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.