumed in this book that at least the structural constraints that handle metrical phonology are the same in all languages. As we will see, this opens up the possibility that a descriptively simple metrical system, like that of Latin, turns out to be surprisingly complicated when described in terms of constraints proposed by linguists on the basis of cross-linguistic typology rather than in terms of constraints tailored to the specific language at hand.

---

[20] Sections 4.1 to 4.6, and 4.10 appeared in Diana Apoussidou & Paul Boersma (2003): The Learnability of Latin stress, *IFA-Proceedings* 25:100-148. Sections 4.7 and 4.8 appeared in Diana Apoussidou & Paul Boersma (2004): Comparing different Optimality-Theoretic learning algorithms: the case of metrical phonology. *Papers from the AAAI Spring Symposium*, Technical Report SS-04-05, 1-7. Section 4.9 appeared in Diana Apoussidou & Paul Boersma (2004): Comparing different Optimality-Theoretic learning algorithms for Latin stress. *Proceedings of the 23rd West Coast Conference on Formal Linguistics*, 29-42.

To obtain information about the universal components of the grammar, one can analyse the actual language acquisition process of infants and children. Such an analysis is quite difficult in the case of phonological perception, since we cannot look inside a speaker's head to see what happens during perception, and speakers themselves, children included, have very little conscious access to the perception process, let alone the capability of reliably reporting on it. The analysis is slightly less difficult in the case of the child's language *production*, since in that situation at least part of the output of the grammar can be observed directly. But even when considering produced forms, the researcher encounters hidden structures like metrical feet, which often remain ambiguous.

Another method for identifying universal aspects of the grammar is to try and simulate the acquisition process with the help of a computational learning algorithm. In that way, the universal principles derived from language acquisition data can be tested with respect to their adequacy. To make this work, a learning algorithm needs to be supplied with the universal ingredients of the grammar, which in the case of OT means that the learning algorithm should be supplied with a universal set of constraints. Simulating learnability has a further benefit for linguistics, namely providing evidence for or against existing analyses in the literature. By means of a learning algorithm that is based on OT, existing OT analyses of a language can be tested with respect to their learnability. If it turns out that an analysis proposed in the literature is not learnable with a certain learning algorithm, then either this analysis or this learning algorithm should be rejected (Boersma 2003).

In this chapter, the learnability of the metrical phonology of Latin word stress is tested. Taking a dead language as the test subject is not as awkward as it may look. Nowadays there are no native speakers of Latin that can tell us how it was pronounced originally; thus, no phonetic analysis is available. Still, its prosodic system is at least partly accessible through analyses of written text such as poems or language descriptions of contemporary witnesses. The decision fell on the stress system of Latin as the test subject because it has been studied by linguists at great length. Latin is often taken as the prototypical example in general studies on metrical phonology when it comes to illustrating phenomena like weight-sensitivity and extrametricality (e.g. Allen 1973b; Hayes 1985, 1987; McCarthy & Prince 1986; Prince 1990). Many of the principles found in Latin word stress have fed ideas about universal constraint sets for metrical phonology in general, and have been used to analyse other languages. In turn, Latin has been analysed with constraints whose cross-linguistic validity has been established in analyses of other languages. Since there exist several OT analyses of Latin word stress, they can be compared with each other with respect to their learnability. We will also see whether different sets of data make a difference in learning, to be able to determine the amount of information needed for a successful simulation. In addition, the simulations are run with two OT based learning algorithms that differ with respect to their way of constraint reranking during the acquisition process.

The chapter is structured as follows. Section 4.2 provides a description of Latin word stress, and outlines the various analyses proposed for Latin. Section 4.3 to 4.5 outline the computer simulations of the acquisition of Latin primary stress with the two learning algorithms. Section 4.6 presents the results of the primary stress simulations, showing that in several respects the GLA performs better than CD. Sections 4.7 to 4.9 provide further simulations of primary and secondary stress. Sections 4.10 and 4.11 place the findings in a larger perspective and discuss their implications for learnability theory and OT.

4.2 Latin main stress

As already pointed out in chapter 2, an important constituent for assigning stress to words is the foot. By causing a rhythmic organization of syllables, the foot underlies the metrical patterns of many languages. Feet are usually *binary*, i.e., they group syllables into pairs, resulting in a pattern of (often alternating) weak and strong syllables. But how does a learner find out that the feet in her language are strong-weak sequences (trochaic) rather than weak-strong sequences (iambic) in the absence of other phonetic cues like e.g. iambic lengthening? The way that syllables are grouped within a word will have an effect on how other words in the language are stressed. The problem is that foot structure belongs to the surface representation and is not contained in the overt form that a learner is actually exposed to. Therefore the learner has to find out by herself whether the syllables in her language group together as trochaic or as iambic feet.

As indicated above, feet are usually binary. However, not all languages count syllables only; some count moras as well. A mora is a smaller unit than a syllable, and determines the *weight* of the syllable: syllables with a long vowel or a diphthong contain two moras (they are *heavy*), while syllables with only a short vowel contain only one mora (they are *light*). Depending on the language, syllables that end in a consonant can also count as two moras. Languages in which the number of moras in a syllable influences stress (or other phonological phenomena) are called *weight*- or *quantity-sensitive*. In such languages, heavy syllables tend to be prominent in the output. In OT, this principle is captured in the constraint WEIGHT-TO-STRESS-PRINCIPLE (WSP), as mentioned in (14) and repeated in (40):

(40) WSP: Heavy syllables are stressed.

In quantity-sensitive languages a foot ideally consists of two moras: either two light syllables or one heavy syllable, as illustrated in (4) and (5) of chapter 2. In quantity-insensitive languages, feet ideally consist of two syllables, regardless of their inner construction, as illustrated in (6) of chapter 2. This binarity is expressed in OT as a constraint FTBIN, defined in (41).

(41) FTBIN: Feet are binary on some level of analysis (mora or syllable).

Note that this constraint allows a foot to consist of three or four moras, as long as these moras are contained in a sequence of two syllables (heavy-light, light-heavy, or heavy-heavy).

Stress assignment is not only determined by foot-internal structure, but by the placement of feet within the word (or phrase) as well. Especially for longer words, the question is at what edge of the word the foot (or the feet) will be constructed. Some languages tend to build feet at or from the left edge of a word (i.e. word-initially), others at or from the right edge (word-finally). We have already met two constraints for foot placement in chapter 2 (AFR and AFL). Further relevant constraints for Latin are discussed in the course of this chapter.

In Latin, stress is handled purely by the grammar: the foot structure of a word is predictable from the syllable structure of the word, and the mental lexicon need not contain any information about where in the word the stress is realized. Thus, a learner of Latin does not have to take into account the complexities that would arise if the language had lexically assigned stress as well. This should make it relatively easy for a learner to figure out the ranking of the relevant constraints. One would think.

Basically, Classical Latin has left-prominent feet (trochees), it is weight-sensitive, and the last syllable in a word is extrametrical (i.e., it never receives stress except if it is the only syllable of a word) (Allen 1978). Syllables ending in a short vowel are light (abbreviated here as 'L'), while syllables with long vowels or diphthongs and syllables that end in a consonant are heavy ('H'). In words with three or more syllables, the penultimate syllable is stressed if it is heavy. If the penultimate syllable is light, the antepenultimate syllable is stressed, regardless of its weight. In words with only one or two syllables, the leftmost syllable is stressed. Some examples are given in (42). The second column represents overt forms. They include phonetic representations with stress (´) and vowel length (ː), enriched with some hidden phonological structure (periods indicate syllable boundaries) but without foot structure. The third column represents these overt forms without segmental information, i.e. the overt stress patterns ('1' for main stress). The issue of secondary stress is ignored at this point.

(42) Weight and stress in Latin

a.	*amice* 'friend'	[a.míː.ke]	[L H1 L]
b.	*rapiditas* 'speed'	[ra.pí.di.taːs]	[L L1 L H]
c.	*misericordia* 'pity'	[mi.se.ri.kór.di.a]	[L L L H1 L L]
d.	*perfectus* 'perfect'	[per.fék.tus]	[H H1 H]
e.	*incipio* 'I begin'	[iŋ.kí.pi.oː]	[H L1 L H]
f.	*domesticus* 'domestic'	[do.més.ti.kus]	[L H1 L H]
g.	*homo* 'man'	[hó.moː]	[L1 H]

As pointed out above, there is some discussion about the details of Classical Latin stress. The different analyses agree on the minimum size of a trochaic foot (two moras), but not on its maximum size. This discussion especially applies to words with three or more syllables; in words with two syllables such as (40g), *homo*, one is bound to either analyze it with a monomoraic foot with a light syllable plus an extrametrical syllable /(L1) H/ or with a trimoraic foot /(L1 H)/. According to some (e.g. Mester 1994, Prince & Smolensky 1993, Hayes 1995), weight-sensitive feet are strictly bimoraic, while according to others (e.g. Hayes 1981, Jacobs 2000), trochees in Latin can be uneven, i.e. consist of up to three moras. The following section describes the analyses in general, as well as the translation of the problem into OT terms.

4.2.1 Linguistic analyses of Latin stress

The examples in (42) are often analysed as resulting from a combination of extrametricality and right-aligned feet. If we ignore for now the possibility of secondary stress, a recipe by Hayes (1995) will add foot structure to the forms in (42), ending up with the full surface structures /L (H1) L/, /L (L1 L) H/, /L L L (H1) L L/, /H (H1) H/, /H (L1 L) H/, /L (H1) L H/, and /(L1) H/. The recipe goes as follows: first make the last syllable extrametrical, i.e. mark it for not being able to be incorporated into a foot, then create a foot as far to the right as possible; this foot has to be bimoraic, and if the foot is disyllabic stress falls on the first syllable. Hayes' (1995) bimoraic analysis is not uncontroversial. A different approach is to propose that the foot always ends just before the extrametrical syllable (Hayes 1981). The result differs from the bimoraic analysis in two forms in (42): it leads to /L L L (H1 L) L/ (*misericordia*) and /L (H1 L) H/ (*domesticus*). The foot (H1 L) has three moras. Hayes (1995) calls it an *uneven trochee*. Such an analysis satisfies the generalized principle of foot binarity: feet consist either of two moras or of two syllables (as introduced in section 2.3).

The choice between the bimoraic analysis and the uneven trochee analysis cannot be made on the basis of the overt stress patterns alone. Some linguists have voted for a strict bimoraic approach in Latin, on the basis of non-stress evidence like iambic and cretic shortening processes (Mester 1994; Hayes 1987, 1995; Kager 1993; McCarthy & Prince 1986, 1990). According to these authors, an unfooted non-final syllable, like *ti* in /do (més) ti.kus/, is better than an uneven trochee, as in /do (més.ti) kus/ (Hayes 1995:91). Iambic and cretic shortening will not be modelled in the simulated language data fed to the child, because they were optional processes. It will be left to the simulated child to construct either a bimoraic or an uneven trochee analysis (though see §4.10 for a brief discussion of iambic shortening).

Disyllabic words cause several complications. The underlying sequence |L L| is pronounced as the overt form [L1 L]. The question is whether it should be footed as /(L1) L/, violating bimoraicity and foot binarity, or as /(L1 L)/, violating extrametricality. Likewise, |L H| is pronounced [L1 H], and the question is whether it is footed as /(L1) H/, violating foot binarity and bimoraicity, or as /(L1 H)/, violating bimoraicity and extrametricality. These sound like questions about the ranking of constraints, so it is natural to express all these conflicting principles in constraints. The ones often seen in the literature are those in (41) and (43) (a specific constraint for foot bimoraicity will be introduced later).

(43) NONFINAL: The last syllable is not contained in a foot.

There have been several proposals for the foot structure of underlying |L L| words in Latin. Prince (1980) and McCarthy & Prince (1986) argue that the structure must be /(L1 L)/. The argument runs as follows. Latin has a so-called *minimal word* requirement: there are no monosyllabic words in Latin that consist of only a light syllable. This observation can be explained by a combination of two requirements: every word must contain at least one foot, and Latin satisfies a ban on degenerate feet, i.e., the (L1) foot is prohibited completely from Latin surface structure. Apart from ruling out monomoraic words, these requirements also demand that words consisting of two light syllables must incorporate the final syllable into the foot: /(L1 L)/. Expressed in a constraint ranking, this would mean that FTBIN would have to outrank NONFINAL (Prince & Smolensky 1993). Such a ranking also predicts that |L H| is footed as /(L1 H)/, because a form with a degenerate foot /(L1) H/ would violate high-ranking FTBIN. Prince & Smolensky (1993:63) abandon feet with the form (H1 L) because this form is “marked or even absent in trochaic systems” (they refer to Hayes 1987, Prince 1990, and Mester 1994); they formulate this as the constraint *(HL) or RHYTHMICHARMONY.

The foot in Latin thus ideally consists of two moras. The analyses also agree that feet containing four moras, like (H1 H), are forbidden in Latin. Jacobs (2000) accepts the uneven trochee (H1 L), but abandons (L1 H) feet.

4.2.2 Latin Stress in OT

For the simulations of Latin main stress the same underlying forms, candidate generator, and set of constraints are used as Tesar & Smolensky (2000) did in their simulations of 124 types of languages with metrical stress. To accommodate analyses voting for uneven trochees, as in Jacobs (2000), and for moraic trochees, as in Prince & Smolensky (1993), some constraint sets are investigated that are slight modifications of the Tesar & Smolensky set, to see whether and how Latin stress can be learned. In total, six different constraint sets are considered, but with the same underlying forms and generator. In the following these ingredients are discussed in detail.

4.2.2.1 Underlying forms

With Tesar & Smolensky (2000), underlying forms are considered that consist of two to seven syllables. For the forms with two to five syllables, all possible sequences of heavy and light syllables are taken into account. Thus, the underlying disyllables are |L L|, |L H|, |H L|, and |H H|. Likewise, there are eight trisyllabic underlying forms: |L L L|, |L L H|, |L H L|, |L H H|, |H L L|, |H L H|, |H H L|, and |H H H|. In the same vein, there are 16 forms with four syllables, and 32 with five. For the forms with six or seven syllables, the ones with heavy syllables are ignored (for computational reasons that can be deduced from table (68), column 6), thus leaving only |L L L L L L| and |L L L L L L L|. In total, therefore, there are 62 different underlying forms, the same ones that Tesar & Smolensky used. Unlike Tesar & Smolensky, who taught the learners all 62 possible overt forms in their simulations, the present learners are taught only the 28 forms (i.e. 28 underlying-surface pairs or 28 overt forms) that have maximally four syllables; if a learner then arrives at a grammar appropriate for these 28 forms, we can have a look at how she generalizes this grammar to the 34 forms that consist of five syllables or more.

4.2.2.2 The candidate generator

In production, each of the 62 underlying forms comes with a tableau. For each of the 62 tableaux, the candidate set (GEN) is restricted in the same way as in Tesar & Smolensky (2000). GEN creates surface forms that meet the following criteria: the sequence of syllables is identical to the sequence of syllables in the underlying form with respect to number, weight, and order; every foot contains exactly one primary-stressed or secondary-stressed syllable; every word contains exactly one foot that contains a primary-stressed syllable; every primary-stressed or secondary-stressed syllable is contained in a foot; no foot contains more than two syllables. For each of the four disyllabic underlying forms, there are six candidates for the surface form. For instance, an underlying |H L| has the following candidates: /(H1) L/, /(H1 L)/, /(H1)(L2)/, /H (L1)/, /(H L1)/, /(H2)(L1)/. Each of the 8 underlying trisyllables has 24 candidates, for instance |H L L| has the candidates /(H1) L L/, /(H1 L) L/, /(H1) L (L2)/, /H (L1) L/, /H L (L1)/, /H (L1 L)/, /(H L1) L/, /H (L L1)/, /(H1)(L L2)/, /(H1 L)(L2)/, /(H1)(L2) L/, /(H1)(L2 L)/, /H (L1)(L2)/, /(H L1)(L2)/, /(H2)(L1) L/, /(H2)(L1 L)/, /H (L2)(L1)/, /(H L2)(L1)/, /(H2) L (L1)/, /(H2)(L L1)/, /(H2 L)(L1)/, /(H2)(L1)(L2)/, /(H1)(L2)(L2)/, and /(H2)(L2)(L1)/. The 16 forms with four syllables have 88 candidates each; the 32 forms with five syllables have 300 candidates each. The single form with six syllables has 984 candidates, the form with seven syllables has 3136.[21]

4.2.2.3 The constraints

The basic constraint set used in the simulations is the one adopted by Tesar & Smolensky (2000), as listed in (14) of section 2.6.1. This constraint set takes into account the restrictions enforced by the generator: since the generator does not generate candidates with trisyllabic feet, we need no constraints against trisyllabic feet;[22] and since the generator does not generate candidates without main stress, we need no constraints to enforce that every word should contain at least one stress (such as the constraint LEX≈PR by Prince & Smolensky 1993).

Apart from the set of 12 constraints that Tesar & Smolensky used (the *T&S set*), five slightly different constraint sets are investigated. These sets additionally involve the three constraints listed in (44).

(44) Additional/alternative constraints for Latin stress
TROCHAIC: The leftmost syllable in a foot is the head syllable.
HEADNONFINAL: The head foot is not aligned with the right edge of the word, and the head syllable is not the last syllable in the word.
FOOTBIMORAIC: Each foot must be bimoraic.

In order to replicate the idea behind the uneven-trochee analysis of Jacobs (2000), the *uneven trochee constraint set* is defined. This set of 12

---

[21] The monosyllabic underlying form |H| has only a single output candidate: /(H1)/. This makes it impossible for the learner to learn anything from such a form. This is the reason why monosyllabic forms are not considered in any of the tableaux or simulations.

[22] This means that the only reason for including FTBIN is that it militates against monomoraic feet.

constraints (Jacobs himself used only six) is similar to the T&S set, but FTNONFIN is replaced with the constraint TROCHAIC that we met with earlier in (8) of section 2.5. The formulation mirrors that of IAMBIC. The difference with the T&S set is that degenerate feet like (L2) violate FTNONFIN, but not TROCHAIC. In the uneven-trochee constraint set,[23] the constraints TROCHAIC and IAMBIC conspire to minimize the number of syllables in a foot, since monosyllabic feet violate neither.

The third constraint set is based on the moraic-trochee analysis by Prince & Smolensky (1993). Like the uneven-trochee set, this *moraic-trochee constraint set* contains TROCHAIC rather than FTNONFIN. Moreover, the constraint NONFINAL is replaced with HEADNONFINAL[24], which demands that neither the head syllable of a foot nor the head foot itself are in word-final position. Both of these conditions can assign a violation mark: HEADNONFINAL (HDNONFIN) is violated once in the form /(H2)(L1 L)/, twice in /(H2)(L L1)/, and not at all in /(H1)(L L2)/.

Since FTBIN does not distinguish between disyllabic and bimoraic feet, it might be worthwhile to investigate the workings of an explicit bimoraic analysis. To this end, three more constraint sets are considered, all consisting of 13 constraints. These sets were constructed by adding to the T&S, uneven trochee, and moraic trochee sets (all of which contain 12 constraints) a straightforward constraint FOOTBIMORAIC (as similarly proposed in the pre-OT approach of Kager 1993a and in the OT-approach of Hewitt 1994). Other than FTBIN, this constraint is assigned a violation for every monomoraic foot such as (L1) or (L2), every trimoraic foot such as (H1 L) or (L2 H), and every quadrimoraic foot such as (H H1).

4.2.3 Assessment of Jacobs' OT analysis of Latin stress

Jacobs (2000) prefers the constraint NONFINAL to Prince & Smolensky's (1993) HDNONFIN because the formulation of NONFINAL is much simpler and because HDNONFIN seems to predict unattested ('quarternary') stress patterns. Jacobs' analysis can be translated to the set of constraints and candidates discussed before, by first noting that one of the constraints

---

[23] Jacobs did not actually include IAMBIC in his set. This will turn out to be crucial in §4.2.3. He also excluded most alignment constraints.

[24] Prince & Smolensky called this constraint NONFINAL, but the name is changed in order to make the meaning behind every constraint name unambiguous.

employed by Jacobs, LX≈PR (“Every lexical word must correspond to a prosodic word”; Prince & Smolensky 1993) is now part of the candidate generator, so that this constraint can be left out of consideration. For right alignment, Jacobs uses a constraint LAST-FOOT-RIGHT “align the last foot with the word, right edge” (LFR; which comes close to EDGEMOST as in Prince & Smolensky 1993). This constraint is assigned one violation mark for every syllable that follows the last foot; it can thus be seen as a gradient version of WFR, and for words with a single foot it has the same number of violations as AFR and MAIN-R. Jacobs’ article happens to contain a ranking that handles all the forms that he considers.[25] This ranking is TROCHAIC >> NONFINAL >> FTBIN >> LFR >> WSP >> PARSE. It correctly predicts the following forms, several of which contain the uneven trochee (H1 L):

(45) Predicted forms

/(L1) L/	*fala*	/(fá) la/	‘siege tower’
/(L1) H/	*fames*	/(fá) meːs/	‘hunger’
/(H1) L/	*fama*	/(fáː) ma/	‘rumor’
/(H1) H/	*fagus*	/(fáː) gus/	‘beech’
/(L1 L) L/	*fabula*	/(fá.bu) la/	‘little bean’
/(L1 L) H/	*fragilis*	/(frá.gi) lis/	‘fragile’
/L (H1) L/	*amice*	/a (míː) ke/	‘friend’
/L (H1) H/	*facultas*	/fa (kúl) taːs/	‘opportunity’
/(H1 L) L/	*fabula*	/(fáː.bu) la/	‘story’
/(H1 L) H/	*flammulae*	/(flám.mu) lai/	‘little flames’

Tableaux (46) to (49) for the disyllabic forms illustrate the ranking of NONFINAL above the four constraints FTBIN, LFR, WSP, and PARSE. If NONFINAL were ranked below any of these constraints, at least one of these tableaux would have had a different winner. These four tableaux show no

---

[25] For Classical Latin, Jacobs (2003:345) actually proposes the ranking { FTBIN, TROCHAIC } >> NONFINAL >> LFR >> WSP >> PARSE. However, this ranking must be incorrect, because it would give final stress in /L (H1)/ in tableau (48), unless PARSE outranks WSP. This latter ranking confusingly occurs in the tableau on Jacobs’ page 342, so that /(L1 H)/ becomes the winning candidate; but the form /L (H1) L/ requires WSP >> PARSE in order to beat /(L1 H) L/, as Jacobs notes himself and the reader can see here in tableau (51). At the very end of his article, Jacobs introduces the ranking TROCHAIC >> NONFINAL >> FTBIN >> LFR >> WSP >> PARSE for Classical Latin.

evidence for the ranking of TROCHAIC, nor for the relative rankings of FTBIN, LFR, WSP, and PARSE with respect to each other.

(46) Extrametricality beats word-foot-right (LFR) and PARSE

|H L|	TROCHAIC	NONFINAL	FTBIN	LFR	WSP	PARSE
☞ /(H1) L/				*		*
/(H1 L)/		*!				
/H (L1)/		*!	*		*	*
/(H L1)/	*!	*			*	

(47) Extrametricality also beats FTBIN

|L L|	TROCHAIC	NONFINAL	FTBIN	LFR	WSP	PARSE
☞ /(L1) L/			*	*		*
/(L1 L)/		*!				
/L (L1)/		*!	*			*
/(L L1)/	*!	*				

(48) Extrametricality also beats the WSP

|L H|	TROCHAIC	NONFINAL	FTBIN	LFR	WSP	PARSE
☞ /(L1) H/			*	*	*	*
/(L1 H)/		*!			*	
/L (H1)/		*!				*
/(L H1)/	*!	*				

(49) Extrametricality beats LFR, WSP, and PARSE

|H H|	TROCHAIC	NONFINAL	FTBIN	LFR	WSP	PARSE
☞ /(H1) H/				*	*	*
/(H1 H)/		*!			*	
/H (H1)/		*!			*	*
/(H H1)/	*!	*			*	
/(H1)(H2)/		*!				

The tableaux for the trisyllabic forms show more detailed evidence for rankings. Tableau (50) shows evidence for the existence of TROCHAIC, which prefers /(L1 L) L/ to /(L L1) L/. But there is no evidence for the *ranking* of TROCHAIC; it could just as well be ranked at the bottom, as far as the pair |L L L| - /(L1 L) L/ is concerned. This freedom of ranking of TROCHAIC is caused, of course, by the absence of the counteracting constraint IAMBIC from the tableau. If IAMBIC had been included, the choice of /(L1 L) L/ instead of /(L L1) L/ would have been direct evidence for the ranking TROCHAIC >> IAMBIC.[26]

(50) Evidence for trochaicity

|L L L|	TROCHAIC	NONFINAL	FTBIN	LFR	WSP	PARSE
/(L1) L L/			*!	**		**
☞ /(L1 L) L/				*		*
/L (L1) L/			*!	*		**
/L (L1 L)/		*!				*
/(L L1) L/	*!			*		*

Tableaux for trisyllabic forms that end in a heavy syllable are not shown here, because the high ranking of NONFINAL ensures that L-final and H-final words are always handled in the same way. The next form to consider, then, is |L H L|. Tableau (51) shows direct evidence that PARSE is dominated by WSP as well as by TROCHAIC. If the ranking of WSP and PARSE had been reversed (with high-ranked TROCHAIC), the candidate /(L1 H) L/ would have won. It is apparently worse to have an unstressed heavy syllable than to have an unfooted light syllable. If TROCHAIC had been ranked below WSP and PARSE, the iambic candidate /(L H1) L/ would have won.

---

[26] Note that without the constraint IAMBIC, the last candidate /(L L1) L/ is harmonically bounded: it could not win under any ranking of the constraints.

(51) Weight-to-stress and trochaicity beat Parse

|L H L|	Trochaic	Nonfinal	FtBin	LFR	WSP	Parse
/(L1) H L/			*!	**	*	**
/(L1 H) L/				*	*!	*
☞ /L (H1) L/				*		**
/L (H1 L)/		*!				*
/(L H1) L/	*!			*		*

The next underlying form to consider is |H L L|. Tableau (52) shows that the winner yields evidence for the existence of the constraints LFR or Parse. Without these constraints, the candidate /(H1) L L/ would have been equally harmonic as /(H1 L) L/. Jacobs' constraint set thus favours the uneven trochee analysis /(H1 L) L/ over the bimoraic analysis /(H1) L L/, irrespectively of the ranking of the constraints, since the violations of /(H1) L L/ form a superset of those of the violations of /(H1 L) L/. In order to turn the bimoraic analysis /(H1) L L/ into a winner, we would need the help of an extra constraint that is ranked above LFR and Parse, perhaps FtBimoraic.

(52) Evidence for LFR or Parse

|H L L|	Trochaic	Nonfinal	FtBin	LFR	WSP	Parse
/(H1) L L/				**!		**
☞ /(H1 L) L/				*		*
/H (L1) L/			*!	*	*	**
/H (L1 L)/		*!			*	*
/(H L1) L/	*!			*	*	*

Until now, all the forms that Jacobs considers have been discussed. Absent from his article, though, is the underlying form |H H L|. Tableau (53) shows that this form is problematic.

(53) A stress clash or a superheavy foot?

|H H L|	TROCHAIC	NONFINAL	FTBIN	LFR	WSP	PARSE
/(H1) H L/				**!	*	**
/(H1 H) L/				*	*!	*
/H (H1) L/				*	*!	**
/H (H1 L)/		*!			*	*
/(H H1) L/	*!			*	*	*
☞ /(H1)(H2) L/				*		*
☞ /(H2)(H1) L/				*		*

In tableau (53), two forms with two feet are optimal. To make the form with penultimate main stress win, the constraints for the placement of main stress would have to be included and ranked in the order MAIN-R >> MAIN-L.[27] The results for the |H H L| forms generalize to quadrisyllabic and longer forms. Because of the ranking WSP >> PARSE, these forms will tend to have secondary-stressed feet around every heavy syllable, and because of the presence of PARSE, light syllables will tend to be footed as well if this does not create iambs. Examples are:

(54) Examples for |H H L| forms

/(L2 L) (H1) L/	*manifesta*	[mà.ni.fés.ta]	'caught in the act'
/(H2 L) (H1) H/	*militaris*	[mìː.li.táː.ris]	'military'
/L (H2) (L1 L) L/	*amicitia*	[a.mìː.kí.ti.a]	'friendship'
/(H2)(H2)(H2)(H1) H/	*definitivus*	[dèː.fiː.nìː.tíː.vus]	'definitive'
/(H2 L) L (H1) L/	*deliciosa*	[dèː.li.ki.óː.sa]	'spoiled'

The choice between /(L2 L) L (H1 L) L/ ([mì.se.ri.kór.di.a]) and /L (L2 L)(H1 L) L/ ([mi.sè.ri.kór.di.a]) would probably have to be made by constraints such as AFL and AFR.

But it is a question whether secondary-stressed forms should be allowed at all, especially those with *stress clashes* (consecutive stressed syllables) like those in tableau (53). The correct form in tableau (53) should therefore be /H (H1) L/, with a single foot. It is possible to get rid of the two

---

[27] As long as MAIN-R is ranked above MAIN-L, it does not matter where in the hierarchy the two constraints are inserted, because they are only violated in the last two candidates, and only decide between those two.

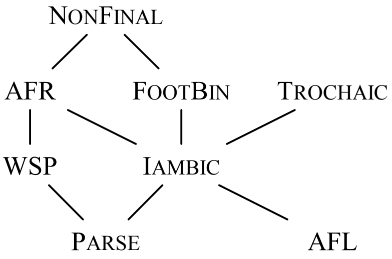
bipedal candidates in (53) by replacing LFR with AFR. This would not change anything in tableaux (46) to (52), but the two last candidates in (53) would get three violations of AFR. If it can be ensured that AFR outranks WSP, the last two candidates in (53) perish. However, the winner will now be the form /(H1 H) L/, with a superheavy foot. This form is observationally incorrect, with its antepenultimate main stress (a speaker of Latin would say [au.díː.re] 'to hear', not [aú.diː.re]); the correct form is /H (H1) L/. However, we can see from (53) that /H (H1) L/ has superset violations when compared with /(H1 H) L/. In order to make /H (H1) L/ more harmonic than /(H1 H) L/, then, an extra constraint would have to be used. An obvious choice is Iambic, and it should be ranked above Parse, as tableau (55) shows.

(55) Jacobs' hierarchy patched up

|H H L|	Trochaic	Nonfinal	FtBin	AFR	WSP	Iambic	Parse
/(H1) H L/				**!	*		**
/(H1 H) L/				*	*	*!	*
☞ /H (H1) L/				*	*		**
/H (H1 L)/		*!			*	*	*
/(H H1) L/	*!			*	*		*
/(H1)(H2) L/				**!*			*
/(H2)(H1) L/				**!*			*

Tableau (55) provides a ranking that will work for all Latin forms. It correctly generalizes to words of more than three syllables, causing all of them to end in /...(H1) X/ or /...(X1 L) X/. The remaining question is where Iambic has to be inserted into the hierarchy. According to (55), it has to outrank Parse. Obviously, it has to be ranked below Trochaic, otherwise /(L L1) L/ would be better than /(L1 L) L/; that would be observationally incorrect, since *iacere* 'to throw' is pronounced [já.ke.re], not [ja.ké.re]. Given the current set of seven constraints, and the low ranking of Parse, Iambic has to be ranked below FtBin, because /(L1 L) L/ has to be better than /L (L1) L/. Finally, AFR has to outrank both WSP and Iambic in order to make /H (L1 L) X/ (e.g. *nobilitas* [noː.bí.li.taːs] 'fame') better than /(H1) L L X/. The complete set of crucial rankings is shown in figure (56). The rankings not marked by lines in this figure are not fixed. Thus, Trochaic could be ranked anywhere between the very top and a position below WSP, as long as it outranks IAMBIC; FTBIN could be ranked above AFR or below WSP, as long as it is ranked below NONFINAL and above IAMBIC; and so on.

(56) The crucial ranking for the uneven trochee analysis without secondary stress



Note the *conspiracy* of the constraints TROCHAIC and IAMBIC. Together they have a preference for monosyllabic feet, since such feet violate neither of these constraints.[28] In the ranking at hand, this monosyllabic bias is just enough to rule out /(H1 H) L/, because TROCHAIC and IAMBIC are both ranked above PARSE. The bias is not enough to rule out /(L1 L) L/ and /H (L1 L) L/, because IAMBIC is still ranked below FTBIN and AFR. This combination of requirements on the ranking of IAMBIC brings about a relatively *deep* grammar: the tree in (56) shows that four levels of constraints are needed to describe the Latin stress rule with its relatively simple formulation of "stress the penultimate if it's heavy, else the antepenultimate".

4.2.4 Assessment of the Tesar & Smolensky constraint set for Latin stress

The effects of the bias of TROCHAIC and IAMBIC for short feet is not found for Tesar & Smolensky's combination of FTNONFIN and IAMBIC. This is because these two constraints have complementary violations on the foot level: monosyllabic and iambic feet, i.e. (X1) and (X X1) (where 'X' stands for either 'L' or 'H'), violate FTNONFIN, while disyllabic trochaic feet, i.e.

---

[28] This is problematic, since factorial typology predicts languages with monosyllabic feet only (René Kager, p.c.).

(X1 X), violate IAMBIC. Thus, the sum of the number of violations of FTNONFIN and IAMBIC is equal to the number of feet in the word. This means that these two constraints together still have the side effect of minimizing the number of feet in a word, but they are not capable of forcing a specific foot form in the way TROCHAIC and IAMBIC could. The reduced power of FTNONFIN as compared to TROCHAIC turns out to make it impossible for Tesar & Smolensky's set of 12 constraints to handle the facts of Classical Latin stress in the uneven trochee analysis without secondary stress. This will become evident in sections §4.6.1 and 4.6.2.

4.2.5 Assessment of a moraic-trochee analysis of Latin stress

Let us now turn to an analysis of Latin stress with moraic trochees. On grounds of optional processes such as iambic and cretic shortening, Mester (1994) argued that Latin employed moraic trochees, and that words with a final H-L-H sequence must be parsed as /(H1) L H/, and not as /(H1 L) H)/, because in the optional shortening process, the final heavy syllable becomes light, and is parsed as /(H1) (L L)/. If the unshortened form had been parsed as /(H1 L) H)/, the shortened form with two feet /(H1) (L L)/ would violate the *Free Element Condition* (Prince 1985, Steriade 1988, Halle & Kenstowicz 1991), which states that newly built prosodic structure cannot overwrite previously established structure. Prince & Smolensky (1993) gave an OT account for an analysis with moraic trochees in Latin. Next to FTBIN, PARSE, TROCHAIC and WSP, they used an extrametricality constraint NONFINAL and an alignment constraint EDGEMOST. Their NONFINAL constraint will be called HEADNONFINAL (HDNONFINAL) here, in order to distinguish it from the NONFINAL constraint in (14). While NONFINAL requires that the final syllable is not included in a foot, HDNONFINAL requires that no head (be it head syllable or head foot) is final. The difference is that in e.g. a candidate like /(H1) (L2)/, NONFINAL is violated, but HDNONFINAL is not. EDGEMOST is a predecessor of the alignment constraints in (14) and requires that the head syllable of a prosodic word needs to be aligned with the right edge. RHYTHMICHARMONY (or *(HL) in short) rules out candidates with (HL)-feet.

(57) Prince and Smolensky’s (1993) additional constraints:
HdNonfinal: No head of PrWd is final in PrWd.
Edgemost (σ’, R): The head syllable is situated at the right edge of PrWd.
RhHarmony/*(HL): Final elements in constituents are long.

Prince & Smolensky (1993) arrived at the following ranking:

(58) A ranking for a moraic-trochee analysis

FtBin, Trochaic, *(HL)
|
HdNonFin
|
Edgemost, WSP
|
Parse

With this ranking, feet are bimoraic, but only where possible: a form like [á.moː], with a light and a heavy syllable, is parsed as a disyllabic foot /(L1 H)/, because the competing candidate /(L1) H/ is ruled out by high-ranked FtBin:[29]

(59) Disyllabic feet in disyllabic words

|L H|	FtBin	Trochaic	HdNonfinal	WSP	Edgem	Parse
/(L1) H/	*!			*	*	*
☞ /(L1 H)/			*	*	*	

According to Prince & Smolensky (1993), [H L] words in Latin are parsed into /(H1) L/ feet; (H1 L) feet are ruled out by *(HL):

---

[29] The interested reader is kindly referred to Prince & Smolensky (1993) for a more detailed analysis.

(60) Stress on the penultimate syllable

|H L|	FtBin	Trochaic	HdNonfinal	WSP	Edgemost	Parse
☞ /(H1) L/					*	*
/(H1 L)/			*!		*	

For [H1 L L] sequences, Prince & Smolensky (1993) suggest that they are parsed as /(H1) (L L)/, and [H H1 L] sequences as /(H) (H1) L/. They do not mention whether the foot preceding or following the stressed syllable carries a secondary stress or not; their representation implies that these feet are stressless.

(61) Stress on the antepenultimate syllable

|H L L|	FtBin	Trochaic	HdNonfinal	WSP	Edgemost	Parse
/(H1) L L/					**	*!*
/(H1 L) L/					**	*!
/H (L1 L)/			*!	*	*	*
☞ /(H1)(L L)/					**	

To enable a comparison between a moraic-trochee analysis of Latin stress, as proposed by Mester (1994) and Prince & Smolensky (1993), with an uneven-trochee analysis, as e.g. proposed by Jacobs (2000), I will give an account for the moraic trochee analysis with some of the constraints listed in (14). In line with Prince & Smolensky (1993), the approach will include the constraints FtBin, Parse, Trochaic and WSP. Departing from the constraint set of Prince & Smolensky (1993), the constraint Edgemost is replaced with the constraint AFR, as established in (14). The performance of Prince & Smolensky's extrametricality constraint HdNonfinal will be compared with the constraint Nonfinal as defined in (14). The straightforward constraint *(HL) is discarded. The resulting constraint set is given in (62).

(62) A new constraint set for the moraic-trochee analysis
AFR
FtBin
HdNonfinal
Parse
Trochaic
WSP

I further deviate from the account given in Prince & Smolensky by only considering feet with stress: where there is a stress, there is a foot, and where there is a foot, there is a stress. For this reason, stressless feet that occur to the right of the head foot are not included in the candidate set.

The moraic-trochee analysis is given in the following, adopting a ranking as closest possible to the one given in (58). For a form like [á.moː], this leads to the incorporation of both syllables into a foot, as shown in tableau (63). A candidate /L (H1)/ crucially violates HdNonfinal twice since both the head foot and the head syllable are word-final.

(63) Exhaustive parsing into a 'wretched' trochee

|L H|	FtBin	Trochaic	HdNonfinal	WSP	AFR	Parse
/(L1) H/	*!			*	*	*
☞ /(L1 H)/			*	*		
/L (H1)/			*!*			*
/(L H1)/		*!	**			

Since HdNonfinal is fine with secondary stress in final syllables, such analyses must pop up, as shown in tableau (64), unless prevented by a constraints against stress clashes.

(64) Stress clash in a bisyllabic word

|H H|	FtBin	Trochaic	HdNonfin	WSP	AFR	Parse
/(H1) H/				*(!)	*(!)	*
/(H1 H)/			*!	*		
/H (H1)/			*!*	*		*
/(H H1)/		*!	**	*		
☞ /(H1)(H2)/						

Analogously, this analysis predicts forms like /(L1 L) (H2)/ for light-light-heavy sequences.

Sequences ending in |…H L L| would result in trimoraic feet /(H1 L) L/, though. This is shown in tableau (65). The candidate with the bimoraic foot, /(H1) L L/, is ruled out by an additional violation of AFR.

(65) An uneven trochee, again

|H L L|	FtBin	Trochaic	HdNonfin	WSP	AFR	Parse
/(H1) L L/					**!	**
☞ /(H1 L) L/					*	*
/H (L1) L/	*!			*	*	**
/H (L1 L)/			*!	*		*
/(H L1) L/		*!		*	*	*
/(H2)(L1 L)/			*!		**	
/(H1)(L2 L)/			*!		**	

But like Jacobs' analysis, the adopted constraint set and ranking predicts a wrong result for underlying |H H L|, shown in tableau (66): if WSP and AFR are crucially tied so that their violation marks add up and the buck is passed to Parse, or if WSP outranks AFR, /(H1)(H2 L)/ wins.

(66) Main stress on the antepenultimate despite a heavy penult

|H H L|	FtBin	Trochaic	HdNonfin	WSP	AFR	Parse
/(H1) H L/				*(!)	**(!)	**
/(H1 H) L/				*	*	*!
/H (H1) L/				*	*	*!*
/H (H1 L)/			*!	*		*
/(H H1) L/		*!		*	*	*
/(H1)(H2) L/					***!	*
/(H2)(H1) L/					***!	*
☞ /(H1)(H2 L)/					**	
/(H2)(H1 L)/			*!		**	

If, on the other hand, AFR outranks WSP, candidate /(H1 H) L/ wins. As with Jacobs' analysis, the analysis can be saved by assuming that AFR outranks WSP and by inserting Iambic into the hierarchy, as (67) shows.

(67) Moraic-trochee analysis patched up

|H H L|	FtBin	Trochaic	HdNonfin	AFR	WSP	Iambic	Parse
/(H1) H L/				**!	*		**
/(H1 H) L/				*	*	*!	*
☞ /H (H1) L/				*	*		**
/H (H1 L)/			*!		*	*	*
/(H H1) L/		*!		*	*		*
/(H1)(H2) L/				**!*			*
/(H2)(H1) L/				**!*			*
/(H1)(H2 L)/				**!		*	
/(H2)(H1 L)/			*!	**		*	

The Latin stress analyses (termed the *uneven-trochee analysis*, the *moraic-trochee analysis*, and the *T&S analysis*) introduced in this section are tested by computer simulations on three training sets of underlying-surface pairs and two training sets of overt forms, for 10 virtual CD learners and 10 virtual GLA learners. This adds up to a total of 6x5x20 = 600 simulated acquisition processes. Sections 4.3-4.5 describe the constraint sets, the training sets, and the acquisition processes.

4.3 The constraint sets

For the simulations, different constraints sets were set up to accommodate the uneven-trochee analysis and the moraic-trochee analysis, with the slightest possible changes. The basis for all constraint sets was the list of constraints as used in Tesar & Smolensky's (2000) learnability simulations, listed in (14). The uneven-trochee set differed from this set in the constraint on trochaicity: FTNONFIN was replaced with TROCHAIC. The moraic-trochee set differed from the original constraint set in that FTNONFIN was replaced by TROCHAIC and NONFINAL replaced by HDNONFIN. These three constraint sets were further modified by adding a constraint FTBIMORAIC. Table (68) summarizes the six constraint sets of section 4.2.2.3. In order that the reader can perform a simple though perhaps tedious check on the correctness of the evaluator, in the last column a count of the total number of constraint violations is included in the 15344 candidates in the 62 tableaux.

(68) Statistics on the six constraint sets

Constraint set	No. of Const.	Trochaicity constraint	Extrametrical constraint	Bimoraicity constraint	Violations
T&S	12	FTNONFIN	NONFINAL	(none)	370404
uneven trochee	12	TROCHAIC	NONFINAL	(none)	340028
moraic trochee	12	TROCHAIC	HDNONFIN	(none)	335932
T&S + FTBIMOR	13	FTNONFIN	NONFINAL	FTBIMOR	398062
uneven trochee + FTBIMOR	13	TROCHAIC	NONFINAL	FTBIMOR	367686
moraic trochee + FTBIMOR	13	TROCHAIC	HDNONFIN	FTBIMOR	363590

4.4 The training data

As mentioned before, every training set contains 28 different forms: no words with five or more syllables are fed to the listener during acquisition.

Each of the first three training sets consists of 28 pairs of given underlying forms together with the fully specified surface forms. The complete list is in table (69); where two or more analyses predict the same form, some ink has been saved. The 'uneven trochee' set is meant to replicate the uneven trochee analysis. The 'at most bimoraic' and 'at least bimoraic' sets are meant to give an analysis that is even more bimoraic than the moraic-trochee analysis in 4.2.5 (which still includes (HL) feet if no special constraints are added). At this point no explicit analyses with secondary stress are included.

The three analyses in table (69) all share the same overt forms, which can be seen in table (70). These are used in the simulations with overt forms.

An important question when dealing with stress systems is whether the language employs secondary stress. For Latin, this question is not trivial (see §4.6.3 and §4.9). *If* Latin had secondary stress, there are many different possibilities to place it. It could be quantity-sensitive or quantity-insensitive, stress clash could be permitted or not, and so on. Thereby many different sets of overt forms with secondary stress are thinkable. At this point, one secondary-stressed overt data set is included, which is shown in table (70). This set has weight-sensitive secondary stress before the main stress: every H is footed,[30] as is every remaining LL; the ambiguity that this will lead to in cases like |L L L H L| will have to be solved by the learner on the basis of her own generalization from the shorter forms in the training set to forms longer than four syllables.

---

[30] Unlike Jacobs (2003), who assumes *CLASH to prevent two heavy syllables next to each other from being both stressed, every heavy syllable to the left of the main stressed syllable is stressed here.

(69) Three training sets with fully structured surface forms

underlying forms	surface forms		
	uneven trochee	at most bimoraic	at least bimoraic
|L L|	/(L1) L/	/(L1) L/	/(L1 L)/
|L H|	/(L1) H/	/(L1) H/	/(L1 H)/
|H L|	/(H1) L/		
|H H|	/(H1) H/		
|L L L|	/(L1 L) L/		
|L L H|	/(L1 L) H/		
|L H L|	/L (H1) L/		
|L H H|	/L (H1) H/		
|H L L|	/(H1 L) L/	/(H1) L L/	
|H L H|	/(H1 L) H/	/(H1) L H/	
|H H L|	/H (H1) L/		
|H H H|	/H (H1) H/		
|L L L L|	/L (L1 L) L/		
|L L L H|	/L (L1 L) H/		
|L L H L|	/L L (H1) L/		
|L L H H|	/L L (H1) H/		
|L H L L|	/L (H1 L) L/	/L (H1) L L/	
|L H L H|	/L (H1 L) H/	/L (H1) L H/	
|L H H L|	/L H (H1) L/		
|L H H H|	/L H (H1) H/		
|H L L L|	/H (L1 L) L/		
|H L L H|	/H (L1 L) H/		
|H L H L|	/H L (H1) L/		
|H L H H|	/H L (H1) L/		
|H H L L|	/H (H1 L) L/	/H (H1) L L/	
|H H L H|	/H (H1 L) H/	/H (H1) L H/	
|H H H L|	/H H (H1) L/		
|H H H H|	/H H (H1) H/		

(70) Two training sets with overt forms

<table>
<tr><th colspan="2">overt forms</th></tr>
<tr><th>main stress only</th><th>secondary stress</th></tr>
<tr><td colspan="2">[L1 L]</td></tr>
<tr><td colspan="2">[L1 H]</td></tr>
<tr><td colspan="2">[H1 L]</td></tr>
<tr><td colspan="2">[H1 H]</td></tr>
<tr><td colspan="2">[L1 L L]</td></tr>
<tr><td colspan="2">[L1 L H]</td></tr>
<tr><td colspan="2">[L H1 L]</td></tr>
<tr><td colspan="2">[L H1 H]</td></tr>
<tr><td colspan="2">[H1 L L]</td></tr>
<tr><td colspan="2">[H1 L H]</td></tr>
<tr><td>[H H1 L]</td><td>[H2 H1 L]</td></tr>
<tr><td>[H H1 H]</td><td>[H2 H1 H]</td></tr>
<tr><td colspan="2">[L L1 L L]</td></tr>
<tr><td colspan="2">[L L1 L H]</td></tr>
<tr><td>[L L H1 L]</td><td>[L2 L H1 L]</td></tr>
<tr><td>[L L H1 H]</td><td>[L2 L H1 H]</td></tr>
<tr><td colspan="2">[L H1 L L]</td></tr>
<tr><td colspan="2">[L H1 L H]</td></tr>
<tr><td>[L H H1 L]</td><td>[L H2 H1 L]</td></tr>
<tr><td>[L H H1 H]</td><td>[L H2 H1 H]</td></tr>
<tr><td>[H L1 L L]</td><td>[H2 L1 L L]</td></tr>
<tr><td>[H L1 L H]</td><td>[H2 L1 L H]</td></tr>
<tr><td>[H L H1 L]</td><td>[H2 L H1 L]</td></tr>
<tr><td>[H L H1 H]</td><td>[H2 L H1 H]</td></tr>
<tr><td>[H H1 L L]</td><td>[H2 H1 L L]</td></tr>
<tr><td>[H H1 L H]</td><td>[H2 H1 L H]</td></tr>
<tr><td>[H H H1 L]</td><td>[H2 H2 H1 L]</td></tr>
<tr><td>[H H H1 H]</td><td>[H2 H2 H1 H]</td></tr>
</table>

Five training sets have been established by now, although one could think of several more, both for the underlying-surface pairs and for the overt forms. All thinkable training sets, however, must be identical with respect to where the main stress falls: on the penultimate syllable if this is heavy, and on the antepenultimate otherwise.

4.5 The acquisition processes

The virtual learners of Latin stress learned from tableaux as presented in chapter 3. When learning from overt data as in table (70), learning took place as described in sections 3.2 to 3.6. Applied to Latin, perception looks as in tableau (71).[31] Imagine that at some point in learning, the learner encounters a [L H1 L] sequence; an overt input that consists of words made up of light and heavy syllables, of which one is marked for stress. In tableau (71), the grammar chooses for the left aligned iambic candidate /(L H1) L/ as the perceived surface structure (overt forms are once more excluded).

(71) Perception in Latin

overt: [L H1 L]	AFL	AFR	TROCHAIC	IAMBIC
a. /L (H1) L/	*!			
👂 b. /(L H1) L/		*	*	
c. /L (H1 L)/	*!			*

In virtual production, shown in (72), a left-aligned trochee is chosen by the constraint ranking. The winning candidate in production is indicated by ‘👄’, and the winning candidate in perception is indicated by ‘ 👂 ’.

(72) Production and error-detection in Latin

underlying: |L H L|	AFL	AFR	TROCHAIC	IAMBIC
a. /(L1) H L/		**!		
👄 b. /(L1 H) L/		*		*
c. /L (H1) L/	*!	*		
d. /L (H1 L)/	*!			*
👂 e. /(L H1) L/		*	*!	
f. /L H (L1)/	*!*			
g. /L (H L1)/	*!		*	

Thereby an error was detected and the constraints will be reranked depending on the reranking strategy.

---

[31] This tableau does not show the starting point of learning; it rather applies the perception process demonstrated in section 3.2 to Latin.

When learning from surface structures as given in table (69), learning proceeded as in section 3.7. In this case, only the computation of surface forms in production had to be learned.

For the two reranking strategies the implementations in the Praat programme (Boersma & Weenink 1992-2006) are used. The evaluation model for CD was OT with crucial ties, i.e. the violations of constraints that are ranked equally high are added to each other as if these constraints formed a single constraint; in Praat, this can be simulated by setting the evaluation noise to zero. As in Tesar & Smolensky (2000), the algorithm was allowed to chew five times on every piece of language data, with backtracking if the quintuple chews did not succeed in making the (alleged) correct adult form optimal in the learner's grammar. A slight difference with Tesar & Smolensky's evaluation model was that when two forms were equally harmonic, a winner was chosen randomly from among them, whereas Tesar & Smolensky somewhat less realistically chose the form that occurred first in the tableau (p.c. between Bruce Tesar and Paul Boersma).

The evaluation model for the GLA was Stochastic OT with an evaluation noise of 2.0. This noise leads to slightly different rankings of the constraints at each evaluation. Within an evaluation of an overt form, however, the ranking stayed constant: the same ranking values drawn from the Gaussian distributions were used first to interpret the overt form into a surface form and an underlying form, then to produce the learner's surface form from the interpreted underlying form.

For each of the 600 virtual learners, the 12 constraints (or 13 constraints, respectively) were initially ranked at a height of 100. After this, language data were drawn randomly with equal probability from the 28 underlying-surface pairs or from the 28 overt forms. All learners therefore heard the forms in different orders and with (very slightly) different frequencies. When a pair or a form caused a mismatch between the learner's own produced surface form and (her guess of) the correct adult form, the CD learner had an adjustment model that would demote the ranking of one constraint by a distance of 1 (e.g. to 99 when a constraint is demoted for the first time), and the GLA learner had an adjustment model that would raise the rankings of some constraints by 0.1 and lower the rankings of some others by 0.1; in the case of the GLA learner, this *plasticity* of 0.1 was further randomized by a relative plasticity standard deviation of 0.1.

A CD learner was allowed to listen to maximally 1 000 pairs of underlying-surface forms or 1 000 overt forms. After every 100 pairs or forms, however, it was checked whether the learner had already arrived at a grammar in which all 28 pairs or forms were *singly grammatical*. An underlying-surface pair is singly grammatical if the surface form is the only optimal candidate for the underlying form, i.e. if it is optimal in its tableau and no other candidate in the same tableau is equally harmonic. An overt form is singly grammatical if for all the tableaux in which it occurs (in the current case this is always a single tableau), this overt form is shared by all optimal candidates. For instance, the overt form [H H1 L L] can be considered singly grammatical if the optimal candidate in the tableau for |H H L L| is /(H H1) L L/, /H (H1) L L/, or /H (H1 L) L/. If all 28 pairs or forms are singly grammatical, it is certain that the learner will not be capable of any more learning with these forms because encountering new forms will not lead to error detection anymore. When this occurred, learning was considered successful and the simulation was stopped (i.e., no more forms were fed to the learner). A CD learner usually either successfully acquired the language within the first or second round of 100 pairs or forms, or she did not acquire the language even after 1 000 pairs or forms; in the latter case it is certain that the learner will never succeed, as is exemplified in the discussions on tables (78) and (85).

GLA learners (who take much smaller steps than CD learners) were allowed to listen to maximally 40 000 underlying-surface pairs or 40 000 overt forms. After every 1 000 pairs or forms, it was checked whether the learner had arrived at a grammar in which all of the pairs or forms were singly grammatical. If so, the simulation was stopped. When deciding whether a pair or form was singly grammatical, the evaluation noise was set to zero and the optimal candidates in the 28 relevant tableaux were computed, then proceeding as above. Although the learner would still be likely in this situation to make several mistakes if the evaluation noise had the usual value of 2.0, it was decided that learning had succeeded, because it was certain that the learner's constraints were already ranked in the correct order and that future learning would reduce the error rate but not change the crucial rankings.

4.6 Results

Table (73) shows the results for the 600 learners. In each cell, the result is indicated as x/y, where *x* is the number of CD learners that succeeded and *y* is the number of GLA learners that succeeded. When none of the 10 learners succeeded, this is indicated by "–"; when all 10 learners succeeded, this is indicated by "√".

(73) Simulation results for 600 learners, in the form "CD/GLA"

	Learning from pairs of underlying and surface forms			Learning from overt forms	
Constraint set	uneven trochees	at most bimoraic	at least bimoraic	main stress only	secondary stress
T&S	–/–	–/–	–/–	–/–	–/√
uneven trochee	√/√	–/–	–/–	–/√	–/√
moraic trochee	√/√	–/–	–/–	1/–	–/–
T&S + FtBim	√/√	√/√	–/–	–/–	–/√
uneven trochee + FtBim	√/√	√/√	–/–	–/7	–/√
moraic trochee + FtBim	√/√	√/√	√/√	9/–	–/–

4.6.1 Informed learning of primary stress in Latin

Table (73) shows that CD and GLA were equally successful in learning from pairs of underlying and surface forms: every cell in the first three columns either contains "√/√" (all 10 CD learners and all 10 GLA learners succeeded) or "–/–" (all 20 learners failed). This is not surprising. CD is a generally applicable OT learning algorithm that when supplied with fully specified underlying-surface pairs, is guaranteed to find a ranking that can generate those forms, if there *is* such a ranking. Thus, from the first "–" in every cell with "–/–" it appears that there is no ranking at all that can generate the 28 underlying-surface pairs at hand with the constraint set at hand. This necessarily means that the GLA will not be able to find an appropriate ranking either (as confirmed by the second "–" in all these cells). From the first "√" in every cell with "√/√" it appears that there *is* a ranking, and the second "√" in these cells tells us that the GLA has also been able to find it. Since there are no cells with "√/–" in the first three columns, we can conclude for the case of Latin stress that in all the cases in which CD works, the GLA works as well.

As expected from the ranking we found in §4.2.3, the uneven trochee analysis was learnable with the uneven trochee constraint set. Two of the CD learners arrived at the ranking in (74).

(74) Idealized results for uneven-trochee CD learning of the uneven-trochee analysis

Constraints	Ranking values
NONFINAL, TROCHAIC	100
AFR, FTBIN, MAIN-R, WFR	99
IAMBIC, WSP	98
AFL, MAIN-L, PARSE, WFL	97

This ranking is exactly what can be predicted from the crucial rankings in figure (56). CD is an algorithm that is claimed to rank every constraint maximally high. When comparing the ranking of the learner in (74)with the ranking in (56), we can see that NONFINAL and TROCHAIC are undominated, so their ranking stays at the original 100 in the simulations. The constraint AFR and FTBIN are outranked only by undominated constraints, so they end up at a height of 99. Each of the constraints WSP and IAMBIC is dominated by a constraint from the second level, so they end up at 98. The deepest constraint in (56) is PARSE; it must end up at 97. The remaining five constraints end up as high as they can without altering any of the optimal candidates: MAIN-R can end up at 99 because it is assigned the same number of violations in all winning candidates as AFR. WFR has to go below NONFINAL, with which it has complementary violations.

But the constraints are not always ranked maximally high. One learner ends up with a ranking similar to (74), but with WFR ranked at 98; and another learner ends up with both WFR and MAIN-R ranked at 98. While this makes no difference in the output forms, the maximally-high-ranking claim of CD is violated here, probably because of the existence of solutions with crucial ties, for which we will now see some more dramatic examples.

Four CD learners ended up with what is probably the minimum number of strata: they collapsed the AFR - FTBIN - MAIN-R - WFR stratum with the IAMBIC - WSP stratum (at 99), and had the four bottom constraints (AFL - MAIN-L - PARSE - WFL) end up at 98. At first sight this violates the crucial rankings established in figure (56). Crucial ties save the analysis, as shown in tableau (75): the three violations of AFR in /(H2)(H1) L/ outnumber the single violations of WSP and AFR in /H (H1) L/.

(75) Crucial ties invalidate crucial rankings

|H H L|	Trochaic	Nonfinal	FtBin	AFR	WSP	Iambic	Parse
/(H1) H L/				**	*(!)		**
/(H1 H) L/				*	*	*(!)	*
☞ /H (H1) L/				*	*		**
/H (H1 L)/		*!			*		*
/(H H1) L/	*!			*	*		*
/(H1)(H2) L/				***!			*
/(H2)(H1) L/				***!			*

What's more, even FtBin can be ranked equally high as Iambic, at least if crucial ties are allowed, as tableau (76) shows.

(76) A crucial tie between FtBin and Iambic at work

|L L L|	Nonfin	Trochaic	AFR	FtBin	Iambic	WSP	Parse
/(L1) L L/			**(!)	*(!)			**
☞ /(L1 L) L/			*		*		*
/L (L1) L/			*	*			**!
/L (L1 L)/	*!						*
/(L L1) L/		*!	*				*

The concept of the crucial tie, inherited from the early days of OT, may not be worth pursuing. After all, how can two violations of the doubly gradient constraint AFR (which counts feet as well as distance) are worse than a single violation of the singly gradient constraint WSP (which counts syllables)? Under a more realistic interpretation of tied constraints, namely that by Anttila (1997), an equal ranking of AFR and WSP in tableau (55) means that both /H (H1) L/ (/au (díː) re/) and /(H2)(H1) L/ (/(aù)(díː) re/) would win in 50% of the cases, and an equal ranking of Iambic and FtBin in tableau (76) would mean that both /(L1 L) L/ (/(já.ke) re/) and /L (L1) L/ (the overtly incorrect /ja (ké) re/) would win in 50% of the cases. This optionality could be introduced in the simulations by taking a tiny evaluation noise, say 0.000001, for the CD simulations performed with Praat. All 10 CD learners would end up in the ranking in (74).

It remains to be said what the remaining two CD learners did. Like the four crucial tie learners just discussed, they had three strata, but one of them had WSP in the bottom stratum (at 98), and the other had MAIN-R, WFR, and WSP in that stratum. Apparently, both of these had managed to learn the language, but again by relying on the crucial tie principle.

The ranking in tableau (74) and those discussed after (74) produce all 28 forms in the first column of table (69). The rankings also correctly generalize to the 34 longer forms that the learner has never heard: they predict, for instance, /H L (H1 L) H/ (/in.di (gén.ti) a/ 'want') and /L L L L (L1 L) L/ (a form that we have not encountered so far).

The next step is to see how the GLA learners have performed. They cannot be bothered by crucial ties, because with a non-zero evaluation noise the probability that two constraints are ranked equally high at evaluation time is practically zero. If two constraints are ranked at nearly the same height, the distribution of outputs of the grammar will be very similar to the Anttila interpretation of a pair of tied constraints. All GLA learners end up with the ranking in table (77), although the precise ranking values differ among the learners, and half of the learners have a reversed ranking for the two bottom-ranked constraints WFL and PARSE.

(77) A typical uneven-trochee GLA learner of the uneven-trochee analysis

Constraints	Ranking values
NONFINAL	110.027
TROCHAIC	105.725
AFR	105.057
FTBIN	104.664
IAMBIC	100.539
WSP	99.984
MAIN-R	99.826
AFL	97.967
MAIN-L	94.105
WFR	89.973
WFL	89.702
PARSE	88.618

The crucial rankings of figure (56) can be found as large ranking distances in table (77). WSP and IAMBIC have stayed where they began, around 100. The three constraints that crucially outrank these two in (56) have been pushed up to about 105. The single constraint that crucially outranks two of the constraints around 105 has been pushed up to a height of 110. The constraint crucially dominated by WSP and IAMBIC (i.e. PARSE) has fallen a double distance, to the region near 90. This deep falling of weak constraints is typical of what the GLA does in general; in this case it is not a result of a domination by MAIN-L or so.

So far, CD and GLA learners have performed equally well, although the CD learners have practiced fancy behaviour by ingeniously inventing analyses with crucial ties whereas the GLA learners have boringly but reliably mimicked the expected ranking of figure (56).

For the T&S constraint set, table (73) shows that there exists no ranking that produces the forms associated with the uneven trochee analysis. According to table (68), this can only be due to a difference between TROCHAIC and FTNONFIN. Indeed we saw in §4.2.4 that the combination of FTNONFIN and IAMBIC is not capable of performing the conspiracy that led the combination of TROCHAIC and IAMBIC to force a winner with a monosyllabic foot in tableau (55). If TROCHAIC is replaced with FTNONFIN in tableau (55) or figure (56), we see that /(H1 H) L/ becomes the winner, because /H (H1) L/ now violates FTNONFIN, which is higher ranked than IAMBIC, which remains the only constraint in (55) and (56) that favours /H (H1) L/ over /(H1 H) L/. But it is still instructive to see how CD and GLA learners perform with the T&S set. Table (78) shows where one CD learner was after the simulations had to stop, i.e. after 1 000 language data.

(78) A T&S CD learner of the uneven-trochee analysis

Constraints	Ranking values
NONFINAL	100
AFR, FTBIN, MAIN-R, WFR, WSP	99
WFL	-102
IAMBIC	-109
AFL, FTNONFIN , MAIN-L, PARSE	-110

This learner has not ended up in a stable grammar. If she encounters more language data, the six constraints at the bottom will continue tumbling down the hierarchy. All ten CD learners have these six constraints ranked in different orders, but all in the vicinity of -110, which will be around -320 after 2 000 language data. At this snapshot in time, the learner of table (78) has iambic forms like /(L L1) L/ (/(ja.ké) re/). When being told that the form should be /(L1 L) L/ (/(já.ke) re/), she will demote IAMBIC to -111. Unfortunately, this will in turn lead her to generate a trochaic /(H1 H) L/ (/(aú.diː) re/). When being told that this should have been /H (H1) L/ (/(au.díː) re/), she will demote WFL, AFL, FTNONFIN, MAIN-L, and PARSE to -112, because all of these constraints prefer /(H1 H) L/ to /H (H1) L/ (and are higher ranked than IAMBIC, the highest constraint that prefers /H (H1) L/). This will go on forever. To measure how well these learners behave as speakers of Latin, their *error rates* were computed in the following way. A 1 000 underlying-surface pairs were randomly drawn, chosen with equal probability from the 28 underlying-surface pairs that have been used in training (therefore, each pair was chosen approximately 36 times on average), and then the learner's surface form for the given underlying form was computed. Each learner's form was then compared with the given adult surface form, and the learner was considered correct if the surface forms were identical. If a learner had e.g. 600 forms correct, her error rate was 40%. Eight of the CD learners turned out to have error rates of approximately 65%, the remaining two had error rates of about 44%.

The GLA learners also fail with the T&S set, but in a different way from the CD learners. The GLA learners all end up in a stable grammar. Table (79) shows the result for one learner.

(79) A T&S/GLA learner, fed with the uneven-trochee analysis

Constraints	Ranking values
NONFINAL	156.752
AFR	150.041
WSP	144.622
IAMBIC	139.944
FTNONFIN	139.618
FTBIN	95.926
MAIN-R	70.001
WFR	43.248
WFL	-401.795
AFL	-1078.115
MAIN-L	-1095.503
PARSE	-1204.395

IAMBIC and FTNONFIN are ranked very close together. Half of the 10 GLA learners have the same ranking as in (79), half have IAMBIC and FTNONFIN reversed. The distance between these two constraints is always small, so that if the learner has evaluation noise during her productions, she will have the ranking IAMBIC >> FTNONFIN approximately half of the time, and FTNONFIN >> IAMBIC the other half of the time. The error rates computed with an evaluation noise of 1.0 (smaller than the noise during training) are between 48% and 58%; the typical errors are that the learners show variation between /(L1 L) L/ and /(L L1) L/ and between /(H1 H) L/ and /(H H1) L/.

Table (73) shows that none of the three constraint sets without FTBIMORAIC is capable of learning a truly bimoraic analysis, like the 'at most bimoraic' and 'at least bimoraic' analyses of table (69). This is not so surprising. We have already seen in §4.2.5 that without constraints that favour strictly bimoraic feet, like *(HL) or FTBIMORAIC, one cannot expect the grammars to be able to learn bimoraic data. Still, the simulations with the moraic-trochee constraint set were successful in learning the uneven trochee analysis. Table (80) shows the resulting grammar for a CD learner.

(80) The generic result for moraic-trochee CD learning of the uneven-trochee analysis

Constraints	Ranking values
HDNONFIN, TROCHAIC	100
AFR, FTBIN, MAIN-R	99
IAMBIC	98
AFL, MAIN-L, WFL	97
PARSE, WFR, WSP	96

This grammar was reached by six of the ten CD learners. This grammar works for both versions of non-stochastic OT: that with crucial ties and that with variationist (*Anttila*) ties. The remaining CD learners had a grammar with a depth of 4, in which IAMBIC was one stratum higher, at 99 (now that the third stratum had been vacated, the ranking of the six constraints dominated by IAMBIC was of course 1 higher as well); this grammar relied on a crucial tie between IAMBIC and FTBIN, as above in the case of the T&S constraint set. Table (81) shows the result for a GLA learner. In (81), no clear layering has yet been established. This could be due to the fact that learning was designed to stop when the error rate was 0% if the evaluation noise was set to zero. With an evaluation of 2.0, i.e. the same as during learning, the error rate for the learner in (81) is still 30%. This means that in the grammar state of (81), the learner will detect mismatches for 30% of the incoming data and therefore take another learning step in 30% of the cases. These learning steps will continue to increase the separation between the constraints in (81).

(81) A moraic-trochee GLA learner of the uneven-trochee analysis

Constraints	Ranking values
HDNONFIN	110.744
TROCHAIC	106.141
FTBIN	105.867
AFR	105.058
IAMBIC	103.180
MAIN-R	102.464
AFL	100.627
WSP	97.805
MAIN -L	88.622
WFL	85.479
PARSE	78.739
WFR	78.037

In order to see whether the crucial rankings had been established in (81), the error rate was computed for an evaluation noise of 1.0. It was 7%; this number tells us something about how the learner will behave after making twice as many learning steps as she has made before reaching the state in (81).

In all cases the resulting ranking for the moraic-trochee constraint set is rather different from Prince & Smolensky's (1993) proposal, which was discussed in §4.2.5. As predicted, a high ranking of AFR rules out secondary stress before the main stress, and AFL >> WSP rules out secondary stress *after* the main stress.

It appears from table (73) that including the FTBIMORAIC constraint improves learnability from underlying-surface pairs. It is not surprising that if the uneven-trochee set and the moraic-trochee set succeeded in learning the uneven trochee analysis, this analysis is still learnable if a constraint is added to these constraint sets. But the addition of FTBIMORAIC seems to be just enough for the T&S set to achieve successful acquisition. The cause of this is that FTBIMORAIC is capable of ruling out (H H) feet but not (L L) feet: FTNONFIN can now outrank IAMBIC in order to produce /(L1 L) L/ rather than /(L L1) L/, without fear of producing /(H1 H) L/, because this form is ruled out by FTBIMORAIC. Otherwise, uneven trochees remain, as in /(H1 L) L/.

In general, the uneven trochee analysis seems to require fewer constraints (twelve) than the bimoraic analyses. The uneven-trochee and the moraic-trochee constraint set seem to be more successful than the T&S constraint set. But the differences between the constraint sets are small, especially regarding the success of the moraic-trochee+FTBIMORAIC set with the 'at least bimoraic' analysis. The next section shows whether there are any differences between the constraint sets when learning from overt forms only. The current section has at least shown that there were *some* combinations of constraint sets and analyses that were capable of learning the Latin stress system, so that we can now turn with confidence to the more realistic simulations, those for learning from overt forms, where hidden structures like feet are not explicitly provided to the learner but where she will have to construct them by herself.

4.6.2 Learning hidden structure and primary stress in Latin

Now we are going to have a look at the results of the simulations with more realistic primary language data. Table (73) shows that the T&S constraint set is not capable of learning a ranking for primary-stress-only overt data. This is not surprising, since the three primary-stress-only analyses (i.e. sets of given underlying-surface pairs) are not learnable with the T&S set either. Of course the learners could have invented a fourth analysis, perhaps one that includes /(L1 L)/ and /(L1) H/ or so, but they did not, so it is possible that there exists no analysis at all for primary-stress-only Latin with the T&S set.

The simulations with the uneven-trochee constraint set are more interesting: CD fails with this constraint set, the GLA succeeds. The first question now is: what analysis did the GLA learners come up with? The answer is that all learners came up with the same analysis, namely the uneven trochee analysis, i.e. for each of the 28 underlying forms in (69) they would produce the corresponding surface form in the 'uneven trochee' column (these surface forms were computed by running the 28 underlying forms through the learner's final grammar with an evaluation noise of zero). These learners ended up with the ranking in (82), sometimes with a different permutation of the very closely ranked constraints FTBIN, WSP, and TROCHAIC, or of IAMBIC, AFL, and MAIN-R.

(82) A typical result for uneven-trochee GLA learning from overt forms: creation of the uneven-trochee analysis

Constraints	Ranking values
NONFINAL	114.290
AFR	108.639
FTBIN	104.784
WSP	104.476
TROCHAIC	104.470
IAMBIC	101.302
AFL	100.739
MAIN-R	99.521
MAIN-L	95.039
WFR	85.710
PARSE	82.381
WFL	82.127

The ranking looks very different from that in (77). Still, (82) satisfies all of the crucial rankings. It will come to no surprise that these learners also correctly generalize the uneven-trochee analysis to forms of more than four syllables. The learners were rather slow in constructing the uneven trochee analysis by themselves. Whereas in the case of the underlying-surface pairs of table (69) all GLA learners had succeeded after the first 1 000 data, the learners of the overt forms needed 3 to 35 rounds of 1 000 data to arrive at an appropriate ranking. But they all succeeded.

Table (73) showed that for the three sets of 12 constraints, 29 out of 30 CD learners of primary-stressed overt forms fail. There is only one CD learner who happens to acquire an appropriate 12-constraint grammar; this learner uses the moraic-trochee set and invents an analysis that has not been considered so far, combining the two ‘at least bimoraic’ forms /(L1 L)/ and /(L1 H)/ of table (69) with the uneven trochees /...(H1 L) X/. This does not sound as a success for CD, since if only 10% of the children had been capable of learning Latin, this language would have perished much faster than it did.

When the constraint sets are enriched with FTBIMORAIC, the performance of CD improves. With the moraic-trochee+FTBIMORAIC set, nine learners managed to construct a functioning analysis. Seven of these came up with the uneven trochee analysis with the ‘at least bimoraic’ form

/(L1 X)/ mentioned before. The rankings of these learners slightly varied, as before. Table (83) shows an example.

(83) A typical result for moraic-trochee+FTBIMORAIC CD learning from overt forms: creation of the at-least-bimoraic uneven-trochee analysis, with empty strata

Constraints	Ranking values
FTBIN, HDNONFIN	100
AFR, FTBIMORAIC, MAIN-R	99
AFL, MAIN-L, WFL	98
PARSE, WFR	97
TROCHAIC, WSP	94
IAMBIC	93

A conspicuous property of seven of the resulting rankings is that they contained *empty strata*. In table (83), which is an average case, strata 95 and 96 are empty. Such empty strata can never occur when CD learns from fully specified underlying-surface pairs because constraints are demoted minimally, but they can when CD learns from overt forms only. Another conspicuous property of the seven rankings is that none of them is correct under the variational interpretation of tied constraints. To see whether there exists such a ranking at all, a simulation would have to be run in which the moraic-trochee+FTBIMORAIC constraint set learns an explicitly given at-least-bimoraic uneven trochee analysis. If so, and if we want to see whether CD can also learn it from overt forms, the simulations that led to table (83) will have to be rerun with a tiny evaluation noise.

Two of the CD learners constructed an at-least-bimoraic analysis. Both relied on crucial ties. Table (84) shows one of the rankings.

(84) Another result for moraic-trochee+FTBIMORAIC CD learning from overt forms: creation of the at-least-bimoraic analysis

Constraints	Ranking values
FTBIMORAIC, FTBIN, HDNONFIN	100
AFR, MAIN-R	99
AFL, MAIN-L, WFL	98
PARSE, TROCHAIC, WFR, WSP	97
IAMBIC	96

The ranking of FTBIMORAIC above AFR and MAIN-R causes the preference for /(H1) L L/ over /(H1 L) L/. We can compare this to the ranking in (83), where the crucial tie between these three constraints favours the uneven trochee /(H1 L) L/ over the bimoraic /(H1) L L/: FTBIMORAIC casts a single vote in favour of /(H1) L L/ whereas AFR and MAIN-R gang up with two votes in favour of /(H1 L) L/. The tenth moraic-trochee+FTBIMORAIC CD learner did not succeed in learning Latin. Her ranking after 1 000 data is given in (85).

(85) The single failure for moraic-trochee+FTBIMORAIC CD learning from overt forms

Constraints	Ranking values
FTBIN	100
HDNONFINAL	99
AFR, MAIN-R	98
AFL, FTBIMORAIC, MAIN-L, WFL	97
PARSE, WFR, WSP	96
IAMBIC	-104
TROCHAIC	-105

This learner has experienced IAMBIC and TROCHAIC tumbling down the hierarchy, alternatingly making the by now usual mistakes of /(H1 H) L/ and /(L L1) L/. To see whether this learner would learn the language later, she was taught 10 000 extra overt forms. This had no other effect than demoting IAMBIC and TROCHAIC down to -2238 and -2239. It appears that this learner, in contrast with the tenth GLA learner of the uneven-trochee set discussed above, has really got trapped in a sequence of grammars that she can never get out of (a ‘non-globally-optimal limit cycle’).[32] This may mean that the ‘9’ in table (73) indicates that primary-stressed-only Latin is not learnable by the whole generation of learners if they entertain the moraic-trochee+FTBIMORAIC constraint set. Whether this situation means that this combination of constraint set, training set, and learning algorithm can be ruled out as a proposal for how Latin children acquired their language, or whether it is just a predictor of sound change, depends on the exact fraction

---

[32] This ‘non-globally-optimal limit cycle’ means that the learner ended up in something like a one-way dead-end street: no matter how much more language data this learner encounters, she will never be able to find a way back, and a way to a correct grammar.

of failures. The best guess at this point is 10%, but this number could be estimated more accurately after a future simulation of, say, 1 000 learners. The 10 GLA learners, by the way, were consistent in not learning with the moraic-trochee+FTBIMORAIC constraint set at all.

The remaining interesting figure for the main-stress-only forms in table (73) is the '7' for the GLA learners with the uneven-trochee+FTBIMORAIC constraint set. Apparently, adding the FTBIMORAIC constraint to the set made the language *less* learnable from overt forms, compared to the uneven-trochee constraint set without FTBIMORAIC. The seven successful learners ended up with rankings that follow the stratification in (86). Interestingly, although FTBIMORAIC is ranked above FTBIN, this grammar renders uneven trochees such as /(H1 L) L/ and /(H1 L) H/, because FTBIMORAIC is dominated by AFR.

(86) A typical success for uneven-trochee+FTBIMORAIC GLA learning from overt forms

Constraints	Ranking values
NONFINAL	120.563
AFR	113.294
WSP	106.877
FTBIMORAIC	105.181
AFL	103.154
TROCHAIC	103.016
FTBIN	102.790
MAIN-R	102.439
IAMBIC	100.746
MAIN-L	97.615
PARSE	81.554
WFR	79.437
WFL	73.763

The remaining three learners were not lucky. Even after 50 000 data, they stuck with grammars like in (87). TROCHAIC and IAMBIC are always ranked very closely. Grammar (87) is of the type that we have seen several times before: since TROCHAIC and IAMBIC are very closely ranked, these learners end up producing one of the two mistakes /(H1 H) L/ or /(L L1) L/. The cause of the problem is that these learners have moved AFL too high up, and not managed to raise FTBIMORAIC above it. If FTBIMORAIC is ranked higher than AFL, it is capable of ruling out /(H1 H) L/, so that IAMBIC is freed from the task of ruling out/(H1 H) L/; this allows IAMBIC to fall below TROCHAIC, so that the learner also stops producing /(L L1) L/ errors. Apparently, adding a constraint does not necessarily improve learnability from overt forms.

(87) A typical failure for uneven-trochee+FTBIMORAIC GLA learning from overt forms

Constraints	Ranking values
NONFINAL	124.255
AFR	115.923
AFL	107.371
WSP	107.226
MAIN-R	103.226
FTBIN	101.215
MAIN-L	100.711
TROCHAIC	98.868
IAMBIC	98.633
FTBIMORAIC	95.829
PARSE	83.892
WFR	75.745
WFL	71.173

4.6.3 Learning hidden structure including secondary stress in Latin

Table (73) showed that CD is not capable of learning from overt forms with the secondary stresses listed in the last column of table (70), with any constraint set. By contrast, the GLA is successful with the T&S and uneven-trochee constraint sets, regardless of whether FTBIMORAIC is included or not. This looks better than the performance with the primary-stress-only forms, which could mean that additional information such as secondary stress does support learning.

Apart from the striking difference between the learning algorithms, the most conspicuous result in table (73) is that the T&S constraint set is successful for the first time. The 10 GLA learners created grammars very similar to the one in table (88).

(88) The result for T&S GLA for secondary-stressed overt forms

Constraints	Ranking values
NONFINAL	108.705
WSP	104.865
MAIN-R	102.437
FTBIN	101.430
WFL	100.773
FTNONFINAL	99.888
AFL	99.852
PARSE	99.273
IAMBIC	97.686
MAIN-L	95.353
AFR	91.682
WFR	91.295

For the 28 overt forms in table (70), this learner constructs an analysis with rather exhaustively parsed syllables and both iambic and uneven trochaic feet, with a preference for trochees:

(89) Mixed foot structure

/(L1) X/
/(H1) X/

/(L1 L) X/
/(L H1) X/
/(H1 L) X/
/(H2)(H1) X/

/L (L1 L) X/
/(L2 L)(H1) X/
/L (H1 L) X/
/(L H2)(H1) X/
/(H2)(L1 L) X/
/(H2 L)(H1) X/
/(H2)(H1 L) X/
/(H2)(H2)(H1) X/

The learner generalizes this exhaustivity to the 32 forms with five syllables, some of them containing both iambs and trochees:

(90) Forms with five syllables

/(L2 L)(L1 L) X/
/(L2 L)(L H1) X/
/(L2 L)(H1 L) X/
/(L2 L)(H2)(H1) X/
/(L H2)(L1 L) X/
/(L H2)(L H1) X/
/(L H2)(H1 L) X/
/(L H2)(H2)(H1) X/
/(H2 L)(L1 L) X/
/(H2 L)(H1 L) X/
/(H2 L)(H2)(H1) X/
/(H2)(H2)(L1 L) X/
/(H2)(H2 L)(H1) X/
/(H2)(H2)(H1 L) X/
/(H2)(H2)(H2)(H1) X/
/(H2 L)(L H1) X/

Importantly, the learner analyses /(H2 L)(L H1) X/ and not /(H2)(L2 L)(H1) X/. Note that in these cases, the learner has created her own patterns of overt forms, e.g. [L2 L L H1 X], which were not in the training set. This means that the learner will produce reasonably good pronunciations for five-syllable forms, even if she has never heard them before; for instance, if the learner is familiar with the nominative /ra (pí.di) taːs/ 'speed', she will come up with the form /(rà.pi) (di.táː) te/ for the ablative singular even if she has never heard that form. For the longest forms consisting of light syllables only, the analyses have a single left-aligned foot that contains a secondary stress: /(L2 L) L (L1 L) L/ and /(L2 L) L L (L1 L) L/. The exhaustivity noted above thus reduces (only in the case of light syllables) to a right-aligned main foot and a left-aligned secondary foot, which is caused by a high ranking of WFL, and a ranking of AFL above PARSE. Three other learners have exactly the same language, and three learners have a slightly different ranking that leads to exactly the same forms as above except that the form with seven syllables scans as /(L2 L)(L2 L)(L1 L) L/. This even more exhaustive parsing of syllables is caused by the ranking PARSE >> AFL. Actually, the speaker in table (88), with her close ranking of AFL and PARSE, can be expected to waver between the two forms with seven syllables. This variation (both between speakers and within speakers) seems to be similar to what real speakers of English, German or Dutch do with longer words (it could even depend on speaking rate, i.e., you could rank PARSE a bit lower when speaking fast). The remaining three GLA learners have /L (H2 L)(H1) X/ instead of /(L H2)(L H1) X/ (with variation in the seven-syllable form) caused by a ranking of FTNONFIN over WFL and PARSE. It is highly questionable whether the difference between those two forms is audible, so large-scale interspeaker variation for such hidden structures within the speech community should come to no surprise.

The results with the uneven-trochee constraint set are quite different. Table (91) shows the final ranking of one learner:

(91) One uneven-trochee GLA learner for secondary-stressed overt forms

Constraints	Ranking values
NONFINAL	108.761
FTBIN	104.053
WSP	103.321
MAIN-R	102.504
TROCHAIC	102.322
PARSE	99.965
AFL	99.818
WFL	97.991
IAMBIC	97.735
MAIN-L	95.671
AFR	94.134
WFR	91.239

This learner avoids iambic forms: she has /L (H1) X/, /L (H2)(H1) X/, /(L2 L) L (H1) X/, /L (H2 L)(H1) X/, and, this time, no other choice than the exhaustive form /(H2)(L2 L)(H1) X/. Since PARSE outranks AFL, the seven-syllable form is /(L2 L)(L2 L)(L1 L) L/. All nine other uneven-trochee GLA learners have AFL >> PARSE, and therefore the forms /(H2 L) L (H1) H/ and /(L2 L) L L (L1 L) L/. Adding FTBIMORAIC to the uneven-trochee constraint set can result in the ranking in (92). The learner in (92) has come up with an analysis that has uneven trochees for main stress (caused by MAIN-R >> FTBIMORAIC), but avoids uneven trochees for secondary stress (caused by the ranking FTBIMORAIC above PARSE and AFR): both phenomena can be seen in /(H2) L (H1 L) L/ . This learner also has /(H2)(L2 L)(H1) X/ and /(L2 L)(L2 L)(L1 L) L/.

Eight other GLA learners arrive in the same language as the learner in (92), except that three of them have a reverse ranking of PARSE and AFL, resulting in forms such as /(H2) L L (H1) X/ and /(L2 L) L L (L1 L) L/.

(92) uneven-trochee+FTBIM/GLA learning from secondary-stressed overt forms

Constraints	Ranking values
NONFINAL	108.926
WSP	103.858
MAIN-R	103.641
FTBIN	103.238
FTBIMORAIC	102.983
PARSE	101.836
AFL	100.866
TROCHAIC	100.785
WFL	97.937
MAIN-L	97.039
IAMBIC	96.771
AFR	94.656
WFR	91.074

The remaining learner, shown in (93), happened to come up with a real bimoraic analysis that avoids all uneven trochees, e.g. /(H2) L (H1) L L/.

(93) uneven-trochee+FTBIM/GLA learning with secondary-stressed overt forms: creation of the at-most-bimoraic analysis

Constraints	Ranking values
NONFINAL	108.711
WSP	103.607
FTBIMORAIC	102.917
MAIN-R	102.891
FTBIN	102.450
TROCHAIC	101.201
PARSE	100.534
AFL	100.430
WFL	97.738
IAMBIC	96.920
MAIN-L	96.906
AFR	94.262
WFR	91.289

The 10 GLA learners with the T&S constraint set and FTBIMORAIC behaved similarly: eight created the bimoraic analysis with the exhaustive forms /(H2)(L2 L)(H1) X/ and /(L2 L)(L2 L)(L1 L) L/, one a bimoraic analysis

with medially unfooted light syllables, i.e. /(H2) L L (H1) X/ and /(L2 L) L L (L1 L) L/, and one allowed uneven trochees in main feet only.

4.6.4 Conclusions

First it has to be said that learning Latin from overt data turns out to be possible. However, it also brings about some instances of the expected failures of CD and GLA (as mentioned in section 3.6), since the overt forms are often ambiguous with respect to their structural analysis. In fact, the only combination of constraint set and algorithm that was capable of learning from primary-stress-only overt forms for all 10 learners was the uneven-trochee set with the GLA. A combination that got close to this performance was the moraic-trochee+FTBIMORAIC set with CD, where nine out of ten learners detected a correct ranking. In order to reliably prove that the former combination is better than the latter, it would be necessary to show that it is nearly 100% correct, for instance by teaching 1 000 learners with the uneven-trochee/GLA combination and computing the percentage correct. This could take two weeks of computer time.

Since both CD and GLA make use of the same interpreting mechanism (Robust Interpretive Parsing), any crucial differences in performance between the two have to be attributed to the different kinds of reranking strategy (demotion-only vs. demotion-and-promotion, and one-shot learning vs. graduality).

Again, learning Latin from overt data turns out to be possible, at least with the GLA algorithm. Whether this means that CD should be ruled out as a candidate for describing Latin with secondary stress remains to be seen, since different secondary stress patterns than tested so far are thinkable, as shown in section 4.9.

The learners came up with ten different analyses for the overt data with secondary-stressed forms, with a total of 109 different surface forms for the 62 underlying forms. For the forms with at most four syllables, the overt forms associated with these ten analyses were (and need to be) identical. Differences between the analyses showed up only in a couple of overt forms with five syllables (namely [H2 L L H1 X] versus [H2 L2 L H1 X]) and in a form with seven syllables (namely [L2 L L L L1 L L] versus [L2 L L2 L L1 L L]), and this difference occurred with all five pairs of analyses that we have seen (i.e. ‘iambic & trochaic’, ‘less iambic’, ‘trochaic’, ‘bimoraic in secondary feet’, ‘bimoraic everywhere’), only depending on the relative ranking of AFL and PARSE, which were always closely ranked. Attested Latin forms with more than four syllables, if weight-sensitively secondary-stressed as here, would therefore give us no information about whether Latin learners used the T&S set or the uneven-trochee set, with or without FTBIMORAIC, and which of the five analysis types they created. Whether other patterns of secondary stress give us such information is investigated in §4.9.

4.7 More learners, different results?

The results in the previous section showed that the GLA learners equipped with the uneven-trochee constraint set (the one with TROCHAIC as the trochaicity constraint) could learn the Latin stress pattern from overt forms with only primary stress, whereas the CD learners could not learn with this constraint combination, but performed best when learning with the moraic-trochee constraint set including FTBIMORAIC. Since the learners were trained on the same word forms, but encountered each learning item in a different order, variation in the outcome is possible. To get a better understanding on whether the CD learners can *consistently* not learn the Latin stress pattern with the uneven-trochee constraint set as opposed to the GLA learners a 100 more learners (50 CD and 50 GLA learners) were run. Again the CD learners did not succeed: they ended up in a grammar that could not reproduce the correct stress pattern. And again the GLA learners did succeed: they came up with an analysis close to Jacobs’ (2000) analysis.

Once more, the learners were trained on all possible sequences of heavy and light syllables. Thus, four patterns of overt disyllables were fed to the virtual child. Likewise, there are eight trisyllabic overt forms, and in the same vein 16 overt forms with four syllables, all following the penultimate/antepenultimate Latin stress rule. The overt forms displayed in (94) are the same as in table (70):

(94) 28 primary-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]

For each of the 100 virtual learners (50 CD learners and 50 GLA learners, once more created in the Praat programme; Boersma & Weenink 1992-2006), all 12 constraints were initially ranked at a height of 100, whereupon 10 000 language data were drawn randomly with equal probability from the 28 overt forms. The evaluation model for CD was OT with crucial ties. As in Tesar and Smolensky (2000), the algorithm was allowed to chew five times on each piece of language data, with backtracking if the pentuple chews did not succeed in making the (alleged) correct adult form optimal in the learner's grammar. When two forms were equally harmonic, a winner was chosen randomly from among them. The evaluation model for the GLA was Stochastic OT with an evaluation noise of 2.0. The CD learners took learning steps of 1.0, and the GLA learners had decreasing learning steps, starting with 0.1; this plasticity of 0.1 was further randomized by a relative plasticity standard deviation of 0.1.

4.8 More learners, the same results

None of the CD learners succeeded in learning the stress pattern of Latin. The ranking after 10 000 data of one showcase CD learner is given in (95).

(95) A failing CD learner, after 10 000 data

Constraints	Ranking values
FTBIN, NONFINAL	100
AFR, MAIN-R, PARSE, WFR	99
AFL, MAIN-L, WFL	98
WSP	–2497
TROCHAIC	–2498
IAMBIC	–2499

At this snapshot in time, this child produces correct forms like /(L1 L) X/ but also incorrect forms like /(H1 H) X/. When hearing the correct overt form [H H1 X], the child will perceive this as /(H H1) X/, given the ranking in (95). This will lead her to demote TROCHAIC below IAMBIC, i.e. to –2500. But this new grammar will incorrectly produce /(L L1) X/, so that when hearing [L1 L X] the learner will demote IAMBIC below TROCHAIC again. These two constraints will continue to tumble down hopelessly along the ranking scale. They will drag along WSP, because when WSP is ranked above TROCHAIC, the learner can make the error /(L H1)/, so that hearing [L1 H] will lead her to demote WSP below TROCHAIC.

In contrast to the CD learners, all 50 GLA learners succeeded (though five of them needed between 10 000 and 200 000 data to converge). Table (96) shows an example.

(96) A successful GLA learner, after 10 000 data

Constraints	Ranking values
NONFINAL	114.290
AFR	108.639
FTBIN	104.784
WSP	104.476
TROCHAIC	104.470
IAMBIC	101.302
AFL	100.739
MAIN-R	99.521
MAIN-L	95.039
WFR	85.710
PARSE	82.381
WFL	82.127

It is now discussed to what forms this ranking leads in production. The top ranking of NONFINAL leads to final-syllable extrametricality: all winners have a final unfooted syllable whose weight does not influence foot structure at all. The disyllables therefore become /(L1) X/ and /(H1) X/, where ‘X’ stands for any final syllable. High-ranked AFR will now make sure that every foot of every word will end after the penultimate syllable. This means that there will only be a single foot in every word, one that ends just before the extrametrical syllable. In forms of more than two syllables, the high ranking of FTBIN will make sure that if the penultimate syllable is light, the antepenultimate syllable will be included in the foot. If this antepenultimate syllable is heavy, WSP will make sure that it is stressed: /...(H1 L) X/; if the antepenultimate syllable is light, it is TROCHAIC that will make sure that it is stressed: /...(L1 L) X/. The situation becomes slightly complicated when we turn to forms ending in |...H X|. Of the three forms /...(L H1) X/, /...(L1 H) X/, and /...L (H1) X/, all of which satisfy FTBIN, only the last one satisfies both WSP and TROCHAIC, so it wins. For |...H H X| the relevant candidates are /...(H H1) X/, /...(H1 H) X/, and /...H (H1) X/. All three are equal as far as FTBIN and WSP are concerned, and the last two satisfy TROCHAIC. The decision between these two will have to be brought by IAMBIC, shown in tableau (97).

(97) A constraint hierarchy that works for all Latin forms

|H H L|	NONFINAL	AFR	FTBIN	WSP	TROCHAIC	IAMBIC	AFL	MAIN-R	MAIN-L	WFR	PARSE	WFL
/(H1) H L/		**!		*				**		*	**	
/(H1 H) L/		*		*		*!		*		*	*	
☞ /H (H1) L/		*		*			*	*	*	*	**	*
/H (H1 L)/	*!			*			*		*		*	*
/(H H1) L/		*		*	*!			*		*	*	
/(H1)(H2) L/		**!*					*	**		*	*	
/(H2)(H1) L/		**!*					*	*	*	*	*	

Figure (56) showed which of the rankings in (97) are crucial (ignoring the four less interesting and low-ranked constraints WFL, WFR, MAIN-L, and MAIN-R). The rankings not marked by lines in this figure are not fixed.

Thus, TROCHAIC could be ranked anywhere between the very top and a position below WSP, as long as it outranks IAMBIC; FTBIN could be ranked above AFR or below WSP, as long as it is ranked below NONFINAL and above IAMBIC; and so on.

The ranking in (56) is close to Jacobs' (2000) ranking: TROCHAIC >> NONFINAL >> FTBIN >> LFR >> WSP >> PARSE, where LFR is a less-gradient version of AFR (it counts the number of syllables from the *last* foot to the end of the word). The crucial difference is the insertion of IAMBIC into the hierarchy. This is required to account for the |...H H X| forms, which Jacobs did not consider in his analysis. Admittedly, it is counter-intuitive that in a language with exclusively trochaic feet, the constraint IAMBIC has to fix the analysis.

As far as longer forms are concerned: the learners have not been trained with forms of five syllables, but we can nevertheless run the 32 possible forms with five syllables through their respective tableaux and see what happens. All forms are handled correctly, for instance /L L (L1 L) H/, /H L L (H1) L/, /L H (H1 L) L/, and /H H H (H1) H/. The forms with six and seven syllables are /L L L (L1 L) L/ and /L L L L (L1 L) L/. Thus, the generalization to longer forms has succeeded.

4.9 Secondary stress in Latin?

Another group of virtual learners was trained on several other 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 CD. The simulations in sections 4.61 and 4.6.2 dealt with primary stress only. Since it is controversial whether Latin had secondary stress, and if so, what it exactly looked like, several different data sets are designed with secondary stress. The virtual learners were then tested whether they were able to learn from these data sets, provided with the basic metrical constraint sets. For each simulation, a number of virtual CD and GLA learners was created, with constraint sets that contained either TROCHAIC or FTNONFIN. As before, the constraints started out with the same ranking heights (100.000). The training data were two to four syllables long and drawn randomly with equal probability from 28 possible overt forms. The CD learners were fed with 1 000 data pieces, while the GLA learners were fed with 10 000 up to 40 000 data pieces (because the GLA learners take smaller reranking steps). All of the simulations were carried out with the Praat programme (Boersma & Weenink 1992-200).

4.9.1 Very weight-sensitive secondary stress

One option would be to have weight-sensitive secondary stress by stressing every heavy syllable and every other light syllable before the main-stressed one. The resulting data set that the learners are fed with is given in table (98). Disyllables were also used in the simulations, but they are suppressed in the table since they do not differ from those in table (94).

(98) Very weight-sensitive secondary stress

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

CD learners training with these 28 overt forms failed: 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 primary language 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)(táː) teːs/. The GLA learners with FTNONFIN produced forms such as /L (H2)(H1) H/ as in /vo (lùp)(táː) teːs/, /L (H2)(L1 L) L/ as in /a (mìː)(kí.ti) a/, and /(H2)(H2)(H2)(H1) H/ as in /(dèː)(fìː)(nìː)(tíː) vus/.

4.9.2 Weight-insensitive secondary stress in Latin?

Another possibility to assign secondary stress is to build a weight-insensitive disyllabic foot at the left edge of the word, resulting in overt forms as in (99) (the 20 forms without secondary stress are suppressed).

(99) 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]

All learners that learn from this data set fail, simply because there *is* no ranking that can describe the data (at least no ranking with the constraints involved here).

The same happens with a training set that has left-aligned, binary weight-sensitive secondary stress, as in (100): there is no OT analysis with the constraint sets involved here that could describe this pattern, so again, all learners fail (data without secondary stress, like [L H1 L H] and [L H1 L], are again suppressed, although they would make the weight-sensitivity more explicit).

(100) 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]

If learning from data that contain left-aligned weight-insensitive secondary stress that is not binary, as in (101), all learners fail, again because there is no constraint ranking that could produce the data.

(101) 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]

4.9.3 Freely assignable secondary stress

A further possibility to assign secondary stress is to let the learners invent it. This is done in the final 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 primary stress, they could construct more than one foot in a word. Given this, GEN would then provide an additional candidate for a form like [H1 L]: /(H1) (L2)/.[33] 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 faithfulness between overt form and surface form for secondary stress.[34] As usual, the choice between the candidates is determined by the ranking. The input to the simulations with freely assignable stress was therefore the same as in table (94): overt forms with primary-stress only. But this time the learners were allowed to invent secondary stress, i.e. their GEN contained not only forms with main stress, but also forms with main *and* secondary stress. Examples for the resulting constraint rankings are listed in tables (102) and (103):

(102) A CD learner

Constraints	Ranking values
FTBIN	100
NONFINAL	99
AFR, MAIN-R	98
AFL, MAIN-L, WFL	97
PARSE, WFR, WSP	96
IAMBIC	-104
TROCHAIC	-105

(103) 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
FTBIN	87.208
WFR	83.038
AFR	80.461

---

[33] This comes close to the forms in GEN of Prince & Smolensky (1993), where forms with feet to the right of the head foot were permitted: /(H1)(L L)/. In their analysis, these feet apparently did not carry secondary stress.

[34] This violation of faithfulness is not modelled with constraints in the simulations, but by adding candidates in GEN.

The CD learners 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/, and furthermore forms like /(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/.[35]

Their generalizations to longer forms were weird but correct: /(L2 L)(L H1) X/ as for /(rà.pi)(di.táː)tem/, and /(L2)(L2 L)(H1 L) X/ as for /(rà)(pì.di)(táː.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/. However, this does not look like a natural pattern.

4.9.4 Summary

Summary of successes. The successful simulations included three very different patterns with secondary stress, which could be learned by GLA learners only. A word like *voluptates* was analysed in the first simulation as /vo.lup (tá:) te:s/, in the second simulation as /vo (lùp)(tá:) te:s/ (with FTNONFIN), and as /(vo.lùp)(tá:) te:s/ (with TROCHAIC; the same in simulation 6). CD learners never converged upon a grammar that rendered the stress patterns in question. The immediate cause for this lies in the behaviour of the constraints for trochaicity and iambicity, since the CD strategy moves them to the bottom of the hierarchy early, while the GLA keeps them ranked in the middle.

Summary of failures. What is missing in the results of the 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. The constraint sets involved here are not capable of producing this pattern, regardless of the input. Furthermore, although forms like

---

[35] This pattern (secondary stress on the initial syllable) comes close to the one proposed in the literature by e.g. Allen (1973:86, and reference therein), except that stresses in consecutive syllables is unlikely: Allen (1973) states that secondary stress only occurred if there was more than one syllable preceding the mainly stressed one.

[L2 H H1 H] were given in simulations 4.9.2 and allowed in the simulation in 4.9.3, no learner came up with the analysis of secondary stress actually proposed in the literature (Allen 1973, 1978), which contains wretched trochees such as /(L2 H)(H1) H/ as for /*(vò.lup)(táː)teːs*/. It is likely that neither constraint set was suitable for such an analysis; the addition of a constraint like *CLASH could improve the results.

4.10 Discussion

In this rather long section on the learnability of Latin stress, the performance of two learning algorithms, six constraint sets, three analyses, and several kinds of overt forms have been investigated. This section reports some results on all these issues and indicates how several more constraint sets, analyses, and kinds of overt forms should be investigated in the future.

Analyses. The present investigation started by giving a couple of analyses that are capable of handling the positioning of Latin main stress correctly. The uneven-trochee analyses (§4.2.3) derived from Jacobs' (2000) analysis, which was augmented with IAMBIC in order to handle |H H L|. The moraic-trochee analysis (§4.2.5) derived from Prince & Smolensky's (1993) analysis, which was augmented with the ranking of AFR >> WSP and with the constraint IAMBIC, again in order to handle |H H L|. The uneven trochee analysis was better learnable than either of the two bimoraic analyses. However, from the simulations with overt forms a fourth analysis transpired that had not been considered before: an analysis with uneven trochees, as in Jacobs (2000), but with at-least-bimoraic feet, so that the light-initial disyllables become /(L1 L)/ and /(L1 H)/.[36] This fourth analysis may well lie at the basis of the process of *iambic shortening* in Pre-Classical Latin (underlying |L H|, e.g. the concatenation of the verb stem |am-| 'love' with the first singular ending |-oː|, becomes /(L1 L)/, e.g. /(á.mo)/ 'I love'), which many authors discuss (Kager 1993a, Prince & Smolensky 1993, Mester 1994, Jacobs 2000). Future research will have to take this analysis into account.

---

[36] These two forms actually occurred in Jacobs' original analysis for Classical Latin, but as shown in §4.2.3, these forms require (with the uneven-trochee constraint set) a ranking of FTBIN >> NONFINAL and of PARSE >> WSP, the latter of which fails to handle |L H L| correctly.

Learning algorithms. CD and the GLA performed strikingly differently in the simulations with overt forms. As summarized in table (73), the GLA succeeded with five combinations of constraint sets and kinds of overt forms. CD performed a bit better on the two constraint sets that contain the perhaps implausible constraint HDNONFIN, although each of these four groups of 10 CD learners had at least one learner who did not acquire Latin (see the rows 'moraic trochee' and 'moraic-trochee+FTBIMORAIC' in table (73), in combination with the columns 'main stress only' and 'secondary stress').

Constraint sets. The simulations seem to reveal that some of the proposed constraint sets are more adequate than others. For instance, TROCHAIC seems to be a better formalization for a trochaic foot pattern than FTNONFIN, which seems to be too restrictive. Also, NONFINAL seems to be a more effective formalization of extrametricality than HDNONFIN. But no constraint set can be ruled out completely yet. As usual in OT, the legitimacy of a constraint set ultimately has to be proven in combination with systems of other languages than the specific language under study.

Overt forms. Learning from forms with a certain type of secondary stress turned out to be easier than learning from forms with primary stress only. There is disagreement in the literature about whether Latin had secondary stresses, and therefore feet, before the primary-stressed foot, and if it had, where these secondary stresses were: they could have been weight-sensitive (Allen 1978) or not (Jacobs 1989). Allen (1978) states that if there was secondary stress, it was on the initial syllable, except if it was pretonic. Therefore it is unlikely that every heavy syllable was stressed in Latin as it was in the present series of forms, the last column in table (70). This series already led to ten different analyses, and other secondary stress patterns will lead to many more. A possible solution to this problem is to *let the learner decide*, as was done in chapter 4.9.3: the learner encountered only overt forms with primary-stressed syllables, but was allowed to invent a full foot structure with secondary-stressed syllables.

Frequency. In the present simulations, the learners were fed every type of underlying form equally often. The typical mistakes of the virtual learners were superheavy trochees in /(H1 H) X/ and iambs in /(L L1) X/. In most cases, the mistakes were caused by a close ranking of TROCHAIC (or FTNONFIN) and IAMBIC without a compensatory ranking somewhere else in the hierarchy. If these learners encountered more |H H X| than |L L X| forms, they would probably end up with a ranking of TROCHAIC slightly above IAMBIC. It is not unlikely that such a ranking would have helped the learner to avoid non-global optima, but this is left for later investigation.

Sound change. There were cases in which a small percentage of the learners did not succeed in acquiring the provided Latin stress pattern, while the great majority of the same type of learners did succeed. Such cases can be predictors of acquisition-induced sound change. It is possible, for instance, that not all constraints are innate, but that they are instead constructed by the learner. In that case, some learners may well entertain constraint sets that have here been shown to lead to unlearnability. The typical mistakes were trochaicity in /(H1 H) X/ and iambicity in /(L L1) X/. The disadvantage of taking a dead language to study acquisition can thus turn into an advantage, since we know a lot about what happened in the daughter languages. With some luck, later investigations may also be able to model the historical change from initial stress in Pre-Classical Latin to weight-sensitive right-aligned stress in Classical Latin.

More realistic models of metrical acquisition. It has simplifyingly been assumed that the learner's productions contained the same number of syllables as their underlying forms and the adult forms. However, it is likely that Latin children were similar to Dutch children (Fikkert 1994) and English children (Gnanadesikan 1995) in that they started out by truncating longer words, e.g. by turning trisyllabic words into disyllabic words consisting of a single foot, and that segmental structure interfered. Such a situation would have strong implications for all of the steps in the modelling of acquisition. For instance, this could mean that learners start out by acquiring everything there is to know about short words, before they go on to consider longer words. Dresher (1999), for instance, provides a non-OT metrical acquisition model that takes into account selective attention to specific structures.

Conspiracies between constraints. As seen in §4.2.2.3, TROCHAIC and IAMBIC conspire to minimize foot size. The alternative are FTNONFIN and IAMBIC. These two have fewer side effects since they have complementary violations on the foot level. This means that doing OT with the pair

FTNONFIN - IAMBIC is close to having a *parameter* “foot direction” in the grammar that is set to one of the values *nonfinal* or *iambic* (but not entirely, because these constraints still conspire to minimize the number of feet).

4.11 Conclusions

Latin stress turned out to be learnable with a limited set of constraints that many OT phonologists nowadays tend to regard as universal (i.e. cross-linguistically valid) as a result of years of typological research on many different stress systems. The virtual learners were tested on two *on-line* learning algorithms, whose only memory of past events is indirectly and concisely stored in the ranking values of the constraints: Constraint Demotion (CD) and the Gradual Learning Algorithm (GLA). The GLA learners turned out to be successful in more constraint set/analysis combinations than the CD learners. It can be argued that the GLA is a more realistic ingredient of human language acquisition than CD, which has been shown earlier: like real children, GLA learners learn gradually rather than abruptly, thus showing realistic gradual learning curves and realistic effects of the distributions of forms in the language data (Boersma and Levelt 2000; Curtin and Zuraw 2001); GLA learning is robust against modest levels of errors in the language data (Boersma 1998); the GLA is capable of handling continuous input data, like auditory cues in L1 and L2 perception (Escudero and Boersma 2004); and, last but not least, the GLA has been able to model language change induced by bidirectional language acquisition (Jäger 2003). Nevertheless, neither CD nor the GLA are capable of learning every metrical system predicted by *factorial typology*, i.e. every metrical system that results from a permutation of the rankings of the twelve constraints provided in (14) (Boersma 2003). Both learning algorithms fail for some rankings, but the rankings for which the two fail are different. If a learning algorithm fails precisely for those rankings that do not correspond to any existing language, this should be regarded as positive evidence for the appropriateness of such a learning algorithm for the description of real language acquisition. For the case discussed in this paper, the results provide more evidence for an appropriateness of the GLA. More languages and, especially, gaps in factorial typology (i.e. expected but non-existent languages) need to be investigated before we can conclude that any OT learning algorithm provides the appropriate model for the acquisition of language.

Looking at the failure of the learners to come up with strictly bimoraic analyses, it appears that something has to be done about FTBIN. It should be split up into separate constraints that refer to moras on one hand and syllables on the other, as e.g. proposed by Hewitt (1994).

Also, if perception precedes lexical access, foot structure has to be assigned before word boundaries are. This order is problematic because some of the used constraints imply a dependence of foot assignment on word boundaries. Consider the overt form [áː.bra.ka.dáː.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 *aː.bra#ka.daː.bra*, the footing would have to be *(áː)bra#ka(dáː)bra*. If it is *aː.bra.ka#daː.bra*, footing would have to be *(áː.bra)ka#(dáː)bra*. This makes the strictly bimoraic analysis more likely, since this bimoraic analysis would predict identical footing in *(áː)bra#ka(dáː)bra* and *(áː)bra.ka#(dáː)bra*, so that feet can be assigned independently from (e.g. before) word boundaries.

A last point is that the learners were given too much information about syllable weight. Children have to e.g. learn the heaviness of CVC syllables. In some languages, CVC is light (e.g. final, monomoraic CVC-feet in Chuukese; Davis 1999, Muller 1999, 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 should be given to the learner than was done in Tesar & Smolensky's (2000) and the present simulations.

The next chapter models weight-insensitive stress in Pintupi, where the learnability of coda moraicity is tested.

5 The learnability of grammatical stress and weight in Pintupi[37]

5.1 Introduction

It is often taken for granted that normally developing children acquiring one and the same language end up with one and the same grammar (e.g. Chomsky & Halle 1968:251). The language-acquiring child is supposed to be capable of creating the adult grammar from the information provided in the speech stream, despite the fact that this information may be incomplete in terms of possible ambiguities or gaps in the data she is exposed to (known as the *poverty of the stimulus* problem; Chomsky 1986:7). In the computer simulations of acquisition here it is shown that final grammars of virtual learners can differ even though they have been trained on the same data and have the same output as given in the training data. This is demonstrated by modelling the word stress pattern of Pintupi, a language spoken in Western Australia (Hansen & Hansen 1969). The learning algorithms applied in the computer simulations are once again Constraint Demotion (CD; Tesar 1995) and the Gradual Learning Algorithm (GLA; Boersma 1997).

The chapter is built up as follows: section 5.2 outlines Pintupi word stress and provides two possible Optimality Theoretic analyses. Section 5.3 outlines the learnability approach for Pintupi. Section 5.4 gives the ingredients to the computer simulations (the training data, GEN, and the constraint sets). Sections 5.5 and 5.6 discuss the resulting grammars of the learners. Section 5.7 discusses the results of a control group, followed by a general discussion of the results in section 5.8 and concluding remarks in section 5.9.

---

[37] I thank two anonymous reviewers of a paper version of this chapter originally submitted to Lingua. This chapter greatly benefited from their comments. Needless to say, any remaining errors are my own.

5.2 Pintupi stress

Pintupi (Western Australia; Hansen & Hansen 1969) was chosen as the target language for the computer simulations because of its very regular and predictable stress pattern. Primary stress is on the first syllable in a word and secondary stress is on every other following syllable except if that syllable is final in the word.[38] Syllables can have the shape CV, CVC, CVV, or CVVC, where 'C' stands for a consonant, 'V' for a vowel, and 'VV' for a long vowel. Pintupi has a phonemic vowel length distinction, restricted to the initial syllable of a word. According to Hayes (1991, 1995) and Kager (1992), Pintupi has a bimoraic word minimum. This indicates that Pintupi is a mora-counting language where long vowels are linked to two moras and short vowels are linked to one mora. Pleading for a weight-sensitive stress assignment in Pintupi is the fact that long vowels only occur word-initially; however, the fact that secondary stress is assigned on every other syllable, and not on the syllable directly following the primarily stressed one, is taken here as the cue that stress is not sensitive to weight.

Traditionally, a stress pattern like the one in Pintupi is analyzed with syllabic trochees iterating from left to right, starting at the left word edge (Hayes 1995, Kager 1999). Final syllables in words with an odd number of syllables are unfooted; this could be due to a prohibition of degenerate feet (feet that contain only one mora or syllable; Prince 1980, Hayes 1995:102).

Some examples are listed in (104) with the corresponding syllable and foot structure. The first column lists some examples of Pintupi in an overt-form fashion. These overt forms (displayed in square brackets) include primary stress ('v́', where 'v' stands for a vowel), secondary stress ('v̀'), syllable boundaries ('.'), and vowel length ('ː'). In the second column the overt forms have been interpreted in terms of hidden metrical structure. These surface forms contain foot structure as well as syllable boundaries ('.'), the head syllable of the head foot ('v́'), and the head syllable of a non-head foot ('v̀'). Long vowels are interpreted as 'vv'.[39]

---

[38] Auditory cues for primary stress in Pintupi are loudness, often along with higher pitch and greater duration of syllables; the auditory cue for secondary stress is slightly increased loudness (Hansen & Hansen 1969).

[39] I chose for a representation of small letters for CV structure in order to be able to indicate stresses in a readable form.

(104) Stress and foot structure in Pintupi

Overt forms	Surface forms	
a. [tʲáː]	/(cv́v)/	‘mouth’
b. [mú.ŋu]	/(cv́.cv)/	‘orphan’
c. [múː.ŋu]	/(cv́v.cv)/	‘fly’
d. [tʲán.pa]	/(cv́c.cv)/	‘evil spirit’
e. [ká.pa.li]	/(cv́.cv) cv/	‘mother’s mother’
f. [míːlʲ.ma.nu]	/(cv́v) cv.cv/	‘whining’
g. [ŋál.ku.nìn.pa]	/(cv́c.cv)(cv̀c.cv)/	‘eating’
h. [pú.ɭiŋ.kà.la.tʲu]	/(cv́.cvc)(cv̀.cv) cv/	‘we (sat) on the hill’
i. [tʲá.mu.lìm.pa.tʲùŋ.ku]	/(cv́.cv)(cv̀c.cv)(cv̀c.cv)/	‘our relation’

5.2.1 Pintupi stress: two possible OT accounts

Analysing the Pintupi stress pattern in Optimality Theoretic terms[40] requires basically the same constraints as in (14) of section 2.5: constraints on foot structure (PARSE, FTBIN), on foot form (TROCHAIC, IAMBIC), and alignment constraints (AFL/AFR, MAIN-L/R), listed in (105).

(105) Constraints

AFL/AFR: The left/right edge of a foot is aligned with the left/right edge of a word.
FTBIN: Feet are either bimoraic or disyllabic.
IAMBIC: The rightmost syllable in a foot is the head syllable.
MAIN-L/R: The head foot is aligned with the left/right edge of the word.
PARSE: Every syllable is included in a foot.
TROCHAIC: The leftmost syllable in a foot is the head syllable.

---

[40] The present OT analysis of Pintupi stress differs slightly from the one by Kager (1999:300) because I include constraints on foot form (TROCHAIC and IAMBIC), and constraints for weight-sensitivity. Kager’s account of the learnability of Pintupi stress (1999:301ff.) involves learning from surface forms instead of learning from overt forms. As outlined in section 3.7, I consider this as informed learning. Since it does not reflect a realistic learning situation I do not want to perpetuate it and rather model the learnability of Pintupi stress from overt forms only.

Let us evaluate the ranking for Pintupi stress step by step, starting with a trisyllabic word *kápali* 'mother's mother', as in tableau (106). The underlying form as the input to the evaluation is given in the upper left cell. TROCHAIC is assigned a violation mark by right-headed binary feet, e.g. */(ka.pá) li/, while IAMBIC is assigned a violation mark by left-headed binary feet, e.g. */(ká.pa) li/. Since the language has a strong-weak pattern, TROCHAIC has to outrank IAMBIC; otherwise the grammar would render */(kapá) li/ instead of /(kápa) li/.

(106)*kapali*: TROCHAIC >> IAMBIC

| ka.pa.li |	TROCHAIC	IAMBIC
☞ a. /(ká.pa) li/		*
b. /(ka.pá) li/	*!	

Tableau (107) shows that FTBIN has to be ranked above PARSE. Otherwise the grammar would yield */(kápa)(lì)/ or */(ká)(pà)(lì)/ or /(ká.pa.li)/[41] instead of /(kápa) li/. PARSE is assigned a violation mark once for every syllable not included in a foot, e.g. */ka (pá.li)/. FTBIN is assigned a violation mark by feet with only one mora, e.g. in the final foot of */(ká.pa)(lì)/.

(107)*kapali*: FTBIN >> PARSE

| ka.pa.li |	FTBIN	PARSE
☞ a. /(ká.pa) li/		*
b. /(ká.pa)(lì)/	*!	
c. /(ká)(pà)(lì)/	*!**	
d. /(ká.pa.li)/	*	

Turning to quadrisyllabic words like *ŋalkuninpa* in (108), we can establish the ranking of MAIN-L >> MAIN-R, to make sure that the left foot in a word is the head foot. MAIN-L and MAIN-R are assigned one violation mark for every syllable between the head foot and the respective word edge.

---

[41] Feet of the size of three syllables or more are banned from GEN in the simulations in section 5.4.

MAIN-L is therefore violated twice in */(ŋàl.ku)(nín.pa)/; MAIN-R is violated twice in /(ŋál.ku)(nìn.pa)/.

(108) *ŋalkuninpa*: MAIN-L >> MAIN-R

| ŋal.ku.nin.pa |	MAIN-L	MAIN-R
☞ a. / (ŋál.ku)(nìn.pa) /		**
b. / (ŋàl.ku)(nín.pa) /	*!*	

PARSE has to be ranked above AFL and AFR in order to allow more than one foot in the word, since AFL and AFR not only cause feet to be aligned with a word edge, but they also favour forms with as few feet as possible. If e.g. AFL outranked PARSE, */(ŋál.ku) nin.pa/ would surface instead of /(ŋál.ku)(nìn.pa)/.

(109) *ŋalkuninpa*: PARSE >> AFL, AFR

| ŋal.ku.nin.pa |	PARSE	AFL	AFR
a. /(ŋál.ku) nin.pa/	*!*		**
b. /ŋal.ku (nín.pa)/	*!*	**	
c. /ŋal (kú.nin) pa/	*!*	*	*
☞ d. /(ŋál.ku)(nìn.pa)/		**	**
e. /(ŋál.ku)(nìn) pa/	*!	**	**

To ensure that feet are built from left to right, AFL has to outrank AFR. The reverse ranking would render */(pú.ɭiŋ) ka (là.tʲu)/[42] instead of /(pú.ɭiŋ)(kà.la) tʲu/. AFL and AFR are gradient and assigned a violation mark for every syllable that is between a foot and the designated word edge.

---

[42] This happens to be the stress pattern of Garawa (Furby 1974); Hayes 1995), which is the same as for Pintupi in even-numbered words, but deviates from Pintupi in odd-numbered words in that it skips the third syllable in footing: /(σ́σ) σ (σ̀σ)/. I thank an anonymous reviewer for pointing this out to me.

(110)*puɭiŋkalatʲu*: AFL >> AFR

| pu.ɭiŋ.ka.la.tʲu |	AFL	AFR
a. /pu (ɭíŋ.ka)(là.tʲu)/	***!*	**
☞ b. /(pú.ɭiŋ)(kà.la) tʲu/	**	****
c. /(pú.ɭiŋ) ka (là.tʲu)/	***!	***
d. /pu (ɭìŋ.ka)(lá.tʲu)/	***!*	**

One can argue whether trisyllabic forms in Pintupi are sufficient to determine the directionality of foot assignment, because main stress can be assigned independently in Pintupi. I argue that the learner can find evidence for the ranking between AFL and AFR in trisyllabic forms simply because it is a fact that AFL is *not* violated in the licit trisyllabic form /(ká.pa) li/, and that AFR *is*, as tableau (111) shows.

(111)*kapali*: AFL >> AFR

| ka.pa.li |	AFL	AFR
☞ a. /(ká.pa) li/		*
b. /ka (pá.li)/	*!	

This will have an effect on the ranking of the two constraints in the learning algorithms that I discuss. Sections 5.5 and 5.6 will show whether this evidence is sufficient for the virtual learners to detect the left-to-right directionality in Pintupi.

What one would not expect is that PARSE also has to outrank IAMBIC: if IAMBIC was ranked above PARSE it would kick out the licit candidate /(ŋál.ku)(nìn.pa)/ and leave the decision to the lower ranked constraint AFL, which would decide in favour of candidate */(ŋál.ku) nin.pa/. Note that candidate (104e) is ruled out by higher-ranked FTBIN: I assume for the moment that coda consonants are not moraic in Pintupi, rendering a syllable such as *-nin-* as monomoraic and therefore light. As shown in (107), monomoraic feet violate FTBIN.

(112) *ŋalkuninpa*: PARSE >> AFL, IAMBIC

| ŋal.ku.nin.pa |	FTBIN	PARSE	AFL	IAMBIC
a. /(ŋál.ku) nin.pa/		*!*		*
b. /ŋal.ku (nín.pa)/		*!*	**	*
c. /ŋal (kú.nin) pa/		*!*	*	*
☞ d. /(ŋál.ku)(nìn.pa)/			**	**
e. /(ŋál.ku)(nìn) pa/	*!	*	**	

So far, I established a suitable partial ranking of the constraints for words with two to five syllables:

(113) Partial rankings:

TROCHAIC, PARSE >> IAMBIC;
MAIN-L >> MAIN-R;
AFL >> AFR;
FTBIN >> PARSE

However, words with a long initial vowel, e.g. *miːlʲmanu* 'whining', cannot be accounted for with this ranking, because it cannot decide between two licit candidates /(míːlʲ.ma) nu/ and */(míːlʲ)(mà.nu)/. The first candidate should win under a syllabic-trochee analysis, while the second candidate should win under a moraic-trochee analysis. The constraint one would regard as competent for the necessary disambiguation, FTBIN, cannot decide between these two candidates because both /(míːlʲ.ma) nu/ and */(míːlʲ)(mà.nu)/ satisfy FTBIN, as shown in (114). Note that FTBIN does not prefer a disyllabic foot (σσ) over a monosyllabic, yet bimoraic foot ($\sigma_{\mu\mu}$), but is equally satisfied by both kinds of feet. In our ranking, the decision between these two candidates is passed on to PARSE, which decides in favour of the wrong candidate, */(míːlʲ)(mà.nu)/:

(114) Long initial vowel

| miːlʲ.ma.nu |	FTBIN	PARSE	AFL
☹ a. /(míːlʲ.ma) nu/		*!	
☞ b. /(míːlʲ)(mà.nu)/			*
| kiː.ki.mi.la.ɳu | 'to kick'			
☹ a. /(kíːki)(mìla) ɳu/		*!	**
☞ b. /(kíː)(kìmi)(làɳu)/			****

One could think of reversing the ranking between PARSE and AFL, but this would lead to problems with forms like *ŋalkuninpa* in (109) or *puɭiŋkalaʈʲu* in (110). This means that there is no ranking with these constraints that can account for the Pintupi pattern. The source of the problem might be the constraint TROCHAIC. In its current definition TROCHAIC does not decide between the two foot forms (σ́ σ) and (σ́). A possible solution could be to replace TROCHAIC with FTNONFIN (Tesar (1998; repeated from (14) of section 2.6.1), even though this constraint did not work as good as TROCHAIC did in the simulations of chapter 4:

(115) FTNONFIN: The foot head is not final in the foot.

In this form FTNONFIN punishes monosyllabic feet like (σ́), and favours /(míːlʲ.ma) nu/ over */(míːlʲ)(mà.nu)/. As a by-product, FTNONFIN is taking over the function of FTBIN (here, at least), therefore FTBIN is left out of the tableau in (116).[43] To be able to unfold its full power, FTNONFIN has to outrank PARSE:

(116) FTNONFIN instead of TROCHAIC

| miːlʲ.ma.nu |	FTNONFIN	PARSE	AFL
☞ a. /(míːlʲ.ma) nu/		*	
b. /(míːlʲ)(mà.nu)/	*!		*
| kiː.ki.mi.la.ɳu | ‘to kick’			
☞ a. /(kíːki)(mìla) ɳu/		*	**
b. /(kíː)(kìmi)(làɳu)/	*!		****

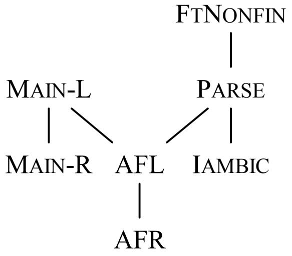
The crucial ranking is displayed in (117). MAIN-L outranks MAIN-R so that the left foot within a word gets main stress. MAIN-L furthermore outranks AFL and AFR. AFL outranks AFR to ensure that feet are iterated from left to right. FTNONFIN outranks PARSE to make sure that final syllables remain unfooted in words with an odd number of syllables and to make sure that in words beginning with a long vowel the first two syllables are parsed in a foot, as was shown in (116). FTNONFIN and PARSE outrank IAMBIC, because else iambic feet would surface (which is comparable to the

---

[43] It should be noted that FTBIN can still play a role in other languages that e.g. have an iambic stress pattern.

situation in tableau (106)) or words with only a single foot would surface, as shown in (112).

(117) A crucial ranking for Pintupi stress with FTNONFIN[44]



5.2.2 Maintaining an analysis with TROCHAIC

One could keep up an analysis with TROCHAIC by including *CLASH (Buckley 1998, Kager 1999, Kager 1992; this constraint traces back to pre-OT approaches by Liberman 1975, Liberman & Prince 1977, Prince 1983, Hammond 1984, and Selkirk 1984).[45] This constraint is commonly employed to prevent stresses in adjacent syllables, i.e. clashes:

(118) *CLASH: No stressed syllables are adjacent.

If this constraint is included dominating PARSE, the attested candidate becomes optimal, as shown in tableau (119).

---

[44] It should also be noted that even with FTNONFIN undominated, monosyllabic forms are guaranteed to surface through a high-ranking constraint LX≈PR (Prince & Smolensky 1993), which entails that content words should be pronounced.

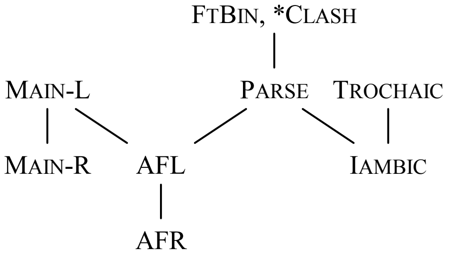
[45] One could also think of keeping up an analysis with TROCHAIC by splitting up FTBIN into FTBISYLLABIC ("feet are bisyllabic") and FTBIMORAIC ("feet are bimoraic") (similar to Kager 1993a, and later Hewitt 1994, who split up FTBIN into FTBIMORAIC$^{min/max}$, FTBISYLLABIC$^{min/max}$, and FTBINNUCLEARMORA$^{min/max}$. In Pintupi, FTBISYLLABIC would then have to outrank FTBIMORAIC, TROCHAIC, and PARSE, in order to ensure that feet are strictly bisyllabic. While the idea of incorporating FTBIMORAIC was pursued for the case of Latin, it is abandoned here in favour of an analysis with *CLASH, which is independently motivated in a series of other stress phenomena.

(119) An analysis with *CLASH and TROCHAIC

| miːlʲ.ma.nu |	TROCHAIC	FTBIN	*CLASH	PARSE
☞ a. /(míːlʲ.ma) nu/				*
b. /(míːlʲ)(mà.nu)/			*!	

In the crucial ranking including TROCHAIC in figure (120), *CLASH outranks PARSE. For the rest of the analysis, the same crucial rankings apply, as in (117). In a constraint set including FTNONFIN, *CLASH would not play a crucial role and could be placed anywhere in the hierarchy.

(120) A crucial ranking for Pintupi stress with TROCHAIC



By now I have established the analysis for Pintupi word stress with two different constraint sets, one including the constraint FTNONFIN in figure (117) and another with a constraint TROCHAIC in figure (120). It is interesting to see which constraint forms the appropriate restriction on trochaic feet by testing the acquisition process using computer simulations.

5.2.3 The weight of CVC syllables in Pintupi

While long vowels in languages with phonemic vowel length always have two moras (under moraic theory; Hyman 1985, Prince 1976, 1983, Hayes 1989, to name just a few), the moraic status of CVC syllables is not always clear. CVC syllables pattern qua weight with CVV syllables in some languages, while they pattern with CV syllables in others (e.g. Zec 1988). Pintupi is a case where it is not clear whether coda consonants are moraic or not. CVC syllables do not *attract* stress in the sense that CVC syllables are more often stressed than unstressed (which would indicate that they are moraic), but they happen to be stressed sometimes. Other evidence for or against moraicity of codas could come from monosyllabic words. Hansen & Hansen (1978:41) report that there are only very few monosyllabic words in Pintupi, and that they always consist of a CV syllable. They represent syllables with a short vowel and syllables with a long vowel both with CV syllable structure. The few examples Hansen & Hansen (1969) and Hansen & Hansen (1978) provide include monosyllabic function words with short vowels (e.g. *ma* 'direction marker') and even fewer examples of content words with long vowels (e.g. tjaː 'mouth'). Hayes (1991, 1995:103) and Kager (1992) inferred from the data that Pintupi has a bimoraic word minimum, and that the minimal word in Pintupi consists of either two syllables or a syllable with a long vowel. The fact that there are no monosyllabic words in Pintupi with a short vowel and a coda consonant indicates that coda consonants are not moraic, because if they were, monosyllabic words consisting of CVC syllables should be possible. This means that only negative evidence is available for a Pintupi-learning child that coda consonants are not moraic. Under the general assumption that learners can only learn from positive evidence, Pintupi-learning children do not know that there are no monosyllabic words consisting of CVC syllables. Their language has coda consonants, and syllables with coda consonants happen to *sometimes* be stressed, for instance in words like *ŋálkunìnpa*. The decision for moraic or non-moraic codas depends on the ranking of the constraints *Cμ (Broselow et al. 1997), WEIGHT-BY-POSITION (Hayes 1989; Sherer 1994), listed in (121), together with WSP. *Cμ militates against moraic coda consonants, while WBP militates against *non*-moraic coda consonants.

(121) Constraints on coda moraicity

*C_μ: Coda consonants are not moraic.
WEIGHT-BY-POSITION (WBP): Coda consonants are moraic.

If *Cμ outranks WBP, coda consonants are not moraic, and the ranking of WSP does not matter (ignoring long vowels for the moment). The decision in tableau (122) is between a candidate with moraic codas, represented with subscript 'μ', and a candidate with codas that are not moraic. The decision is made by high-ranked *Cμ.

(122) Codas not moraic

| ŋal.ku.nin.pa |	*Cμ	WBP	WSP
/(ŋál$_{\mu}$.ku)(nìn$_{\mu}$.pa)/	*!		
☞ /(ŋál.ku)(nìn.pa)/		*	

If, however, WBP outranks *Cμ, coda consonants are moraic. If WBP outranks both *Cμ and WSP, all codas surface as moraic:

(123) All codas moraic

| pu.ɭiŋ.kal.pi |	WBP	*Cμ	WSP
☞ /(pú.ɭiŋ$_{\mu}$)(kàl$_{\mu}$.pi)/		**	*
/(pú.ɭiŋ)(kàl$_{\mu}$.pi)/	*!	*	
/(pú.ɭiŋ)(kàl.pi)/	*!*		

If WSP outranks WBP, only codas in stressed syllables surface as moraic. This has an stress-to-weight effect (Myers 1987; Prince 1990), as already stated by Morén (2000):[46]

(124) Stress-to-weight effect

| pu.ɭiŋ.kal.pi |	WSP	WBP	*Cμ
/(pú.ɭiŋ$_{\mu}$)(kàl$_{\mu}$.pi)/	*!		**
☞ /(pú.ɭiŋ)(kàl$_{\mu}$.pi)/		*	*
/(pú.ɭiŋ)(kàl.pi)/		**!	

Equipped with the constraints *Cμ, WBP and WSP as part of the universal constraint set, and given *Freedom of Analysis* (Prince & Smolensky 1993:6, McCarthy & Prince 1993b:21), children can choose between representations with moraic codas, and representations with non-moraic codas. This is tested with the computer simulations in section 5.4.[47]

---

[46] One feels tempted to discard STRESS-TO-WEIGHT ("if stressed, then heavy"; e.g. Kager 1999) as an independent constraint at this point. However, the ranking of WSP >> WBP >> *Cμ gives the stress-to-weight effect only for coda consonants, and the effect shows not in the case long vowels are involved.

[47] The reason why I do not model the learnability of vowel length and stress here is that this would involve faithfulness to the underlying moraicity of vowels. The topic for this chapter is the learnability of grammatical stress without the involvement of faithfulness. Modelling stress that involves faithfulness is tackled in chapter 6.

In the remainder of this chapter, moraic codas are represented with subscripted ‘μ’ as in /(ŋál$_{\mu}$.ku)(nìn$_{\mu}$.pa)/, and non-moraic codas as /(ŋál.ku)(nìn.pa)/, without a subscript ‘μ’. I assume that vowels in Pintupi are always moraic, and chose not to mark them with ‘μ’ for readability reasons. We will see in section 5.5 whether my virtual learners found enough evidence in the training data to infer that coda consonants should not be moraic in Pintupi.

5.3 Modelling Pintupi stress

In section 5.2 I established a constraint ranking that accounts for Pintupi stress. Learnability provides a tool for testing the viability of a linguistic analysis. The learnability of metrical structure is intriguing, since the interpretation of stress in the speech stream regarding structure can be ambiguous. If a listener hears a trisyllabic form with stress on the middle syllable like [σ σ́ σ], she might interpret this form as having either an iambic foot /(σ σ́) σ/ or a trochaic foot /σ (σ́ σ)/. Adult speakers of a language ideally know how to interpret the form, based on their knowledge about whether the language is trochaic or iambic, but learners of a language do not know this yet.

The former OT-based simulations on the acquisition of metrical stress demonstrate the learnability from overt forms, i.e. forms that are marked for stress but not for foot structure. The learning data consisted of strings of light (‘L’) and heavy (‘H’) syllables, e.g. [L1 L L], [L H1 L], [H H1 L L2].[48] In Pintupi, stress can appear on syllables with coda consonants, giving room for the interpretation that these syllables are heavy and attract stress. But there are also syllables with coda consonants that are unstressed, indicating that these syllables are analyzed as being not heavy since they do not attract stress. It is interesting to see whether virtual learners of Pintupi analyze the language as completely weight-insensitive or as partly weight-sensitive, or even as completely weight-sensitive, because the long vowels in the language occur only word-initially and are always stressed. Long vowels, which are assumed to have two moras, do not shed light on the question

---

[48] The ‘1’s refer to primary stress, and the ‘2’s refer to secondary stress.

whether Pintupi stresses weight-sensitive or not, because they only occur in the first syllable of a word, which happens to be primarily stressed anyway.

Another question is whether the virtual learners acquire the Pintupi-like stress pattern, i.e. whether they will have primary stress on the initial syllable of the forms and secondary stress on every other following syllable, with the exception of final syllables in words with an odd number of syllables. Whether the learners will choose the foot structure shown in table (104) or whether they will assign feet differently is subordinate to the question whether stress is assigned to the correct syllable of the word. I gave the virtual learners a head start: they know that their language, Pintupi, has weight characteristics in that it has phonemic vowel length.

In section 5.2, two analyses of Pintupi stress were given. They differ in their interpretation of the constraint on trochaicity. To see whether there exists a difference in the learnability of these constraints one half of the virtual learners is equipped with FTNONFIN and the other half is equipped with TROCHAIC. If it turns out that e.g. the learners with FTNONFIN can learn the stress pattern more easily or can learn it at all, and the learners with TROCHAIC fail to learn the pattern it will indicate that TROCHAIC is not a good formulation of the constraint on trochaicity. The same holds for the reverse case where the learners with FTNONFIN fail and the ones with TROCHAIC succeed.

A further issue is how uniform the grammars of the learners are in the final state. If the learners come up with different constraint rankings, they could come up with different foot structure as compared to the ones established in §5.2. It is also interesting to see to what extent the learners are able to transfer the learned stress pattern to forms that they have not encountered up to that point. Finally, one half of the learners has CD (Tesar 1995) as the reranking strategy, while the other half has the GLA (Boersma 1997) to compare the learning performances of both strategies.

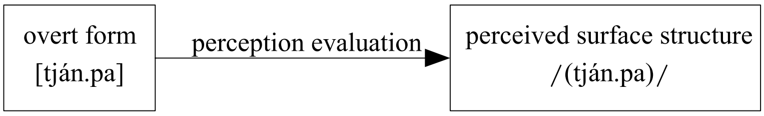
The simulations are carried out with the Praat programme (Boersma & Weenink 1992-2006). The virtual learners created by the programme have to find out for themselves whether coda consonants are moraic or not, and whether stress is sensitive to the weight of a syllable.

The upcoming sections outline the perception process in 5.3.1, the production process in 5.3.2, and the reranking strategies CD (5.3.3) and GLA (5.3.4). The ingredients to the computer simulations are given in section 5.4, followed by the results in section 5.5. The generalizations to forms that the learners have not been trained on are given in section 5.6. The results of a control group are given in section 5.7. A discussion of the results and the conclusions are given in sections 5.8. and 5.9.

5.3.1 Perceiving Pintupi stress

Perception in Pintupi works essentially the same as already demonstrated in sections 3.2 and 4.5, but for the purpose of testing the learnability of coda moraicity I unravelled the L/H syllable structure we saw in the approach to the learnability of Latin stress. The overt forms no longer consist of words with L/H syllable structure, but of words with CV-, CVV, and CVC-syllables. As before, syllable boundaries are given, with the consequence that the learners do not have to find out whether a word-medial consonant belongs to the coda of the preceding syllable or to the onset of the following. Primary stress is indicated by ‘´’, secondary stress by ‘`’. Stresses are given in the overt form, foot structure has to be imposed by the listener.

(125) Perception of Pintupi stress



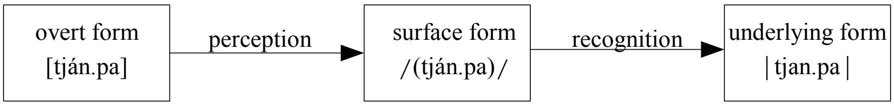
The tableau evaluation of this overt form is shown in (126) with an overt form [tʲán.pa]. Consider a Universal Grammar that consists of the constraints PARSE, AFL, AFR, FTBIN, IAMBIC, TROCHAIC, WSP and NONFIN (NONFIN as defined in (14) of section 2.6.1). Imagine that a listener has the grammar in tableau (126). The listener assigns foot structure to the incoming form [tʲán.pa] by applying her grammar. Since the overt form already contains stress, the possibilities to assign foot structure are limited, and only two candidates are considered for evaluation. Candidate (117a), /(tʲán.pa)/, has a disyllabic foot, while candidate (117b), /(tʲán) pa/, has a monosyllabic foot. I assume here that there cannot be a candidate like /(tʲan.pá)/ since stress would be on a different syllable. The decision for candidate /(tʲán.pa)/, as the optimal one (marked with a ‘ ☞ ’ in the tableau) is taken by high-ranked PARSE. Candidate /(tʲán) pa/ violates PARSE, because the final syllable is not footed. Thus, the listener has interpreted the overt form [tʲán.pa] as having the structure /(tʲán.pa)/.

(126) Perception: stress to foot

[t^{j}án.pa]	PARSE	AFL	AFR	FTBIN	IAMBIC	TROCHAIC
👂 a. /(t^{j}án.pa)/					*	
b. /(t^{j}án) pa/	*!		*	*		

Now that the listener has determined the surface structure /(t^{j}án.pa)/ of the overt input [t^{j}án.pa] she maps this perceived form onto a form in her lexicon. This means for the purposes of this chapter that the surface form is trivially mapped onto the underlying form by stripping off all metrical structure like feet and stress marks, leaving only the segmental and syllabic structure |t^{j}an.pa|.[49] The whole process of perception and recognition is comprehension.

(127) Comprehension



As outlined in §3.2 to 3.6, the comprehension process is the same for adults and language learning children, implying that the learner already knows the underlying form. This will do for the moment, because we deal with grammatically assigned stress only at the moment. However, this is not a realistic learning situation: there are languages that have lexically determined stress, i.e. where stress specifications in the lexicon interact with the grammar (e.g. in Modern Greek of chapter 6). A language learner does not know initially whether the target language has grammatical or lexical stress, and will have to learn the underlying forms as well. Chapter 6 gives an account for the case where the underlying form itself is not given, but has to be created by the learner. For now, it will suffice to assume that the learner knows the underlying forms. In the following section we will see how the learner is able to evaluate her own production by applying her current grammar.

---

[49] In real life the acquisition of metrical structure interacts with the acquisition of segmental and syllabic structure and with the creation of lexical entries. Once again these issues are left out of consideration since the discussion here is limited to the acquisition of metrical structure alone.

5.3.2 Virtually producing Pintupi stress

The previous section showed how the learner perceives the overt form [tʲán.pa] as having the surface structure /(tʲán.pa)/. From listening alone the learner will not arrive at an adult-like grammar. She needs to compare her perception to her production in order to be able to learn. Each time she encounters an incoming form, she will compute what she herself would say. In the production evaluation in tableau (128) are more candidates than in the perception evaluation, because the learner has to add foot structure *and* stress to the form.[50] Candidate (119a), /(tʲán.pa)/, is the perceived form of (126) and is marked with '👂'. This perceived form has trochaic foot structure, and is ruled out by IAMBIC. The form that the learner would produce, /(tʲan.pá)/, (marked by '👄') satisfies IAMBIC.

(128) Virtual production: foot to stress

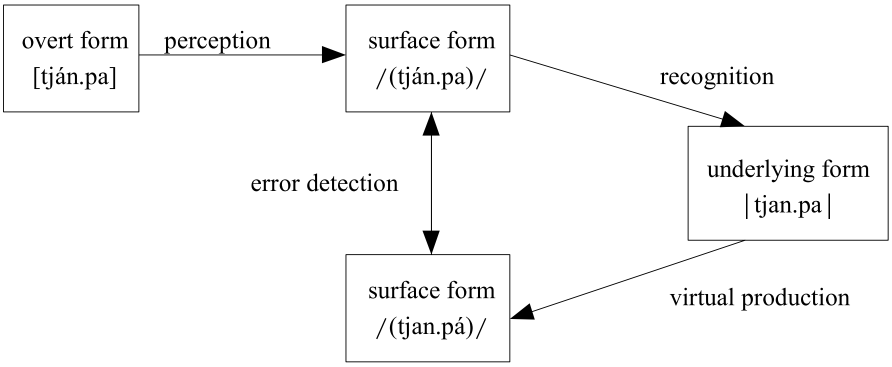
| tʲan.pa |	PARSE	AFL	AFR	FTBIN	IAMBIC	TROCHAIC
👂 a. /(tʲán.pa)/					*!	
b. /(tʲán) pa/	*!		*	*		
c. /(tʲán)(pà)/		*!	*	**		
d. /tʲan (pá)/	*!	*		*		
👄 e. /(tʲan.pá)/						*
f. /(tʲàn)(pá)/		*!	*	**		

Now that the learner computed what she would produce for the form she perceived, she can compare the two forms and notice an error between them. This discrepancy between the perceived form and the produced form impels the learner to change her grammar, until her perceived surface form and her produced surface form are identical.

---

[50] GEN dictates many more candidates than the ones shown here, but for clarity's sake I stick to just a few of them.

(129) Error detection



Two strategies for constraint reranking are applied in the computer simulations here: Constraint Demotion (Tesar 1995) and the Gradual Learning Algorithm (Boersma 1997). They are outlined in sections 5.3.3 and 5.3.4.

5.3.3 Learning with Constraint Demotion

As mentioned in section 3.4, the learner regards the perceived form as the target she wants to match her production to. When using Constraint Demotion (henceforth CD) as the learning strategy she will adjust her constraint ranking by looking up all constraints that prefer her produced form and lower them below the highest ranked constraint that prefers the form she perceives. This will make it more likely that in a future evaluation of this form, the produced form matches the perceived form. The CD makes use of crucial ties: the violations of all constraints within one stratum are summed up, and the candidate with the least violations in that stratum is the most harmonic one.

In our example learning tableau in (130), IAMBIC is the constraint preferring the learner’s produced form /(tʲan.pá)/, and TROCHAIC is the constraint preferring the learner’s perceived form /(tʲán.pa)/. To make it more likely that the perceived form /(tʲán.pa)/ will also be the winner of the production evaluation, IAMBIC is demoted below TROCHAIC, into a new stratum:

(130) CD at work in Pintupi

| tʲan.pa |	Parse	AFL	AFR	FtBin	Iambic	Trochaic	
👂 a. /(tʲán.pa)/					*!		
b. /(tʲán) pa/	*!		*	*			
c. /(tʲán)(pà)/		*!	*	**			
d. /tʲan (pá)/	*!	*		*			
👄 e. /(tʲan.pá)/						*	
f. /(tʲàn)(pá)/		*!	*	**			

The next time the learner encounters the form [tʲán.pa], she will still perceive /(tʲán.pa)/, even with the new ranking:

(131) Perception with the new ranking

[tʲán.pa]	Parse	AFL	AFR	FtBin	Trochaic	Iambic
👂 a. /(tʲán.pa)/						*
b. /(tʲán) pa/	*!		*	*		

When computing her production anew, she no longer detects an error:

(132) Production anew

| tʲan.pa |	Parse	AFL	AFR	FtBin	Trochaic	Iambic
👄 👂 a. /(tʲán.pa)/						*
b. /(tʲán) pa/	*!		*	*		
c. /(tʲán)(pà)/		*!	*	**		
d. /tʲan (pá)/	*!	*		*		
e. /(tʲan.pá)/					*!	
f. /(tʲàn)(pá)/		*!	*	**		

In our example, the perceived and produced form have been brought to agreement now; the produced form is identical to the perceived form: /(tʲán.pa)/.

When this happens with all the forms that the learner encounters, learning is terminated. However, it is possible that the reranking of constraints leads to a different interpretation of the same form, which in turn potentially leads to a new mismatch of the interpreted and produced form. The intermediate encounter of other forms might change the constraint ranking in a way that interpretation and production are out of tune again. In general, though, the encounter of different forms should help the learner to come up with a ranking that creates forms matching the adult output.

5.3.4 Learning with the Gradual Learning Algorithm

While CD makes use of an ordinal ranking of constraints, the Gradual Learning Algorithm (henceforth) GLA makes use of Stochastic OT (Boersma 1998), as outlined in section 3.6.2. Constraints are assigned real numbers (*ranking values*) on the ranking scale as a measurement of the distance between constraints. In each evaluation of a given form, a little bit of noise is added to the ranking value of each constraint, with the consequence that constraints close to each other can swap their order for this specific evaluation. In addition to that, grammar adjustment in the GLA is a bit different. In the GLA, all constraints that prefer (i.e. not violated in) the perceived form (IAMBIC) and all constraints that prefer the produced form (TROCHAIC) are looked up. Consider the same grammar as in tableau (128), with the same constraints and ranking, repeated in (133). After realizing that the surface form in perception and the surface form in production do not match, i.e. after detecting an error, the learner adjusts her grammar.

(133)Grammar adjustment with the GLA

| tʲan.pa |	PARSE	AFL	AFR	FTBIN	IAMBIC	TROCHAIC
👂 a. /(tʲán.pa)/					*!→	
b. /(tʲán) pa/	*!		*	*		
c. /(tʲán)(pà)/		*!	*	**		
d. /tʲan (pá)/	*!	*		*		
👄 e. /(tʲan.pá)/						←*
f. /(tʲàn)(pá)/		*!	*	**		

The constraints preferring the perceived form are shifted upwards, while the constraints preferring the produced form are shifted down the hierarchy by a specified amount. Irrespectively of TROCHAIC, IAMBIC is lowered on the ranking scale; irrespectively of IAMBIC, TROCHAIC is shifted up on the ranking scale. The dashed line in (133) reads as a beginning overlap of the two constraints, and the arrows indicate the direction the constraints take ('→' means downwards, and '←' means upwards the constraint hierarchy).

After some learning has taken place, i.e. after the learner encountered more forms like [tʲán.pa], TROCHAIC will swap places with IAMBIC, bringing perception and production into agreement, as shown in (134).

(134) An adjusted GLA-grammar

| tʲan.pa |	PARSE	AFL	AFR	FTBIN	TROCHAIC	IAMBIC
👄 👂 a. /(tʲán.pa)/						*
b. /(tʲán) pa/	*!		*			
c. /(tʲán)(pà)/		*!	*	*		
d. /tʲan (pá)/	*!	*		*		
e. /(tʲan.pá)/					*!	
f. /(tʲàn)(pá)/		*!	*	*		

In the simulations, the two learning strategies are compared with respect to their overall success in learning from the forms they are confronted with, but also with respect to their success in transferring what they have learned to forms they were not trained on.

5.4 Simulating the acquisition of Pintupi stress

The ingredients to the computer simulations are as before training data (5.4.1), a candidate generator (5.4.2), a set of constraints (the add-ons for Pintupi stress are given in 5.4.3), and the reranking strategies CD and GLA. The learners encountered the training data in a randomized order. The learners started out with an initial ranking where all constraints were ranked at the same heights (at 100.000). The CD learners had a plasticity of 1, meaning that constraints were reranked by 1 (e.g. first to 99 on the ranking scale, then to 98 etc.). They learned with zero evaluation noise and encountered 1 000 forms. The GLA learners had decreasing plasticity, starting out by 1, with four times a decrement of 0.1.[51] This means that in the

---

[51] 'Decreasing plasticity' means that the learners shifted constraints by 1.0 points on the ranking scale during the encounter of the first 10 000 data, by 0.1 points during encountering the second 10 000 data, by 0.01 points during encountering the third 10 000 data, and by 0.001 points the last 10 000 data, meaning that the learners took

beginning, the GLA learners took ranking steps as big as the CD learners (rather large ones), but decelerated their learning pace in the course of time. They were fed with 40 000 training forms.

5.4.1 The training data

The learning data set consists of 17 word-like forms, listed in (135). These forms are two to four syllables long and are made up of syllables and stress marks. Monosyllabic forms are excluded from the set since there is only one possibility to stress a monosyllabic word, therefore the virtual child cannot learn much about stress placement from it. The length of words is limited to at most four syllables for the following reasons: to account for the claim that child-directed speech are often simplified utterances (Phillips 1973) and to see what the learners will do when they are asked to produce forms that they have not been trained on. It was also noted by Hansen & Hansen (1969:162) that words consisting of two to four syllables are more frequent than monosyllabic words or words with more than four syllables. The learning set covers only a selection of all possible combinations of syllables in Pintupi.[52] All forms are overt, i.e. they contain stress marks, but neither foot nor moraic structure. Because I focus on the acquisition of stress here, the overt forms furthermore contained syllable boundaries.[53] The learners will know that each vowel in a syllable is inherently moraic, but they will not know that of coda consonants. GEN contains candidates with or without moraic coda consonants, and I expect the learners to decide for either one possibility or the other in response to the data they are trained on.

---

smaller and smaller learning steps. This is to approach a realistic learning curve; in the beginning, learning takes place rather fast, while it slows down in the course of time.

[52] I chose only forms whose syllable structure matched the examples in the Hansen & Hansen (1969) paper for the learning data set.

[53] In fact, children also have to learn to set syllable boundaries as well as what kind of syllable structure their language allows, e.g. whether their language allows for codas and/or consonant clusters etc. This issue is beyond the scope of this dissertation and will have to be modelled elsewhere.

(135) The training data

2-syllable forms	3-syllable forms	4-syllable forms
[cv́.cv]	[cv́.cv.cv]	[cv́.cv.cv̀.cv]
[cv́.cvc]	[cv́.cv.cvc]	[cv́.cvc.cv̀.cv]
[cv́v.cv]	[cv́.cvc.cv]	[cv́.cvc.cv̀c.cvc]
[cv́c.cv]	[cv́c.cv.cv]	[cv́c.cv.cv̀.cv]
[cv́c.cvc]	[cv́c.cv.cvc]	[cv́c.cv.cv̀c.cv]
	[cv́c.cvc.cv]	[cv́c.cvc.cv̀c.cv]

5.4.2 The candidate generator

As we have seen in (135), the training data consist of overt forms with two, three, or four syllables. The syllables can have the shape of CV, CVV, or CVC; syllable boundaries, primary and secondary stress are given. This holds also for the candidates provided by GEN: the candidates are surface forms that consist of corresponding strings of CV-, CVV-, and CVC-syllables. Syllable boundaries are indicated. The candidates contain foot structures and stress marks for primary and secondary stress on various positions within the word. Feet may span over one or two syllables, but never over more than two syllables. In the case that there are coda consonants in a word, the candidate list will contain candidates with moraic representations, e.g. the candidate list for a word like |cvc.cvc| may contain candidates such as /(cv́c) cvc/, /(cv́c$_{\mu}$.cvc$_{\mu}$)/, /(cv́c$_{\mu}$.cvc)/, and the like.

5.4.3 The constraints

The constraints used for the simulations on Pintupi stress consist of the set in (14) of section 2.6.1 with some add-ons. To be able to model the learning data that contain syllable structure instead of syllables already labelled for light or heavy, *Cμ and WBP are added. To maintain an analysis with TROCHAIC, *CLASH and *LAPSE are added (*LAPSE as in Elenbaas & Kager 1999 is violated in forms that have more than two consecutive unstressed syllables, and is added as a counterpart to *CLASH). Once more I will test two different constraint sets, that only differ in their implementation of the trochaicity constraint:

(136) Constraint sets for the simulations

1) TROCHAIC learners	2) FTNONFINAL learners
AFL/AFR	AFL/AFR
FTBIN	FTBIN
TROCHAIC	FTNONFIN
IAMBIC	IAMBIC
MAIN-L/R	MAIN-L/R
NONFIN	NONFIN
PARSE	PARSE
WFL/WFR	WFL/WFR
WSP	WSP
*Cμ	*Cμ
WBP	WBP
*CLASH	*CLASH
*LAPSE	*LAPSE

The list for general metrical phenomena is complemented, but remains far from complete. My interest was to let the learners decide whether coda consonants are moraic or not; this constraint set enables them to do so. In the initial state all constraints are ranked equal.

This results in four different types of learners. Since each learner encounters the data in a different order, variation in the results is anticipated (as we have seen in chapter 4 on Latin stress). To check whether differences in the learning results will arise, 50 learners of each learning type were created, resulting in a total of 200 virtual learners:

(137) Learning types:

	TROCHAIC set	FTNONFIN set
CD learning strategy	50 learners	50 learners
GLA learning strategy	50 learners	50 learners

Apart from the rankings established in (117) and (120) it is expected that the constraints *LAPSE and WFL/WFR show little effect in the outcome, since an analysis of Pintupi stress is not depending on them. WSP should not show an effect because stress assignment should be weight-insensitive. *CLASH and FTBIN should only play a role in learners that have the TROCHAIC constraint set; as I have indicated above (section 5.2.2), FTNONFIN could in principle take over some of the functions of *CLASH and FTBIN. One can argue whether it should, though.

5.5 Results for Pintupi stress

The virtual learners with the CD strategy learned from tableaux as described in (126), (128), and (130), while the virtual learners with the GLA strategy learned from tableaux as described in (126), (128), and (133). The results were taken from the learners after their training on 1 000 (for the CD learners) and 40 000 forms (for the GLA learners) respectively. All 200 learners produced the same overt output, listed in table (138). Learning can be considered successful because primary stress is correctly on the first syllable in all forms and secondary stress is on the third syllable in forms with four syllables. The forms that occurred in the training data are printed in bold; all the other forms had to be created by the learners themselves.

(138) The overt output of all 200 virtual learners

disyllables	trisyllables	quadrisyllables	
[cv́.cv]	**[cv́.cv.cv]**	**[cv́.cv.cv̀.cv]**	[cv́v.cvc.cv̀.cv]
[cv́.cvc]	**[cv́.cv.cvc]**	[cv́.cv.cv̀.cvc]	[cv́v.cvc.cv̀.cvc]
[cv́v.cv]	**[cv́.cvc.cv]**	[cv́.cv.cv̀c.cv]	[cv́v.cvc.cv̀c.cv]
[cv́v.cvc]	[cv́.cvc.cvc]	[cv́.cv.cv̀c.cvc]	[cv́v.cvc.cv̀c.cvc]
[cv́c.cv]	[cv́v.cv.cv]	**[cv́.cvc.cv̀.cv]**	**[cv́c.cv.cv̀.cv]**
[cv́c.cv]	[cv́v.cv.cvc]	[cv́.cvc.cv̀.cvc]	[cv́c.cv.cv̀.cvc]
	[cv́v.cvc.cv]	[cv́.cvc.cv̀c.cv]	**[cv́c.cv.cv̀c.cv]**
	[cv́v.cvc.cvc]	**[cv́.cvc.cv̀c.cvc]**	[cv́c.cv.cv̀c.cvc]
	[cv́c.cv.cv]	[cv́v.cv.cv̀.cv]	[cv́c.cvc.cv̀.cv]
	[cv́c.cv.cvc]	[cv́v.cv.cv̀.cvc]	[cv́c.cvc.cv̀.cvc]
	[cv́c.cvc.cv]	[cv́v.cv.cv̀c.cv]	**[cv́c.cvc.cv̀c.cv]**
	[cv́c.cvc.cvc]	[cv́v.cv.cv̀c.cvc]	[cv́c.cvc.cv̀c.cvc]

However, the learners came up with five different ways of assigning foot structures that resulted in the correct overt stress pattern. An overview is given in table (139). The numbers in the last but one column refer to the number of learners that came up with the analysis at hand, the numbers in the last column give the percentages.

The 103 learners in table (139) came up with the foot structure proposed by linguists (“linguist’s analysis”). Examples for this foot structure are /(cv́c.cv) cvc/ and /(cv́c.cv)(cv̀c.cv)/. The 5 GLA learners in (130b) analyzed coda consonants consistently as moraic, e.g. /(cv́c$_{\mu}$.cv) cvc$_{\mu}$/ and /(cv́c$_{\mu}$.cv)(cv̀c$_{\mu}$.cv)/. The 24 learners in (130c) analyzed only codas in stressed syllables as moraic, e.g. /(cv́c$_{\mu}$.cvc) cvc/ and /(cv́c$_{\mu}$.cvc)(cv̀c$_{\mu}$.cv)/. The 40 learners in (130d) analyzed final syllables consistently as being extrametrical, e.g. /(cv́c.cv) cvc/ and /(cv́c.cv)(cv̀c)cv/. Last but not least, 28 learners in (130e) analyzed codas as being moraic and final syllables as being extrametrical, e.g. as /(cv́c$_{\mu}$.cv) cvc$_{\mu}$/ and /(cv́c$_{\mu}$.cv)(cv̀c$_{\mu}$) cv/.

(139) Distribution of analyses

Analysis		GLA learners		CD learners		Total	%
		FTNONF	TROCH.	FTNONF	TROCH.		
a.	Linguist’s analysis:	23	3	30	47	103	51.5
b.	Moraic codas:	5				5	2.5
c.	Moraic codas/ stressed syllables:	1		20	3	24	12
d.	Extrametricality:	11	29			40	20
e.	Extrametricality/ moraic codas:[54]	10	18			28	14

The different analyses are discussed in sections 5.5.1-5.5.5. The variation is not only due to evaluation noise in the GLA learners, as we can see from the different analyses of the CD learners. Each learner encountered the data in a different order, and I consider that to be the reason for the variation in the results

The virtual learners invented five different ways to realize the Pintupi overt stress pattern. The GLA/FTNONFIN learners invented all five different ways. The learners with the GLA/TROCHAIC combination came up with three of these analyses, while the CD learners came up with two analyses.[55]

---

[54] One learner analysed final syllables as extrametrical and codas as moraic, but was not consistent in that within one form, some codas were moraic and some were not, independently of whether the codas occurred in stressed syllables. This learner is subsumed under category (130e).

[55] For the determination of the GLA learners’ outputs the evaluation noise was set to zero, meaning that the ranking they displayed on check-up after learning was frozen in as the final ranking.

5.5.1 The linguist's analysis

Back in section 5.2 I established the foot structure linguists would assign to the Pintupi stress pattern: disyllabic, left-headed feet assigned iteratively from left to right, unfooted final syllables in odd-numbered words, and coda consonants as not moraic. 103 learners came up with this pattern, as shown in (140):

(140) Examples for the foot structure in the linguist's analysis

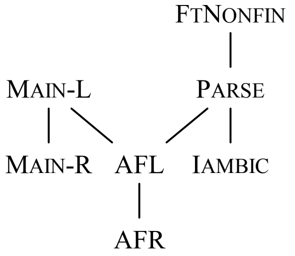
overt forms	surface forms
[cv́.cvc]	/(cv́.cvc)/
[cv́v.cv]	/(cv́v.cv)/
[cv́c.cv.cvc]	/(cv́c.cv) cvc/
[cv́v.cv.cv̀.cvc]	/(cv́v.cv)(cv̀.cvc)/

The grammars that the learners came up with to produce this pattern are discussed next. Section 5.5.1.1 discusses the FtNonfin learners, and section 5.5.1.2 discusses the Trochaic learners.

5.5.1.1 FtNonfin learners

23 GLA/FtNonfin learners and 30 CD/FtNonfin learners came up with a constraint ranking that assigned feet in the way described above. In order to be able to compare the rankings of the learners with that established in section 5.2, the ranking in figure (117) is repeated here as figure (141):

(141) A crucial ranking for an analysis of FtNonfin learners



GLA/FtNonfin learners. A final constraint ranking typical for a GLA/FtNonfin learner is shown in (142). For once, I show the complete hierarchy with ranking values, and will for the remainder of the chapter restrict myself to crucial differences in the learners' constraint rankings. Important is whether the crucial rankings are maintained: Main-L outranks Main-R and AFL; FtNonfin outranks Parse and Iambic. Parse in turn outranks Iambic and AFL, and AFL dominates AFR. *Cμ is ranked above WBP, therefore coda consonants do not surface as moraic.

(142) The ranking of a GLA/FtNonfin learner

FtNonfin	116.116
*Lapse	111.136
Main-L	110.243
WFL	108.154
FtBin	106.977
Parse	105.028
*Cμ	104.139
*Clash	104.089
Nonfin	101.947
AFL	101.028
WSP	100.915
WFR	98.053
WBP	95.861
AFR	95.017
Main-R	93.846
Iambic	81.759

We can evaluate this ranking with e.g. *ŋalkuninpa*. In tableau (143), the first three candidates contain a single foot and are ruled out. The two candidates that do not align this foot with the left word edge are ruled out because they violate high-ranking Main-L. The very first candidate is ruled out by lower-ranked Parse, because it has two unfooted syllables. Candidate (134e) is ruled out by top-ranked FtNonfin, because it contains a monosyllabic foot. Candidate (134f) is ruled out by Main-L; it has two proper trochaic feet, but the head foot is not aligned with the left word edge. That leaves candidate (134d) as the winner: it has two proper trochaic feet, and the head-foot is aligned with the left word edge. Neither does it violate any of the other high-ranking constraints which I left out of the tableau to focus on the ranking proposed in figure (141). For clarity's sake, I will only include the relevant constraints in the upcoming tableaux. For instance, the weight constraints are left out until effects of weight show up in the learners data.

(143) A GLA learner producing /(ŋál.ku)(nìn.pa)/

| ŋal.ku.nin.pa |	FtNonfin	Main-L	Parse	AFL	AFR	Main-R	Iambic
a. /(ŋál.ku) nin.pa/			*!*		**	**	*
b. /ŋal.ku (nîn.pa)/		*!*	**	**			*
c. /ŋal (kú.nin) pa/		*!	**	*	*	*	*
☞ d. /(ŋál.ku)(nìn.pa)/				**	**	**	**
e. /(ŋál.ku)(nìn) pa/	*!		*	**	***	**	*
f. /(ŋàl.ku)(nín.pa)/		*!*		**	**		**

In sum it can be said that the losing candidates were filtered out by the constraints which *should* have filtered them out, according to the crucial constraint ranking given in figure (117) (or in figure (141), respectively).

Next, I will discuss a case of a CD learner who came up with the same foot structure, yet a different ranking.

CD/FtNonfin learners. A constraint ranking typical for the CD/FtNonfin learners that came with said foot structure is shown in figure (144):

(144) The ranking of a CD/ FtNonfin learner

*Clash, *Lapse, AFR, FtBin, FtNonfin, Main-L, WFL, WSP
\>>
*Cμ, AFL, Iambic, Main-R, Nonfinal, Parse, WFR
\>>
WBP

We can see that the crucial ranking of Main-L over Main-R and AFL is accomplished, as well as the ranking of FtNonfin above Iambic and Parse. However, Parse is ranked on the same stratum as AFL and Iambic, and AFR even outranks Parse and AFL. Other learners that came up with this foot structure had the three constraints ranked on the same stratum. How come that the desired foot structure still shows? If we only pick out the constraints that we defined as being responsible for the foot structure described above, we do not end up with the desired candidate: in tableau (145), the desired candidate (marked with a '☹') is ruled out because it has more violations in the lower stratum than the competing candidates (136a), (136b) and (136c). The candidates (136e) and (136f) are ruled out because they have more violations than the others in the first stratum.

(145) A CD learner producing *ŋalkuninpa*

<table>
<tr><th>| ŋal.ku.nin.pa |</th><th>AFR</th><th>FtNonFin</th><th>Main-L</th><th>AFL</th><th>Iambic</th><th>Parse</th><th>Main-R</th></tr>
<tr><td>☞ a. / (ŋál.ku) nin.pa /</td><td>**</td><td></td><td></td><td></td><td>*</td><td>**</td><td>**</td></tr>
<tr><td>☞ b. / ŋal.ku (nín.pa) /</td><td></td><td></td><td>**</td><td>**</td><td>*</td><td>**</td><td></td></tr>
<tr><td>☞ c. / ŋal (kú.nin) pa /</td><td>*</td><td></td><td>*</td><td>*</td><td>*</td><td>**</td><td>*</td></tr>
<tr><td>☹ d. / (ŋál.ku)(nìn.pa) /</td><td>**</td><td></td><td></td><td>**</td><td>**</td><td></td><td>**!</td></tr>
<tr><td>e. / (ŋál.ku)(nìn) pa /</td><td>***!</td><td>*</td><td></td><td>**</td><td>*</td><td>*</td><td>**</td></tr>
<tr><td>f. / (ŋàl.ku)(nín.pa) /</td><td>**(!)</td><td></td><td>**(!)</td><td>**</td><td>**</td><td></td><td></td></tr>
</table>

It is the notion of crucial ties (e.g. Tesar & Smolensky 2000:38) which enables the CD learners to come up with the Pintupi stress pattern although the constraints are not totally ranked. If we take into account all constraints that were implemented in the learning process and that remained high-ranking, we get the correct candidate, as shown in tableau (146). The reason why /(ŋál.ku)(nìnpa)/ surfaces instead of e.g. */(ŋál.ku) nin.pa/, although AFR outranks Parse *and* AFL, is that /(ŋál.ku)(nìnpa)/ violates only AFR of all the constraints in that stratum (two violations in total), while */(ŋál.ku) nin.pa/ violates AFR and *Lapse (three violations in total), */ŋal.ku (nín.pa)/ violates Main-L and WFL (also three violations), and */ŋal (kú.nin) pa/ violates AFR, Main-L and WFL. These CD learners apparently did not find evidence in the data to demote AFR below AFL.

(146) A decision based on crucial ties

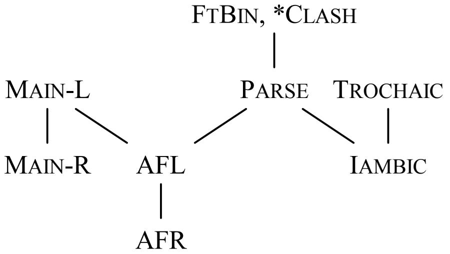
| ŋal.ku.nin.pa |	*LAPSE	AFR	FTBIN	FTNONFIN	MAIN-L	WFL	AFL	IAMBIC	MAIN-R	PARSE
a. /(ŋál.ku) nin.pa/	*(!)	**(!)						*	**	**
b. /ŋal.ku (nín.pa)/					**(!)	*(!)	**	*		**
c. /ŋal (kú.nin) pa/		*(!)			*(!)	*(!)	*	*	*	**
☞ d. /(ŋál.ku)(nìn.pa)/		**					**	**	**	

In sum, although AFR paradoxically outranks AFL, the wished-for foot structure could still surface thanks to the other constraints ranked on that stratum.

5.5.1.2 TROCHAIC learners

Three GLA/TROCHAIC learners and 47 CD/TROCHAIC learners came up with the linguist's analysis. The ranking of figure (120) is for convenience repeated as figure (147):

(147) A crucial ranking with TROCHAIC



First I discuss the GLA learners and then the CD learners.

GLA/TROCHAIC learners. In the grammar of a GLA learner shown in tableau (148), MAIN-L properly outranks MAIN-R and AFL. TROCHAIC outranks IAMBIC, while FTBIN and *CLASH outrank PARSE. PARSE in turn dominates AFL that dominates AFR. The numbers above the columns indicate the actual ranking values of the constraints in the end-grammar. None of the shown candidates violates TROCHAIC. MAIN-L rules out candidate (139b) with a right-aligned foot. FTBIN rules out candidate (139d), which has a monosyllabic foot in final position. *CLASH rules out candidate (139c), which has a binary (since bimoraic) foot. This leaves candidate (139a) as the winner: /(mí:l^j.ma) nu/ has a left-aligned, disyllabic foot.

(148) A GLA/Trochaic learner producing *mil*^{j}*manu*

	117.881	117.027	116.740	112.413	107.834	94.073	92.415	90.349	57.590
|mi:l^j.ma.nu|	TROCHAIC	MAIN-L	FTBIN	*CLASH	PARSE	MAIN-R	AFL	AFR	IAMBIC
☞ a. /(mí:l^j.ma) nu/					*	*		*	*
b. /mi:l^j (má.nu)/		*!			*		*		*
c. /(mí:l^j)(mà.nu)/				*!		**	*	**	*
d. /(mí:l^j.ma)(nú)/			*!			*	**	*	*

CD/TROCHAIC learners. In the CD/TROCHAIC learners, e.g. the showcase in tableau (149), the crucial rankings are not all borne out. MAIN-L is ranked above MAIN-R, *Cμ is ranked above WBP, and TROCHAIC is ranked above IAMBIC; but FTBIN, *CLASH, PARSE, AFL and AFR are ranked on the same stratum. Again, the exclusion of the competing candidates is taken care of by the crucial ties on the first stratum.

(149) An CD /TROCHAIC learner producing *mil*^{j}*manu*

|mi:l^j.ma.nu|	TROCHAIC	MAIN-L	FTBIN	*CLASH	PARSE	AFL	AFR	MAIN-R	IAMBIC
☞ e. /(mí:l^j.ma) nu/					*		*	*	*
f. /mi:l^j (má.nu)/		*(!)			*(!)	*(!)			*
g. /(mí:l^j)(mà.nu)/				*(!)		*(!)	**(!)	**	*
h. /(mí:l^j.ma)(nú)/			*(!)			**(!)	*(!)	*	*

5.5.1.3 Summary

The ranking of FTBIN over PARSE is not always maintained: some learners had these constraints reversed (some GLA/FTNONFIN learners), or ranked on the same stratum (some CD/TROCHAIC learners). High-ranking FTNONFIN will ensure that trochaic feet are disyllabic, and prohibit degenerate feet such as (cv́). In this way FTNONFIN takes over the function of FTBIN and even *CLASH, so the ranking between PARSE and FTBIN becomes irrelevant. This also explains why FTBIN and PARSE are often very close to each other in terms of ranking values across learners of all conditions. In the CD learners, it is due to the crucial ties that the required forms still surface.

So much for the learners that created foot structure like the one proposed by e.g. Hayes (1995) or Kager (1999) for Pintupi stress. In the next section, I discuss the analyses of the learners that interpreted coda consonants as moraic. Remember that their overt forms nevertheless show the Pintupi stress pattern.

5.5.2 Moraic coda consonants

Five of the 50 GLA/FTNONFIN learners analyzed coda consonants as being moraic in the surface form (regardless whether they occur in stressed syllables or not). None of the other types of learners came up with this analysis. The foot structure that these learners assign is the same as for the linguist's analysis, but coda consonants are consistently analyzed as being heavy, marked with a subscript 'μ' as illustrated in (150). Stress assignment in these forms is clearly weight-insensitive.

(150) Coda consonants analyzed as moraic

overt forms	surface forms
[cv́.cvc]	/(cv́.cvc$_\mu$)/
[cv́c.cvc.cv]	/(cv́c$_\mu$.cvc$_\mu$) cv/
[cv́.cvc.cv̀c.cv]	/(cv́.cvc$_\mu$)(cv̀c$_\mu$.cv)/

The ranking of WBP over *Cμ and WSP is responsible for the moraic analysis of these learners. The ranking of these constraints with respect to the others is shown in figure (151).

(151) A crucial ranking for moraic codas

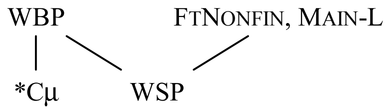


Tableau (152) demonstrates how a GLA/FtNonfin learner with such a ranking would produce a word like *ŋalkuninpa*. I only include four constraints and three candidates for the sake of clarity: FtNonfin filters out the last candidate */(ŋál$_{\mu}$)(kú.nin$_{\mu}$) pa/ because of its bimoraic foot on the first syllable. WBP filters out the candidate which surfaced as optimal in the tableaux (143) and (146). The candidate that has the same foot structure, but moraic coda consonants, surfaces as optimal. The numbers above the constraint columns are the actual ranking values of this learner and indicate the distances between the constraints.

(152) A GLA/FtNonfin learner producing *ŋalkuninpa*

	115.277	102.396	97.604	96.065
|ŋal.ku.nin.pa|	FtNonfin	WBP	*Cμ	WSP
/(ŋál.ku)(nìn.pa)/		*!*		
☞ /(ŋál$_{\mu}$.ku)(nìn$_{\mu}$.pa)/			**	
/(ŋál$_{\mu}$)(kú.nin$_{\mu}$) pa/	*!		**	*

In contrast to the learners with the linguist's analysis, these learners ranked WBP above *Cμ. Remember that WBP requires coda consonants to be moraic, while *Cμ militates against moraic codas. WSP, the constraint favouring stressed heavy syllables, is crucially ranked below the constraints FtNonfin and Main-L. A reverse ranking would bring about stress sensitivity to heavy syllables (i.e. syllables with coda consonants). These learners analyzed coda consonants as being always moraic, even in unstressed position. This is shown in tableau (153). The first candidate */(já.lin)(t^{j}à.ra)/ is ruled out by WBP. The second candidate /(já.lin$_{\mu}$)(t^{j}à.ra)/ wins because of its moraic coda.

(153) *jalint^{j}ara* ‘north’

| ja.lin.t^{j}a.ra |	WBP	*Cμ	WSP
/ (já.lin)(t^{j}à.ra) /	*!		
☞ / (já.lin$_\mu$)(t^{j}à.ra) /		*	*

The learners of this analysis assign stress clearly weight-insensitively, since the syllables with moraic codas do not attract stress. In the following section the learners are presented that analyzed codas only as moraic when they occur in a stressed position.

5.5.3 Moraic coda consonants in stressed syllables only

One GLA/FtNonfin learner, 20 CD/FtNonfin learners and 3 CD/Trochaic learners analyzed codas only in stressed syllables as moraic, as illustrated in table (154). Foot structure was perfectly disyllabic and trochaic, assigned from left to right, just like in the linguist’s analysis.

(154) Moraic codas in stressed syllables

overt forms	surface forms
[cv́c.cvc]	/(cv́c$_\mu$.cvc)/
[cv́c.cv.cvc]	/(cv́c$_\mu$.cv) cvc/
[cv́v.cvc.cv̀c.cv]	/(cv́v.cvc)(cv̀c$_\mu$.cv)/

These learners drew the conclusion that only codas in stressed positions are heavy. This effect smells like the ‘stress-to-weight’ principle (Myers 1987; Prince 1990), and has its cause in the ranking of WSP above WBP, which has to be ranked above *Cμ in turn, as argued in §5.2.3. The other constraints are ranked as in the other analyses:

(155) The crucial ranking for moraic codas in stressed syllables

WSP >> WBP >> *Cμ

We have already observed that only if WBP is ranked above *Cμ, coda consonants can be moraic. If WSP now outranks WBP and *Cμ, only coda consonants that occur in a stressed syllable are moraic, because WSP would cause codas in unstressed syllables not to be moraic.

Taking a GLA learners as a basis for the evaluation in tableau (156), we can see that a candidate where all codas are moraic is ruled out by WSP, while the candidate without moraic codas is ruled out by the second violation of WBP. The candidate with a moraic coda in stressed position violates WBP only once, and surfaces as optimal:

(156)*puɭiŋkalpi* ‘(he fell) finally at the hill’

	115.849	110.192	102.254	100.892	99.108
| pu.ɭiŋ.kal.pi |	FTNONF	MAIN-L	WSP	WBP	*Cμ
/(pú.ɭiŋμ)(kàlμ.pi)/			*!		**
☞ /(pú.ɭiŋ)(kàlμ.pi)/				*	*
/(pú.ɭiŋ)(kàl.pi)/				**!	

The CD learners show the same ranking of WSP >> WBP >> *Cμ.

5.5.4 Final syllable extrametricality

As established in section 5.2.1, final syllables in Pintupi are unfooted if the word has an odd number of syllables, because a monosyllable is too small to be parsed into a foot. This results in some forms with syllable extrametricality “by accident”, as it were, as in /(cv́v.cv) <cv>/. It might not be straightforward to analyze forms like that as being the result of the involvement of a constraint like NONFINAL. However, learners of the language could misinterpret those forms as an occurrence of extrametricality caused by NONFINAL. And indeed, eleven of the 50 GLA/FTNONFIN and 29 of the GLA/TROCHAIC learners analyzed *all* final syllables as being extrametrical, even at the cost of having degenerate feet, as in table (157).

(157) All final syllables extrametrical

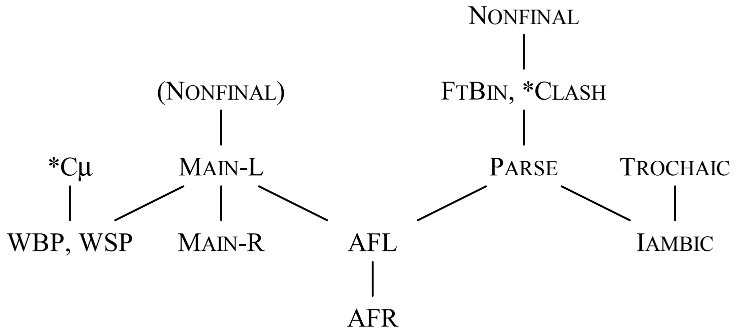
	overt forms	surface structures
a.	[cv́.cv]	/(cv́) cv/
b.	[cv́v.cv.cvc]	/(cv́v.cv) cvc/
c.	[cv́v.cv.cv̀.cvc]	/(cv́v.cv)(cv̀) cvc/

This is not surprising if one considers the Pintupi data and bears in mind that learners have the constraint NONFINAL to their disposal: the learners might interpret forms as having syllable extrametricality, because all words have unstressed final syllables. So some of the learners ended up with a ranking that put NONFINAL above FTNONFIN/TROCHAIC:

(158) A crucial ranking for final syllable extrametricality
NONFINAL >> { FTNONFIN, PARSE, FTBIN }

The ranking of MAIN-L >> MAIN-R, FTNONFIN >> IAMBIC, and PARSE >> AFL >> AFR is nevertheless borne out. *Cμ outranks WBP, so that coda consonants are not moraic.

(159) A crucial ranking for extrametricality with TROCHAIC



The effect of this ranking is exemplified in tableau (160). The candidate with disyllabic feet is ruled out by top-ranking NONFINAL, rendering the candidate with an extrametrical syllable as the winner:

(160) *ŋalkuninpa* 'eating'

	117.690	107.054	101.627	90.762
| ŋal.ku.nin.pa |	NONFINAL	PARSE	FTNONF	FTBIN
/ (ŋál.ku)(nìn.pa) /	*!			
☞ / (ŋál.ku)(nìn) pa /		*	*	*

None of the CD learners came up with an analysis like this. The answer is probably again to be found in the fact that CD learners learn with crucial ties. Candidates that satisfy NONFINAL are probably harmonically bounded, most of the time, and ruled out by the violations of the other constraints on that stratum, in the interpretive parse, seen in tableau (161), as well as in production, seen in tableau (162). FTBIN is violated in the candidates with monosyllabic feet, because codas are analyzed as not being moraic by these learners (candidates with moraic codas are left out).

(161) A CD learner interpreting *ŋalkuninpa*

[ŋál.ku.nìn.pa]	NONFINAL	PARSE	FTNONF	FTBIN
☞ /(ŋál.ku)(nìn.pa)/	*			
/(ŋál.ku)(nìn) pa/		*(!)	*(!)	*(!)

(162) A CD learner producing *ŋalkuninpa*

|ŋal.ku.nin.pa|	NONFINAL	PARSE	FTNONF	FTBIN
☞ /(ŋál.ku)(nìn.pa)/	*			
/(ŋál.ku)(nìn) pa/		*(!)	*(!)	*(!)
/(ŋál.ku) nin.pa/		**!		

A number of GLA learners combined the extrametricality and moraic coda analyses. This is shown in the next section. Since there are no CD learners that found the extrametricality analysis, no CD learners have found the combined analysis of extrametrical syllables and moraic codas.

5.5.5 Moraic codas and final syllable extrametricality

10 of the 50 GLA/FTNONFINAL learners analyzed coda consonants as being moraic and final syllables as extrametrical. This results in degenerate feet even in words with an even number of syllables:

(163) Moraic codas and syllable extrametricality

overt forms	surface forms
[cv́v.cvc]	/(cv́v) cvc$_\mu$/
[cv́c.cv]	/(cv́c$_\mu$) cv/
[cv́c.cvc.cvc]	/(cv́c$_\mu$.cvc$_\mu$) cvc$_\mu$/
[cv́.cvc.cv̀.cvc]	/(cv́.cvc$_\mu$)(cv̀) cvc$_\mu$/

Final syllable extrametricality is due to the ranking of NONFINAL above FTNONFIN, FTBIN, and PARSE. Moraic codas come along with the ranking of WBP above *Cμ and WSP.

While none of the GLA/TROCHAIC learners came up with an analysis of moraic codas, quite a number of them (17 learners) came up with an analysis that had both moraic codas and extrametrical syllables. As shown in tableau (164), NONFINAL is ruling out all candidates that have the final syllable footed. The decision between the candidates with extrametrical syllables is made by WBP, leaving the candidate with moraic codas as the optimal output.

(164) *puɭiŋkalpi* '(he fell) finally at the hill'

	114.347	105.539	103.617	101.062	98.938	96.217	94.256
| pu.ɭiŋ.kal.pi |	NONFIN	TROCHAIC	PARSE	WBP	*Cμ	FTBIN	WSP
/(pú.ɭiŋ$_\mu$)(kàl$_\mu$.pi)/	*!				**		*
/(pú.ɭiŋ)(kàl$_\mu$.pi)/	*!			*	*		
/(pú.ɭiŋ)(kàl.pi)/	*!			**			
☞ /(pú.ɭiŋ$_\mu$)(kàl$_\mu$) pi/			*		**		*
/(pú.ɭiŋ)(kàl$_\mu$) pi/			*	*!	*		
/(pú.ɭiŋ)(kàl) pi/			*	*!*		*	

The GLA/FTNONFINAL learners showed a similar ranking; if we replaced TROCHAIC with FTNONFIN in the tableau above, it would still render the same candidate as optimal.

5.5.6 Summary

To sum up shortly, all 200 virtual learners created grammars that describe the data they heard during the training phase. Nevertheless they created different grammars which is reflected in different surface structures. This does no harm, since all these surface structures translate to uniform overt outputs. Communication between these learners is guaranteed, because when talking to each other they wouldn't realize that their grammars differ. Let us now have a look at what the learners do when they have to abstract away from the familiar forms, and have to determine the stress pattern of words they have not been trained on.

5.6 Generalizations to unheard forms

In line with the tradition in computational linguistics (e.g. Manning & Schütze 1999:577) the virtual learners were asked to make generalizations, i.e. predict the stress pattern of words they were not trained on. This provides evidence for to what extent the learners are able to abstract away from the type of forms they heard in the training phase to a grammar accounting for the language. As outlined in 5.4.1, the learners have been trained on a set of 17 types of di- to quadrisyllabic words. After the training phase, the virtual learners were tested on what stress and foot structure they assign to forms they had not heard before. Among them were all di- to quadrisyllabic combinations of CV-, CVC- and CVV-syllables that are allowed in Pintupi, as listed in (138). The learners were then tested what stress and foot structure they assign to forms that contain long vowels in any syllable within the word, like [cv́.cvv.cvc] or [cv́.cv.cvv] (discussed in 5.6.1). Furthermore they were tested on what stress they assign to forms with more than four syllables (discussed in 5.6.2). The five-syllable forms consist of all possible combinations of syllable forms (forms like *puɭiŋkalatʲu*, but also non-attested forms of Pintupi with long vowels in non-initial position), and the six- and seven-syllable forms consist of CV-strings. 86 learners transferred the stress pattern they have been trained on to forms with more than four syllables and to forms with long vowels in other syllables than the initial one (these forms are actually not attested in Pintupi, but one could imagine a real-life scenario where speakers of Pintupi are confronted with loanwords with that kind of syllable structure). All of them were GLA learners. The CD learners mostly generalized to weight-sensitive forms. An overview of the results is given in (165):

(165) Summary of generalizations

Generalizations	GLA learners		CD learners		Total
	FtNonf	Trochaic	FtNonf	Trochaic	%
a. Linguist's analysis:	14	0	0	0	7.0 %
b. Moraic codas:	5	0	0	0	2.5 %
c. Moraic codas /stressed syllables only:	0	1	0	0	0.5 %
d. Extrametricality	10	29	0	0	19.5 %
e. Extrametricality/moraic codas:	10	17	0	0	13.5 %

= 43%

In total, 43% of all 200 learners came up with a Pintupi-like pattern; all of them GLA learners. Of the 50 GLA/FtNonfin learners, 78% (19.5% of all 200 learners) came up with a Pintupi-like pattern. Of the 50 GLA/Trochaic learners, 94% (23.5% of all 200 learners) came up with a Pintupi-like pattern.

5.6.1 Generalizations to unattested forms

Some examples are given in (166) for generalizations to forms with non-initial long vowels that the GLA learners produced:

(166) Generalizations to forms with non-initial long vowels

overt forms	surface structures
[cv́.cvv]	/(cv́.cvv)/ or /(cv́) cvv/
[cv́c.cvv]	/(cv́c.cvv)/ or /(cv́c) cvv/
[cv́.cv.cvv]	/(cv́.cv) cvv/
[cv́.cvv.cvc]	/(cv́.cvv) cvc/
[cv́.cvc.cvv]	/(cv́.cvc) cvv/
[cv́v.cv.cvv]	/(cv́v.cv) cvv/
[cv́.cvv.cv̀.cvc]	/(cv́.cvv)(cv̀.cvc)/ or /(cv́.cvv)(cv̀) cvc/
[cv́.cvc. cv̀.cvv]	/(cv́.cvc)(cv̀.cvv)/ or /(cv́.cvc)(cv̀) cvv/

The overt forms that these learners produced had primary stress on the initial syllable, and secondary stress on every other following syllable (except when that syllable was final in the word). It would be interesting to see whether real Pintupi-speakers would stress loanwords that contain long vowels in non-initial position like that.

The constraint ranking responsible for the weight-insensitive treatment of forms with non-initial long vowels is Main-L and FtNonfin above WSP in the group of FtNonfin learners and Main-L and *Clash above WSP in the group of the Trochaic learners. Main-L >> WSP guarantees that the foot with main stress will be aligned to the left edge of the word, so that a heavy syllable cannot attract stress away from the edge. FtNonfin >> WSP prevents stress clashes. In the Trochaic group, *Clash takes care of that.

Consider a learner of the GLA/FtNonfin-group that was able to generalize to this weight-insensitive pattern. In tableau (167) we can see for the first time an effect of WFL, ruling out a candidate */cv (cv́v.cv) cvc/ with stress on the heavy syllable. Even without this constraint, the same candidate would be ruled out by Parse. The optimal candidate is (158a), /(cv́.cvv)(cv̀.cvc)/, which has not a moraic coda. Its direct competitor, */(cv́.cvv)(cv̀) cvc/, is ruled out by *Cμ, which is ranked above WBP.

(167) Generalization to unattested forms

| cv.cvv.cv.cvc |	FtNonfin	Main-L	FtBin	WFL	Parse	*Cμ	WSP	WBP
☞ a. /(cv́.cvv)(cv̀.cvc)/							*	*
b. /(cv́.cvv)(cv̀) cvc/	*!		*		*		*	*
c. /(cv́.cvv)(cv̀.cvc$_\mu$)/						*!	**	
e. /cv (cv́v)(cv̀.cvc)/	*!			*	*			*
g. /cv (cv́v.cv) cvc/				*!	**			*
h. /cv (cv́v)(cv̀.cvc$_\mu$)/	*!			*	*	*	*	

5.6.2 Generalizations to longer forms

The learners were also tested what stress pattern they would assign when asked to produce words with more than four syllables. Some examples for

Pintupi-like generalizations to forms with five to seven syllables are given in (168):

(168) Generalizations to five-syllable forms and longer

overt forms	surface forms
[cv́.cvv.cv̀v.cv.cvc]	/(cv́.cvv)(cv̀v.cv) cvc/
[cv́.cvv.cv̀.cvv.cvc]	/(cv́.cvv)(cv̀.cvv) cvc/
[cv́c.cvv.cv̀v.cvc.cvv]	/(cv́c.cvv)(cv̀v.cvc) cvv/
[cv́.cv.cv̀.cv.cv̀.cv]	/(cv́.cv)(cv̀.cv)(cv̀.cv)/ or /(cv́.cv)(cv̀.cv)(cv̀) cv/
[cv́.cv.cv̀.cv.cv̀.cv.cv]	/(cv́.cv)(cv̀.cv)(cv̀.cv) cv/

Consider the grammar of a GLA/Trochaic learner with such a pattern. The form with six syllables has an extrametrical syllable due to high-ranking NonFinal, but stress is nonetheless Pintupi-like: the first syllable has primary stress, the third and fifth syllables have secondary stress. Main-L makes sure that stress is aligned with the left word edge, while Parse makes sure that there are three feet in the word (the numbers in the cells indicate the number of constraint violation):

(169) Generalization to forms with six syllables

| cv.cv.cv.cv.cv.cv |	NonFin	Main-L	*Lapse	Trochaic	WFL	Parse	FtBin	AFL	AFR
☞ a. /(cv́.cv)(cv̀.cv)(cv̀) cv/						*	*	6	7
b. /(cv́.cv)(cv̀.cv)(cv̀.cv)/	*!							6	6
c. /cv (cv́.cv)(cv̀.cv) cv/		*!			*	**		4	4
d. /(cv́.cv) cv (cv̀.cv) cv/						**!		3	5
e. /(cv́.cv) cv.cv (cv̀.cv)/	*!		*			**		4	4

It can be concluded that most of the GLA learners found enough evidence in the data to rank AFL above AFR, resulting in a left-to-right directionality of foot assignment. The implication of this finding would be that when learning from words with up to four syllables, the language learners are able to generalize to a Pintupi-like pattern in words with five and more syllables, but not to a Garawa-like pattern.

5.6.3 Generalizations to other stress patterns

Of the 200 learners, 57% generalized to patterns quite different from the weight-insensitive, left-aligned pattern proposed for Pintupi. Among them were all CD learners. Many of them displayed a strong tendency for weight-sensitivity in forms with long vowels in non-initial position and produced forms like /(cv.cv́v) cv (cv̀v.cvv)/. From the data they were trained on they did not infer the ranking of MAIN-L, *CLASH and FTNONFIN above WSP, that would be crucial for a weight-insensitive analysis. Once again, the reason are the crucial ties. One could say that CD learners would show a different stress pattern for loanwords than most of the GLA learners. Which one is the correct pattern attested by real Pintupi speakers is yet to be shown.

Another difference was the alignment of feet. Many of the longer forms contained feet that were not properly stringed together, but left out syllables. This comes about with the equal ranking of AFL and AFR. AFL would have to outrank AFR in order to properly align the feet. Learners with an equal ranking of the two constraints therefore produced forms like /(cv́.cv) cv (cv̀.cv).cv/ and even forms with iambic feet for the form as /(cv.cv́) cv.cv.cv.cv.cv/, with seven syllables. The iamb could occur because the constraints in the first stratum could not evaluate an optimal candidate, and the decision was left to the lower ranked IAMBIC constraint. However, only one single CD learner came up with a Garawa-like pattern (where primary stress is assigned to the initial syllable, but secondary stress is iteratively assigned from the right). Most of the other CD learners showed a strong tendency to weight-sensitivity. This could mean that there is a learning path to Pintupi stress when learning from two- to four syllable forms, but not to Garawa.[56]

A general reason for this deviation could be that any of the ingredients in this modelling of stress is deficient. The constraints might be poor descriptions of their function, the learning algorithms could be wrong, the training set could have been too impoverished or OT as a theory of learning

---

[56] It is likely that it was only accidental that one learner out of 200 found the Garawa pattern.

could be inadequate. That all of the learners came up with an overt stress pattern that is Pintupi-like suggests that the ingredients are quite sufficient, though.

5.7 The control group: learning from polysyllabic forms

One point of discussion is whether the learners would be able to create a more uniform foot structure when trained on a more complete set of data. I therefore tested the four types of learners (10 GLA/FTNONFIN learners, 10 GLA/TROCHAIC learners, 10 CD/FTNONFIN learners, 10 CD/TROCHAIC learners) on a bigger training set consisting of two- to five-syllable forms with all combinations of syllable structures that are licit in Pintupi, plus six- and seven-syllable forms consisting of CV-syllables only. The results are summed up in table (170). Almost all CD learners (47,5% of all the 40 learners of the control group) came up with left-aligned, disyllabic feet as in (161a), while only 5% of the GLA learners chose this analysis. Most of the GLA learners (35% of all the 40 learners of the control group) chose to analyze all final syllables as extrametrical, as in (161d). Some GLA learners (10%) chose to analyze the Pintupi pattern as having final syllables extrametrical and codas as moraic (161e). One CD learner (2.5%) chose to analyze codas in stressed syllables as being moraic (161c). No learner of the control group analyzed the Pintupi pattern as having left-aligned, disyllabic feet with moraic codas (161b).

(170) Resulting foot structures when learning from complete data

Analysis		GLA learners		CD learners		Total	%
		FTNONF	TROCH.	FTNONF	TROCH.		
a.	Linguist’s analysis:	1	1	9	10	21	52.5
b.	Moraic codas:						
c.	Moraic codas/ stressed syllables:			1		1	2.5
d.	Extrametricality:	7	7			14	35.0
e.	Extrametricality/ moraic codas:	2	2			4	10.0

The results roughly correspond to the analyses of the tested group in (139): the majority of the test group chose for the left-aligned, disyllabic pattern (51.5% vs. 52.5% in the control group), while the next biggest group chose for the extrametrical analysis (20% vs. 35% in the control group). A smaller percentage of the test group chose for the combined extrametrical/moraic-coda analysis (14% vs. 10% in the control group). An even smaller percentage chose for the moraic-codas-in-stressed-syllables-only analysis (12% vs. 2.5% in the control group). In the test group, 2.5% learners came up with the moraic-codas-everywhere analysis, while none of the learners in the control group came up with this analysis. This gap is probably accidental; if I ran a bigger control group, some learners might pop up analyzing codas as being always moraic. As for the distinction between GLA and CD learners, we also find a rough correspondence to the distribution in (139). None of the CD learners in the test group came up with the moraic-coda analysis, or with the extrametrical analysis, or with the combined extrametrical/moraic-coda analysis, and none of the CD learners of the control group came up with any of those analyses. Therefore I conclude that the imperfect training data shown in (135) for the test group were sufficient information to deduce a Pintupi adult-like grammar from.

5.8 Discussion

In the simulations of this chapter, all learners acquired the stress pattern in the sense that they produced stress on the correct syllable within a word, i.e. their overt production was the same as the overt forms in the target language. Despite the fact that the overt forms were the same, the learners came up with different analyses, though. More than half of the learners (all in all 103) came up with an analysis similar to the one linguists have come up with for Pintupi stress (e.g. Hayes 1995, Kager 1999). Five learners analysed coda consonants as moraic, but apart from that assigned the same foot structure as a linguist would assign it. Twenty-four learners treated only stressed codas as moraic, nonetheless assigning disyllabic feet from left to right. These three groups of learners can be clustered together as one group in terms of foot structure, resulting in a total of 132 learners that came up not only with the desired stress pattern but also with the desired foot structure. The remaining learners assigned a different foot structure in that they always left final syllables unfooted. Some of them assigned this foot structure in combination with moraic codas. Stress assignment was nevertheless Pintupi-like and weight-insensitive.

There are several reasons why different analyses were possible. One is the fact that the learners encountered the same data, but in a different order. Depending on which forms you encounter a lot in the beginning, your perception changes to the extent your grammar changes. This applies even more for the GLA learners, since they learned with a plasticity decrement, i.e. they took bigger learning steps in the beginning and were slowing down over time.

Moreover, the data that the learners encountered do not give explicit evidence as to whether codas in Pintupi are moraic or not, so some learners interpreted codas as being moraic (WBP >> *Cμ), while others did not (*Cμ >> WBP). Evidence for/against syllable extrametricality in Pintupi is not unambiguous, either. The data in Pintupi do not show explicit evidence for or against the footing of final syllables. Syllable extrametricality comes about with a ranking of NONFINAL above FTNONFIN and FTBIN.

A further reason for the variation in analyses lies in the different characters of the learners. The GLA/FTNONFIN learners came up with five different analyses. The GLA/TROCHAIC-group came up with three analyses (a subset of the ones that the GLA/FTNONFIN learners came up with), while all the CD learners came up with two analyses (again a subset). This suggests that the kind of learning strategy has a bigger impact in the resulting variation than the difference in constraints.

While all of the learners were able to assign stress correctly to forms of two to four syllables, they did not uniformly transfer this stress pattern to forms that they have not been trained on. Eighty-six of the GLA learners transferred the stress pattern they applied to di- to quadrisyllabic forms to longer forms (forms with five to seven syllables) and to forms that are not attested in Pintupi (forms with long vowels in non-initial position). None of the CD learners applied this stress pattern to unheard forms. Most of them tended to stress long vowels that occurred anywhere in the word, i.e. their grammar showed a tendency towards weight-sensitivity. These generalizations could occur because the data that the learners have been trained on do not give enough evidence for a complete weight-insensitive analysis of Pintupi, since primary stress is always on the initial syllable of a word, and long vowels only occur in initial position and are therefore always stressed.

5.9 Conclusions

With respect to the different constraint sets, neither FTNONFIN nor TROCHAIC can be excluded as the constraint on trochaic feet. TROCHAIC learners came up with less different analyses than FTNONFIN learners.

Another effect became apparent in the simulations of acquisition here. An extensive list of constraints can complicate learning (as could be seen in the simulations with FTBIMORAIC in chapter 4), but it can also facilitate learning: one constraint not directly applying to the phenomenon can substitute the effect of another constraint crucial for the analysis: an analysis using TROCHAIC needs FTBIN and *CLASH to be higher-ranking so that disyllabic feet can surface. An analysis with FTNONFIN can take over the role of FTBIN: if FTNONFIN is ranked above PARSE feet are naturally disyllabic. Final syllables are unfooted not because they are too small, but because they would violate high-ranking FTNONFIN. This applies only to languages with trochaic feet; in iambic languages, FTBIN might still make a difference. It looks like OT constraints in their present form overlap to a certain extent, resulting in redundancy.

Simulations on the learnability of languages in an OT framework gives linguists the possibility of testing claims made in cross-linguistic research and the study of child language acquisition. Learnability limits the amount of e.g. possible stress systems in a different way than factorial typology does. While factorial typology is the set of all possible rankings of constraints that result in different languages, learnability limits the typology of all possible languages by restricting the range to the constraint rankings that are learnable (see also Boersma 2003).

In sum it can be said that learners of one and the same language may not end up with exactly the same grammar. This partly meets claims made by e.g. Mohanan (1992) and Yip (2003), who propose that speakers may vary in their grammars, yielding slightly different overt outputs, and still be able to communicate. The virtual learners here were exposed to the same data tokens, with the only difference that they encountered the data in a different order, and ended up with five different analyses of hidden structures. Their overt output was the same, though, which means that not only communication is guaranteed, but also that the speakers still speak the same variety of their language.

6 The learnability of grammatical and lexical stress in Modern Greek

6.1 Introduction

This chapter shows how underlying forms are learned by an OT on-line learning algorithm. The proposed algorithm is “stupid”: it processes one form at a time under one grammar at a time. “Stupid” is good and effective: no extra learning mechanisms are required than the ones already involved in a general grammar learning model of OT (e.g. Boersma 1997). Interpretation of incoming forms and constraint reranking as a reaction to error detection is enough. Surface and underlying forms are connected via faithfulness constraints, and underlying forms are connected to meaning via lexical constraints. It is shown that the learner can acquire grammar and underlying forms concurrently, and creates an economical lexicon. This is exemplified with the learning of lexical stress. Lexical stress (in opposition to grammatically assigned stress) is information that is not predictable by the grammar (i.e. the constraint ranking of a language), and is therefore stored in the lexicon as some sort of underlying representation.

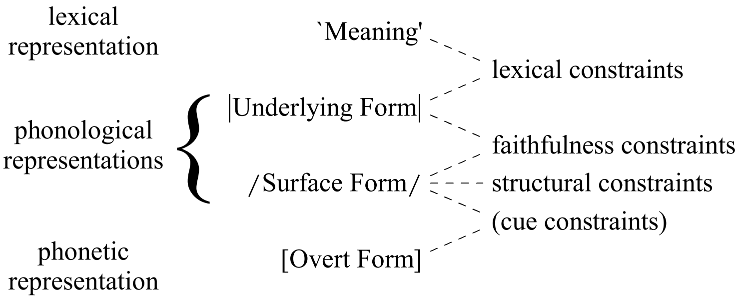
In the learnability approaches of chapters 4 and 5 it was not necessary to stipulate a learning mechanism for underlying forms, since stress was always assigned by structural principles, and was not influenced by the lexicon. Underlying forms were always trivially derivable from the surface form and merely had to be disengaged from metrical structure in order to be looked up in the lexicon. If a language learner knew in advance whether she had to learn a language with grammatically assigned stress that would be all there is to it. If it were the case that stress was always determined by the lexicon, the learning of underlying forms would be trivial as well: the learner would simply have to map surface forms faithfully onto underlying forms. However, determining underlying forms is not always straightforward. First of all, the learner does not know whether her language has grammatical or lexical stress, and second, the language can have a mix of both. The learner needs to be able to handle this. A case in point is lexical stress in Modern Greek, where stress is not predictable by grammatical principles alone, but is highly influenced by the lexicon. Morphemes can contrast in being underlyingly stressed or unstressed, and can even be specified for being preor post-stressing. Modern Greek stress therefore provides a challenging test ground for a learnability account of underlying forms; however, the model introduced here should be applicable to any other area where the learning of underlying forms is involved.

Proposals for the learning of underlying forms in OT are plenty, starting with e.g. Prince & Smolensky's Lexicon Optimization (1993; enhanced by Itô et al. 1995, and Tesar & Smolensky 2000), Tesar (2004, 2006), Tesar et al. (2003), or Jarosz (2006). A discussion of some of these alternatives is given in section 6.7. What they have in common is that they make use of paradigmatic information: the learner collects all the possible surface instantiations of a morpheme, and on the basis of the collected data adjusts the grammar (i.e. the constraint ranking). This kind of learning is called *off-line*: the gathered data are stored for further processing and can be accessed repeatedly in the course of learning. While the proposed algorithms for the learning of underlying forms are successful in what they are supposed to do, this kind of learning feels unnatural. When would a child know that she gathered enough information to be able to analyze the data? It is more natural to assume that a child processes a form the moment she is confronted with it, and discards the form immediately after processing. This is called *on-line* learning. By learning on-line, the child does not have to store any surface form. In the remainder of the chapter it is shown how such an on-line learning of underlying forms can be modelled in OT.

The proposed grammar model (as shown in figure (1) of chapter 1) is based on Boersma's grammar model (2005), but extended with an additional representational level 'Meaning', as also discussed in Boersma (2006ab). The idea is that we still have overt forms, surface forms, and underlying forms, but need an additional level of *meaning* to model the acquisition of underlying forms. Intuitively it means that you create a form only because you have a meaning that you want to express. Formally, the notion of meaning will help the learner to recognize identical morphemes, and reduce lexical allomorphy.

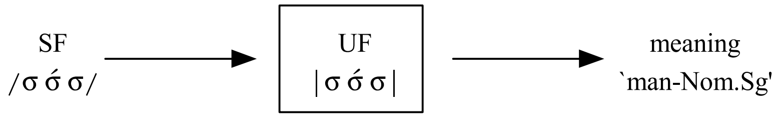
In this model, the different levels of representations are connected through different groups of constraints (shown in figure (171), which is repeated from figure (13) in chapter 2). In this chapter I will focus on all four levels of representation shown in (171): *meaning*, *underlying form*, *surface form* and *overt form*. The connection between overt form and surface form is expressed by *cue constraints* (Boersma 1998), which I ignore, because I model the relation between overt and surface forms with *structural* *constraints* alone, as in chapters 4 and 5. The connection between surface and underlying forms is expressed by *faithfulness constraints*. The connection between form and meaning is expressed by a new family of lexical constraints (Boersma 2001). The term 'meaning' as used here conflates both semantic and syntactic information.

(171) The grammar model:



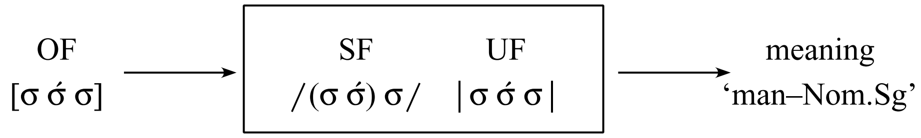
First, I provide an OT account of on-line learning of underlying forms where surface form and meaning are given, and the learner has to create the corresponding underlying form. The learner will deal with one level of hidden structure, as addressed in step 3 of the introduction and implemented in the learning approaches of surface forms in chapters 4 and 5. But instead of modelling hidden surface forms the learner will have to infer hidden underlying forms:

(172) Learning hidden underlying forms



As addressed in step 4 of the introduction and outlined in chapter 3, I push the proposal as far as to the learning of both surface and underlying form, given overt form and meaning. The learner then has to deal with two levels of hidden structure:

(173) Learning hidden surface forms and hidden underlying forms



The outline of the chapter is as follows: the next section provides a description and an analysis of lexical stress in Modern Greek. Section 6.3 outlines the learning model and its ingredients. The model is put to the test in computer simulations in sections 6.4, 6.5, and 6.6. Section 6.7 discusses some alternatives, and section 6.8 provides a general discussion.

6.2 Grammatical and lexical stress in Modern Greek

The learning of underlying forms is illustrated with a simplified version of Modern Greek stress. I will focus on the nominal paradigm, which will not weaken our claim made here, since nouns show all the stress patterns that are possible in Modern Greek. Modern Greek stress is weight-insensitive and vowels do not differ in phonological length.[57] Words in Modern Greek can be stressed on any of the last three syllables of a word (Joseph & Philippaki-Warburton 1987):

(174) A trisyllabic window for stress in Modern Greek

astráɣalos	'ankle-Nom.Sg.M'
ániksi	'spring-Nom.Sg.F'
mitéra	'mother-Nom.Sg.F'
mixaní	'machine-Nom.Sg.F'

---

[57] They might differ in phonetic duration, though. Stressed syllables tend to be phonetically longer than unstressed ones. In fact, stress in Modern Greek is mainly phonetically realized as a combination of duration and intensity (McKeever Dauer 1980; Botinis 1989). Nevertheless, there is no phonemic vowel length distinction.

Some words keep their stress throughout the paradigm:

(175) Fixed stress
γóndola ‘gondola-Nom.Sg’ vs. *γóndolon* ‘gondola-Gen.Pl’

Others shift stress:

(176) Stress shift
θálasa ‘sea-Nom.Sg’ vs. *θalasón* ‘sea-Gen.Pl’

The common view on Modern Greek is that stress is mainly assigned through specifications in the lexicon (e.g. Philippaki-Warburton 1970, 1976, Ralli 1988, Malikouti-Drachman & Drachman 1989, Touratzidis & Ralli 1992, Drachman & Malikouti-Drachman 1996, Revithiadou 1999). Roots as well as suffixes can contrast in stress. I follow Revithiadou (1999) in the classification that roots and suffixes can be unstressed or stressed. For computational reasons I simplify the specifications for morphemes that happen to push stress onto another morpheme: roots that are “unaccentable” under Revithiadou’s analysis will simply be specified as post-stressing, and suffixes that are specified for having a “weak accent” in her analysis will be specified for being pre-stressing. In Revithiadou’s account, structural constraints are responsible for unaccentable roots; in my account given here, these roots are subject to faithfulness constraints.

I first focus on the contrast between underlyingly stressed and unstressed morphemes. A word like *γóndola* ‘gondola-Nom.Sg’ retains stress on the root when inflected with the genitive plural suffix -*on*: *γóndolon*, as in example (175). But a word like *θálasa* ‘sea-Nom.Sg’ shifts stress to the suffix when inflected: *θalasón*, as in (176). The root of a word like *γóndola* is analyzed as being under-lyingly stressed |ɣóndol-|, and the root in a word like *θálasa* is analyzed as being underlyingly unstressed |θalas-|. The genitive plural suffix -*on* is underlyingly stressed, as becomes apparent when attached to an unstressed root like in the case of *θalasón*: only then can it surface as stressed. The nominative singular suffix -*a* is under-lyingly unstressed, as becomes apparent when combined with an unstressed root *θalas*-: then the phonological default stress on the antepenultimate syllable is assigned. In the case that both root and suffix are specified for stress, as in *γóndolon*, the root stress is preserved.

I will ignore foot structure for the moment (but will come back to it in section 6.6), and represent surface forms (SF) as in column a. of table (177). Underlying forms (UF) are represented as in column b. of table (177), and meaning as in column c. of table (177). The notion ‘meaning’ here refers thus to both the semantic content of a root (e.g. *γondol-* expresses the concept ‘gondola’) and to the syntactic content of a suffix (e.g. *-on* expresses case ‘genitive’ and number ‘singular’).

(177) A simple contrast[58]

a. surface forms	b. underlying forms	c. meaning
/ɣóndola/	|ɣóndol+a|	‘gondola-Nom.Sg’
/ɣóndolon/	|ɣóndol+ón|	‘gondola-Gen.Pl’
/θálasa/	|θalas+a|	‘sea-Nom.Sg’
/θalasón/	|θalas+ón|	‘sea-Gen.Pl’

We can distil four underlying morphemes from the table in (177): an underlyingly stressed root |ɣóndol-|[59], an unstressed root |θalas-|, a stressed suffix |-ón|, and an unstressed suffix |-a|. The learner is expected to eventually arrive at the four proposed morphemes, but cannot know in advance whether e.g. the root *γondol-* is specified for stress or not. The two-way contrast in morphemes means for the learner that she can choose between a stressed |γóndol-| and an unstressed |γondol-|. The same applies to every other morpheme, which gives our learner a pool of eight possible underlying forms to choose from:

(178) Possible underlying forms:

|ɣondol-|	|θalas-|	|-a|	|-on|
|ɣóndol-|	|θálas-|	|-á|	|-ón|

My goal is to establish only a basic analysis for Modern Greek stress that can easily be implemented into the computer simulations, because I want to focus on the learning of underlying forms, and not on a detailed analysis of Modern Greek stress. The reader is kindly asked to forgive me if I therefore

---

[58] The contrast described in this section is nothing else than the language in example (5) of the PAKA-world in Tesar et al. (2003:480).

[59] If I were to model the acquisition of stops in Modern Greek, I would analyze the coronal stop /d/ as underlyingly voiceless (Arvaniti 1999). Since this is beyond the scope of the dissertation, I stick to the voiced surface form.

leave out issues such as secondary stress (which is controversial in Modern Greek) and the influence of derivational suffixes (which draw stress away from roots; see Revithiadou 1999 for a detailed analysis, and Apoussidou 2003 for an alternative account). Section 6.2.1 provides an analysis of the phonological default stress. Section 6.2.2 provides an analysis of lexically stressed words.

6.2.1 Analyzing grammatical stress in Modern Greek

The default stress on the antepenultimate syllable (or on the penultimate syllable if the word consists of two syllables) applies when none of the morphemes of a word are underlyingly stressed. To account for the limitation of stress to the last three syllables of a word, and to make sure that stress will be on the antepenultimate syllable in a word without lexical stress, we need the following universal structural constraints (which is a limited choice from the constraints I used in chapters 4 and 5):

(179) Structural constraints for Modern Greek

AFL/AFR: Feet are aligned at the left/right word edge
FTBIN: Feet are disyllabic.
IAMBIC: Feet are right-headed.
NONFINAL: No foot is final in a prosodic word.
TROCHAIC: Feet are left-headed.

AFR needs to be ranked above AFL, because stress is not left-aligned:

(180) AFR >> AFL

|a.stra.γal+os|	AFR	AFL
☞ /a (strá.γa) los/	*	*
/(á.stra) γa.los/	**!	

NONFINAL has to outrank AFR, because the final syllable is extra-metrical:

(181) NONFINAL >> AFR

|θa.las+a|	NONFINAL	AFR
☞ /(θá.la) sa/		*
/θa (lá.sa)/	*!	

TROCHAIC outranks IAMBIC, because feet are left-headed:

(182) TROCHAIC >> IAMBIC

|θa.las+a|	TROCHAIC	IAMBIC
☞ /(θá.la) sa/		*
/(θa.lá) sa/	*!	

FTBIN outranks TROCHAIC, to rule out monosyllabic feet:

(183) FTBIN outranks TROCHAIC

|θa.las+a|	FTBIN	TROCHAIC
☞ /(θá.la) sa/		
/(θá) la.sa/	*!	

If we wrap it up in one tableau, it still works:

(184) NONFINAL >> AFR; FTBIN >> TROCHAIC >> IAMBIC

|θa.las+a|	NONFIN	AFR	AFL	FTBIN	TROCH	IAMBIC
☞ /(θá.la) sa/		*				*
/θa (lás.a)/	*!		*			*
/θa.la (sá)/	*!		**	*		
/(θa.lá) sa/		*			*!	
/θa (la.sá)/	*!		*		*	
/(θá) la.sa/		**!		*		

6.2.2 Analyzing lexical stress in Modern Greek

If any morpheme in a word has lexical (= underlying) stress, surfacing stress is determined by faithfulness, which outranks the structural constraints. Or rather, MAX(Stress) outranks the structural constraints; DEP(Stress) can be violated (as we can see from the default case of /(θá.la) sa/, where an accent is inserted on the root morpheme) and has to be ranked low. The high ranking of MAX(Stress) is illustrated in tableau (185), where I only include those structural constraints in the tableau that are high-ranked and violated. The first candidate is ruled out because it violates MAX(Stress): the underlying stress specification of the suffix is not realized on the suffix in the output form. The winning candidate violates a number of structural constraints because it puts a monosyllabic foot into final position, but it satisfies the highest-ranked MAX(Stress) constraint.

(185) MAX(Stress) outranks the structural constraints

|θa.las+ón|	MAX(Stress)	NONFIN	FTBIN
/(θá.la) son/	*!		
☞ /θa.la (són)/		*	*

When a lexically stressed root and a lexically stressed suffix are combined, only one of them can surface. The root retains its stress, as we can see from *γóndolon* in tableau (186). Therefore, the MAX constraint needs to be split up into MAX(Root) and MAX(Affix). Likewise, DEP is split up into DEP(Root) and DEP(Affix). The DEP constraints are ranked below the MAX constraints.

(186) MAX(Root) >> MAX(Affix)

|γón.dol+ón|	MAX(Root)	MAX(Affix)
☞ /(γón.do) lon/		*
/γon.do (lón)/	*!	

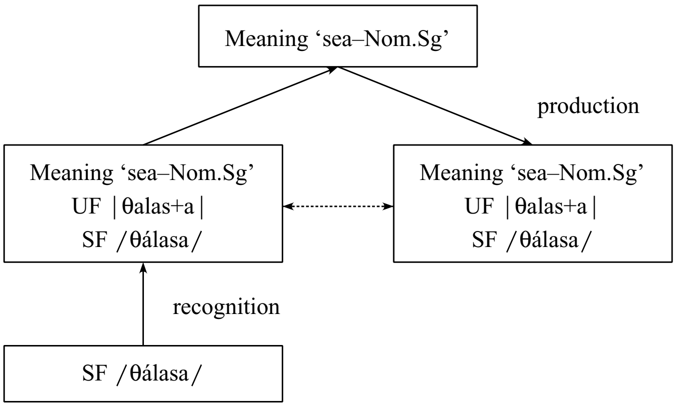
The next section describes the proposed model for the learning of underlying forms.

6.3 The model for the learning of underlying forms

In the proposed model the learner learns the underlying forms of a language by ranking the relevant constraints.

For the purposes of 6.4 and 6.5, I ignore foot structure. Thereby the surface forms (henceforth 'SF') as in column a. of table (177) are observable to the learner. The mapping between SFs and underlying forms ('UF's as in column b. of table (177)) is regulated by the faithfulness constraints listed in (16) of section 2.6.2, and the mapping between UFs and meaning (meaning as in column c. of table (177) above) is regulated by the lexical constraints discussed in 2.6.3. The learning process is error-driven like described in section 3.5.2, in the sense that a learner only changes her grammar if she detects a mismatch between the form she recognizes and the form she would produce for that word. In recognition, the SF serves as input to the OT evaluation, and the candidates are triplets of meaning, UF, and SF. In virtual production, i.e. the computation of the form that the learner would have produced herself, meaning serves as input to the evaluation, and the candidates are the same kind of triplets as in recognition. If the meaning/UF/SF triplet of the virtual production step is identical to the one in recognition, nothing is changed in the grammar. If there is a mismatch (i.e. an error is detected) the grammar (i.e. the constraint ranking) is changed.

(187) The processing model



I propose that at least two kinds of constraints are involved in the learning of underlying forms: faithfulness and lexical constraints. Faithfulness constraints are necessary to establish a correspondence between SF and UF, in this case faithfulness constraints on stress. The faithfulness constraints need to be split up into MAX and DEP since deletion of stress in the surface form (incurring a violation of MAX) as well as insertion of stress can occur (incurring a violation of DEP). MAX and DEP have to be furthermore split up into faithfulness for roots and faithfulness for affixes, as shown in tableau (186); in my simplified version of Modern Greek, faithfulness to the root outranks faithfulness to the affix. We have already seen these constraints in section 2.6.2. The following section goes into detail about lexical constraints.

6.3.1 Lexical constraints

Lexical constraints establish the link between UF and meaning. Boersma's (2001) approach to the phonology-semantics interaction in OT makes use of lexical constraints to distinguish between two homophonous forms with different meanings in comprehension, such as Dutch [rɑt] 'wheel' and [rɑt] 'rat'. The overt form [rɑt] 'wheel' has a voiced coda underlyingly, |rɑd|, while the overt form [rɑt] 'rat' is underlyingly |rɑt|. If both forms equally violate any structural constraints, and if the decision between the two forms is solely left to faithfulness constraints in comprehension, the form [rɑt] 'wheel' will never be chosen, because it incurs a faithfulness violation to the coda consonant, which is underlyingly voiced. The choice would always fall on [rɑt] 'rat', because this form does not incur a faithfulness violation. Boersma therefore proposed lexical access constraints militating against each possible underlying form in order to model the access of either [rɑt] 'wheel' or [rɑt] 'rat' depending on the ranking of these constraints and the semantic context in which the two forms are encountered.

Escudero (2005:220ff.) similarly modelled *message-driven learning in recognition* involving faithfulness constraints and lexical constraints. She proposed that the perception grammar of a (for instance, Dutch) speaker learning a new language (for instance, Spanish) can be adjusted by the influence of the semantic context in which a given form occurs. Initially, the second-language learner brings her native grammar and lexicon and applies them to the second language. In Escudero's account, learners of a new language can adjust their (already present) phonological categories with the help of semantic context. Learners learn from comprehension alone; the recognition process influences the perception process.

The situations in Boersma's (2001) and Escudero's (2005) approaches to the determination of underlying forms differ from the L1 learning situation assumed here. A child learning her first language does not initially have any phonological categories, as little as she has underlying forms. She is not (only) confronted with the problem of homonymy, but with the problem of determining underlying forms in the first place. Given the overt form that already includes stress, her recognition grammar cannot have any influence on her perception.[60] I therefore assume constraints against the

---

[60] In the case of acquiring hidden surface forms in a second language, Boersma's or Escudero's approach will not help, either: lexical stress is phonetically not distinguishable from grammatical stress. Their proposals work if one models the

creation of underlying forms that will enable the learner to establish the appropriate underlying forms for her language. The lexical constraints I employ are formulated as follows:

(188) Lexical constraints on underlying stress specifications:
Don't connect the meaning 'xy' to the form |XY| that is specified/unspecified for stress.

For each meaning (or morpheme, for that matter) several constraints are induced that militate against specified corresponding underlying forms. In the case at hand, the constraints militate against the connection of a meaning to an underlying form that is specified for being stressed or unstressed. For a form like *θalasón*, there are constraints on roots and affixes. Two constraints on the relation between a meaning and a root morpheme are given in (189):

(189) Lexical constraints on the root *θalas-*:
*|θalas-| 'sea': *Don't connect the meaning 'sea' to an unstressed root |θalas-|.*
*|θálas-| 'sea': *Don't connect the meaning 'sea' to a stressed root |θálas -|.*

Likewise, there are several constraints on the affix *-on*:

(190) Lexical constraints on the suffix *-on*:
*|-on| 'Gen.Pl': *Don't connect the meaning 'Gen.Pl' to an unstressed suffix |-on|.*
*|-ón| 'Gen.Pl': *Don't connect the meaning 'Gen.Pl' to a stressed suffix |-ón|.*

As already mentioned in section 2.6.3, these constraints are not necessarily innate, but could be induced by the learner. Whenever a learner encounters a new meaning and a new form, she invokes a constraint against the connection between the two. For this she needs to be equipped with the ability to invoke constraints against any possible underlying form. In lack of

---

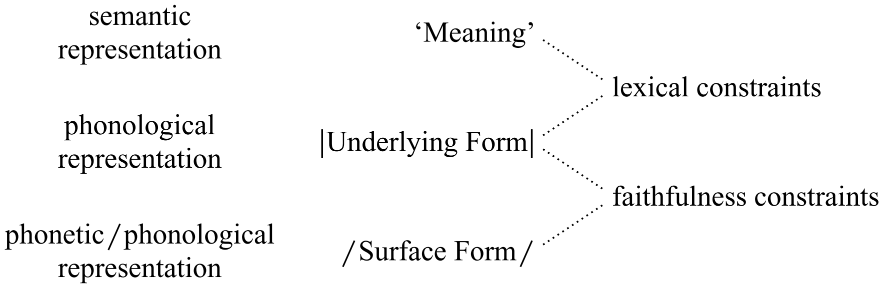
acquisition of the phonetic correlates for stress, or in the case of second language acquisition, when the phonetic correlates of the second language differ from the ones in the first language; then message-driven recognition learning can help to adjust the correlates for stress.

an algorithm that does so, the required constraints are implemented in the simulations.

6.3.2 Recognition

I first focus on the mapping of the surface form onto an underlying form and meaning here. This implies that overt forms are set aside for the moment. Figure (191) illustrates this part of the bigger model given in figure (171).

(191) One level of hidden structure



The recognition mapping proceeds as follows: the listener interprets an incoming SF as an UF by applying her grammar.[61] This means that the SF becomes the input to an OT evaluation, with triplets of meaning/UF/SF as candidates (provided by GEN). A ranking of faithfulness constraints in interaction with lexical constraints selects the optimal meaning/UF/SF triplet. In (192), the evaluation of the incoming SF /θalasón/ is shown. The candidate meaning/UF/SF triplets only differ in their UFs. The UFs are split up into roots and affixes, and since there is a two-way contrast both in roots and in affixes (stressed/unstressed), there are four possible candidates that have the same surface form (because this is the input) and the same meaning (because I ignore the possibility of homonymy).[62] Consider a listener with

---

[61] This is in fact the same function as Robust Interpretive Parsing or the perception process in the mapping of overt forms to surface forms that we have seen in sections 3.2, 4.5, and 5.3.1, except that this time, the input is the surface form, and the form to be interpreted is the underlying form.

[62] I simplifyingly assume at this point that only the initial syllable of the root can be specified for stress. In real Modern Greek, however, any of the syllables in disyllabic roots can be specified for stress.

the constraint ranking as in tableau (192). All faithfulness constraints are ranked high. This rules out any candidate that is not faithful to the underlying form: the second and the last candidate are excluded by high-ranked MAX(Root), because the underlying stress on the root is not realized in the surface form; the first candidate is ruled out by high-ranked DEP(Affix), because the surface form has stress on the affix, but the underlying affix is unstressed. The listener recognizes the third candidate with an underlyingly unstressed root and an underlyingly stressed affix, |θalas+ón|.

(192) Recognition of *θalasón*

SF /θalasón/	MAX(R)	MAX(A)	DEP(R)	DEP(A)	*|θalas-| ’sea’	*|θálas-| ’sea’	*|-on| ’Gen.Pl’	*|-ón| ’Gen.Pl’
‘sea-Gen.Pl’ |θalas+on| /θalasón/				*!	*		*	
‘sea-Gen.Pl’ |θálas+on| /θalasón/	*!			*		*	*	
☞ ‘sea-Gen.Pl’ |θalas+ón| /θalasón/					*			*
‘sea-Gen.Pl’ |θálas+ón| /θalasón/	*!					*		*

An adult listener will ideally recognize the same form that she would produce. A learner is already a listener, too, and will proceed the same way in recognition as an adult speaker. However, as outlined in sections 3.3 to 3.5, a learner has an under-developed grammar and might compute a different form in the production step. This is demonstrated in the next section.

6.3.3 Virtual production

As discussed in chapter 3, a language learner will run a check on her grammar as soon as she recognized a word by virtually producing it. The meaning becomes the input to the production evaluation. Again, the learner can choose between meaning/UF/SF triplets by applying her grammar (the same constraint ranking as in the recognition step). The upper left cell, the place reserved for the input to an evaluation, contains now ‘meaning’, based on the idea that if you intend to produce an utterance, you start out with the meaning you want to express, and hence choose a form to say it with. GEN provides all possible combinations of meaning, underlying forms and surface forms as candidates. The underlying forms are combinations of root and suffix, and differ with respect to their stress specification. Unstressed roots are combined with unstressed or stressed suffixes, as are stressed roots. There are two different surface forms, one realizing stress on the root and one realizing stress on the suffix. I do not model structural constraints on the surface forms (yet). Since for the moment I only want to model a two-way contrast in morphemes, I exclude the possibility of stress on the second syllable of the root; I simply bar this possibility from GEN. There are more candidates in the production evaluation than in the recognition evaluation because in production both the UFs and the SFs can vary. If the optimal meaning/UF/SF triplet is identical to the one in recognition, no reranking takes place. If the two triplets are not identical, the resulting mismatch is deemed an error (as outlined in section 3.5), and the learner will adjust her constraint ranking. This is shown in tableau (193): the candidate with an underlyingly stressed root and an underlyingly unstressed suffix is the winner of the virtual production step, (indicated by '👄'), and differs from the winner in recognition (indicated by ' 👂 ').

(193) Production of 'sea-GEN.PL'

'sea-Gen.Pl'	MAX(R)	MAX(A)	DEP(R)	DEP(A)	*|θalas-| 'sea'	*|θálas-| 'sea'	*|-on| 'Gen.Pl'	*|-ón| 'Gen.Pl'
'sea-Gen.Pl' |θalas+on| /θalasón/				*!	*		*	
'sea-Gen.Pl' |θalas+on| /θálason/			*!		*		*	
'sea-Gen.Pl' |θálas+on| /θalasón/	*!			*		*	*	
👄 'sea-Gen.Pl' |θálas+on| /θálason/						←*	←*	
👂 'sea-Gen.Pl' |θalas+ón| /θalasón/					*!→			*→
'sea-Gen.Pl' |θalas+ón| /θálason/		*!	*		*			*
'sea-Gen.Pl' |θálas+ón| /θalasón/	*!					*		*
'sea-Gen.Pl' |θálas+ón| /θálason/		*!				*		*

This brings about constraint reranking. Any difference in the triplets of the recognized candidate from the virtually produced candidate (i.e., if the UF that is chosen in production differs from the UF in recognition, *or* the SF chosen in production differs from the SF in recognition) elicits error detection and subsequent constraint reranking. The reranking makes it more likely that in the future, recognition and production are brought into harmony.

The recognition/production step that models three levels of representation, i.e. surface form, underlying form and meaning, is put to the test in the next section.

6.4 Testing a two-way contrast

For the purpose of testing a two-way contrast in stress, I assume that the learner only has to find out whether a morpheme is underlyingly stressed or not. Therefore only faithfulness and lexical constraints are included in the simulations, and no structural constraints. I put 10 virtual learners (created in the Praat programme; Boersma & Weenink 1992-2006) to the test. The virtual learners reranked the constraints in a GLA-fashion (Boersma 1997) in the sense of section 3.6.2, with the slight difference that constraints were reranked with weighted uncancelled.[63] All constraints were initially ranked at the same heights, and the learners learned from tableaux as shown in (192) and (193). Learning was done with an evaluation noise of 2.0 and an initial plasticity of 1.0 with a decrement of 0.1. The group of learners was homogeneous, i.e. there was no variation in the reranking strategy or constraint set.

6.4.1 The training data

All of the virtual GLA learners had access to the same training data consisting of four different surface forms listed in (194), which we have

---

[63] ‘Weighted uncancelled’ means that the ranking is lowered for all constraints that are violated more in the recognized form than in the learner’s production, and the ranking is raised for all the constraints that are violated more in the learner’s production than in the recognized form, and the size of the learning step is divided by the number of constraints that move in the same direction. This makes sure that the average ranking of all the constraints is constant.

already encountered in section 6.2. The learners encountered the data in a different order. For each learner a total of 1 000 000 forms was drawn randomly from the training set, i.e. each of the four forms was presented approximately 250 000 times.

(194) The training data:
/ɣóndola/ /ɣóndolon/ /θálasa/ /θalasón/

The learners were furthermore equipped with the knowledge that corresponding underlying forms are composed of roots and suffixes, and with the meaning of these morphemes. The possible underlying forms that the learners could create are provided in the next section.

6.4.2 The pool of underlying forms to choose from: GEN

For testing the two-way contrast, GEN contained eight possible underlying forms that the learners could choose from, listed in (195). Each root and each suffix exists as underlyingly stressed or unstressed.

(195) Possible underlying forms:
|ɣondol-| |θalas-| |-a| |-on|
|ɣóndol-| |θálas-| |-á| |-ón|

Simplifyingly I assume that GEN contains only candidates where roots can have stress on the first syllable, not the second. The following section provides the constraint set.

6.4.3 The constraint set

The learners of the two-way contrast are equipped with a constraint set including faithfulness and lexical constraints. The faithfulness constraints are listed in (196), and have been discussed in section 2.6.2.

(196) The faithfulness constraints on stress
Max(Root) Dep(Root)
Max(Affix) Dep(Affix)

The lexical constraints have been discussed in section 2.6.3. The full list for the simulations of the two-way contrast is given in (197).

(197) Lexical constraints:
*|ɣondol-| ‘gondola’: Don’t connect the meaning ‘gondola’ to an unstressed root |ɣondol-|.
*|ɣóndol-| ‘gondola’: Don’t connect the meaning ‘gondola’ to a stressed root |ɣóndol-|.
*|θalas-| ‘sea’: Don’t connect the meaning ‘sea’ to an unstressed root |θalas-|.
*|θálas-| ‘sea’: Don’t connect the meaning ‘sea’ to a stressed root |θálas -|.
*|-a| ‘Nom.Sg’: Don’t connect the meaning ‘Nom.Sg’ to an unstressed suffix |-a|.
*|-á| ‘Nom.Sg’: Don’t connect the meaning ‘Nom.Sg’ to a stressed suffix |-á|.
*|-on| ‘Gen.Pl’: Don’t connect the meaning ‘Gen.Pl’ to an unstressed suffix |-on|.
*|-ón| ‘Gen.Pl’: Don’t connect the meaning ‘Gen.Pl’ to a stressed suffix |-ón|.

The list should in principle also contain constraints such as *|ɣondol-| ‘sea’ (‘don’t connect a meaning ‘sea’ to the root |ɣondol-|’). I saved some ink and computation time by not including these constraints (and they would always end up top-ranked in the computer simulations anyway, because the forms are always given along with their meaning).

6.4.4 Results: the chosen underlying forms

The final ranking of one example learner is shown in (198) (the constraints at the top are ranked higher than the constraints at the bottom):

(198) Ranking learner No. 1:
*|ɣondol-| ‘gondola’
*|-á| ‘Nom.Sg’
MAX(R)
DEP(A)
*|-on| ‘Gen.Pl’
*|θálas-| ‘sea’
*|-a| ‘Nom.Sg’
*|ɣóndol-| ‘gondola’
DEP(R)
MAX(A)
*|-ón| ‘Gen.Pl’
*|θalas-| ‘sea’

If we evaluate e.g. *γóndolon* with this ranking in tableau (199), we can see that the candidate with an underlyingly stressed root and an underlyingly stressed affix is chosen, together with the correct surface form (I leave out meaning in the candidate cells because it is always the same anyway). The numbers above the constraint columns indicate the final ranking values of the constraints. Lexical constraints militating against other morphemes are left out for better readability. All candidates with an underlyingly unstressed root are ruled out by high-ranking *|γondol-| ‘gondola’, irrespective of their surface form. The next-ranking constraint MAX(R) rules out the candidates that have an underlyingly stressed root, but where the root stress is not retained in the surface form. That leaves two competing candidates with identical surface stress on the root: one where the root is underlyingly stressed and the suffix is unstressed, and one where both root and suffix are underlyingly stressed. The constraint against an unstressed suffix *|-on| ‘Gen.Pl’ is ranked higher than the constraints against a stressed suffix *|-ón| ‘Gen.Pl’. This results in the winning candidate |γóndol+ón| /γóndolon/, where both morphemes are underlyingly stressed and where stress is on the root in the surface form.

(199) A meaning-to-form evaluation

	105.655	105.597	105.571	103.153	96.454	95.721	95.711	93.523
‘gondola-Gen.Pl’	*|ɣondol-| ‘gondola’	MAX(R)	DEP(A)	*|-on| ‘Gen.Pl’	*|ɣóndol-| ‘gondola’	DEP(R)	MAX(A)	*|-ón| ‘Gen.Pl’
|γondol+on| /γóndolon/	*!			*		*		
|γondol+on| /γondolón/	*!		*	*				
|γóndol+on| /γóndolon/				*!	*			
|γóndol+on| /γondolón/		*!	*	*	*			
|γondol+ón| /γóndolon/	*!					*	*	*
|γondol+ón| /γondolón/	*!							*
☞ |γóndol+ón| /γóndolon/					*		*	*
|γóndol+ón| /γondolón/		*!			*			*

I checked how stably this learner would choose her candidates. Out of 1 000 trials with an evaluation noise of 2.0, the learner virtually produced in 99.9% of the cases |γóndol+ón| /γóndolon/ for the meaning ‘gondola-Gen.Pl’. In 0.1% of the cases, however, MAX(A) outranked *|-on| ‘Gen.Pl’, and she therefore chose for the competing form |γóndol+on| /γóndolon/, where the surface form is the same, but the underlying form consists of a stressed root and an unstressed suffix. I regard this as a licit slip of the mind. The overall percentages for the forms that this learner chose are given in (200):

(200) Percentages for learner No. 1

‘gondola-Nom.Sg’ → |γóndol+a| /γóndola/	100%
‘gondola-Gen.Pl’ → |γóndol+ón| /γóndolon/	99.9%
‘sea-Nom.Sg’ → |θalas+a| /θálasa/	99.5%
‘sea-Gen.Pl’ → |θalas+ón| /θalasón/	99.9%

All the learners created grammars with the crucial rankings displayed in (201), and correct percentages similar to learner No. 1. Note that it is not important whether a constraint like *|ɣondol-| ‘gondola’ is ranked above a constraint *|θálas-| ‘sea’, but only whether they are ranked with another constraint referring to the same meaning.

(201) Crucial rankings

*|ɣondol-| ‘gondola’ >> *|ɣóndol-| ‘gondola’
*|θálas-| ‘sea’ >> *|θalas-| ‘sea’
*|-á| ‘Nom.Sg’ >> *|-a| ‘Nom.Sg’
*|-on| ‘Gen.Pl’ >> *|-ón| ‘Gen.Pl’
MAX(R) >> MAX(A)
DEP(A) >> DEP(R)

Since the learners choose to almost a 100% of the times the underlying forms as established (by linguists) in section 6.2, I find this an encouraging result and will check what happens when there is a three-way contrast in underlying forms involved: stressed/unstressed and pre-/post-stressing forms.[64] This is a step closer to real Modern Greek and is shown in the next section.

6.5 Testing a three-way contrast: learning pre- and post-stressing morphemes

As already mentioned in section 6.2, there are morphemes in Modern Greek that push stress onto another morpheme, instead of being stressed themselves. Examples of post-stressing roots are *uranós* ‘sky-Nom.Sg’ and *aɣorá* ‘market-Nom.Sg’. The suffixes *-os* and *-a* behave as underlyingly unstressed in combination with underlyingly unstressed roots such as *ánθropos* ‘man-Nom.Sg’ and *θálasa* ‘sea-Nom.Sg’, therefore it is unlikely (but not completely unreasonable to assume) that the suffixes in *uranós* and *aɣorá* are underlyingly stressed allomorphs of the underlyingly unstressed suffixes in *ánθropos* and *θálasa*. Rather, the stress pattern of these words is a property of the roots. I analyze pre- and post-stressing morphemes in the same way as I analyze underlyingly stressed morphemes: The root is lexically specified for being post-stressing, indicated with ‘→‘, and MAX is responsible for the surfacing stress pattern. The winning candidate in (202) is the one with stress on the suffix.

---

[64] This looks like a four-way contrast, but is not really: I assume post-stressing underlying forms only for roots, and pre-stressing underlying forms only for suffixes.

(202) Post-stressing roots

|a.γor→+a|	MAX(R)	NONFINAL	FTBIN	DEP(A)
/(á.γor) a/	*!			
☞ /a.γor (á)/		*	*	*

Likewise with pre-stressing suffixes: they impose a stress on the preceding syllable with the help of high-ranking MAX(A), as illustrated in tableau (203). Pre-stressing suffixes are indicated by a '←'. MAX(A) is violated whenever the underlying stress specification of the suffix is not fulfilled in the surface form; for tableau (203) it means that the constraint is violated when stress is too far to the left in the surface form, as in the first candidate, and also when the stress is realized on the suffix itself, as in the second candidate.[65]

(203) Pre-stressing suffixes

|an.θrop+←u|	MAX(A)	NONFINAL	FTBIN	DEP(A)
/(án.θro) pu	*!			
/an.θro (pú)/	*!	*	*	*
☞ /an (θró.pu)/		*		

For the purpose of testing the three-way contrast in stress, I assume that the learners have to find out whether a root is underlyingly post-stressing, stressed, or unstressed and whether a suffix is underlyingly pre-stressing, stressed, or unstressed. Again, only faithfulness and lexical constraints are included in the simulations, and no structural constraints. I put another 10 virtual GLA learners (created in the Praat programme; Boersma & Weenink 1992-2006) to the test. All constraints were initially ranked at equal heights, and the learners learned from tableaux introduced in (192) and (193). Learning was done with an evaluation noise of 2.0 and an initial plasticity of 1.0 with a decrement of 0.1. The group of learners was homogeneous, i.e. there was no variation in the reranking strategy or constraint set.

---

[65] This means here that MAX, together with the specification, is also responsible for the exact place of the stress within the word.

6.5.1 The training data

The training data consisted of seven surface forms that included stress, but excluded foot structure. All learners encountered the data in different orders. For each learner a total of 1 000 000 forms was drawn randomly from the training set, i.e. each of the seven forms was presented approximately 140 000 times.

(204) The training data[66]

γóndola	θálasa	ánθropos	aγorá
γóndolon	θalasón	anθrópu	

The learners were again equipped with the knowledge that the underlying forms are composed of roots and suffixes, and with the meaning of these morphemes. The possible underlying forms that the learners could create are provided in the next section.

6.5.2 The pool of underlying forms to choose from: GEN

Our regular virtual learner in this section does not know whether the root of *γóndola* is underlyingly stressed, unstressed or post-stressing, but she knows that it *could* be stressed, unstressed or post-stressing. I add post-stressing versions of all the roots in (195), e.g. |aγor→|, which combines with unstressed |-a|, and a pre-stressing version for every suffix in (195), e.g. |←on|. I also add a suffix |-u| with the meaning 'Gen.Sg.' and a root |anθrop-|, meaning 'human', to my universe. For each root, GEN therefore provides three possible allomorphs either stressed, unstressed, or post-stressing. For suffixes, GEN provides three possible allomorphs either stressed, unstressed, or pre-stressing. The list is given in (205).

---

[66] The feminine nouns ending in *-a* do not take the pre-stressing suffix *-u* as the genitive singular form (they take *-as*). Masculine nouns in *-os* take *-on* as the genitive plural suffix, but show different stress behaviour, e.g. *ánθropos* becomes *anθrópon*. It seems that there are indeed allomorphs for the suffix *-on*: an underlyingly stressed one *-ón*, and underlyingly pre-stressing one, *←on*. Since the roots cannot be freely combined with any of the suffixes, only a subset of all theoretically thinkable combinations is included in the simulations.

(205) Underlying forms as of ‘freedom of analysis’

<table>
<tr><td>|ɣondol-|</td><td>|θalas-|</td><td>|-a|</td><td>|-on|</td></tr>
<tr><td>|ɣóndol-|</td><td>|θálas-|</td><td>|-á|</td><td>|-ón|</td></tr>
<tr><td>|ɣondol→|</td><td>|θalas→|</td><td>|←a|</td><td>|←on|</td></tr>
<tr><td>|aγor-|</td><td>|anθrop-|</td><td>|-u|</td><td>|-os|</td></tr>
<tr><td>|áγor-|</td><td>|ánθrop-|</td><td>|-ú|</td><td>|-ós|</td></tr>
<tr><td>|aγor→|</td><td>|anθrop→|</td><td>|←u|</td><td>|←os|</td></tr>
</table>

For each possible underlying form there are lexical constraints that militate against it. The relevant ones are listed in the next section.

6.5.3 The constraint set

The learners are again equipped with the faithfulness constraints listed in (196) and the lexical constraints listed in (197). Since there is a three-way contrast this time, more lexical constraints have to be added. The relevant ones are listed in (206). This results in a set of 24 lexical constraints in total, plus four faithfulness constraints (in total 28 constraints). The learners learned from tableaux as shown in sections 3.3 to 3.5.

(206) Additional lexical constraints

*|γondol→| ‘gondola’: Don’t connect the meaning ‘gondola’ to a post-stressing root |ɣondol→|.

*|θalas→| ‘sea’: Don’t connect the meaning ‘sea’ to a post-stressing root |θalas→|.

*|aγor-| ‘market’: Don’t connect the meaning ‘market’ to an unstressed root |aγor-|.

*|áγor-| ‘market’: Don’t connect the meaning ‘market’ to a stressed root |áγor-|.

*|aγor→| ‘market’: Don’t connect the meaning ‘market’ to a post-stressing root |aγor→|.

*|anθrop-| ‘human’: Don’t connect the meaning ‘human’ to an unstressed root |anθrop-|.

*|ánθrop-| ‘human’: Don’t connect the meaning ‘human’ to a stressed root |ánθrop-|.

*|anθrop→| ‘human’: Don’t connect the meaning ‘human’ to a post-stressing root |anθrop→|.

*|←a| ‘Nom.Sg’: Don’t connect the meaning ‘Nom.Sg’ to a pre-stressing suffix |←a|.

*|←on| ‘Gen.Pl’: Don’t connect the meaning ‘Gen.Pl’ to a pre-stressing suffix |←on|.

*|-os| ‘Nom.Sg.M’: Don’t connect the meaning ‘Nom.Sg.M’ to an unstressed suffix |-os|.

*|-ós| ‘Nom.Sg.M’: Don’t connect the meaning ‘Nom.Sg.M’ to a stressed suffix |-ós|.

*|←os| ‘Nom.Sg.M’: Don’t connect the meaning ‘Nom.Sg.M’ to a pre-stressing suffix |←os|.

*|-u| ‘Gen.Sg’: Don’t connect the meaning ‘Gen.Sg’ to an unstressed suffix *|-u|.

*|-ú| ‘Gen.Sg’: Don’t connect the meaning ‘Gen.Sg’ to a stressed suffix |-ú|.

*|←u| ‘Gen.Sg’: Don’t connect the meaning ‘Gen.Sg’ to a pre-stressing suffix |←u|.

6.5.4 Results: the chosen underlying forms

When we check the outputs of the 10 different learners, we can see that the learners acquired the correct surface forms, but often chose between two different underlying forms for a meaning or between two surface forms for a meaning. This is illustrated with learner No. 1's ranking, listed in (207)[67] and evaluated in tableau (208).

(207) The ranking values for learner No. 1

*|γondol→| 'gondola' 119.7
*|γondol-| 'gondola' 119.0
MAX(R) 118.7
MAX(A) 111.8
*|aγor-| 'market' 117.2
*|áγor-| 'market' 116.9
DEP(A) 109.4
*|-on| 'Gen.Pl' 108.9
*|-á| 'Nom.Sg' 107.606
*|←on| 'Gen.Pl' 104.3
*|←a| 'Nom.Sg' 102.1
DEP(R) 98.5
*|-a| 'Nom.Sg' 90.3
*|-ón| 'Gen.Pl' 86.9
*|aγor→| 'market' 65.8
*|γóndol-| 'gondola' 61.1

For the meaning 'gondola-Gen.Pl', all candidates with a post-stressing or an unstressed root are ruled out by the high-ranked lexical constraints that militate against these forms. All candidates with stress on the affix in the surface form are ruled out by the high-ranked MAX constraints. The selected surface form is always /γóndolon/, therefore I exclude for brevity's sake the candidates with the alternative surface forms, because they never win anyway (as a result, MAX(R) and DEP(A) are never violated in the tableau). The candidates with an underlyingly stressed suffix are ruled out not because of the lexical constraint *|-ón| 'Gen.Pl' that militates against it, but because of MAX(A), which outranks the lexical constraint militating against an underlyingly *un*stressed suffix, *|-on| 'Gen.Pl'.

---

[67] For space reasons, I only list the constraints I discuss in tableaux (208) and (211).

(208) Meaning-to-form evaluation

‘gondola-Gen.Pl’	*|γondol→| ‘gondola’	*|γondol-| ‘gondola’	MAX(R)	MAX(A)	DEP(A)	*|-on| ‘Gen.Pl’	*|←on| ‘Gen.Pl’	DEP(R)	*|-ón| ‘Gen.Pl’	*|γóndol-| ‘gondola’
☞ 84.9% |γóndol+on| /γóndolon/						*				*
|γondol+on| /γóndolon/		*!				*		*		
|γondol→+on|/γóndolon/	*!					*		*		
15.1% |γóndol+ón| /γóndolon/				*!					*	*
|γondol+ón| /γóndolon/		*!		*				*	*	
|γondol→+ón|/γóndolon/	*!			*				*	*	
|γóndol+←on|/γóndolon/				*!			*			*
|γondol+←on|/γóndolon/		*!		*			*	*		
|γondol→+←on|/γóndolon/	*!			*			*	*		

Due to the probabilistic ranking of constraints in the GLA, the ranking between MAX(A) and *|-on| ‘Gen.Pl’ can switch from time to time. Therefore, the winning candidate under the ranking in tableau (208) is in 84.9% of the cases |γóndol+on| /γóndolon/, with an underlyingly unstressed affix, but in 15.1% of the cases |γóndol+ón| /γóndolon/, with an underlyingly stressed affix. This happens to be the form that I (and other linguists) have assumed for Modern Greek in table (177) and (178). In the evaluations where the underlying form |γóndol+ón| is chosen, *|-on| ‘Gen.Pl’ outranks MAX(A). Looking at the ranking of lexical constraints separately, we can see that the constraints clearly decide for one underlying form, |γóndol+ón|:

(209) Separate ranking of lexical constraints

{*|γondol→| ‘gondola’, *|γondol-| ‘gondola’} >> *|γóndol-| ‘gondola’
{*|-on| ‘Gen.Pl’, *|←on| ‘Gen.Pl’} >> *|-ón| ‘Gen.Pl’

Looking at the ranking of faithfulness constraints separately, I can state that the ranking I stipulated in (186) is also achieved:

(210) Separate ranking of faithfulness constraints
MAX(R) >> MAX(A) >> DEP

A case of a post-stressing form is provided in tableau (211). Due to the ranking of DEP(A) >> *|-á| ‘Nom.Sg’ in 73.9% of the cases, the candidate with an underlyingly post-stressing root and an underlyingly stressed suffix |aγor→-á| /aγorá/ wins. In 26.1% of the cases, the constraints swap places in the hierarchy and the candidate with the underlyingly post-stressing root and the underlyingly unstressed suffix |aγor→-a| /aγorá/ wins.

(211) Virtual production of *aγorá*

‘market-Nom.Sg’	MAX(R)	*|aγor-| ‘market’	*|áγor-| ‘market’	MAX(A)	DEP(A)	*|-á| ‘Nom.Sg’	*|←a| ‘Nom.Sg’	DEP(R)	*|-a| ‘Nom.Sg’	*|aγor→| ‘market’
|aγor-a| /aγorá/		*!			*				*	
|áγor-a| /aγorá/	*!		*		*				*	
26.1 % |aγor→-a| /aγorá/					*!				*	*
|aγor-á| /aγorá/		*!				*				
|áγor-á| /aγorá/	*!		*			*				
☞ 73.9% |aγor→-á| /aγorá/						*				*
|aγor-←a| /aγorá/		*!		*	*		*			
|áγor-←a| /aγorá/	*!		*	*	*		*			
|aγor→-←a| /aγorá/				*!	*		*			*

Again, if we separate the lexical constraints from the faithfulness constraints, we get the underlying forms as assumed in table (178).

(212) Separate ranking of lexical constraints for *aγora*
*|aγor-| ‘market’ >> *|áγor-| ‘market’ >> *|aγor→| ‘market’
*|-á| ‘Nom.Sg’ >> *|←a| ‘Nom.Sg’ >> *|-a| ‘Nom.Sg’

This result is only partially satisfying; the learners end up with allomorphy: for a given meaning, they choose for one underlying form, e.g. |γóndol+on| in one evaluation, but for |γóndol+ón| another time. This does not affect their communication skills, since the surface forms are uniform, but violates a principle such as lexical economy. A possible solution to the problem could be to split up the virtual production step into the computation of the underlying form from meaning as a first step, and subsequently the computation of surface form from the computed underlying form as a second step (Paul Boersma, p.c.). Another solution could be to split up the acquisition process: first, the mapping from surface form to underlying form could be modelled, and then the mapping from underlying form to meaning.

The next step will push the model one step forward: section 6.6 shows how both underlying forms and surface forms can be learned, given meaning and overt forms.

6.6 Modelling comprehension: two levels of hidden structure

This section shows how cases with two hidden levels of representations can be modelled. I now combine the two approaches of learning surface forms, as addressed in step 2 of the introduction, and learning underlying forms, as addressed in step 3 of the introduction, resulting in step 4 of the introduction. Imagine that a mother is pointing out a boat to her child, pronouncing the overt form [γóndola] at the same time: thus I stipulate that meaning and overt form are given to the learner, and that both a surface form (in the form of /(γón.do)la/) and an underlying form (in the form of |γóndol+a|) have to be created, as was illustrated in figure (173).

6.6.1 The training data

The training data are given in (213). I will limit myself to the two-way contrast in morphemes, in order to avoid an explosion of candidates and constraints. From the stress in the overt form, the learner has to infer foot structure in the surface forms and stress marks in the underlying forms.

(213) Training data

[γóndola]	[θálasa]	[astráɣalos]
[γóndolon]	[θalasón]	[astraɣalón]

The learner does not know whether the language at hand has lexical or grammatical stress, therefore three kinds of constraints are involved: faithfulness constraints, structural constraints, and lexical constraints. The faithfulness constraints are as in (196), the lexical constraints are as in (197), and the structural constraints are as in (179). For the moment, all representations will be processed in a parallel fashion; the candidates for evaluation are therefore quadruplets consisting of meaning / underlying form / surface form / overt form.

6.6.2 The pool of underlying forms to choose from: GEN

The possible underlying morphemes that GEN provides are listed in (214). For the purposes of this section I only assume a two-way contrast in the underlying forms.

(214) Possible underlying morphemes

|ɣondol-|	|θalas-|	|astraγal-|	|-a|	|-on|	|-os|
|ɣóndol-|	|θálas-|	|astráγal-|	|-á|	|-ón|	|-ós|

GEN also has to provide possible surface forms, this time candidates with varying foot structures:[68]

(215) Possible surface structures

/(γón.do)la/	/(θá.la)sa/	/(á.stra)γa.los/	/(á.stra)γa.lon/
/(γón)do.la/	/(θá)la.sa/	/(á)stra.γa.los/	/(á)stra.γa.lon/
/γon(do.lá)/	/θa(la.sá)/	/a(strá.γa)los/	/a(strá.γa)lon/
/γon.do(lá)/	/θa.la(sá)/	/a(strá)γa.los/	/a(strá)γa.lon/
/(γón.do)lon/	/(θá.la)son/	/a.stra(γa.lós)/	/a.stra(γa.lón)/
/(γón)do.lon/	/(θá)la.son/	/a.stra.γa(lós)/	/a.stra.γa(lón)/
/γon(do.lón)/	/θa(la.són)/		
/γon.do(lón)	/θa.la(són)/		

---

[68] GEN is pretty restricted here: there are no secondary stresses, i.e. more than one foot, in the candidates, and stress can never occur on the second syllable of the root.

6.6.3 The constraint set

The constraints are once more the faithfulness constraints of (196), plus the structural constraints of (179), plus lexical constraints militating against each possible underlying morpheme of (197) with additional constraints against ‘ankle’-forms:

(216) Additional lexical constraints:
*|astraɣal-| ‘ankle’
*|astráɣal-| ‘ankle’
*|-os| ‘Nom.Sg.M’
*|-ós| ‘Nom.Sg.M’

6.6.4 Results: the chosen underlying forms

The virtual learners learned as shown in the tableaux in sections 3.3 to 3.5, applying the GLA reranking strategy. They are successful with respect to the overt forms they learn to produce: these are the same than they were fed in the training phase. Communication is therefore guaranteed. However, surface forms and underlying forms are not always unique, as we can see from table (217). The meaning ‘gondola-Nom.Sg’ takes always the underlying form |γóndol+a|, the surface form /(γón.do)la/, and the overt form [γóndola], throughout all learners. For all the other forms, there is variation, but only in one of the two hidden forms, either in the surface form or in the underlying form. For instance, the meaning ‘gondola-Gen.Pl’ always takes the surface form /(γón.do)lon/, but varies between the underlying forms |γóndol+on| and |γóndol+ón|. The meaning ‘sea-Gen.Pl’ always takes the underlying form |θalas+ón|, but varies between the surface forms /θa.la(són)/ and /θa(la.són). We can observe that for the root |γóndol-|, the stressed underlying form is uniformly chosen. For the suffix |-a|, always the unstressed form is chosen; the same with the suffix |-os|. This is because there is no alternation in the overt forms: The root |γóndol-| is always stressed, and the two suffixes are always unstressed. The morphemes that have alternating stress in the overt forms display allomorphy.

(217) Results with two hidden levels (underlying and surface forms)

Meaning	Underlying form	Surface form	Overt form	L 1 %	L 2 %	L 3 %	L 4 %	L 5 %	L 6 %	L 7 %	L 8 %	L 9 %	L 10 %	Total %
‘gondola-Nom.Sg’	|γóndol-| + |-a|	/(γón.do)la/	[γóndola]	100	100	100	100	100	100	100	100	100	100	100
‘gondola-Gen.Sg’	|γóndol-| + |-on|	/(γón.do)lon/	[γóndolon]	42.9	48.2	58.2	45.1	51.5	54.1	41.5	57.7	67.1	49.8	51.61
	+ |-ón|			57.1	51.8	41.8	54.9	48.5	45.9	58.5	42.3	32.9	50.2	48.39
‘sea-Nom.Sg’	|θalas-| + |-a|	/(θá.la)sa/	[θálasa]	25.8	31.7	49.0	63.6	36.3	3.6	20.0	49.5	24.3	38.3	34.21
	|θálas-|			74.2	68.3	51.0	36.4	63.7	96.4	80.0	50.5	75.7	61.7	65.79
‘sea-Gen.Pl’	|θalas-| + |-ón|	/θa.la(són)/	[θalasón]	3.3	91.6	10.5	63.6	37.0	94.8	3.7	3.5	3.5	79.9	39.14
		/θa(la.són)/		96.7	8.4	89.5	36.4	63.0	5.2	96.3	96.5	96.5	20.1	60.86
‘ankle-Nom.Sg.M’	|astraγal-| + |-os|	/a(strá.γa)los/	[astráγalos]	34.5	68.8	15.4	33.2	53.8	6.3	32.1	61.0	18.2	16.1	33.94
	|astráγal-|			65.5	31.2	84.6	66.8	46.2	93.7	67.9	39.0	81.8	83.9	66.06
‘ankle-Gen.Pl’	|astraγal-| + |-on|	/a.stra.γa(lón)/	[astraγalón]	3.3	89.3	12.4	63.6	36.0	94.5	3.3	3.5	31.5	81.4	41.88
		/a.stra(γa.lón)/		96.7	10.7	87.6	36.4	64.0	5.5	96.7	96.5	68.5	18.6	58.12

However, it is not the case that overt forms are always faithfully mapped onto underlying forms. If underlying forms were completely faithful to the corresponding surface forms, it would imply that the genitive plural suffix *-on*, when combined with the root |γóndol-|, would always take the underlyingly unstressed form |-on|, but when combined with the root |θalas-|, would always take the underlyingly stressed form |-ón|. This is not the case; in many (not necessarily the majority) of the cases, the 'correct' underlying form (i.e. the underlying form I have assumed for Modern Greek in section 6.2) is chosen.

The problem is once again the interference of faithfulness constraints with lexical constraints, as we can see in (220). Max(A) outranks *|-on| 'Gen.Pl', and by that rules out the candidate that we would prefer. If the two constraints switch their place in the hierarchy at evaluation time (as they do from time to time, because they are ranked close to each other), the candidate I prefer would be chosen: |γóndol+ón|. If we separate the ranking of the lexical constraints from the faithfulness constraints, we can see that the selection of underlying forms is uniform:

(218) Ranking of lexical constraints
*|γondol-| 'gondola' >> *|γóndol-| 'gondola'
*|-on| 'Gen.Pl' >> *|-ón| 'Gen.Pl'

(219) Ranking of faithfulness constraints
Max(R) >> Max(A) >> Dep

Tableau (220) shows the evaluation of *γóndolon* with the relevant constraints (i.e. the lexical constraints on other morphemes than *γondol-* and *-on* are excluded).

I furthermore tested whether different initial rankings such as *Lex >> { Faith, Structural constraints } or { *Lex, Structural constraints } >> Faith would lead to uniform underlying forms and uniform surface forms, but they did not improve learning. During the learning process, faithfulness is needed. I also tested whether other input frequencies would change the result, but they did not. As already addressed in section 6.5, it will be worth investigating whether a more serial approach (modelling first perception, then recognition; or modelling the mapping from meaning to UF and then the mapping from UF to SF in production) instead of the fully parallel approach would make a difference.

(220) Evaluating ‘gondola-Gen.Pl’

<table>
<tr><th></th><th>142.6</th><th>139.6</th><th>135.3</th><th>134.2</th><th>130.9</th><th>118.8</th><th>118.2</th><th>111.6</th><th>110.7</th><th>110.0</th><th>101.0</th><th>93.7</th><th>82.9</th><th>66.8</th><th>64.5</th></tr>
<tr><th>`gondola-Gen.Pl'</th><th>* | γondol- | 'gondola'</th><th>Max (R)</th><th>Max (A)</th><th>* | -on | 'Gen.Pl'</th><th>Dep (A)</th><th>Dep (R)</th><th>Nonfinal</th><th>Parse</th><th>FtBin</th><th>Trochaic</th><th>AFR</th><th>AFL</th><th>Iambic</th><th>* | -ón | 'Gen.Pl'</th><th>* | γóndol- | 'gondola'</th></tr>
<tr><td>`gondola-Gen.Pl' | γondol+on | /(γóndo)lon/ [γóndolon]</td><td>*!</td><td></td><td></td><td>*</td><td></td><td>*</td><td></td><td>*</td><td></td><td></td><td>*</td><td></td><td>*</td><td></td><td></td></tr>
<tr><td>`gondola-Gen.Pl' | γondol+on | /(γón)dolon/ [γóndolon]</td><td>*!</td><td></td><td></td><td>*</td><td></td><td>*</td><td></td><td>**</td><td>*</td><td></td><td>**</td><td></td><td></td><td></td><td></td></tr>
<tr><td>`gondola-Gen.Pl' | γondol+on | /γondo(lón)/ [γondolón]</td><td>*!</td><td></td><td></td><td>*</td><td>*</td><td></td><td>*</td><td>**</td><td>*</td><td></td><td></td><td>**</td><td></td><td></td><td></td></tr>
<tr><td>`gondola-Gen.Pl' | γondol+on | /γon(dolón)/ [γondolón]</td><td>*!</td><td></td><td></td><td>*</td><td>*</td><td></td><td>*</td><td>*</td><td></td><td>*</td><td></td><td>*</td><td></td><td></td><td></td></tr>
<tr><td>☞`gondola-Gen.Pl' | γóndol+on | /(γóndo)lon/ [γóndolon]</td><td></td><td></td><td></td><td>*</td><td></td><td></td><td></td><td>*</td><td></td><td></td><td>*</td><td></td><td>*</td><td></td><td>*</td></tr>
<tr><td>`gondola-Gen.Pl' | γóndol+on | /(γón)dolon/ [γóndolon]</td><td></td><td></td><td></td><td>*</td><td></td><td></td><td></td><td>**!</td><td>*</td><td></td><td>**</td><td></td><td></td><td></td><td>*</td></tr>
<tr><td>`gondola-Gen.Pl' | γóndol+on | /γondo(lón)/ [γondolón]</td><td></td><td>*!</td><td></td><td>*</td><td>*</td><td></td><td>*</td><td>**</td><td>*</td><td></td><td></td><td>**</td><td></td><td></td><td>*</td></tr>
<tr><td>`gondola-Gen.Pl' | γóndol+on | /γon(dolón)/ [γondolón]</td><td></td><td>*!</td><td></td><td>*</td><td>*</td><td></td><td>*</td><td>*</td><td></td><td>*</td><td></td><td>*</td><td></td><td></td><td>*</td></tr>
<tr><td>`gondola-Gen.Pl' | γondol+ón | /(γóndo)lon/ [γóndolon]</td><td>*!</td><td></td><td>*</td><td></td><td></td><td>*</td><td></td><td>*</td><td></td><td></td><td>*</td><td></td><td>*</td><td>*</td><td></td></tr>
<tr><td>`gondola-Gen.Pl' | γondol+ón | /(γón)dolon/ [γóndolon]</td><td>*!</td><td></td><td>*</td><td></td><td></td><td>*</td><td></td><td>**</td><td>*</td><td></td><td>**</td><td></td><td></td><td>*</td><td></td></tr>
<tr><td>`gondola-Gen.Pl' | γondol+ón | /γondo(lón)/ [γondolón]</td><td>*!</td><td></td><td></td><td></td><td></td><td></td><td>*</td><td>**</td><td>*</td><td></td><td></td><td>**</td><td></td><td>*</td><td></td></tr>
<tr><td>`gondola-Gen.Pl' | γondol+ón | /γon(dolón)/ [γondolón]</td><td>*!</td><td></td><td></td><td></td><td></td><td></td><td>*</td><td>*</td><td></td><td>*</td><td></td><td>*</td><td></td><td>*</td><td></td></tr>
<tr><td>`gondola-Gen.Pl' | γóndol+ón | /(γóndo)lon/ [γóndolon]</td><td></td><td></td><td>*!</td><td></td><td></td><td></td><td></td><td>*</td><td></td><td></td><td>*</td><td></td><td>*</td><td>*</td><td>*</td></tr>
<tr><td>`gondola-Gen.Pl' | γóndol+ón | /(γón)dolon/ [γóndolon]</td><td></td><td></td><td>*!</td><td></td><td></td><td></td><td></td><td>**</td><td>*</td><td></td><td>**</td><td></td><td></td><td>*</td><td>*</td></tr>
<tr><td>`gondola-Gen.Pl' | γóndol+ón | /γondo(lón)/ [γondolón]</td><td></td><td>*!</td><td></td><td></td><td></td><td></td><td>*</td><td>**</td><td>*</td><td></td><td></td><td>**</td><td></td><td>*</td><td>*</td></tr>
<tr><td>`gondola-Gen.Pl' | γóndol+ón | /γon(dolón)/ [γondolón]</td><td></td><td>*!</td><td></td><td></td><td></td><td></td><td>*</td><td>*</td><td></td><td>*</td><td></td><td>*</td><td></td><td>*</td><td>*</td></tr>
</table>

In the following sections I discuss some alternatives to the on-line learning approach of underlying forms I proposed here.

6.7 Alternative approaches to the learning of underlying forms

In section 6.7.1 I discuss how Constraint Demotion (Tesar 1995) fares in my on-line learning approach, as opposed to the GLA. Section 6.7.2 discusses Lexicon Optimization as a means to determine underlying forms. Section 6.7.3 explores the off-line learning approach *inconsistency detection* and *surgery* by Tesar et al. (2003). Section 6.7.4 outlines probabilistic unsupervised learning of underlying forms (Jarosz 2006) that makes use of the Expectation Maximization Algorithm (Dempster et al. 1977).

6.7.1 Learning underlying forms with Constraint Demotion

To compare whether CD (Tesar 1995) would fare better as a reranking strategy in than the GLA, I ran a simulation with 10 virtual learners for the two-way contrast outlined in section 6.4 where everything was kept the same except for the reranking strategy. It turned out that the 10 CD learners arrived at a ranking that rendered the correct SFs in the production step. However, they decided to create faithful lexical allomorphs: instead of choosing just one morpheme for each meaning, as the GLA learners in their place did, they sometimes chose two, namely in the cases where the surface forms yielded alternation. The root |θalas-| occurred as consistently underlyingly unstressed when combined with the affix |-ón|, and as consistently underlyingly stressed when combined with the affix |-a|. The affix |-on| occurred as underlyingly stressed when combined with the root |θalas-|, and as underlyingly unstressed when combined with |ɣóndol-|:

(221) The resulting lexicon of CD learners:

|ɣóndol-| |θálas-| |-a| |-on|
              |θalas-| |-ón|

This was due to the fact that the CD learners were able to establish a ranking between the lexical constraints, but failed to rank the faithfulness constraints:

(222) The ranking of an example CD learner:

{MAX(R), MAX(A), DEP(R), DEP(A), *|-á| 'Nom.Sg', *|ɣondol-| 'gondola'}
\>>
{ *|-a| 'Nom.Sg', *|-on| 'Gen.Pl', *|θálas-| 'sea', *|ɣóndol-| 'gondola' }
\>>
{ *|θalas-| 'sea', *|-ón| 'Gen.Pl' }

This is a possible solution; however, it is not the most restrictive lexicon that can be found (if you assume that there can only be one underlying form for every lexical item, as listed in (178) of section 6.2). It means that in the case of alternation, SFs are always faithfully mapped onto UFs. If every encountered item is stored as is in the lexicon, large parts of the grammar can be considered superfluous. Like in the case with the GLA, it is worthwhile investigating whether a more serial processing approach can make a difference.

6.7.2 Lexicon Optimization

In Lexicon Optimization, as proposed in Prince & Smolensky (1993:191) and further developed in Itô et al. (1995), the underlying form of a word is determined by evaluating different possible underlying forms with respect to a surface form which is optimal in the ranking of the language. The optimal surface form is determined by the ranking of structural constraints, and the appropriate underlying form for this surface form is determined by faithfulness: the most faithful underlying-surface pair is the most harmonic one, and chosen as the optimal pair. Applying this to the case in point, Modern Greek, it shows that Lexicon Optimization is problematic: even given the language's constraint ranking, the decision between two underlying forms, the stressed |γóndol+on| and the unstressed |γondol+on|, cannot be made by the grammar, as shown in tableau (223).

(223) Lexicon Optimization in Modern Greek

<table>
<tr><th>UF</th><th>SF</th><th>Max(R)</th><th>Max(A)</th><th>Nonfin</th><th>AFR</th><th>AFL</th><th>FtBin</th></tr>
<tr><td>☞ |γóndol+on|</td><td>/(γóndo)lon/</td><td></td><td></td><td></td><td>*</td><td></td><td></td></tr>
<tr><td>|γóndol+on|</td><td>/γondo(lón)/</td><td>*!</td><td></td><td>*</td><td></td><td>**</td><td>*</td></tr>
<tr><td>☞ |γondol+on|</td><td>/(γóndo)lon/</td><td></td><td></td><td></td><td>*</td><td></td><td></td></tr>
<tr><td>|γondol+on|</td><td>/γondo(lón)/</td><td></td><td></td><td>*!</td><td></td><td>**</td><td>*</td></tr>
<tr><td>|γóndol+ón|</td><td>/(γóndo)lon/</td><td></td><td>*!</td><td></td><td>*</td><td></td><td></td></tr>
<tr><td>|γóndol+ón|</td><td>/γondo(lón)/</td><td>*!</td><td></td><td>*</td><td></td><td>**</td><td>*</td></tr>
<tr><td>|γondol+ón|</td><td>/(γóndo)lon/</td><td></td><td>*!</td><td></td><td>*</td><td></td><td></td></tr>
<tr><td>|γondol+ón|</td><td>/γondo(lón)/</td><td></td><td></td><td>*!</td><td></td><td>**</td><td>*</td></tr>
</table>

The difference could be made by Dep(R), which would decide that the unstressed underlying form is the optimal one. This would put the root *γondol-* on a par with the root *θalas-*, which is underlyingly unstressed, too. Stress in *γóndolon* would then be the result of the phonological default assignment. However, when combining *γondol-* and *θalas-* with *-on*, the difference between the two roots becomes apparent: *γondol-* maintains stress (becoming *γóndolon*), while *θalas-* loses it to *-on* (becoming *θalasón*). The lexicon-optimization process is therefore not sufficient.

Tesar & Smolensky (1996, 2000:77) extend Lexicon Optimization by combining the evaluation of different input-output pairs for a given word with input-output pairs of different words, i.e. including paradigmatic comparison. Paradigmatic comparisons involve off-line learning: the learner gathers informative data until she changes her grammar. I argue that this is problematic in a more realistic learning situation, because the learner does not know when she has gathered enough data, and whether these data are informative. This is not the case for an on-line learning algorithm as the GLA: a GLA learner will adjust her grammar as long as incoming forms trigger an error detection. If incoming forms do not trigger any error detection any more, learning stops, and the grammar is not changed any further.[69]

---

[69] The learning curve of the constraints reaches a plateau in this case, and the constraints do not alter their ranking values any more.

6.7.3 Comparison to inconsistency detection and surgery

Inconsistency detection and surgery (Tesar et al. 2003) makes use of paradigmatic comparison to determine underlying forms. The point of departure in Tesar et al.'s approach (2003) was similar to the one in section 6.4: surface forms are given to the learner, as is the morphemic distinction between roots and affixes. As a first step, the learner tries to find a ranking for the language data. If there is no such ranking, meaning that there is inconsistency in the data, she will modify her lexicon. To be able to do so she will gather paradigmatic information. Forms that do not show any alternation in the paradigm are faithfully mapped onto underlying forms. All alternating forms are listed as underlyingly unstressed. She will then randomly modify one of the unstressed forms and try again to find a ranking for the data. If she does not find one, she will reset the modified form to unstressed and try to modify another form she listed as unstressed. She will proceed with modifying lexicon and ranking in turns until she finds a lexicon consistent with a ranking. Via *surgery* the learner is able to remember which pairs of forms have already been tested and discarded. This algorithm proceeds in an off-line fashion: when no ranking is found that is consistent with all the data, the lexicon is modified, after gathering all possible surface forms. But a learner does not know when she gathered enough data to go on with learning. Furthermore, she has to remember all the forms she discarded as not optimal.

6.7.4 Considering multiple grammars

Another approach to the learning of underlying forms is considering multiple grammars (in fact, *all* grammars possible) at a time, as in Jarosz (2006). It makes use of the Expectation Maximization Algorithm (Dempsey et al. 1977) applied to Optimality Theoretic grammars. The proposed approach is similar to the one proposed here in 6.4: all possible underlying forms and all rankings of constraints are initially equally probable. The probability of the underlying forms and the probability of the rankings are computed in combination with the probability of observed forms.

In this approach, the probability of all possible constraint rankings and underlying forms is computed, consulting the distribution of surface forms iteratively. Every iteration step consults all possible grammars. This means that, at all times in learning, all constraint rankings (i.e., all possible grammars *N!*) are present. This does not scale well: a constraint set of e.g. 30 constraints yields 30! possible grammars: every constraint added to the possible set of UG lets the number of possible grammars grow exponentially. While this might be a proper mathematical model for finding rankings and underlying forms given a distribution of surface forms, it is not suitable as a learnability approach.

6.7.5 A note on Richness of the Base

What I proposed here is something which some OT-ists think is forbidden: I put restrictions on what is generally referred to as the *input* in OT, namely the underlying form. According to Prince & Smolensky's (1993) principle *Richness of the Base*, there should be no restrictions on the input, but at the same time they acknowledge the need for concrete underlying forms. As Tesar & Smolensky (2000:30) put it:

(224) Tesar & Smolensky's definition of Richness of the base

> "Richness of the base: The set of possible inputs to the grammars of all languages is the same. The grammatical inventories of languages are defined as the forms appearing in the outputs that emerge from the grammar when it is fed the universal set of all possible inputs (P&S section 9.3)."

This means that just as constraints are universal in OT, so is the pool of possible underlying forms. This does not mean that a given language makes use of all possible underlying forms, just as little as it makes use of all constraints. The pool of underlying forms I assumed for Modern Greek in e.g. (195) is only a small part of a universal set of possible underlying forms. A child learning Chinese could choose from the forms in (195), but probably wouldn't pick any of them, because they do not relate to any form she is exposed to in her language. So what 'input' in production-directed OT refers to is the set of possible underlying forms. For the lexicon of a given language it is necessary to make a selection among these possible forms. Prince & Smolensky (1993) proposed Lexicon Optimization as the mechanism for selecting concrete underlying forms: pairs of possible underlying forms and surface forms (since the grammar evaluates just one surface form as correct, the surface form always stays the same in the pairs, and only the underlying forms vary) are compared to each other. The most harmonic pair that is chosen by the constraint ranking contains the best underlying form for that surface form. As I argued in section 6.7.2, Lexicon Optimization can lead to abundant allomorphy: the underlying forms for given surface forms are the most faithful ones. This is undesirable, because if the underlying forms are always the most faithful ones, we end up with having an underlying form for each different surface form (e.g. German |tak| for 'day-Sg', and |tag+ə| for 'day-Pl'). This entails a lot of redundant information in the lexicon, which one would like to exclude. Itô et al. (1995) and Tesar & Smolensky (2000) got around that problem by including paradigmatic comparison. I already argued that this implies off-line learning, which I want to discard as a possibility. Instead of assuming that the pool of possible underlying forms is made available by Richness of the Base, I argue that it is Freedom of Analysis, and therefore GEN, which provides the pool. Underlying forms are an *output*, and the choice for an underlying form over another is then made by constraint evaluation. I still embrace the idea of Richness of the Base. In my account, meaning and overt forms are inputs, and are therefore the *rich base*: overt forms are the input to the comprehension process (Boersma 2000), and meaning is the input to the production process.

6.8 Discussion

The alternatives discussed in 6.7.1 to 6.7.4 have in common that they are all off-line learning approaches: the learner first has to gather paradigmatic information before she can begin to modify her lexicon. This is not a very natural approach to language acquisition. It is not clear at what point the learner knows that she gathered enough information and will not encounter any further alternations. Moreover, she has to maintain access to all the observable forms or even to all possible grammars at all times in learning. This implies unlikely mnemonic processes: incoming forms are not processed and then discarded, but stored for later consultation. Furthermore, the concept of surgery relies on backtracking, which might not be possible in a real learning situation. The multiple-grammar approach of Jarosz (2006) relies on the availability of all possible grammars during the whole learning process, which is not very realistic to assume, either.

In this chapter, it has been shown how underlying forms can be learned by a rather "stupid" on-line learning algorithm that takes meaning into account. An on-line approach of learning is better than an off-line approach, because one form is processed at a time, under one ranking at a time. Former processed forms or rankings do not have to be remembered, because their occurrence is implicitly stored in the ranking of the constraints. The ranking is adjusted systematically. No extra learning mechanisms are required than the ones already involved in a general grammar learning model of OT (e.g. Boersma 1997): interpretation of incoming forms and constraint reranking as a reaction to error detection. The learning of underlying forms takes place by learning the grammar. This resolves the problem of whether it is the grammar or the lexicon that has to be learned first.

Why is it feasible to only have a few underlying forms and not many? If underlying forms would always be faithful to surface forms, this would render the concept of a grammar superfluous: everything would be in the lexicon. Psycholinguistic evidence furthermore indicates that words are composed of and stored as parts (e.g. roots and suffixes) in the lexicon (e.g. Chomsky & Halle 1968:12), and not as forms as a whole. This means that speakers of a language should be able to decompose incoming forms and find connections between them. In the proposed model, this is ensured by linking forms to meaning in the learning process. The learning approach of underlying forms makes use of grammatical restrictions on the lexicon in form of lexical constraints. This means in effect that a strict demarcation between grammar and lexicon cannot be uphold; the lexicon becomes a part of the grammar.

7 Conclusions

This dissertation shows on the one hand how computational learnability can be applied to compare different Optimality Theoretic analyses of one and the same phenomenon. This is shown in chapter 4 on Latin stress: there are learning paths to some analyses of Latin stress (in this case Jacobs' 2000 analysis on Latin stress), but not to others. This has consequences for linguistic analyses: they should not only be coherent, but also learnable. On the other hand the dissertation shows that the grammars of individual speakers of a language (and even of one and the same variety of that language) do not have to be uniform. This is shown in chapter 5 on the learnability of Pintupi stress. Furthermore it was shown that not only learning from overt forms is possible (shown in chapters 4 and 5), which is not new (see Tesar & Smolensky 1996, 1998, 2000), but that learning from overt form is still possible if underlying forms are not given, but have to be learned as well (shown in chapter 6 with Modern Greek).

7.1 Learning hidden structures

The first applications of algorithms on the learning of metrical structures (Tesar 1995, Tesar & Smolensky 1996, 1998, 2000) started out with learning from pairs of surface and underlying forms. In the case of metrical phonology this means that the learning algorithm processed forms that contained metrical structures such as feet and stress. It soon became apparent that this is not a natural learning situation: foot structure is not directly audible in the speech stream, and has to be inferred by the language learner herself. Tesar & Smolensky (1996, 1998, 2000) therefore saw the need for learning from *overt forms*, an abstraction of the auditory speech signal. They tested 124 languages of the languages that their constraint set could generate, with one learner per language that encountered the data always in the same order. I focussed on the learning of 3 languages, with many learners that encountered the data in a random order (because this is a more natural learning situation). This resulted in considerable variation in the final grammars of the simulations on Latin and Pintupi. The learnability simulations on Latin furthermore showed that if given the possibility, the virtual learners would invent their own analyses. In section 4.9 on secondary stress it was demonstrated that if provided with main-stress-only forms, usually assumed as an indication that there is only one foot per word, the learners would process the forms as if there was potentially more than one foot per word. This had a positive effect on the learnability. The fact that the learners varied from each other in their secondary stress assignment is not absurd: for instance, native speakers of English are reported to vary in their assignment of secondary stress as well.

In chapter 6 on Modern Greek stress I argued that two levels of hidden structures have to, and can be learned: surface forms and underlying forms are mental constructs and have to be created by the learner. The learner is capable of learning them from overt forms if she takes the meaning of the forms into account. The learning mechanism I proposed is parallel and proceeds very much like Robust Interpretive Parsing (Tesar & Smolensky 1996, 1998, 2000) or Boersma's (1998:296) perception model, except that in the learning process, perception and recognition are processed in parallel. It might in fact be serial: instead of evaluating surface form and underlying form at the same time, processing therefore both perception and recognition concurrently, it might be the case that first perception is processed (i.e. first the surface form is evaluated) and then recognition (the evaluation of the underlying form). It might also be the case that after learning, production is processed serially: first, the underlying form is computed, and then the corresponding surface structure (Paul Boersma, p.c.). In any case, underlying forms are subject to *Freedom of Analysis*, and not *Richness of the Base*, and are evaluated by lexical constraints. The implication is that there is no strict demarcation between grammar and lexicon, and that at least a part of the lexicon becomes a part of the grammar.

The idea behind the learning of underlying forms with the help of meaning is that a learner of a language may find it necessary to create a form only if it has a meaning or a function. This hooks up phonology to semantics. The question can be raised, then, how autonomous the modules are. Connecting phonology to phonetics and semantics as in Boersma's programme for bidirectional phonetics and phonology (2006b) gives room for an explanation of various effects attributed to the interfaces between the modules, such as bootstrapping in language acquisition.

7.2 Why there can be different grammars for the speakers of the same language

In the simulations on Pintupi stress it becomes even more apparent that variation in the grammars of different speakers can vary, without having an effect on communication. As long as the overt forms of listeners and speakers are identical, they will not be aware of the fact that they have different grammars. The results of the Pintupi simulations also raises the question about an evaluation metric for OT grammars. The grammars of the different Pintupi learners were equally restrictive in the sense that they generated the appropriate overt forms, and only the appropriate ones. So which one of them is the most restrictive? If this cannot be determined, it is licit to state that all the grammars the learners came up with are equally appropriate.

The variation in the grammars of the different learners has a number of causes. First of all, the reranking strategies are different. Apparently, it makes a difference whether a learner reranks her constraints in a Constraint Demotion fashion or in a Gradual Learning Algorithm fashion. Second, the GLA takes evaluation noise into account; this is probably the cause why the learners in the Pintupi simulations came up with more variation in the resulting grammars than the CD learners. Third, the specific constraint sets that the learners used made a difference. Throughout the simulations it made a difference whether the learners used TROCHAIC or FTNONFIN as the constraint on trochaic feet. Unfortunately it cannot be concluded that one constraint yields better learning results than the other: in the case of Latin, TROCHAIC seemed to be the better embodiment, while in the case of Pintupi it is difficult to decide: learning was possible with either constraint. Overall it can be stated that OT constraints might have a too general formulation, and that it would be better to break down the constraints into more restrictive versions. This was done e.g. for the alignment constraints (McCarthy & Prince 1993a), which I incorporated in my simulations, but it has also been proposed for FOOTBINARITY (Hewitt 1994) and for constraints on syllable weight (Morén 2000). A fourth cause of the variation in the final grammars is that the learners encountered the data in a random order: each learner processed the training forms in a different order. It may be that the forms a learner encounters in the beginning of the learning process have a bigger impact on the learning course than forms the learner encounters later in time.

This holds especially for the GLA learners because they learned with a plasticity decrement, meaning that their learning process slowed down over the course of learning by taking smaller and smaller learning steps. This is one more cause why GLA learners displayed more variation in grammars.

7.3 The innateness of constraints

While the innateness of constraints, and OT as a grammar model of universality, can be questioned as such, this book takes the innateness of constraints as a working hypothesis. It may turn out, though, that e.g. faithfulness as such is part of Universal Grammar, but that specific faithfulness constraints have to emerge in response to the data, since faithfulness constraints refer to categories, and categories themselves have to be learned (Boersma 1998:275). In the case of lexical constraints, the innateness can be questioned even more: the concept of having lexical access at all may be innate, in that learners will invoke constraints militating against the forming of underlying forms. But they will only invoke these constraints at the moment they are confronted with a specific form. A child learning Chinese probably does not have a constraint against an underlying form |γóndol-|, because she is never confronted with a form like that.

7.4 Constraint Demotion vs. the Gradual Learning Algorithm

The guaranteed convergence of Constraint Demotion when learning from pairs of surface and underlying forms is a nice mathematical tool to determine whether there is a constraint ranking for the language data, given a set of constraints. However, this makes no statement about the appropriateness of Constraint Demotion as a learning algorithm for natural language acquisition, because learning from pairs of surface and underlying forms is an unnatural learning situation (learning in this case is “too informed”). I therefore argue that Constraint Demotion is not superior to the Gradual Learning Algorithm. In the simulations on Latin stress, the GLA fared better than CD. In the simulations on Pintupi stress, both algorithms fare equally well with respect to the primary language data the learners have been trained on. When it came to generalizations to longer forms of Pintupi that they had not been trained on, most of the GLA learners were able to transfer the weight-insensitive stress pattern that they showed in shorter forms to forms of five syllables and more. The CD learners were not consistent with their analyses, and none of them transferred the learned pattern to longer forms. One could argue that the primary language data they had been trained on provided insufficient evidence to shape an appropriate grammar; however, the GLA learners were able to find a consistent analysis to transfer the stress pattern. Conclusive information on whether real speakers of Pintupi behave like GLA learners or like CD learners when it comes to generalizations to unattested forms in Pintupi may only be gained from the investigation on behaviour of loanwords with non-initial long vowels in Pintupi.

7.5 The logical problem of language acquisition

It has been argued that children can only learn from positive evidence in the data, and that this is problematic in the case that the target language is a subset of a hypothetical language a child might entertain in the course of learning. Only negative evidence could tell the child a way out, but negative evidence is not provided by the target language: if the child never hears a certain form in a language, she can not infer that this form is impossible. The proposed learning model shows the way out of this problem by letting the child provide her own negative evidence: any candidate that is not chosen as optimal by her current grammar constitutes implicit negative evidence (Tesar & Smolensky 2000:33). She will exactly know how to change her grammar, because she has access to the constraint violations of the perceived and produced forms.

The logical problem of language acquisition might not be so problematic after all: Hendriks (2000) argues against the “logic” in the logical problem of language acquisition, because it hinges on the notion that language learners deduct hypotheses on their language by logical reasoning. As was shown by experiments on human reasoning of Wason (1966) and Griggs & Cox (1982), among others, humans deduce hypotheses not by logical reasoning. Hendriks (2000) concluded that transferred to language learning, the results of the experiments renders the notion of the logical problem of language acquisition as futile. Moreover, motherese does not seem to be very impoverished (concerning qualitative and quantitative defective nature of the stimulus), e.g. Pullum 1996, Sampson 1997; Newport, Gleitman & Gleitman (1977) report that motherese is quite well-formed.

7.6 Future research

Several possibilities for further research can be derived form the results of the present work. One is a step towards a more realistic modelling of language. One of the things that the present work took as a working hypothesis is that the virtual learners already knew syllable and word boundaries. This is not realistic. A proper model of learnability should take into account that syllable and word boundaries and the full syllable structure needs to be acquired alongside with the stress pattern. Further simulations on stress should incorporate the learning of boundaries (e.g. by incorporating phrasal stress) and full syllable structure. For modelling weight-sensitive languages, differences in weight have to be considered by e.g. modelling the sonority of segments.

A further point is that it is yet to be shown to what extent constraints have to be innate. Although this dissertation makes strong use of constraints that are hard-wired into the grammar, it is not evidence against a more emergentist approach. For instance, faithfulness constraints apply to categories (there is faithfulness to features, segments, stress, etc.), but one could argue that categories themselves have to be learned. The lexical constraints proposed in the present account need not be present in all languages. As addressed in section 7.3, a child learning another language than Modern Greek might have other constraints than *|-ón| ‘Gen.Pl’ to her disposal when learning underlying forms. A future model of learnability needs to account for how constraints themselves can be learned.

Concerning linguistic theory, it is worth to further investigate how much “allomorphy” in the form of underlying forms language users have in their minds. Native speakers of Modern Greek often hesitate, when being asked to produce a word in either genitive singular or plural case, because they are not sure about the correct stress pattern, especially in infrequent words. There are many declension classes in Modern Greek that behave all differently across the paradigm when it comes to stress, making it impossible for a speaker to predict stress for new or unfamiliar words. My proposed model predicts that learners can have variation in both surface forms and underlying forms. Further investigation on how consistently or inconsistently real Modern Greek speakers stress infrequent words could provide insight into how lexical stress is processed in interaction with the grammar.

In sum, it can be said that the specific goal of this dissertation, as defined in the introduction, is achieved: I showed that a language learning child can bootstrap into the phonology of her language, if given a sufficient number of informative pairs of overt forms and meaning, i.e. by using phonetic and semantic information. The proposed model for the learning of underlying forms may be improved by extending it to a more serial approach, where either perception and recognition are processed serially or where production is split up into a serial computation of the underlying form first, followed by the computation of the surface form. The proposed model is moreover a step towards whole language simulations of acquisition and evolution (Boersma 2006b).

References

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

Allen, W. Sydney (1978). *Vox Latina. The pronunciation of Classical Latin*. Second edition. Cambridge University Press.

Anttila, Arto (1997). *Variation in Finnish phonology and morphology*. Doctoral dissertation, Stanford University.

Apoussidou, Diana (2001). Optimaler Wortakzent im Neugriechischen. M.A. thesis, Düsseldorf.

Apoussidou, Diana (2006a). Learning to be insensitive to weight in Pintupi. Ms., ROA 799.

Apoussidou, Diana (2006b). On-line learning of underlying forms. Ms., ROA 835.

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

Apoussidou, Diana & Paul Boersma (2004a). Comparing different Optimality-Theoretic learning algorithms: the case of metrical phonology. *Papers from the AAAI Spring Symposium*, Technical Report SS-04-05, 1-7.

Apoussidou, Diana & Paul Boersma (2004b). Comparing two Optimality-Theoretic learning algorithms for Latin stress. *Proceedings of the 23rd Western Coast Conference on Formal Linguistics*. Eds. Vineeta Chand, Ann Kelleher, Angelo J. Rodríguez and Benjamin Schmeiser. 29-42.

Arvaniti, Amalia (1999). Greek voiced stops: prosody, syllabification, underlying representations or selection of the optimal? *Proceedings of the 3rd International Linguistics Conference for the Greek Language*. Athens: Ellinika Grammata. 383-390.

Blutner, Reinhard (2000). Some aspects of Optimality in natural language interpretation. *Journal of Semantics* 15:115-162.

Boersma, Paul (1997). How we learn variation, optionality, and probability. *IFA Proceedings* 21:43-58.

Boersma, Paul (1998). *Functional phonology: formalizing the interactions between articulatory and perceptual drive*s. Doctoral dissertation, University of Amsterdam.

Boersma, Paul (1999). On the need for a separate perception grammar. Ms., ROA 358.

Boersma, Paul (2000). The OCP in the perception grammar. Ms., ROA 435.

Boersma, Paul (2001). Phonology-semantics interaction in OT, and its acquisition. *Papers in Experimental and Theoretical Linguistics* 6. Eds. Robert Kirchner, Wolf Wikeley & Joe Pater. Edmonton: University of Alberta. 24-35. [ROA 369, 1999].

Boersma, Paul (2003). Review of Tesar & Smolensky (2000), *Learnability in Optimality Theory*. *Phonology* 20:436-446.

Boersma, Paul (2005). Some listener-oriented accounts of hache-aspiré in French. Ms., ROA 730.

Boersma, Paul (2006a). The acquisition and evolution of faithfulness rankings. Handout 14th Manchester Phonology Meeting, May 2006.

Boersma, Paul (2006b). A programme for bidirectional phonology and phonetics and their acquisition and evolution. Handout LOT summer school 2006, Amsterdam. http://www.fon.hum.uva.nl/paul/lot2006.html.

Boersma, Paul, Paola Escudero & Rachel Hayes (2003). Learning abstract phonological from auditory phonetic categories: an integrated model for the acquisition of language-specific sound categories. *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona. 1013-1016.

Boersma, Paul & Bruce Hayes (2001). Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32:45-86.

Boersma, Paul & Claartje Levelt (1999). Gradual constraint-ranking learning algorithm predicts acquisition order. *Proceedings of Child Language Research Forum* 30. Ed. Eve Clark. Stanford, California: CSLI. 229-237.

Boersma, Paul & Claartje Levelt (2003). Optimality Theory and phonological acquisition. *Annual Review of Language Acquisition* 3:1-50.

Boersma, Paul & David Weenink (1992-2006). *Praat: doing phonetics by computer (Version 4.4)*. [Computer programme]. Retrieved from www.praat.org.

Bosson, James E. (1964). Modern Mongolian. *Uralic and Altaic Series* 38. Bloomington: Indiana University.

Botinis, Antonis (1989). *Stress and prosodic structure in Greek*. Doctoral Dissertation, Lund University Press.

Broselow, Ellen, Su-I Chen & Marie Huffman (1997). Syllable weight: convergence of phonology and phonetics. *Phonology* 14:47-82.

Chomsky, Noam (1986). *Knowledge of language: its nature, origin, and use*. New York: Prager.

Chomsky, Noam & Morris Halle (1968). *The sound pattern of English*. New York, Evanston and London: Harper & Row Publishers.

Clark, Robin & Ian Roberts (1993). A computational model of language learnability and lanugage change. *Linguistic Inquiry* 24:299-345.

Curtin, Suzanne & Kie Zuraw (2001). Explaining constraint demotion in a developing system. *Proceedings of the 26th annual Boston University Conference on Language Development*. Eds. Barbora Skarabela, Sarah Fish & Anna H.-J. Do. Somerville, MA: Cascadilla Press. 118-129.

Daelemans, Walter, Steven Gillis & Gert Durieux (1994). The acquisition of stress: a data-oriented approach. *Computational Linguistics* 20:421-451.

Davis, Stuart (1999). On the representation of initial geminates. *Phonology* 16(1):93-104.

Dempster, Arthur P., Nan M. Laird & Donald B. Rubin (1977). Maximum likelihood from incomplete data via the EM Algorithm. *Journal of Royal Statistics Society* 39(B):1-38.

Dresher, Elan B. (1999). Charting the learning path: cues to parameter setting. *Linguistic Inquiry* 30:27-67.

Dresher, Elan B. & Jonathan Kaye (1990). A computational learning model for metrical phonology. *Cognition* 34:137-195.

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

Escudero, Paola & Paul Boersma (2003). Modelling the perceptual development of phonological contrasts with Optimality Theory and the Gradual Learning Algorithm. *Proceedings of the 25th Annual Penn Linguistics Colloquium (Penn Working Papers in Linguistics* 8.1). Eds. Sudha Arunachalam, Elsi Kaiser & Alexander Williams. 71-85.

Escudero, Paola & Paul Boersma (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition* 26(4):551-585.

Everett, Daniel (1988). On metrical constituent structure in Pirahã. *Natural Language and Linguistic Theory* 6:207-46.

Everett, Daniel & Keren Everett (1984). Syllable onsets and stress placement in Pirahã. *West Coast Conference on Formal Linguistics* 3:105–16.

Fikkert, Paula (1994). *On the acquisition of prosodic structure*. Doctoral dissertation, Leiden.

Furby, Christine (1974). Garawa phonology. *Papers in Australien Linguistics* 7:1-11.

Gibson, Edward & Ken Wexler (1994). Triggers. *Linguistic Inquiry* 24:407-454.

Gnanadesikan, Amalia (1995). Markedness and faithfulness constraints in child phonology. Ms., University of Massachusetts, Amherst.

Gordon, Matthew (2005). A perceptually-driven account of onset-sensitive stress. *Natural Language & Linguistic Theory* 23:595-653.

Gordon, Matthew (2002). A phonetically-driven account of syllable weight. *Language* 78:51-80. [Also in *UCLA Working Papers in Phonology 2*].

Green, Thomas & Michael Kenstowicz (1995). The lapse constraint. Ms., MIT. ROA 101.

Griggs, Richard A. & James R. Cox (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology* 73:407-420.

Halle, Morris & Michael Kenstowicz (1991). The Free Element Condition and cyclic versus noncyclic stress. *Linguistic Inquiry* 22:457-501.

Halle, Morris and Jean-Roger Vergnaud (1978). Metrical structures in phonology. Ms., Cambridge, Massachusetts.

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

Hammond, Michael (1984). Constraining metrical theory: a modular theory of rhythm and destressing. Doctoral dissertation, University of California, Los Angeles. [Published 1988, New York: Garland.]

Hansen, Kenneth C. & Lesley E. Hansen (1969). Pintupi Phonology. *Oceanic Linguistics* 8(2):153-170.

Hansen, Kenneth.C. & Lesley E. Hansen (1978). *The core of Pintupi grammar*. Alice Springs, NT: Institute for Aboriginal Development.

Hayes, Bruce (1981). *A metrical theory of stress rule*s. Doctoral dissertation, MIT.

Hayes, Bruce (1985). Iambic and trochaic rhythm in stress rules. *Berkeley Linguistics Society* 13. Eds. Mary Niepokuj, Mary Van Clay, Vassiliki Nikiforidou & Deborah Feder. 429-446.

Hayes, Bruce (1987). A revised parametric metrical theory. *Northeastern Linguistic Society* 17:274-289.

Hayes, Bruce (1989). Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20:253-306.

Hayes, Bruce (1991). *Metrical stress theory: principles and case studies*. Ms., UCLA.

Hayes, Bruce (1995). *Metrical stress theor*y. Chicago: University of Chicago Press.

Hendriks, Petra (2000). The problem with logic in the logical problem of language acquisition. *Proceedings of the 22nd Annual Meeting of the Cognitive Science Society*. Eds. Lila Gleitman & Aravind Joshi. 220-225.

Hewitt, Mark S. (1994). Deconstructing Foot Binarity in Koniag Alutiiq. Ms., ROA 12.

Hyman, Larry (1985). *A theory of phonological weight*. Dordrecht: Foris.

Idsardi, William J. (1992). The computation of prosody. Doctoral dissertation, MIT.

Inkelas, Sharon (1998). Exceptional stress-attracting suffixes in Turkish: Representations vs. the grammar. *The Prosody-Morphology interface*. Eds. René Kager, Harry van der Hulst & Wim Zonneveld. Cambridge: Cambridge University Press. 134-187.

Itô, Junko Armin Mester & Jaye Padgett (1995). Licensing and underspecification in Optimality Theory. *Linguistic Inquiry* 26:571-613.

Jacobs, Haike (1989). *Nonlinear studies in the historical phonology of French*. Doctoral dissertation, Katholieke Universiteit Nijmegen.

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

Jacobs, Haike (2003). The emergence of quantity-sensitivity in Latin. *Optimality Theory and language change*. Ed. D. Eric Holt. Dordrecht: Kluwer. 229-247.

Jäger, Gerhard (2003). Learning constraint sub-hierarchies. The bidirectional gradual learning algorithm. *Pragmatics in OT*. Eds. Reinhard Blutner and Henk Zeevat. Basingstoke: Palgrave McMillan.

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

Jarosz, Gaja (2006). Probabilistic unsupervised learning of Optimality Theoretic Grammars. Handout, 3rd Old World Conference in Phonology, Budapest, January 2006.

Joseph, Brian D. & Irene Philippaki-Warburton (1987). *Modern Greek*. London: Croom Helm.

Kager, René (1992). Shapes of the generalized trochee. *Proceedings of the 11th WCCFL*. Ed. Jonathan Mead. Stanford: CSLI Publications. 298-312.

Kager, René (1993a). Alternatives to the Iambic-Trochaic Law. *Natural Language and Linguistic Theory* 11:381-432.

Kager, René (1993b). The moraic iamb. *CLS 27: Papers from the 27th Regional Meeting of the Chicago Linguistic Society. Volume One: The General Session*. Eds. Lise M. Dobrin, Lynn Nichols & Rosa M. Rodriguez. Chicago: Chicago Linguistic Society. 291-305.

Kager, René (1994). Ternary rhythm in alignment theory. Ms., University of Utrecht. ROA 35.

Kager, René (1996). Stem disyllabicity in Guugu Yimidhirr. *Dam phonology. HIL Phonology Papers II*. Eds. Marina Nespor & Norval Smith. The Hague: Holland Academic Graphics. 59-101.

Kager, René (1999). *Optimality Theory*. Cambridge University Press.

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

Kenstowicz, Michael (1995). Base-identity and uniform exponence: alternatives to cyclicity. Ms., MIT. ROA 103.

Liberman, Marc (1975). The intonational system of English. Doctoral dissertation, MIT.

Liberman, Mark & Alan Prince (1977). On stress and linguistic rhythm. *Linguistic Inquiry* 8:249-336.

Manning, Christopher D. & Hinrich Schütze (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.

McCarthy, John (1979a). On stress and syllabification. *Linguistic Inquiry* 10:443-465.

McCarthy, John (1979b). *Formal problems in Semitic Phonology and Morphology*. Doctoral dissertation, MIT, Cambridge. [Distributed by Indiana University Linguistics Club, Bloomington.]

McCarthy, John & Alan Prince (1986). Prosodic morphology. Ms., University of Massachusetts at Amherst and Brandeis University.

McCarthy, John & Alan Prince (1990). Foot and word in prosodic morphology: the Arabic broken plural. *Natural Language and Linguistic Theory* 8:209-283.

McCarthy, John & Alan Prince (1993a). Generalized alignment. *Yearbook of Morphology 1993*. Eds. Geert Booij & Jaap van Marle. Dordrecht: Kluwer. 79-153.

McCarthy, John & Alan Prince (1993b). Prosodic Morphology I: constraint interaction and satisfaction. *Technical Report* 3, Rutgers University Center of Cognitive Science. ROA 482.

McCarthy, John & Alan Prince (1995). Faithfulness and reduplicative identitiy. *University of Massachusetts Occasional Papers in Linguistics 18: Papers on Optimality Theory*. Eds. Jill Beckmann, Laura Walsh Dickey & Suzanne Urbanzyk. Graduate Linguistic Student Association, University of Massachusetts, Amherst. 249-384.

McKeever Dauer, Rebecca (1980). Stress and Rhythm in Modern Greek. Doctoral dissertation, University of Edinburgh.

McQueen, James M. & Anne Cutler (1997). Cognitive processes in speech perception. *The handbook of phonetic sciences*. Eds. William J. Hardcastle & John Laver. Oxford: Blackwell. 566-585.

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

Mohanan, Karuvannur Puthanveettil (1992). Emergence of complexity in phonological development. Phonological development. Eds. Charles A. Ferguson, Lise Menn & Carol Stoel-Gammon. Timonium, Maryland: York Press, Inc. 635-662.

Morén, Bruce (2000). The puzzle of Kashmiri stress: implications for weight theory. *Phonology* 17:365-396.

Myers, Scott (1987). Vowel shortening in English. *Natural Language & Linguistic Theory* 5:485-518.

Muller, Jennifer S. (1999). A unified mora account of Chuukese. *Proceedings of the 18th Western Coast Conference on Formal Linguistics*. Eds. Sonya Bird, Andrew Carnie, Jason D. Haugen and Peter Norquest. 393-405.

Newport, Elissa, Henry Gleitman & Lila Gleitman (1977). Mother, I'd rather do it myself: some effects and non-effects of maternal speech style. *Talking to children: language input and acquisition*. Eds. Catherine Snow & Charles Ferguson. Cambridge: Cambridge University Press.

Pater, Joe (2005). Non-convergence in the GLA and variation in the CDA. Ms., ROA 780.

Phillips, Juliet R. (1973). Syntax and vocabulary of mother's speech to young children: age and sex comparisons. *Child Development* 44:182-185.

Pinker, Steven (1997). *How the mind works*. New York & London: Norton.

Powers, William T. (1973). *Behavior: The control of perception*. Chicago: Aldine.

Prince, Alan (1976). 'Applying' stress. Ms., University of Massachusetts, Amherst.

Prince, Alan (1980). A metrical theory of Estonian quantity. *Linguistic Inquiry* 11:511-562.

Prince, Alan (1983). Relating to the grid. *Linguistic Inquiry* 14:19-100.

Prince, Alan (1985). Improving Tree Theory. *Proceedings of the Berkeley Linguistics Society* 11:471-490.

Prince, Alan (1990). Quantitative consequences of rhythmic organization. *CLS 26-II: Papers from the Parasession on the Syllable in Phonetics and Phonology*. Eds. Karen Deaton, Manuela Noske & Michael Ziolkowski. Chicago: Chicago Linguistic Society. 355-398.

Prince, Alan & Paul Smolensky (1993). Optimality Theory: constraint interaction in generative grammar. Ms., Rutgers University, New Brunswick, NJ, and University of Colorado, Boulder. Technical Report RuCCS-TR-2. Rutgers Center for Cognitive Science.

Pullum, Geoffrey K. (1996). Learnability, hyperlearning, and the poverty of the stimulus. *Proceedings of the 22nd Annual Meeting: General Session and Parasession on the Role of Learnability in Grammatical Theory*, Berkely Linguistic Society. 498-513.

Revithiadou, Anthi (1998). *Headmost accent wins*. Doctoral dissertation, LOT.

Sampson, G. (1997). *Educating Eve: The 'Language Instinct' debate*. London: Cassell.

Sapir, Edward & Morris Swadesh (1960). Yana dictionary. *University of California Publications in Linguistics* 22. Berkeley: University of California Press.

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

Selkirk, Elisabeth (1984). *Phonology and syntax: the relation between sound and structure*. Cambridge, MA: MIT Press.

Sherer, Tim (1994). Prosodic phonotactics. Doctoral dissertation, University of Massachusetts, Amherst. ROA 54.

Smolensky, Paul (1996). On the comprehension/production dilemma in child language. *Linguistic Inquiry* 27:720-731.

Steriade, Donca (1988). Greek accent: a case for preserving structure. *Linguistic Inquiry* 19:271-314.

Tesar, Bruce (1995). Computational Optimality Theory. Doctoral dissertation, Department of Computer Science, University of Colorado, Boulder. ROA 90.

Tesar, Bruce (1997). An iterative strategy for learning metrical stress in Optimality Theory. *Proceedings of the 21st annual Boston University Conference on Language Development*. Eds. Elizabeth Hughes, Mary Hughes & Annabell Greenhill. Somerville, MA: Cascadilla Press. 615-626.

Tesar, Bruce (1998a). An iterative strategy for language learning. *Lingua* 104:131-145.

Tesar, Bruce (1998b). Using the mutual inconsistency of structural descriptions to overcome ambiguity in language learning. *Proceedings of the North East Linguistic Society 28*. Eds. Pius N. Tamanji and Kiyomi Kusumoto. GLSA, University of Massachusetts, Amherst. 469-483.

Tesar, Bruce (1999). Robust interpretive parsing in metrical stress theory. *Proceedings of the 17th WCCFL*. Eds. Kimary Shahin, Susan Blake, & Eun-Sook Kim. Stanford: CSLI Publications. 625-639.

Tesar, Bruce, John Alderete, Graham Horwood, Nazarré Merchant, Koichi Nishitani & Alan Prince (2003). Surgery in language learning. *Proceedings of WCCFL* 22:477-490. ROA 619.

Tesar, Bruce (2004). Contrast analysis in phonological learning. Ms., ROA 695.

Tesar, Bruce (2006). Learning from paradigmatic information. Ms., ROA 844. To appear in *NELS* 36.

Tesar, Bruce & Paul Smolensky (1993). The learning of Optimality Theory: An algorithm and some basic complexity results. Ms., ROA 52.

Tesar, Bruce & Paul Smolensky (1996). Learnability in Optimality Theory (long version). *Technical Report JHU-ogci-96-4*. Department of Cognitive Science, Johns Hopkins University, Baltimore.

Tesar, Bruce & Paul Smolensky (1998). Learnability in Optimality Theory. *Linguistic Inquiry* 29:229-268.

Tesar, Bruce & Paul Smolensky (2000). *Learnability in Optimality Theory*. Cambridge, MA: MIT Press.

Van de Vijver, Ruben (1998). *The iambic issue: iambs as a result of constraint interactio*n. Doctoral dissertation, Free University of Amsterdam. HIL Dissertations 37.

Walker, Rachel (1995). Mongolian stress: Typological implications for nonfinality in unbounded systems. *Phonology at Santa Cruz* 4:85-102.

Walker, Rachel (1997). Mongolian stress, licensing, and factorial typology. Ms., ROA 171.

Wason, Peter C. (1966). Reasoning. *New Horizons in Psychology*. Ed. B.M. Foss. Harmondsworth: Penguin Books.

Wexler, Kenneth & Peter W. Culicover (1980). *Formal principles of language acquisition*. Cambridge, MA: MIT Press.

Yip, Moira (2003). Casting doubt on the onset-rime distinction. *Lingua* 113(8):779-816.

Zec, Draga (1988). Sonority constraints on prosodic structure. Doctoral dissertation, Stanford University.

Samenvatting

In dit proefschrift wordt de leerbaarheid van grammaticale en lexicale klemtoon onderzocht. Het leerproces wordt binnen het kader van Optimaliteitstheorie computationeel gemodeleerd. De resultaten tonen aan dat zelfs als taalleerders dezelfde taal verwerven, ze toch kunnen eindigen met verschillende grammatica's. Verder tonen de resultaten aan dat eerste-taalverwervers de fonologie van hun taal leren door versterkt gebruik te maken van de betekenis van een vorm en de bijbehorende fonetische informatie.

Het voorgestelde taalmodel werkt met vier representatieniveaus: een fonetische representatie, een fonologische oppervlakterepresentatie, een fonologische onderliggende representatie en een representatie voor betekenis. Deze verschillende niveaus zijn door verschillende constraint groepen met elkaar verbonden. In het comprehensie proces, dus bij het waarnemen en herkennen van gesproken taal, wordt de verbinding tussen de fonetische vorm en de fonologische oppervlaktevorm bepaald door structurele constraints. In zowel het comprehensieproces als ook het productieproces wordt de verbinding tussen de fonologische oppervlaktevorm en de fonologische onderliggende vorm bepaald door getrouwheidsconstraints. De verbinding tussen de fonologische onderliggende vorm en de betekenis wordt door lexicale constraints bepaald, wat impliceert dat er geen strenge scheiding tussen grammatica en lexicon bestaat.

Curriculum Vitae

Diana Apoussidou was born on the 28th of April 1975 in Mönchengladbach, Germany. She studied general linguistics, educational sciences and developmental psychology at the Heinrich-Heine-Universität Düsseldorf, where she graduated as Magistra Artium in 2001. From 1998 to 2002, she was involved in the SFB 282 “Theorie des Lexikons”, first as a student assistant, then as a research assistant. It is there where she developed a deeper curiosity for phonology, and metrical phonology in particular. After an internship at the linguistic publisher Mouton de Gruyter in Berlin, she started her PhD in the year 2002 at the Institute of Phonetic Sciences, which is part of the ACLC at the Universiteit van Amsterdam. The present NWO-founded dissertation is the result.