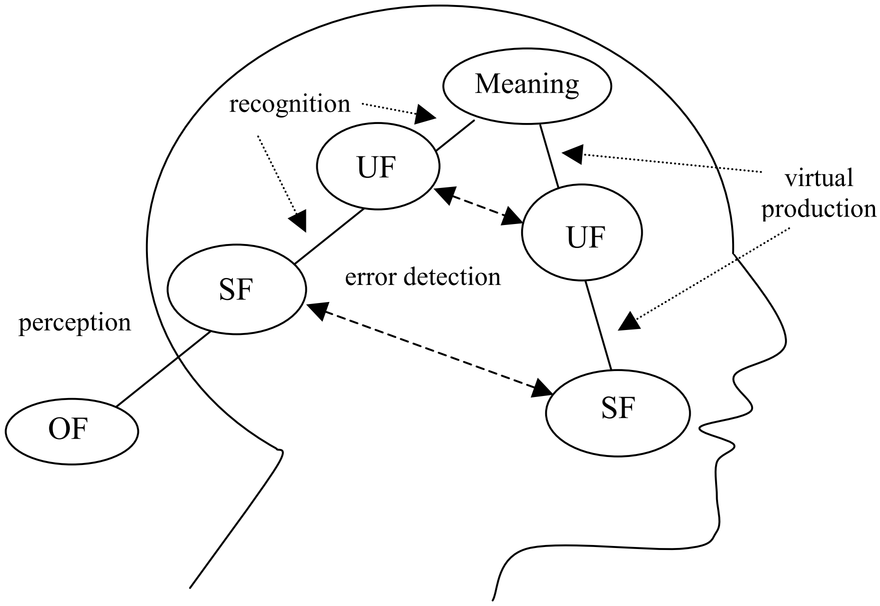
3 The learnability of hidden structure and the grammar

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.

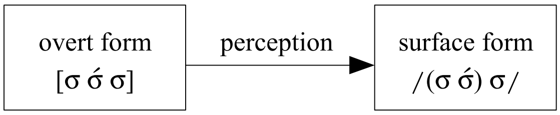
3.2 Perception

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

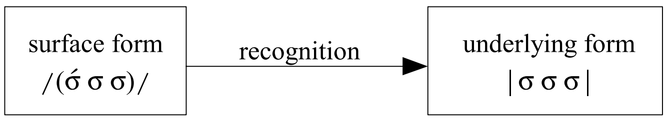
overt form: [σ σ́ σ]	AFL	AFR	TROCHAIC	IAMBIC
a. /σ (σ́) σ/ [σ σ́ σ]	*!	*		
👂 b. /(σ σ́) σ/ [σ σ́ σ]		*	*	
c. /σ (σ́ σ)/ [σ σ́ σ]	*!			*

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:

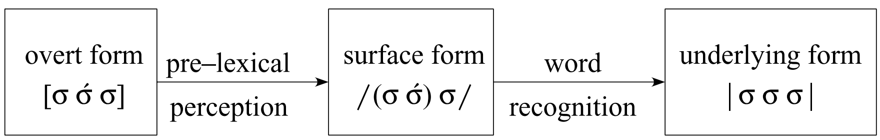
(20) Recognition



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

OF		SF		UF		Meaning
[σ σ́ σ]	→	/(σ σ́) σ/	→	|σ σ́ σ|	→	‘man–Nom.Sg’

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

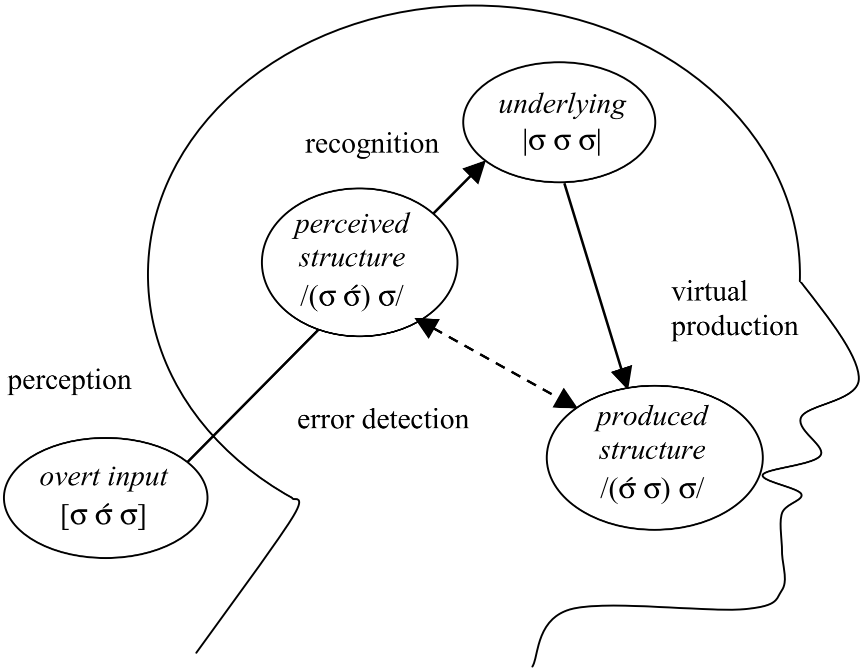
In virtual production from underlying form, the underlying form is the input to the production evaluation in tableau (24), which I repeat as tableau (26). In this virtual production step, the grammar of the learner gives out a triplet with a trochaic, left-aligned foot /(σ́ σ) σ/, indicated by ‘👄’. The 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

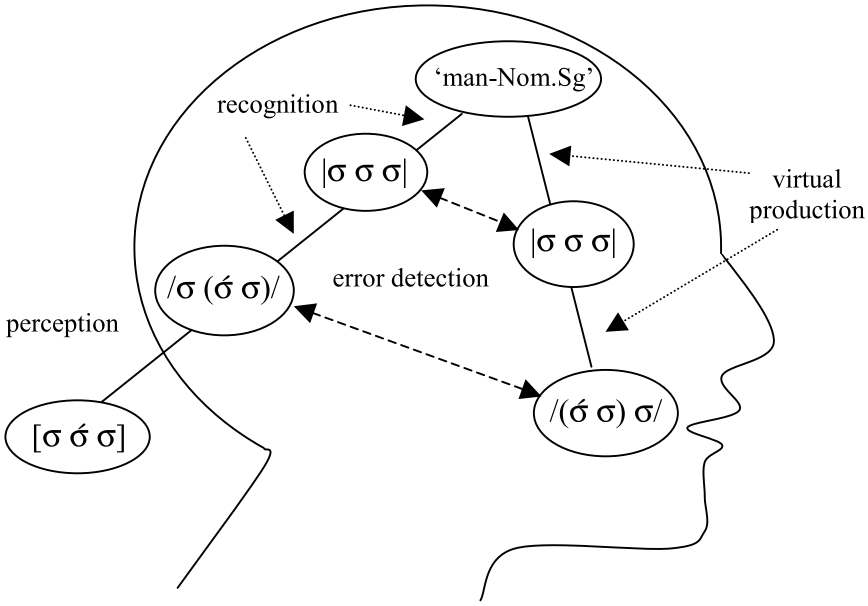
In (28) we see the same tableau as in (25), except that this time both the winning candidate in production (marked with ‘👄’) as well as the candidate of the comprehension process (marked with ‘👂’) are marked. The winning candidate in comprehension differs from the one in production only in the surface 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.

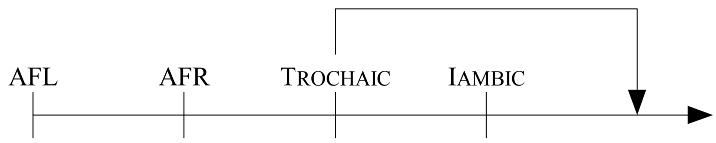
3.6 The reranking strategies

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.

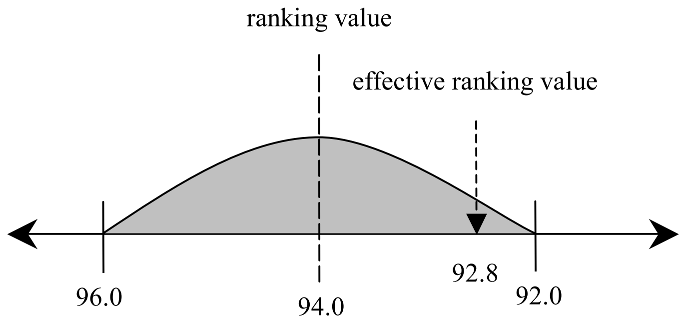
In real life, there is a considerable amount of optionality, both in what children produce and in what they encounter in the data. In this respect CD is not a realistic model of acquisition, since it cannot handle this phenomenon. A development towards solving this problem is the GLA, which is described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

---

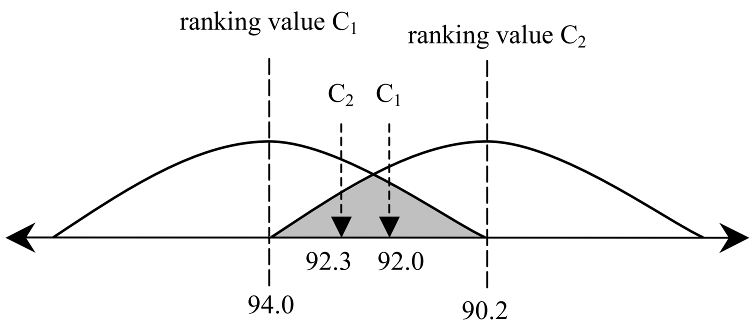
18 In fact, a Gaussian or *normal* distribution only approaches zero, but never touches it, contrary to what the figure in (34) indicates.

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):

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

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

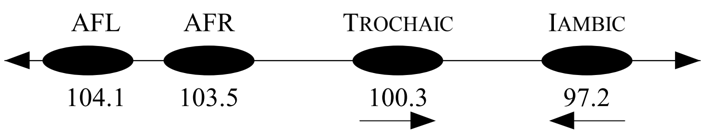
(35) Two constraints on a continuous ranking scale with overlap



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).