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 |ɣóndol-|, 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).