International Journal
of
English Studies

IJES

UNIVERSITY OF MURCIA

www.um.es/engphil/ijes

REVIEW ARTICLE

Learning the Phonology of a Language: An Optimality Theory Approach

JUAN ANTONIO CUTILLAS-ESPINOSA'
Universidad de Murcia

Tesar, Bruce & Smolensky, Paul (2000) *Learnability in Optimality Theory.* Cambridge, Mass: MIT Press. 140 pp. (ISBN 0-262-20126-7 hardback).

I. OPTIMALITY THEORY AND GENERATIVE GRAMMAR

Optimality Theory (Prince & Smolensky 1993, McCarthy & Prince 1993; OT henceforth) has experimented a spectacular development in the last decade, exerting an influence on both phonology and syntax. The first introductory texts have already appeared (Archangeli & Langendoen 1997, Dekkers et al 2000, Kager 1999, McCarthy forthcoming) as well as the first serious attacks against the theory (McMahon 2000). Both things could well be seen as indicators of OT's successful development over a relatively short period of time. OT could well be seen as a development of traditional generative grammar. There is not a complete break between the two, but rather a set of noticeable differences in the approach to the basic oppositions universal *vs* language-specific and constraints *vs* rules. Both OT and generative linguistics accept the concept of Universal Grammar (UG) but they diverge from each other, among other things, in the interpretation of cross-linguistic variation. The Principles and Parameters Theory is the standard

• ***Address forcorrespondence:*** Juan Antonio Cutillas-Espinosa, Departamentode Filología Inglesa, Facultad de Letras, Campus de La Merced, Universidad de Murcia, 30071 Murcia (Spain). E-mail: jacuti@um.es account of language-specific differences within the generative framework: "language knowledge consists of principles universal to all languages and parameters that vary from one language to another" (Cook & Newson 1996: 2). Chomsky summarises the relationship between language-specific grammars and Universal Grammar as follows:

> The grammar of a language can be regarded as a particular set of values for the parameters, while the overall system of rules, principles and parameters, is UG which we may take to be one element of human biological endowrnent, namely the "language faculty".
>
> *Chomsky(1 982: 7)*

OT assumes the existence of a Universal Grammar understood as a set of universal constraints which are *violable.* Language-specific differences arise from different hierarchies of constraints: some languages will regard some particular constraints as more important than others, so that whenever it is necessary to incur the violation of some constraint, the one which is lower ranked will be chosen.

Leamability questions have always been inherent to generative grammar. In fact, data from L1 acquisition is on the basis of many of its assumptions, to the extent that the very existence of a 'language faculty' anda 'universal grammar' are directly linked to facts about first language acquisition. OT also had to offer an answer to the question of how we learn a grammar and Tesar and Smolensky's book is a coherent account of the learning process within the optimality frarnework. The importance of their endeavour can only be measured if we consider that no grammatical model can be plausible if it is not reasonable to assume that the logic of its machinery can be mastered by a six-year-old child. The credibility of OT as a grammatical model will depend on the theorists'ability to show how a language can be acquired easily, efficiently and even in non-optimal environments.

II. READING *LEARNABZLZTY IN OPTZMALZTY THEORY:* SOME REMARKS

The general tone of the book is highly technical and those with no previous background in OT might find it not accessible. However, the basic idea (the so called RIP/CD leaming algorithm) is presented quite straightforwardly. The refinements to this basic idea are more problematic: sometimes the notation (which in some cases resembles mathematical formulae) becomes an obstacle, rather than a means of explanation. This notational complexity is further reinforced by the fact that the influence of computational linguistics is present throughout the book (Smolensky was originally a computer scientist). Perhaps that is why some of the concepts in the book seem to have been phrased in order to make them understandable for the computers where the CD algorithm was going to be tested. The problem is that this involves a degree of abstraction and specific notation which may be difficult to follow.

The structure of the book is not completely clear, insofar as it sometimes retums to previously discussed issues and develops different bits ofthe same theoretical aspects in different parts of it. This is precisely the reason why we might get the impression that it is actually a collection of papers on leamability rather than a coherent whole.

It is also essential to make clear that most of the principles about learnability presented in the book are purely theoretical: they do not emerge from empirical work on phonological acquisition. Demonstrations (where provided) are computer-based: they just show that the proposed algorithms work quite well in computers (although depending on the initial constraint hierarchy, failure to achieve the target ranking can reach 39.5%, see page 69). It is doubtful that the assumption "if it works in computer simulations, it must be at work in the human mind" can actually be defended. Finally, all explanations deal with first language acquisition, no reference is made to second language acquisition processes.

III. THE MAIN TENETS

In this section we shall proceed to discuss the main tenets of *Learnability in Optimality Theory* before we move on to discuss the different topics included in each chapter.

III.1. The problem of learning underlying forms

The discussion about leamability starts from the very nature of OT as an input-output device. We know that leamers have access to overt forms which are presented to them as a string of sounds: overt forms can be pronounced and heard. However, OT´s production of candidates relies on an input which is not necessarily equivalent to an overt form; the theory assumes that there are ***underlying forms*** (for instance, /tapóns/ for [tapónes] in Spanish or /kætz/ for [kæts] in English). Thus, OT distinguishes between the overt part of grammatical forms (which the leamer can actually hear) and full structural descriptions (which include overt and hidden forms). It is not obvious how the leamer can get to these underlying forms and this poses an important problem for leamability.

In order to leam a grammar, we need to have access to both overt and underlying forms (i.e. to ***full structural descriptions)***. The problem is that the procedure to establish the nature of underlying forms (Robust Interpretive Parsing) requires a grammar. At this point we seem to have arrived at an insurmountable difficulty, a circular situation which we have represented in figure 1.

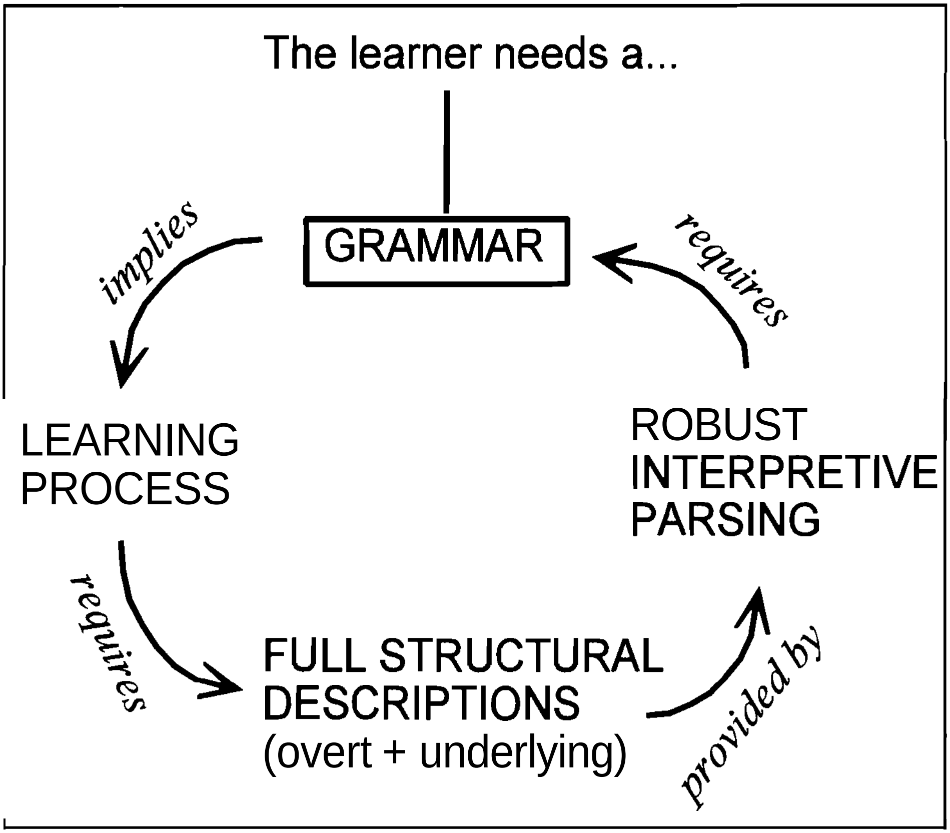

Figure 1. The problem of learning underlying forms in an Optimality Theory approach to grammar

Tesar & Smolensky suggest an ***iterative*** approach to the problem based on solutions devised for speech recognition programmes. These programmes were able to recognise a given sound and, at the same time, improve the recognition criteria with the new data provided by each occurrence of the sound. Thus they could both perform the task and improve accuracy of performance in each operation, until convergence with optimal feature specifications occurred. Applying this logic to grammar learning, we assume that the leamer starts from a provisional constraint hierarchy (grammar) which is used to analyse overt forms and get full structural descriptions. The information provided by this analysis is then used to modify the existing hierarchy and subsequently robust interpretive parsing starts again. The process is repeated until the target hierarchy is finally found. We shall now focus on how constraint rankings are modified.

1112. Constraint Demotion

In considering the process whereby constraints are ranked, we shall assume that leamers start without a fixed initial order in their innate set of constraints, although we shall also discuss the proposal of some researchers in the direction of assuming that *markedness* constraints are ranked higher *thanfaithfulness* constraints in the initial hierarchy. Thus, we start with all constraints placed in a single *stratum:*

$$\{C_1, C_2, C_3 .... C_n\}$$

The leamer perceives the leaming data and, by applying robust interpretive parsing, assigns hidden structure to the overt forms (φ). By doing this, the learner does not only get positive evidence about the nature of optimal candidates, but also negative evidence about what cannot be an optimal candidate. Leamability in Optimality Theory is based on these two sources of evidence. The leamer forms *mark datapairs,* that is to say, comparisons between the optimal candidate (and the constraints which it violates) and a suboptimal candidate and its list of violations. The relationship between them is expressed in the format *loser* $\prec$ *winner (1)*

(1)

loser $\prec$ *winner*	*marks (loser)*	*marks (winner)*
a $\prec$ b	C_1, C_2	C_2, C_3

The next step to take is to disregard those violations ofconstraints which both winnerand loser have in common. This process is called *mark cancellation.* Those marks which are cancelled are crossed off the list (2):

(2)

Loser $\prec$ *winner*	*marks (loser)*	*marks (winner)*
a $\prec$ b	$C_1, \cancel{C_2}$	$\cancel{C_2}, C_3$

Then, the learner checks that the winner mark (violation of constraint C_3) is dominated by the constraint violated by the loser (C_1) in his provisional constraint ranking. In other words, once we have deleted the violations of constraints shared by loser and winner (i.e. those which assess them as equally 'bad'), the remaining constraint(s) will have to favour the winner, that is to say, the violations of the loser have to be more important than those incurred by the winner. If this is not the case in the current ranking, it will have to be changed to match the learning data.

Let us come back to the data in (2). Each candidate violates one constraint, but in spite of this b is the winner. The only possible interpretation is that violating $\mathbb{C}_1$ is worse than violating $\mathbb{C}_3$. Let us now imagine that our learner has the following constraint ranking:

$$\{\mathbb{C}_1, \mathbb{C}_2, \mathbb{C}_3\}$$

When she realises that in her grammar $\mathbb{C}_1$ and $\mathbb{C}_3$ are equally ranked, she applies constraint demotion: she proceeds to demote the winner mark minimally, placing it in the next stratum (creating it if necessary). Thus, the new ranking, given the first mark-data pair, is the following:

$$\{\mathbb{C}_1, \mathbb{C}_2\} \gg \{\mathbb{C}_3\}$$

This is how we change the initial ranking, moving towards the target one. However, we may find cases where taking demotion decisions is not so simple:

(a) It may happen that the winner marks are already dominated by those of the loser. In this case we have been presented with a non-informativepair and the ranking will not be changed.
(b) We realise that after mark cancellation, more than one winner marks are not dominated by the loser marks. In that case one single mark data pair may produce more than one constraint demotion. If all the loser marks are placed in the same stratum, we may start considering any of them; otherwise we have to start with the highest-ranked one. Let us assume a grammar with the following constraint ranking and a mark-data pair like the one in (3):

$$\{\mathbb{C}_3, \mathbb{C}_5\} \gg \{\mathbb{C}_1\} \gg \{\mathbb{C}_{2,} \mathbb{C}_4\}$$

(3)

loser ≺ winner	marks *(loser)*	marks (winner)
a ≺ b	$\mathbb{C}_1, \mathbb{C}_{2,} \cancel{\mathbb{C}_4}$	$\mathbb{C}_3, \cancel{\mathbb{C}_4}, \mathbb{C}_5$

First, we check which is the highest-ranked loser mark; in our ranking it is $\mathbb{C}_1$. Next, we check if the winner marks are dominated by $\mathbb{C}_1$. After realising that this is not the case, we demote these constraints to the stratum immediately below $\mathbb{C}_1$:

First demotion: $\{\mathbb{C}_1\} \gg \{\mathbb{C}_3, \mathbb{C}_5, \mathbb{C}_{2,} \mathbb{C}_4\}$

Subsequently we check if the remaining loser mark ($\mathbb{C}_2$) dominates the winner marks. As $\mathbb{C}_2$ is in the same stratum as $\mathbb{C}_3$ and $\mathbb{C}_5$, we have to apply constraint demotion again, leaving this ranking:

$$\text{Second demotion: } \{\mathbb{C}_1\} \gg \{\mathbb{C}_{2,}\mathbb{C}_4\} \gg \{\mathbb{C}_3, \mathbb{C}_5\}$$

III.3. Error Driven Constraint Demotion (EDCD)

Originally presented in Tesar (1998), ***Error Driven Constraint Demotion,*** (EDCD henceforth) is a refinement on the previous Constraint Demotion algorithm (CD). The aim is to alter the procedure of search of new mark data pairs so that these are always informative. We already know that CD analyses overt forms, completing them with hidden structure and considering that these are the optimal candidates (winners). Subsequently, CD generates a suboptimal candidate (a loser), chosen at random among the endless list submitted by ***Gen.*** The problem of chosing ***any*** suboptimal candidate is that it may not be informative, that is to say, that it may not provide information which can be used in the leaming process. The solution provided by Tesar & Smolensky resorts to the following mechanisms:

- **Interpretive parsing:** It takes an overt form (φ) which has been perceived by the learner and provides a full structural description including hidden structure.
- **Production-directed parsing:** The evaluation of the different candidates which aims at selecting one of them (the optimal candidate).
- **Provisional ranking:** It is needed by the learner in order to approach the target constraint hierarchy.

The procedure performed by the algorithm is quite simple. The learner perceives an overt form (φ) and analyses it using interpretive parsing, thus achieving a full structural description including underlying / hidden structure. This is positive evidence: we know that φ, the form we have perceived, is the optimal output candidate. But is this evidence consistent with our ranking?. In order to check on this, the algorithm takes the underlying form of φ, which serves as the basis for production-directed parsing. The question is quite simple: given this input and my current grammar, which overt form is optimal? Is it the same as the one I have perceived?. If the answer is "yes", the ranking does not undergo any change, because it is consistent with the leaming data we have. What we have perceived is the same as what we would have produced. However, if the answer is "no", there is something wrong with our constraint ranking. We have to make changes so that our grammar correctly selects as the optimal output the candidate which we know is optimal. Here is where 'traditional' Constraint Demotion starts. We take the optimal output which has been perceived and the candidate which our grammar (erroneously) regards as optimal, and by contrasting them we get a mark data pair which is going to be informative. This first pair causes the demotion of one or more constraints. If this is not enough, i.e. if the candidate provided by interpretive parsing and that ofproduction-directed parsing are not the same one yet, the whole EDCD algorithm starts again. We have summarised the process in figure 2.

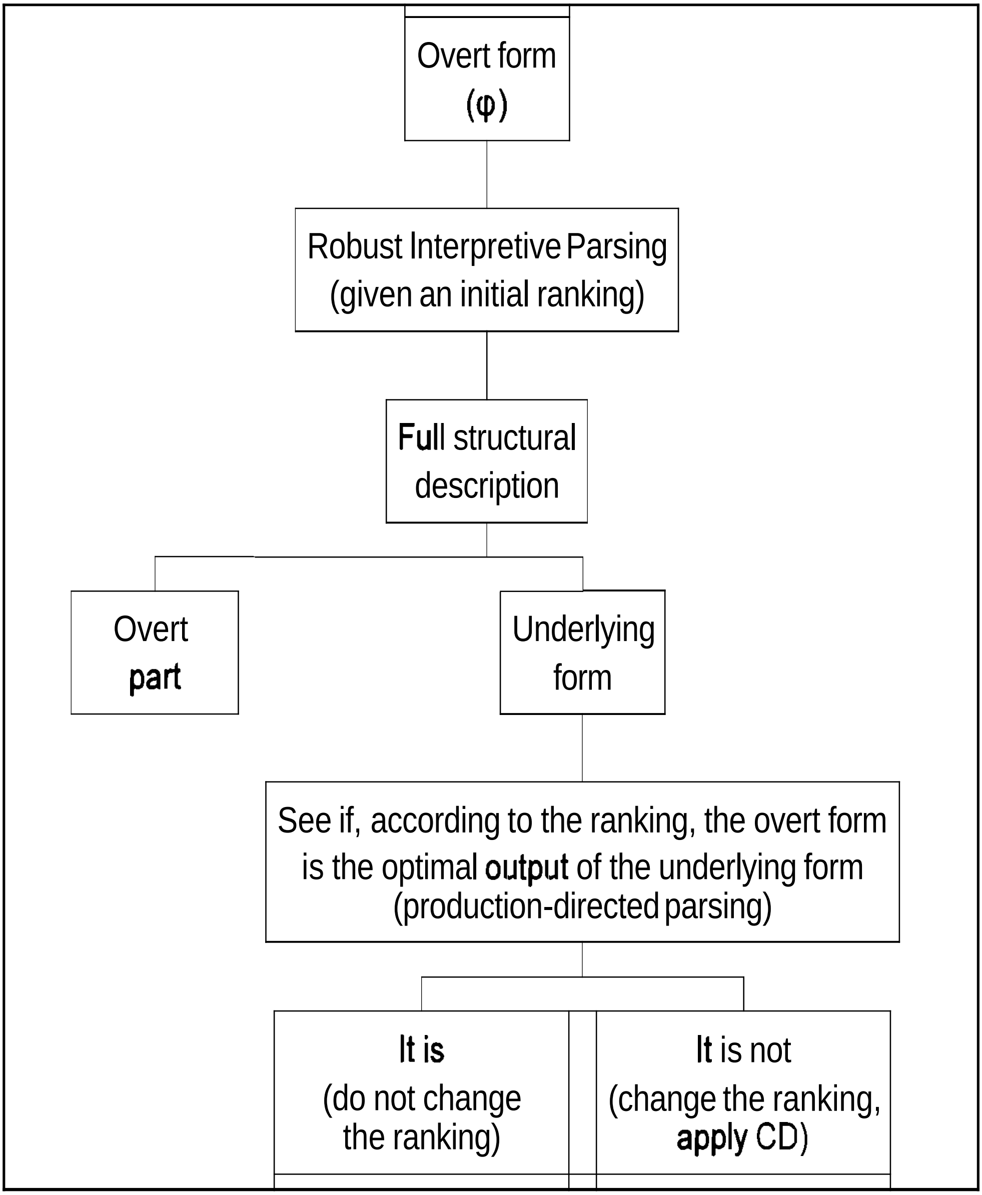
Overt form
(φ)
Robust Interpretive Parsing
(given an initial ranking)
Full structural description
Overt part
Underlying form
See if, according to the ranking, the overt form is the optimal output of the underlying form (production-directed parsing)
It is
(do not change the ranking)
It is not
(change the ranking, apply CD)

Figure 2. Simplified representation of the EDCD algorithm as explained in Tesar and Smolensky (pp. 60-62)

Let us take the example of the process of leaming Spanish plurals. It is a well-known fact that they are formed by adding the suffix *-s* to the stem. Thus, the plural of the word *casa* is *casas.* However, when the final segment of the stem is a consonant, epenthesis takes place to avoid violation of a basic phonotactic principle of Spanish which militates against word-final coda clusters: *tapón* - *tapones,* not **tapóns* (but see Alarcos (1994: 63-64) for exceptions in words such as *bíceps* or *tórax).*

We shall put forward two constraints, WF-CLUSTER (which demands that no more than one consonant can appear in word-final position) and DEP, which militates against insertions. Let us assume that the leamer has not yet ranked them, so that both constraints are placed in the same stratum:

{WF-CLUSTER, DEP)

Our learner can make do with that ranking provided that she only finds singular and plural forms of the type *casa* - *casas* and *fuerte –fuertes.* As we show in (4), the constraint ranking with no hierarchical implications for W-F CLUSTER and DEP does the job and chooses the optimal candidate:

(4)

/kása+s/	DEP	W-F CLUSTER
[kásaes]	*!	
☞ [kásas]		

The conflict arises when we face an input of the type /salóns/. In this case, we find that there is a tie between the two different candidates under consideration (5):

(5)

/salón+s/	DEP	W-F CLUSTER
? [salónes]	*	
? [salóns]		*

Let us now suppose that the learner actually perceives that the people around him actually pronounce [salónes], rather than [salóns]. After applying robust interpretive parsing to the overt phonetic form (φ = [salónes]), she gets a full structural description including the underlying form /salóns/. Subsequently, the leamer submits that underlying form (input) to production-directed parsing, getting the result which we have presented in (5). This is how she realises that there is a mismatch between perceived form ([salónes]) and the grammar's lack of arguments to choose between [salónes] and [salóns], which would probably lead to altemations between both forms. As a result of this error, the leamer gets a mark data pair (6) which, as opposed to the random procedure of selection of suboptimal candidates in CD, will always be informative, because it originates in a conflict between the grammar and phonetic (perceived) 'reality'.

(6)

*Loser ≺ **winner***		*marks (loser)*	*marks **(winner)***
b ≺ a	salóns ≺ salón	WF-CLUSTER	DEP

In (6) there is a winner mark (DEP) which is not dominated by the loser mark (W-F CLUSTER), so that the leamer proceeds to apply constraint demotion, thus leaving the following hierarchy:

WF-CLUSTER » DEP

This ranking already selects the forms [kásas] y [salónes] as the optimal outputs for the inputs /kásas/ and /salóns/, respectively (7, 8).

(7)

Input: /kása+s/	WF-CLUSTER	DEP
☞ [kásas]		
[kásaes]		*!

(8)

Input: /salón+s/	WF-CLUSTER	DEP
[salones]		*
[salóns]	*!	

To sum up, EDCD is a useful instrument to guide the search of informative mark data pairs which can help the leamer to get to the target ranking with the minimal computational effort.

IV. LEARNABZLZTY IN OPTZMALZTY THEORY, CHAPTER *BY* CHAPTER

Chapter 1 is an introduction to the contents of the book. It offers a first approach to the comparison between Principles & Parameters acquisition theory and OT. Tesar & Smolensky argue that Principles & Parameters provide either too general or too specific accounts of the learning process, whereas optimality approaches can offer theories which are both general and linguistically informed. Chapter 1 also introduces some of the basic terminology of OT and formulates the basic learning problem of acquiring hidden structures.

Chapter 2 provides a short introduction to OT. Some basic concepts are defined:

- **Constraint ranking,** which is the ranking of the universal constraints in a...
- **Dominance hierarchy,** in the sense that any constraint *dominates* all those placed below it in the ranking (in other words, it is more important than *all* the others below it).
- **Richness of the base,** whereby possible inputs are the same for all the languages in the world, so that differences between languages arise after applying a constraint ranking; in other words: "no constraints hold at the level of underlying forms" (Kager 1999: 19).
- **Harmonic ordering** of structural descriptions, implying that the one which incurs the least serious violations of constraints is the most harmonic.

Chapter 3 develops the concept of constraint demotion. It provides an example of a possible application to syllabic theory. Furthermore, it introduces other basic concepts:

- **Mark cancellation:** If two candidates violate the same constraint (ℂ), the mark of this violation will be cancelled when comparing them to decide which one is the optimal output.
- **Stratified hierarchies:** During the learning process it is possible to find that two or more constraints have the same importance in the ranking: they are said to belong to the same *stratum*. When the hierarchy develops so that there is just one constraint per stratum we say that it is totally ranked. Tesar & Smolensky argue that adult grammars are totally ranked, although the reader may think that rather than a statement about the structure of adult grammars what we are getting is a necessary condition for the successful application of Optimality Theory to the learning process. In other words, we know that OT's accounts of learnability seem to be problematic if we assume a target hierarchy which is not totally ranked, but does this necessarily mean that all adult grammars share this property?.

This chapter also includes some interesting considerations about data complexity in constraint demotion, which is of great importance if efficient and feasible learning is supposed to derive from it.

Chapter 4 presents us with the results of the application of the RIP/CD algorithm to metrical stress grammars. It is shown that the leamer's initial hierarchy has an influence on the success of the algorithm and two different solutions are proposed: either we assume that the learner tries different initial hierarchies until she finds one which makes the algorithm work or we constrain robust interpretive parsing limiting its possibilities to provide suboptimal candidates. Some possibilities for future work are also evaluated. In addition, this chapter develops the concepts of interpretive parsing and production-directed parsing and their relation to constraint demotion.

Chapter 5 is short but particularly dense. Two main issues are dealt with: i) the nature of the leamer's input and his initial constraint hierarchy and ii) the leaming of the underlying forms ofmorphemes. As far as the first question is concemed, Tesar & Smolensky insist on the concept of richness of the base: all languages share a set of possible inputs, constraints account for language-specific differences. With regard to constraint hierarchies, they make some general comments about faithfulness constraints (those which make sure that meaning differences are preserved) and structural constraints (those which disallow the presence of marked forms). It is suggested that, in the absence of any further evidence, leamers assume that structural (markedness) constraints dominate faithfulness constraints so that only ifmarked overt forms are found markedness will be demoted below faithfulness. Thus, leamers start assuming a simple system and only include marked elements as the result of overt leaming data.

As far as the leaming of the underlying forms of morphemes is concemed, the basic proposal is **Paradigmatic Lexicon Optimization** (PLO). Lexicon optimization is aprocess for the selection of the underlying form of morphemes: "the underlying form of a morpheme is the one, among all those that give the correct surface forms, that yields the maximum-Harmony paradigm" (Tesar & Smolensky: 80). In practical terms, this usually means that we minimize as far as possible the divergence between output and input forms: "Wherever the leamer has no evidence (from surface forms) to postulate a specific diverging lexical form, (s)he will assume that the input is identical to the surface form" (Kager 1999: 33). Tesar & Smolensky add that lexicon optimization has to be applied to ***completeparadigms,*** not just isolated elements, in order to account for lexical altemations. Useful examples from the devoicing of word final stops in German are provided. The concept of ***lexicon optimization*** is essential because an extreme interpretation of the richness of the base principle could lead us to infinite possible inputs, which is not feasible for leaming and computational purposes.

Chapter 6 is basically a comparison of Principles & Parameters theory and OT. It is suggested that in the former there is no consideration of the interaction between different parameters. Furthermore, parameters have to have restricted effects, which is convenient for leaming purposes but problematic for explanatory purposes. On the other hand, OT is in fact based on the ***interaction*** between constraints and Tesar & Smolensky argue that it is useful both for learning and descriptive purposes.

Chapter 7 could be seen as a schematic summary of the basic principles put forward in the book in the form of theorems and proofs, lemmas and definitions. Chapter 8 discusses a possible solution for computational problems in production-directed parsing (what Tesar & Smolensky call dynamicprogramming), although the authors claim that it is also applicable to interpretive parsing.

V. SOME RECENT ALTERNATIVES TO (ED)CD

In this section we have a look at some of the recent altemative proposals to the model presented in Tesar & Smolensky. We shall focus our attention on Prince & Tesar's Biased Constraint Demotion (1999) and Boersma & Hayes Gradual Learning Algorithm (2001).

V.1. Biased Constraint Demotion (BCD)

The BCD model, proposed by Prince & Tesar (1999) does not only place all markedness constraints in a privileged position, but also keeps their status actively. This is why it has to be regarded as an algorithm by itself, independent ofthe (ED)CD proposal. We shall not discuss the details of how the algorithm works, but rather focus our attention on some of its most important features.

Perhaps one of the most remarkable aspects of BCD is that, in practical terms, it implies the absence of an initial hierarchy. Although Prince & Tesar (1999) do not emphasise this particularly unorthodox aspect of their approach, they argue that their algorithm "places" constraints in the hierarchy and it evenpromotes some of the constraints, specially markedness ones (as opposed to the exclusively demoting technique advocated in Tesar & Smolensky). As Prince & Tesar (1999: 13) remark "Under BCD, the initial state is not arbitrary, nor does it require special stipulation". The algorithm is based on two basic principles:faithfulness *delay* and avoid the inactive.

> (a) Faithfulness **delay:** On each pass, among those constraints suitable for membership in the next stratum, if possible place only markedness constraints. Only place faithfulness constraints if no markedness constraints are available to be placed in the hierarchy (Prince & Tesar 1999: 10).
>
> (b) Avoid the inactive: When placing faithfulness constraints into the hierarchy, if possible only place those that prefer *some* winner. If the only available faithfulness constraintsprefer no remaining winners, then place all of them into the hierarchy.
>
> Prince & Tesar (1999: *11)*

Prince & Tesar suggest that faithfulness constraints should be dominated by as many markedness constraints as possible. In order to measure the degree of compliance with this principle, they propose what they call the ***r measure*** of a hierarchy: "The r-measure for a constraint hierarchy is determined by adding, for each faithfulness constraint in the hierarchy, the number of markedness constraints that dominate that faithfulness constraint" (Prince & Tesar 1999: *6).*

V.2. The Gradual Learning Algorithm

The Gradual Leaming Algorithm (Boersma & Hayes 2001) is based on an altemative account of the nature of constraints and a previous model of learnability within the model of Functional Phonology (Boersma 1997, 1998).

Hayes (2000) published a paper called «*Gradient well-formedness*» where he deals with the problem of coping with those areas of language where there is variation. He assumes that strict domination and selection of candidates is not always at work and acknowledges the need to incorporate the concept of 'preference' to any realistic grammatical model. Hayes makes reference to the traditional idea of free ranking, which implies that two constraints have exactly the same value and consequently neither of them is dominant. This is traditionally represented in Optimality Theory by a dotted line in a tableau; in the case of a tie between two constraints, each one is chosen 50% of the times (9)

(9)

Candidates	CONSTRAINT 1	CONSTRAINT 2
(50%) candl		
☞ (50%) cand2		*

Unfortunately, this idealised view does not seem to fit in with linguistic reality. Very often we find that choices are not strict, but in spite of this they reflect very clear pattems of preference of the type shown in (10).

(10)

Candidates	CONSTRAINT 1	CONSTRAINT 2
☞ (85%) candl	*	
☞ (15%) cand2		*

In order to account for this fact, Hayes suggests that constraints should be understood as ***strictness bands*** where we can find some potential ***selectionpoints.*** When two strictness bands overlap, variation appears: depending on the selection point chosen by the speaker in each strictness band, domination relations may change. The probability that one candidate is preferred to the other(s) will depend on the constraints'position on the domination continuum:

> It will be useful in what follows to consider rankings notas simple arrangements of constraint pairs but rather as the result of the constraints'each taking on arange of values on an abstract continuum [...] We can speak of each constraint possessing a ***strictness band*** [...]. Within each band, I have given a ***selection point,*** which is defined as the particular value of strictness taken on by a constraint on a given speaking occasion.
>
> *Hayes (2000: **89-90)***

Thus, constraints cannot be understood as discrete entities in a perfect domination relation, but rather as a group of domination values which can overlap with those of other constraints. In each constraint evaluation, the learner will assign an exact value to each constraint (the selection point) and this possibility of moving within a band helps us to explain probability distributions in the selection of optimal forms.

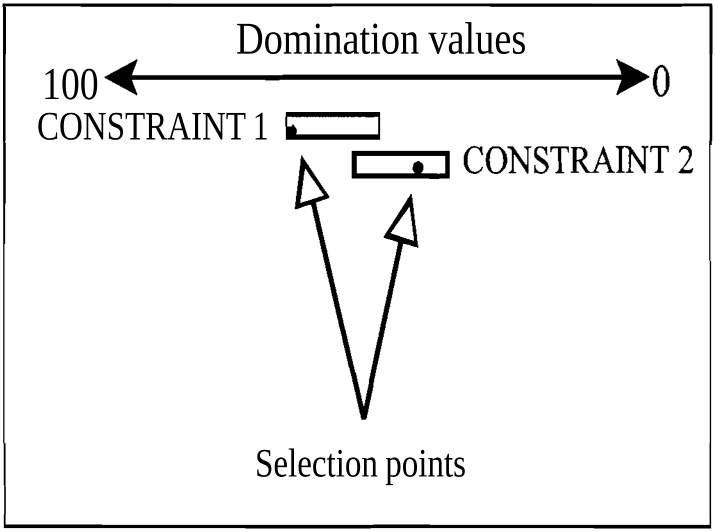

Figure 3. Constraint ranking represented as strictness bands. Given the selection points within each band, constraint 1 dominates constraint 2.

In figures 3 and 4 we show how the concept of strictness bands can account for probability distributions. Given the fact that most of the area of constraint 1 has higher domination values than constraint 2, we should expect that most of the times constraint 1 will dominate constraint 2 (figure 3). However, it is also true that both constraints overlap and consequently we may also find some cases (a minority) where constraint 2 dominates constraint 1 (figure 4).

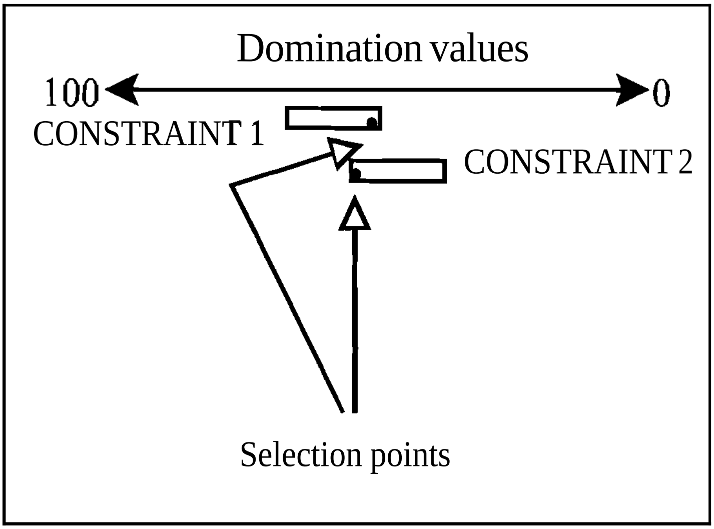

Figure 4. Constraint ranking represented as strictness bands. Given these selection points, constraint 2 dominates constraint 1 (variation).

This new approach to constraint interaction also implies a different learning algorithm. There are two slightly different versions: firstly, the Maximal Gradual Learning Algorithm (MGLA) proposed by Boersma (2000), which is a serious departure from traditional OT learning theory based on the principles of functional phonology; secondly, the Gradual Leaming algorithm by Boersma & Hayes (2001), which favours some compromise with traditional views of learnability. We shall focus our attention on the latter.

The algorithm's initial state places all constraints at the top ofthe scale with a domination value of 100. Like Tesar & Smolensky, Boersma & Hayes assume that learners have access to underlying forms. The basic mechanism is very similar to EDCD: the conflict between learning data and the learner's provisional grammar prompts changes in the ranking of constraints. The difference is that conflicts between learning data and grammar do not lead to immediate constraint demotion, but rather to a slight movement in the position of strictness bands so that the result ofthese changes is notas dramatic as in Tesar & Smolensky's model. Furthermore, the strictness bands violated by the winner will be demoted, but the ones violated by the loser will not remain unchanged, they will be promoted. These are the two greatest differences between EDCD and GLA: changes in constraint rankings are gradual and they involve both demotion and promotion.

The possibility of moving a strictness band depends on the degree of plasticity of a constraint ranking. The higher its plasticity, the more radical changes affecting the ranking will be and consequently the whole process will take place in a shorter period of time. On the other hand, a low level of plasticity helps reduce the possibility of leaming being affected by erroneous data. Boersma & Hayes proposal is that the leamer will start with a high level of plasticity in her ranking and this plasticity will gradually decrease as the leaming process progresses. Thus, as the leamer grows older it will be more diffícult to introduce drastic changes in her constraint ranking, which fits in quite well with what we know about second language acquisition.

The algorithm has been applied to different situations where EDCD seems to have problems and the results obtained have been satisfactory. Firstly, GLA is able to cope with cases of *free* variation. It changes the ranking minimally, thus managing to reflect different distributions where variation exists. Secondly, the algorithm is robust when it faces erroneous data. Both Constraint Demotion and its error driven version carry out drastic changes in constraint rankings, so that one single slip of the tongue taken as a leaming datum by the child affects her grammar dramatically. As constraint promotion is not allowed in these approaches, we cannot simply 'undo' the harm done by erroneous data: the whole constraint demotion process must start again to restore the initial state (probably after many operations and much trouble). In the case of the Gradual Learning Algorithm such a problem does not exist: changes are minimal and imply both promotion and demotion. An isolated erroneous leaming datum could only produce a small change in the constraint's domination value, a change which can easily be corrected when data which are consistent with the correct grammar are found. In (11) we have contrasted the drastic effects of applying EDCD to erroneous data with the minimal variation performed by GLA affecting domination values (in brackets). In spite of the changes in these values, one single erroneous datum does not alter the hierarchy:

(11)

EDCD: $C, \gg C_2$ ⇨ Erroneous data ⇨ $C_2 \gg C,$

GLA: $C_{1(77)} \gg C_{2(44)}$ ⇨ Erroneous data ⇨ $C_{1(72)} \gg C_{2(49)}$

The Gradual Leaming Algorithm has also been successful in coping with questions regarding*frequency* ofselection of different altemative candidates and gradient *well-formedness*, that is to say, those cases where one form is not seen as completely wrong but rather inappropriate given one's own linguistic behaviour (Hayes (2000) applies this concept to the study of the altemation between dark and light '1' in English).

Research in gradual learning is of the utmost importance. It is an attempt to adapt linguistic theory to the actual leaming process and linguistic behaviour, which imply variability and gradation. It is also a call of attention to linguists, who resort to idealised data too often. As Hayes remarks, "there is little point in analysing overidealized data [...] if you possess a theory that permits you to analyze accurate data. [...] There is good evidence that at present linguistics is not difficult enough" (Hayes 2000: 117-118). Finally, the concept of *plasticity* can account for the consolidation of the adult's grammar, for fossilisation processes in second language acquisition and for the observed interaction between age and successful L2 acquisition.

VI. THE APPLICATION TO SECOND LANGUAGE ACQUISITION RESEARCH

Second language researchers have focused their attention on the possibilities of applying OT principles to traditional problems (such as the acquisition of syllable structure or prosody), although little attention has been paid to the implications of second language acquisition studies for the formulation of leaming algorithms.

Hancin-Bhatt & Bhatt (1997) focus their attention on the acquisition of English syllabic structure by native speakers of Spanish and Japanese. In their paper they relate certain key issues in Optimality Theory to Major's Ontogeny *Model* (1987): the high level of transfer at the beginning of the leaming process may be related to the use of the constraint ranking of the leamer's mother tongue in the new L2 situation; the eventual decrease of transfer may be seen as the result of reranking.

Broselow, Chen & Wang (1998) also look at syllable structure in the interlanguage of some learners of English as a second language, resorting to the familiar OT concept of the *emergence* of the unmarked. According to their findings, the selection of different 'repair' strategies for syllable configurations which are not allowed in the leamer's L1 depends on a group of markedness constraints. These constraints are low ranked in the leamer's initial hierarchy, but as the result of the need to cope with foreign forms and unfamiliar syllable structures they become active, thus conditioning the shape of unfaithful candidates. Assuming that a violation of faithfulness has to take place (otherwise L2 syllable structure would already have been acquired), these markedness constraints make sure that at least unfaithfulness does not result in unnecessarily increased markedness.

The application of learning algorithms to second language acquisition research is specially interesting, although we can only point out some possible directions for future work. The Gradual Leaming Algorithm offers interesting insights for researchers interested in the age variable. An investigation of the concept of 'plasticity' is needed: is it adevelopmental universal or an individual characteristic? Can it be consciously altered? How could we measure it?.

A second question to be considered is how lexicon formation takes place in second language acquisition. Can we really assurne that the learner has access to underlying forrns? To what extent can we argue that, in spite of having learnt a first language, her interpretive parsing is equally robust?. If this is not the case, are all errors grammatical or are they more closely related to issues such as erroneous lexical entries based on rnisperceptions?.

Finally, second language acquisition poses sorne problerns for the concepts of *demotion* and *promotion*. When we apply CD / EDCD, alterations to the L1 ranking rnay have very drastic effects, which are sornetirnes unattested in any interlanguage. For instance, the dernotion of sonority sequencing constraints would involve acquiring, all of a sudden, different groups of clusters which, in principle, are not even related. In these cases it rnay be more reasonable to assurne that sorne constraints can be modified, rather than dernoted or prornoted, thus rninirnising the effects of learning operations. Second language research should be a valuable source of inforrnation about how theoretical generalisations about learnability fit in with actual data.

VII. CONCLUSION

Tesar & Srnolensky's book is a valuable reference for traditional approaches to leamability within an Optirnality Theory frarnework. It surnrnarises the reflections of two of the 'founder rnernbers' of the discipline. However, other researchers have developed altemative algorithms based on the previous work presented in Tesar & Smolensky which, in our opinion, are more *realistic* insofar as they can cope with variation and developrnental instability (Boersrna & Hayes 2001).

Another interesting question, which affects all cornputational approaches to leamability, is whether such theories are really grounded or not. Now we know that cornputers can actually work out the constraint ranking of a language starting frorn *some* (but not all) initial hierarchies, provided that they are given sufficient overt information. Does that really rnean that this is the way the hurnan rnind works?. We cannot be satisfied with a simple staternent of the type 'if the hurnan rnind performed these operations, it would acquire a language'. The only possible answer is that further research on phonological acquisition must be carried out in order to test whether RIP/CD and EDCD are indeed at work in phonological acquisition by hurnan beings.

Tesar & Srnolensky's Learnabiliíy in *Optimality* Theory is possibly a rnust for phonologists. But insofar as the developrnent of OT seerns to have wider implications to the extent of having becorne a revolution in linguistic theory, it is also recornrnended for linguists in general and specially for applied linguists with sorne interest in phonological acquisition.

REFERENCES

Alarcos Llorach, E. (1994). ***Gramática de la lengua española.*** Real Academia Española de la Lengua, Colección Nebrija y Bello. Madrid: Espasa Calpe.

Archangeli, D. & Langendoen, D.T. (Eds.)(1997). ***Optimality Theory: An introduction.*** Oxford: Blackwell.

Boersma, P. (1997). How we learn variation, optionality and probability. Rutgers Center for Cognitive Sciences and Rutgers University, New Brunswick. [http://ruccs.rutgers.edu/roa.html], ROA-221.

Boersma, P. (1998). ***Functional phonology: formalizing the interactions between articulatory and perceptual drives.*** LOT dissertations 11. The Hague: Holland Academic Graphics.

Boersma, P. (2000). Learning a grammar in functional phonology. In: J. Dekkers, F. van der Leeuw and J. van de Weijer (eds).

Boersma, P. & Hayes, B.P. (2001). Empirical tests of the gradual learning algorithm. ***Linguistic Inquiry*** 32(1): 45-86.

Broselow, E., Chen, S-I. & Wang, C. (1998). The emergence of the unmarked in second language phonology. ***Studies in Second Language Acquisition***, 20: 261-280.

Chomsky, N. (1982). ***Some concepts and consequences of the Theory of Government and Binding.*** Cambridge, Mass.: MIT Press.

Cook, V.J. & Newson, M. (1996). ***Chomsky's Universal Grammar. An introduction.*** Oxford: Blackwell.

Dekkers, J., Leeuw, F. & Weijer, J. (2000). ***Optimality theory: phonology, syntax and acquisition.*** Oxford: OUP.

Hancin-Bhatt, B. & Bhatt, R. (1997). Optimal L2 syllables. lnteractions of transfer and developmental factors. ***Studies in Second Language Acquisition,*** 19: 331-378

Hayes, B.P. (2000). Gradient well-formedness in Optimality Theory. In: J. Dekkers, F. van der Leeuw and J. van de Weijer.

Kager, R. (1999). ***Optimality theory.*** Cambridge: CUP.

Major, R.C. (1987). A model for interlanguage phonology. In: G. Ioup and S. Weinberger (eds)

Interlanguage phonology: the acquisition of a second language sound system (pp. 101-124). New York: Newbury House.

McCarthy, J. (forthcoming). *Optimality theory: A thematic guide*. (Manuscript).

McCarthy, J. & Prince, A. (1993) ***Prosodic morphology 1: constraint interaction and satisfaction.*** Ms., University of Massachussetts, Amherst and Rutgers University.

McMahon, A. (2000) *Change, chance and optimality*. Oxford: OUP.

Prince, A. & Smolensky, P. (1993) *Optimality Theory: constraint interaction in generative grammar*. Ms., Rutgers University, New Brunswick and University of Colorado, Boulder.

Prince, A. & Tesar, B. (1999). Leaming phonotactic distributions. ROA-353.

Tesar, B. (1998). Error-driven learning in Optimality Theory via the efficient computation of optimal forms. In P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis & D. Pesetsky (eds) ***Is the best good enough? Optimality and competition in syntax*** (pp. 421-435). Cambridge, Mass: MIT Press and ***MIT Working Papers in Linguistics***.