Optimality Theory

An Overview

edited by Diana Archangeli
and D. Terence Langendoen

BLACKWELL Publishers

1

Optimality Theory: An Introduction to Linguistics in the 1990s*

Diana Archangeli

Optimality Theory (henceforth "OT") is THE Linguistic Theory of the 1990s. It made its public debut at the University of Arizona Phonology Conference in Tucson in April 1991, when Alan Prince and Paul Smolensky presented a paper entitled simply 'Optimality'. In the spring of 1993, linguists around the world found in their mailboxes a pair of hefty and convincing manuscripts: *Optimality Theory: Constraint Interaction in Generative Grammar* by Alan Prince and Paul Smolensky and *Prosodic Morphology I: Constraint Interaction and Satisfaction* by John McCarthy and Alan Prince. Research in Optimality Theory, especially in the area of linguistics known as phonology (see Section 2), has grown tremendously ever since, and is coming to dominate the world of linguistic research as presented at conferences, workshops, seminars, and colloquia; and the Rutgers Optimality Archive is perhaps the most active and extensive of the various electronic publication outlets in linguistics (see Foreword). The impact of OT in the areas of linguistics outside of phonology has not been as dramatic, but it has been significant, and is likely to rival its impact in phonology before long.

Since OT is a theory of generative linguistics and has had its greatest impact so far in phonology, the next two sections present brief summaries of the goals of generative linguistic theory, and more specifically of the goals of phonological theory. Readers who are familiar with this material can skip directly to Section 3, where discussion of OT begins.

*Special thanks to Michael Hammond, D. Terence Langendoen, Dirk Elzinga, Keiichiro Suzuki, and Margaret Speas for their careful reading and suggestions which led to improvements in this chapter. Work on this chapter was supported in part by NSF grant BNS-9023323 to the author.

1 What is Linguistics?

There are two central research objectives in linguistics. The first is to determine and characterize **universal properties of language**, the properties that are shared by all languages. Although the manifestation of a specific universal in a particular language may not be the same as it is in the language next door, such universals are thought to be present in some regard in every language. This leads to the second research objective in linguistics, to determine and characterize the range of **possible language variation**.

> **Linguistics is the study of...**
> 1. **language universals**: the range and type of properties shared in some way by all languages.
> 2. **language variation**: the range and type of variation possible between languages.

By the definitions of language–universal and language–specific properties given above, one might imagine that there is a continuum between the two. The term **markedness** is used to refer to this continuum, with completely unmarked properties being those found in virtually all languages and extremely marked properties being found quite rarely. Language universals must be formulated in a way that is able to characterize this distribution.

The central hypothesis driving generative linguistic research today, due to Noam Chomsky (Chomsky 1965, 1975, 1986; see also Pinker 1994), is that these universals are part of the genetic inheritance of every normal human being. Thus, not only do human beings have an innate ability to learn language, but this innate ability is limited, so that not all strings of sounds can be learned as a language, just as not all strings of words can be put together as a sentence of a language. Universal properties of the world's languages result from inflexibility in this innate language capacity; language variation arises from its flexibility. Linguists use the term **universal grammar** to refer to the innate language knowledge that humans have, including both the flexibility and the inflexibility. In our discussion of Optimality Theory, we will see how the model encompasses both universal and language–specific properties, and how markedness is expressed.

> **Universal Grammar...**
> is the innate knowledge of language that is shared by normal humans — it characterizes both the universal properties of language and the variation tolerated among specific languages

In studying a language, the linguist finds evidence to show that there is a pattern to study, then figures out what the nature of the pattern is, and, finally, determines a formal characterization of the pattern. In each of these efforts, linguists maintain a fairly broad approach. When finding a pattern, the concern is not simply "does this pattern exist?" but also "how does this pattern interact with other patterns in the language?" and "how does this pattern compare to similar patterns in other languages?"

For example, in English there are adjectives like *active*, *tangible*, and *possible*. A negative form of each adjective can be created by adding a prefix, resulting in *inactive*, *intangible*, and *impossible*. The linguist notes that the negative prefix takes the form *im-*; which ends with a labial nasal (m), whenever it precedes an adjective which begins with a labial stop (*p*, *b*), otherwise it takes the form *in–*: *imbalance*, *impolite*, but *inoperative*, *intangible*, *infallible*, *inviolable*. The prefix, then, is analyzed as having an input form, *in–*, which relates to two different output forms, *in–* and *im–*, depending on the context in which the prefix is placed. (See also Chapter 3 for more about this sort of sound change and Chapter 4 on the standard generative phonology relation between a single input and a variety of output forms.) In characterizing patterns, whether phonological, morphological, or syntactic, linguists try to determine the input form, the output form(s), and the nature of the relation between input and output. Optimality Theory offers a specific view of the nature of that relation.

Studies that focus on a single language explore the patterns that exist within that language. Studies that focus on comparable phenomena across languages examine the range of variation possible within natural human language. By understanding the variation that does occur, we are also able to determine those areas where there is no variation. The more common properties or patterns are thought to be universal, part of our innate language endowment. Not all universals are manifested in the same way in all languages however, due to variation. The more robust a universal is in a particular language, the less marked the language is in that respect. A highly marked property is one which has minimal (or no) claims to universality.

Linguists look for...		**to determine...**
a.	**patterns**	their existence and characteristics
b.	**variation**	differences among the patterns of different languages
c.	**universals**	the properties that are part of our innate language endowment
d.	**markedness**	the robustness of a given property within a language

These methods and goals can be more concretely understood by working through particular language data. For example, consider the phonological universal that words start with a consonant–vowel ("CV") sequence. (Ultimately, we refine this notion in terms of syllables and onsets; for the moment "words start with a CV sequences" is adequate.) In English *sing*, *like*, *wish* all start with a "CV" sequence. Languages share this property to different degrees. For instance, in Yawelmani (a language we examine in some detail below) *every* word starts with such a sequence. By contrast, the English pattern shows variation in two ways: on one hand, it allows words to start with more than one consonant, e.g. *stripe*, *gleam*, *smooth*, while on the other hand, some words start with a vowel (and no consonant): *apple*, *important*, *up*. In this regard, the syllables of Yawelmani are less marked than are those of English.

Within linguistics there are four major subdisciplines: phonology, morphology, syntax, and semantics, defined in (1.1). The first three are topics of chapters in this book. There are other subdisciplines as well, including psycholinguistics, sociolinguistics, and phonetics. However, the four areas mentioned here are the core disciplines within formal linguistics.

Each chapter discusses the application of Optimality Theory in a specific subdiscipline in linguistics; in each, we explore the way in which OT characterizes the universals, variation, and markedness in that subdiscipline. OT began its life as a theory of phonology; this introductory chapter follows suit to a large extent. However, the points made extend to other subdisciplines, as is demonstrated in Chapter 4 for morphology and Chapters 5 and 6 for syntax. In this chapter, sound patterns are used simply as a vehicle for better understanding how the model works.

(1.1) The four major subdisciplines in linguistics

a.	**phonology**	The study of how sounds combine to make morphemes and words, e.g. *in–active*, but *im–polite*, not *in–polite*
b.	**morphology**	The study of how morphemes combine to make words, e.g. *act–ing, in–act–ive*, but not *in–act–ing* 'not acting'
c.	**syntax**	The study of how words combine to make sentences, e.g. *I saw the dog* is good English, *I saw dog the* is not.
d.	**semantics**	The study of how meanings of subparts combine to make meaning of the whole.

2 An Extended Example: Syllable Structure

To make our discussion of patterns, variation, universals, and markedness concrete, some properties of the cross–linguistic distribution of consonants and vowels of words are illustrated in (1.2), with examples from Hawaiian, English, Berber, and Yawelmani.

A simple pattern of consonants and vowels is found in Hawaiian (1.2a). Hawaiian allows no more than one consonant in a row so we find words like *kanaka* 'man' with three singleton consonants: *kanaka*. However, Hawaiian has no sequences of consonants. In fact, when borrowing words from another language, any consonant sequences are altered to fit the Hawaiian pattern: English *flour* becomes *palaoa*; English *velvet* becomes *weleweka*, etc.

English illustrates the opposite extreme, for it allows long strings of consonants in the middle and at the edges of words, as in *construct* and *sprig*, illustrated in (1.2b). An even more extreme case is illustrated by Berber, a language spoken in Morocco, which does not require vowels at all in its words, *txdmt* 'gather wood' along side *ildi* 'pull'.

Finally, a middle ground is struck in Yawelmani, a Native American language that was once spoken in California (Newman 1944). This language allows at most two consonants in a sequence within a word, as in *xa[th]in*, where the sequence *th* represents two consonants, *t* and *h*. Additionally, Yawelmani tolerates at most one consonant at the beginning and one at the end of a word: *xathin* starts with a single *x* and ends with a single *n*.[1]

[1]Since phonology studies the sounds of words, it is important not to get confused by the orthographic conventions of a particular language. For example, the symbol [θ] is used to represent the sound spelled *th* in an English word like *sixth* or *ether*. The symbol sequence [th] as in *xathin*

(1.2) Example: cross–linguistic distribution of consonants and vowels in words

a.	Hawaiian	allows no more than one consonant in a row	*wahine* 'woman' *alapine* 'often'
b.	English	allows long strings of consonants…	construct; sprig *sɪksθs* (*sixths*)
		but doesn't require them.	*maven*
c.	Berber	allows words to consist solely of consonants…	*trglt* 'lock' *txdmt* 'gather wood'
		but also allows vowels in words.	*ildi* 'pull'
d.	Yawelmani	allows up to two consonants in the middle of words…	xa*th*in 'ate'
		but allows at most one consonant at word edges.	*x*athi*n* 'ate'

The four languages illustrated here demonstrate that there is a wide range of ways in which consonants and vowels distribute themselves within words in the world's languages. Significantly, there are also many patterns of consonants and vowels that you can think of that simply do not occur in natural languages. One such imaginable but non–occurring language would stack up all the consonants at the beginning of the word and all the vowels at the end of the word (1.3a). Words like *mrnaia* would exist, but no words like *marina*. A more "language–like" example would be comparable to English except that it *requires* all words to start with two or more consonants (1.3b). Words like *sprig* would be well–formed in this language, but not a word like *construct*, for *construct* begins with a single consonant.

(1.3) Some imaginable but non–occurring languages

a.	All consonants are in a sequence at the left edge of the word, followed by all vowels.	OK: not OK:	*spree, blue, mrnaia* *sprig, lube, marina*
b.	Every word begins with a string of consonants, otherwise like English.	OK: not OK:	*string, sprig, blue* *ring, pig, every*

There are no languages like those sketched in (1.3) and yet *it is not hard to describe such patterns*. In fact, many of the nonexistent patterns are *easier* to describe than some of the patterns found in natural languages, such as those in (1.2).

Through this extended discussion of consonant and vowel distribution, we have arrived at the central issues facing students of language. Although our example has been in terms of the sound systems of languages, the questions themselves are general and extend to all domains of language study.

(1.2d) is two sounds, as in *hot headed*; not one, as in *ether*. Finally, the sound symbolized by [x] in (1.2c, d) is a voiceless velar fricative, the final sound in the German pronunciation of a name like *Bach*. Following conventions of the field, square brackets are used to enclose symbols which represent sounds directly, such as [θ], [th], and [x] above.

The three central questions in linguistic research:
1. What are the patterns that occur in natural languages?
2. How do we characterize the occurring patterns?
3. How do we exclude the patterns that we do not find and that we think we never will find?

In the next section, we explore the distribution of consonants and vowels in Yawelmani to illustrate the way linguists try to answer these three questions. We first examine the general answers to these questions, then turn to the Optimality Theoretic answers to questions of variation and universals to illustrate the workings of that model.

Yawelmani CV Distribution

There are four basic facts about the distribution of consonants and vowels in Yawelmani. First, words must begin with a consonant (1.4a). For example, the word *xathin* begins with a single [x]. By contrast, the Hawaiian word *alapine* 'often' starts with a vowel; a sequence like *alapine* is not a possible word in Yawelmani. In Yawelmani, a word like *alapine* would have the same "feel" to it as [bnɪk] *bnick* has in English: the sounds are all acceptable in the language but their organization is wrong. (Contrast this with the sequence [blɪk] *blick*, where the sounds and the organization are acceptable for English words, even though it happens not to be a word in English. See Chapter 2 for further discussion of words, possible nonwords, and impossible sequences.)

Second, at word edges, Yawelmani allows no more than one consonant (1.4b). This contrasts with words like [strɛŋθ] 'strength' in English where several consonants in sequence are found at both edges. (Note that Yawelmani words are not *required* to end in a consonant, unlike the word–initial requirement already discussed. However, if a Yawelmani word does end in a consonant, it may have only *one* consonant at the end.)

Third, there is no more than one vowel in a sequence in Yawelmani (1.4c). In *xathin*, both the a̲ and the i̲ are flanked by consonants. This contrasts with a language like Hawaiian or English, which both allow two vowels in a row. Finally, in the middle of words you can get at most two consonants together, although one consonant by itself is also permitted (1.4d). In this regard as well, Yawelmani differs from English, for English allows longer medial strings of consonants as in *pe<u>rsn</u>ickety*. (Examples in the "not OK" column are drawn from real words of Hawaiian, English, and Berber.)

(1.4) Additional Yawelmani facts

		OK	NOT OK
a.	Words begin with exactly one consonant.	*x*athin	*a*loha 'greeting'; *a*pple, *o*dd
b.	At word edges, only one C is allowed.	*x*athi*n*	*str*eng*th*
c.	No more than one vowel occurs in sequence.	x*a*th*i*n	ke*i*ki 'child', l*eo* 'voice'; al*iə*n (al*ie*n)
d.	Word–internally, CC is OK, but not necessary.	xa*th*in, xa*t*en	i*nstr*uct, co*nsp*ire; *trglt* 'lock'

As stated, the observations listed above do not reveal any obvious pattern. Other statements could be constructed, for example statements like those in (1.4) except replacing the word "consonant" by "vowel" and vice versa. That is, not only is the existence of a pattern unclear from the above, but so is its relation to universals governing the arrangement of consonants and vowels in words.

The statements in (1.4) are apparently unrelated observations about the placement of consonants and vowels in words. Although these locations can be stated clearly, it is not obvious from the list why these particular patterns are found and not others. In order to make sense of such observations about different languages, linguists have proposed that consonants and vowels are organized into constituents composed of consonants and vowels, called **syllables**, and that *words are composed of syllables*. That is, the distribution of consonants and vowels is characterized in terms of where each occurs in a syllable, a chunk smaller than a whole word and whose properties are easier to characterize. The distribution of consonants and vowels in words follows from the patterns that result when syllables are strung together.

In the next section, we examine the way in which phonologists characterize these facts to reveal the essential organization of consonants and vowels in Yawelmani words.

Words are Composed of Syllables

Under the assumption that words are composed of syllables, the linguist characterizes possible syllables, rather than possible words, both universally and for a given language. (See also Chapter 2.) In (1.5), I list certain general tendencies of syllables.

The terms used in the right–hand column of (1.5) are standard for referring to the parts of a syllable. A "CVC" syllable like [kæt] 'cat' has an **onset** [k], the initial consonant; it also has a **peak** [æ], the vowel; and it has a **coda** [t], the syllable–final consonant. A **complex onset** and a **complex coda** are found in [klæsp] 'clasp', which begins with two consonants, [kl], and ends with two, [sp]. The symbol "*" is used by linguists to indicate unacceptability. For example, placing a * at the beginning of a sentence indicates that the sentence is ungrammatical: **John seems that he ran.* Thus, *COMPLEX is shorthand for "complex onsets and complex codas are unacceptable".

(1.5) Typical properties of syllables

a.	Syllables begin with a consonant.	ONSET
b.	Syllables have one vowel.	PEAK
c.	Syllables end with a vowel.	NOCODA
d.	Syllables have at most one consonant at an edge.	*COMPLEX
e.	Syllables are composed of consonants and vowels.	ONSET & PEAK

There are two points of significance here. First, these statements are general tendencies, not absolute laws. Thus, there are syllables in languages which violate some of these properties, a point that OT exploits as we will see below. Second, the standard definition of a syllable, a constituent composed of at least one consonant followed by a vowel, results from combining (1.5a) and (1.5b): if a syllable starts with a consonant, it satisfies ONSET and if it has a vowel, it satisfies PEAK. This is one example of the observation that, by characterizing syllables in terms of the four simple properties in (1.5a–d) (which must be stated at any rate), further properties are also characterized.

In the next section, we see how each of these properties is manifested in Yawelmani. We also see how sequences of well-formed Yawelmani syllables result in the distribution of consonants and vowels in Yawelmani words, listed above in (1.4).

Explaining Yawelmani Consonant and Vowel Distribution Using Syllables

Figure (1.6) shows how the general tendencies of syllables (given in (1.5)) are realized in Yawelmani. The only one of these tendencies that does not hold absolutely in Yawelmani is NOCODA (1.6c) since some syllables do end with consonants.

(1.6) Properties of Yawelmani syllables

	general tendency		*Yawelmani*
a.	PEAK	Syllables have one vowel.	always
b.	ONSET	Syllables begin with a consonant	always
c.	*COMPLEX	Syllables have at most one consonant at an edge.	always
d.	NOCODA	Syllables end with a vowel.	sometimes

The chart in (1.7) shows that if the only violable constraint in Yawelmani is NOCODA, then two types of syllables result, a CV syllable (1.7a) and a CVC syllable (1.7b). (We postpone discussion of how to characterize which constraints are violated in a particular language until Section 3.) Other imaginable syllable types, such as CVCC or CC (1.7c,d), are impossible in this language. CVCC syllables do occur in other languages, for instance in English *cart*, *desk*, and *tact*. English, then, tolerates violations of *COMPLEX. CC syllables occur in Berber; Berber tolerates violations of PEAK. A language which allows no violations of syllable constraints whatsoever has only CV syllables, (1.7a).

(1.7) How the Yawelmani syllable properties give rise to syllables

		PEAK	ONSET	NOCODA	*COMPLEX
☞	CV	OK	OK	OK	OK
☞	CVC	OK	OK	FALSE	OK
*	CVCC	OK	OK	OK	FALSE
*	CC	FALSE	OK	OK	OK

The virtue of proposing that words are composed of syllables is that once we characterize the syllables of a language, lists of observations such (1.4) are seen to be exactly the properties we would expect. The discussion below shows how the observations about Yawelmani that are listed (1.4) are formally characterized by the properties given in (1.5), schematized below.

Yawelmani words begin with exactly one consonant (1.4a). Since each *syllable* necessarily begins with a consonant (ONSET), each *word* also begins with a consonant. Neither a vowel alone nor a vowel–consonant sequence is a syllable and so no word in the language can start with a vowel. By comparing the syllabification of *xathin* with that of the English word *aha*, we see that violations of ONSET are not tolerated in Yawelmani, although they are allowed in English.

In (1.8), the violation of ONSET is indicated by an asterisk (*). The pictures here use a standard notation for syllables, where "σ" stands for "syllable" and syllable membership is shown by a triangle between a σ at the top and a sequence of consonants and vowels at the bottom. Another notation, used elsewhere in this chapter and book, uses dots to separate syllables: *xat.hin* and *a.ha*.

(1.8) Words begin with exactly one consonant: ONSET

		ONSET	
☞	σ σ xat hin		
	σ σ a ha	*	not a possible word in Yawelmani

At word edges in Yawelmani, only one C is allowed (1.4b). The longest string of consonants that Yawelmani allows at the beginning of a word is the single consonant which satisfies ONSET. Similarly, a word might end with a single consonant, or with a vowel, because that is how syllables end. But a word cannot begin or end with two or more consonants. Our characterization of syllables explains this point: due to *COMPLEX, syllables do not end with more than one consonant. Since words are composed of syllables, words cannot end that way either.

(1.9) At word edges, only one C is allowed: *COMPLEX

		*COMPLEX	
☞	σ σ xat hin		
	σ striŋ	*	not a possible word in Yawelmani

In Yawelmani, no more than one vowel occurs in a row (1.4c). With CV and CVC syllables, it is impossible to get two vowels in a row. Consonants must always intervene between vowels due to the necessary syllable–initial consonant. To get two vowels in a row, a syllable would have to be able to start with a vowel, a violation of ONSET, as shown in (1.10). In such a case, the second syllable is defective: unlike some languages (like English) in which syllables may start with vowels, no Yawelmani syllable does so.

In Yawelmani, word–internally, CC is OK (1.4d). The longest possible uninterrupted sequence of consonants in Yawelmani is two: this occurs if the syllable on the left ends with a consonant — recall that the one on the right must begin with a consonant. Since syllables cannot begin or end with a sequence of consonants, no more than two consonants in sequence arises from two syllables in sequence. Since syllables must have a peak (a vowel), the single C cannot be a syllable as in the third candidate in (1.11).

(1.10) **No more than one vowel occurs in sequence: ONSET**

	ONSET	
☞ σ σ xat hin		
σ σ xa in	*	not a possible word in Yawelmani

(1.11) **Word–internally, CC is OK: *COMPLEX and PEAK**

	*COMPLEX	PEAK	
☞ σ σ xat hin			
σ σ log whin	*		not a possible word in Yawelmani
σ σ σ log w hin		*	not a possible word in Yawelmani

To summarize, we have made two proposals. First, words are composed of syllables. Second, syllables in Yawelmani are limited by the criteria in (1.6a–c). An immediate result of these proposals is that the list of facts about the distribution of consonants and vowels in Yawelmani words follows; nothing further need be said.

The role of syllable structure

1. Words are composed of syllables.
2. The facts about the distribution of consonants and vowels in a language follow from the structure of syllables in that language.

It is important to examine more closely the syllable properties given in (1.5) and (1.6). These sets of statements are stated as constraints on specific aspects of a syllable. Each of these statements expresses a strong universal tendency. For example, although it is not the case that all languages require onsets (ONSET), it is the case that *every language allows onsets* and *no language disallows onsets*. By allowing these constraints to be violated in some languages, two results are accomplished. First, language specific patterns and variation between languages are admitted into the model through such violations. Second, markedness is admitted into the model: each constraint violation indicates markedness in that respect. Employing constraints as we have done so far addresses the issues central to linguistic analysis: patterns, variation, universals, and markedness.

Constraints characterize universals.

Constraint violations characterize markedness, patterns, and variation.

3 Constraint Ranking and Faithfulness

Optimality Theory proposes that Universal Grammar contains a set of violable constraints. The constraints, as noted above, spell out universal properties of language. OT also proposes that each language has its own ranking for these constraints. Differences between constraint rankings result in different patterns, giving rise to systematic variation between languages.

Optimality Theory views...

Universal Grammar as a set of violable constraints

the grammars of specific languages as the language–particular ranking of those constraints

The constraints include ones governing aspects of the phonology, such as the syllabification constraints just examined. The constraints also include ones governing morphology, including constraints which determine the appropriate position of morphemes (see Chapter 4). Finally, the constraint set includes constraints which determine the correct syntactic properties of a language, such as "a noun phrase must have case" (see Chapters 5 and 6). There is one family of constraints whose properties cut across all subdisciplinary domains, namely the **FAITHFULNESS constraints**, which say that the input and output are identical. Violations of FAITHFULNESS lead to differences between input and output, such as the difference seen in the prefix *in–/im–*, discussed in Section 1 above.

FAITHFULNESS constraints require that the output be identical to the input.

Violation of constraints is tolerated in a very limited context. A constraint may be violated successfully only in order to satisfy a higher ranked constraint.

In this section, we examine how constraint ranking and violation handles the Yawelmani data. As already seen, in Yawelmani only one of the basic syllabification constraints may be violated, NOCODA. As a violable constraint, NOCODA must be outranked by some more important constraint(s). In this case, the relevant constraints are FAITHFULNESS constraints, one requiring faithfulness of consonants between input and output (FAITHC) and the other requiring faithfulness of vowels (FAITHV). In order to understand the necessary constraints, it is useful to explore the options available to avoid a NOCODA violation.

Syllabify the consonant as a peak. The first possibility is simply to syllabify the offending consonant as a syllable in and of itself, thereby violating PEAK. In this way, the offending consonant is now a peak, and so no longer violates NOCODA. Again, Yawelmani does not take this option: *xa.ten*, not **xa.te.n*, 'will eat'. (The dots refer to syllable boundaries, e.g. in *xa.ten*, the first syllable is *xa* and the second is *ten*, while in *xa.te.n*, there are three syllables, *xa*, *te*, and *n*. The * indicates that the syllabification *xa.te.n* is ill–formed.)

Delete the offending consonant. A second possibility is to simply delete the consonant which would otherwise be syllabified as a coda, thereby violating FAITHC. In Yawelmani, this does not happen. In the word *xaten* 'will eat', composed of *xat–* 'eat' and *–en* 'future tense', the word–final *–n* is also in coda position. It does not delete: *xa.ten*, not **xa.te*, 'will eat'.

Insert a vowel. A final possibility is to add a vowel after the offending consonant, constituting a violation of FAITHV. In this way, the offending consonant is now an onset for the added vowel, and so no longer violates NOCODA. Again, Yawelmani does not take this option: *xa.ten*, not **xa.te.ni*, 'will eat'.

The figure in (1.12) illustrates the above discussion in the manner common to most work in Optimality Theory. The figure is called a **tableau**; the constraints are ranked across the top, going from highest ranked on the left to lowest ranked on the right. Solid lines between constraints indicate crucial rankings while dashed lines indicate that the ranking is not (or not yet) crucial. In this example, for instance, it is crucial that NOCODA be subordinate to all other constraints. (Ultimately, we will see that FAITHV ranks below the other constraints and above NOCODA.)

The top left–hand cell shows the input representation for which candidates are being considered. Candidates show up in the leftmost column, with the optimal candidate indicated by the symbol "☞". The optimal candidate is the one with the fewest lowest violations. Violations are indicated by asterisks (*), and an exclamation point highlights each "fatal" violation, i.e. the violation that eliminates a candidate completely.[2] Shaded areas indicate constraints that are irrelevant due to the violation of a higher ranked constraint. (Only NOCODA gets shaded since it is the only constraint that must be dominated given the data considered so far.)

(1.12) A tableau showing [xa.ten] is the optimal candidate

/xat–en/	PEAK	ONSET	*COMPLEX	FAITHC	FAITHV	NOCODA
☞ xa.ten						*
xa.te.n	*!					
xa.te				*!		
xa.te.ni					*!	

The first candidate in (1.12), [xa.ten], is the optimal one. Its only violation is that of NOCODA, the lowest ranked constraint. If the final consonant is syllabified as a peak, as in [xa.te.n] (the second candidate in (1.12)), the NOCODA violation is avoided, but only at the cost of a fatal PEAK violation.

The role of FAITHFULNESS becomes apparent when we consider the next two candidates, both of which avoid the NOCODA violation at the expense of a FAITH violation. Candidate [xa.te] in (the third candidate in (1.12)) has lost the final consonant thereby incurring a FAITHC violation. Since FAITHC is higher–ranked than NOCODA, this violation is fatal, and eliminates *xa.te* from consideration. The form [xa.te.ni] in (the fourth candidate in (1.12)) suffers a comparable fate. In this form, an extra vowel allows the last consonant to syllabify as an onset, which eliminates the NOCODA violation. Due to the extra vowel, however, the form incurs a fatal violation of FAITHV.

[2]Given that GEN creates an infinite set of candidates, a necessary strategy when presenting tableaux is to restrict the candidates presented in the tableau to those which are critical to the point being made — the infinite set could not possibly be considered! Thus, the candidates in (1.12) are limited to four relevant ones. A second, similar, strategy is to omit from tableaux the candidates which violate undominated constraints in the language. The tableau in (1.12) contains an example: since it has been established that Yawelmani never violates ONSET, [xat.en] (with an ONSET violation) need not be included in subsequent tableaux.

> Satisfying some higher ranked FAITHFULNESS constraint (s) in Yawelmani may force the violation of lower ranked constraint(s).

4 Optimality Theory

At this point, we have worked through a specific example in order to understand the way in which Optimality Theory works. In this section, I take a more formal approach in laying out properties of the model.

Optimality Theory, like other models of linguistics, proposes an input and an output and a relation between the two. In transformational (or derivational) views, which have been the dominant paradigm in linguistic research since the mid 1960s, the input is the starting point, there is a series of operations performed on the input, and the result of these operations is the output. Crucially, if an operation makes some change in the input, that changed form serves as the input to the next operation. See especially Chapters 3 and 4 for discussion of this point.

In OT, the relation between input and output is mediated by two formal mechanisms, **GEN** and **EVAL**. GEN (for **Generator**) creates linguistic objects and notes their faithfulness relations to the input under consideration. EVAL (for **Evaluator**) uses the language's constraint hierarchy to select the best candidate(s) for a given input from among the candidates produced by GEN. The constraint hierarchy for a language is its own particular ranking of **CON**, the **universal set of constraints**.

The roles of GEN, EVAL, and CON are illustrated in (1.13), which schematically presents how OT determines the optimal syllabification for the input /xat–en/. The input feeds into GEN, which creates candidates. The candidates are considered by EVAL, which selects the optimal candidate from the set.

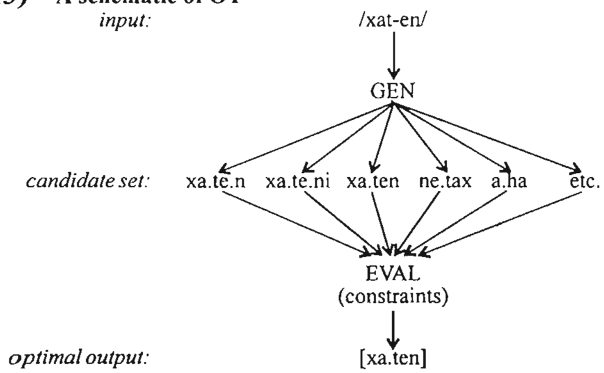
This specific example is the same one we have just worked through, namely determining the optimal syllabification for /xat–en/. The tableau in (1.12) makes explicit the role of EVAL. However, as can be seen by inspecting (1.13), there is quite a bit more to the formal aspects of OT than simply reading a tableau. In this section, we examine the Input, GEN, CON, and EVAL.

The Input

Universal Grammar provides a vocabulary for language representation; all inputs are composed from this vocabulary. As a result, inputs are linguistically well–formed objects in the sense that the input does not contain nonlinguistic objects. This is the sole restriction imposed on the input: all other constraints are found in EVAL. In the specific example we have been examining, the vocabulary provided by Universal Grammar must include consonants, vowels, and syllables. For other examples discussed in this book, this vocabulary must include phonological features and the categories *noun*, *verb*, etc.

The Formal Model		
1.	**GEN**	for a given input, the **Generator** creates a candidate set of potential outputs
2.	**EVAL**	from the candidate set, the **Evaluator** selects the best (optimal) output for that input
3.	**CON**	EVAL uses the language particular ranking of constraints from the **universal set of constraints**

(1.13) A schematic of OT



GENerator

GEN is a function which relates the input to a set of candidate representations, any one of which may be the optimal output form for the specific input. GEN is restricted in that it can only generate linguistic objects, ones composed from the universal vocabulary that similarly restricts inputs.

GEN is quite creative, being able to add, delete, and rearrange things without restriction. Since there are no restrictions, the candidate set created by GEN for any given input is infinite. (One candidate has one added element, another has two added elements, etc.) This particular property is a serious problem for those who wish to implement Optimality Theory either as a production and processing model or as a computational model, although efforts have been made to surmount the challenges.

GEN also has the job of indicating correspondences between input and output representations. These correspondences are crucial in evaluating the FAITHFULNESS constraints, such as FAITHV and FAITHC. How correspondence is encoded by GEN and how it is treated by EVAL is a subject of controversy in current Optimality research; however,

all researchers agree that such correspondences must be encoded in some way. See Chapter 4 for relevant discussion.

The Universal CONstraint Set

CON, as a universal set of constraints, is posited to be part of our innate knowledge of language. What this means is that every language makes use of the same set of constraints. This assumption leads directly to a characterization of the universal aspects of human language: all languages have access to exactly the same set of constraints. This is the formal means by which *universals* are encoded.[3]

Whether stated in a positive fashion (e.g. ONSET) or negatively (e.g. *COMPLEX), any constraint may end up being violated in some language: the potential for being violated is a result of the position of a constraint in a particular language's hierarchy, rather than a property of the constraint itself. In this way, the constraints also provide a measure for markedness: the higher ranked constraints (and so rarely violated) indicate the ways in which the language is unmarked while the lower ranked (and so frequently violated) constraints indicate the ways in which the language is marked. As such, markedness is encoded directly in the model. This is an important result, for earlier models have required separate theories of markedness.

EVALuation

EVAL is the mechanism which selects the optimal candidate (s) from the candidate set created by GEN. EVAL makes use of a ranking of the violable constraints. The optimal output, the one that is selected by EVAL, is the one that best satisfies these constraints.

EVAL is at the heart of Optimality Theory

1. The constraints in CON are violable.
2. The constraints are ranked.
3. EVAL finds the candidate that *best satisfies* the ranked constraints.
   - **a.** Violation of a lower ranked constraint may be tolerated in order to satisfy a higher ranked constraint.
   - **b.** Ties (by violation or by satisfaction) of a higher ranked constraint are resolved by a lower ranked constraint.

Best satisfaction can be achieved in two ways. Violations of lower ranked constraints are tolerated in the optimal form provided that they help avoid violation of some higher ranked constraint. Lower ranked constraints adjudicate when all viable candidates tie on

[3]The ideal which Optimality research aims for (and sometimes appears to fall short of) is to provide evidence of the universality of each constraint necessary for some particular language. For constraints such as the ones posited for syllabification in this chapter, universality is readily motivated; there are numerous analyses involving constraints whose status as a universal is minimal at best. At this point, it is unclear whether this is a weakness of the model itself, or a weakness of the analyses.

some higher ranked constraint, either because all candidates satisfy the higher constraint or, more interestingly, because all candidates violate the higher ranked constraint. In this way, unmarked patterns that are encoded in low–ranked constraints occasionally emerge despite those patterns not being observed throughout the language. See especially McCarthy & Prince (1994).

Summary

In this section, we have briefly explored the basic parts that Optimality Theory proposes for grammars, input, GEN, EVAL, and CON. We have seen two ways in which *universal* properties of language are encoded: (i) inputs and GEN are limited by a universal linguistic vocabulary; (ii) CON contains a universal set of constraints which all languages use. These properties are illustrated with the Yawelmani example already discussed: ONSET, PEAK, and *COMPLEX are all considered very strong universal properties.

How Optimality Theory works
1. *Universal Grammar includes*
   a. a linguistic alphabet
   b. a set of constraints, **CON**
   c. two functions, **GEN**(erate) and **EVAL**(uate).
2. *The grammar of a particular language includes*
   a. basic forms for morphemes (from which inputs are constructed)
   b. a **ranking** for the constraints in **CON**
3. *For each input,*
   a. **GEN** creates a candidate set of potential outputs
   b. **EVAL** selects the optimal candidate from that set

We have seen that *markedness* is encoded via constraints and constraint violations. Yawelmani is unmarked with respect to ONSET, PEAK, and *COMPLEX, since these constraints are never violated. However, since some Yawelmani syllables end with consonants, NOCODA can be violated. Yawelmani is marked in this respect. Under OT, markedness arises when a constraint, such as NOCODA, is violated: such violation occurs only in order to satisfy higher ranked constraints.

The *patterns* found in languages are characterized by the language–specific ranking of the universal constraints. This connects closely to language *variation*, which also arises from the different ways in which languages rank the constraints in CON. This characterization of patterns and variation is one of the most exciting and intriguing aspects of OT. The next section provides an illustration, comparing the effects of different rankings of *COMPLEX, PEAK, FAITHC, and FAITHV.

How Optimality Theory addresses the issues that concern linguists.
1. **Language variation** is characterized as different rankings of the same set of constraints.
2. **Specific patterns** are derived from the language–particular rankings of these constraints.
3. **Universals** are present in the universal — but violable — constraints.
4. **Markedness** is inherent in the model.
   a. Each constraint is a markedness statement.
   b. Specific aspects of markedness result from the ranking.

5 Language Variation as Differences in Constraint Rankings

Language variation, as already noted, also follows from the role of the constraints within particular languages. Two constraints A and B may be ranked A » B in one language and B » A in another. Each ranking characterizes the distinctive patterns of the two languages and leads to variation between them.

To illustrate language variation via differences in constraint rankings, we return to the Yawelmani data and examine some forms related to the two we have already considered, *xathin* 'ate' and *xaten* 'will eat'. The additional data supports the partial constraint ranking FAITHC, PEAK, *COMPLEX » FAITHV. We then consider alternative rankings of these four constraints, and see that different rankings result in different patterns of syllabification. Spanish, English, and Berber provide examples.

Additional Yawelmani Data

As already seen with *xaten*, adding *–en* to the end of a verb marks the future tense. Given that *logwen* means 'will pulverize', the verb root for 'pulverize' must be *logw–*, as shown in (1.14b).

(1.14) Yawelmani future tense: a V–initial suffix

	word	*morphemes*	*gloss*
a.	xaten	xat–en	'will eat'
b.	logwen	logw–en	'will pulverize'

Stripping off the *–en* suffix reveals the bare verb roots, *xat–* 'eat' and *logw–* 'pulverize'. Significantly, the root *logw–* is an unsyllabifiable sequence in Yawelmani because it ends with two consonants. Because the future suffix *–en* begins with a vowel, the consonant sequence can safely syllabify: the /g/ is the coda of the first syllable and the /w/ the onset of the second syllable, (1.15b). However, this type of root presents a problem if there are any consonant–initial suffixes: attaching a suffix which begins with a consonant to a verb root like *logw–* will necessarily create a sequence of *three* consonants. As we have already seen, no more than *two* consonants occur in a sequence in the words of this language.

(1.15) Syllabification of *xaten* and *logwen*

a. xa.ten b. log.wen

In fact, Yawelmani does have consonant–initial suffixes, for example, the past tense suffix *–hin*. When *–hin* is added to the verb root *xat–*, the two consonant sequence *th* is created. These two consonants can each find a position in a syllable, the *t* is the coda of the first syllable while the *h* is the onset of the second syllable.

(1.16) Yawelmani past tense: a C–initial suffix

word	*morphemes*	*gloss*	*syllabification*
xathin	xat–hin	'ate'	xat.hin

There are four logically possible alternatives, illustrated in (1.17). The first three solutions to the problem have outputs that are segmentally faithful to the input but have odd syllables. A complex coda (1.17a–i) or a complex onset (1.17a–ii) might be created, or the extra consonant might simply form a syllable by itself, (1.17b). The remaining two solutions each involve a mismatch between the input and the output: something is either added or lost. In both cases the resulting syllables are well–formed. First, the extra consonant might be deleted, indicated by the ~~w~~ in (1.17c). Second, a vowel might be added, (1.17d). Adding a vowel allows the extra consonant to syllabify without creating a complex onset or coda.

(1.17) Syllable well–formedness and the extra consonant

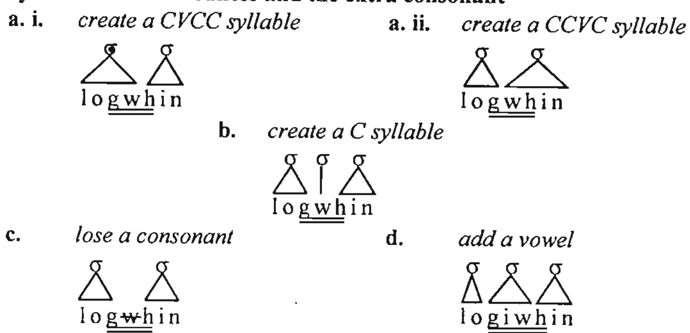


Figure (1.18) shows how the relevant constraints considered thus far stack up with respect to the logical possibilities given in (1.17). The chart shows that whenever an input representation has a three consonant cluster, one of these constraints must be violated. The question is "what do you do?" The very interesting answer is that what you do depends on what language you speak. There are languages of each type.

The chart in (1.18) shows that when a form is completely faithful (i.e. there is nothing added or deleted), one of two bad results occurs: either a violation of *COMPLEX (the first two candidates in (1.18)) or a violation of PEAK (the third candidate in (1.18)). The

unfaithful candidates avoid those two violations in one of two ways: by deleting a consonant and so incurring a FAITHC violation (the fourth candidate in (1.18)), or by adding a vowel, at the cost of violating FAITHV (the fifth candidate in (1.18)).

(1.18) An input /...CCC.../ sequence

/logw–hin/		*COMPLEX	PEAK	FAITHC	FAITHV
logw.hin	BAD	*			
log.whin	BAD	*			
log.w.hin	BAD		*		
log._hin	BAD			*	
lo.giw.hin	BAD				*

> Every output resulting from an input /...CCC.../ sequence must violate some constraint!

FAITHV is Outranked in Yawelmani

In Yawelmani, we discover that a vowel, [i], is added to rescue an otherwise unsyllabifiable consonant, a signature property of this language. We have already seen this pattern with *logwen* 'will pulverize' and *logiwhin* 'pulverized'; (1.19) provides an additional example, with the verb /ʔilk–/ 'sing'.[4] In all pairs of Yawelmani words related by the presence or absence of some vowel, the vowel [i] is always the relevant one.

(1.19) Words related by the presence or absence of [i]

	attested word	*not...*	*gloss*
a.	logiwhin	*logwhin, *loghin	'pulverized'
	logwen	*logiwen	'will pulverize'
b.	ʔilikhin	*ʔilkhin, *ʔilhin	'sang'
	ʔilken	*ʔiliken	'will sing'

Once we accept that roots like *logw–* 'pulverize' and *ʔilk–* 'sing' end with two consonants, we must explain why the vowel [i] is inserted in certain forms but not in others. The short answer is that the vowel is added in order to allow an extra consonant to be syllabified (1.17d). This prevents the loss of a consonant, shown (1.17c) as well as the ill–formed syllables, as in (1.17a,b).

The challenge when addressing such facts is to formally characterize the relation between the input *logw–hin* and the output *logiwhin* in a manner that expresses both universals and variation. The Optimality Theory response is that such patterns are stated in terms of constraint satisfaction and violation, where the constraints themselves express the universals and the particular constraint rankings express variation.

---

[4]The symbol [ʔ] is a glottal stop. This sound is produced by some speakers of English between the two words in "a apple" (not "*an* apple") and in some dialects of English in place of the "t" in words like *bottle*.

The option preferred in Yawelmani is to insert a vowel, thereby saving all consonants and avoiding complex syllable margins (onsets and codas). The first two forms in (1.20) have retained all consonants by putting two at the edge of one of the syllables, either [gw] closing the first syllable or [wh] starting the final syllable, fatally violating *COMPLEX. The third form in (1.20) is completely faithful in terms of segments, but has syllabified a consonant as the peak of a syllable, thereby violating the high–ranked PEAK. The fourth form in (1.20), has lost one of the three consonants and so violates FAITHC, also a fatal violation due to its high ranking.

(1.20) Yawelmani: *COMPLEX, FAITHC, PEAK » FAITHV

/logw–hin/	*COMPLEX	FAITHC	PEAK	FAITHV
logw.hin	*!			
log.whin	*!			
log.w.hin			*!	
log . hin		*!		
☞ lo.giw.hin				*

The final form in (1.20) includes an extra vowel, which allows syllabification of the third consonant, but which violates FAITHV. By hypothesis, though, FAITHV is the lowest ranked of these constraints in Yawelmani: for the input /logw–hin/, [logiwhin] is selected as the optimal output form.

Three other rankings remain to be considered, those with each of FAITHC, *COMPLEX, and PEAK as the subordinate constraint. Although a formal possibility with constraint ranking is that these four constraints are crucially ranked with each other (giving 4!, or 24 possible rankings) we need not consider all of the logically possible rankings. Constraint rankings are crucial only when they decide between competing candidates, but the data and constraints under consideration here do not have that level of complexity. (On this point, see the discussion of syllable typology in Chapter 2.)

FAITHC is Violated in Spanish

The next example we consider illustrates the "lose–a–consonant" option. In such cases, FAITHC is subordinate: in a three consonant sequence, the best thing to do is to leave a consonant out. In this way, PEAK, FAITHV and *COMPLEX are satisfied, at the cost of violating FAITHC. An example is found in Spanish. A caveat is in order here. Even a small amount of familiarity with Spanish will reveal that the facts presented here are simplified somewhat. In particular, I ignore the well–known fact that Spanish inserts a vowel in front of *sC* clusters (e.g. *esfera* 'sphere'; compare *hemisferio* 'hemisphere', not **hemiesferio*.)

The first column in (1.21) shows Spanish verbs in the infinitive form. Each of the verb roots ends with two consonants when followed by an infinitive suffix, *–er* or *–ir*. for example, in (1.21a) the verb root ends with the two consonant sequence *–rb–*: *absorber*, and in (1.21b) it ends with the two consonant sequence *–lp–*: *esculpir*.

(1.21) Spanish data

	infinitive	*adjective or noun*	
a.	absorb–er	absor–to	*absorbto
b.	esculp–ir	escul–tor	*esculptor
c.	distingu–ir	distin–to	*distingto

However, as the second column shows, where an adjective–forming suffix *–to* is added, the verb root appears to end with a single consonant. For instance, in (1.21a) the verb root ends with the single consonant *–r–*: *absorto*. What happened to the *–b–*? We can ask similar questions for (1.21b,c): what happened to the *–p–* and the *–g–*?[5] The answer is clear if we focus on the phonological shape of the suffixes, rather than their morphological function. The infinitive suffix is vowel initial. As such, the final consonant of the verb root (*–b–*, *–p–*, *–g–*) is syllabified as the onset for that vowel's syllable (1.22a). By contrast, the noun– and adjective–forming suffixes begin with consonants. The final consonant in the verb root cannot syllabify without creating a CC onset or a CC coda, thereby violating *COMPLEX, (1.22c).

(1.22) Syllabification of *absorber* and *absorto*

a.	ab.sor.ber	b.	ab.sor.to	c.	*ab.sorb.to	*ab.sor.bto
	es.cul.pir		es.cul.tor		*es.culp.tor	*es.cul.ptor
	dis.tin.[gi]r		dis.tin.to		*dis.ting.to	*dis.tin.gto

Spanish apparently does not allow *COMPLEX violations. In this way, Spanish is like Yawelmani. Spanish does not allow a consonant to syllabify by itself, giving *[ab.sor.b.to], again like Yawelmani. Nor does it adopt the Yawelmani option, of inserting a vowel, resulting in *[ab.so.reb.to] or *[ab.sor.be.to].

Significantly, the two languages differ in their resolutions to the "extra consonant" problem. As already seen, in Yawelmani the added vowel [i] rescues the unsyllabifiable extra consonant. In Spanish, however, the unsyllabifiable consonant is not rescued: it is simply deleted.

The tableau in (1.23) shows that ranking FAITHV, PEAK, and *COMPLEX above FAITHC results in consonant deletion. The first form in (1.23), which surfaces as [absorto] since the *–b–* is not syllabified, is the correct form. All the input vowels surface and no vowels are inserted, satisfying FAITHV. There are no CC onsets or codas, satisfying *COMPLEX. There are no syllables composed solely of consonants, satisfying PEAK. These results are achieved at cost, however: the root–final consonant *–b–* is not syllabified, incurring a violation of FAITHC. Due to ranking the other constraints above FAITHC, however, the FAITHC violation is not fatal.

This contrasts with the results in the second, third and fourth forms in (1.23), each of which incurs a fatal violation of some constraint ranked above FAITHC. In the second candidate in (1.23), FAITHC is satisfied because the *–b–* is syllabified. However, satisfaction of FAITHC comes at a perilous cost, the violation of *COMPLEX. Similarly, in the third candidate in (1.23), the *–b–* is again syllabified, this time as the onset to the syllable

[5]In Spanish orthography, a *u* follows the *g* in *distinguir* to indicate that the *g* is "hard", that is, pronounced like the *g* in English *get* and *Abigail*.

of an inserted vowel [absorbeto].[6] Satisfaction of FAITHC again comes at cost, here a violation of FAITHV due to the inserted vowel. Finally, in the fourth candidate in (1.23) satisfaction of FAITHC is attained through violation of PEAK. By ranking FAITHC below the other constraints, the correct form is selected.

(1.23) **Spanish: FAITHV, PEAK, *COMPLEX » FAITHC**

	/absorb–to/	FAITHV	PEAK	*COMPLEX	FAITHC
☞	ab.sor . to				*
	ab.sorb.to			*!	
	ab.sor.be.to	*!			
	ab.sor.b.to		*!		

In both Yawelmani and Spanish, *COMPLEX and PEAK are high–ranked. The languages differ in which FAITHFULNESS constraint is more critical, FAITHV or FAITHC. In Spanish, FAITHV is more important: no vowel may be inserted to rescue the extra consonant. The consonant is simply lost. In Yawelmani, FAITHC is more important: violating FAITHV is countenanced as long as the effect of the violation protects a consonant from deletion.

We now turn to an example of a third type of language, one in which FAITHFULNESS is more important than keeping syllables simple. Our example is English.

**COMPLEX is Violated in English*

In English, we add a suffix *–ness* to adjectives in order to create nouns: *happy, happiness*; *sad, sadness*; etc. Since *–ness* begins with a consonant, the critical environment is found when *–ness* is added to an adjective that ends in two consonants, such as *limp*. Rather than *limness*, with consonant loss, or *limp[i]ness,* with vowel addition, English creates a complex coda, resulting in *limpness*. In fast and/or casual speech, people may omit the *t* in the cluster *...ftn...* and say *so*[fn]*ess* instead of *so*[ftn]*ess*. How best to account for this has yet to be resolved satisfactorily. One possibility within OT is variable ranking of constraints, or variable values in constraints depending on speech rate/style. Another possibility, discussed in Chapter 2, is the use of correspondence constraints between careful and casual speech representations.

The forms in the noun column are formed by simply adding *–ness* to the adjectives, and syllabifying, without deleting or adding anything, exactly the pattern expected if other constraints outrank *COMPLEX. This is illustrated by the tableau in (1.25). The correct form, *limpness* is the first candidate in (1.25). There are no violations to PEAK, FAITHV or FAITHC; the sole violation is to the subordinate *COMPLEX. (In fact, the very existence of words like *limp*, *soft*, and *crisp* attest to the low ranking of *COMPLEX in English.)

[6]The alert reader will have noticed that I have changed the vowel which is inserted: in the Yawelmani example, the vowel [i] is inserted while in Spanish the vowel is [e]. In the next example, from English, the inserted vowel is [i]. Why are the vowels different? The answer is that this is yet another type of variation found between languages. Vowel qualities of inserted vowels vary across languages. The vowel [i] is the vowel that is inserted in the Yawelmani example, while in Spanish the vowel [e] is sometimes inserted (as in the *esfera* 'sphere' example), although it is not inserted in the environment under discussion. In English, the vowel [i] is inserted in certain contexts: when the past tense or plural suffix adds a syllable to the word as in *kisses* [kɪsis] and *hinted* [hɪntid]. (See Chapter 4.) Under Optimality Theory, the quality of the vowel is also determined by constraint ranking, a complexity that is ignored in our discussion.

(1.24) **English data**

	adjective	*noun*		
a.	limp	limpness	*lim_ness	*limp[i]ness
b.	soft	softness	*sof_ness	*soft[i]ness
c.	crisp	crispness	*cris_ness	*crisp[i]ness
d.	strange	strangeness	*stran_ness	*strang[i]ness

(1.25) **English: FAITHV, FAITHC, PEAK » *COMPLEX**

	/lɪmp–nɛs/	FAITHV	PEAK	FAITHC	*COMPLEX
☞	lɪmp.nɛs				*
	lɪm.nɛs			*!	
	lɪm.pi.nɛs	*!			
	lɪm.p.nɛs		*!		

Each of the failing candidates violates one of the other three constraints. The adjective–final consonant *–p–* is lost in the second candidate in (1.25), incurring a FAITHC violation in order to achieve universally better syllables. Similarly, the added vowel in the third candidate in (1.25) results in universally better syllables, but only at the cost of a FAITHV violation. Finally, the fourth candidate in (1.25) shows the extra consonant syllabified by itself, incurring a PEAK violation. The English pattern is characterized by ranking *COMPLEX below the other three constraints, the third of the four rankings we consider.

PEAK is Outranked in Berber

A striking fact about Berber is the long sequences of consonants that surface. The Berber data is taken from Dell & Elmedlaoui (1985), a discussion of the Imdlawn Tashlhiyt dialect. See Prince & Smolensky (1993) for a complete reanalysis of the data in terms of Optimality Theory.

A striking fact about Berber words is that they do not even need to contain vowels: *trglt* 'lock', *txdmt* 'gather wood', *trkst* 'hide', all in the second person singular perfective (e.g. 'you have locked'). The question, of course, is how such sequences are arranged into syllables. Berber accomplishes this via PEAK violations.

The chart in (1.26) contrasts two morphological categories, the third person masculine singular which consists of a vowel prefix *i–* and the third person feminine singular, a consonant prefix *t–*. When *i–* is added to a verb which starts with two consonants, nothing surprising happens: the first consonant closes the first syllable while the second is the onset for the second syllable, as in *il.di* 'pull'. The surprise occurs when the consonantal prefix *t–* is added: in this case, the verb's initial consonant syllabifies as the peak of a syllable: *tl.di* 'pull'. (A comparable situation is found in English, although it is perhaps not nearly so dramatic as the Berber pattern. For example, in *cylinder*, the final consonant, *r* serves as the peak of the final syllable while it is clearly in the onset of its syllable in the related word *cylindrical*.)

The pattern is exactly the one we expect if PEAK is outranked by the other constraints. In this event, the best solution to the problem of "too many consonants, not enough vowels" is to allow one of the consonants to assume the syllabic position normally reserved for vowels.

(1.26) Berber data

	3 m sg	*3 f sg*	*gloss*
a.	iz.di	tz.di	'put together'
b.	if.si	tf.si	'untie'
c.	is.ti	ts.ti	'select'

As illustrated by (1.27), each of the alternative syllabifications results in the violation of a higher ranked constraint. Creating a single syllable with the prefix violates *COMPLEX (the first candidate in (1.27)); deleting consonants violates FAITHC (the second, third and fourth candidates in (1.27)); and adding a vowel violates FAITHV (the fifth candidate in (1.27)).

(1.27) Berber: FAITHV, FAITHC, *COMPLEX » PEAK

/t–fsi/	FAITHV	*COMPLEX	FAITHC	PEAK
t.fsi		*!		*
_ .si			*!*	
_f.si			*!*	*
t_.si			*!*	*
tif.si	*!			
☞ tf.si				*

Berber resolves the extra consonant problem by allowing consonants to syllabify as syllable peaks, a position that most languages reserve for vowels.

Summary

In this section, we explored the "extra consonant" problem: syllabification of a sequence of three consonants *must* violate one of four different constraints. Different rankings of the four constraints predict four different resolutions to the problem, depending on which constraint is lowest ranked. The syllabification patterns of four languages, Yawelmani, Spanish, English, and Berber, are exactly the four patterns predicted by the model.

6 Conclusion

This chapter began with a very brief introduction to linguistics, and a slightly longer introduction to phonology (Sections 1 and 2 respectively). We then explored the structure

of Optimality Theory. Section 3 sketched an analysis of Yawelmani syllabification within OT and introduced the "tableau", an expository device used to demonstrate the effect of EVAL.

Section 4 provided a more formal discussion of the components of OT, including the input, GEN, EVAL, CON, and the output. A schematic summary picture is given below of the functions in a grammar under OT. The subscript "L" on the function EVAL indicates that EVAL is a *language–particular* ranking of CON, the universal set of constraints.

(1.28) Optimality Theory (Prince & Smolensky 1993)

GEN (Input_k) → {Candidate_1, Candidate_2, Candidate_3,...}

EVAL_L{Candidate_1, Candidate_2, Candidate_3,...} → Output_k

We then considered how Optimality Theory accounts for the central issues in linguistics: *universals, markedness*, patterns and *variation*. Universals are represented in OT by CON, the universal set of constraints. Markedness is represented in OT by constraint violation while constraint satisfaction corresponds to unmarked properties. Patterns are the result of the interplay between a particular constraint hierarchy and the inputs provided by the language. Variation results from differences in the constraint rankings selected by specific languages, illustrated in Section 5. Each language deals with ...CCC... sequences in a different way, and each way is characterized by a distinct ranking of the relevant constraints.

In addition to knowing how OT works and how OT accounts for the central linguistic issues, it is also useful and interesting to examine some other aspects of the model. Why has the model become so popular so fast? What areas of study are particularly suited to OT analysis? Does the model change what we thought we knew about language? What challenges does OT face?

In this closing section, I explore why OT caught on so rapidly, to the point that six years after its inception, it is the dominant paradigm in formal phonology, and is rapidly gaining ground in both morphological and syntactic analysis. I then consider two types of issues that remain to be explored, those concerned with the nature of the theory itself, and those addressing different empirical domains that may be amenable to OT analysis.

The Rise of OT

To understand the rapid and widespread acceptance of Optimality Theory, one must understand the state of theoretical research in linguistics in the late 1980s. In many ways, it was foundering. Consider phonology. Great advances in our understanding of representations were made throughout the late 1970s and continuing into the early 1980s, resulting in the nonlinear representations that are now widely assumed. There was great hope that as our understanding of representations improved, the characterization of alternations would be simplified. This simplification did not happen.

Efforts were also directed specifically at formally restricting the possible types of alternations. Efforts in this domain, too, were unsuccessful: the alternations permitted by every formal model unfortunately also include alternations that are both unattested and thought to be unlikely. There were always counterexamples.

Finally, constraints were being used — or rather, over–used. The standard of the underlying–surface relation included (i) the abstract underlying form of the morphemes of a language (ii) which concatenate and then (iii) undergo a series of rules. When no further rules apply, (iv) the surface form has been attained. (See also Chapters 3 and 4.)

(1.29) **Generative Phonology**

underlying representation
⇩
morpheme concatenation
⇩
rules
⇩
surface representation

This picture looks neat and tidy — until the role of constraints is added. Constraints entered this picture at all stages. Constraints hold of the underlying forms, in terms of what sounds are permitted and what sequences of sounds are licit. Constraints hold of morpheme concatenation, restricting how morphemes may combine. Constraints hold of rule application, limiting both how rules can apply and what types of sounds or sound sequences can be produced. Constraints hold of outputs, prohibiting patterns that do not occur at the surface. Unlike in OT, all of these types of constraints have been viewed as *inviolable* within the relevant domain.

(1.30) **Phonology in the 1970s and 1980s**

constraints hold here ⇨	underlying representation
	⇩
constraints hold here ⇨	morpheme concatenation
	⇩
constraints hold here ⇨	rules
	⇩
constraints hold here ⇨	surface representation

The frustrations in syntax bear some similarities, but also some differences. Syntactic representations, too, have evolved into increasingly elaborate hierarchical structures which in turn have required the increasing use of empty terminal nodes. Inviolable constraints, generally called **conditions** or **principles**, have played a more dominant role in syntax than in phonology however, for syntacticians to a greater extent than phonologists have attempted to minimize the rule component.

The "inviolable" principles of syntax have themselves proved to be problematic in that inviolability has been purchased at the cost of a variety of types of hedges. As shown in detail in Chapter 6, some principles are "parameterized", holding in one way in one language and in another way in another language. Other principles have peculiar restrictions built–in. For example, the Extended Projection Principle begins with the strong claim "All clauses must have a subject", such as *John* in *John ran* and *it* in *It rained all night.* However, this principle is weakened by the codicil "*unless the language lacks expletives*", in order to account for subjectless sentences in languages like Yaqui

such as *Yooko yukne* 'It will rain tomorrow' (literally 'Will rain tomorrow'; see Chapter 6, Section 2 for more discussion of this point.)

In both areas, research results have indicated that the general analytic strategy has been on the right track; at the same time, there had been growing dissatisfaction in two ways. First, despite continued innovations in theories of rules and of representations, certain types of data remained unexplained. Second, the prevailing belief about constraints — that they are inviolable — resulted in a continuing frustration with their role in grammar, for it is exceedingly difficult to find a constraint that is never violated.

Optimality Theory redefines the role of constraints and in so doing redefines the research focus. All constraints are violable. Grammars define the relative significance of violating specific constraints. Constraints are present only in the constraint hierarchy: there are no separate constraints on inputs nor on outputs. There are two powerful implications for linguistic analysis here. First, there simply is no rule component at all. Second, the constraint hierarchy must be constructed to return some result regardless of the input (the result may be a **null parse**, that is nothing at all). Examining the nature of underlying representations and of rules has been core to linguistic research: OT changes the focus of this research.

A desirable aspect of this change is that research focuses directly on universal properties of language: since the constraints are hypothesized to be universal, OT redirects our research focus towards language universals. This aspect of grammar must be central to any OT analysis, regardless of how language–specific the phenomenon is. OT has not yet been able to answer all the unanswered questions; however, it has provided a dramatically different approach to accounting for both universals and variability and to the input–output relationship.

Optimality Theory addressed these problematic issues.

1. **It defines a clear and limited role for constraints.**
   a. Each constraint is universal.
   b. Constraints are ranked in EVAL.
2. **It eliminates the rule component entirely.**
   Different constraint rankings in EVAL express language variability.
3. **It focuses research directly on language universals.**
   Each constraint is universal.
4. **It resolves the "nonuniversality of universals" problem.**
   Universals don't play the same role in every language.

Remaining Issues

Issues abound in each of the components of the OT model. OT challenges the way in which we think about linguistic representations and relations such that virtually every aspect of previously held assumptions must be reconsidered.

The input. Linguists are only beginning to explore the nature of the input under OT. There are four classes of problems here. First is the issue of the lack of constraints on the input, an aspect of the theory known as Richness of the Base. In standard generative phonology, numerous constraints were imposed on the input. For example, an analysis would frequently begin by defining the vowel and consonant inventories of a language in terms of permitted feature combinations. See Chapter 3 for the analysis of inventories in OT. Under OT, such constraints on the sounds that make up the input are impossible. Inputs are potentially as infinite as the candidate set; the constraints in EVAL must be ranked so that impossible sounds or sound sequences never surface.

Since there are no constraints on the input, it is easy to construct multiple inputs that converge on a single output. Which of the multiple inputs is the best one? A variety of strategies are imaginable. For example, an algorithm called **lexicon optimization** is introduced in Prince & Smolensky (1993) and developed in Itô, Mester, & Padgett (1995). Lexicon optimization examines the constraint violations incurred by the winning output candidate corresponding to each competing input. The input–output pair which incurs the fewest violations is considered the optimal pair, thereby identifying an input from the output.

Second is the issue of what exactly goes into an input? Most people assume that some kind of phonological representation is the input for each morpheme. One challenge is how exceptional phonological properties are to be expressed, since the standard model is that EVAL will select the form that least violates the constraints, thereby normalizing at least some types of aberrant patterns. One possible approach to representing irregularities is to include constraint violations as part of the input representation, to indicate which constraints the input fails to satisfy. (This approach is taken to the limit in Golston 1996, who argues that inputs consist solely of the relevant constraint violations.)

In syntax the "content of the input" question is more puzzling. One possibility is that the input is extremely enriched, containing words, argument structure, and indications of which word is the subject, the object, etc. An even more enriched input would include some degree of syntactic structure assigned to the sequences of words. Inputs might also be impoverished. For instance, the input might include words but not their order or grammatical relations. In this case, a single input will correspond to a variety of outputs with different meanings (e.g. input {*dog*, *man*, *bites*} would correspond to both *Dog bites man* and *Man bites dog*. The view taken in both Chapters 5 and 6 is that the input consists of an ordered sequence of words, but not their grammatical structure. The extreme along these lines would be that the input contains no words at all; that words are inserted by GEN or even after EVAL. Under the latter view, EVAL's role is simply to determine which syntactic structures are well–formed, not whether specific instantiations of those types are well–formed.

Third is the issue of faithfulness between the input and the output. In this chapter, for example, we conflated two aspects of faithfulness under FAITHV and FAITHC. These constraints prevented both the *addition* and the *removal* of elements. However, in some languages these two sides of faithfulness may be ranked independently of each other for the same class of elements: DON'T DELETE: "input elements are in the output" vs. DON'T ADD: "output elements are in the input". Furthermore, faithfulness is a relation found not only between input and output, but also between other pairs, such as the **base** and the **reduplicant**, a point illustrated at some length in Chapter 4. The current most prevalent view of faithfulness is **Correspondence Theory**, laid out in McCarthy & Prince (1995) and McCarthy (1995).

Finally, there is the question of whether there is an input at all. Some works have argued that instead of input representations, **morphemes** are best expressed as constraints

themselves (Hammond 1995, Russell 1995). As such, they may be ranked with respect to other (nonmorphemic) constraints. An advantage of this approach is the ease with which exceptional behavior is expressed; a disadvantage is that it does not obviously extend to include forms, like sentences, that are larger than the morpheme. The nature of the input is discussed in greater detail in the Afterword .

GEN. GEN's function is to produce a candidate set for every input. There are two aspects of GEN that raise concern. First, in the purely formal model, for every input, an infinite candidate set is generated. Although this does not raise serious problems for formal research, it does hamper efforts to explore psycholinguistic and computational models of language, since neither responds happily to infinite sets. The second problem area is understanding the types of manipulations that GEN can make. It is widely assumed that GEN can only create universally well–formed linguistic objects, that is ones which do not violate any universally inviolable constraints. This assumption requires that we distinguish between universally inviolable constraints and those which are violated, even if only rarely.

CON. The central issues involving the constraint set CON revolve around the question "what are the constraints on the constraints?" Proposals about CON include the idea that certain constraints contain variables which are filled on a language particular basis. (For example, the ALIGN constraints match edges of a pair of elements, where the pair is named in each constraint. See McCarthy & Prince 1993b on Alignment.) Another proposal is that constraints may be conjoined, for instance by logical operators, to make more complex constraints. Yet another challenge, already discussed, is to establish that each proposed constraint is a universal. Chapter 4 raises questions about the universality of each constraint.

EVAL. EVAL evaluates candidates in terms of a particular ranking of constraints, so better understanding of EVAL involves a better understanding of constraint ranking. Typically an analysis will include two or more constraints whose ranking is irrelevant, yet OT assumes that all constraints are ranked with respect to each other. In some work it has been argued that constraints are **tied**, which may allow more than one candidate to be selected as optimal. See Chapter 6 for an example of tied constraints. Another issue involving constraint ranking is "inherent" ranking, where the substantive properties of the constraints themselves determining their ranking with respect to certain other constraint. For example, constraints governing the syllabic positions of elements with different degrees of sonority have been proposed as an instance of inherently ranked constraints (Prince & Smolensky 1993).

The output. The central concern with the output is "what happens next?" In the standard generativist view, the output of one component serves as the input to the next. The metaphor assumes a modular picture of language, where the output of one module serves as the input to others. Is this the best metaphor under OT? For example, the discussion above has focused solely on phonological properties of the string. Phonetic properties might be analyzed through a separate constraint hierarchy, but they might also be analyzed through constraints that intermingle with the phonological constraints. Work by Donca Steriade and her students at the University of California in Los Angeles explores this view.

The modular nature of language. A widely held assumption about linguistic representations and relations is modularity (Fodor 1983). The basic idea of modularity is that the principles responsible for different aspects of an utterance are themselves structured differently. This view is greatly exaggerated under OT, since each constraint can be viewed as an independent entity with its own internal structure.[7] Concerns about modularity arise in another way, too. A logical extension of the OT model for language is that there is a single constraint hierarchy, which internally ranks all constraints, whether syntactic, morphological, phonological, phonetic, or semantic. This possibility predicts interaction between components (**modules**). For example, particular syntactic constraints might be violated in order to satisfy a phonological or morphological constraint, or vice versa. This contrasts sharply with the view of grammar as having a separate and independent syntactic component, phonological component, etc.

Extensions

Within linguistics, some of the most interesting research areas opened up by OT are the interface areas, just mentioned. OT allows the possibility of a single constraint hierarchy, with constraints of all types potentially mingled together. In particular, constraints from different components may be crucially ranked with respect to each other. This possibility provides a new framework for exploring the interfaces between components of the grammar, for example morphology and phonology, syntax and morphology, phonology and phonetics.[8]

There are numerous other domains that may be fruitfully explored using OT, beyond simply the characterization of core language phenomena. Some of these are sketched below.

Poetics. An intriguing domain in which linguistic studies have long played a significant role is the exploration of what is or is not significant to a particular type of poetry. For example, the prototypical relation between poetic meter and spoken rhythm is that strong matches strong and weak matches weak. Studying the matches and mismatches between the strong or weak positions in poetic meter and the strong or weak positions in a nonpoetic rendition of a line of poetry leads to a very precise characterization of the poet's "voice". Recent work by Bruce Hayes and his students at the University of California at Los Angeles suggests that OT offers exciting new insights into the relation between word stress and metrical structure of English folk verse.

Behavior of borrowed words. When words are borrowed from one language to another whose sound patterns are different, the word typically is modified. For example, as noted in our discussion of Hawaiian, that language adapts words borrowed from English, a language which violates Hawaiian syllable constraints, by adding extra vowels: *weleweka* 'velvet'. The expectation under OT is that the borrowing language's constraint hierarchy will evaluate the candidates produced by GEN, taking as input the source language's output of the borrowed word. Exploration of how well this hypothesis works, and where it fails, may lead to significant new insights into what happens when words are borrowed.

[7]There are also "families" of constraints, such as the FAITHFULNESS family, for example FAITHV and FAITHC. These constraints all have the same structure and refer to the same type of element. A more precise statement, then, is that each constraint *family* can be viewed as an independent entity with its own internal structure. See especially Chapter 3 for discussion of constraint families.

[8]Works in these areas can be found in the Rutgers Optimality Archive; see the Foreword.

Second language acquisition. When an adult learns a second language, typically the second language is spoken with some degree of "accent". Understanding the nature of that accent is complex, depending on a multitude of variables (such as familiarity with the language) which are difficult to measure. OT provides one guide to identifying patterns we might expect in specific accents, by identifying the constraint rankings of the native language and the second language.

The empirical problem of (first) language acquisition. Under OT, part of acquiring a language is acquiring the critical constraint rankings of that language. Since constraints interact, it is reasonable to assume that evidence for a particular ranking of constraints is not always noticed by the learner, so some constraints are ranked incorrectly, to be reranked when further data is available. This predicts specific stages that a child might go through, each of which would reflect the incorrect dominance of some universal constraint. This prediction is quite different from that of a rule–based model, in which a child might incorrectly learn a language–particular rule, which in itself may have little claim to universality.

The logical problem of language acquisition. This point refers to the challenge of understanding how a child might acquire language under a specific formal model of language. It answers the question of whether a grammar is *in principle* learnable, rather than addressing the issue of how a language is acquired. There is already a small but growing body of work in this area indicating that OT does provide a learnable model of grammars, in particular by Paul Smolensky at Johns Hopkins University and his student Bruce Tesar at the University of Colorado at Boulder, and by Douglas Pulleyblank and his student William Terkel, both at the University of British Columbia.

Language change. Under OT, the formal characterization of language change through time is that constraints are reranked. A prevalent view of diachronic language change is that change occurs when there is imperfect transmission from one generation to the next. Combining these two claims implies that constraints can only be reranked when the evidence for a particular ranking is not very robust. Thus, OT makes clear predictions both about the effects of change and about the type of change that might occur.

Natural language perception. As noted earlier in this section, OT works both from the input to the output and from the output to the input. Under OT it is possible to examine an output and determine the optimal input, something that is not possible under rule–based views of language. For language perception, this is an exciting result, for OT offers a formal theory of language which is able to use outputs to access inputs, crucial to any complete model of language perception.

Natural language production. The standard generative model of phonology, in which an input is manipulated by a series of rules to produce an output, is not readily translated to a model of natural language production. There are two types of challenges, and under OT, both of these problems are solved. First, the formal device of a series of rule applications does not carry over easily into a model of production. Under OT, the input–output relation is mediated in one step, by EVAL, not in multiple steps by a series of rules: this issue, then, does not even arise. Second, the types of rules necessary under such a model include ones which can operate from the end of the word towards the beginning of the word, yet evidence shows that planning for word production (as well as the actual articulation of a word) starts at the beginning of the word, not the end. Under

OT, such apparently directional patterns are determined by inspection of the input and output representations, not as the result of specific operations on representations.

Computational modeling of language. Except for the problem of the infinite candidate set, OT is particularly conducive to computational modeling. Already there are a variety of efforts at developing computational models of specific aspects of OT grammars; OT work in language learnability also relies heavily on computational modeling. This contrasts sharply with other models of phonology, in which each rule of a language expresses an idiosyncratic property of that language, making computational modeling highly idiosyncratic as well.

In closing, I have tried to show that OT not only is gaining wide acceptance among formal linguists, but also that it has extensions into numerous related domains, some of which have already proven to be fruitful. For many, the extensions are as exciting as the successes in formally accounting for core language phenomena.