A Short Treatise of Optimality Theory

Gisbert Fanselow & Caroline Féry

University of Potsdam

2002

Introduction*

This short treatise is meant to be the first part of a future book on the strengths and shortcomings of Optimality Theory (OT) as applied to syntax, morphology and phonology. The first four chapters of that book, included in the present volume, introduce the main theoretical claims of OT, its predictions and strengths, and, as such, can be read by themselves.

OT is a theory of grammar which focuses on conflicts between the different ideals that linguistic utterances try to attain. A structure is grammatical if it manages to get as close to fulfilling these principles as possible. OT's main contribution to the theory of grammar is the way it makes this basic idea precise, viz., by the introduction of the concept of universal, violable and rankable constraints affecting linguistic structures, which may be in conflict with each other. OT resolves these conflicts by ranking the relevant constraints. The higher a constraint is placed in the hierarchy, the more important it is that it be respected by grammatical forms. All but the highest of the constraints are dominated and all constraints can be violated by grammatical linguistic structures if their violation implies the fulfilment of higher ranking principles.

The set of universal constraints, which make up Universal Grammar, is ideally motivated on the basis of independent principles, such as ease of articulation, perceptual contrastivity, typological generalizations, economy or analogy. Depending on their ranking, the constraints decide on the well-formedness of so-called candidates, which are competing structures for one and the same grammatical output. Taking a simple example, suppose there are two rivalling candidates for a single output form, $Cand_1$ and $Cand_2$, as well as two conflicting constraints, A and B. A could be a constraint requiring that a linguistic structure must be simple, and B could be one appealing to conservatism and militating against change. Suppose further that $Cand_1$, the faithful candidate, violates constraint A (and fulfills B) and $Cand_2$, the simpler but changed candidate, violates constraint B (but fulfills A). Depending on the ranking among A and B, a different one of the two candidates wins the competition. If A is higher-ranking than B, $Cand_2$ wins, since this candidate violates the higher-ranking constraint and is eliminated from the competition. If the order between the two constraints is reversed, $Cand_1$ is the winner.

* We thank Kathleen Kaltz, Anja Mietz and Ewa Trutkowsky for their help with the formating of this treatise, as well as Kirsten Brock and Tonio Green who checked the English.

There has been a number of attempts to explain the working of OT by means of conflicts external to the grammar: traffic rules and traffic signs are a favorite way of outlining the basics of OT. Here, we choose to illustrate the working of OT with simple examples involving Grice's Cooperative Principle and Conversational Maxims, which constitute a linguistic domain yet different from the grammatical facts lying at the heart of our interests in the following chapters. We take a naive, simplistic and playful stand on the way the conversational maxims are envisaged as tools in the description of real communication. As will become clear to pragmaticians (and others), it is neither the content nor the proper (philosophical) interpretation of the maxims that we are interested in, but just the conflicts they trigger and the way the conflicts are resolved (see Blutner 2000 and Krifka 2002 for more serious attempts of integrating OT into pragmatics). Grice's (1975) main interest has been to describe situations in which speakers (or writers) blatantly fail to observe a conversational maxim, but still obey the cooperative principle. Failures of observance of maxims happen in situations in which speakers have good reasons to do so. In OT, as shown above, this can be expressed by ranking constraints in a way that the violated maxim is ranked lower than the other ones, for the sake of another constraint triggering the violation.

According to Grice, the general aim of communication is expressed by the Cooperative Principle (1), a straightforward principle, vague enough to cover all kinds of situations. This principle assumes that protagonists in a conversation cooperate to render the communication optimal. Furthermore, a number of Conversational Maxims, formulated in (2) decompose the cooperative principle in as many subgoals. We have adorned each of the maxims with a name in small caps, turning them into constraints which can be ranked in an OT fashion.

(1) Grice's Cooperative Principle

Make your contribution as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged.

(2) Grice's Conversational Maxims

a. Maxims of quantity (QUANTITY)
1. Make your contribution as informative as required. (INFORMATION)
2. Do not make your contribution more informative than is required. (*DETAIL)

b. Maxims of quality (QUALITY)
1. Do not say what your believe to be false. (*FALSE)

2. Do not say that for which you lack adequate evidence. (*FABULATE)

c. Maxim of relation (RELATION)
   1. Be relevant. (RELEVANCE)

d. Maxims of manner (MANNER)
   1. Avoid obscurity of expression. (*OBSCURITY)
   2. Avoid ambiguity. (*AMBIGUITY)
   3. Be brief. (BEBRIEF)
   4. Be orderly. (ORDER)

QUANTITY, QUALITY, RELATION, and MANNER are cover constraints, and they stand for the more specific constraints that they subsume. Apparently, Grice's maxims apply perfectly in written communication of the type typically found in cooking recipes or (well-composed!) software installation handbooks. In this kind of texts, the communication works optimally in the sense of (2), with maximal informativeness, clarity and no redundancy. Ideally, such instructions are orderly and brief; they avoid ambiguity and obscurity, they say only what is relevant, they tell no lies, avoid unnecessary lengths, and of course, they are as informative as required. In an optimality-theoretic kind of representation, all maxims appear to be equally ranked. They do not seem to conflict with each other, maybe because such planned communication is a thoughtful action, in which writers have the opportunity to organize what they want to say in an ideal way. There is no time constraint, and except in a situation in which the recipe has to be compressed to make it fit on a file card, there is no spatial or temporal limitation. Redundancies are avoided because the reader can reread the text until everything is clear, and the writer, knowing that, has no reason to repeat some part of the information, even if it is crucial. Apparently, the constraints do not get into conflict with each other, presumably because of the absence of limitations on text construction that would render the simultaneous fulfilment of the constraints impossible.

We will now discuss a number of examples of violations of maxims for the sake of respecting others, and show how these examples can be accounted for in an OT-like fashion. After that, we will return to the informationally ideal text type recipe.

First, external constraints may make one goal more important than others. Due to these external constraints, it is no longer the case that texts (candidates) can be constructed that fulfill the requirements of all constraints. The maxims get ranked, and when in conflict, the one with the highest rank decides. A first instance of conflicts among the maxims arising in this way may involve the requirement to be unambiguous (*AMBIGUITY) and the need to be brief (BEBRIEF), both being

maxims of manner. Texts on a sign must be legible at greater distances, and there are trivial size restrictions on signs. In this situation, brevity is a key goal, which may turn out to be often incompatible with avoidance of ambiguity. In a communication which has to be concise, ambiguities are not to be avoided entirely. Consider Halliday's well-known example 'Dogs must be carried,' and the man in the London tube who, having read the sign, wondered whether he had to carry a dog before he was allowed to enter the train. The sign posted in German libraries *Bitte leise sprechen* 'please speak softly' is ambiguous in the same way. It can either mean: "please speak, and in a soft way so" or "if you must speak, please do it softly". If pronounced, stress assignment would disambiguate (*bitte leise SPRECHEN* forcing the former, and *bitte LEISE sprechen* the latter interpretation), but signs carry no intonation. German *Nichtraucher* has a preferred (a person who never smokes) and a dispreferred (a person who is not smoking presently) interpretation. In spite of this ambiguity, signs in German trains simply restrict certain coaches to *Nichtraucher*, and they definitively do not exclude smokers from sitting there, as long as they have not lit a cigarette. For ambiguities in pictograms, see von Heusinger (2000).

We may call the sign constellation a "text grammar", on analogy to the kind of systems we study in the following chapters, and contrast it with other "text grammars." In our analogy, grammars are just types of text construction. Above, in the recipe situation, all maxims are apparently equally ranking, but in sign situation, *AMBIGUITY is violated for the sake of brevity. In OT, such a situation is expressed by ranking higher the constraint responsible for the violation of another constraint. We have the ranking expressed in (3a).

(3) a. BEBRIEF >> *AMBIGUITY
b. *AMBIGUITY >> BEBRIEF

All other maxims play no role in the conflict, and, provided there is no evidence to the contrary, they are still ranked equally. For other text sorts, the ranking in (3b) seems relevant. Official documents, laws, by-laws, and other juridical texts are cases in point. At least in Germany, they tend to be long and complex, because the need to avoid ambiguity has the highest priority. In such a case, it takes more time to express what is meant. If written by a lawyer the sign in the London tube could possibly read "In case you have a dog with you, please carry it."

Returning to the cooking recipe, it is not at all clear whether the ranking (3b) should be preferred in this situation as well. After all, the recipe could be formulated with a list of key words in order to respect BEBRIEF, but then of course a good deal of ambiguous situations are prone to arise.

The relative ranking of the maxims in (2) is also relevant when the "grammar" of the texts is enriched by further content constraints on text composition. Such a situation is discussed by Grice himself. Suppose you have to write a text in which you must not say anything negative – because you are legally bound, as you are when you write official reports about employees in Germany, or because you have to write a letter of recommendation (for a candidate who you know is inappropriate). Given a high rank to the principle NOTHINGNEGATIVE, the maxims *FALSE and INFORMATION/RELEVANCE are likely to get into conflict. In such a situation, one might write that the candidate has a nice handwriting, or that he is always punctual. This piece of information may be true and therefore fulfill *FALSE, but it clearly violates INFORMATION and RELEVANCE since one does not deliver as much information as is required, and also irrelevant pieces of information In OT, this text type may be characterized by a ranking in which *FALSE dominates INFORMATION and RELEVANCE.

(4) Report/Letter of recommendation situation

NONEGATIVE > *FALSE >> INFORMATION, RELEVANCE

The text type of dinner table politeness also has a high rank of NONEGATIVE. Even if it is true, you should not say: "That soup was too salty!" Now, RELEVANCE ranks higher than *FALSE, and given the rules of politeness are what they are, it is considered more appropriate to say "Oh, what a tasty soup" than "The soup has just the correct temperature."

The examples discussed so far may be considered text grammars, and they resemble the phonological or syntactic systems to be discussed below in a number of respects: they are governed by a set of maxims/constraints that seem to have a grounding in external factors, and different text sorts/languages may be described as arising from different rankings of these constraints. But the examples discussed so far differ from syntax and phonology in a crucial way: differences in pre-existing goals to be achieved (be friendly to your hosts, do not lie to your colleagues) make different rankings intrinsically more appropriate for the relevant situation. Typically, this is not true for grammars: there are no reasons apart from historical accident that make the grammar of comparative formation in German and Fula different.

Our final example is similar to the one just discussed. The maxim of INFORMATION requires that the speakers make their contribution as informative as required. It does not say anything about the truth of the information, but just about the quantity. In other words, it says that more information is better than less. On the other hand, *DETAIL says that you should not give too much detail. If there is

a scale of informative content, the higher you are on the scale, the better you satisfy QUANTITY, and the less you say the better you satisfy *DETAIL. In a way these two constraints counterbalance each other. Imagine a situation where you are asked where you spent your last vacation, and let's decide that you spent it in Paris. In the context of the question, you may answer "Paris", but there are higher spots in the scale of informativeness, like Hotel de la Poste in Paris. If you choose to answer by the latter, your answer is better on QUANTITY. There are also lower spots on the scale, like France, or Europe. These satisfy *DETAIL better, even if they are not very good on informativeness. Having to choose between answering the above mentioned question with "Hotel de la Poste in Paris" and "France," there is no doubt that giving additional information is the better choice. So that the ranking of the two relevant constraints is as in (5).

(5) Vacation text grammar
INFORMATION >> *DETAIL

In the recipe situation, it is absolutely crucial that the constraints are ranked in the order given in (5). Otherwise you may be confronted with vague descriptions of how to bake a birthday cake. And probably in nearly all situations, except for the pictogram one, the order (5) makes sense.

By closer examination, the first impression that all maxims can be fulfilled at a time has turned out to be wrong. We have no place to develop this here, but it is pretty clear that there can be no communication situation in which all maxims can be ranked on the same level.

Turning now to the content of the treatise and the place it is meant to occupy in the larger book, the first chapter reviews the reasons why a theory of grammar should integrate conflicts as one of its most important components. We will show that, indeed, even though many grammatical models have been constructed in such a way that they could avoid the explicit acknowledgement of conflicts, conflicts and their resolution have in fact accompanied linguists in one form or another since the very beginning of grammar. This chapter also introduces the architecture of OT with the help of an example coming from the sentence accent in German. The second chapter reviews the kinds of conflicts which OT expresses best. The first kind of conflict is the competition between markedness and faithfulness constraints, the former driving the linguistic forms into the direction of structural simplicity and the latter being the ones responsible for lexical contrasts. The second kind of conflict involves the irreconcilable requirements of different sorts of markedness. The third kind is the alignment principles which account for linearization and boundaries. Universality and free reranking is the topic of the third chapter. By defining a set of universally valid constraints, OT

predicts that all languages use the same ingredients, but in different constellations. Individual grammars are the products of reorderings of the universal constraints. Finally, in the fourth and last chapter, an equally important component of OT is discussed w.r.t. its predictions, namely the way OT takes decisions as to which candidates are the winners of the linguistic competitions. Decisions are categorical: candidates are just winners or losers. Furthermore, there is just one optimal and grammatical candidate, which is by definition the candidate which fulfills the constraint ranking best.

While the first part of the book focuses on motivating the key aspects of OT, the purpose of its second part is a different one. When OT is applied to a larger domain of data, a number of data classes can be identified for which OT, apparently, offers no satisfactory solution. Several theoretical amendments have been proposed for OT, such as bi-directional optimisation, sympathy, tied constraints, etc. In the second part, we try to assess the necessity and feasibility of such extensions of classical OT – by discussing empirical issues such as ineffability, gradedness, and opacity, that is, by focusing on empirical domains about which classical OT has little to say if anything. We will argue for a very conservative view of Optimality Theory, which avoids compromising on OT's crucial characteristics: hierarchy-based resolution of conflicts between universal constraints.

Chapter 1

Fundamentals of the OT approach to grammar

Summary of the Chapter

In this chapter, we introduce the overall architecture of OT. In order to do so, we motivate the core assumption of Optimality Theory (OT), that grammatical rules or principles are violable and that they stand in conflict with each other.

Conflicts among rules or principles are ubiquitous in language and it is only natural that they have been part of the history of grammar from its very beginning, when Pānini wrote the first treatise on grammatical problems we know of. However, apart from explicit discussions of the role specificity plays, linguistics never really tried to formulate a general model of conflict resolution for language (but the rule ordering component of early generative grammar may be an exception). OT makes the claim that the resolution of the different types of conflicts identified in 1.1 follows a general scheme, but in addition, OT also subscribes to further views such as the claim that all principles used by natural language grammars are universal, and that languages differ only in the way they resolve the conflicts between these universal principles. These fundamental assumptions of OT will be sketched in section 1.2: The grammar generates candidates for inputs, and the choice of the optimal candidate, the grammatical one, is made on a "lexicographic" base, using constraint ranking only. In order to make OT accessible even for linguists with little or no background in OT, we develop the argumentation and the methods of this model with the help of an example progressively increasing in complexity: default sentence stress in German.

1.1 Motivating a conflict-tolerant type of grammar

The main characteristics which distinguishes the optimality-theoretic grammatical model from others is the explicit conflict resolution component. Optimality Theory makes the fundamental claim that no linguistic object, no syllable, no word, and no sentence, manages to satisfy all requirements imposed by the principles or rules of grammar. Universal Grammar consists of a set of principles, called constraints, expressing universal linguistic tendencies and included in all

languages. These principles are as simple and general statements as possible, and may be in conflict with each other. This is so because these principles may impose incompatible demands on specific linguistic entities. Individual grammars must resolve these conflicts, and they do so by ranking the constraints. OT's basic insight is that even if grammars are driven by the same principles, these principles are ranked in different ways in different languages. A constraint A may be ranked very high in some language L1, so that linguistic outputs always or nearly always fulfil A, and ranked lower in another language L2, constraint A can be crucially dominated by a constraint B, conflicting with A, to the effect that linguistic outputs fulfil B and violate A. In such a case, A and B conflict with each other and the conflict is resolved differently in L1 and L2. This kind of conflict is usually visualized by means of so-called tableaux in the OT literature. Tableau (1) shows the ranking of A and B in L1. A is ranked higher than B. Suppose now that several candidates compete for the best output. Candidate 1 fulfils A but violates B and candidate 2 violates A and fulfils B. Violation of constraints by candidates is shown by an asterisk in the corresponding cell. There may be other candidates participating to the competition, which violate or fulfil both constraints, or which violate the constraints more than once, but we concentrate here on Candidates 1 and 2. In L2, the ordering of the two constraints is reversed: B dominates A. What does the ordering of A and B mean for L1 and L2? In L1, Candidate 1 is the grammatical output, the optimal candidate, whereas in L2, it is candidate 2 which wins, in each case the candidate fulfilling the highest constraint. Thus, even if both linguistic principles expressed by constraints A and B are present in both languages, OT's prediction is that their ranking has an influence on the choice of the best candidate.

(1)

	A	B
$Cand_1$		*
$Cand_2$	*	

(2)

	B	A
$Cand_1$	*	
$Cand_2$		*

Concrete examples for this kind of conflicts, where languages make different choices, will be discussed in detail in the next chapter, where it will become apparent that faithfulness and markedness constraints can influence the inventory

of segments and of other structures. Some languages have marked segments, like nasal vowels, affricates, gutturals, or different Cases, whereas some other languages chose to eliminate these marked patterns. The difference between languages is a consequence of the ranking of the faithfulness constraints for marked structures w.r.t. markedness constraints against them. The conflictual nature of the constraints implies that they are violable. If optimal candidates are to be identified at all, and if constraints are ranked, their violability is a necessary property of the theory.

Since violability may be considered a weakening of the empirical import of principles, the conflictual nature of the grammatical principles should be well-motivated. Thus, before we give a general overview of OT in 1.2, we focus on the necessity of introducing conflicts as a central architectural device into the grammar.

Conflict resolution components have been part of grammar from its very start. We discuss Pānini's grammar below, but other linguists have also focused on conflicts, like Bech (1955/1957) for instance. Some grammatical models mention conflicts explicitly, like OT and other constraint-based models, whereas others do not highlight them as a crucial part of their linguistic approach and confine themselves to using technical means implying the resolution of conflicts. In reviewing some conflicts found in language, it is useful to distinguish at least two types: (a) the so-called "elsewhere" organization of rules or constraints, for which there seems to exist a principled answer to the question as to how the conflict should be resolved, and (b) those conflicts in which the particular choice among the conflict resolution options seems arbitrary.

The 'Elsewhere' case (also called *proper inclusion* (Anderson 1983, Fanselow 1991), *Paninian conflict resolution* (Prince & Smolensky 1993, McCarthy & Prince 1993a), *intrinsic rule ordering* (Bach 1964, Kenstowicz & Kisseberth 1977), *specific before general*, *specificity principle*)) refers to situations in which a well-defined subclass of linguistic expressions is affected by a certain process, but elsewhere – in most environments – a different, general process applies. Such cases, some well-known, abound at the interface between morphology and phonology, like the following:

- The indefinite article in English is *a*, except if the word following it begins with a vowel. In this case it is *an* (*a pear, an apple*).
- The singular definite article in French is *le* or *la*, except if the following word begins with a vowel. In this case it is *l'* (*la poire* 'the pear', *l'orange* 'the orange').

- The Dutch diminutive is *-tje*, except if the base ends with a syllable containing a short vowel and closed by a sonorant. In this case it is *-etje*. (*banann-tje* 'little banana' vs *bol-etje* 'little cup')
- The first obstruent in the second member of a compound in Japanese becomes voiced (Rendaku), except if the morpheme already contains a voiced obstruent. In this case it remains voiceless (/ori-kami/ –> [ori-gami] 'folding paper' /yama-tera/ –> [yama-dera] 'mountain temple').
- The German dorsal fricative is realized as the palatal fricative [ç], except after a back vowel. In this case, it surfaces as the velar [x][1] (*Buch* [x] 'book', *ich* [ç] 'I').

In principle, these facts may be captured in a variety of ways. Taking the German dorsal fricative as an example, the contexts in which the two variants appear might be just listed in two different sub-rules, as in (3) Given that their domains of application are disjoint, they do not stand in conflict with each other.

(3) Distribution of the palatal and velar dorsal fricative in German

a. The German dorsal fricative is realized as a velar [x] after a back vocoid (vowel and glide).

b. The German dorsal fricative is realized as a palatal [ç] after a front vowel, a consonant, and at the beginning of a word.

From a descriptive point of view, (3) is unobjectionable, since the surface facts are captured. But there are other reasons that militate against (3). The format of the description makes it appear an *accidental* property that the set of environments listed in (3) accounts for all possible contexts, and that the application domains do not overlap, so that two sounds are in complementary distribution. Furthermore, while the context of rule application in (3a) is a natural one from a phonological point of view, the composition of application domains in (3b) does not reflect this property. By working with rules such as (3), one would have to concede that phonological processes may be conditioned in a purely arbitrary fashion, in a manner we cannot hope to understand, an unsatisfactorily conclusion.

Obviously, we can do much better! The complementary distribution of the dorsal fricative's variants becomes apparent when the contexts in which the rules are applied are organized along the following line: a distinction is made between a *particular* case accounting for the more specific environment (4a), and a *contextless*, 'Elsewhere' case (4b). If we understand (4) as a system of generative

[1] Or as the uvular [χ], depending on the vowel involved (see Wiese 1996) or Féry (2000).

rules, we derive the correct results if the more specific rule is applied before the general one. If we understand (4) as a system of constraints, correct predictions are made if a more general principle is inapplicable in the domain in which it competes with a more specific statement.

(4) Distribution of the palatal and velar dorsal fricative in German

a. The German dorsal fricative is realized as velar [x] after a back vocoid.

b. Otherwise it is realized as [ç].

Complementarity and exhaustiveness follow as well because (4b) implies that (4) affects all instances of the dorsal fricative, and because of the rule/constraint interaction just mentioned. The arbitrarity problem for the non-particular rule has also disappeared: the set of relevant contexts in fact needs not be listed in a phonological rule. The context set on (4b) is no natural class by itself – it arises when a natural process carves out a set of environments from the totality of possibilities.

The two rules in (4) may be said to be *in conflict* with each other for certain elements in the following sense. If we drop the explicit "elsewhere/otherwise" restriction in (4b), rules (4*a) and (4*b) impose different and incompatible requirements on how a dorsal fricative should be realized when if follows a back vowel. The general rule is unrestricted, it could also apply in the domain of the more specific rule.

(4*) Distribution of the palatal and velar dorsal fricative in German

a. The German dorsal fricative is realized as velar [x] after a back vocoid.

b. The German dorsal fricative is realized as [ç].

The conflict is then resolved by the 'Elsewhere Condition' – a principle of grammar with the sole purpose of resolving conflicts among rules and principles. The two formulations in (5) reflect the derivational and representational interpretations it can be given, respectively.

(5) Elsewhere Condition

If the domain of application of rule/constraint R is properly included in the domain of application of S, then S cannot be applied where R can be (then R must be applied before S has a chance to be applied).

It is important to note that the problem addressed by the 'Elsewhere Condition' is not confined to a tiny aspect of German phonology. The list of examples given above already shows that "elsewhere" phenomena are ubiquitous in phonology and at the phonology-morphology interface. They rather seem to characterize the sound system quite generally. A principled solution is called for, and the "elsewhere" principle is a concrete and successful proposal. But note that we have thereby motivated the existence of *violable* constraints in grammar (what the more general rule requires is not always respected), and of *conflicts* between rules and constraints.

It comes as no surprise that 'Elsewhere' cases can be found in the other domains of language as well. Consider e.g., plural formation in Polish, an instance of core morphology. For grammatically non-neuter nouns, the primary distinction is whether the noun ends in a soft (palatalized) consonant (then, plural is formed by adding *–e* [e]) or not (then, the ending is *–y* [ɨ]. This need not involve an "elsewhere" situation, but note that there is an exception to the latter rule: masculine personal nouns form their plural by adding –i [i]. We certainly prefer (6) to (6*) as a characterization of Polish plural formation.

(6) Plural Rules for non-neuter nouns ending in a hard (nonpalatalized) consonant

a. Add -*i* if the noun is masculine personal.

b. Add -*y*.

(6*) Plural Rules for non-neuter nouns ending in a hard consonant

a. Add -*i* if the noun is masculine personal.

b. Add -y, if the noun is feminine, or if it is masculine and not personal.

Classical examples of 'Elsewhere' effects in the syntax involve the influence of the lexicon on syntactic structure. Thus, simple transitive verbs combine with accusative objects in German, but there are lexical exceptions (for verbs like *helfen* "help"). Again, an elsewhere formulation of the case assignment rules as in (7), in which (7d) states the "elsewhere" situation, is superior to a listing as exemplified in (8), where (8d) lists the verbs assigning accusative in the same way as verbs assigning another case. (8) does not express at all that there is an accusative assignment *rule* in German that is quite different in nature from the other case marking options.

(7) Case Rule for Objects in German

a. the object case is nominative for *bleiben* "remain", *sein* "be"and *werden* "become"

b. the object case is genetive for *gedenken* "commemorate"… *bedürfen* "require" (7 verbs)

c. the object case is dative for *helfen* "help" …. *gehören* "belong" (perhaps 100 entries)

d. the object case is accusative

(8) Case Rule for Objects in German

a. the object case is nominative for *bleiben* "remain", *sein* "be"and *werden* "become"

b. the object case is genitive for *gedenken* "commemorate"… bedürfen "require" (7 verbs)

c. the object case is dative for *helfen* "help" … *gehören* "belong" (perhaps 100 entries)

d. the object case is accusative, for *lieben* „love" … eruieren "find out" (perhaps 25,000 entries) and all other newly formed verbs

One need not confine one's attention to irregularities in order to find more cases of the 'Elsewhere Condition' in the syntax. Direct objects show no morphologically visible case marking in Hindi or in Spanish – from a theoretical point of view, they seem to bear accusative case. Particles *–ko* and *–a* are added in Hindi and Spanish, respectively, when the direct object is animate and specific. This marking is otherwise used for the dative.

(9) a. Juan busca un libro
John looks for a book

b. Juan busca a una secretaria
John looks for a(specific) secretary

c. Juan da el libro a uns secretaria
John gives the book to a secretary

(10) represents the "elsewhere" version of the case rules for direct objects, while (11) is a rule system that tries to avoid conflicts.

(10) Direct Object Case Rule for Hindi and Spanish

a. Assign dative case to specific animates.

b. Assign accusative case.

(11) Direct Object Case Rule for Hindi and Spanish

a. Assign dative case to specific animates.

b. Assign accusative case to animates if not specific, and to inanimates.

That the syntax-semantics interface may be particularly prone for exemplifying 'Elsewhere' effects has been suggested frequently. For the interaction of reflexive and personal pronouns, this has, e.g., been proposed by Bouchard (1983), and his ideas were taken up in one of the other form in the subsequent literature (Koster 1988, Fanselow 1991, Burzio 1998, Reuland 2001, Wilson 2001, among many others). The key observation is that reflexive pronouns ("anaphors") and personal pronouns ("pronominals") are in complementary distribution when it comes to expressing a coreference relation, as (12) illustrates, where co-indexation is to be interpreted as expressing presupposed co-reference of two argument positions.

(12) a. John_i saw himself_i/*him_i in the mirror

b. John_i prefers very much for himself_i/*him_i to win the race

c. John_i prefers very much for Mary to caress him_i/*himself_i

d. John_i hopes that he_i/*himself_i will win

It is a commonplace that anaphors must find their antecedent in a local domain. When the antecedent is too far away syntactically, a personal pronoun must be used. (11) might seem fine, coming close to what Chomsky (1981) proposed.

(13) Coreference Rule

a. Use an anaphor to express coreference between a and b, if a and b are close enough

b. Use a pronominal to express coreference between a and b, if a and b are not close enough

Consider now German in this respect. First, we observe that German has no genitive anaphor, and it has no anaphoric version of the possessive pronoun.

(14) a. er gedenkt *sich seiner (selbst)
he commemorates himself his-gen

b. er liebt *sich's seine Frau
he loves refl his wife

The rule in (13) would thus have to be modified along the lines given in (15).

(15) Coreference Rule

a. Use an anaphor to express coreference between a and b, if a and b are close enough, and if b is neither a genitive nor a possessive.

b. Use a pronominal to express coreference between a and b, if a and b are not close enough, or if b is a genitive or a possessive.

Furthermore, since first and second person paradigms have no anaphor, a pronoun is used instead (16a), and in certain dialects, the polite form is constructed with a pronominal as well, since there is no polite anaphor. Thus in Bavarian, the anaphor *se* cannot be used for the polite form, and is replaced by the pronoun *eana* in the standard anaphoric contexts. As a consequence, (15) should be elaborated as in (17).

(16) a. ich liebe mich
I love me
'I love myself'

b. hom's eana/*se hi-gsetzt?
have-you.polite you/*yourself seated
'Did you take a seat?'

(17) Coreference Rule

a. Use an anaphor to express coreference between a and b, if a and b are close enough, and if b is neither a genetive nor a possessive, and if a is not 1st or 2nd person, or a polite 3rd person.

b. Use a pronominal to express coreference between a and b, if a and b are not close enough, or if b is a genetive or a possessive, or if a is 1st or 2nd person, or a polite 3rd person.

We could continue along these lines: when the antecedent is not a subject, there are special conditions for the use of the anaphor, and typically, these special conditions imply that a pronoun replaces the anaphor when the anaphor is blocked. Instead of adding more and more complications to *both* rules (17), Bouchard proposes an extremely simple and attractive idea that can be formulated as in (18): the pronoun is the "elsewhere" default:

(18) Coreference Rule

a. Use an anaphor to express coreference between a and b, if a and b are close enough, to the extent that an anaphor with the relevant feature specification can be found in the lexicon.

b. Use a pronominal to express coreference between a and b.

The domain of coreference illustrates a further aspect of specificity: the "elsewhere-interaction" need not be confined to two rules or constraints. More rules may interact in a nested fashion. Thus, consider the distribution of Dutch pronouns and anaphors, as discussed in Koster (1988) and Reuland (2001), which makes use of three indexical expressions to express coreferentiality: anaphor *zichzelf*, reflexive *zich* and pronoun *hem*.

(19) a. *Oscar* haat *zichzelf*/**zich*/**hem*
Oscar hates himself

b. *Oscar* voelde [*zich*/*zichzelf/ *hem* wegglijden]
Oscar felt SE/him slide away

c. hij zegt dat Marie van *zich/*zichzelf/hem houdt
he said that Mary loves him

Zichzelf is used when the corefential elements are strictly local (roughly, when they are co-arguments) and only if the anaphor is referential *Zich* is used whenever the anaphor has no referential content, and when slightly more relaxed locality conditions are fulfilled (19a,b). With a number of exceptions that can easily be explained away (as Everaert 1988 has shown), *zich* cannot intrude into the domain of *zichzelf*. The relation between the pronoun and the two anaphors

adds little new to our picture. If the rules/constraints in (20) are applied in the given order, a neat description arises.

(20) Coreference Rule for Dutch

a. Use *zichzelf* for expressing coreference between a and b if they are clausemates and if b is referential.

b. Use *zich* for expressing coreference between a and b if a and b are close enough.

c. Use a pronoun for expressing coreference.

Obviously, the domain of application for *zichzelf* is a proper subset of the application domain of *zich,* and the pronominal's domain is a superset of the latter. Therefore, the 'Elsewhere Condition' makes apparently correct predictions concerning which means of expressing coreference must be used in which context.

In the past years, Blutner (1999) and others have proposed to analyze certain effects of pragmatic implicature in terms of conflictual rules. Consider the following example: when we say that we are happy, we claim that our emotional state (serotonine level?) is above a certain threshold s. When we say that we are unhappy, we claim that our emotional state is below a certain threshold k. When we say that we are not unhappy, we typically intend to express that our emotional state lingers somewhere between s and k. How does that come about? If *unhappy* means being below k, the negation of *unhappy* should be able to refer to *any* state above k, not just those below s. But note we have a word (a lexicalized expression) for everything that is above s, viz. *happy*.

In its literal interpretation, *not unhappy* is applicable in a larger domain (everything above k) than *happy* (everything above s), so an 'Elsewhere' effect might explain why we cannot use *not unhappy* in all cases where *happy* is applicable, too.

The insight that principles covering a specific set of data are applying before more general ones has been acknowledged in most linguistic theories, and it has been so from the very beginnings of the scientific inquiry of language. The Indian linguist Pānini, who is sometimes regarded as one of the world's first grammarians, has based part of his theory on conflicts between application of rules and environments where rules were blocked. The following discussion is based on Kiparsky's (2002) insightful interpretation of Pānini's grammar, and more particularly of Astādhyāyī, a system of about 4000 grammatical rules of

Sanskrit. The rules of Astādhyāyī are grouped together, so as to build classes, and the expressions which they have in common are omitted from the particular rules and are instead stated for the whole group of rules at the beginning of a heading. This permits the formulation of very simple rules, but also implies that rules are not understandable in isolation.

Kiparsky illustrates simplicity in Panini's grammar with the formation of patronymics, the derived nouns which designate the descendant of the person expressed by the base. For our goals, it is sufficient to examine the way blocking is accounted for in Astādhyāyī, and to show how conflicts are explicitely identified and resolved in the Paninian rules. The general (elsewhere) patronymic suffix is *–aN*, phonologically *–a*, with a diacritic *N* which causes strengthening of the stem's initial syllable; general rules accent the suffix, and truncate the stem-final *–a* before it. A descendant of Upagu is called *Aupagavá*. *–aN* is just the most general (elsewhere) patronymic suffix. Morphological and phonological effects of suffixing *–aN* are expressed by a number of rules, themselves parts of a more general suffixation process. We list the rules here for completeness. Rules (21a,b) govern suffixes in general, (21c) states that the first syllable of the suffix is accented. This latter rule is the 'Elsewhere' case, and is active only in case no particular rule is applicable.

(21) Rules for suffixation

a. 3.1.1 pratyayah
suffix-Nom
An item introduced in (earlier) rules is (termed) *pratyayah*
'suffix'

b. 3.1.2 para ca
following-Nom and
'and (an item introduced in earlier rules) follows'

c. 3.1.3 ādyudāttás ca
initial-accent-Nom and
'and has initial accent'

The next set of rules in (22) governs the specific properties of *taddhita* suffixes ('secondary' denominal derivational suffixes) and their patronymic subclass. Rules (22a–c) contribute to rule (22d) that suffixes *–aN*.

(22) Rules for *taddhita* suffixes

a. 4.1.1 nyāpprātipadikāt
Ni- āP-stem-Abl
'after (an item ending in the feminine suffixes) *Ni, āP ,* or (after) a nominal stem'

b. 4.1.76 taddhitāh
taddhitah-NomPl
'denominal suffixes'

c. 4.1.82 samarthānām prathamād vā
semantically-related-GenPlfirst-Abl optionally
'After the first semantically related stem [marked by a pronoun in the genitive case in each rule], optionally [preferably].

d. 4.1.83 prāg divyato 'n
up-to divyati-Abl *aN*-Nom.
Up to rule 4.4.2. the accented *taddhita* suffix *aN* is added after the first semantically related nominal stem [marked by a pronoun in the genitive case in each rule].'

Exceptions to suffixing *–aN* to express patronymicity are for example a group of stems ending in *–pati* 'lord', which form their patronymics with the suffix *–Nya* (*Prajāpati –> Prāyāpatya*). This class has also an exception: a class of compounds in *–pati*, which require *–aN* again, rather than *–pati* (*Á svapati –> Ásvapatá*). Pānini groups the two *–aN* rules together, orders the *–Nya* rule afterwards, and achieves in this way maximal concision.

(23) Rules for the exceptions

a. 4.1.84 ásvapatyādibhyás ca (83 aN) (82 samarthānām…)
Ásvapati -etc-Abl
'The *taddhita* suffix *–aN* is also added after the first syntactically related stem which belongs to the class *Ásvapati* etc.'

b. 4.1.85 dityadit ī adityapatyuttarapadān nyah (82 samarthānām…) …
diti-aditi-aditya-pati-second-word-Abl nya-Nom
The *taddhita* suffix *–Nya* is added after the first syntactically related stem *Diti* … and after the compounds in *–pati*.

Grammatical architectures allowing for "elsewhere" types of conflict resolution seem widely accepted. The 'Elsewhere' principle is, however, not sufficient to account for all conceivable cases of rule interaction in language. We will see below that the 'Elsewhere Condition' cannot explain all kinds of rule interaction because the processes in question often do not have nested but rather merely overlapping domains of application. But let us focus first on more disturbing facts pointing to the conclusion that 'anti-elsewhere' effects (the reverse of 'Elsewhere' effects) are very common: the more general rule applies in the domain of the more restricted one. This is a surprising situation if specific conditions always take precedence over general ones on principled grounds, for reasons intimately linked to the architecture of grammar. One well-known example of 'anti-elsewhere' effects from phonology comes from the variation in German between two pronunciations of a word ending with underlying /ng/. In one variant, the word *Zeitung* 'newspaper' is pronounced [tsaɪ̯tʊŋ], as the result of assimilation of /n/ to the dorsal articulation of /g/ and deletion of /g/. This happens in standard German and in most other dialects of German. In the other variant, *Zeitung* is pronounced [tsaɪ̯tʊŋk], as a consequence of assimilation of /n/ to the dorsal articulation of /g/ and Final Devoicing.

Part of the processes leading to the two surface forms [tsaɪ̯tʊŋ] and [tsaɪ̯tʊŋk] are conflicting with each other. Dorsal assimilation of the nasal ([n] –> [ŋ]) applies in both cases, but the fate of the dorsal stop depends on additional factors. Either it deletes or it is devoiced. In derivational phonology, this variation has been analyzed as a consequence of a conflicting ordering of the rules of Final Devoicing and g-deletion. If g-deletion applies first, we are left with just [ŋ] and nothing else happens. In the alternative ordering, Final Devoicing applies first, leading to [ŋk]. In this case, nothing else happens either since the environment of g-deletion is not present after having changed [g] to [k]. The two orderings are shown in (24) and (25) (see also Wurzel 1980).

(24) Derivation of [tsaɪ̯tʊŋ] (more frequent in the standard variant of German)

a. Nasal assimilation: [nasal, coronal] –> [dorsal]/ _ [–cont, dorsal

tsai͡tuŋ –> tsai͡tuŋg

b. g- deletion: g –> ø / ŋ _ $\left\{ \begin{matrix} C_o \\ \text{schwa} \end{matrix} \right\}$

tsai͡tuŋg –> tsai͡tuŋ

c. Final Devoicing: [+voiced, –cont, –son] –> [–voiced] / _]$_\sigma$

(25) Derivation of [tsai͡tuŋk] (more frequent in the Northern variant of German)

a. Nasal assimilation: [nasal, coronal] –> [dorsal]/ _ [–cont, dorsal]

tsai͡tuŋ –> tsai͡tuŋg

b. Final Devoicing: [+voiced, –cont, –son] –> [–voiced] / _]$_\sigma$

tsai͡tuŋg –> tsai͡tuŋk

c. g- deletion: g –> ø / ŋ_ $\left\{ \begin{matrix} C_o \\ \text{schwa} \end{matrix} \right\}$

The contexts in which g-deletion applies affecting only [g], a voiced obstruent, can be understood as being a subset of the situations in which final devoicing applies, affecting syllable-final voiced obstruents in German in general. In other words, an 'Elsewhere' effect is to be expected. The rule of g-deletion should always take precedence over final devoicing – which it does not in those dialects in which [tsai͡tuŋk] is acceptable.[2] In (24) the rules of g-deletion and of Final Devoicing are ordered as expected: particular before general. But in (25), the relevant rules apply in the reverse ordering: general before particular.

OT has no problem with these 'anti-elsewhere' effects. Since both the general and the specific conditions are expressed by means of constraints, and since

[2] The context of g-deletion, as posited in (24) and (25), also includes pre-schwa environments, but in the derivational analyzes of Hall and Wiese (1996), this context is syllable-final at a certain point of the derivation, before schwa-insertion.

constraints can be ordered differently in different languages, both "elsewhere" and "anti-elsewhere" are predicted to be possible outcomes. But a grammar in which the "elsewhere" facts are necessarily ordered after the particular cases have applied cannot account for "anti-elsewhere" results straightforwardly.

A further kind of conflict resolution arises when two rules that potentially apply to the same element are not in an "elsewhere" relationship, but in another type of relation: optionality. From the perspective of Universal Grammar, this third category of conflictual cases involves *true optionality* (*free variation*), and these cases have also figured prominently in discussions concerning conflicts in language. The conflict resolution happens in an arbitrary manner.

Truly optional cases are, perhaps, hard to come by within a single language, since it is generally possible to find some contextual effect influencing the choice of one or the other option. In his influential study, Labov (1966) showed that free variation should be understood as an artefact of sociological features like social status, age, sex, etc, and that the realization of allophones in free variation are largely predictable on the basis of statistical calculation. In this book, we are interested in accounting for free variation, but not in the factors influencing the alternation, or the diachronical evolution, even though we acknowledge the interest of such studies. In chapter 10, some remarks are introduced about which kind of influence on the allophonic variation should be part of the grammar proper. At this point, it is enough to mention that free variation is common in language or dialect variation (and may be difficult to deal with in Optimality Theory).

An example of intralinguistic optionality comes from the Ukrainian paradigm in (26) illustrating the point for case assignment (Sobin 1985, Shevelov 1963). The 'Elsewhere Condition' leads us to expect that the more general case assignment rule must not apply in the domain of more specific rules. This prediction is not borne out in Ukrainian. Simplifying a bit, Ukrainian is like any other Slavic language in allowing that subjects bear nominative case. In a passive, the subject may bear accusative case. As in Russian or Polish, there are special case rules for negative clauses, according to which direct objects and subjects in passive clauses may bear genitive. All rules predict different outcomes, they make conflicting statements on what the surface form of a case in a positive and a negated passive should be. The three rules stand in an "elsewhere" relation, with the genitive contexts being a subset of the accusative contexts, and the latter being a subject of the constellations that accept nominatives. One would thus expect that subjects of negated passive can show up with genitive case only. This expectation is not borne out, as (25) illustrates. The conflict between the rules is resolved by attributing them equal importance – the choice between the various case possibilities is optional (Paslawska, p.c.)

(26) a. Cervku bulo zbudovano v 1640 roc'i.
church -fem.acc be-past.neuter build-part-neut in 1640

b. Cervka bula zbudovana v 1640 roc'i.
church -fem.nom be-past.fem build-part-fem.sg in 1640

c. Cervkvy ne bulo zbudovano v 1640 roc'i.
church -fem.gen be-past.neuter build-part-neut in 1640

d. Cervka ne bula zbudovana v 1640 roc'i.
church -fem.nom be-past.fem build-part-fem.sg in 1640

A second example brings us back to the syntax-semantics interface. Above, our conclusion concerning the distribution of anaphors and pronouns was that they are in complementary distribution, reflecting an 'Elsewhere' effect. This is true for standard anaphors only, however. In certain languages like Icelandic, Chinese, or Japanese, reflexive pronouns may have a (subject) antecedent in a higher clause – but this enlargement of the domain of anaphoric binding does not necessarily lead to a corresponding reduction in the options for pronominals!

(27) Xiaoming$_i$ yiwei Xiaohua bu xihua ziji$_i$/ta$_i$
Xiaoming think Xiaohua not like self/him
'Xiaoming thinks that Xiaohua does not like self'

These examples suggest again that the general rule *does* sometimes apply in the domain of the particular rule. The "elsewhere" approach therefore cannot be the *only* solution for conflicts in languages. Whether a conflict is solved in terms of the 'Elsewhere Condition' is not determined on a principled basis – rather it is subject to interlinguistic variation.

Further examples of optionality within single languages are listed here:

- In a set of environments, the auxiliary may but need not be contracted in English (*he will kiss Mary – he'll kiss Mary*)
- The question phrase may but need not be preposed in French matrix questions (*tu as vu qui? – qui as-tu vu?* 'Who have you seen')
- The location of the negation on the auxiliary alternates with its location on the complement in sentences like *Mary did not see anything* vs. *Mary saw nothing*

- In a variety of languages, high vowels in the first position in a hiatus can be alternatively realized as a full vowel or as a glide: The realization of *Radio* in German alternates between [ʀa.di.o] and [ʀa.djo], nuage 'cloud' in French between [ny.aʒ] and [nɥaʒ].

True optionality in the context of typological variation is a frequent phenomenon and has occupied an important place in the OT literature up to now. We will sketch only two cases of this kind here, as further examples will appear at numerous places in this book.

The first example is the well-known variability in the tolerance languages have for codas in syllabification. English has numerous examples of syllables with codas (cap, hat, lamp, etc.) whereas Hawaiian has not a single one. This language does not tolerate codas at all, and furthermore has a very limited segment inventory. It accordingly changes the segmental and syllabic structure of loanwords in a considerable way. The following examples come from Gussenhoven & Jacobs (1998:43).

(28) Adaptation of English loanwords in Hawaiian

a.	Albert	–>	ʔalapaki
b.	ticket	–>	kikiki
c.	wharf	–>	uapo

Prince & Smolensky (1993) assume that the constraint responsible for the dispreference for codas (called NoCoda) conflicts with the tendency for segments to be realized without change, as well as without deletion or insertion of additional segments. In Hawaiian, clearly, it is NoCoda which wins the competition, since vowels are inserted after segments which are in the coda position in the original language. In English, codas are allowed and it is the tendency for underlying segments to be realized without change which wins.

It is important to notice that these two tendencies are not in an "elsewhere" relationship. It is not the case that being truthful to an underlying segment is more general or more specific that the prohibition against codas. In fact the two needs are expressed as unrelated, and it is only in some situations that they conflict and can lead to opposite results.

The second case of typological variation is the well-known difference between languages which systematically locate wh-phrases at the front of the sentence and languages with wh-phrases in situ. A language of the first kind is English and a language of the second kind is Japanese.

(29) Wh-Phrases

a. English: What did you tell me?

b. Japanese: John-wa nani-o kaimasita ka
John-TOP what-Acc bought Q
'What did John buy?'

The example has been discussed a number of times in the OT literature, first by Grimshaw (1997). The conflict observed in this case is whether the need to place the wh-word or phrase in the position in the sentence in which its scope is most clearly visible is higher ranking than the desire to avoid movement and traces. In English, the former solution is chosen, whereas in Japanese the opposite ranking is the right one.

We will return to these examples in more detail in chapter 2.

If one concedes that principles may stand in conflict with each other, and may be violated, because 'Elsewhere' effects, 'anti-elsewhere' effects and true optionality imply just that, it is a natural idea to extend this analysis to other types of rule interaction. The next chapter will concentrate on the third kind of conflicts between grammatical principles and propose a typology of conflicts along the lines of OT conflict resolution.

1.2 Overview of OT

Having motivated the presence of conflicts in grammar, we now give a sketch of the architecture of Optimality Theory as it was developed in Prince & Smolensky (1993). Some of OT's architectural decisions derive in straightforward way from the need to resolve grammatical conflicts. Others involve matters of execution that might have also been arranged differently. Thus, OT is a constraint-based/representational model of grammar. As Prince & Smolensky point out in their 1993 manuscript introducing the theory, major aspects of OT might be formulated in rule based/derivational systems, too. Further aspects of the organization of human languages have not been addressed in a principled way at all. Thus, OT can be applied as a grammatical model for phonology, morphology, syntax, semantics, and pragmatics, as well as for facts concerning language acquisition, language loss, diachronic change, etc. The nature of the interaction of, say, the morphological and syntactic constraints is not the subject of *uniform* modeling in Optimality Theory, so that approaches in which morphological and syntactic constraints are part of a single set of constraints (see Bresnan 1999 for

instance) coexist with approaches in which syntactic structures are built on the basis of syntactic principles only, and are later interpreted morphologically by a separate constraint system (see Trommer 2002 for instance).

Let us now introduce some basic technical aspects of OT. Imagine L, a native speaker of Mandarin Chinese, who has learnt German from books and grammars but who has never heard it spoken. She only knows that German is a stress-based language and that main accent in a declarative sentence is realized with a falling tone. She would like to find out how regular sentence stress is assigned, something that her textbooks and grammars have not taught her. She comes across a German speaker, S, who utters single sentences, sometimes a bit out of context. S first says (30) (the word bearing the falling tone for main accent is written in small caps).

(30) Guten TAG 'Hello'

From this utterance, L deduces that noun phrase accent is final and formulates an according OT principle (31).

(31) ALIGN-R -NP (NP, main stress, Right)
In a noun phrase, accent is final.

This constraint expresses that the right edge of a noun phrase falls together with the right edge of a main stress. Formally, (31) might be interpreted as a generative statement that picks any noun phrase and guarantees that the accent goes to the rightmost position. In the OT model developed by Prince and Smolensky, the approach is truly constraint based. The grammatical entity is picked by the *evaluation component* of the grammar, called EVAL. EVAL consists of a set of constraints, by which possible outputs are evaluated. These possible outputs are called *candidates.* In the context of (30), the candidates would seem to be *Guten Tag*, GUTEN TAG, *Guten TAG* and *GUTEN Tag*. OT candidates are generated by the generative part of the grammar, the ***GEN*** component. Thus the grammar, as conceived by Prince & Smolensky consists of two steps. First, candidates are generated by the function GEN, which delivers for each input a certain number, possibly an infinite number, of candidates. The input, for the moment best conceived as a kind of underlying representation, can have little structure, possibly underspecified, but which can also be completely specified. The principle Richness of the Base (Prince & Smolensky 1993), guarantees that the input can take any form it wants, as long as it is a linguistic entity. It is the task of the second component of the grammar, EVAL, to eliminate bad inputs, and to

determine, according to the constraint ranking of the language under consideration, which are the grammatical forms.

Principle (31), together with a constraint to the effect that just a single word has main stress in an NP, correctly picks (30) as the winning candidate. The other candidates violate the constraint and are eliminated.

The next sentence uttered by S (32) confirms (31) since accent in the NP *zwölf Stunden Verspätung* is final.

(32) Mein Flugzeug hatte zwölf Stunden VERSPÄTUNG
my plane had 12 hours delay
'my plane was 12 hours late '

Sentence (32) allows L to hypothesize a further constraint to the effect that sentence accent is also final. L formulates constraint ALIGN-R.

(33) ALIGN-R (sentence, main stress, Right)
In a sentence, main stress is final.

S goes on with his monologue and the next sentence forces L to revise her grammar.

(34) Ich bin nämlich gestern von Berlin nach BEJING geflogen
I am Particle yesterday from Berlin to Bejing flown
'I flew from Berlin to Bejing yesterday'

The main stress in this sentence is not compatible with ALIGN-R, since the penultimate word bears main accent, but according to ALIGN-R, the last word *geflogen* should be stressed. What could be the difference between (32) and (34) leading to the different position of main stress? The sentence (34) is in the present tense and has an inflected verb in the V2-position, whereas (34) is in the perfect tense, with a final unstressed past participle, and the participle is unstressed. One idea might be to split (33) into two different principles. (33) itself would be restricted to sentences with a simple tense, whereas a further principle restricted to sentences with complex tenses would place the main stress into the position immediately preceding the verb complex at the end of the sentence. L, however, decides to keep the predictions of ALIGN-R but adds a new constraint to her grammar. She observes that there is a similarity between both sentences. It lies in the fact that the accented word is an internal argument of the verb. L formulates

STRESS-ARGUMENT in (35), which requires that an argument of a verb bear main accent.

(35) STRESS-ARGUMENT
Accent is on a verbal argument.

At this point, L has already constructed a series of hypotheses of how sentence stress is assigned in German. Her system derives the correct predictions concerning (34) if ALIGN-R must be respected to the extent only that it does not get in conflict with STRESS-ARGUMENT. When the two constraints predict divergent results, STRESS-ARGUMENT decides. The constraints of grammar thus have a different weight. EVAL is not just a set of principles and constraints, it consists of *a hierarchy (an exhaustive total ordering) of the constraints in question*. The notation 'a >> b' means 'a has a higher rank than b'. (36) is thus what L needs in her account of German stress placement.

(36) STRESS-ARGUMENT >> ALIGN-R

The candidate structures are evaluated relative to the hierarchy EVAL, and the candidate with the best violation profile is the grammatical one. A first formal definition of grammaticality can be found in (37).

(37) A candidate c generated by GEN from Input I is *grammatical* iff all candidates c' it competes with are such that c' violates the highest constraint c from EVAL on which c and c' differ, whereas c does not.

Concentrating on S's last sentence "*Ich bin nämlich gestern von Berlin nach Bejing geflogen*", *a tableau*, like the ones we saw in (1) and (2), can be drawn which visualizes the evaluation procedure and identification of the grammatical candidates in a transparent form. In an OT tableau like (38), constraint ordering is represented by linear organization. The leftmost constraint is the highest one; to its right comes the next one in the hierarchy, and so on. Since L has only two constraints to rank, this presents no particular problem. The next step is to mark violations of the constraints by the candidates with the help of asterisks. Candidate a violates ALIGN-R, but not STRESS-ARGUMENT. Candidate b violates STRESS-ARGUMENT. Such a tableau allows us, as well as L, to evaluate candidates and decide which one is optimal. STRESS-ARGUMENT is the highest-ranking constraint, and for this reason, candidate b, violating it, is eliminated from the competition. This is indicated in tableau (38) with an exclamation mark following the asterisk. Candidate a is identified as grammatical. It violates a constraint, yet

respects the highest constraint on which the two candidates introduced so far differ.

(38) Tableau

	STRESS-ARGUMENT	ALIGN-R
a. Ich bin […] von Berlin nach BEJING geflogen		*
b. Ich bin […]von Berlin nach Bejing GEFLOGEN	*!	

Obviously, however, the set of candidates to be considered is not exhausted by what we find in tableau (38). There are many other options for locating stress in the sentence in question, as indicated in tableau (39). Candidates d through g are correctly eliminated by STRESS-ARGUMENT, but with candidate c, we seem to run into a problem, since its constraint violation profile does not differ from a, the only grammatical option.

(39) Tableau

	STRESS-ARG	ALIGN-R
a. Ich bin nämlich gestern von Berlin nach BEJING geflogen		*
b. Ich bin nämlich gestern von Berlin nach Bejing GEFLOGEN	*!	
c. Ich bin nämlich gestern von BERLIN nach Bejing geflogen		*
d. Ich bin nämlich GESTERN von Berlin nach Bejing geflogen	*!	*
e. Ich bin NÄMLICH gestern von Berlin nach Bejing geflogen	*!	*
f. Ich BIN nämlich gestern von Berlin nach Bejing geflogen	*!	*
g. ICH bin nämlich gestern von Berlin nach Bejing geflogen	*!	*

This problem is solved in OT by assuming that some constraints are *gradient* in the sense that we can determine how often they are violated. Main stress is one position away from the right edge in a, but is separated by two elements from clause final position in c. Assume that this difference matters – we might e.g., interpret ALIGN-R as meaning: an element E must not follow the main stress of a sentence. We enter a violation (mark) for each phrase that violates ALIGN-R under that interpretation. The result is represented in tableau (40). The candidate c

through g successively accumulate violation marks for principle ALIGN-R, because the distance of the main stress to the clause final position increases. For candidate c, the additional violation of ALIGN-R is critical (as indicated by the exclamation mark), while the fate of the other candidates has already been settled by the higher constraint stress-argument. It is a useful custom to shade those cells in a tableau which are irrelevant for the outcome of the evaluation procedure. The winning candidate is identified by the sign ☞.

(40) Tableau

	STRESS-ARG	ALIGN-R
☞ a. Ich bin nämlich gestern von Berlin nach BEJING geflogen		*
b. Ich bin nämlich gestern von Berlin nach Bejing GEFLOGEN	*!	
c. Ich bin nämlich gestern von BERLIN nach Bejing geflogen		**!
d. Ich bin nämlich GESTERN von Berlin nach Bejing geflogen	*!	***
e. Ich bin NÄMLICH gestern von Berlin nach Bejing geflogen	*!	****
f. Ich BIN nämlich gestern von Berlin nach Bejing geflogen	*!	******
g. ICH bin nämlich gestern von Berlin nach Bejing geflogen	*!	*******

Of course, we need to slightly change the definition of grammaticality in order to formally arrive at the result that we have motivated informally with tableau (40).

(41) A candidate c generated by GEN from input I is *grammatical* iff all candidate c' it competes with are such that c' violates the highest constraint c from EVAL on which c and c' differ *less often* than c does.

With tableau (40), we introduced several properties of OT constraints. Constraints are *ranked*. OT assumes that the ranking is *always* determined on a language-particular basis. We saw that German has the ranking shown in (36), but in a language with regular final stress, like French, the two constraints STRESS-ARGUMENT and ALIGN-R would be ranked the other way round, thus ALIGN-R >> STRESS-ARGUMENT, in order to guarantee that candidates with nonfinal accent are eliminated before STRESS-ARGUMENT has a chance to choose among the remaining candidates. In a sentence like (42), the participle is final and stressed. The argument *l'argent* 'the money' does not bear the nuclear stress because it is not final. This is illustrated in tableau (43).

(42) French final stress

Marie a rencontré le peintre auquel l'argent a été REMBOURSE.
Mary has met the painter to whom the money has been reimbursed

(43) A French tableau

	ALIGN-R	STRESS-ARGUMENT
☞ a. auquel l'argent a été REMBOURSE.		*
b. auquel L'ARGENT a été remboursé.	*!	
c. AUQUEL l'argent a été remboursé.	*!*	

The constraint reranking faculty is a crucial property of OT which will be the theme of chapter 3. In fact, all possible rankings derivable from some set S of constraints are claimed to possibly figure as natural languages.

(44) Let S be a set of constraints used in language L. Then any ordering of S yields a possible natural language L'.

In other words, constraints are freely rerankable.

The third property of constraints, which has been introduced in the context of tableaux/s (40) and (43), is their violability. In principle, all constraints are violable, even if they are the highest in a hierarchy. It was also demonstrated with ALIGN-R that constraints can also be gradiently violated, though most constraints induce binary decisions on violations. A candidate that violates a gradient constraint more often than another candidate looses to this second candidate.

We can now return to an observation that has already been made several times: the decision as to which candidate is optimal is made on a *lexicographic* basis, called like that because it is reminiscent of the way a lexicographer orders the words in a dictionary. There is an ordering of the letters of the alphabet (a, b, … x, y, z), and if two words differ by their first letter, it is this letter that decides which word comes first in the lexicon. If they have the first letter in common, the second letter determines their order – unless the two first letters of the words are once more identical. In such a situation, the third letter will determine order if it can do so, etc.. In OT, the decision between two candidates follows exactly the same strategy. If c and c' differ on the highest constraint C, it is their performance relative to C which determines which of the two is better. If they do not differ with respect to C, the next principle C' in the hierarchy is chosen. If the two

candidates do not have the same number of violations with respect to C', then C' will determine between then, otherwise, one proceeds to the next lowest constraint C'' etc.

Envisaging the entire competition, pairwise evaluations is able to identify one optimal candidate. Let us limit the discussion to the first three candidates of tableau (40). In tableau (43) a wins over b, c wins over b, and a wins over c. The winner of the entire competition is a.

(45)

	STRESS-ARG	ALIGN-R
a. Ich bin […] von Berlin nach BEJING geflogen ~ b. Ich bin […]von Berlin nach Bejing GEFLOGEN	b	a

(46)

	STRESS-ARG	ALIGN-R
a. Ich bin […] von Berlin nach BEJING geflogen ~ c. Ich bin […] von BERLIN nach Bejing geflogen		a cc

(47)

	STRESS-ARGUMENT	ALIGN-R
b. Ich bin […]von Berlin nach Bejing GEFLOGEN ~ c. Ich bin […] von BERLIN nach Bejing geflogen	b	cc

Let us now return to our imaginary interaction of S and L. S goes on with her monologue and says the following sentence.

(48) Man hat mich KONTROLLIERT.

This sentence has a verbal argument, but it is not stressed. The argument is a reflexive pronoun and as such, a function word with less prosodic weight than a content word. L decides on the basis of sentence (48) that such words repel stress. She needs a new constraint ranked higher than STRESS-ARGUMENT, since otherwise an argument would be stressed regardless of its status as function or content word.

(49) UNSTRESSEDFUNCTIONWORD
Function words are unstressed.

Tableau (50) illustrates how this new constraint forces stress to be final again. Since *mich* cannot be stressed because of high ranking UNSTRESSEDFUNCTIONWORD, and there is no other internal argument which could be stressed, the decision is taken by ALIGN-R.

(50)

	UNSTRFUNCWORD	STRESS-ARG	ALIGN-R
☞ a. Man hat mich KONTROLLIERT		*	
b. Man hat MICH kontrolliert	*!		*
c. Man HAT mich kontrolliert		*	*!

An important characteristic of OT competitions is visible here. A candidate can be chosen on the basis of a relatively low-ranking constraint, here ALIGN-R, which was shown in to play no role in the decision between candidates containing content words as arguments. The situation illustrated has been called Emergence of the Unmarked (TETU) by McCarthy & Prince (1994). It will be illustrated in more detail in chapter 3, but now it suffices to say that even low-ranking constraints can take decisions given that all higher-ranking ones are not in a position to decide, be it because more than just one candidate fulfil them, or the reverse situation is the case, the candidates that could still win are doing equally bad on them.

Coming back to (32), we observe that the optimal candidate for this competition does not violate any constraint, as testified by tableau (51). This is of course because only three constraints are active. As soon as more constraints are involved, winning candidates always violate some constraints.

(51)

	UNSFUNCWRD	STRESS-ARG	ALIGN-R
a. Mein Flugzeug hatte 12 Stunden VERSPÄTUNG			
b. Das Flugzeug HATTE Verspätung		*!	*
c. Das FLUGZEUG hatte Verspätung			*!*

So far, we have left it open which candidates c, c', ... compete with each other relative to EVAL, in other words, whether we opt for a containment of a correspondence version of OT. In the correspondence theory, every output can in principle be evaluated for every input (see chapter 2 for a detailed exposition of the two versions). But of course, the well-formedness of "*Ich muss mich BEEILEN*" does not interfere at all with "*Ich bin nämlich gestern von Berlin nach BEJING geflogen*", although the former sentence violates less principles than the latter. This is because these two sentences are optimal candidates of different inputs, and for this reason, even if they participate in the same competition, as would be predicted in the correspondence theory, they do not really compete with each other. High-ranking constraints eliminate all candidates which differ in an obvious way from the input. In the containment version of OT, the two sentences do not even compete with each other. Only those candidates which contain the input are part of the evaluation set.

Is the result obtained in this section at odds with our reflection concerning the 'Elsewhere'effect? Consider first a situation in which the more specific principle S outranks the general principle G, as was shown with the distribution of the German dorsal palatal and the distribution of the reflexives and the pronouns in several languages in section 1.1. Whenever S is applicable, it governs the wellformedness of candidates, whenever it is not, G decides. This is the typical "elsewhere" constellation. But if G outranks S, S has *never* a chance to exert any visible effect in the language in question. We called this effect 'anti-elsewhere'. Thus, G is always applicable when S is, but since it has a higher rank, it is always G that decides. It would be represented, e.g., by a language in which reflexive pronouns never surface, because the principle that licenses them is outranked by the general principle that coreference be expressed by the use of pronominals. The language in question cannot be distinguished empirically from a language in which S is non-existent. Thus, assuming that S nevertheless exists (but does no harm) is sufficient to maintain the claim of free rerankability, although it may not always be forced upon us on empirical grounds. To conclude this chapter, we

illustrate 'Elsewhere' and 'anti-elsewhere' effects and show that the same grammar that was illustrated in detail for true optionality also account for these effects.

First, consider 'Elsewhere' effects as those illustrated in section 1.1 for pronouns and anaphora, which both can be used to express coreference. Imagine a (possibly fictive) situation in which anaphora (reflexives) are used in a smaller set of domains than pronouns. A constraint is necessary to express that anaphors cannot refer to an antecedent across specified domain boundaries (a finite clause for instance). In other words, anaphora can corefer inside of certain grammatical domains but not beyond. This is expressed by (52a), where [ ] stand for variable domains according to the language considered. Leaning on insights formulated by Burzio (1988), Fanselow (1991), and translated by Burzio (1998) into OT, we may postulate that reflexives are less costly than pronouns, because they have less grammatical features. Constraints a, b and c ranked in that ordering express the "elsewhere" relationship: Since anaphora are less costly, they are preferred as long as they are licenced in the domain in which they stand. Otherwise (elsewhere), pronouns are chosen.

(52) Constraints expressing pronouns and anaphora coreferentiality

a. BARRIER: A reflexive must be bound within the domain of [ ]

b. *PRONOUN: No pronoun (Pronouns are costly.)

c. *ANAPHOR: No anaphors (Anaphors are costly.)

Tableaux (51) and (53) illustrate some language in which anaphors have a limited range. Tableau (53) shows how the constraints allow an anaphor to express coreferentiality as long as the anaphor and its antecedent a are in the limit of the domain defined by BARRIER. When the antecedent and the coreferent are separate by the crucial barrier, the candidate using an anaphor is no longer allowed because it violates BARRIER, and candidate b, using a pronoun, is optimal.

(53) Tableau: Coreferentiality by means of reflexives or pronouns

	BARRIER	*PRONOUN	*ANAPHOR
☞ a. [a… anaphor]			*
b. [a…pronoun]		*!	

(54) Tableau: Coreferentiality by means of reflexives or pronouns

	BARRIER	*PRONOUN	*ANAPHOR
a. [a… [anaphor …]]	*!		*
☞ b. [a…[pronoun …]]		*	

In a language without anaphora, the same constraints are used, but with a different ranking. Now the particular constraint BARRIER has no effect anymore, regardless of its ranking, because the crucial ranking between *ANAPHOR and *PRONOUN disallow the emergence of anaphora altogether. Since it should be evident that across the boundaries specified by [ ], anaphora are also disallowed, we so not draw the tableau for this case. The particular case – anaphor under special circumstances – is obliterated entirely as a consequence of the prohibition of anaphora in all domains.

(55) Tableau: Coreferentiality by means of pronouns only

	BARRIER	*ANAPHOR	*PRONOUN
a. [a… anaphor]		*!	
☞ b. [a…pronoun]			*

'Elsewhere' and'anti-elsewhere' effects are thus accounted for by the same means: ranking of the relevant constraints, though the rankings are different. Furthermore the cases just illustrated for the distribution of pronouns and anaphora are similar to what has been illustrated for stress assignment in German and French, a true optionality effect. Here too, it was shown that reranking is all we need to obtain the desired effect.

Conflicts and their resolution have been shown to be pervasive in all domains of linguistics, and no grammatical theory has managed to avoid reference to them, (see chapter 5 on the precursors of OT). OT puts the emphasis of the analysis on exactly this aspect and accounts for different kinds of conflicts by the same means: reranking of the constraints. The next chapter examines the constraints and provide a classification as well as a review of their interactions.

Chapter 2

Conflicts

Summary of the Chapter

The issue of the kind of conflicts that are pervasive in natural language grammars is a good starting point for a detailed consideration of Optimality Theory. There are various driving forces in the constitution of grammars: lexical contrasts must be maintained in the interest of expressivity and contrast (this yields a set of *faithfulness* constraints), linguistic structures differ in terms of simplicity, possibly with respect to a number of dimensions (the realm of *markedness* constraints), and finally, the individual elements in a structure must be ordered relative to each other, and different levels of representations must be tied to each other (the domain of *alignment*). Conflicts between faithfulness and markedness, and among markedness and alignment principles, simply cannot be circumvented. Because it focuses on the resolution of these conflicts, Optimality Theory is the proper architecture for a theory of natural language.

2.1 Faithful or simple? A first source for conflicts

The exploitation of the descriptive potential of conflicts among linguistic principles and their resolution is the key feature of Optimality Theory. Although conflicts have always figured in linguistic analyses (see the preceding chapter and chapter 5), one may wonder why a theory of language should focus on them so heavily.

In a certain sense, conflicts always reside in the eye of the beholder only: the preceding chapter has revealed that conflict-free formulations of grammatical principles are certainly possible, to the extent that one is willing to complicate each linguistic rule by a list of exceptions. Admitting conflicts and optimality in the theory of grammar may make its overall architecture more complex, but the spirit of Optimality Theory is that this is a price one should be willing to pay: it allows a considerable simplification of the formulation of the individual principles of grammar.

But there is more that motivates conflicts among grammatical principles than just overall simplicity considerations. Quite independent of whether we prefer

lists of exceptions to conflicts, one may wonder why there is this potential of contradictions among the constraints of grammar, which seems absent in most (if not all) artificial languages like the ones used in computer sciences. A simple consideration suggests that conflicts are unavoidable at least in phonology, but this line of reasoning can be extended to syntax – at least to a certain extent. It is the conflict inherent to the interaction between the need to maintain lexical contrasts and simplicity. We will illustrate this with the example of final devoicing in German.

Sound and sound combinations differ with respect to articulatory and perceptive difficulty. It requires some effort by the speaker to maintain a voicing contrast in syllable final position on articulatory grounds. The activities of the vocal cords which yield voice cannot fully unfold unless a vowel follows. Furthermore, producing voice on word-final obstruents is not really worthwhile, since voicing in this position is typically not perceived well by the hearer.

Quite in line with this, German has Final Devoicing (FD) of stops and fricatives. Consider, for example, the data in (1) and (2) involving predicative and feminine singular nominative forms of adjectives in German. The predicative adjectives in the second column end in an obstruent, and these obstruents are always voiceless. The feminine adjective is formed by adding a schwa-vowel to the predicative form. When we consider the fem. sg. column of (1) and (2), we realize that both voiced and voiceless obstruents precede the added word-final schwa, depending on the adjective considered.

(1)

	predicative	fem. sg.
cold	kalt [kalt]	kalte [kaltə]
ill	krank [kʀaŋk]	kranke [kʀaŋkə]]
informal	salopp [zalɔp]	saloppe [zalɔpə]
ripe	reif [ʀa͡if]	reife [ʀa͡ifə]
hot	heiss [ha͡is]	heisse [ha͡isə]

(2)

	predicative	fem. sg.
cowardly	feig [faɪ̯k]	feige [faɪ̯gə]
nice	lieb [li:p]	liebe [li:bə]
stupid	doof [do:f]	doofe [do:və]
silent	leis (colloq.) [laɪ̯s]	leise [laɪ̯zə]

Words pronounced with a final voiced stop, like /g/ or /b/, or with a final voiced fricative, like /z/ or /v/, are absent in German, and they seem to be so on principled grounds: there is a simplicity principle in the theory of the sound system that rules out voiced obstruents in the syllable final position. It can be formulated as in (3).

(3) FINALDEVOICING (FD)
No voiced obstruent at the end of a syllable.

As for (2), the following description is standard in generative phonology (Wurzel 1970, Wiese 1986, among others): the "underlying" form of the adjective *lieb* 'nice' ends in a voiced bilabial stop /b/. Whether [b] can be realized on the phonetic surface, in the light of (3), is a function of the result of syllabification. In case a schwa follows, the principles of syllable formation yield [li:bə], which respects (2) on trivial grounds. If /li:b/ is used in isolation, [b] would be syllable final. In this case, it is replaced by [p].

If these facts are linked to simplicity, one would expect that German is not the only language that shows this type of alternation, and this expectation is borne out. (4a) illustrates the effects of FD in Russian, Dutch FD is exemplified in (4b), and Polish data can be found in (4c).

(4) Final devoicing in other languages

a. Russian nouns

Nom. sg.	Gen. pl.	
ryba	ryp	'fish'
pobeda	pobet	'victory'
groza	gros	'storm'
lyʒa	lyʃ	'soul'

b. Dutch nouns

Nom.sg	Nom. pl.	
web	webben	‘web, webs’
pat	padden	‘toad, toads’
kluif	kluiven	‘bone, bones’
muis	muizen	‘mouse, mice’

c. Polish verbs

1sg.	Imper. sg.	
rob’e	rup	‘do’
vodze	vutʃ	‘lead’
otvoʒe	otvuʃ	‘open’

Many further examples come from languages from all over the world: Catalan, Turkish, Indonesian, Ngizim and Czech all have final devoicing for example. Many more languages can be added to this list when other kinds of neutralization are considered, such as the contrast between plain, aspirated, voiced and glottalized obstruents which is neutralized to plain (or to a subset of the contrasts) in the syllable (or stem or word) final position. Korean is a well-known case of final neutralization, as are Sanskrit and the Athapaskan languages (e.g., Navajo, Slave). Final devoicing, envisaged as a reduction of the contrasts made on the obstruents, is therefore by no means an idiosyncrasy of German. It reflects a very frequent pattern of natural language. Recall that there are articulatory and perceptual reasons for not maintaining a voicing contrast in a syllable-final position. The question arises as to why we find languages in which voiced obstruents do surface in the coda of a syllable. The answer lies in the need to realize some lexical contrasts. As we shall see below, if articulatory and perceptual simplicity always determined what we can say, the expressive power of language would be unduly reduced.

FINALDEVOICING is a *markedness* constraint. Voiceless obstruents in syllable-final position are easier to produce and better to perceive than their voiced counterparts, they replace them in quite a number of languages and they appear earlier in language acquisition. Let us therefore assume that the grammars of *all* natural languages contain the constraint FD (3), which penalizes syllable-final voiced obstruents.

However, FD is not truly universal in the sense of being an unviolable principle of grammar. In some languages, its effects are confined to stops and affricates, while fricatives are not affected. In other languages like Turkish, final devoicing applies to the native vocabulary without exception, but can be inactive for loanwords, as (5) illustrates.

(5) Turkish

FD in native words	[kɑnɑt ~ kɑnɑdɯ]
No FD in foreign words	[etyd ~ etydy]

And finally, in a language like English, FD never triggers voicing alternations: *has*, *big*, and *love* are pronounced with a final voiced obstruent. While FD is not truly universal in the sense that it is always respected at the phonetic surface of every word in every language, some its effects can even be observed for speakers of languages like English that maintain the voicing contrast in syllable-final position. Some children acquiring English start with a phonological system in which they *devoice* syllable-final consonants (Ingram 1974, 1989, Smith 1973)! Thus, final devoicing is not "learnt" in German, rather, it is *unlearnt* in English. Even in English, the expression of the voicing contrast is sometimes shifted to a length difference in the vowel preceding the consonant (see Repp 1982).

Markedness principles like FD cannot apply globally in all languages for the following functional reason. A markedness constraint demands that a certain phonological dimension be realized with a specific feature. In a syllable final position, the feature [+voice] should be absent. When this markedness constraint applies, the relevant dimension can no longer be used for creating and maintaining lexical contrasts. As an example, the contrast between /rad/ *Rad* 'wheel' and /rat/ *Rat* 'counsel' is abolished in German: both words are realized as [rat]. Generalizing to other markedness constraints, *all* phonological dimensions seem to involve a simplicity scale in one way or the other. Thus, we can observe that all languages have CV syllables, while other syllable types are less common. At the beginning of the acquisition of phonology, children often go through a phase in which they utter CV syllables only. For these reasons, among others, CV is likely to be the simplest syllable type and a markedness principle thus requires syllables to be of the CV type. Moreover, the best onset for a syllable is a plosive, and the best vowel is an /a/. Consequently, the set of acceptable syllables would be reduced to a few items if *all* markedness conditions had to be respected at once (presupposing that this is possible at all, but see next section) and since words should not consist of too many syllables (they are probably best consisting of just a bisyllabic foot), unconditional respect for markedness would severely restrict the expressive power of language to just a handful of words. Some additional examples of markedness constraints are provided here, which will be introduced systematically in the pages to follow.

(6) Markedness constraints
   a. *NASALVOWEL: No nasal vowel.
   b. *VOICEDOBSTRUENT: No voiced obstruent.

c. ONSET: A syllable has an onset.
d. NOCODA: A syllable has no coda.
e. NOM: Each sentence contains a nominative noun phrase.
f. SUBJECT: All clauses have a subject.

There is an inherent conflict between markedness/simplicity and expressivity in language, and at least in phonology, this conflict affects every single item. It can even be considered to be one of the driving forces of phonology.

The maintenance of expressivity is coded relative to a lexicon, the set of morphemes of a language. Let us suppose for the moment, as sketched above, that the lexicon specifies the *underlying representations* (UR) of the words, which can be conceived of as matrices of phonological features. The task of the phonological part of the grammar is to compute the phonetic form corresponding to each UR.

Expressivity can be maintained if the construction of phonetic representations attempts to preserve the contrasts specified in the underlying representations. The output representations should be faithful to the underlying representations – or, in OT terms, the output should be faithful to the input. We may start with a very general principle like FAITH, which requires that phonetic realization of a word to respect the specification of features in the input. FAITH, has a lower rank than FD in German, because voiced obstruents of the input fail to be realized overtly in word- and syllable-final position. The ranking is just the other way round in English; here, FD never has any visible effect, because FAITH is more important.

(7) FAITH
Respect lexical specifications

To be more precise, the lexical representations of morphemes (or morpheme combinations) constitute the *input* of the generation process GEN. Inputs (*lieb, lieber, hard, harder*) are subjected to the GEN component, which consists of a set of rules performing operations like change the voicing specification of a segment, or syllabify the sequence of sounds. The rules of GEN are applied to the input in all conceivable combinations, yielding at least the candidates listed in the left row of the tableau in (8) and (9). FD and FAITH are ranked differently in English and German. Thus, as (8) illustrates, FD violations can be fatal for a candidate in German, but, as (9) shows, they are not in English.

(8) FINALDEVOICING is ranked higher than FAITH in German

/lieb/	FD	FAITH
li:b	*!	
li:p		*

/liebe/	FD	FAITH
li:.bə		
li:.pə		*!

(9) FAITH is ranked higher than FINALDEVOICING in English

/hard/	FAITH	FD
hard		*
hart	*!	

/harder/	FAITH	FD
har.der		
har.ter	*!	

(8) and (9) pretend that there is only *one* faithfulness constraint, FAITH, which may conflict with markedness constraints like FD. This simplistic approach is incompatible with the fact that the generation procedure mapping lexical inputs onto phonetic representations can do more than just change a voicing feature of an input. These operations will have different effects on the output; many of them will yield different ways of circumventing an FD violation in German. If there were only one general constraint that penalized all deviations from the input to the same degree (as FAITH does), we could not make a choice between the different ways of circumventing the FD violation. Empirically, this would not be adequate.

What are the other operations that GEN can perform, in addition to changing the specification of the voicing feature? For example, consonants may be *deleted* by GEN, as an examination of Catalan adjectives reveals. Like German, Catalan also forms feminine adjectives by adding a schwa to the underlying representation of the masculine form. Thus, we can observe pairs such as *petit – petit´* 'small', *curt -curt´* 'short', and *blank–blankə* 'white'. Not all pairs involve an added *–ə* only. There are also pairs such as *pla-planə* 'plane', *al-altə* 'high', and *ket-əkestə* 'this' in Catalan. The feminine version of the adjective not only has an additional vowel, but it also differs from the masculine form by an additional consonant. Note that the different behavior of *əket* and *pətit* suggests that we cannot compute

the shape of the feminine form by considering the output of the masculine form. Rather, we need to assume that the input of *pətit* is petit, while the input of *əket* is something like *akest.* When the feminine vowel is added, nothing happens, but in case it is not, a further constraint of Catalan phonology comes into play: there are no complex consonant clusters in the coda. This is expressed by the constraint NoComplexCoda (NCC) in (10). The ranking NCC >> Faith yields the correct results (as the tableau in (11) shows) if the complete deletion of phonemes is what Gen does to an input. We generate the pair *pla-plan* if syllables cannot end in an *n* in Catalan.

(10) NoComplexCoda (NCC)
*CC]

(11) Final Devoicing in Catalan

akest	NCC	Faith
əket		*
əkest	*!	

əkestə	NCC	Faith
əkes.tə		*!
əke.tə		

malalt	NCC	Faith
məlal		*
məlaltə	*!	

malalta	NCC	Faith
mə.lal.tə		
mə.la.lə		*!

Likewise, vowels may be added by Gen, as illustrated by plural formation alternations in English. The English plural is formed by adding an anterior coronal fricative to the stem. This consonant can be realized as /z/ or as /s/, depending on the voicing specification of the preceding consonant or vowel. Relevant examples are *kid/kidz*, and *cat/cats.* There is a further alternation that is of particular import here: we find [veiz-veizɪz] for *vase-vases* and [pleis pleisɪz] for *place-places*. Obviously, English phonology does not allow geminate consonants, and in order

to avoid a violation of the corresponding constraint, expressed in (12), a vowel is inserted between a stem-final coronal sibilant and the coronal fricative of the plural morpheme.

(12) NoGeminate
*CC, if the two consonants are identical.

The Catalan plural is also formed by adding *s* (see *llop-llops*, *torre-torres*), but [ʊ] rather than [ɪ] is added when two [s] would be neighbors in the plural: *gos-gosus, tros-trosus*.

We have observed, then, that Gen may perform different kinds of operations. It is obvious that the phonological problems discussed so far could be solved in various ways, given the richness of operations allowed by Gen. Thus, the FD problem can be circumvented by devoicing (yielding [li:p]), but no FD violation would be present either if the final consonant were deleted ([li:]), or if we added a vowel ([li:bə]). There are many ways in which *repair* of a structure violating a constraint may be effected. Grammars typically select one type of repair as the grammatical one. Only the first solution is the one German has opted for, but we cannot capture this if we only have one categorial Faithfulness constraint, as (13) shows:

(13) Final Devoicing in German

/li:b/	FD	Faith
li:b	*!	
li:p		*
li:.bə		*
li:.pə		*
li:.		*
i:		*
i.pə		*

Obviously, the problem with Faith stems from the fact that it is insensitive to the degree and nature of deviations from the lexical form. We can remedy this situation by assuming that there is a *family of different types of faithfulness*

constraints. They penalize specific aspects in which a candidate may be unfaithful to an input. As a first case, we may distinguish the MAX family of constraints from the DEP family. The MAX (for maximize) constraint MAX(A) requires that for every element of some type A in the input, there must be a corresponding element in the output. MAX(voice) requires that we should not delete any voicing specification in the input. DEP constraints (dependency or 'don't epenthesize'), on the other hand, rule out the insertion of elements of a certain type. DEP(ˊ) penalizes structures in which a schwa has been inserted. More formally, we may state the overall structure of the two constraint families as in (14), where IO stands for faithfulness between **I**nputs and **O**utputs. (This view presupposes that there are other types of faithfulness as well. This will be the topic of a later chapter.)

(14) a. MAX-IO (No Deletion):
Each segment of S_1 has a correspondent in S_2 (S_1 is input and S_2 is output).

b. DEP (A)-IO (No Epenthesis):
Each segment of S_2 has a correspondent in S_1.

MAX(voice) penalizes deletion of the voicing feature of a segment. The grammaticality of [li:p] suggests, then, that FD >> MAX(voice) in German.

(15) Addition of MAX(voice)

/li:b/	FD	MAX(voice)
li:b	*!	
li:p		*
li:.bə		
li:.pə		*!
li:.		*!
i:		*!
i:.pə		*!

Replacing FAITH by MAX(voice) is not sufficient to account for all possible repairs of final voiced obstruents, as (15) shows. The ungrammaticality of [li:] as

an output of /li:b/ suggests that MAX(C) (don't delete a consonant) is more important in German than FD.

(16) Addition of MAX(C)

/li:b/	MAX(C)	FD	MAX(voice)
li:b		*!	
li:p			*
li:.bə			
li:.pə			*!
li:.	*!		*
i:	*!		*
i:.pə	*!		*

(16) comes close to the desired result, but we still need to exclude epenthesis of, e.g., a schwa as a solution to the German FD problem. Placing DEP(ə) higher than FD yields the correct result: an output element must have a corresponding element in the input, which is not the case for the epenthesized schwas in (17). Note that the ranking of MAX(C) and DEP(ə) is not crucial for the data in (17). The only important ranking is that they are both higher than FINALDEVOICING. An irrelevant or not yet determined ranking of constraints is indicated by a dotted line between the columns.

(17) Addition of DEP(ə)

/li:b/	MAX(C)	DEP(ə)	FD	MAX(voice)
li:b			*!	
li:p				*
li:.bə		*!		
li:.	*!			*!
i:.pə				

The ranking we have arrived at is summed up in (18). It reflects the fact that a syllable-final voiced obstruent is repaired minimally, just by deleting a feature: [voice] is simply deleted from the featural representation of the obstruent. Other repairs, like epenthesis or deletion of a whole segment, are more costly since they would introduce more massive structural changes, and are thus eliminated earlier in the evaluation. How OT accounts for such preferences is the subject of chapter 3.

(18) MAX(C), DEP(ə) >> FINALDEVOICING >> MAX(voice)

Of course, our treatment of German FD is too simple in many respects. What happens to *ambisyllabic* consonants that occupy syllable-final and syllable-initial position at the same time? They do *not* undergo devoicing, and the account sketched here would not yield that result. But a detailed treatment of FD is not the issue here; FD just serves to illustrate the interaction of faithfulness and markedness principles (see Féry 2002 for a detailed analysis).

Faithfulness constraints are multileved principles that make simultaneous reference to inputs and outputs.Their role and scope is thus particularly clear in phonology, the part of grammar for which there is some consensus concerning the nature of inputs. It is relatively safe to assume that that inputs are made up of the lexical specification of the segmental or suprasegmental properties of the morphemes (the "underlying representation" of classical generative phonology). In the syntax, faithfulness principles may play a role as well, but it is more difficult to show that the contrast between faithfulness and markedness is as clear-cut as it is in phonology. There are two reasons for this, one of which is that the nature of inputs is less clear in the syntax. It is hard to be faithful if one does not know what to be faithful to.

There is, however, a domain in syntax in which faithfulness effects are obvious: respect for lexical idiosyncrasies. These are syntactically relevant properties that must be represented in the lexicon, and one would expect these to be reflected in actual syntactic representations. As predicted by OT, languages differ with respect to the extent that they are faithful to lexical exceptions, as a consideration of overt case marking suggests.

Many theories of case assume that the case of a noun phrase is licensed (checked/assigned) by some other element in the clause (see, e.g., Chomsky 1981). Several categories in a clause can do so. First, Infl licenses nominative case: in a finite clause, the subject bears nominative case. Below, we will call the constraint that forces subjects to bear nominative case AGREE, since a subject agrees with Infl in at least some morphological features, like person and number. It is exactly the argument that agrees with Infl which is in the nominative. Second,

there is a structural rule to the effect that verbs may license the accusative case of their complements, and third, individual verbs may combine with specific cases exceptionally: *helfen* 'help' requires dative case, and *gedenken* 'commemorate' genitive case.

(19) a. (dass) der Mann schläft
that the.nom man sleeps

b. dass er den Mann kennt
that he the.acc man knows

c. dass er dem Mann hilft
that he the.dat man helps

How can we describe the situation in (19)? By many criteria, the nominative is the most unmarked case, at least in a language of the German type. It is the most frequent case (it is used when there is only one argument to pick up case), it is the default case that shows up when no other case can be assigned by more specific rules, its formal marking is weak, etc. By these criteria, accusative is more marked, but it is still less marked than the dative. In German, regular dative shows up only when nominative and accusative have already been assigned. There are many languages that do not have a dative case distinct from the other two cases, and the dative is always highly marked morphologically. We may therefore posutlate two markedness constraints *ACC and *DAT (see Woolford 2001) .

(20) *ACC: Do not use an accusative case.
*DAT: Do not use a dative case.

Not only is the nominative case licensed by the Infl node but it is also the most unmarked case, the one which does not violate either of the two markedness constraints in (20).We therefore understand why nominative but not accusative or dative show up in (19a). On the other hand, there is no second nominative in double object structures. This is explained by UNIQUECASE/OCP (see Woolford 2001, Stiebels 2001). With the ranking *DAT >> *ACC, we can derive that objects are normally marked for accusative in simple transitive structures in German. It is only with ditransitive verbs that (regular) dative shows up.

(21) UNIQUECASE
Do not use the same Case twice within a single clause.

(22) Uniqueness of Case in German with one argument

	UNIQUENESS	*DAT	*ACC
ich kenne der Mann I.nom know the.nom man	*!		
ich kenne den Mann I.nom know the.acc man			*
ich kenne dem Mann I.nom know the.dat man		*!	
mich kenne dem Mann I.acc know the.dat man		*!	*

(23) Uniqueness of Case in German with two arguments

	UNIQUENESS	*DAT	*ACC
ich gebe ihm den Wagen I.acc give him.dat the.acc car		*	*
ich gebe ihn den Wagen I give him.acc the.acc car	*!		**
ich gebe er der Wagen I.give him.nom the.nom car	*!*		
mich gebe ihn den Wagen me.acc give him.acc the.acc car	*!*		**
…			

Quite a number of verbs impose lexical requirements on the case shape of their arguments, however. German verbs like *helfen* 'help' oder *folgen* 'follow' construct with dative rather than accusative objects. For this to be possible, we need to assume that such verbs bear a case specification in their lexical entry, and that MAX(CASE) (=Faith-lex of Woolford 2001) is ranked higher than the markedness constraints introduced so far. As mentioned above, *helfen* 'to help' is such a verb, assigning a lexical case in dative.

(24) Lexical Case

	MAX(CASE)	UNIQUENESS	*DAT	*ACC
ich helfe du I help you.nom	*!	*		
ich helfe dich I help you.acc	*!			*
ich helfe dir I help you. dat			*	
mich helfe dir me.acc help you.dat			*	*!

That MAX(Case) >> UNIQUENESS can be seen from the existence of verbs like *lehren* "teach" or *kosten* "cost" that are constructed with two accusative arguments.

German is a language in which MAX(CASE) has a very high rank. In other languages like Japanese (see Woolford 2001) or Faroese (see Fanselow 2000) it is less important. For example, this becomes evident when one considers passive formation. The crucial German examples are given in (25). The remaining argument of *kennen* 'know' switches from accusative to nominative, because the *ACC violation would no longer be justified by avoiding a Uniqueness violation. On the other hand, lexical dative case *is* retained in the passive because of MAX(CASE).

(25) a. dass der Mann gekannt wird
that the.nom man known was

b. dass dem Mann geholfen wurde
that the.dat man helped was

(26) Lxical Case in passive

	MAX(CASE)	UNIQUENESS	*DAT	*ACC
dem Mann geholfen wird the.dat man helped is			*	
den Mann geholfen wird the.acc man helped is	*!			*
der Mann geholfen wird the.nom man helped is	*!			

Faroese is diffferent. We observe a shift from lexical dative to nominative case in this language.

(27) Faroese Case

a. Teir hjálpa honum
they help him.dat

b. hann varð hjálptur
he.nom is helped

The markedness and faithfulness principles introduced so far do not yield this result, but the key observation can already be made: apparently, MAX(CASE) is overridden by some other constraint in Faroese. Faithfulness to lexical specifications is not absolute in all languages, not even in the domain of lexical exceptions. The markedness constraint that forces the shift to nominative in Faroese is easy to identify: it is the requirement that a sentence have a (nominative) subject, that is, that there be a noun phrase bearing nominative case with which the verb agrees for categories such as person and number. For example, we may formulate such a principle as in (28). The following tableaus show that the two different rankings of AGREE and MAX(CASE) yield the German (25b) and Faroese (27b), respectively.

(28) AGREE
In a finite clause, the verb must agree with a nominative noun phrase.

(29) Agree in German

	MAX(CASE)	AGREE
dem Mann geholfen wird		*
der Mann geholfen wird	*!	

(30) Agree in Faroese

	AGREE	MAX(CASE)
hann varð hjálptur		*
honum varð hjálptur	*!	

Lexical idiosyncrasies have to be captured by faithfulness constraints, and sometimes they are overriden by markedness constraints, but in general it seems that a situation in which markedness principles override lexical idiosyncrasies does not figure prominently among the constitutive aspects of syntax, though the role of faithfulness in syntax is, obviously, determined by decisions concerning what a syntactic input is.

For example, it can be argued that faithfulness constraints are responsible for what other approaches capture as *economy constraints*. Economy manifests itself in at least two respects: Expletives such as *there* or *it* cannot be inserted freely into syntactic structures, but only when they are unavoidable. Likewise, phrases and heads are not arbitrarily moved around in sentences – their movement always fulfills a purpose. Under the presupposition that the insertion of meaningless elements and movement are "costly" operations, the term "economy constraint" becomes interpretable. How are such economy constraints linked to faithfulness?

Inputs determine which candidates compete with each other in Optimality Theory. In syntax, constructions must have the same lexical meaning in order to belong to the same competition, so a syntactic input will have to specify the content words to be used in a sentence.

Which *content* words we use is *not* dictated by the grammar of the language, but rather by what we want to say. Their choice is not part of the evaluation component of grammar. In contrast, which function words we use is not so much determined by what we want to say, but rather by what the grammar prescribes. Functional elements that do not contribute to the meaning of a sentence are *not* part of the input. They are inserted by GEN. Consider, e.g., the data in (31) and (32) in this respect.

(31)	(qu')	il	a	été	dansé
	(that)	it	has	been	danced
	*(que)		a	été	dansé

(32)	*dass	es	getanzt	wurde
	that	it	danced	was
	dass		getanzt	wurde

German and French allow the formation of passive intransitive verbs. When an active verb has one argument only, the passivized counterpart will have no argument at all. In German, this leads to subjectless constructions (just as in (25b)), and furthermore, we can observe that the insertion of an expletive subject leads to ungrammaticality. Assuming that GEN is, in principle, capable of placing expletives into various positions, DEP(pronoun) >> AGREE yields the correct predictions concerning (32). French, on the other hand, does not tolerate subjectless constructions. An expletive must be inserted to fulfill the needs of AGREE.

(33) DEP(pronoun): Do not insert pronouns.

(34) German passive

	DEP(pronoun)	AGREE
dass getanzt wurde		*
dass es getanzt wurde	*!	

(35) French passive

	AGREE	DEP(pronoun)
il a été dansé		*
a été dansé	*!	

DEP(pronoun) penalizes structures that contain pronouns that have not been part of the input. It is a faithfulness constraint – spelling out the content of "economy constraints" such as Full Interpretation (Chomsky 1995) in just a slightly different way. This result is different in an approach in which Richness of the Base is taken seriously. In such an approach we can have inputs with and inputs without pronouns. Whether it is an output with or one without a pronoun which is taken as optimal depends on the constraint hierarchy.

If syntactic inputs are confined to arrays of content words, the sentences *who did you see*? and *I did not see her* violate the faithfulness principle DEP(aux): the output contains an auxiliary (*do*) that was not part of the input. The ungrammaticality of **I did see her* (without focal stress on *did*) and of **who did see you* indeed show that the pleonastic verb *do* may not be freely inserted into structures. Its use must be warranted by the need to respect a markedness principle (see Grimshaw 1997).

Do-insertion into a constituent question may be captured along the following lines. Apparently, the head position must be overtly filled in all English main clauses. Let us call the pertinent principle OBLHD(clause), taking up and slightly modifying the description developed by Grimshaw (1997).

(36) OBLHD(clause)
The head position of a clause must be phonetically filled.

Suppose, following Chomsky (1986) and Grimshaw (1997), that wh-subjects do not need to move to the specifier position of CP. They occupy the highest specifier position in the clause, and we may assume that the verb fills the head

position of this projection (perhaps because it is a VP, as Grimshaw 1997 argues). Due to DEP(aux), **who did see her* is blocked by *who saw her*.

When an object is questioned, it moves to the specifier position of CP. This CP must have a head. The following tableau shows that *do* must be inserted in object questions if OBLHD(clause) >> DEP(aux).

(37) OBLHEAD in English

	OBLHD	DEP(AUX)
who did you see		*
who you saw	*!	

According to economy considerations, *movement* should also be restricted to those contexts in which it cannot be avoided. Indeed, there is no language in the world in which word order is free in a literal sense. Where we find reordering, it always serves a function – the need to express scope differences, different information structure packagings – or the need to fill a subject position because of AGREE in, say, a passive clause when no expletive can be inserted for reasons that need not concern us here. Movement always fulfills a purpose; if no such requirements must be met, movement is illegal.

(38) Bill invited Mary
Mary was invited _
*was invited Mary
*there was invited Mary
*it was invited Mary

Let us consider the formation of questions in somewhat more detail, showing that the ban on unnecessary movement can be conceived of as a faithfulness constraint. In English, German, Bulgarian, Italian and many other languages, constituent questions must begin with a wh-phrase. This may be captured by a markedness constraint such as (40):

(39) a. I do not care [who [you have invited]]
b. *I do not care [[you have invited who]]
c. I hope you have invited someone
d. *I hope someone you have invited

(40) WH-CRIT
The specifier position of a question must be filled by a wh-phrase.

Chinese is different. As (41) shows, the wh-phrases remain in their canonical positions.

(41) Zhangsan xiangxin **shei** mai-le shu
Zhangsan believe who bought books
'who does Zhangsan believe bought books?'

If we want to capture (41), WH-CRIT must be counteracted by a further principle, and this principle is the ban against unnecessary movement. The principle is often called STAY, and in the formulation offered by Grimshaw (1997) it counts as a faithfulness constraint. STAY must penalize movement, so that the two different rankings in the following tableaus yield the two language types:

(42) WH-CRIT in English

	WH-CRIT	STAY
who you have invited		*
you have invited who	*!	

(43) WH-CRIT in Chinese

	STAY	WH-CRIT
Shei Zhangsan xiangxin mai-le shu	*!	
Zhangsan xiangxin **shei** mai-le shu		*

How could this be expressed – in particular in the light of the fact that OT is a representational grammar that should not directly talk about moving an element in the constraints? Many syntactic models assume that the displacement of phrases and heads is a process that not only shifts an element from position A to position B, but that also creates a "trace of movement", an inaudible copy of the moved element, in position A. The abstract representation of (39a) would thus be something like (44).

(44) I do not care [who [you have invited t]]

The relevant principle can then make direct reference to traces. STAY could be formulated as a constraint to the effect that traces are ungrammatical, or, expressed in more standardized terms:

(45) STAY = DEP(trace)

In (44), an element is present that was not part of the input, namely the trace. Thus, STAY = DEP(trace) blocks movement because insertion of a trace violates faithfulness. If WH-CRIT >> STAY, movement to clause-initial position is licensed when WH-CRIT cannot be satisfied in a cheaper way. In the reverse ranking the situation arises that we find in Chinese. In the formulation offered by Grimshaw, STAY is a faithfulness constraint because the trace is not part of the input. Under this perspective, syntactic economy effects *always* instantiate faithfulness considerations.

The correctness of this assessment of the role of faithfulness constraints in syntax obviously depends on a model-specific assumption concerning movement. Traces of movement figured prominently as a descriptive device in syntax in the Government and Binding Model (Chomsky 1981), out of which the system proposed by Grimshaw (1997) grew, but they have been replaced by the concept of a copy left behind by movement in Chomsky (1995) – and the role played by these copies in grammatical descriptions is fairly low. Furthermore, if the output of a syntactic competition is not an abstract syntactic representation (which may contain traces) but a surface structure in the strict sense of the word (a Phonetic Form of the sentence), then traces cannot be present in outputs at all, so that STAY could not be interpreted as a genuine faithfulness constraint. If PFs are outputs of syntactic competition, (45) would not be the proper formulation for the ban against movement, that is, movement economy would not illustrate faithfulness in a straightforward sense. It might rather reflect the interaction of different alignment constraints regulating the linear ordering of sentential constituents (see 2.3.).

A final remark on syntactic inputs seems necessary. If an input consisted of a set of content words only, then a question such as *who did Mary see* could not arise in English, given the constraints formulated so far. The input of this sentence would be {who, see, Mary}. The words can be combined in various ways. Consider which structure STAY, DEP(aux) and WH-CRIT predict as a winner:

(46) WH-CRIT, STAY and DEP in English

	WH-CRIT	STAY	OBLHD	DEP(aux)
who saw Mary				
who did see Mary				*!
who Mary saw		*!	*	
Mary saw who	*			
Mary did see who	*!			*
who did Mary see		*!		*

Who saw Mary is better than *who did Mary see* in a number of respects. It does not epenthesize an auxiliary. It does not have to *move* the wh-phrase to the front. But subject questions do not block object questions in English, although they have a better constraint violation profile. They simply do not compete with each other, because they mean different things. Thus, inputs cannot be just *sets* of words.

Grimshaw (1997) solves the problem by assuming that inputs are *predicate-argument structures* (PAS), that is, pairings of predicates with fillers of argument roles. They may look like in (47). Since (47a) (corresponding to *who did Mary see*) and (47b) (*who saw Mary*) have different PASs, the two sentences would not compete with each other if PASs are inputs, as required.

(47) a. [PRED: see
Agent:Mary
Patient:who]

b. [PRED: see
Agent:who
Patient:Mary]

The more semantic information is added to inputs, the larger the role played by faithfulness in syntax will be. Legendre et al. (1998) propose that the semantic scope of question words is part of the syntactic input. If this view is correct, the need to move wh-phrases to their scope position will be driven by faithfulness considerations. If scope is not part of the input, however, wh-phrases raise to the front position because of a markedness principle like WH-CRIT. We need not (and cannot) decide the issue here; it suffices to make the fact transparent that what is a faithfulness principle, and what is not, is a matter of the proper delineation of inputs.

Markedness and faithfulness principles interact in a further very interesting way related to the nature of the input. More traditional approaches to phonology or syntax assume that one needs to complement the grammatical descriptions of languages by systematic statements on the nature of lexical specifications. There are segmental inventories in phonology, which defining the set of phonemes that exist in a certain language and seem to be the building blocks for further phonological operations. Similarly, languages differ with respect to the set of syntactic features they use (ergative case plays no role in French), the functional elements they lexicalize (Russian and Hindi have no determiners) and the nature of lexical specifications (no mono-argumental verb assigns dative case in Japanese) in a very systematic way. Do we need a special grammatical component for expressing these regularities?

OT's answer is (in principle) negative: markedness constraints may be ranked relative to faithfulness constraints in such a way that certain featural specifications could never surface in the language in question. Thus, nasal vowels are more marked than oral ones – they appear later in language acquisition, the contexts in which they show up may be restricted – and they do not occur in all languages (see also next chapter). If the corresponding constraint NONASAL (48) is ranked higher than MAX(nasal), then nasal vowels have a hard time showing up among the sounds of that language, though they may surface as a result of assimilation. If in a language L, NONASAL is ranked higher than the constraint requiring assimilation of nasality between adjacent vowels and consonants, no nasal vowel will ever arise, no matter what the origin of the nasality. The ranking of the markedness constraint and faithfulness is sufficient to define this aspect of an inventory.

(48) NONASAL
Vowels are not nasal.

Recall also that the use of cases is governed by a set of markedness constraints Suppose that *ERG >>*DAT >> *ACC holds in a certain language, and suppose that (as seems to be the case) simple verbs never have more than three arguments. Then, at best, UNIQUENESS may imply that the three least marked cases are used in the maximal situation (nominative, accusative, dative) – ergative has no chance of showing up. Thus, there is no principled need to specify a segmental inventory of the set of relevant syntactic features separately from the rest of the grammar. The ranking of markedness constraints determines which features and segments can play a role at all in the language in question.

So far, we have framed our discussion of faithfulness in terms of the so-called *containment* version of Optimality Theory proposed in Prince & Smolensky (1993). In this approach, the computation starts with a given input, to which GEN applies, projecting (generating, deriving) the input onto several output candidates on the basis of this specific input. All outputs must contain the input.This is a view that would be most appropriate for a derivational model of grammar. This approach leaves some room for the influence of properties of GEN on what can be grammatical in a certain language: only the structures S' onto which GEN maps the input I take part in the competition for the best form realizing I.

Prince & Smolensky (1995) have proposed a slightly different model, called *correspondence theory*. This model assumes that the output candidates are generated independently of inputs. There is a set of (generally) possible output structures O generated by GEN. Faithfulness constraints then control the degree of correspondence between certain properties of these output candidates and the input. The difference between these two approaches can be illustrated as follows.

In Containment Theory, we start out with, say, *rad*, and compute various output candidates from the input by changing features (*rat, rod*), deleting material (*ra*), adding material *(rade)*, changing order (*dar*), or by not doing anything at all. The hierarchy of constraints then determines which of the candidates is the best one. In Correspondence Theory, all these forms are generated independently, with no special reference to *rad*. When we try to figure out the optimal realization of *rad* we do so in a system in which correspondences are established between parts of the input and parts of the output representation. We can do so by co-subscripting, as exemplified in (49) – a relation that expresses that the two elements bearing the same index are those that need to be checked for the degree of correspondence.

(49) a. $r_1a_2d_3$

b. $r_1a_3d_2$

c. $r_1a_2t_3$

d. $d_3r_1a_2$

e. $r_2a_3t_1$f. $p_1i_6n_{349}$

EVAL then applies as in a containment model: Candidate c is unfaithful to the input with respect to the voicing specification of element 3. Note that candidate e has the same surface form as candidate c, but involves radical faithfulness violations: element 2 has moved to initial position and completely changed its featural makeup – just like its other segments. Note that the correspondence view minimizes the possible role played by GEN in grammatical description. In containment, a certain structure C may fail to be the optimal representation of input I because GEN cannot map I onto Σ, while it may be the case that a different input leads to Σ (a fact irrelevant for EVAL). In the Correspondence Theory, on the other hand, because Σ is thus generatable as such, it *is* a candidate to be evaluated with respect to any input. Thus, Σ now has a chance of winning the competition for the best output of I. This difference in the predictions made by the two versions of the theory will be important when we discuss bidirectionality.

Correspondence Theory originally arose from the need to account for cyclic phenomena, and most of all, from the need to have output forms as the base for class II affixation, hypochoristic formation or reduplication. In reduplication, the reduplicant (the part of the word which is copied, or reduplicated), usually reproduces a portion of its base in its surface form. The base thus serves as a kind of input, to which the reduplicant can be faithful. McCarthy & Prince first called this relation "output-output correspondence." In a second step, because the input-output relationship is indistinguishable from the output-output one, they showed

that input-output faithfulness can be accounted for by the same constraints as those needed in the correspondence relationship. Further correspondence constraints proposed by McCarthy & Prince (1995) are listed in (50). S_1 and S_2 can be input or output.

(50) Faithfulness constraints

a. IDENT(F):
Let α be a segment in S_1 and β be any correspondent of α in S_2.
If α is [γF], then β is [γF].
(Correspondent segments are identical in feature F).

b. LINEARITY "No Metathesis":
The output is consistent with the precedence structure of the input.

c. CONTIGUITY "No Skipping":
The portion of S_1 which is in correspondence with S_2 forms a continguous string.

d. ANCHOR:
Each element at the left/right edge of S_1 has a correspondent at the left/right edge of S_2.

This section has shown that the tension between markedness constraints reflecting universal tendencies of language and faithfulness to lexical specifications is an important source for building up conflicts in the description of natural language. Markedness constraints imply that lexical specifications be changed in the interest of simplifying language outputs, while faithfulness conditions try to maintain the differences between different words or different arrays of words – they try to maintain the lexical contrasts. The conflicts are not solved the same way in all languages. Faithfulness to the input may be responsible for some, but perhaps not all aspects of *economy* in language. The definition of faithfulness implies that we have a clear concept of what is a possible input, and what is a possible output.

2.2 Markedness competing with Markedness

The tension between faithfulness and markedness requirements is not the only source of conflicts among grammatical principles. A simple reflection suggests that the interaction of markedness constraints themselves will not always be free of conflicts either. Markedness constraints reflect the shaping force that several factors may exert upon language (see our remarks on grounding in the next chapter). There is no reason to expect that such factors should always pull language into the same direction. Language is organized on different levels, and

its grammar is organized along several dimensions. Except for very simple and unmarked inputs, simplicity metrics of these different levels need not and cannot be identical. Conflicts between the constraints encoding these simplicity metrics are inevitable.

A good place to look for conflicts between markedness constraints is metrical phonology. It introduces organizational aspects for the sound system of language which go much beyond segmental or syllable structural considerations. One important generalization is the
Trochaic-Iambic Law (Hayes 1995). This claims that a canonical trochee (σ́σ) consists of two equal syllables, in which both syllables have the same weight (mono- or bimoraic, or just syllables in quantity-insensitive systems). In contrast, a canonical iamb ($\sigma_1\acute{\sigma}_2$) must consist of unequal syllables: σ_1 is lighter than $\acute{\sigma}_2$. This last requirement encoding a metrical optimum is potentially in conflict with the markedness constraint encoding that all syllables have the structure CV. The conflict between the two markedness principles is resolved differently in different languages. Some languages value the metrical restriction higher than the syllabic one. To produce good iambs, French, Hixkaryana and Yupik Eskimo *lengthen* the last syllable, but Eastern Ojibwa (Odawa dialect), and Beduin Arabic *reduce* the first syllable. A word like *kitib* becomes *k.tib* (McCarthy 2000). The first syllable loses its vowel and becomes badly marked. But in doing so, the first syllable becomes lighter than the second one, and the result is a good iamb. In the conflict between markedness principles of syllable and metrical structure, the latter wins. It must be noted that final [b] is extrametrical, and plays no role in the weight computation of the last syllable.

In other iambic languages, syllable manipulation is only possible in some restricted contexts. Unami and Munsee (Kayes 1995, Goddard 1979) are such languages. Canonical iambs are formed only when either vowel deletion is possible in the weak member of an iamb, or when gemination takes place in the strong syllable (see 51c). There remain numerous iambs, however, in which no change takes place, as shown in (51a and b).

(51) Munsee iambs

a. /wəlamaləsəw/	–> [wəlamaləsəw]	‘he is well’
b. /mətəme:w/	–> [mətəme:(w)]	‘he follows a trail’
c. /nəmətəme:/	–> [nəməttəme:]	‘I follow a trail’

Other straightforward examples of conflict between markedness considerations involve an even smaller amount of structure. Sonority should increase in the onsets of syllables as one goes from left to right, so that sequences such as [tr] or

[pl] are good onsets, while [rt] and [lp] are not. The corresponding markedness principles easily run into conflicts with the principle NoCoda when it comes to syllabifying longer sequences of sounds, as in Spanish *centro* /tsentro/, where *se.ntro* avoids a coda for the first syllable but makes the second bad in terms of the sonority of its onset. Spanish, like English, German, Italian and others, chooses the alternative syllabification *sen.tro* with a coda, but with fulfillment of the sonority hierarchy. Of course, faithfulness comes into play here as well: in terms of syllabification principles the output *te.se.ne.te.ro* is quite unfaithful to the input. Notice in passing that we are again confronted with the issue of repair. Here we see that in many Indo-European languages, it is better to fill a coda than to violate the sonority hierarchy or to change the segmental make-up of words. In other languages, like Bantu languages, Japanese or Hawaiian, which have severe restrictions on their codas, other options would be preferred. The case is different from what we observed in final devoicing, where the universally, preferred repair was to delete [+voice] from the final obstruent. In the case of syllabification, several options, and thus several rankings of the constraints, are equally good. This is illustrated in detail in the next chapter.

While undisputable examples of faithfulness vs. markedness conflicts are rare in syntax and its interface levels (maybe they are confined to lexical exceptions, depending on the selected input concept), markedness vs. markedness conflicts are easier to identify. Consider again the assignment of case. We have already argued that nominative is less marked than accusative and dative, with the latter being the most marked instance of case. Recall that we may assume that there are markedness principles *Acc and *Dat which penalize the use of accusative and dative case, respectively. They successfully explain why intransitive (mono-argumental) clauses are normally constructed with nominative, but they would also lead one to expect that nominative should be the only case for both arguments in the transitive clause as well. While there are some such languages, the majority chooses a nom-acc pattern (or nom-ergative pattern), which motivates a further markedness principle that has also already been introduced: Uniqueness requires that within a single clause, each case should be associated with at most one noun phrase. *I saw him* respects one markedness principle (Uniqueness) by violating another (*Acc). English ranks Uniqueness higher than *Acc, but it does not do so with respect to *Dat – at least according to some analyses that take both objects. That is why two accusatives are fine in *I give her it*. German opts for the opposite ordering, which means that the highly marked dative shows up whenever the two less marked cases nominative and accusative are already "used up", so that Uniqueness dictates that a third case be employed.

2.3 Alignment Constraints

There is a final class of grammatical principles that may render conflicts in natural language inevitable: the demands of linearization, and of glueing various levels of representations together at the correct points. It will become obvious that the distinction between *alignment* constraints and markedness constraints may sometimes be blurred, but the crucial point does not, of course, lie in classification but rather in identifying different sources of conflicts in natural language.

Let us begin with a simple example. When we look at the English IP, we observe that its specifier, the subject, occurs at the left periphery of the category. Likewise, in a wh-question (a CP), its specifier, the wh-phrase, is at the left edge. The genitive specifier of a noun phrase is not linearized differently. We may express this by using an *alignment constraint* that tells us how to locate the edges of different categories with respect to each other. In its most general form, it would take the shape of (52):

(52) Generalized Alignment
Align (Cat_1, $Edge_1$, Cat_2, $Edge_2$)

For all Cat_1 ($\forall Cat_1$) there is a Cat_2 ($\exists Cat_2$) such that $Edge_1$ of Cat_1 and $Edge_2$ of Cat_2 coincide.
. Cat_1 and Cat_2 are prosodic and grammatical categories. Edge is Left or Right.
In order for (52) to be true, the left/right edge of Cat_1 must fall together with the left/right edge of Cat_2. The two categories mentioned in the definition (52) are variables. The order of the arguments in an alignment constraint is not indifferent. The first argument is universally quantified and the second existentially.

Some examples of the categories used in McCarthy & Prince's original work are given in (53). When the Edge of Cat_1 and the Edge of Cat_2 are the same, they are mentioned just once, in agreement with newer conventions. (53a) is active in English stress, and (53b) in German syllabification, as will be exemplified below.

(53) Examples of alignment constraints

a. ALIGN-Ft: Align(PrWd, Foot, Left)
A morpheme begins with a syllable.

b. ALIGN-R: Align(Affix, PrWd, Right)
A morpheme ends with a syllable.

The syntactic observations discussed in section 2.1 could be expressed in the specific constraints given in (54) – or they even suggest the more general constraint (55).

(54) Align (Specifier-IP, Left, IP, Left)

Align (Specifier-CP, Left, CP, Left)

(55) Align (Specifier-CP, Left, CP, Left)

Align (Specifier-XP, Left, XP, Left)

Align (Specifier, Left, XP, Left), which requires the left edge of a specifier to coincide with the left edge of a maximal projection, is an alignment principle to which English attributes a high rank. Let us give it the name SPECLEFT. It has an obvious touch of markedness principle as well: since specifiers rarely occur at the rightmost position at the syntactic surface in any language, it would be a mistake to assume a principle SPECRIGHT with the opposite properties of SPECLEFT. At least in syntax, there is no symmetry in alignment (see also Kayne 1994, Haider 1996, and Grimshaw 2001). In phonology, as shown below, things are different.

However, SPECLEFT is not the only alignment constraint at work. English, French, German and Turkish clauses begin with their specifiers (at least in the unmarked case), but not all languages behave like that. Beside SVO and SOV languages, there are also VSO and VOS languages. In Irish or Niuean, the finite verb is the initial element of the IP.

(56) a. Chuala Róise go minic roimhe an t-amhr·n sin
heard Róise often before-it that song
'Róise had often heard that song before' (McCloskey 1996:269)

b. Gheall sé go bhfillfeadh sé ar an bhaile
promised he that return he on home
'He promised that he would return home'

This suggests that HEADLEFT (=Align (Head, Left, XP, Left)) is an alignment principle (reflecting an unmarked option), too. Due to simple laws of geometry reflecting temporal organization, only *one* element can appear at the left edge of a

category, however, so that SPECLEFT and HEADLEFT are *intrinsically* in conflict with each other.

At the IP level, English and Irish solve this conflict differently. In English, SPECLEFT >> HEADLEFT forces the subject into the initial position, whereas the Irish ranking HEADLEFT >> SPECLEFT excludes the appearance of finite SVO clauses and favors the VSO arrangement.

Grimshaw (2001) generalizes the idea that syntactic alignment principles require the leftmost realization of elements only. In her model, no right-alignment constraint is ever needed. All categories try to be aligned at the left edge. We may assume this is due to the special role the initial position plays perceptually. As soon as there is more than one word in a clause, alignment conflicts are unavoidable, and they are resolved differently in different languages. If we assume a further principle COMPLEFT (complements must appear at the left periphery of XP), the difference between OV (Japanese) and VO languages (English) can be captured, too. At first glance, it seems quite easy to compute the set of language types generatable under such premises (the factorial typology, see below) from the different rankings.

(57) SPECLEFT >> HEAD LEFT >> COMP LEFT: tolerates SVO only
SPECLEFT >> COMP LEFT >> HEAD LEFT: tolerates SOV only
HEAD LEFT>> SPEC LEFT >> COMP LEFT: tolerates VSO only
HEAD LEFT>> COMPLEFT >> SPEC LEFT: tolerates VOS only
COMPLEFT >> SPECLEFT >> HEAD LEFT: tolerates OSV only
COMPLEFT >> HEADLEFT >> SPECLEFT: tolerates OVS only

The specification of the options in (57) goes together with the assumption that the most general principles that are conceivable (specifiers are always to the left) determine serialization in natural language. This is not correct, however: German adjective phrases have the head precede the complements (*der seiner Frau treue König* the his.dat wife.dat faithful king, 'the king faithful to his wife'), while complements follow nouns in the noun phrase (*der Gatte der Königin*, 'the husband of the queen'). This can be expressed only if we allow for constraints like those in (58) – and if these can outrank the very general statements used in (57). These posit that the complement of an adjectival phrase is aligned to the left, whereas the complement of a noun phrase is positioned to the right of the respective phrase.

(58) a. Align (Complement-AP, Left, AP, Left)

b. Align (Complement-NP, Right, NP, Right)

Since the subject is the specifier of IP in some languages (English) but may fail to move out of VP in others (German), and given that alignment constraints may be quite specific (as (58)) shows, the order SV may in fact be the result of quite a number of different rankings. We leave it to the reader to work out the effect of these constraints on the word order.

(59) SPECLEFT >> HEAD LEFT

Align (Spec-IP, Left, IP, Left)
>> Align (Head-IP, Left, IP, Left)

Align (Spec-IP, Left, IP, Left)

>> Align (Head-VP, Left, VP, Left)

Align (Spec-VP, Left, VP, Left)

>> Align (Head-VP, Left, VP, Left)

Obviously, the same holds for other order relations, as well. Alignment principles may trigger movement, and a conflict among alignment constraints is an alternative way of capturing the default "ban" against movement, as was explained in section 2.1 above with respect to wh-movement. The wh-criterion originally introduced by Rizzi (1990) turned out to be a driving force for movement: the left edge of a wh-question must begin with a wh-phrase. In terms of alignment, we can formulate this constraint now as in (60).

(60) WH-CRIT: Align (CP[+wh], Left, wh-phrase, Left)

But the wh-phrase mentioned in (60) has further grammatical properties. For example, it may be the complement of a verb phrase. As such, it must be (left-) aligned in the VP – the category it was originally merged in. This is expressed in (61).

(61) ALIGN (Complement-VP, Left, VP, Left)

Thus, (61) tries to prevent the movement of the wh-phrase to the left edge of the CP in an example such as *I wonder* [$_{CP}$ *who you* [$_{VP}$ *saw*]], while (60) attracts it to the clause-initial position. If (61) >> (60) the Chinese system discussed above arises (no movement to clause initial position); if (60) >> (61), the English

constellation obtains. Observe that we have made reference neither to markedness nor to faithfulness considerations in this description of the triggered nature of movement.

Another example of linear ordering comes from morphology. Prefixation, suffixation and infixation can be seen as the results of alignment constraints. Prefixation is expressed by a constraint like (62a), which follows the alignment schema and which posits that an affix is always at the left edge of the word it is forming with its stem. Suffixation is as in (62b), the mirror alignment. Thus, morphology appears to be less asymmetric than syntax: there are also principles that require a morpheme to be right-aligned. Languages like English, German and French rank these constraints very high, not allowing phonological constraints to interfere in the linear ordering of stems and affixes.

(62) a. Prefixation: ALIGN-L: Align(Affix, PrWd, Left)
English: *un-true*; German: *Ge-lände* 'ground', *un-reif* 'immature'

b. Suffixation: ALIGN-R: Align(Affix, PrWd, Right)
English: *instrument-al, instrument-less*; German*: kind-isch* 'childish'

Linear ordering of the suffixes among each other is regulated by independent principles, which are largely language dependent, and thus not really interesting for OT. More interesting are infixation facts, as for instance infixation in Tagalog. In Tagalog (McCarthy & Prince 1993b) the infix *-um-* is located after the onset of the first syllable, if there is one, otherwise at the left edge of the word (see also Orgun & Sprouse 1999 for different examples).

(63) Infixation in Tagalog

Root	um + Root		
aral	um-aral	'to teach'	
sulat	s-um-ulat	'to write'	(*um-sulat)
gradwet	gr-um-adwet	'to graduate'	

In this language, it is more important to fulfill the constraint against codas than to align a prefix with the left edge of a word. We thus find the ranking NOCODA >> ALIGN-L(Prefix, PW, L), as illustrated in the following tableaus.Violations of Align-constraints are gradient, but as always violations are minimal (see chapter 4 for a discussion of gradient constraints).

(64) Infixation in Tagalog (um-aral)

/ um-aral /	NoCoda	Align-L
☞ um-aral	*	
a-um-ral	**!	*
ar-um-al	*	*!*

(65) Infixation in Tagalog (s-um-ulat)

/ um- sulat /	NoCoda	Align-L
um-sulat	**!	
☞ s-um-ulat	*	*
su-um-lat	**!	**
sul-um-at	*	**!*

(66) Infixation in Tagalog (gr-um-adwet)

/um-gradwet/	NoCoda	Align -L
um-gradwet	***!	
g-um-radwet	***!	*
☞ gr-um-adwet	**	**
gra-um-dwet	**	***!

In addition to governing linearization, alignment constraints may be employed for other tasks as well:

- Separation of domains: for example “crisp” syllabification.
- Association of different kinds of entities with each other, e.g., of stress or tones with syllables.

In fact, in contrast to syntax and morphology, the most prominent effect of alignment in phonology is the requirement that grammatical constituents have clear prosodic boundaries. A prototypical example first introduced by McCarthy & Prince (1993b) is the observation that morpheme edges should fall together with syllable edges. This effect can be felt in many languages, though some languages blur their morpheme or even their word edges in having larger domains

of syllabification. French, for example, allows a great deal of syllabification across word boundaries. However, no resyllabification is found across boundaries of Phonological Phrases (PhP). Compare [$_{PhP}$ *les-enfants*] [$_{PhP}$ *sont-allés nager*] 'the children went for a swim', where liaison applies between *les* and *enfants* as well as between *sont* and *allés*. But in the sentence [$_{PhP}$ *les-enfants*] [$_{PhP}$ *ont mangé du chocolat*] 'The children have eaten chocolate', no liaison applies between *enfants* and *ont*, since these words belong to different Phonological Phrases.

German, on the other hand, is a good example of a language which tries to let its morpheme boundaries coincide with syllable boundaries (and, in doing so, to clearly delimit morpheme boundaries), though it crucially does not always succeed. In general, suffixation in German implies syllable boundaries between stem and suffix, as shown in (67a–c), except in case the stem ends in a consonant and the suffix begins with a vowel. These two segments are syllabified together, as shown in (67d).

(67) Suffixation in German

a. C+C: faul/Faul-heit	[faʊ̯l.haɪ̯t]	'lazy-laziness'
b. V+V: Ruhe/ruh-ig	[ʀuː.ɪç]	'quietness-quiet'
c. V+C: froh/fröh-lich	[fʀøː.lɪç]	'joyful-joyful'
d. C+V: Kind/kind-isch	[kɪn.dɪʃ]	'child-childish'

The two constraints in (68) compete with each other for (67d), but not in the other cases.

(68) a. ONSET: Syllables have onsets

b. ALIGN-R (stem, syllable, R):
the right edge of a stem falls together with the right edge of a syllable (for all right edges of stems there is a right edge of a syllable, so that both edges fall together).

In suffixation, the need to satisfy the unmarked syllable structure is higher than the need to separate morphemes. In the case of *Faulheit*, *ruhig* and *fröhlich*, ONSET and ALIGN-R do not compete, since both constraints can be fulfilled at the same time: the morphemes are separated by a syllable boundary. In *ruhig* or *böig*, the second syllable has no onset, but a syllable boundary separates the two morphemes all the same. As shown in (70) it is more costly to insert a consonant than to violate ONSET. In *Faulheit* and *fröhlich*, morpheme structure and syllabification fall together. Since the suffix begins with its own consonant, this

consonant serves as the onset of the syllable of the suffix. But in *kindisch* (and also in words like *sonn-ig* 'sunny', *Lad-ung* 'cargo', *lach-en* 'to laugh') there is a conflict between ONSET and ALIGN-R. The suffix has no consonant of its own. In order to get an onset it must take the last consonant of the stem or epenthesize a consonant. Epenthesis is excluded, since DEP ranks higher than ONSET, but since ONSET is ranked higher than ALIGN-R, the first solution is chosen. In such a case, the morpheme boundary does not coincide with a syllable boundary. In other words, ALIGN-R is dominated and crucially violated.

(69) Syllabification of *kindisch* 'childish'

/kind+isch/	DEP	ONSET	ALIGN-R
☞ kin.disch			*
kind.isch		*!	
kind.lisch	*!		

(70) Syllabification of *böig* 'windy'

/bö+ig/	DEP	ONSET	ALIGN -R
☞ bö.ig		*	
bö.tig	*!		

It must be noticed that a candidate like *b-ig-ö*, with infixation of the suffix in order to fulfill ONSET – in the same vein as what was observed in Tagalog – does even better in such a ranking. What is needed to eliminate this candidate is an additional faithfulness constraint, CONTIGUITY, formulated in (50c) above, to the effect that contiguous segments in the input are contiguous in the output. In German, this constraint is undominated, whereas it is crucially violated in Tagalog.

A consequence of ALIGN-R is that two adjacent vowels that could fuse together into a diphthong refrain from doing so because of morpheme edges. This is shown in (71). Even though the diphthong [ai^{a}] exists in German, as in words like *Hai* 'shark' or *Fleisch* 'meat', diphthongization is blocked when the two vowels come from different morphemes, as in *prosa-isch*.

(71) No glide formation across morpheme boundaries

prosa-isch [a.i] ‘prosaic’ (vs. Fleisch [flaɪ̯ʃ] ‘meat’)
ruh-ig [u.i] ‘quiet’ (vs. Pfui [pfuɪ̯] ‘yuck’)

Turning now to prefixation and compounding, the need for ‘crisp’ syllabification (a term from Ito & Mester 1994) is even greater than in suffixation. Even when the prefix ends with a consonant and the stem begins with a vowel, or when the same situation arises between two elements of a compound, there is no resyllabification across the morpheme boundary.

(72) Prefixation in German

verärgern [vɐ.ɐ.gɐn] *[və.ʁɐ.gɐn] ‘to annoy’
unartig [ʊn.aɐ̯tɪç] *[ʊ.naɐ̯..tɪç] ‘naughty’

(73) Compounding in German

Stockente	[ʃtɔk.ɛn.tə]	*[ʃtɔ.kɛn.tə]	‘mallard’
Seeadler	[ze:.ɑ.dlɐ]		‘sea eagle’

A second Alignment constraint is needed which also aligns the stem with a syllable, but from the left side. This new constraint, ALIGN-L, is higher ranking than ONSET.

(74) ALIGN-L (stem, syllable, L):

The left edge of a stem falls together with the left edge of a syllable (for all left edges of stems there is a left edge of syllable, so that both edges fall together).

The final ranking is ALIGN-L >> ONSET >> ALIGN-R. Both prefixes and compounds have a clear syllable boundary (though some lexicalized elements seem to be able to trigger resyllabification in fast speech, as in a word like *unerhört* ‘unheard of’, which can be pronounced as *u.ner.hört*). The syllable boundary between *n* and *a* in *unartig* is signalled by a glottal stop.

(75) Syllabification of *unartig* ‘naughty’

/un+ar+tig/	ALIGN-L	ONSET	ALIGN-R
☞ un.ar.tig		**	*
u.nar.tig	*!	*	*
un.art.ig		***!	

As for the third effect of alignment, the coinciding of different kinds of linguistic entities, phonology can require that a phonological (or grammatical) element fall on another constituent. For instance, features, tones or stress are associated with special units bearing them (‘bearing units’). Features are associated with segmental roots, tones with syllables or moras, and stress can be multiply associated since it is typically realized on a syllable, but is also at the same time the head of a foot, of a word, and so on. Thus, at the phonetic level it is realized on segments and syllables, but at the interface with syntax, stress can be associated with lexical elements or with XPs, and in the semantics, stress can also mark elements in the scope of focus operators.

To end this review of the effects of alignment, let us briefly show how OT accounts for lexical stress. Stress in suprasegmental phonology is best conceived of as the grouping of constituents. It is often peripheral in the domain considered, like Feet, Prosodic Word, Intonation Phrase, etc. It is generally final, penultimate or initial, which speaks for an analysis in terms of alignment. At the lower level, syllables are grouped into feet. Feet are trochaic (left-headed) or iambic (right-headed). Note that OT does not need to postulate a universal inventory of feet, like the one elaborated by Hayes (1995) for instance. The inventory of feet is just a consequence of the universal constraints. If feet are binary at the syllabic or moraic level, and if they are either right- or left-headed, only binary trochaic and iambic feet are allowed. This is again in line with the general remarks about the fact that OT defines inventories of linguistic elements with the help of constraints licensing them (faithfulness) or prohibiting them (markedness).

(76) Trochaic-Iambic Pattern

ALIGN-L/R(Foot,Head, L/R)

Align the left/right edge of a foot with its head.

(77) FOOT-BINARITY
Feet are binary.

Since the constraints regulating the unmarked stress pattern are violable, marked feet are possible, like ternary or unary feet, feet without a head, etc. Other properties of stress are accounted for in the same way.

Also at higher levels, in a Prosodic Word or a phrase, stress can be interpreted as standing for groupings of constituents, and there, too, it is peripheral. However, since feet are the relevant constituents, when feet are trochaic and aligned with the right edge of a Prosodic Word, stress is penultimate. As an illustration, consider the stress pattern of English, as proposed by McCarthy & Prince (1993b).

(78) English stress (McCarthy & Prince 1993b)

a. ALIGN(PrWd, Ft, L): All Prosodic Words start with a left-aligned foot.
b. ALIGN(Ft, PrWd, R): All feet are right-aligned with the right edge of the word.
c. ALIGN(Head, PrWd, R)
d. PARSE-SYLLABLE: Syllables are parsed into feet.
e. (Tàta)ma(góuchee) *Ta(tàma)(góuchee)

(78a and b) are formulated differently. (78a) says that for each prosodic word there is one left-aligned foot. (78b) quantifies over the arguments the other way round. It specifies that all feet are right-aligned with a prosodic word. If the constraint PARSE-SYLLABLE is high-ranking, the effect of this constraint together with (78b) will be that foot formation is iterative from right to left, up to the beginning of the word, where it is more important to have a foot left-aligned with the edge of the word. Main stress is penultimate, because the last foot of the word also contains the main stress.

With stress, we have illustrated a property of OT called Generalized Template. Language-dependent choices among possible patterns need not be independently posited, but derive from constraint ranking. It does not need to be specified that English and German use moraic trochees. This is a consequence of the ranking of the constraints responsible for stress, like ALIGN, FOOT-BINARITY, HEAD, etc. Generalized Template has been important for morphological templates like those postulated for reduplication or hypochoristic formation, since it allows them to be eliminated from the theory entirely.

Align constraints have been shown to be applicable to a multitude of cases, since every positional phenomenon can be expressed in terms of ALIGN, like ONSET and NOCODA for instance (Ito & Mester 1994). To conclude this chapter,

let us come back to the facts of final devoicing, and account for them with alignment-like constraints.

Positional faithfulness has recently been introduced into OT phonology by Beckman (based on work by Steriade, Flemming and Lombardi, among others). Its effects are comparable to the effects of alignment, though the basic idea is different. It claims that faithfulness is more prone to be fulfilled in prominent positions than in less prominent ones. In order to be effective, a context-dependent faithfulness constraint has to be ranked higher than the constraint militating against the element or property in question. This latter constraint in turn has to dominate the corresponding contextless constraint (see (79) and (80) for an illustration). Once again, we are confronted with an elsewhere effect. The insight behind positional faithfulness is that prominent positions allow more contrasts, and, as we saw before, contrasts are a consequence of being faithful to input lexical specifications.

Consider final devoicing in German once again. There have been at least two different ways of looking at final devoicing. First the standard explanation claims that it is a neutralizing process which takes place at the end of syllables. This is the view we worked with at the beginning of the chapter. A second way of looking at these facts was first proposed by Lombardi. In her approach, voiced obstruents are licensed in a certain position, and more precisely, before tautosyllabic sonorants. German allows voiced obstruents in those positions which can be summed up as syllable initial (though being before a sonorant is not necessarily equivalent to being in the onset). At the end of a syllable, only voiceless obstruents are possible, because in this position obstruents are never before a sonorant (at least in German). Positional faithfulness accounts for this contrast with two constraints, ordered in an elsewhere way. (79) is the general constraint, and (80) the specific one. If (80) is ranked higher than FINALDEVOICING, as in tableaux (81) and (82), its effect will be to allow voicing in onsets. If (79) is ranked lower than FD, again as in tableaux (81) and (82), the result is that only onsets contain voiced obstruents. If (79) is the higher ranking constraint, voiced obstruents are allowed in all positions, as illustrated in (83) for English.

(79) IDENT(VOICE)
Correspondent segments must agree in voicing.

(80) Positional faithfulness constraint (Beckman 1998:38)
$\underline{\text{IDENT(VOICE)}}_{\underline{\text{Onset}}}$
Onset segments and their input correspondents must agree in voicing.

(81)

li:b	IDENT(VOICE)$_{Onset}$	FD	IDENT(VOICE)
li:b		*!	
li:p			*

(82)

ball	IDENT(VOICE)$_{Onset}$	FD	IDENT(VOICE)
ball			
pall	*!		*

(83)

bed	IDENT(VOICE)	FD	IDENT(VOICE)$_{Onset}$
bed		*	
bet	*!		*

Positional faithfulness has been very effective in accounting for phonetic contrasts which are realized in certain contexts and not in others (in the vicinity of certain segments, in stressed positions, and the like) due to differences in the perceptual cues.

This chapter has shown how conflicts in grammar are conceived of as conflicts between different kinds of constraints in OT. We have concentrated on classes of conflicts, and their resolution. In the next chapter we show how typological variation is interpreted as constraint reranking of the universal constraints.

Chapter 3

Universality and Free Ranking

Summary of the Chapter

Several of its design properties render Optimality Theory a very attractive model of the linguistic faculty. By assuming that *all* constraints have to be universal, OT severely restricts the degrees of freedom in model formulation in linguistics (one of the core problems of linguistic description). Its strong reference to markedness shows that OT cannot possibly deny its roots in phonology, but in doing so, and by stressing the role *grounding* plays for the identification of universal constraints, it sets even tighter limits for linguistic theorizing.

OT furthermore offers a restrictive theory of linguistic variation: differences between languages can arise only a different rankings of universal principles in different languages.

All these advantages are made possible by the violability of constraints. Constraints do not need to be fulfilled in all contexts. Constraints can thus be formulated in as simple and straightforward a form as required for them to be compatible with grounding.

After some introductory remarks concerning the strong connection between the universality of principles and their violability, this chapter continues with a review of the role of markedness in OT - and largely ignores faithfulness issues. It is shown which kinds of facts motivate the postulation of markedness in OT, and how reranking accounts for typological diversity. Typological variation in OT is the result of constraint reranking: it is shown in the next section that factorial typology is an important component of OT. Some formal aspects of markedness hierarchies, in particular "harmonic alignment" are discussed in detail in the last section.

3.1 Universality

Everyone who has tried to formulate a grammatical rule or principle for some linguistic phenomenon in a certain language knows about a central problem of linguistics: the empirical facts can be described by a multitude of very different

grammatical rules or principles. Even if one applies common standards of simplicity and elegance, a particular choice among these possible descriptions can be hard to motivate.

Linguistic theorizing would be much more constrained if one requires that *all* grammatical rules and principles be *universal* in the sense that they correctly apply in *all* existing and possible natural languages. In the optimal state of affairs, one would find positive evidence for the validity of the constraints in question in all languages. A less strict but still highly predictive requirement would postulate that the constraints must at least be compatible with the grammatical systems of all languages.

At the outset of generative grammar, the identification of a universal set of principles seemed to be a goal much beyond reach. Theoretical research in generative grammar focused on the basic architecture and other formal properties (the distinction of underlying representation and phonetic representations, the difference between the deep and the surface structure of a sentence, the overall format of phonological rules or syntactic transformations, rule ordering), while the substantive parts of grammatical description were formulated in a language particular form.

This state of affairs changed gradually. In the syntax, Chomsky (1981) is the first full-fledged attempt of a grammatical description that works with *universal principles only* though these are complemented by language-particular parameters. At roughly the same time, phonology postulated universal principles concerning syllable structure (Clements & Keyser 1983), the association of tones with tone bearing units (Goldsmith 1976), and the prohibition of identical adjacent elements in a morpheme (the Obligatory Coutour Principle, Leben 1973, Goldsmith 1976, McCarthy 1986), and developed a substantial theory of feature geometry by which one could assess the plausibility of phonological processes (Clements 1985, Sagey 1986).

While the universality of constraints is certainly a highly valued *goal*, postulating the necessity of universal principles and actually identifying them are, of course, two different things. That the conflictual nature of OT is particularly helpful in this domain is easy to see when one considers a simple example from syntax. In English, French, or the Mainland Scandinavian languages, finite clauses require the presence of a subject. If the verb or predicate in question requires no subject argument, a formal subject like *there*, *it*, or *il* must be inserted, as (1) illustrates, Chomsky (1981, 1982), therefore postulated the Extended Projection Principle or the EPP as part of universal grammar. It does the job of the constraint AGREE introduced in the preceding chapter.

(1) a. there was a moose shot

b. it seems that John loves Mary
c. il a été dansé
   it has been danced
   'one danced'

(2) Extended Projection Principle (EPP, informal version)
All (finite) clauses have a subject

However, one can observe that *other* languages fail to respect the EPP - at a superficial level at least. German examples such as (3a-b) seem to be cases in point.

(3) a. dass getanzt wird
   that danced was
   'that one danced'

b. dass mir schlecht ist
   that me.dat sick is
   'that I feel sick'

c. dass mir geholfen wurde
   that me.DAT helped was
   'that one helped me'

Does that mean one has to abandon the EPP in a grammatical description of German – thereby rendering the constraint non-universal? In OT, the answer is negative. The EPP fails to be respected in German in a well-defined set of contexts only. Subjectless clauses arise when the verb has no argument at all, due to passivization (as in (3a), or (b) when the verb has a single argument, which bears an exceptional, lexically controlled case, such as the dative in (3b), or when passivization implies that the only verbal argument left bears an exceptional case, as in (3c). At least (3b,c) can be related to the constraint MAX(CASE) of chapter two (an (exceptional) Case specified in the lexical input must be represented in the surface structure)- a background assumption made in the Chomskyan tradition as well (see Chomsky 1981, Chomsky 1998):

In Optimality Theory, one can maintain that *both* the EPP and MAX(CASE) are *universal* principles. The EPP is related to the Agree constraint introduced in section 2 as well if subjects are nominative noun phrases that agree with the finite verb; Agree then requires that each (finite) clause possesses such as subject (=EPP). In Chomsky (1981,1982), the EPP is furthermore linked to a positional

requirement for subjects (they must appear in Spec,IP), but this issue need not concern us at the present moment.

AGREE/EPP and MAX(CASE) have to be respected in all languages, but to the extent only that there aren't any more important (i.e. higher ranking) principles that force their violation. If MAX(CASE) ranks higher than AGREE/EPP in German, we expect that clauses may fail to have a subject (a nominative noun phrase) when the only argument bears a lexical case. Furthermore, if it is generally true (and at least for German, this holds without exception) that there is only one exceptional case per predicate (see REF (Jackendoff. Yip) Fanselow, 2000, but see also Wunderlich 2002), the ranking MAX(CASE) >> EPP also implies that clauses with more than one argument do have a subject: then, there are enough noun phrases present for fulfilling MAX(CASE) and AGREE/EPP simultaneously.

On the other hand, given the logic of OT, one would expect there to be languages in which AGREE/EPP >> MAX(CASE). In such a language, the single argument of construction with one argument only should invariably bear nominative Case – because of the high rank of AGREE/EPP, lexical prespecifications could not be maintained when AGREE/EPP is at stake. However, as is the case with the German ranking, a construction with two arguments might show one NP with an exceptional case – because the EPP can be fulfilled by the other. Japanese may be a case in point, according to the description proposed by Woolford (2001). In Japanese, transitive verbs are able to govern exceptional dative case, but interestingly, unlike what we find in German, there are no mono-argumental predicates at all which assign an exceptional case. Thus, German and Japanese are similar in the transitive case. When there is only one argument, the two Case conditions in question cannot be met at the same time. German and Japanese illustrate that the two conceivable ways of resolving the conflict between MAX(CASE) and Agree are indeed realized.

Thus, in OT, universal principles may sometimes be suspended in certain constructions in certain languages – but only if that is warranted by the need to fulfil more important, equally universal, principles. In the ideal case, one can still observe the effects of the violated universal constraint in those domains in which it cannot be overridden by more important principles. In German, AGREE/EPP becomes visible whenever the need to fulfil lexical requirements is irrelevant, while in Japanese, the importance of MAX(CASE) is obvious whenever the EPP is satisfied otherwise. No constraint is obliterated completely, but they can be inactive, if they are too low in the ranking. Even in languages where the effect of a particular constraint is often rendered invisible higher ranked constraints, the effect of the universal tendency (AGREE/EPP) is observable in the default case, as a consequence of The Emergence of The Unmarked (TETU). This can happen in a

set of particular grammatical contexts, insensitive to the effect of the higher-ranking specific constraints.

The absence of a subject in (3a) seems related to a further principle, Full Interpretation (FI) of Chomsky (1995), formulated as DEP(pronoun) in the preceding chapter, that penalizes structures into which expletive pronouns lacking an interpretation (they were not part of the input) have been inserted. The crucial observations for (3a) have also been discussed in chapter 2: French ranks AGREE/EPP >> DEP(pronoun) , so that expletive *il* can and must fill the subject position when no other noun phrase could do so. In German, on the other hand, DEP(pronoun)>> AGREE/EPP, so that *es* must not be inserted to save a clause from an EPP violation (though *es* may be used in other contexts, when more important constraints such as the one forcing the second position of finite verbs, are at stake).

Classical generative syntax had invented the tool of a "*parameter"* in Universal Grammar, which can turn a principle on or off in some language, either in general, or restricted to certain environments. A well-known example is the pro-drop parameter that regulates whether agreement is "strong" enough to license null pronominals. For the EPP, Chomsky (1995) proposes a parametrization that can render it ineffective (but with consequences different from the ones discussed so far). He postulates a feature for Infl that requires the overt (visible) presence of a subject at surface structure if the feature is "strong", but only the covert presence of subjects at later levels of syntax if the pertinent feature is weak. Chomsky (1993) works with the idea that nodes like Infl (which is responsible for nominative case, and which forces the presence of a subject) may become 'inactive' under different conditions in different languages. The conditions specified in the constraints are, under this alternative view, universal, but they are set into effect in very different ways in different languages. In the worst case, a parameter would simply state that something "must/need not" be the case, depending on the language. Such a parameter has little empirical content, but it at least restricts what kind of conditions a language *may* have to satisfy. More convincing proposals for parameters specify additional conditions for the application of the constraint. Thus, for case marking/checking, the direction of case marking (does the object precede or follow the verb) and adjacency (is it possible for an adverb to intervene between the verb and the object) may be subject to parametrization.

Local constraint deactivation in terms of parametrization in the GB model differs from local deactivation by higher principles in OT in an obvious way: there is no theoretical limit to parametrization in the GB-model. In principle, whenever a constraint C motivated in languages L_1 … L_m is disrespected in language L_n, C may be "parametrized" accordingly. Thus, there is no data that

could refute parametrization theory on principled grounds. This is different in OT: if C motivated in L_1 … L_m is violated in L_n one must be able to identify the constraints D_1 … D_k that override C in L_n – and these constraints must be compatible with what we know about $L_1 \ldots L_m$ as well! Thus, some possible constellation of facts could not be reconciled with OT (C being violated without there being identifiable universal and higher ranked principles responsible for it), which means that OT makes *empirically stronger claims* than parametrized UG.

Another reaction open to classical generative grammar to the problem of identifying universal principles would be the addition of more and more complications to the principles of grammar. If one *really* wants to formulate the EPP such that it holds universally and in an unviolated fashion, one will certainly find a way to do so - but one then needs to restrict the effects of the EPP to (a) SVO languages (because subjects are not always mandatory in the SOV language Dutch), to (b) constructions without the need to fulfil lexical requirements (because SVO Russian may very well have subjectless clauses with a dative argument only), to (c) languages with an impoverished inflectional system (because of Italian *piove*, 'it rains'), and (d) to active agentive clauses (because accusative objects are licensed as an alternative to subjects in passive and non-agentive clauses in Hebrew). It is not clear what interest such a constraint may have. One extreme reaction to the descriptive diversity of natural language lies in the identification of very complex empirical generalizations. According to OT, this is not the correct approach: it is not the constraints that are complex[1] – it is their interaction. The conflictual nature of constraints in OT allows to identify

---

1 "Classical" generative syntax reacts to the difficulties discussed above in further ways also. It postulates more abstract structural representations with "invisible" elements, so that one can argue that the constraints is not violated at all under a closer look. Thus, following insights or Rizzi (1981, 1986) , it is customary to assume inaudible subjects in such examples (see e.g. Grewendorf 1989 for German):

(i) dass pro mir geholfen wird
that me.dat helped is
dass pro getanzt wird
that danced is

Likewise, recall that Rizzi (1990) argues for a constraint (the wh-Criterion) according to which the Spec,CP of a question must be filled by a wh-phrase (occupying their scope position). Where this constraint is, apparently, violated, as in Chinese (ii), one can again argue that an *invisible* wh-operator fills Spec,CP position in Chinese questions.

(ii) Zhangsan xiangxin **shei** mai-le shu
Z believe who bought books
"who does Zhangsan believe bought books"

The Chomskyan tradition is rich in attempts to show that these empty elements are independently motivated, but whether these attempts have always been successful is an issue we will not go into here.

well-motivated and simple universal principles much more easily than in alternative approaches.

3.2 Markedness as a basis for the universal component of OT

The constraints of the EVAL component of OT are universal, but universality does not rule out violability. In fact, violability is a prerequisite for universality. Universal constraints thus represent, from a descriptive point of view, tendencies that languages try to follow whenever they can. This concept of universality is more reminiscent of *markedness theory* than of non-violable universality, it is thus also more related to phonological and morphological than to (mainstream) syntactic theorizing (although the syntactic types of typology also represent an unmarked ideal that languages can meet up to a certain degree only. By its strong generative orientation, OT consequently offers a way to reconcile more typological approaches working with statistical tendencies with hardcore generativism (see, e.g., Haspelmath 1999c).

The notion of markedness has arisen in linguistics in the twenties (of the last century) in the framework of the Prague's school and has from the start been related to *neutralization*: A neutralized phoneme loses or lacks a property, expressed as a feature, and becomes or is less marked than a non-neutralized one. In pairs of marked vs. unmarked segments, the marked segment always contains an extra property like roundedness, nasalization, voiceness or aspiration. In the pair [t/d], [d] has the property of being voiced, and is thus more marked than [t]. In other words, *the segment with the fewer features*, in this case, [t], *is the unmarked one*, because it is less complex. Markedness is a relative concept which makes only sense in a set of elements contrasting with each other. According to Trubetzkoy's (1939) original definition, both [t] and [d] can be neutralized in the syllable-final position as a consequence of Final Devoicing: the feature [voice] is lost at the end of a syllable (as in German *Kin[t]* vs. *Kin[d]er*).2

Greenberg (1966), inspired from Zipf's observations, consolidated the terminology and applied the markedness concept to syntactic, morphological, and other fields. He found that *the unmarked members* of the marked-unmarked pairs *are more frequent* than the marked elements. Unaspirated, non-glottalized, unvoiced, oral, and short phonemes are more frequent than their aspirated, glottalized, voiced, nasal and long counterparts. In morphology, singular is less

2 Trubetzkoy thought first that the neutralized segment resulting from Final Devoicing was an extra-segment, called archiphoneme, though he abandoned the idea of a third segment resulting from neutralization later.

marked than plural, and present tense less marked than past tense (see also Wurzel 1980). In syntax, the order subject-object is much more frequent than object-subject. Thus, SVO, SOV, VSO are frequent as canonical orders, but OVS, OSV, VOS are rare.

In semantics, prototype theory can also be considered an instance of markedness theory. Some semantic features are more unmarked than others which make some representatives of a class better prototypes. Greenberg also introduced the *implication* criterion with respect to the universality of unmarked elements as compared to the relative rarity of marked ones: If the existence of some segment, feature or property always implies the existence of another one, the latter is more unmarked than the former. Languages have basic color terms for orange only if they have also words for red and blue (but not vice versa, see Berlin & Kay 1969), verbs show agreement with the direct object only if they also agree with the subject (but not vice versa REF), languages that have nasal vowels also possess oral ones (Ladefoged & Maddieson 1996). All these facts constitute evidence for the relative markedness of the categories involved.

In a similar vein, the results of assimilation and other phonetically motivated processes go in the direction of markedness. The unmarked feature neutralizes and assimilates more easily, a fact correlating with the greater stability of the marked features and segments, and the default articulation of the unmarked ones.

Articulatory ease is also used to predict which of two segments is the unmarked one: an unmarked segment should be easier to articulate than a marked one. This is probably not a very reliable criterion, since articulatory ease is difficult to define in the absence of secondary articulation. The segments [y] and [ø] are neither more nor less difficult to articulate than [u] or [e], but the former segments are rarer than the latter ones. This fact correlates with the frequency of unrounded front vowels and rounded back ones as opposed to rounded front vowels (or the even rarer unrounded back vowels). Moreover, clear cases of articulatory ease go together with featural complexity: a consonant with a secondary articulation is probably more ‘difficult’ to articulate than a consonant with a single articulation. But then, it is also more complex from the point of view of the number of features. For syntax, the ease or difficulty of producing certain structures has recently become the topic of intensive research in psycholinguistics. Results such as the greater processing difficulty of object-subject order in German (as compared to subject-object-order) suggest there might be something to the idea that more marked syntactic constellations are also more difficult to produce and perceive.

Finally, Jacobson observes that the same segments which are universally acquired first are often lost last in aphasic disorders. *Order of acquisition and of loss of linguistic elements* thus became a further criterion for markedness.

Some examples of markedness are given in (4). We will see below that these markedness hierarchies are expressed by inherent internal rankings of the relevant constraints.

(4) a. Oral vowels are less marked than nasal ones (all languages that have nasal vowels also have oral ones).

b. Voiceless obstruents are less marked than voiced ones (obstruents have a tendency to be voiceless).

c. Voiced sonorants are less marked than voiceless ones (sonorants have a tendency to be voiced).

d. Open syllables are less marked than closed ones (all languages have open syllables, but not all have closed syllables).

e. Nominative is less marked than accusative case (when only one structural case is needed, then nominative).

f. Present tense morphology is less marked than past (present morphology is usually non-existent).

g. Active is less marked than passive (passive arises only under special circumstances).

h. Subject initial order is less marked than object initial order

In OT, the role of markedness principles is considerably extended. They represent the tendency of structures to fulfil formal regularities and predictability, even relatively marked ones. As shown in the preceding chapter, OT uses the inherent conflict between markedness and faithfulness constraints as one of its most fundamental formal tenets. Since markedness requires simple structures and faithfulness requires that inputs be realized the way they come, regardless of their complexity, the two kinds of constraints often impose diverging demands on outputs. Individual grammars differ as to which option they choose.

In section 2.2 it was furthermore shown that two different kinds of markedness constraints can compete with each other. For instance, best foot and best syllable are sometimes incompatible, a fact leading to the important observation that markedness is not an absolute notion but is often defined with respect to a certain structure in another related domain. In the case mentioned, a syllable can be marked or unmarked in terms of the segments composing it, their linear organisation, or relative to the foot structure it appears in. Further criteria enter the decision of what a best syllable should look like. For instance, a word-initial syllable may need more or different segments than a word-final one. Thus,

unmarked structure is generally not definable in absolute terms, but needs a point of reference. It will be shown below how the technique of harmonic alignment makes use of this important observation.

OT's architecture predicts that a marked structure can win over a less marked one under two circumstances: either faithfulness considerations are involved (in which case a faithfulness constraint dominates a constraint against a marked structure), or a higher-ranking markedness constraint forces the emergence of the marked structure. For example, nasal vowels are universally more marked than oral ones (5a). However, they can be part of the inventory of a language, as in Polish and French, a fact accounted for by highly ranked faithfulness to input nasal vowels, as shown in (5b). Alternatively they may emerge as a consequence of assimilation, as in English (5c), see Kager (1999) for a detailed exposition of how OT deals with such cases. The constraints used in (5) should be self-explanatory.

(5) Oral and nasal vowels
a. Universal: *NASALVOWEL >> *ORALVOWEL
b. Polish, French: MAX(NasalVowel) >>* NASALVOWEL >> *ORALVOWEL
c. English: AGREE(Nasal) >> * NASALVOWEL >> *ORALVOWEL

Both in (5b) and in (5c) the universal ranking (5a) is respected, but a higher ranking constraint forces the realization of the marked segments (beside the unmarked ones). In Polish and in French, *input* nasal vowels are protected by MAX(NasalVowel), a faithfulness constraint. In English, nasal vowels arise as a consequence of AGREE(Nasal), a high-ranking markedness constraint, requiring nasal vowels as a result of *assimilation* with adjacent nasal consonants. Notice that AGREE(Nasal) is low ranking in French, since oral vowels never become nasal as the result of assimilation (though they can fuse with an adjacent nasal consonant, as in *bon* [bɔ̃] as opposed to *bonne* [bɔn] 'good, masc./fem.' in which the vowel is oral). In English, MAX(NasalV) is ranked below *NASALVOWEL, prohibiting the realization of input nasal vowels. Compare the following tableaux.

(6) Nasal vowels in French

sɛ̃k *cinq* 'five'	MAX (NasalV)	* NASAL VOWEL	AGREE (NasalV)	*ORALVOWEL
☞ a. sɛ̃k		*		
b. sink	*!		*	*

(7) Nasal vowels in English

can't	AGREE (NasalV)	*NASAL VOWEL	MAX (NasalV)	*ORALVOWEL
☞ a. cãn't		*		
b. can't	*!			*

Consider next further examples for the interaction of markedness. To this aim, let us return to China and meet L again, at a later point in her life. She has learned to place sentence stress and made considerable progress in German grammar. She is now trying to figure out in which ways Mandarin Chinese and German differ and compares what seems natural to her in both languages in concentrating on the following two aspects: first syllable structure, especially codas and final devoicing, and, second, word order.

Mandarin Chinese syllables have no codas. As observed by Broselow et. al (1998), when speakers of this language learn a language with codas, they typically apply several repair strategies. One repair consists in adapting the syllable structure of the target language to the Chinese pattern, either by deleting the coda or by adding an epenthetic segment after the offending consonant. This leads to the elimination of codas. Another strategy endorsed by speakers learning English is to devoice final voiced obstruents, and this leads to universally unmarked codas. This is remarkable since Mandarin Chinese, having no codas with obstruents, also has no final devoicing, and English tolerates voiced codas. Thus neither the target nor the source language has final devoicing. However, it is such a natural process that it emerges spontaneously in an intermediate grammar such as those observed in L2 acquisition. In German, final devoicing is part of the grammar, which eases L's task, as compared to Chinese speakers learning English who have to go one step further away than L from what they do in their own language. To see how markedness violations can be calculated, compare the

overview in (8) in which is is shown that Chinese is the less marked language, German is in-between, and English is more marked than than both Chinese and German - at least as far as codas are concerned. This is because English has codas which can be voiced obstruents, German has codas, but does not accept voiced obstruents, and Mandarin, having no codas, also has no voiced obstruents. In the overview (4), 'yes' indicates markedness, and 'no' unmarkedness.

(8) Markedness in codas

	Codas	Voiced Obstruents in the Coda
Mandarin	No	No
German	Yes	No
English	Yes	Yes

In OT, avoidance of both codas and voiced obstruents in the codas is achieved by markedness constraints, the effect of which are felt universally, but in different proportions in individual languages, as illustrated in (5). In Chinese, both prohibitions are high-ranking, which means that the relevant markedness constraints are undominated (5a). In German, codas are avoided when possible, but there is no absolute prohibition against them. Words like *Auto* 'car' or *Opa* 'grandfather' are syllabified as [au.to] and [o.pa], without coda, and not [aut.o] and [op.a] (see Prince & Smolensky 1993, who cite Jacobson, for the observation that this is the case in all languages), but words like *Bein* 'leg' and *rot* 'red' are syllables with codas: [baɪ̯n], [ro:t]. However, in German, Final Devoicing is high-ranking, as has been discussed extensively in Chapter 2. In English, neither NoCoda nor FinalDevoicing is high-ranking, though morpheme-internal syllabification proceeds like in German in avoiding codas, and the markedness of voiced obstruents is not cancelled, but noticeable in language acquisition. As always, **Faith** stands for the responsible faithfulness constraints requiring marked inputs to be present in the output. (9) illustrates how the three languages rank the relevant constraints, and thus how the differences in markedness noticed in (8) is reflected by the ranking of the relevant constraints. Chinese forbids codas altogether and, appropriately, ranks NoCoda at the top position. German has the ranking motivated in Chapter 2, with Faith to the voicing sandwiched between FinalDevoicing and NoCoda, and finally English, allowing all kinds of codas, ranks Faith at the top position.

(9) Typology arising from NoCoda and FinalDevoicing
   a. Chinese: NoCoda >> **Faith,** FinalDevoicing
   b. German: FinalDevoicing >> **Faith** >> NoCoda

c. English: **FAITH** >> FINALDEVOICING, NOCODA

In the syntactic domain of word order, we have already introduced the major principles distinguishing the two languages. Both Chinese and German are subject initial – that is, in neutral sentence, the subject comes first. Given that subjects are specifiers of the clause, it seems that both Chinese and German attribute a high rank to Align (Specifier, Left, XP, Left) (SPECLEFT) – which represents a markedness constraint in the alignment domain, as discussed above, because, as has been motivated in Chapter 2, rightward alignment constraints seem non-existent.

(10) Align (Specifier, Left, XP, Left)

Objects typically follow the verb in Chinese, while they precede it in German (complement) clauses. Recall from the preceding chapter that this variation can be captured in terms of a difference in the ranking of leftmost head alignment and leftmost complement alignment

(11) a. Align (Head, Left, XP, Left) (HEADLEFT)
b. Align (Complememt, Left, XP, Left) (COMPLEFT)

In Chinese, the ranking HEADLEFT > COMPLEFT implies that verbs should precede their objects, the opposite (German) ranking yields the inverted order object-verb. Apparently, leftmost linearization is unmarked for *all* elements. The conflict between the principles in (11) is thus one among two markedness constraints. Which of the two constraints is ranked higher seems arbitrary from the point of view of universal grammar – there seem to be as many languages in which the head precedes the complement as there are languages in which the reverse holds. On the other hand, the high rank of SPECLEFT represents a universal tendency (nearly 90 percent of the languages of the work are subject initial).

(12) Chinese ranking
SPECLEFT > HEADLEFT > COMPLEFT

German ranking
SPECLEFT > COMPLEFT > HEADLEFT

The ranking difference in (12) is not the only factor that distinguishes Chinese from German word order. There is a fundamental difference between the two languages as to their readiness to linearize phrases on grounds different from (10) and (11). Chinese does not show much rearrangement whereas German organizes

word order according to a number of further factors, some of which already mentioned in Chapter 1 and 2. Wh-words can always remain *in situ* in Chinese but appear in the leftmost position in German. This was explained in terms of the wh-criterion (13), which is ranked higher that the principle Stay forbidding movement in German, but which is lower in Chinese.

(13) WHCRIT
The specifier position of a question must be filled by a wh-phrase!

Again, it seems quite hard to establish which of the two possible rankings of STAY and WHCRIT is the "simpler" one in natural language. Nevertheless, Chinese question formation *is* more marked than the German one in a certain sense. Note that WHCRIT can be satisfied not only by moving a phrase into Spec.CP (violating STAY) – as Müller (1999) observes, the insertion of an expletive wh-phrase would do the same job. It is an option also used in German.

(14)	was	denkst	du	wen	sie	liebt
	what	think	you	who	she	loves

A wh-expletive in Spec.CP does not violate STAY, but fails to respect DEP(wh-phrase): do not insert wh-phrases that have not been in the input. Apparently, Chinese ranks *both* DEP(wh-phrase) and STAY above WHCRIT, because nothing needs to change in a question clause, as illustrated in (15).

(15) Zhangsan xiangxin **shei** mai-le shu
Zhangsan believe who bought books
'who does Zhangsan believe bought books'

As Cheng (1991) observes, languages that do not move wh-phrases (STAY > WH-CRIT) usually use scope markers (wh-elements inserted into Spec.CP) in questions, implying that WH-CRIT > DEP(wh-phrase). Likewise, languages that move wh-phrases in certain contexts (WH-CRIT > STAY) rarely employ additional wh-elements as scope markers in these constructions (DEP(wh-phrase) > WH-CRIT). Thus, both the rankings wh-crit > {STAY, DEP(wh)} (represented by German) and {STAY, DEP(wh)} >> WH-CRIT (represented by Chinese) are rare among the world's languages – perhaps not because of truly formal but rather for functional reasons – (the semantic scope of the wh-phrase fails to be indicated in Chinese).

Monolingual speakers may well not be aware of linguistic tendencies. It is only in comparing languages that such generalizations come to light. L, an alert second

language learner, is in a good position to take a glance at the diversity of languages, as well as their similarities since Chinese and German make different choices in many respects, a few of which have been introduced in the preceding pages. An evident question that L is likely to ask concerns the origin of markedness, as well as the best method to formulate principles in the form of universal constraints.

At the present state of knowledge, a complete list of markedness constraints is not a realistic goal, but some trends can be identified, and they suggest a very interesting perspective on the nature of markedness constraints.

In phonology, many processes and alternations, as well as inventories are motivated by phonetic facts, of acoustic, perceptive and articulatory nature. Markedness principles reflect states of affair that are easier to produce, perceive or acquire for reasons deeply rooted in articulatory and perceptory phonetics. Constraints that have this property are said to be *grounded*. We have already presented some suggestive evidence for this in the context of final devoicing: the production of a voiced obstruent is more difficult than the one of its voiceless counterpart in a coda position, and it is not really worth while because of phonetic facts. Likewise, the constraint *NASALVOWEL is motivated on articulatory grounds. Activity of the soft palate is involved in the production of nasal vowels: it is lowered in order to let the airstream escape through the nose. This additional articulatory gesture as compared to plain oral vowels renders nasal vowels more marked than oral ones which explain the relative rarity of nasal vowels in the languages of the world. There are not only constraints that have a bearing on the inventories of linguistic elements and structures. A second group of constraints, active in phonology, express assimilation between adjacent segments which have an obvious articulatory explanation as well: One articulatory gesture is more economical than two. Assimilation between adjacent, and even non-adjacent segments in the case of vowel harmony, is readily interpreted as coarticulation in phonetics, which itself can be interpreted as the avoidance of effort that results from rapid movements required to move from one articulation to another (see e.g. Lindblom 1989, Flemming 1995).

(16) Assimilation
AGREE(Nasal): A vowel preceding a nasal consonant is nasal.
VOICINGASSIMILATION: Obstruent clusters must agree in voicing

It may, thus, not be too far off the track to suspect that most if not all markedness constraints of phonology are grounded in articulatory or perceptive laws in one way or the other. A similar picture may emerge for some parts of the syntax, too.

The standard view held in generative syntax is that its principles are irreducible fundamental laws of the language faculty. Typologists would rather say that the laws languages tend to observe are derivable from consideration of simplicity in processing and other respects. Consider, in this context, the determination of scope in natural language. Simplifying matters a bit, we may observe first that the scope relations among phrases that are not moved syntactically correspond to their surface structure hierarchies in German clauses with "neutral" intonation (17).

(17) a. dass jeder Mann eine Frau liebt
that every.nom man a woman loves
"that every man loves some woman"
Only: ∀x∃y (man(x) -> (woman(y) & loves (x,y))

b. dass ein Mann jede Frau liebt
that a.nom man every woman loves
"that some man loves every woman"
Only: ∃x (man(x) & ∀y woman(y) → loves (x,y))

The data in (17) suggest that natural language obeys a scope principle such as (18)

(18) X has semantic scope over Y, iff X c-commands Y in the syntax
(X c-commands Y iff the node immediately dominating X also dominates Y)

In (17) the subject takes scope over the object, and it c-commands this object as well, as subjects usually do. Languages seem to differ as to whether (18) must always be respected (Hungarian, partially Chinese), or whether certain environments license violations of (18) (there are more of these in English than in German). This issue need not concern us here – more important in the present context is that (18) prescribes a strict correspondence between a semantic relation (scope) and a syntactic one (c-command). One may say then that the syntactic law (18) is grounded in semantics – perhaps, because (18) allows that the element that takes wide scope can be processed first in online language understanding. If X c-commands Y, it will normally precede Y. Thus, X is perceived before Y in online processing. Since processing is incremental, it is advantageous if the interpretation of an element X does *not* depend on the interpretation of an Y *following* X. Therefore, it is advantageous if X is not in the scope of Y when X precedes Y. This is, however, essentially, what (18) states. Similarly, WHCRIT may be understood as a response to the functional need of setting questions apart from assertions, and of marking the scope of a wh-phrase. There is a universal tendency

to place topical material at the left edge of the clause, which can also be interpreted as being functional.

It has also been shown in this chapter and the preceding one that Alignment constraints target the initial (left) position of sentences rather than their final (right) position. In other words, all constituents compete with all others in their desire to be realized as early as possible. On the other hand, temporal constraints on the articulatory and perceptive tracts severely restrict what can be uttered at one time. Psychological grounding can be used to explain the earliness requirement of production: communication should happen as quickly as possible, without delay. Since this is true of every part of every sentences, we have an explanation of the left-orientation of the alignment constraints.

OT syntax thus offers a perspective on explanation in syntax that models like the Government and Binding Theory did not possess: laws of syntax may be explained in terms of more fundamental organizational principles of language and cognition. It thus comes close to typological models of language, but offers a formal way of capturing what a "tendency" like (18) is. If laws of syntax cannot be violated, the existence of ambiguous structures such as German (19) or English (20) would force a complex formulation of (18), for which a grounding in processing would then at least be much more difficult to achieve.

(19) dass einen Mann jede Frau liebt
that a.acc man every woman loves
'that every woman loves a man'
Both ∀x∃y (man(x) -> (woman(y) & loves (x,y))
and ∃x (man(x) & ∀y woman(y) → loves (x,y))

(20) that some voter expects every Republican to win

Whether a substantial portion of the syntactic principles can and should partly be grounded in semantics or other parts of perception is a open issue that we cannot deal with here. Discussions in OT syntax typically do not highlight the issue of grounding. Furthermore, OT offers quite a different perspective on the processing-syntax relation, as Smolensky & Stevenson 1995, Fanselow et al. 1999 and others have observed: if the online processing of sentences involves a process that is reminiscent of the architecture of OT, parsing ease may be explained in terms of syntactic laws (and not vice versa, as in grounding approaches) – as was originally envisaged by Pritchett 1996 within the limits of the Government and Binding approach. Suppose that violable principles such as EPP/AGREE are directly applied in online processing, and suppose that they try to be satisfied as early as possible – in order to maximize the harmony of the partial structures computed

incrementally. Then it will follow that one tries to locate subjects in a clause immediately, because EPP/AGREE will be violated unless such a subject is found. The preference for subjects appearing as early as possible in sentences in one of the earliest laws of psycholinguistics (see Frazier & Flores d'Arcais 1989).
In other words, the violable nature of OT syntax principles implies that they can be formulated in a fashion which makes grammar and processing laws appear more intimately connected than in other approaches. Whether this corroborates a grounding view of grammar, or whether it implies a grammatical explanation of processing difficulty, or whether "grammar" and processing merely reflect different perspective on the same object is still an open question that deserves more attention than it currently gets.

3.3 Factorial Typology

While grammatical constraints are universal in OT, constraint ranking is not. Rather, each of the logically possible rankings constitutes a possible natural language grammar, and natural language grammars differ by the ranking of the same set of principles.

When one considers a set of constraints, one can therefore figure out which kinds of languages are made possible by the different rankings, in other words, identify the *factorial typology*, viz. the total amount of permissible permutations of interacting constraints. This section illustrates factorial typology with well-known facts from syllable structure, but we return in section 3.4 to an in-depth reflection on the role of markedness considerations in OT and universal grammar, and how they affect factorial typology.

Jacobson has observed that, universally, the most unmarked syllable is an open one with an onset (CV). This kind of syllable can be found in all languages. Even if syllables without onsets and with codas are tolerated in many languages, such syllables only appear if a marked syllabification cannot be avoided, given the constraint hierarchy. Thus the fact that no language disallows CV syllables or prefers syllables with codas over those without codas must follow from universal grammar.

A sequence CVCV is in the default case and in all languages syllabified as CV.CV. But languages differ as to what they do with inputs consisting of CVC or VC. Some languages avoid syllables without onsets, or such with codas to a much greater extent than English, and systematically use strategies to repair 'bad' syllables, like those coming from loanwords for instance. Japanese augments *Christmas* to *kurisumaisu* and *Arbeit* 'homework' to *arubaito* and in Hawaiian, the expression *happy Christmas* is *mele kalikimaka*, *ticket* is *kikiki* and *Samuel* is

Kamuela. As should be clear, Japanese and Hawaiian use the unmarked option. At the other end of the scale, some languages are faithful to the input's segments. Extreme cases of faithfulness are German with very complex syllables *Strumpfs* [ʃtχʊmpfs] 'sock, gen.', and Polish with violations of the sonority hierarchy, as in *zmierzch mglisty* [zmʲêx mglist] 'misty dusk'. In these languages, marked syllables are chosen over the option to delete or epenthesize segments. Intermediate solutions between Polish and Hawaiian are also represented in some languages, like the prohibition against complex onsets and codas or the prohibition against just certain classes of segments in the syllable edges.

As mentioned before, French, German, and English tolerate syllables without onsets and with codas, but only if no other syllabification is forced by higher ranking faithfulness constraints. Thus, monomorphemic CVCV are always syllabified as CV.CV or $CVC_{ambysyll}V$ but never as CVC.V. Compare the syllabification of *coma* and *comma* in English, German and French.

(21) Prefered syllabification

English: *coma* (CV.CV), *comma* ($CVC_{ambisyll}V$)
German: *Koma* [kó:.ma] (CV.CV), *Komma* [kɔ́ma] ($CVC_{ambisyll}V$)
French: *coma* [kɔma] (CV.CV) (no ambisyllabicity)

The following discussion concentrates on different choices made by individual languages' grammars about what they do with marked inputs, like /V/ and /VC/. We follow Prince & Smolensky's (1993) discussion of syllabification, and provide additional examples of hiatus avoidance

Let's consider first the input /V/ and how it is syllabified. Three options of how to deal with CV+V emerge in a context in which only the constraints ONSET, MAX and DEP are considered. The relevant environment is one where the input /V/ is just after a syllable CV, thus in a hiatus position. With 3 relevant constraints, the factorial typology, that is the total number of permutations, delivers 6 rankings (3! = 3 x 2 x 1), but only 3 different syllabifications emerge from these 6 rankings. The first option is that the syllabification respects the input and nothing happens: a hiatus is realized. This is illustrated in (22a) with an example from Maori. Second, one of the vowels can be deleted, and just one syllable CV is pronounced, like in the French examples in (22b). The third option is one in which a consonant is inserted between the two vowels, delivering an output consisting of two consecutive CV syllables, as in the German example (22c).

(22) Hiatus tolerance or avoidance

a. Faithfulness in Maori: /puea/ –> [puea]
b. Deletion in French: /lə ami/ –> [la.mi] 'the friend',
/jə la adɔr / –> [jə.la.dɔr] 'I adore her'
c. Epenthesis in German: /beamtə/ –> [bəʔamtə] 'civil servant',
/ge+atmet/ –> [gəʔatmət] 'breathed'

In the following tableaux, the two constraints leading to fatal violations are shown as tied constraints, as a shortcut for two rankings. Thus MAX >> DEP >> ONSET and DEP >> MAX >> ONSET deliver the same optimal output, namely the one which violates the lowest ranking constraint, in this case ONSET. In the example shown here, violations are always double because there are two hiatus, of course the same result obtains with just one hiatus, since the first candidate, the optimal one, does not violate one of the higher ranking constraints.

(23) Two rankings delivering faithfulness (as in Maori):

DEP >> MAX >>ONSET
DEP >> MAX >> ONSET

/V/ puea	DEP	MAX	ONSET
☞ V puea			**
Ø pu		*!*	
CV puteta	*!*		

(24) Two rankings delivering deletion (as in French):

ONSET >> DEP >> MAX
DEP >> ONSET >> MAX

/V/ le.ami	DEP	ONSET	MAX
V le.ami		*!	
☞ Ø lami			*
CV letami	*!		

(25) Two rankings delivering epenthesis (as in German):

ONSET >> MAX >> DEP
MAX >> ONSET >> DEP

/V/ gəatmət	MAX	ONSET	DEP
V gəatmət		*!	
Ø gatmət	*!		
☞ CV gəʔatmət			*

If we consider four constraints, viz., ONSET, NOCODA, MAX, and DEP, there are 24 (4!) possible permutations. But not all differences are of interest, because, as shown by McCarthy and Prince 1993, some constraints never conflict with each other.

- ONSET and NOCODA, for instance, do not conflict (they both impose .CV.)
- If both MAX and DEP dominate ONSET and NOCODA, the ordering of the faithfulness constraints is indifferent. In other words, MAX >> DEP >> { ONSET, NOCODA } and DEP >> MAX >> { ONSET, NOCODA } are equivalent.
- MAX conflicts with DEP only if ONSET or NOCODA dominates one of them

Consequently, there are exactly three different types, as before: deletion, epenthesis and faithfulness. Since ONSET and NOCODA do not conflict, all rankings in which the relevant one of the two is ranked lowest lead to faithfulness. (26) lists all rankings delivering deletion.

(26) All rankings delivering deletion:

DEP >> ONSET >> NOCODA >> MAX
DEP >> NOCODA >> ONSET >> MAX
ONSET >> NOCODA >> DEP >> MAX
NOCODA >> ONSET >> DEP >> MAX
NOCODA >> DEP >> ONSET >> MAX
ONSET >> DEP >> NOCODA >> MAX

We refrain from listing all rankings delivering deletion and epenthesis, since the result is trivially simple to calculate. Important is which is the lowest ranking constraint, since it is the one which can be violated by the optimal candidate.

Since all other constraints eliminate candidates, it does not matter in which order they come.

Consider next an input which has two medial consonants of which the first one is expected to be a coda, if parsed, because of the sonority relationships between the two consonants. Again three cases, deletion, epenthesis and faithfulness, are shown to be predicted by the factorial typology.

(27) Deletion

/takti/	ONSET/ NOCODA	DEP	MAX
a. tak.ti	*!		
b. ta.ka.ti		*!	
☞ c. ta.ti			*

(28) Epenthesis

/takti/	MAX	ONSET/ NOCODA	DEP
a. tak.ti		*!	
b. ta.ka.ti			*
☞ c. ta.ti	*!		

(29) Faithfulness

/takti/	MAX	DEP	ONSET/ NOCODA
☞ a. tak.ti			*
b. ta.ka.ti		*!	
c. ta.ti	*!		

Notice that given the set of constraints, no language can exist that has only syllables with codas but without onsets. This has nothing to do with the constraint ranking, but with the kinds of constraints used: onset are required and codas are forbidden. And, as desired, given the fact that all languages have consonants and vowels, there is also no language without CV syllables.

Returning to the hiatus avoidance data in (22), we add some further options and observe that we need more constraints to account for these new data. In (22d), a glide is inserted between the two offending vowels, in (22e) the first (high) vowel is turned itself into a glide, whereas in (22f), it is shown that vowel deletion is sometimes accompanied by compensatory lengthening. One of the vowel is elided and the other is lengthened, thus occupying prosodically the vacant place left by the elided vowel.

(22) More on hiatus avoidance

d. Glide formation in French: */ʒɔliə/* 'pretty' –> [ʒɔlijə] in songs and poems, or in emphasis.
e. Glide formation in German: /radio/ –> [ra.djo]
f. Compensatory lengthening in LuGanda: *bu ami* 'the chiefs' –> [baami]

In terms of constraints, the three additional options fulfill ONSET, but not by inserting an unmarked consonant, as was the case in the German examples, and also not by just deleting one of the vowels without leaving a trace, as was the case in French. Still the options shown in (22d-f) are relatively unmarked, and well-documented cross-linguistically. Glide formation out of one of the vowel (generally the first of the hiatus sequence) is also attested in Dutch (bioscoop –> bi[j]oscoop 'movie'), and in Hawaian (/puaʔohi/ –> [puwaʔohi] 'chatter'), for instance. Glide formation of the first high vowel is an extremely common process, also found in Dutch, English, French, Italian, etc. Similarly, compensatory lengthening accompanying vowel deletion is a universal phenomenon, documented a.o. in Turkish, Old English and a number of Bantu languages. To account for them, additional constraints are needed, to the effect that ONSET will be fulfilled in glide formation 1 by allowing some vocalic features to spread into an inserted consonantal position, in glide formation 2, that vocalic features are changed into consonantal ones and that, in the case of CL, faithfulness will be respected, if not in the segmental input, at least in the moraic or skeletal input. This of course increases the factorial typology by at least 3 constraints: 7! (4 old + 3 new) = 28 and increases the number of possible grammars accordingly. We have shown 3 options, on top of the 3 we had before, but more options are available, like combinations of the simple cases. However, since at least some of the complex options, say epenthesis plus glide formation, are not attested universally, the question arises how to get rid of some of the possible grammar.

Before we leave the thorny problem of factorial typology for the moment, a final remark should be made. As observed in the context of syllabification, total reranking of all concerned constraints predicts repairs which do not exist. For

instance, marked sequences of segments with respect to one features (nasality, voicing …) are repaired minimally, by changing the problematic feature, though in principle more drastic changes are possible, like metathesis, deletion or even the transformation of unconcerned features etc. Many of these repairs are never attested. It seems that reranking the constraints obeys some limiting principles. In the last section of this chapter, we turn to proposals how how to constrain this reranking power.

Until now, we have considered factorial typologies which deliver fewer results than the number of possible rankings. But this is an artefact from the data considered. Recall the factorial typology from the domain of syntax that was addressed in 2.4. It was shown there that if we work with the alignment constraints HEADLEFT, SPECLEFT, and COMPLEFT, we in fact arrive at 6 different language types corresponding roughly to the six logically possible permutations of the constraints (provided that the verb and the subject may either be found in VP or in IP). This shows that each ranking may correspond to a language type of its own, or the number of systems predicted may be much smaller than the number of rankings, because, as we have seen, different rankings may favor the same constructions.

3.4 Markedness hierarchies and Harmonic Alignment

OT constraints serve the task to describe relative markedness, and when the relevant dimension is a binary one, markedness can be expressed fairly easily. Thus, recall that nasal vowels are more marked than oral ones, a fact we can express by postulation a constraint *NASALVOWEL, universally ordered before *ORALVOWEL. Marked nasal vowels can thus appear only if higher constraints license the violation of this principle.

But markedness distinctions are not *always* binary. When we considered case in German, we introduced the principle *ACCUSATIVE (don't use accusative case), following Woolford (2001) and other proposals, since nominative is the unmarked case. But dative is more marked than accusative – simple transitive verbs never form constructions with the dative – only three place verbs assign dative case regularly. This state of affairs is derived only if *Dative from above is *always* ranked above *Accusative. If this ranking holds, constructions will avoid using the dative for an NP X if accusative marking of X is possible (because it does not violate principles with an even higher rank). But a *universal* ranking of *DATIVE and *ACCUSATIVE is not really in the spirit of OT! This situation may be dealt with in three different ways. First, one might say that the perspective we took on the problem was simply wrong. "Dative" and "Accusative" are mere descriptive

labels for case forms – there is no inherent meaning to them. **Both** markedness constraints *CASE1 and *CASE2 for case yield the case frames <nom>, <nom, CASE1/2>, and <nom, case1, case2> for intransitive, transitive, and ditransitive predicates, respectively, and whether we call the second case appearing with transitive verbs "accusative" or "dative" is just a matter of convention. Cases are arbitrarily ordered according to their markedness. The markedness constraint will than yield the abstract patterns just mentioned, which are the only 'empirical' data we have. Accusative is the label we attach to the secondmost unmarked case, whatever other empirical import that case might bear. Our problem would then just be a terminological one.

It is not likely, however, that this explanation generalizes to other grammatical features. Singular, Plural, and dual stand in a markedness relation, and it is difficult to maintain that plural and dual are terms we may interchange at our discretion in the way we have just discussed for case. Plural is not just the verb form less marked than a dual, it has a clear and grammatically relevant content. The first solution thus must be dismissed, because it is not applicable in related domains. Furthermore, the label-change approach may even be wrong for case. There ARE substantial differences between accusative and dative: the former can freely be replaced by a nominative in a passive and by a genitive in a nominalization, while the latter can be so only in marked languages (Icelandic); dative often needs more morphological realization than accusative, etc.

A second interpretation may be correct for case, but it is again not clear if it can be generalized to all other areas with n-ary markedness dimensions. According to an idea that goes back to Kiparsky and that was worked out in detail by Wunderlich (1997) and Stiebels (2000), one may analyze the different cases as complex rather than primitive entities: nominative case can be found in all grammatical constellations, so we may say it is featureless. Accusative case does not show up on an NP unless there is a "higher" argument in the semantic representation of the predicate – let us assume this is coded by the feature "+higher". A dative does not show up unless there is both a higher and a lower argument – consequently, dative is defined as [+higher, + lower].

Irrespective of their ranking, the pertinent constraints *[+higher], *[+lower] *imply* that a dative (violating a superset of what accusative violates) can be used only when the accusative option is not applicable. Thus, we *derive* the markedness hierarchy nom > acc > dat from the featural composition of the cases. Similar considerations may apply to other markedness scales in the syntax. For instance, analyses such as Sg = {}, Plural = {+plural} and dual = {+plural, + two}, just yield the right results.

Whether this featural approach to markedness hierarchies (the element with less featural content is also less marked) applies in all relevant cases is difficult to

decide, however. In most phonological markedness hierarchies, the constituents acquire gradually more of some property, as will be illustrated by the following examples. In the sonority hierarchy (30a) going from left to right, the segments are becoming more and more sonorous: Stops are not sonorous at all, fricatives are slightly more sonorous, then nasals and so on. Syllables structures in (30b) are also becoming gradually more complex, since more consonants are added at each step. And in (30c) the consonants with secondary articulation have an evident additional component as compared to those with just one primary articulation (see also Padgett 2002).

(30) Markedness hierarchies

a. Sonority hierarchy:
   Stops < Fricatives < Nasals < ... < Vowels
   0 1 2 n
b. Syllable structure
   CV > (C)VC, (C)VC > CCVC, (C)VCC
c. Primary and secondary articulation
   C > C^{j}, C^{g}

Only for (30b) and (30c), an analysis along the lines sketched for Case and number seems obvious: a secondary articulation involves more features, the simpler syllable structures have less structural slots. For sonority, an analysis in terms of differences in the number of features used has been proposed by Clements (1990) who proposes that sonority is measured by the major class features. The more plusses the segments have for these features, the more sonorous they are. In his system, the major class features are the features syllabic, vocoid, approximant and sonorant. Vowels have 4 plusses, glides 3, liquids 2, nasals 1 and obstruents 0. In this system, sonority hierarchy is thus also to be derived from a hierarchy based on independent features. *All* hierarchies in (30) are nonpermutable, representing "meta-constraints", as McCarthy calls them. The constraints forcing these markedness relations cannot be reranked –in contrast to OT's fundamental assumption of a free reorderning of constraints in the hierarchies. We need to identify a theoretical means to capture this situation.

The idea that such universally valid hierarchies– the *markedness hierarchies* – should occupy an important place of their own is very popular in new developments of OT, but the notion was already present in Prince & Smolensky's (1993) original script, as well as the concept of harmonic alignment on which most of the recent literature on the theme grounds. As the term already indicates, markedness hierarchies organize linguistic elements (features, segments,

properties and the like) in hierarchies from the most to the less unmarked. Extending it, the notion has been applied to all kinds of properties, not only to those in which clear markedness in Trubetzkoy's sense is recognizable.

Prince & Smolensky (1993) introduce "Harmonic Alignment", a notion taking advantage of the fact that markedness must be relativized to relevant contexts and that hierarchies have two ends, and thus two directions. In the sonority hierarchy for instance, the segments are clearly ranked with respect to their intrinsic sonority, but *sonority itself* is not marked. In other words, a vowel is by itself neither more nor less marked than a consonant. But it is marked for a syllable peak to be a consonant, and it is marked for a syllable margin to be a vowel. The mirror-like pattern of the kind of segments which may constitute a peak or a margin is expressed with the help of two hierarchies using the same phonetic material but organized in two divergent directions, depending on the kind of syllabic elements looked at, peak or margin.

(31)

Peak: From less sonority to more sonority ('<' means 'more harmonic'):
Stops < Fricatives < Nasals < Liquids < Vowels
Margin: From more sonority to less sonority:
Vowels < Liquids < Nasals < Fricatives < Stops

Before continuing our discussion of harmonic alignment, let us analyze some typological variation of the kind of segments which can be a syllable peak, since these data will be relevant to the discussion that follows. (32a) is universally valid (every language has syllables with vowels as peaks), but (32b) is not (not all languages have syllables with consonants as peaks).

(32) a. Vowels as syllable peaks

b. Consonants as syllable peaks

The question of which kind of segments may appear in C can only be answered in a language-particular fahsion. A first class of languages, exemplified by most Romance languages as well as Hindi, Japanese, etc.accepts vowels only as syllable peaks. French – as Hindi – goes a step further than most other languages in that words like those in (33) prefer to have a coda with increasing sonority rather than to make the liquids [r] or [l] syllable peaks. Though such codas are

marked and impossible in most languages, this is still better than a syllabic sonorant.

(33) French only vowel as syllable peak
ocre [ɔkr] 'ochre' or *siècle* [sjɛkl]

A second class of languages, to which many Germanic languages belong, German and English for instance, have vowels and sonorants as syllable peaks. In English, the words in (34a) have a syllabic sonorant, as is the case in the German words (34b).

(34) Germanic syllabic sonorant
a. English: *eagle, meter, button.*
b. German: *Himmel* [hɪm̩l]'sky'*, eben* [e:bn̩] 'even'*, nieder* [ni:dɐ] 'low'

The third class of languages is much rarer and is exemplified by Imdlawn Tashlhiyt Berber with data from Dell & Elmedlaoui (1985). The data in (35) illustrate the fact that this language tolerates *all* kinds of segments as syllable peaks, including obstruents. All syllables have an onset, except for phrase initial ones, which can be onsetless. This is due to an alignment constraint which is not relevant here.

(35) Syllabification in Imdlawn Tashlhiyt Berber

Voiceless stop	.rA.t**K**.tI.ra-t-kti	'she will remember'
Voiced stop	.b**D**.dL.bddl	'exchange'
Voiceless fricative	.t**F**.tKt.t-ftk-t	'you suffered a strain'
	.t**X**.zNt.t-xzn-t	'you stored'
Voiced fricative	.tx**Z**.nAkk^{w}. t-xzn#nakk^{w}	'she even stockpiled'
Nasal	.tz**M**t.t-zmt	'it (f.) is stifling'
Liquid	.t**R**.g**L**t.t-rgl-t	'you locked'
High vowel	.**I**l.d**I.**i-ldi	'he pulled'
	.rat.l**U**lt.ra-t-lul-t	'you will be born'
Low vowel	.tR.b**A**.t-rba	'she carried on her back'

The hierarchy (36a) from Prince & Smolensky (1993) ranks the segments in a hierarchy reflecting their readiness to be a margin, and (36b) does the same for the peak. For instance *M/a expresses the fact that [a] is not allowed to be a margin. The two hierarchies mirror each other, and are replicating the two hierarchies established in (31).

(36) a. *M/a >> *M/i >> *M/l >>… *M/t
b. *P/t >> … >> *P/l >> *P/i >> *P/a

The scales reflect the observation that in all languages, more sonorous segments make better peaks and less sonorous segments better margins. No language allows obstruents but not vowels, as syllable peaks, for instance. The difference between languages as to which segments can be peaks is explained in the following way. Languages identify a position in the hierarchies, above which no segment can be a peak or margin, as shown in (37). The selection of the turnoff point is a function of the place of the faithfulness constraint MAX. The higher MAX is in the hierarchy, the less tolerant the language is with respect to the kind of segments it admits as peaks or margin.

If MAX is ranked low, a language like Berber comes into being, in which all segments can be peaks. (37a) shows that in Berber, every segment can be a margin, except for [a] which is always a nucleus. (37b) shows that syllable margins in English and in German cannot be of a higher sonority than a glide. In (38a) it is shown that Berber admits all segments as peaks, even voiceless obstruents, whereas English and German tolerate nasals as peaks but no fricative and no stop, as shown in (38b).

(37) Untenable margins:

a. *M/a >> MAX >> *M/i >> *M/l >>… *M/t
The maximum sonority of possible onsets/codas cannot be higher than the sonority level |i|. (Berber)

b. *M/a >> *M/i >> MAX >> *M/r >>… *M/t
The maximum sonority of possible onsets/codas cannot be higher than the sonority level |l|. (German, English)

(38) Untenable peaks:

a. MAX >> *P/t >> *P/n >> *P/l >> *P/i >> *P/a
The minimum sonority of possible nuclei cannot be less than the sonority level of |t|. (Berber)

b. *P/t >> *P/n >> *P/l >> MAX >> *P/i >> *P/a
The minimum sonority of possible nuclei cannot be less than the sonority level of |i|. (French, Hindi)

The individual hierarchies can be combined in one, depending on the language.

(39) Combined hierarchies

a. Berber
*M/a >> MAX >> *P/t, *M/i >> *P/n, *M/l >> … >> *P/a *M/t
b. English, German
*P/t, *M/a >> *P/n, *M/i >> MAX >> *P/l, *M/j >>… >> *P/a*M/t

Faithfulness constraints (as well as other kinds of markedness constraints) can be reordered freely with respect to the constraint sequence in (39). The crucial point is that the scales in (36) cannot be reordered, and that they are mirror images of each other.

Let us consider now how these descriptive tools are formally defined. One presupposes the existence of scales like the sonority hierarchy, (called "dimensions") that are established on a grammar-external basis. The degree of sonority is, essentially, a phonetic fact. From such dimensions, constraint hierarchies may be computed, as given in (40), by *harmonic alignment*. Take the sonority hierarchy as an example: X and Y are peak and margin, and {a,b,...,z} segments. D1 and D2 are the opposite rankings of the segments as they are hierarchically organized in peaks and margins. H_X and H_Y are the hierarchies expressed in the form of harmony scales.

(40) Harmonic alignment (Prince & Smolensky (1993):

Given a binary dimension D_1 with a scale X > Y on its elements {X,Y}, and given another dimension D_2 with a scale a > b > ... > z on its elements {a,b,...,z}, then the *harmonic alignment* of D_1 and D_2 is the pair of harmony scales H_X, H_Y:
a. H_X: X/a > X/b > ... > X/z
b. H_Y : Y/z > ... > Y/b > Y/a

Thus, H_{peak} is, essentially what we find in (36): Peak/a > Peak/i …. > Peak/t. The harmonic alignment translates into a constraint hierarchy by entering a constraint *X/j for each element in the harmony scales. Since X/a is better than X/b, the constraint *X/b banning b as an X must be ranked higher than *X/a:

The constraint alignment is the pair of constraint hierarchies C_X, C_Y:

(41) Constraint alignment

a. C_X: *X/z >> ... >>*X/b >> *X/a
b. CY : *Y/a >> *Y/b >> ... >> *Y/z

Harmonic Alignment is not just a matter of phonology, it has also been applied in different domains of syntax, like the realization of arguments as subjects and objects and the concomittant modus (Aissen 1999), weak and strong pronouns (Müller 1999), pronouns and anaphors (Wilson 1999), null and overt pronouns (Artstein 2000) to cite a few examples. An interesting property of interleaving different markedness hierarchies is that it can replace faithfulness. It is thus a step in the direction of eliminating properties in the input, and the need to be faithful to them.

Artstein (2000) uses markedness hierarchies and harmonic alignment to account for the licensing of null subjects in Hebrew, Irish and Italian, as well as complex cliticization in Sesotho. In many languages, subject pronouns can be left unrealized quite freely, independent of their featural content (as in Italian or Spanish). As (13) illustrates, only a local subject pronoun (=1st or 2nd person) can be dropped in the past and in the future in Hebrew, while a nonlocal (3rd person) one cannot.[3] There seem to exist no languages in which only non-local pronouns can be dropped, but not 1st or 2nd person ones.

(42) a. axalti banana
ate.1s banana ('I ate a banana')
b. *axla banana
ate.3fs banana ('she ate a banana')
c. ata/*pro* toxal et ha-banana
'you will eat the banana'
d. hi/**pro* toxal et ha-banana
'she will eat the banana'

This implicational relation can be captured as follows. First, the phonetic (non-)realization of a pronoun clearly constitutes a dimension like (43a). Furthermore, there is a person scale such as (43b), which manifests itself, e.g., in

3 The present has no pro-drop, probably due to the homophony between the 1st and the 3rd person; even the inflectional prefixes are becoming homophonous (the 1s glottal stop in the inflectional prefix becomes homophonous to the inflectional 3s glide). Pro-drop is also less frequent in the future than in the past tense because the imperative is identical to the future except for the absence of the pronoun (Edith Doron p.c.).

cutoff points in the so-called "ergative split" in many languages (see Stiebels 2001), in principles of subject selection (see Aissen 1999), or in the availability of reflexives. Artstein observes that the hierarchies in (43) all reflect the likelihood of an argument of being a topic. Subjecthood and topichood are traditionally associated with saliency, and it is not surprising that the most salient an element is, the more probable it is that a pronoun refering to it can be dropped completely.

(43) a. Null > Overt
b. 1st/2nd Person > Proper Noun 3rd > Human 3rd > Animate 3rd > Inanimate 3rd
c. Subject > Object
d. Agent > Beneficiary > Experiencer/Goal > Instrument > Patient/Theme > Locative

By aligning (43a) and (43b), we generate two contraint hierarchies. When (44a) >> (44b), a language like English arises, in which the high-ranked ban against null pronouns rules it out that any pronoun be left phonetically unrealised. Likewise, when (44b) >> (44a), Italian comes into being, in which (topical) pronouns cannot be phonetically overt. When the two hierarchies are interleaved, the grammar of languages like Hebrew comes into being, in which *Null/3 >> *Overt/1,2 >> *Null/1,2 guarantees that pro-drop is confined to 1st and 2nd person. There is no way by which the mirror image of Hebrew might arise – quite a correct prediction.

(44) a. *Null/3 >> *Null/1,2
b. *Overt/1,2 >> *Overt3

In a similar vein, the harmonic alignment of (43a) and (43c) implies that there are languages in which pro-drop is confined to subjects, and languages in which it affects subjects and objects at the same time. Pro-drop restricted to objects is not attested, and the pertinent grammar cannot be generated on the basis of (43) and harmonic alignment,

The realization of a pronoun as null or overt is thus not interpreted as faithfulness or betrayal of an input pronoun but the result of markedness hierarchies.

To sum up the discussion of harmonic alignment so far, we have seen that several unrerankable markedness hierarchies, which make use of the same material, can interleave with each other. The fact that hierarchies are implemented differently in different languages is an important source of typological variation, though this variation is not unconstrained. For example, the fact that a language

like Italian allows only pro-drop of a topic subject is a natural consequence of harmonic alignment of different markedness hierarchies. Another advantage of an approach with harmonic alignment is that it partly eliminates the need for faithfulness constraints and thus for inputs.

The goal of this chapter has been to illustrate the universality of the constraints and to show how typological variation arises from reranking. We have shown that constraints should be grounded on the basis of principles which are partly independent of grammar, and that they ideally only deliver possible grammar and repairs of marked structures.

Chapter 4

Uncompromising decision taking

Summary of the Chapter

After having discussed the conflict resolution orientation of OT and the universality of constraints in the preceding chapters, in this chapter we focus on the last uncontroversial property of OT, viz. its lexicographic way of taking decisions, anchored in the EVAL component of the grammar. The total ranking of constraints delivers exactly one winning candidate for each evaluation. There is no place for an optional choice among various candidates, and also no place for gradient judgments. All candidates which are not winners are equally losers, no matter how many constraints they violate. The same lexicographic approach predicts the absence of compensatory effects. The fulfillment of lower ranking constraints can never compensate for the violation of a higher ranking one.

At the end of the chapter, we introduce some mechanisms which seem at first sight to be able to express compensation: local constraint conjunction, multiply violable (or gradient) constraints, and self-conjunction. Whereas local constraint conjunction is another kind of elsewhere effect, the other two kinds of mechanisms are as lexicographic as the usual simple constraints.

4.1 Unique winners as a consequence of the lexicographic decision-taking

Every time we talk or write, we choose between alternatives, and the decisions we take depend on the grammar of the language we speak, that is, in OT terms, on the ranking of the constraints on the one hand, and on the inputs we want to turn into outputs on the other hand. Since speech takes place rapidly, several choices must be made as fast as possible, which means that the grammar must be organized in such a way that it can be processed in an extremely efficient way. It may thus be a desirable property of a grammar that decisions can be taken on the basis of as few criteria as possible. Because of its architecture, OT is such a grammar. Recall from chapter 1 that, given some input I, some GENerator function GEN and some hierarchy H of constraints $C_1,\ldots,C_n$, the grammaticality of a construction $c \in GEN(I)$ is determined as in (1).

(1) c is grammatical (relative to I and H) iff there is no c' ∈ GEN(I) such that c' violates the highest constraint C_j on which c and c' differ less often than c does.

The competition between possible outputs can be envisaged on a pairwise basis, in which the highest constraint C for which two candidates have a different number of violations decides between them. We do not need to consider any constraint lower than C once we have discovered this decisive C. For the total competition, the optimal candidate Cd_{winner} is the one with the best violation profile in the sense that Cd_w violates the constraint hierarchy less than all other competitors in the evaluation set. 'Less' means that there is no candidate Cd_{loser} which does better than Cd_w with respect to the highest constraint on which they differ.

It does not matter whether the highest constraint for which two candidates diverge is among the top ones in the hierarchy or whether the violation profile of two candidates is the same until one considers the hundredth constraint. Going down the constraint hierarchy, as soon as two candidates differ on a single constraint, a decision between them is taken: one candidate survives, the other is eliminated. This can happen at any place in the hierarchy. Of course, the violation profile of two candidates can be different on further constraints, which need not favor Cd_w. Since it is the highest constraint which eliminates a candidate, it does not matter whether and how often the winner beats the loser on single constraints. The winner might even be better than the loser on a single constraint only, while all other constraints favor the loser. If the single constraint favoring the winner is the highest on which they differ, the competition between winner and loser stops at this point. Remember from Chapter 1 that this way of taking decision is called lexicographic, because of the analogy to the way words are ordered in a lexicon.

For obvious reasons, lexicographic conflict resolution is not the only one conceivable. Suppose there is a set of constraints/cues C_i, which are possibly assigned different weights $W(C_i)$. Then a candidate c might be considered grammatical if

- it fulfils a certain number n of constraints (threshold)
- if fulfils more constraints than any other candidate (majority)
- the sum of the weights of the constraints it fulfils exceeds some n
- the sum of the weights of the constraints it fulfils exceeds the sum of weights of any other constraint.

Some of these different ways of resolving conflicts have been proposed in linguistics. For example, in his discussion of word order in German, Uszkoreit

(1984) identifies a number of constraints on word order. Noun phrases with nominative Case should precede noun phrases with accusative or dative Case, noun phrases in focus should follow those which are not focused, etc. According to Uszkoreit, a sentence is grammatical, then, whenever it manages to fulfill *at least of* the relevant constraints. Consider, e.g., the set of constraints in (2) that may be deemed appropriate for German. As we can see from the data in (3) and (4), it indeed seems to be the case that serializations are well-formed as soon as they manage to fulfill at least constraint, no matter which.

(2) a. Nominative precedes dative/accusativ
b. Animate precedes inanimate
c. Non-focus precedes focus

(3) a. dass das Buch einer Frau gefällt (Satisfies 1a and 1c, violates 1b)
b. dass einer Frau das Buch gefällt (Satisfies 1b, violates 1a and 1c)
c. dass Bücher der Frau gefallen (Satisfied 1a, violates 1b and 1c)

(4) a. dass der Mann einen Fehler macht (Satisfies all constraints)
b. dass den Fehler ein Mann machte (Satisfies 1c, violates 1a,b)
c. *dass einen Fehler der Mann macht (Satisfies no constraint at all)

OT claims that conflicts are never resolved in these or similar ways in natural language grammars. Conflict resolution is always hierarchy-based. Indeed, conflict resolution in natural language seems to follow the OT pattern quite in general. When a principle has a certain rank and decides the evaluation in favor of a given candidate because of this rank, this constraint will *always* be the one that decides, irrespective of how many other constraints must be sacrificed in its favor – but of course only as long as it is not defeated by a constraint with an even higher rank.

Consider, e.g., question formation in English again. English constituent order is fairly strict – a set of principles P and their ranking (such as the alignment constraints introduced in Chapter 2) imply that the object follows the verb in a standard English clause. When the object is the only wh-word of an indirect question, these principles P are defeated by WH-IN-SPEC, the constraint requiring that complement questions begin with a wh-phrase. This constraint is responsible for the grammaticality of (5b) as opposed to (5a), with the wh-complement in situ.

(5) a. *I wonder [Bill met whom]
b. I wonder who Bill met

Matrix clauses in English are governed by a constraint called OBLHD by Grimshaw (1997) that forces the presence of a head in each projection and thus implies (given a number of further assumptions) that a finite auxiliary be moved in front of a subject in a matrix wh-question in order to satisfy the need of IP to have a head. This movement violates a principle such as STAY which militates against movement in general. WH-IN-SPEC is again the decisive factor for determining grammaticality among (6a-c), In order to keep it unviolated, violations of STAY and of the principles P mentioned above must be accepted. (6a) violates WH-IN-SPEC, while (6b) violates OBLHD. Only (6c) fulfills both constraints at the cost of a Stay and some P violations

(6) a. *Bill has met who
b. *Who Bill has met
c. Who has Bill met

Lexical verbs must not move in English. When there is no auxiliary in a main question clause, an auxiliary must be inserted (7e), which induces a violation of FULLINTERPRETATION (FI) in the system of Grimshaw (1997) – a DEP constraint formulated as "Do not insert material into the output that was not present in the input" or as "Do not insert meaningless elements." In short, in the winning candidate (7e), WH-IN-SPEC forces the violation of at least two principles, FI and STAY. Since (affirmative) declarative main clauses neither violate the principles P nor FI nor STAY (see (7a)), we deduce that WH-IN-SPEC is responsible for the set of violations induced in (7e). The examples in (7b-e) illustrate what has been shown until now. WH-IN-SPEC and OBLHD are high-ranking and their fulfillment is the origin of the violations of lower-ranking constraints like P, FI and STAY.

(7)	a. Bill met Sue	No violation
	b. *Bill met who	Violation of WH-IN-SPEC
	c. *who Bill met	Violation of STAY, P and OBLHD
	d. *who met Bill	Violation of STAY and OBLHD (in the intended interpretation that he (Bill) met someone)
	e. who did Bill meet	Violation of FI and STAY

Finally, in more complex structures, resumptive pronouns are inserted into the object position of a wh-question in certain varieties of English, in order to avoid an island violation (see Pesetsky 1998). Once more, the whole trouble caused by inserting pronouns (DEP(pronoun)) that double an already existing argument is created by the need to respect WH-CRIT.

(8) who do you wonder whether Mary likes him

Similar examples are harder to come by in phonology. Here is one from the interface between phonology and morphology that illustrates the same point. Recall from chapter 2 the necessity for German morphemes to be aligned at their left edge. This has been accounted for by a constraint requiring alignment of morphemes with syllables at their left edge, which we called ALIGN-L. When a prefix is added, several violations can pile up in order to respect this requirement.

First, a past participle like *gearbeitet* 'worked' is pronounced with a glottal stop between the inflectional prefix *ge-* and the verb stam *-arbeit-*, as in (9a). This is because the stem begins with a vowel. Alternatively, the same state of affair can be accounted for by a violation of ONSET, if the presence of the glottal stop is felt to be a boundary marker. When the prefix ends with a consonant, a second violation arises, which has to do with the fact that the prefix cannot get rid of its coda. This is shown in (9b). Fianlly, in a case where the prefix ends with a consonant and the stem begins with a vowel, both constraints are violated in order to let ALIGN-L be fulfilled.

(9)	a. ge-arbeitet 'worked'	Violation of DEP(ʔ)/ONSET
	b. ent-rüsten 'be indignant'	Violation of NOCODA
	c. ent-arten 'degenerate'	Violation of DEP(ʔ)/ONSET and violation of NOCODA

Lexicographic conflict resolution of the style used in OT implies a number of interesting and far-reaching consequences for natural language grammar:

- There is always at least one winner in each competition.
- Given total ranking, there is at most one winner in each competition. In other words, there is no optionality.
- Optimal candidates are always predicted to be fully grammatical. There is no difference in the relative acceptability of candidates.
- There are no compensatory effects in language. No number of low ranked principles can ever defeat a higher ranked constraint.

In this chapter , we only begin a discussion of these predictions, which lie at the core of what constitutes Optimality Theory – many issues will then be taken up again in the second part of the book.

4.2 There is always a winner

As shown in (1) and in chapter 1, a candidate c is grammatical whenever there is no better alternative c'. Grammaticality is a ***relative*** notion – the least offending structure is chosen as the grammatical output for any input. Such a least offending structure always exists for obvious reasons. Thus, there should always be ***at least one grammatical candidate*** for each input[1]. Lexicographic conflict resolution is a sufficient condition for obtaining this consequence – but obviously not a necessary one. All procedures in which *relative* properties are decisive yield the same result – unlike what holds in models in which grammaticality depends on some absolute criterion – such as the need to reach a certain threshold value of fulfilling a certain number of constraints, or such the need to fulfil a pre-defined set of inviolable constraints. For an example of the last kind, suppose a model of grammar in which all principles are obligatory and unviolable, and which use OT-like constraints. The grammatical candidate has to fulfill all constraints, and as a result, a situation like the following would deliver no winner: for a set of candidates, say a., b., c. and d., and a set of constraints C_1, C_2, C_3 and C_4, there is no constraint which is violated by no candidate. In (10), a. violates C_1, b. violates C_2, etc.

In the ordinary OT model, where constraints are violable, if the constraints are organized in the ranking shown in (10), candidate d. wins, but any other ranking also defines a winner. The violation is not a problem, as long as we find a candidate which does better than all others. More complex patterns also deliver an optimal candidate, like the one illustrated in (11), in which C2 is violated by all candidates remaining in competition after C1 has eliminated candidate a. Thus, in order for a grammar to always deliver an optimal candidate, the violability of the constraints is a prerequisite.

(10) An abstract tableau

	C_1	C_2	C_3	C_4
a	*			
b		*		
c			*	
d				*

[1] Of course, this holds only if the set of output candidates produced by GEN for an input I is non-empty. Furthermore, one must assume that no constraints exist that favor unlimited growth in structure.

(11) Another abstract tableau

	C_1	C_2	C_3	C_4
a	*			
b		*		
c		*	*	
d		*		*

As was observed in Chapter 1, violability of constraints is an important departure from older generative approaches which have assumed rules to be obligatory, and which have developed an intricate pattern of specialized rules overriding each other in order to avoid violability. In syntax, a principle like wh-in-Spec would have to be applicable only in some languages, or would have to be active under different circumstances in different languages. The same is true for Final Devoicing, which makes different predictions according to the language under consideration. But, due to the aerodynamic restrictions discussed in chapter 2, all languages require their obstruents to be voiceless, because voiceless obstruents are more convenient to articulate, and whether they are devoiced on the surface or not is expressed by the way constraints are ranked. Some languages require all obstruents to be voiceless, some are content with Final Devoicing, and still other do not devoice obstruents at all. Thus the fact that OT identifies one winner in each evaluation set is entirely compatible with the violability of constraints.

When one considers, for example, the effects that automatic phonology has on the underlying representation (the input)of a word in phonology, it is indeed ***always*** the case that a phonetic form realizing this input can be identified, at least given that basic requirements on prosodic minimality and segmental well-formedness are given. Constraints to the effect that strings of segments are syllabified, that a syllable has an onset, should not have a coda, that its nucleus be more sonorant than its onset, etc., **never** lead to a situation in which a certain word could not be pronounced. Similarly, in the syntax, there are, for example, various conditions on the well-formedness of a binding relation between a noun phrase and a coreferent element – there are locality requirements for reflexive pronouns, there may be case restrictions for the reflexive itself, and constraints on the case relations between the reflexive and its antecedent, but coreference and binding relations can always be expressed. Alignment constraints also never seem to block the formation of a grammatical structure, even if they are badly violated.

The prediction of the existence of a winning candidate for all competitions (for all inputs) thus seems be borne out in a large enough number of cases for making OT's architecture worth considering, but there are also some at least *prima facie* instances of what Pesetsky (1998) has called "ineffability" – inputs that lead to no

grammatical output. Many examples from syntax involve island constraints – there is simply no grammatical way of asking what (12) unsuccessfully tries to express in English.

(12) *how many kilos do you still refuse a diet although it is obvious that you will weigh if you continue eating as much as you do now?

Not all morpheme combinations yield grammatical outputs – the causativizing suffix *–ize* does not combine with adjectives stressed on their final syllable (*corruptize, from corrúpt). The ineffability of some inputs constitutes a major problem for the architecture of OT. We will take up the issue in chapter 7.

4.3 No optionality

The OT conflict resolution architecture not only predicts that there is at least one winner, it also predicts that *exactly one* candidate is optimal for each input. The result of the competition between outputs is in all cases the candidate which best fulfills the constraints in the language under consideration. The winner's uniqueness is a direct consequence of the constraint ranking and of the lexicographic decision taking. More precisely, OT predicts that there is exactly *one constraint violation profile* that wins a competition. It may (in principle) be the case, however, that several candidates share a violation profile, and may thus come out as winners in the same competition. If no special devices such as tied constraints (see chapter 8) are introduced, such candidates would be able to differ in dimensions only for which no grammatical constraint is relevant. Such dimensions may indeed be very rare, because all factors that ever decide on the grammaticality in some language of the world matter in the EVAL component, and this for each linguistic object.

As in the previous section, it is the relative nature of the concept of grammaticality that implies that there should not be more than one winner, and again, lexicographic decision taking is a sufficient (but not necessary) condition for that outcome. If being grammatical only means to satisfy a certain number of constraints, there are many *different* ways by which this criterion can be met, so that, potentially, many structures with a different violation profile should be grammatical. Similar results obtain when the weights of constraints are added, or when other weight-based computations are carried out. Lexicographic decision taking is particularly prone to yield a single winner only.

This prediction OT makes again does not seem to be too far off the track. Given a certain arrangement of segments or of words with a particular meaning (given a

certain input), there is, in most cases, no more than one possible grammatical realization corresponding to it. Individual grammars tolerate only a very small amount of true synonymy or true free variation. Sometimes, the impression of optionality may arise because the linguistic descriptions employed are too coarse. It is not too long ago that constituent order in languages like German, Russian and Warlpiri was described as being free, because the focus was on properties such as Case, grammatical function, or thematic role. This view has given way to a conception in which fine-grained pragmatic distinctions favor different linear arrangements of the constituents in different circumstances. Word order in German, Russian, and Warlbiri, is no longer considered free nowadays – there is no (or very little) optionality involved. Likewise, when several words compete for one concept, like *eye-doctor* and *oculist*, or *Frenchmen* and *frogs*, they more often than not have different connotations and are consequently used in different contexts.

Furthermore, apparent variation in a language may be better explained as the co-existence of different dialects, sociolects, etc., and many native speakers may master *several* of these different subvarieties at the same time, so that their language may show free variation or optionality because they are bilinguals, in a certain sense. E.g., auxiliary choice for perfect participles of individual verbs in German is subject to semantically triggered variation (with a verb like *hit*, *have* is the default auxiliary choice, while *be* signals an event that was not caused by the grammatical subject) and subject to regional variation (*he has swum* vs. *he is swum*). There is no optionality for those who master a single regional variety only, but those who are more flexible have a choice (not fundamentally different in nature from the choice between speaking German or English!).

Variationist sociolinguistics is the discipline that studies how linguistic change and variation differ in various linguistic communities, and it recognizes that historical change in progress is manifested as synchronic variation. In this approach, optionality is understood as being driven by sociocultural differences between individuals, and as a process towards diachronic changes. It is impossible to ignore this vast field of research if one intends to take a serious look at optionality. Diachronic variations can be (and have been) conceptualized as changes in the constraint ranking implying that optionality is just an intermediate step toward a linguistic change: it is the result of a period of instability between the old form and the new one. If one agrees with this view of optionality, the problem with which grammarians are confronted, and which we address in this section, is how to account for this unstable phase. If one does not agree, the problem remains largely equivalent for the grammar, since what we want to account for is synchronic variability.

In spite of the fact that the restriction of the set of winners to exactly one is a perfect result for most domains in natural language, and seems defensible in the sense just discussed in others, true optionality in the sense that a single grammar yields two outputs for the same domain cannot be denied in others, so that a further problem for OT has been identified.

Optional forms can be in free variation, in which case they are interchangeable in all contexts, or their shared distribution can be limited to some patterns or contexts. Then, the variation can be called partial free distribution. We will not try to make a formal distinction between the two cases, but let us review a few cases. Complement negation, as in *Mary saw nothing*, and auxiliary verb negation, as in *Mary didn't see anything*, seem at first sight to be in complete free variation, at least as far as their meaning is concerned. Another example of free variation is the presence or absence of complementizer in a sentence like *I think* (*that*) *Mary is smart*. Many segmental alternations also seem to be completely optional, leading to free variation between application of the alternation or not, some examples of which are given below. In short, free variation has been observed in a number of domains of linguistics; in phonology, segmental alternations are not the only examples which have been described in these terms, but also many stress systems display free variation of all or some metrical patterns, see below for some examples.

A pattern that can be used to exemplify partial free variation is Finnish genitive plural, as discussed by Anttila (1997). Genitive plural in Finnish has a strong variant, typically *-iden*, and a weak variant, typically *-jen*. Stems ending in a heavy syllable show no optionality, but stems ending on a light syllable can take both variants. The free variation is thus limited to the stems ending on a light syllable. The form *korjaamo* 'repair shop', a word ending with a light syllable, has both *kór.jaa.mo.jen* and *kór.jaa.mòi.den*. Variability is also found in *naapuri* 'neighbor' which allows both endings, *náa.pu.ri.en* and *náa.pu.rèi.den*, but *puu* 'tree' does not allow **púi.den* and *potilas* 'patient' not **pó.ti.lài.den*. These words have only one form because they end on a heavy syllable. Anttila calculates the probability to have one or the other form in case of optionality form a large corpus. As an example *kór.jaa.mòi.den* appears in 20 percent of the cases, and *kór.jaa.mo.jen* in 80 percent. He develops a model in which the variability is statistically measurable from the grammar itself (though the model is highly dependent on the constraints he uses), a point which will be ignored here.

Accounting for optionality and free variation has not been a major concern in the mainstream linguistic literature[2], and, despite the existence of some important

[2] For grammatical problems such as the GB-framework or derivational phonology, optionality is not a technical problem at all, so that the issue did not have to be discussed.

proposals that we will discuss in Chapter 8, OT has not much contributed to increase interest in the issue although standard OT does not predict their existence straightforwardly. Since OT always identifies exactly one winning candidate, optionality between linguistic forms is not easily captured in this type of grammar. Note, however, that what OT really predicts is that there is exactly one winner *per input*, and a number of cases involving optionality in a pre-theoretic sense will thus not constitute a *technical* problem for OT. The *meanings* expressed by two different inputs may be identical, and these two different inputs may yield different outputs. Semantically, the two outputs with the same meaning may fall under the label "free variation", but no technical problem is involved. E.g., the choice between *Mary did not see anything* and *Mary saw nothing* may be of that kind.

Likewise, we have already seen that there is one situation involving a single input in which two candidates can tie: this is when they have exactly the same violation profile. For this situation to arise, two candidates differ somehow, in their phonetic representation, in the presence of a complementizer, in the featureal make-up of a segment or whatever, but there is no constraint choosing between the options. As a result, the two candidates are doing exactly alike with respect to the constraints. Whether we accept such a situation is a function of the constraints participating to the grammar. Grimshaw (1997) analyzes the presence or absence of a complementizer in sentences like *I think Mary is smart* and *I think that Mary is smart* in this way. In her analysis (which she revises in Grimshaw 2001), these two candidates have the same violation profile, which means that they tie on all constraints. Her reflection is based on the observation that the complementizer is a pure function word that is not part of the input and is as such invisible for the faithfulness constraints. It can arise or not under the pressure of markedness constaints, but whether this happens or not is not taken into account by the evaluation procedure.

A review of the literature on optionality does not suggest, however, that optionality can always be understood as arising from identical violation profiles, or from different inputs. Rather, free variation like the one in Ilokano which has been described by Hayes & Abad (1989) and Boersma & Hayes (1999) is representative, though their - not so representative - analysis in terms of the gradient learrning algorithm deserves a detailed discussion that we delay until Chapter 8. Consider the algorithm in (13) showing an optional metathesis between a glide and a glottal stop.

(13)

daʔo ‘kind of tree’ / pag- daʔo - an /	–>	pagdaʔwan, pagdawʔan	‘place where daʔo ’s are planted’
baʔo ‘rat’ / pag- baʔo - an /	–>	pagbaʔwan, pagbawʔan	‘place where rats live’
taʔo ‘person’ / taʔo - en /	–>	taʔwen, tawʔen	‘to repopulate’
ʔaggaʔo ‘to dish up rice’ / pag- gaʔo - an /	–>	paggaʔwan, paggawʔan	‘place where rice is served’
ʔagsaʔo ‘to speak’ / pag- saʔo - en /	–>	pagsaʔwen, pagsawʔen	‘to cause to speak’

The stem-final [o] becomes a glide when it is in a hiatus position, and as a result becomes the onset of a following syllable which would otherwise begin with a vowel. The former syllable onset, the glottal stop, now becomes the coda of the preceding syllable, a dispreferred pattern in Ilokano (laryngeal segments are not allowed in the coda in many languages), and to avoid it, metathesis can happen, which switches the role of the glottal stop and of the glide. Even without going into the details of Boersma & Hayes’s model for optionality, it can be observed that standard constraints are at play, like LINEARITY, a constraint against metathesis, and *ʔ]$_{\sigma}$, a constraint against syllable-final glottal stop. An analysis in terms of identical violation profile is excluded.

Other standard cases of optionality, like the glide-formation in some French words (*Ouest* can be bisyllabic [u.ɛst] or monosyllabic [wɛst]) or the realization of stem final /n/ in Dutch as in *spelen* ‘to play’, *Groningen* ‘name of a city’ and *kinderen* ‘children’, schwa formation in words with a final syllabic sonorant, like German *Himmel* ‘sky’ which can be pronounced as [hɪm̩l] or [hɪmʌl], etc. cannot be analyzed in terms of identical profiles either. Returning to the above examples, optional word order or optional metrical patterns have also been accounted for in terms of alternative ranking of constraints, and not in terms of identical violation profile. For the former case, see Choi (1996) and Büring (2001) for instance, for the latter one, see Bye (1996) and Kager (1994) among others on Estonian and Pater (2000) on English.

Most often, free variation in OT has been accounted for in terms of tied constraints. When two candidates that are in free variation have an identical profile except for just two constraints which favor one or the other member of the free variation, it has been proposed, first by Prince & Smolensky (1993), that the constraints on which they differ have the same rank in the hierarchy, and the rest of the hierarchy is identical. To illustrate this with just one example, consider the Ilokano data in (13) and how tied constraints can account for them in (14). The three highest ranking constraints are not violated in the two optional candidates. They guarantee that no glottal stop is part of the onset, that no segment is deleted and that every syllable has an onset (forcing glide formation). Candidates c to e are eliminated by these constraints. The remaining candidates a. and b. are both winners, since they both violate just one of the tied constraints. Candidate a. violates *ʔ]σ, the constraint against glottal stop in the coda, and candidate b. violates LINEARITY, the constraint against metathesis. Both violate the low-ranking constraint requiring faithfulness of the vocality of input [o].

(14) Glide formation with optional metathesis in Ilokano (from Boersma & Hayes 2000)

/taʔo - en /	*[σ ʔC	MAX-IO(V)	ONSET	LINEARITY	*ʔ]σ	IDENT-IO (syllabic)
☞ a. taʔ.wen					*	*
☞ b. taw.ʔen				*		*
c. ta.ʔo.en			*!			
d. taʔen, taʔon		*!				
e. ta.ʔwen	*!					*

It can be debated whether tying two constraints is an abbreviation for two grammars. Also controversial is the possibility of defining exactly two constraints which should be tied, as well as the exact number of involved constraints. A further problem is the issue of whether differing candidates also diverge on other, otherwise irrelevant and low ranked constraints, which would invariably incorrectly imply a decision between the candidates which are equally grammatical – in which case tying the constraints was of no use: One of the optional candidates is eliminated at a lower point in the hierarchy. We return to these problems and others, in Chapter 8. Summing up at this point, it can be

remarked that OT does not in general allow for optionality, except in the case of identical violation profile. Tied constraints have been the favorite way of capturing optionality, but this move is in blatant contradiction with the requirement of total constraint ranking.

4.4 No gradient judgements

Grammatical architectures have been proposed in which the principles of grammar have different weights, so that in case of a conflict, the principle with a higher weight wins. Since grammars consist of quite a number of constraints with such weights, algorithms need to be identified that allow the computation of some numerical value linked to a structure from the weights of the constraints favoring the structure, and the weights of the constraints disfavoring it. In the case of *what did he say* we therefore need to assume that the (added) weight of the principles favoring the placement of the object behind the verb is smaller than the weight of WH-CRIT. Typically, this is carried out in such a way that a value between 0 and 1 is computed for, e.g., *what did he say.* Since, as we have seen above, the violation profile of winners such as *(I do not care)* what he said, what did he say, what has he said, differ substantially, the numerical values attributed to them by the relevant algorithm will most probably differ, too. There is a problem with this consequence: there is no observable difference in grammaticality or acceptability between these three winning structures!

At least the practice of classical descriptive and traditional generative grammar implies the claim that grammaticality always IS a clear-cut issue. A sentence either belongs to the language or not, a phonetic shape is either totally well formed or ill formed. There is no grey-zone in language, and the lexicographic way of conflict resolution implies just that background assumption of linguistic practice: the best candidate is the winner, there is no theoretically important (or meaningful) way by which a structure might be a better winner than another.

Good linguistic practice is certainly not unfounded. For each language, it seems not too difficult to come up with a list of generalizations that hold categorically in the language. There is simply no doubt that complement constituent questions begin with a wh-phrase in English, German, and French, there is no doubt that the verb must agree with the subject in German, voiceless stops in initial syllables simply have to be aspirated in English and German. There is, however, another kind of data where judgments are not so clear-cut. This is the domain for which linguists eliciting judgements from native speakers get answers like 'I don't know. I have no intuition about these data.' Or 'I would never say that, but if someone does, it is all right.'

Schütze (1996) assumes that the reason why some native speakers lack clear intuitions about parts of grammar is mainly a consequence of the scarcity of the data considered, not only in the course of acquisition, but also afterwards. It is the absence of stimuli or reinforcement which renders speakers unsure about the grammaticality of sentences or words. Hayes (2000) partially shares Schütze's (1996) opinion about the origin of gradient or unsure judgments. He proposes that cases inducing clear judgments are either part of universal grammar or have been reinforced so often that speakers feel perfectly confident about them. As an example, he mentions the present participle formation *-ing* in English, which he illustrates with a nonce formation: Everyone would agree that a new verb *to blick* would produce *blicking*. But in the parts of the grammar for which there is no or not much evidence, gradient judgements are the rule rather than the exception.

Going a step further than Hayes, we suspect that all regular or automatic phonology and morphology elicit categorical judgements, and that the domain of irregular or semi-automatic word formation correlates with a lack of firm intuitions, at least in forms which are not frequently established in the language. In a study of diminutive formation in German, in which *-chen* often but not always correlates with stress and umlaut, Féry (1994) found that speakers are reluctant to accept or refute formations like *Wermuthchen/Wermüthchen, Autochen/Autöchen* and the like, which do not conform to the regular stress pattern of stem-final stress for a diminutive (these words have main stress on the first syllable). Similar experiences have been reported with a large number of affixes in many different languages: *-ize* and *-ate* in English, for example. Speakers are uncertain about the grammatical value of *serbize*, and *obfuscable* (from *obfuscate*), etc. As far as new words are concerned, their introduction is often accompanied with a lack of confidence of how to use them, as we have experienced at the time of the introduction of the monetary *cent* in Germany. Before its daily use, both [sɛnt] and [tsɛnt] were heard, the former one because it corresponds to the pronunciation of this word in the anglophone world (from which German public language continues to borrow tons of words), the latter one because it fits the German phonology, which has no word-initial [s] in its core phonology. Only a few days after the official introduction of the Euro in Germany, the pronunciation [sɛnt]was used as the official, normative pronunciation, but both realizations are still heard in the daily use. Compound verbs like *e-mailen*, *bausparen, in-line-skaten* also trigger gradient judgments. Past participle formation is not totally settled and there exists some variations between *ge-e-mailt*, *e-gemailt* or no participle, for example.

Gradiency in judgements may or may not correlate with the two other problems mentioned above, optionality and ineffability. Often, when the grammatical system seems to offer no way of dealing with a given input, intuitions fail to be

sharp. Reciprocal pronouns must find their antecedent within the same finite clause in English, so that **Mary and Bill wonder whether each other will win* is a good example of ineffability: if a structure is sought which expresses only the reciprocal meaning, English seems to offer no alternative to the ill formed example. In fact, the structure is not *totally* rejected by all speakers of English, and this uncertainty may be a natural reaction to a situation in which something that should be expressible cannot be so – the grammatical architecture implies that there should be a winner, so one cannot be too comfortable with a constellation in which nothing seems to work. Optionality has exactly the same effect. In the case of diminutive formation in German just discussed, the conflict among the various possible forms appears unresolved – again a situation in which there can be no clear judgments on what is grammatical and what not.

However, while the three domains overlap, they are certainly not identical. There are domains with gradient judgements and uncertainty but lacking ineffability. The proper choice among the two different adjectival inflections (strong and weak) in German may be a case in point: in some contexts, in particular with the dative, speakers are very reluctant to give firm judgements on which form would be appropriate (*mit dunklem bayrischen Bier* vs. *mit dunklem bayrischem Bier* "with dark.dat. Bavarian.dat beer"; *mir alten Frau* vs. *mir alter Frau*, me.adt old.dat woman.unm) without there being ineffability. Ineffability can, but need not imply graded judgements. In contrast to German, one cannot form a passive from intransitive verbs in English (**there was danced yesterday, *there was worked a lot*), and native speakers seem to have little doubt about this fact.

Gradiency also cannot be completely reduced to variation. Where there is variation, a set of related data compete, and one variant may be the best in terms of, say, weighted grammaticality. In free variation, it can be the case that some speakers use only one variant and do not accept the other, whereas the remaining speakers find both variants perfect (though this is not a necessary condition for free variation). If gradiency would reduce to variation, we would expect the same kind of distribution, namely that some speakers definitely accept a gradient sentence or a gradient word, some others definitely reject it, with an equal conviction but in the opposite direction, and some speakers accept both variants. Gradiency would then arise as a consequence of the number of persons accepting the structure as compared to the number of people rejecting it. But this is not necessarily what we find. When we asked Dutch speakers to rate *wat heeft wie gekocht* "what has who bought", some people were happy to accept it, others rejected the sentence, but a large number just rated it as questionable. The same pattern was found in our survey of the diminutive formation in German.

Gradiency is a phenomenon found at the level of the individual speaker, not at the level of the speech community (variation).

As an example of gradiency, consider the variation between dark or light *l* in English as discussed by Hayes (2000) (see also the brief discussion in Boersma & Hayes 2001). He used his own pronunciation of light and dark *l* in different environments and asked test persons to give their opinion on the well-formedness of the resulting data. He found out that light *l* is preferred word-initially, as in *lorry* and in the onset of a stressed syllable, as in *balloon*, whereas dark *l* is preferred in the coda as in *bell* and *tilt*, both finally or preconsonantly. In the intervocalic ambisyllabic position, before an unstressed syllable, both light and dark variants are acceptable, though not always in the same measure. The morphology may influence the judgment, as in *touchy- feel-y* and *mail it* with an admissible dark *l* by analogy with the dark *l* of *feel* and *mail* vs. *grayling*, *gaily*, and *freely* , where the realization with the light variant was judged better. In this case, gradiency is explained as a consequence of conflicting factors influencing the pronunciation of a segment in one or the other direction, and the judgments given by native speakers reflect their preference for one of the factors. It is clearly not the case that a strict constraint ranking can account for such patterns.

An alternative that one needs to consider seriously, though it is refuted by both Schütze and Hayes as implausible, is that gradiency of judgment is a consequence of what they call 'performance factors.' Some data are difficult to judge in the absence of an appropriate frame and become better if inserted in a adequate context. Psycholinguistic experiments show that people often fail to detect the wellformedness of a sentence (garden-path sentence), and the reverse situation ("illusion of grammaticality") may also occur (people overlook the factor that renders a structure ungrammatical). Judgment errors are due to processing difficulty, which is continuous by its various nature. Furthermore, from a logical point of view, a distinction must be made between the content of the judgment and the degree of certainty of the judgments, but it is far from being clear that the two aspects are NOT mixed up in graded acceptability judgments. Though the performance explanation will not apply to all cases - certainly not to the light vs. dark l examined by Hayes - it does not strike us as so absurd as to be rejected as a whole. Gradiency can, presumably, not be completely eliminated from grammar, but processing problems may mislead on experimental subjects in some domains.

Standard OT has no method to account for gradient data. This is again a consequence of the absence of compromise and of the lexicographic decision-taking. There is only one winner, and all other candidates are equally bad and eliminated. OT does not make a difference between a loser that violates only one low-ranking constraint more as compared to the winner and candidates which are eliminated because they violate all high-ranking constraints. It does not

distinguish between winners either. If we added a mechanism to measure suboptimality in order to grade candidates, we would be confronted with the problem that many, probably most data are not gradient but categorical, and crucially, that it is not necessary the case that the constraint ordering is the one needed to derive the expected gradiency. For most data, gradiency is not necessary. Again, we return to the issue of gradiency in Chapter 9 where we discuss some of the proposals found in the literature.

4.5 No compensatory effects

"Compensation" is a type of conflict resolution in which the violation of a higher ranking constraint is tolerated in order to satisfy two or more lower-ranking principles.This cannot happen in OT, since, as we saw in section 4.1, decisions are taken on a lexicographic basis.

In the domain of non-linguistic decision taking, however, compensatory effects seem to be an everyday situation. As an example, imagine that buying a red car is top-ranking for some buyer, but a low price and four doors on a green car is even better, even if individually, neither the price nor the number of doors is more important than the color. This kind of compensatory effect is called Multi-Attribute Utility in the relevant psychological literature (Jungermann, Pfister & Fischer 1998).[3]

In order to evaluate how such effects could be handled in OT, if necessary, two situations must be distinguished: first, the high-ranking constraint is violated by all candidates (no red car is available) and second, some candidates fulfill the high ranking constraint (a red car is available). The first situation is manageable in OT. Since the high ranking constraint cannot take the decision, decision taking power is passed on to lower constraints: in the car example introduced above, the price and the number of doors. The winning candidate fulfills both lower conditions, but the color must be one other than red, say green, the next favorite color of the buyer. This kind of compensation is not a problem for OT, in fact it is the essence of OT. The other situation, however, in which fulfilling both lower constraints is better than fulfilling the higher one even though the higher constraint can be fulfilled, cannot be captured in standard OT. In this case, there are red and green cars. A red car with a low price and four doors would be the best option, but there is no such car. Only green ones are affordable and have more than three doors and it is this latter option that is chosen over red cars with low price and red cars with

[3] We have also not found compensatory effects in the domain of other social convention, like laws or rules of games or sports. Here, too, properties that come near to compensation are better analyzed as elsewhere effects.

four doors. In OT, such a situation would mean that fulfilling two lower constraints is better than fulfilling a higher-ranking one.

An obvious strategy to introduce the compensatory aspect of Multi-Attribute Utility into OT is *constraint conjunction*, in which the conjoined effect of two lower constraints is itself a constraint which can be ranked even higher than a constraint dominating each one individually. The constraint A & B is violated iff the same element violates A and B at the same time. If A dominates both B and C, but B and C taken together dominate A, compensatory effects are captured: A must not be sacrificed in the interest of B or C alone, but B and C together manage to defeat A. This is illustrate in (15).

(15) Constraint Conjunction

a. A >> B, C
b. B & C >> A >> B, C

Smolensky (1995) was the first to propose constraint conjunction, and it has been widely used in the literature since then. Examine first the predictions made by constraint conjunction in our car example. If there is no cheap red car with four doors, the next best option is the green car with these attributes, as shown in tableau (16). Of course, as soon as there is such a red car, it is chosen as optimal since it does not violate the second constraint A. This is shown in tableau (17).

(16) Green car

	B (price) & C (doors)	A (red)	B (price)	C (doors)
cheap green car with 4 doors		*		
red car with 3 doors	*!		*	*
expensive red car	*!		*	*
expensive green car	*!		*	*
green car with 3 doors	*!		*	*

(17) Red car

	B (price) & C (doors)	A (red)	B (price)	C (doors)
cheap red car with 4 doors				
cheap green car with 4 doors		*!		

red car with 3 doors	*!			*
expensive red car	*!		*	
expensive green car	*!		*	
green car with 3 doors	*!			*

It is to be observed that OT has no problem expressing constraint conjunction. The conjoined constraint B & C is just an ordinary constraint with a certain rank. It does not have to be high-ranking but can be dominated by A, or even by all single constraints, in which case it would never have an effect. The OT scale, being composed of discrete elements organized in a total ranking, ranks a complex constraint like B & C in the same way as it does a simple one, and the decision as to which candidate is optimal takes place in a lexicographic fashion.

The question arises whether constraint conjunction is a grammatical reality. The introduction of conjoined constraints (plus, possibly, the addition of ties) allows to capture all kinds of compensatory effects in OT. Is this an advantage? Not necessarily, since the move increases the descriptive power of OT substantially – as long as we have no restricted theory of which constraints may be conjoined at all. This drawback could be accepted if it could be shown that the step consisting in conjoining constraints is really necessary, in other words, if it could be shown that true compensation exists in grammar. Instead, there is ample evidence for the empirical claim of OT concerning the absence of compensation in natural language grammars. To illustrate this claim, let us return for a moment to the German Case marking rules. It was shown in Chapter 2 that the following constraints, which appear in the order of their likely ranking, are active in this language.

(18) a. Lexically determined exceptional case must be respected
b. Do not use one case more than once within a clause
c. Avoid use of dative case
d. Avoid use of accusative case

Logically, it might be that (18a) wins over (18b) and (18c-d) in pairwise competitions, but that (18b) and (18c-d) are jointly able to defeat (18a). In such a situation, the use of lexically governed Cases would e.g. be avoided, for example, in an active ditransitive structure (in which the other object necessarily bears either accusative or dative Case), but lexical Case could show up in the passive (when the other object switches to nominative). While such a constellation is conceivable, it is certainly NOT found in German (nor in any other language we are aware of).

In the realm of the sciences of language in general, the effects resembling compensation most are the so-called phonetic trading relations in perception (Repp 1982). The idea is that a change in one phonetic cue can be compensated by a change in another phonetic cue so as to maintain the original phonetic percept. As an example, the voiced-voiceless distinction is attained by varying multiple cues. For stop consonants in initial position, for instance, both Voice Onset Time (VOT) and the first formant (F1) transition are relevant. If the voice onset is delayed (and thus the VOT is increased), the onset frequency of F1 must be lowered, in order for the phonetic percept to be equivalent. Another cue that can be traded for VOT is the amplitude of the aspiration. If the amplitude of the noise created by aspiration before voice onset is higher, the duration of VOT must be decreased to get the same perceptual impression. Other trading effects have been observed in numerous other parts of articulation, such as place of articulation and manner cues. Even if these effects are compensatory, the trading of one cue for another is not truly compensatory in the sense described above. What happens in the trading relations is that a loss of a kind of cue can be compensate by another cue. Moreover, these effects have usually been explained as being the consequence of the way our auditory perception works, and are to be located in the domain of psychoacoustics rather than in linguistics.

In sum, we can be confident that what has been called compensatory effects are of a different kind from the kind of compensation predicted by conjoining constraints. When we discuss constraint conjunctions in detail in Chapter 8 we will argue that they are just other cases of elsewhere effects, of specified environments or rules having priority over more general cases, or, following Padgett (2002), that they can be replaced by hierarchies. Conjoining constraints is another mean of expressing a prohibition of a very marked structure. In other words, we will propose that conjoined constraints are to be interpreted as single constraints, and are better formulated as such.

In the remainder of the chapter, we discuss cases which look like compensation, but are not compensation: multiply violable constraints and self-conjunction of constraints.

4.6. Multiply violable constraints

Up to now, our discussion of lexicographic conflict resolution in OT has focused on situations in which a certain constraint C is violated or not. In this simple case, a candidate c satisfying C will win the competition, provided that C is the relevant constraint in the evaluation. The definition in (1) allows conflict resolution in further situations, however. The crucial constraint C may be violated by candidate cd1 three times, and by candidate cd2 five times. In OT, cd1 will win the

competition – not because it respects constraint C, but because it avoids unnecessary violations of it. Violations of a constraint that cannot possibly be improved do not matter for grammaticality.

The standard perspective would seem to be that a constraint may be violated n times by a candidate c because there are n elements in c that violate it. But under a different way of computing constraint violations incurred by a structure, there need not be a 1:1 correspondence between number of constraint violations and number of elements violating the constraint. The existence of multiply violable constraints, also called "gradient" (not to be mixed up with constraints that lead to gradiency in the sense of vague judgments) which can be violated several times by a single candidate also introduces a new perspective in our exploration of compensatory effects in language and OT.

Alignment constraints are prototypical examples of multiply violable constraints, if one interprets them (following standard practice) as measuring the distance between a certain element, like a syllable, a foot, a complementizer, a head, etc., and a specific edge. When an alignment constraint is interpreted in a gradient fashion, the number of elements (of a certain type) separating X from the target edge determines the number of violations incurred. If B should be at the left edge of A according to some alignment constraint C, the position of B in the pattern [$_A$ X Y Z B …] incurs *three* violations of C by B under a gradient interpretation (but only one on a non-gradient interpretation).

Under this gradient interpretation of constraints, it is often the case that an element that is not perfectly aligned with the edge wins the competition. This happens when overriding constraints eliminate all the candidates which would fulfil the alignment requirement better, or when several elements compete for the same position. In this latter case, they cannot all be aligned with a specific edge. It is the number of violations of a single constraint by a single element that is responsible for multiply violable constraint violation, making such constraints different from binary ones which can be violated only once by each element.

We can work out the difference between multiply violable and binary constraints more clearly by considering a word with several syllables closed by a coda. Every syllable violates the constraint NoCoda and, as a result, the tableau evaluating the whole word contains as many violations of NoCoda as there are syllables with a coda. However, each syllable violates NoCoda just once, and this is the reason why NoCoda is binary, not gradient. Turning to real-life multiply violable constraints, consider the lexical stress pattern often found in Germanic languages, with trochaic feet (σ́ σ), main stress on the penultimate syllable and otherwise left-aligned alternating secondary stresses. This pattern is illustrated in (19a) with a word of six and one of seven syllables and four relevant constraints

in the line of McCarthy & Prince (1993b) and Féry (1998). The relevant constraints are listed in (19b).

(19) Trochaic foot pattern

a. (σ́ σ)(σ́ σ)(σ́ σ)
(σ́ σ)(σ́ σ) σ(σ́ σ)

b. ALIGN (PW, R, Ft, R): A Prosodic Word ends with a foot.
ALIGN (Ft, L, PW, L): All feet are Pw-initial.
FOOT-BIN: Feet are binary (under a syllabic analysis).
PARSE-SYLL: Syllables are parsed into feet.

The first two constraints of interest in (19b) must be ranked as shown in (20), with ALIGN (PW, R, Ft, R), abbreviated as ALIGN-R, higher-ranking than ALIGN (Ft, L, PW, L), abbreviated as ALIGN- L, so that a foot will be present word-finally, as shown in candidate a. in Tableau 8. In the opposite ranking, in which candidate b. would be the winner, all feet would be as far left as possible, and the consequence would be that the last syllable would be unparsed and not the antepenult as in (19a). FOOT-BIN and PARSE-SYLL are also crucial since they force feet to be binary and syllables to be parsed. PARSE-SYLL is not undominated since the optimal candidate in our example, consisting of a word with an odd number of syllables, has an unparsed syllable in order to satisfy FOOT-BIN, but still, a candidate with non-binary feet, like (20d), and a candidate with just one or two parsed syllables, like (20e), are eliminated by these constraints.

(20) Multiply violable constraints

σ σ σ σ σ σ σ	FOOT-BIN	PARSE-SYLL	ALIGN-R	ALIGN-L
☞ a. (σ́ σ)(σ́ σ) σ(σ́ σ)		*		2 + 5
b. (σ́ σ)(σ́ σ)(σ́ σ) σ		*	*!	2 + 4
c. (σ́ σ) σ́ σ σ(σ́ σ)		**!*		5
d. (σ́ σ)(σ́ σ) (σ) (σ́ σ)	*!			2 + 4 + 5
e. (σ́ σ) σ σ σ σ σ		**!***	*****	

Beside the obvious difference in directionality, ALIGN-L and ALIGN-R make further distinct predictions. Align-L quantifies universally over feet and posits that every foot must be left-aligned with a prosodic word. It thus counts each foot individually and counts the number of syllables separating it from the left edge. In

candidate a., the first foot is perfectly aligned and does not violate the constraint. The second foot is separated from the left edge of the prosodic word by the first foot, thus two syllables, and the last foot, which is right-aligned, is five syllables away form the left edge of the prosodic word. As a result, it has five violations. If segments are counted and not syllables as in our example, violations become more numerous but the result of the computation is identical: it is still candidate a. which wins the competition. The same calculations can be applied to the other candidates. Align-R quantifies universally over prosodic words and existentially over feet and requires that prosodic words always end with a foot. Only the rightmost foot counts, since it is the one which is mostly apt to fulfill this constraint. Candidate a. fulfills the constraint perfectly, candidate b. violates it once because its rightmost foot is one syllable away from the right edge, and finally, candidate e. violates the constraint 5 times, since the rightmost syllable is five syllables away from the right edge.

However, to return briefly to the theme of the preceding section, even if the optimal candidate has more violations of ALIGN-L, as compared to candidates b., c. and e. in (8), it is not the case that any candidate with fewer violations of this constraint can compensate for its ungrammaticality. Candidate b. fails because it has an additional violation of ALIGN-R, candidate c. and e. because they violate PARSE-SYLL more often than candidate a., and candidate d. is excluded because it contains a too light foot.

The upshot is that multiply violable constraints are treated in a purely lexicographic way as well. The fact that a constraint is multiply violable does not change anything the nature of decision taking. The number of violations of a multiply violable constraints can be decisive, but the decision itself is as categorical as in a non-gradient constraint. For the conflict resolution algorithm, it is unimportant by which mechanism a constraint is violated n times by a candidate.

McCarthy (2002) argues that graded constraints can be eliminated from OT, and be replaced by categorial constraints, which he called quantized constraints. His proposal is to replicate the effect of gradiency in a tableau like (20) by replacing ALIGN-R and ALIGN-L by pairs of constraints, one accounting for the presence of syllables between the relevant foot and the edge under consideration, and one for the presence of feet in the same position. If there are any syllable (one or more), the first constraint is violated. If there are any feet (also one or more), the second constraint is violated. These two new constraints are categorical. It does not matter how many feet or how many syllables separate the foot form the edge, but as long as there are any the relevant constraint is violated. (21) shows how ALIGN-R is replaced by these two quantized constraints.

(21) Quantized ALIGN (Ft, Wd, R)

a. ALIGN-BY-σ(Ft, Wd, R)
No syllable stands between the right-edge of Ft and the right-edge of Wd.
b. ALIGN-BY-FT(Ft, Wd, R)
No foot stands between the right-edge of Ft and the right-edge of Wd.

Likewise, gradient interpretations of alignment constraints can be avoided by reformulating them slightly. If an alignment constraint C "X must be at the left edge of Z" is reformulated as a constraint C* "a W must not intervene between an X and the left edge of Z", then any number of violations of C incurred by X under a gradient interpretation is matched by an equal number of violations of C* incurred categorically by the elements that separate X from the relevant border. Thus, there is no *technical* need for gradient constraints, a fact that confirms the lexicographic nature of conflict resolution in natural language.

In both interpretations of the align effects, 'gradiency' does not lead to compensatory effects. There is however a situation in which decomposing the multiply violable constraints in several non-gradient ones can imitate compensatory-like effects, and we address this point in the next section.

4.7 Self-Conjunction of Constraints as Multiple Violation

Self-conjunction of constraints is a special case of constraint conjunction, which is often used to express OCP violations. A single constraint is conjoined with itself, and this a certain, precise number of time. There are constellations in which one element can violate a single constraint several times but nevertheless a multiply violable interpretation of the relevant constraint is not the ideal solution, as for instance when another, independent constraint, must be inserted between n and n-1 violations in order to correctly describe the grammatical facts of the language. There is thus a universal hierarchy along the line of (22), which says that it is worse to violate C n times than to violate it n-1 times.

(22) Self-conjunction of constraints

$C^n >> C^{n-1} >> .... C^2 >> C$

A different way of representing the same state of affairs is to postulate three constraints C^1, C^2 and C^3. C^1 is violated by a candidate with just one violation (or more), C^2 by a candidate with two violations (or more) and C^3 by a candidate violating the constraint three times (or more).

As an example, consider Rendaku in Japanese, the process of voicing the first obstruent in the second part of a compound, illustrated in (23a). The application of Rendaku is restricted by the so-called Lyman's Law which limits the number of voiced obstruents to one per morpheme (an OCP restriction). The effect of Lyman's Law is shown in (24). When the second word of the compound already contains an underlying voiced obstruent, in other words, when Rendaku would trigger a second voiced obstruent, its effect is blocked.

(23) Rendaku (examples from Ito & Mester 2002)

a. **t**ama	'ball'	teppoo+**d**ama	'bullet'
b. **s**ono	'garden'	hana+**z**ono	'flower garden'

(24) *"Lyman's Law"* (examples from Ito & Mester 2002)

a. taba	'bundle'	satsu+taba 'wad of bills'	*satsu-daba
b. sode	'sleeves'	furi+sode 'long-sleeved kimono'	*furi-zode

The prohibition of more than one voiced obstruent per morpheme can be expressed by a single locally bounded constraint against voiced obstruents, a constraint which can be violated several times, according to the number of voiced obstruents appearing in one morpheme. However, adopting this solution poses ranking problems. If the constraint necessary to cause Rendaku, (which we call here RENDAKU) is ranked higher than the gradient constraint against voiced obstruents, RENDAKU would always apply, regardless of the number of voiced obstruents in the surface form, whereas if it is lower ranking, no voiced obstruent would ever emerge as a consequence of RENDAKU. This is shown schematically in (25) and (26). In other words, the effect of RENDAKU would always be overridden by the prohibition against voiced obstruents, at least against those not protected by faithfulness constraints. In (25) *teppodama* is rightly predicted to undergo RENDAKU, whereas *satsutaba* is wrongly predicted to underly RENDAKU as well. In (26) RENDAKU applies in neither form, which is right for *kamikaze*, but wrong for *teppodama*.

(25) Rendaku (failed)

	RENDAKU	*VOICEDOBSTR
satsu-taba	*!	*
satsu-daba		**!
teppo-tama	*!	
teppo-dama		*

(26) Rendaku (failed again)

	*VOICEDOBSTR	RENDAKU
satsu-taba	*!	*
satsu-daba	*!*	
teppo-tama		*
teppo-dama	*!	

An alternative solution is to decompose *VOICEDOBSTR into two similar constraints, one against a single occurence of a single voiced obstruent and another one against two voiced obstruents, and to guarantee that the constraint applies inside of a morpheme. This is the solution of Ito & Mester (1998) who propose to express Lyman's Law with the help of a constraint *VOICEDOBSTR^2 ranked higher than RENDAKU. Both RENDAKU and *VOICEDOBSTR^2 (Lyman's Law) are higher ranking than the markedness constraint against one voiced obstruent. *VOICEDOBSTR^2 is a locally bounded constraint prohibiting two voiced obstruents per morpheme. The tableaux in (27) and (28) show the effect of Rendaku and its blocking by Lyman's Law in these terms.

(27) RENDAKU applies

/ore-kami/	*VOICEDOBSTR^2	RENDAKU	*VOICEDOBSTR
ore-gami			*
ore-kami		*!	

(28) RENDAKU is blocked by Lyman's Law

/kami-kaze/	*VOICEDOBSTR^2	RENDAKU	*VOICEDOBSTR
kami-kaze		*	*
kami-gaze	*!		**

A constraint like *VOICEDOBSTR^2 is called a *self-conjoined constraint*, because it conjoins a constraint prohibiting a marked pattern with itself. But it can also be understood as part of a gradient constraint (or a markedness hierarchy) split up in several constraints. Its function is similar to that of a multiply violable constraint, because it allows multiple violations of a constraint by a single element. It is also different from gradient constraints because it specifies exactly the number of violations which are prohibited. It states that violating a markedness constraint at least n times in a certain domain is bad. Since a self-conjoined constraint occupies a rank of its own in the hierarchy, there is a further and more important difference from multiply violable constraints. If A > B, EVAL will select for c rather than c' if c violates A less often than c', independent of the number of violations of B (see above). A self-conjoined constraint B&B will be decisive when B&B > A > B, so that the number of violations of B now becomes important for the outcome of the competition. After all, constraint conjunction always allows for expressing compensatory effects.

A further example of self-conjoined constraints comes from syntax. Chomsky (1986) proposes that the distance between a phrase and its trace is measured by the number of barriers that have been crossed. Structural complements are not barriers in the normal case, but they may inherit barrierhood. On the other hand, specifiers and adjuncts are barriers *per se*, and CP may be a barrier if it dominates certain types of IP in certain constellations. Details are unimportant for the argument we wish to present, so we confine ourselves to giving some examples in (29).

(29) Number of crossed barriers

a. No barrier crossed: *what do you fix t*
b. One barrier crossed: *what do you wonder [how to fix t*
c. Two barriers crossed: ??*what do you wonder [how [one should fix t*

If Chomsky 1986 is correct in claiming that languages and/or construction types differ as to how many barriers may be crossed by the moved wh-phrase in question formation, we need to be able to rank the constraint responsible for Wh-fronting, Wh-in-Spec (see above), between the constraints expressing the number

of barriers that have must not be crossed. A pattern like the one given in (30) arises.

(30) CROSSEDBARRIER^{k} >> WH-IN SPEC>> CROSSEDBARRIER^{k-1}

(30) expresses the situation in which a movement that crosses k-1 barriers is grammatical, while a movement that crosses k barriers leads to ungrammaticality (see Legendre et al 1998). In English, one barrier may be crossed, but not two, so that the ranking CROSSEDBARRIER^{2} >> WH-IN-SPEC >> CROSSEDBARRIER is obtained.

Self-conjunction of constraints can be compared to OCP (Obligatory Contour Principle) in phonology which prohibits proximity of identical elements. Like self-conjunction, OCP is also locally bounded, and like self-conjunction it militates against an increase of markedness.

Conjoined constraints may, however, imply a problem related to the free-ranking axiom of OT. Since it should be possible to re-rank all constraints relatively to each other, it is not necessary that the hierarchy (14) proposed by Smolensky (1995) necessarily holds. The individual constraints C^{k} composing it can be freely ranked relatively to each other. Does the ranking $C^{i} > C^{i+k}$ imply that sometimes it can be better for an element to violate a constraint more often? Not necessarily! Note first that C^{i} means that C is violated at least i times. When a structure violates C^{i+k} it also violates C^{i}. Therefore, when two structures c and c* differ by C^{i} and C^{i+k} only, it is always C^{i+k} that settles the conflict, irrespective of its rank.

There is another aspect in which self-conjoined constraints differ from truly multiply violable constraints, which has to do with the elements triggering the violations. In the Align constraints, it is really one syllable, or one foot, or one Specifier which is separated from an edge by several elements of the same type or of a slightly different type. In *VOICEDOBSTR^{2}, even if the violations happen in a well-defined domain, a word or a sentence, they are induced by different elements in each case. The two violations of *VOICEDOBSTR^{2} come from two different voiced obstruents. In the case of CROSSEDBARRIER^{k} it must be the same element in the same domain, but the barriers crossed are different in each case. McCarthy (2002) makes a distinction between 'horizontal' gradiency, alignment effects, and 'vertical'gradiency, all other gradient cases, and observes that only horizontal gradiency goes together with unboundedness. The other cases of gradiency, to which the examples of this section belong, are bounded by the number of elements able to violate the multiply violable constraint. There is another true difference between alignment effects and all other gradient effects, which rests on the fact that only alignment constraints can be violated several times by the same

element. It is thus not an accident that gradiency has been mainly used in relationship with alignment, and self-conjoined constraints are found mainly in domains implying pure markedness, since the difference is related to the nature of the different patterns. However, the relationship is not compulsory. Recall the Tagalog infixation examples which showed that alignment can be interrupted by markedness, as well: an affix is as far to the left as possible, but only as long as it satisfies the need of NoCoda as well as it can.

In the same vein, notice also that in both examples of self-conjunction it is important to limit the violation of the relevant constraint to a specific domain. If violations of *VoicedObstr occurs twice, but by different morphemes, this does not affect *VoicedObstr^2. The same is true of CrossedBarrier. If this constraint is violated by different phrases, this does not involve self-conjunction, but rather multiple violations of the single constraint CrossedBarrier. Of course, one of the wh-phrases could also violate CrossBarrier^2 on top of violating the single constraint. We illustrate this with the sentence *what does he wonder how to fix*, which contains two different occurrences of a fronted wh-word. The total number of crossed barriers is three. *What* has crossed two barriers and violates CrossedBarrier^2 once and CrossedBarrier twice, as shown in Tableau 10. *How* has crossed just one barrier. Altogether there are 3 violations of CrossedBarrier but only one of CrossedBarrier^2.

(31)

	Cross Barrier^2	Wh-Fronting	Crossed Barrier
*how do you wonder what to fix t	*!		***
Do you wonder what to fix how		*!	

Summing up this chapter, fulfillment of constraints happens in a yes or no fashion. No compensatory effect in which two lower ranking constraints override a higher one is possible. Some cases, like multiply violable constraints and self-conjoined constraints which seem to function like compensation turn out to be normal lexicographic effects after all.

It has also been shown that the purely lexicographic decision-taking proper to OT excludes optionality and gradient judgments. There is always just one optimal candidate, and all others are declared to be equally bad. Whether this is really the last answer of OT will be addressed again in this book.

References

Aissen, J. (1999): Markedness and Subjekt Choice in Optimality Theory. *Natural Language Theory* 17. 773–771.

Anderson, S. R. (1983): *The Architecture of Cognition.* Cambridge, MA: Harvard University Press.

Anttila, A. (1997): Deriving variation from grammar: A study of Finnish genitives. In: Hinskens, F., Hout, R. van & Wetzels, L., eds. (1997): *Variation, change and phonological theory.* Amsterdam: John Benjamins.

Artstein, R. (2002): *Phonological Decomposition: Semantic Interpretation for Parts of Words*. PhD dissertation, Rutgers University: New Brunswick.

Bach, E. (1964): Syntactic Theory. In: Rinehart, H. & Winston, P. H., eds. (1964): *An Introduction to Transformational Grammars*. New York, NY. [Revised Version (1974). Lanham, MD: University Press of America.]

Bech, G. (1955/1957): *Studien über das deutsche verbum infinitum.* Kopenhagen: Munksgaard.

Berlin, B. & Kay, P. (1969): *Basic Color Terms: Their Universality and Evolution.* Berkeley: University of California.

Blutner, R. (1999): *Some aspects of optimality in natural language interpretation.* MS, Berlin.

Boersma, P. (1999): *On the need for a separate perception grammar*. Ms, University of Amsterdam.

Boersma, P. & Hayes, B. (2000): Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32. 45–86.

Boersma, P. & Hayes, B. (2001): Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32:1.

Bouchard, D. (1983): *On the content of empty categories.* Dordrecht: Foris.

Bresnan, J. (1999): The Emergence of the Unmarked Pronoun. In: Legendre, G., Vikner, S. & Grimshaw, J., eds. (1999): *Optimality-theoretic Syntax.* Stanford University: MIT Press.

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

Burzio, L. (1994): *Principles of English stress.* Cambridge University Press.

Burzio, L. (1998): Anaphora and Soft Constraints. In: Barbosa, P., Fox, D., Hagstrom, P., McGinnis, M. & Pesetsky, D., eds. (1998): *Is the best good enough? Optimality and competition in syntax.* MIT Press & MIT Working Papers in Linguistics, Cambridge, MA. 193–219.

Büring, D. (2001): *Let's Phrase It! - Focus, Word Order, and Prosodic Phrasing in German Double Object Constructions.* In: Müller, G. & Sternefeld, W. eds. (2001): Competition in Syntax. *Studies in Genitive Grammar* 49. Berlin & New York: de Gruyter. 101–137.

Bye, P. (1996*): Correspondence in the prosodic hierarchy and the grid.* PhD thesis, University of Tromsø.

Cheng, L. (1991): *On the typology of wh-questions.* PhD dissertation, MIT.

Choi, H.-W. (1996): *Optimizing Structure in Context: Scrambling and Information Structure.* PhD dissertation. Stanford University Department of Linguistics.

Chomsky, N. (1981): *Lectures on Government and Binding.* Dordrecht: Foris.

Chomsky, N. (1982). *Some concepts and consequences of the theory of government and binding.* Ms, Cambridge. MIT Press.

Chomsky, N. (1986): Barriers. *Linguistic Inquiry Monograph* 13. Cambridge: MIT Press.

Chomsky, N. (1993): A Minimalist Program for Linguistic Theory. In: Hale, K. & Keyser, J., eds. (1993): *The View From Building* 20. Cambridge: MIT Press.

Chomsky, N. (1995): *A Minimalist Program for Linguistic Theory.* Ms, Cambridge. MIT Press.

Chomsky, N. (1998): *Minimalist inquiries: the framework.* Ms, Cambridge. MIT Occasional Papers in Linguistics.

Clements, G. N. (1985): The geometry of phonological features. *Phonology Yearbook* 2. 225–252.

Clements, G. N. (1990): The Role of the Sonority Cycle in Core Syllabification. In: Kingston, J. & Beckman, M., eds. (1990): *Papers in laboratory Phonology I: Between the grammar and Physics of Speech.* Cambridge: Cambridge University Press. 283–333.

Clements, G. N. & Keyser, S. J. (1983): *CV phonology. A generative theory of the syllable.* MIT Press.

Dell, F. & Elmedlaoui, M. (1985): Syllabic consonant and syllabification in Imdlawn Tashlhiyit Berber. *Journal of African Languages and Linguistics* 7. 105–130

Everaert, M. (1986): *The Syntax of Reflexivization*. Dordrecht: Foris.

Fanselow, G. (1991): *Minimale Syntax.* Habil., University of Passau. Erschienen als Sonderheft der Groninger Arbeiten zur germanistischen Linguistik.

Fanselow, G. (2000): Optimal Exceptions. In: Stiebels B. & Wunderlich D., eds. (2000): *The Lexicon in Focus*. Berlin: Akademie Verlag. 173–209.

Fanselow, G. (2001): Features, q-roles, and free constituent order. *Linguistic Inquiry* 32:3.
Féry, C. (1994): *Umlaut and inflection in German.* Otimality Archive. Rutgers University.
Féry, C. (1998): German Word Stress in Optimality Theory. *Journal of Comparative Germanic Linguistics*. (ROA-301-0399; http://roa.rutgers.edu).
Féry, C. (2000): Phonologie des Deutschen: eine optimalitätstheoretische Einführung. *Linguistics in Potsdam* 7.
Flemming, E. (1995): *Auditory Representations in Phonology.* PhD dissertation, UCLA.
Frazier, L. & Flores d´Arcais, G. (1989) : Filler-driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language* 28. 331–344.
Goddard, C. (1979): Particles and illocutionary semantics. *Papers in Linguistics* 12:1–2. 185–229.
Goldsmith, J. (1976): *Autosegmental phonology*. PhD dissertation, MIT, Cambridge. New York: Garland Press.
Greenberg, J. H. (1966): *Language Universals with Special Reference to Feature Hierarchies.* The Hague: Mouton.
Grewendorf, G. (1989): Ergativity in German. In: Muckel, S. & Pechmann, T., eds. (1989): *Predictive antecedent reactivation: Evidence from the processing of German verb-final sentences.* Dordrecht: Foris.
Grimshaw, J. (1997): Projection, Heads and Optimality. *Linguistic Inquiry* 28:3. 373–422.
Grimshaw, J. (2001): Economy of Structure in OT. (ROA 434-0601; http://roa.rutgers.edu).
Grimshaw, J. & Vieri, S.-L. (1998): Optimal subjects and subject universals. In: Barbosa, P., Fox, D., Hagstrom, P., McGinnis, M. & Pesetsky, D., eds. (1998): *Is the best good enough? Optimality and competition in syntax.* MIT Press & MIT Working Papers in Linguistics, Cambridge, MA. 295-326.
Gussenhoven, C. & Haike, J. (1998): *Understanding Phonology*. London: Arnold.
Haider, H. (1996): Downright down to the right. Lutz, U. & Pafel, J., eds. (1996): *On Extraction and Extraposition*. Amsterdam: Benjamins. 145–271.
Hayes, B. (1995): *Metrical stress theory: Principles and case studies.* University of Chicago Press.
Hayes, B. (2000): Phonological Acquisition in OT: The early stages. In: Kager, R. & Zonneveld. W., eds. (2000): *Phonological Acquisition and Typology.*

Hayes, B. & Abad, M. (1989): Reduplikation and Syllabification in Ilokano. *Lingua* 77. 331–374.
Haspelmath, M. (1999c): Optimality and diachronic adaptation. *Zeitschrift für Sprachwissenschaft* 18:2. 180–205.
Ito, J. & Mester, A. (1998): *Markedness and Word Structure: OCP Effects in Japanese.* Ms, University of California at Santa Cruz.
Ito, J. & Mester, A. (1998): The Phonological Lexicon. In: Tsujimura, N. ed. (1998): *A Handbook of Japanese Linguistics.* Oxford: Blackwell.
Ito, J. & Mester, A. (1994): Reflections on CodaCond and Alignment. In: Merchant, J., Padgett, J. & Walker, R., eds. (1994): *Phonology at Santa Cruz* 3. University of California, Santa Cruz. 27–46.
Jacobson, P. (1987): Phrase Structure, Grammatical Relations, and Discontinous Constituents. In: Huck, G. & Ojeda, A., eds. (1987): *Syntax and Semantics* 20. San Diego: Academic Press. 27–69.
Jungermann, H., Pfister, H.-R. & Fischer, K. (1998): *Die Psychologie der Entscheidung.* Heidelberg: Spektrum.
Kager, R. (1994b): *Ternary rhythm in alignment theory.* Ms: Utrecht University.
Kager, R. (1999): *Optimality Theory. A Textbook.* Cambridge: Cambridge University Press.
Kayne, R. (1994): *The Antisymmetry of syntax.* Cambridge, MA: MIT Press.
Kenstowicz, M. & Kisseberth, C. (1977): *Topics in Phonological Theory.* New York: Academic Press.
Kiparsky, P. (1973): Phonological Representations. In: Fujimura, O., ed. (1973): *Three Dimensions of Linguistic Theory.* Tokyo: TEC Company. 1–136.
Kiparsky, P. (2002): On the Architecture of Panini's Grammar. CIEFL Hyderabad. Conference on the Architecture of Grammar. January 2002.
Koster, J. (1987). *Domains and Dynasties.* Dordrecht: Foris.
Labov, W. (1966): Hypercorrection as a factor in linguistic change. In: Bright, W. ed. (1966): *Sociolinguistics.* Janua Linguarum XX. The Hague: Mouton.
Ladefoged, P. & Maddieson, I. (1996): *The sounds of the world's languages.* Oxford: Blackwell Publishers.
Leben, W. (1973): *Suprasegmental Phonology*. PhD dissertation, MIT, Cambridge.
Legendre, G., Smolensky, P. & Wilson, C. (1998): When is less more? Faithfulness and minimal links in wh-chains. In: Barbosa, P., Fox, D., Hagstrom, P., McGinnis, M. & Pesetsky, D., eds. (1998): *Is the best good enough? Optimality and competition in syntax.* MIT Press & MIT Working Papers in Linguistics, Cambridge, MA. 249–289

Lindblom, B. (1989): Phonetic invariance and the adaptive nature of speech. In: Elsendoorn, Ben A.G. & Bouma, H., eds. (1989): *Working Models of Human Perception.* San Diego: Academic Press. 139–174.
McCarthy, J. J. (1986): OCP Effects: Gemination and Antigemination. *Linguistic Inquiry* 17:2. 207–263.
McCarthy, J. J. & Prince, A. S. (1993a): *Prosodic Morphology: Constraint Interaction and Satisfaction.* RuCCS-TR-3. (ROA-482-0102; http://roa.rutgers.edu).
McCarthy, J. J. & Prince, A. S. (1993b): Generalized Alignment. In: Booij, G. & Marle, Jaap van, eds. (1993): *Yearbook of Morphology.* Dordrecht: Kluwer. 79–153.
McCarthy, J. J. & Prince, A. S. (1995): Faithfulness and reduplicative identity. In: Beckman, J. N., Walsh; L. & Urbanczyk, S., eds. (1995): Papers in Optimality Theory. *University of Massachusetts Occasional Papers* 18. 249–384.
McCarthy, J. J. & Prince, A. S. (1995): Prosodic Morphology. In: Goldsmith, J., ed. (1995): *Handbook of Phonology.* Oxford: Basil Blackwell. 318–366.
McCarthy, J. J. & Prince, A. S. (1994): The Emergence of the Unmarked: Optimality in Prosodic Morphology. In: Gonzalez, M., ed. (1994): *Proceedings of the North-East Linguistics Society* 24. Amherst, MA: Graduate Linguistic Student Association. (ROA-13-0594; http://roa.rutgers.edu).
Mc Carthy, J. J. (2000): Faithfulness and Prosodic Circumscription. In: Dekkers, J., Leeuw, F. van der & Weijer, J. van de, eds. (2000): *Optimality Theory: Syntax, Phonology, and Acquisition.* Oxford: Oxford University Press. 151–189.
McCarthy, J. J. (2002): *Against Gradience.* Ms, Rutgers Optimality Archive [ROA-510-0302; http://roa.rutgers.edu].
McCloskey, J. (1996): Subjects and Subject Positions in Irish. In: Borsley, B. & Roberts, I., eds. (1996): *The Syntax of Celtic Languages.* Cambrigde: Cambridge University Press. 241–283.
Müller, G. (2000): Order Preservation, Parallel Movement, and the Emergence of the Unmarked. In: Legendre G., Grimshaw, J. & Vikner, S., eds. (2000): *Optimality Theoretic Syntax.* MIT Press. (ROA-275-0798).
Orgun, C. O. & Sprouse, R. L. (1999): From MPARSE to CONTROL: Deriving ungrammaticality. *Phonology* 16. 191–224.
Padgett, J. & Chiosáin, M. N. (2002): Markedness, Segment Realization, and Locality in Spreading. In: Lombardi, L. (ed.) Constraints and

Representations: Segmental Phonology in Optimality Theory. Cambridge University Press.
Pesetsky, D. (1998): Some Optimality Principles of Sentence Pronunciation. In: Barbosa, P., Fox, D., Hagstrom, P., McGinnis, M. & Pesetsky, D., eds. (1998): *Is the best good enough? Optimality and competition in syntax.* MIT Press & MIT Working Papers in Linguistics, Cambridge, MA. 337–383.
Prince, A. & Smolensky, P. (1993): *Optimality Theory: Constraint interaction in generative grammar.* University of Colorado: Boulder.
Prince, A. & Smolensky, P. (1993): Optimality Theory: Constraint Interaction in Generative Grammar. RuCCS *Technical Report* 2. Piscataway, NJ: Rutgers Center for Cognitive Science, Rutgers University, and Boulder, CO: Department of Computer Science, University of Colorado. (ROA-537-0802; http://roa.rutgers.edu).
Pritchett, B. (1992): *Grammatical Competence and Parsing Performance.* Chicago: University Press.
Repp, B. (1982): Phonetic Trading Relations and Context Effects: New Experimental Evidence for a Speech Mode of Perception. *Psychological Bulletin* 92.1. 81–110.
Reuland, E. (2001): Primitives of Binding. *Linguistc Inquiry* 32:3.
Rizzi, L. (1982): Issues in Italian Syntax. *Studies in Generative Grammar* 11. Foris: Dordrecht.
Rizzi, L. (1986): Null objects in Italian and the theory of pro. *Linguistic Inquiry* 17. 501–557.
Rizzi, L. (1990): *Relativized Minimality.* Ms, Cambrigde: MIT-Press.
Sagey, E. (1986): *The Representation of Features and Relations in Non-Linear Phonology.* PhD dissertation, MIT, Cambridge.
Schütze, C. T. (1996): The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology. University of Chicago Press.
Shevelov, G. Y. (1964): *A prehistory of Slavic: the historical phonology of common Slavic.* Heidelberg.
Smith, N. (1973): The acquisition of phonology: a case study. Cambridge University Press.
Smolensky, P. (1995): On the internal structure of the constraint component Con of UG. Talk given at the University of Arizona.
Stevenson, S. & Smolensky, P. (1997): Optimal Sentence Processing. Conference on Computational Psycholinguistics.
Sobin, N. (1985): Case Assignment in Ukrainian Morphological Passive Constructions. *Linguistic Inquiry* 16:4.

Stiebels, B. (2000): *Typologie des Argumentlinking.* Unpublished PhD dissertation, Düsseldorf.
Trommer, J. (2002b): *Distributed Optimality.* PhD dissertation, University of Potsdam.
Trubetzkoy, N. S. (1939): Grundzüge der Phonology. *Travaux du Cercle Linguistique de Prague* 7. Göttingen: Vandenhoek & Ruprecht.
Uszkoreit, H. J. (1984): *Word order and constituent structure in German.* PhD dissertation, University of Austin, Texas. [first published in: (1987). CSLI publications. California, Standford.]
Wiese, R. (1986): Schwa and the structure of words in German. *Linguistics* 24. 697–724.
Wiese, R. (1986): *The phonology of German.* Oxford: Clarendon Press.
Wiese, R. (1996): i-Bildungen im Deutschen: Prosodischen Morphologie und Repräsentationsebenen in der Phonologie. Vortrag gehalten bei der DGfS-Tagung in Freiburg.
Wilson, C. (2001): Bidirectional Optimization and the Theory of Anaphora. In: Legendre, G., Vikner, S. & Grimshaw, J., eds. (2001): *Optimality-theoretic Syntax.* Stanford University: MIT Press.
Woolford, E. (2001): Case patterns. In: Legendre, G., Vikner, S. & Grimshaw, J., eds. (2001): *Optimality-theoretic Syntax.* Stanford University: MIT Press.
Wunderlich, D. (1997): Cause and the Structure of Verbs. *Linguistic Inquiry* 28. 27–68.
Wunderlich, D. (2002): Optimal case patterns: German and Icelandic compared. In: Brandner, E. & Zinsmeister, H., eds. (2002): *New Perspectives on Case Theory.* CSLI Publications.
Wurzel, W. U. (1970): Studien zur deutschen Lautstruktur. *Studia Grammatica* VIII. Berlin: Akademie-Verlag.
Wurzel, W. U. (1980): Phonologie. In: Heidolph, K. E., Flämig, W., Motsch, W. et al, eds. (1980): *Grundzüge einer deutschen Grammatik*. Chapter 5. Berlin: Akademie-Verlag.