Table of Contents

Notes on Contributors

Diana Archangeli (MA, University of Texas; PhD, MIT) is Professor of Linguistics at the University of Arizona. She taught previously at the University of Illinois, Champaign–Urbana. She is the coauthor with Douglas Pulleyblank of *Grounded Phonology* and is now engaged in a number of projects involving Optimality Theory.

Michael Hammond (PhD, UCLA) is Associate Professor of Linguistics at the University of Arizona. He has also taught at the University of Minnesota and at the University of Wisconsin, Milwaukee. He has published extensively on phonological theory and metrical analysis.

D. Terence Langendoen (PhD, MIT) is Professor and Head of the Department of Linguistics at the University of Arizona. He taught previously at Ohio State University, Brooklyn College, and the Graduate Center of the City University of New York. He has published on a wide variety of linguistics topics.

David Pesetsky (PhD, MIT) is Professor of Linguistics at the MIT. He taught previously at the University of Massachusetts at Amherst, and is well known for his wide–ranging research in syntax.

Douglas Pulleyblank (PhD, MIT) is Associate Professor of Linguistics at the University of British Columbia. He taught previously at the University of Southern California and the University of Ottawa. He is known particularly for research on the phonology of African languages.

Kevin Russell (PhD, University of Southern California) is Assistant Professor of Linguistics at the University of Manitoba. He is currently working on the phonology and morphology of Arabic and Cree.

Margaret Speas (MA, University of Arizona; PhD, MIT) is Associate Professor of Linguistics at the University of Massachusetts, Amherst. She has written extensively on problems of phrase–structure analysis and the analysis of pronominals.

Foreword

The goal in creating this volume has been to offer an accessible introduction to Optimality Theory, a powerful new model of grammar. Our intended audience is anyone with a serious interest in language who desires to understand this model, regardless of their background in formal linguistic theory itself.

What is a grammar and how does it work?

People who know a language are able to produce and recognize a huge number of intricately structured expressions (words, phrases, sentences, etc.). Moreover, they are able to distinguish those expressions which belong to a particular language from possibly very similar expressions which do not. Linguists, the scientists who study language, have assumed that these abilities are accounted for by a mechanism, called a **grammar**, which relates the expressions of a language to the elementary parts of which they are made.

Linguists are thus faced with two related problems. One is to ensure that the grammar of a particular language is able to encompass all of the expressions that can reasonably be supposed to belong to that language. The other is to ensure that the grammar is able to distinguish those expressions which belong to the language from those which do not.

The problem can be compared to that of a fisherman trying to catch in a net all the fish of certain types in a certain area, but nothing else (no other types of fish, no other creatures, etc.). The ideal net would be large and fine enough to gather all the desired fish (the desirables), and be designed to allow the undesired fish and other creatures (the undesirables) to escape. But it may not be possible to construct such a net. Any net which is large and fine enough to catch all the desirables may of necessity also catch some undesirables.

If that is the case, one would need a device (a separator) to remove the undesirables once the catch has been taken, no matter how effective the net is in allowing the undesirables to escape. One might therefore decide to put one's energies more into designing an effective separator than into refining the capabilities of the net to allow the undesirables to escape. The ideal separator is one which always succeeds in removing the undesirables, no matter how many the net retains. If one could design an ideal separator,

then one might be content with a net which catches everything in the area, allowing nothing to escape, leaving the job of removing the undesirables entirely to the separator.

The ideal net corresponds to the original idea of a generative grammar (as in Chomsky 1957) that accounts directly for (i.e. generates) *all and only all* the expressions of a given language with no auxiliary devices to remove ungrammatical expressions. Because of the enormous complexity of the grammar which results from trying to put that idea into practice, many linguists chose to drop the *only all* proviso for the generative mechanism (the technical description of this state of affairs is that the grammar **overgenerates**), and to add devices, called **filters**, to eliminate the ungrammatical expressions that the generative mechanism allows; see Chomsky and Lasnik (1977) for a proposal along these lines.

The resulting theory divides the task of separating the grammatical from the ungrammatical sentences to two parts of the grammar: the generative component, which accounts for all the grammatical expressions, allows some ungrammatical expressions, and rejects others (i.e. the net); and the filtering component, which removes all the ungrammatical expressions let in by the generative component (i.e. the separator).

This situation, in turn, has been viewed as unsatisfactory: why have two components of the grammar responsible for separating out the expressions which are ungrammatical in a particular language? Chomsky (1995b:223) states this view as follows:

> The worst possible case is that devices of both types are required: both computational [generative] processes that map symbolic representations to others and output conditions [filters].

In phonological research in the late 1980s and early 1990s, analyses including both generative processes and filters were prevalent. Moreover, in many cases, the same facts might be covered by process or by filter, with no empirical consequences. Optimality Theory was introduced in response to this situation. Optimality Theory opts for the 'ideal separator': a very simple generative mechanism (GEN; see Chapter 1) that allows ungrammatical expressions to be created essentially without restriction, leaving all the work of separating out the ungrammatical ones to filtering devices (EVAL; also see Chapter 1). Because the need was so apparent in phonology, the Optimality Theoretic model has rapidly gained the attention of phonologists worldwide.

In syntactic research, Optimality Theory again is the ideal separator. But the research climate is less receptive to such a model: in general, syntactic analyses have not made rampant use of both processes and filters. For example, Chomsky's Minimalist Program (see Chapter 6) represents a return to the idea of the 'ideal net': a generative mechanism that allows the ungrammatical expressions to escape, permitting only the grammatical ones to be accounted for. Consequently, the Minimalist Program and Optimality Theory can be seen as attempts to avoid the worst-case scenario in opposite ways.

An overview of the book

This book is organized to present an introduction to Optimality Theory, and to demonstrate its workings in phonology, morphology, and syntax. Chapter 1, by Diana Archangeli, first summarizes the goals of formal linguistic research, then introduces

Optimality Theory, showing how it addresses these goals. The reader who has little or no understanding of Optimality Theory would do well to start with this chapter. It serves as a preface to the remaining chapters, since the concepts it introduces are assumed in each of the other chapters. The reader who is already familiar with the basics of Optimality Theory might prefer to go directly to one of the following specialized chapters: Chapters 2 and 3 on phonology, Chapter 4 on morphological issues, and Chapters 5 and 6 on syntax. The book concludes with an Afterword, concerning the nature of the input. A summary of Chapters 2 through 6 follows, including comments on which chapters serve as background for subsequent chapters.

Chapter 2, by Michael Hammond, provides an introduction to syllables and feet, the two central constituents in discussions of prosody. The chapter illustrates how Optimality Theory accounts for a variety of prosodic phenomena. It also provides an excursus into psycholinguistics, with a discussion of how some surprising patterns of speech perception are explained under Optimality Theory, patterns which constitute a serious challenge to derivational models of language. This chapter is particularly useful for the non–phonologist because virtually all of the examples are from English. This chapter relies heavily on the analysis provided in Chapter 1, as well as making use of the theoretical points introduced there; it is also useful to the understanding of Chapter 4, which is about morphology.

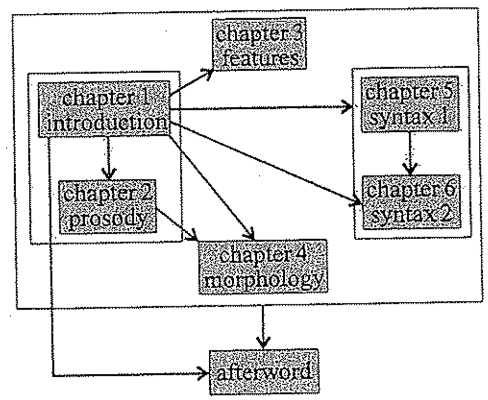
Chapter 3, Douglas Pulleyblank's chapter on phonological features, explains the concept of phonological features and illustrates a variety of feature patterns found in different languages. The cross-linguistic sketch of how different languages resolve nasal-obstruent sequences (e.g. *nt*, *ms*, *nb*, etc.) illuminates one of the main advantages of Optimality Theory, its ability to precisely characterize formal differences between languages. The chapter also addresses the issue of how a "segmental inventory" is expressed within a model which allows for no restrictions on the inventory of segments in underlying representation.

Kevin Russell introduces key questions in the study of word formation, or morphology, in Chapter 4. This chapter focuses primarily on the phonological, or pronunciation, aspects of morphology: it does not address the syntactic and semantic reasons why certain morphemes may combine with each other while others may not. Chapters 1 and 2 form a useful introduction to Section 3 in particular, which explores reduplication and infixation phenomena. In Section 4, he turns to English, providing an account of the "multiple use" of the suffix *s* in English. In English, both the possessive and the plural forms of most nouns sound alike: *book's/books*, *tool's/tools*, *judge's/judges*. Interestingly, a possessive plural is formed exactly the same way: *books'*, *tools'*, *judges'*, not **books's* ([bookss] or [booksəs]), etc.

The remaining two chapters explore syntactic problems in terms of Optimality Theory. David Pesetsky begins Chapter 5 with a beautifully clear introduction to the essence of current syntactic theory, elucidating both the phenomena and the formal explanations. This part of the chapter is an excellent introduction for Chapter 6 as well as for the rest of Chapter 5, while the reader who is already familiar with current syntactic theory might wish to skip the introductory section and begin directly with Section 2, comparing standard theory and Optimality Theory in syntax, or Section 3, an exploration of the distribution of *that* and of relative pronouns. In English, we can say *the man who I saw*, *the man that I saw*, and even *the man I saw*, but we don't say **the man who that I saw*.

Pesetsky shows that the facts for the comparable sentences in French are subtly and interestingly different, and provides an Optimality Theoretic account of each pattern.

In Chapter 6, Margaret Speas first evaluates the standard "principles and parameters" theory of syntax, and shows that the inviolable principles of this theory are inviolable simply because each such principle includes an "escape hatch" for when it does not hold. She then shows that by adopting Optimality Theory, the principles can be expressed more generally, the escape hatches being eliminated in favor of constraint ranking. The discussion centers on the analysis of "null pronouns", occurring in the position of the underscore in sentences like *Mary expects __ to promote Bill* and __ *To behave in public would enhance Bill's reputation*. In the first, it can only be Mary who will do the promoting, whereas in the second, the one being admonished to behave might be Bill, but might also be some other person. After formulating constraints and constraint rankings to explain these facts, she analyzes the properties of null pronouns in a number of other languages to show that OT also insightfully accounts for the cross-linguistic patterns these pronouns display.



Kudos

A number of people worked very hard to bring this volume to publication. Several of these individuals are identified in the chapters for their contributions to the development of their respective content. We were ably assisted in the formatting, editing, and indexing of this book by Laura Moll–Collopy, Keiichiro Suzuki, and Dirk Elzinga. We are very proud of the results they achieved.

Funding for this book came in part from National Science Foundation grant BNS–9023323 to Diana Archangeli, for which we are grateful.

Finally, for their patience and moral support throughout the duration of this project we thank Dante Archangeli and Nancy Kelly. Special thanks go to Marina and Amico Archangeli for always being there.

Electronic Access to Optimality Theory

Readers who are interested in accessing more material on Optimality Theory have three options. The first is to look to the published literature. A good start is this book. However, at the time this book is going to press, there is very little published work available on Optimality Theory. By contrast, electronic access to a wide variety of works is possible. The Rutgers Optimality Archive (ROA) is a well–maintained electronic repository of unpublished works in OT, which is accessible through the World Wide Web. The ROA includes abstracts of most entries.

URL of the Rutgers Optimality Archive on the WEB
`http://ruccs.rutgers.edu/roa.html`

The third option is to join the optimality net. This is an electronic bulletin board which posts information about additions to the archive and archive maintenance. It also occasionally serves as a forum for discussion of issues in OT. The instructions for joining this discussion group are available in the ROA homepage.

Tucson, Arizona, USA
October 1996

Afterword

The six chapters of this volume lay out basic properties of Optimality Theory and present examples to illustrate how the model works in different linguistic domains. Numerous important issues arose in these discussions, with most of them being left unresolved pending further research. This uncertainty represents the state of OT itself, which is still in its infancy. In our final comments here, we explore one of the unresolved issues which is central to OT research, and to linguistic research in general: *what is the nature of the input?* As with other aspects of linguistic theory, the nature of the input takes on new form and new significance when viewed from the OT perspective.

We do not need at this point to raise the parallel issue concerning the output. The output is easy to understand, at least intuitively: it is what people say, though as the reader will have discovered by now, it may have far more structure than meets the eye (or ear). By contrast, the input, being a hypothetical construct, is a harder concept to grasp. Nevertheless, and despite its output orientation, OT relies heavily on the input since FAITHFULNESS constraints require identity of input and output. If we do not know what the input is, we do not know how to evaluate those constraints.

We can approach the question of what is in the input by asking three questions: (i) what *must* be present in the input? (ii) what *cannot* be present in the input? and (iii) what *might* be present in the input? Broadly speaking, the answers to these questions categorize linguistic information into information that maintains distinctions, information that erases distinctions, and information that simply cannot make any distinctions.

What must be present?	information that maintains distinctions between distinct linguistic entities
What cannot be present?	information that removes distinctions between distinct linguistic entities.
What might be present?	information that cannot distinguish between distinct linguistic entities.

The Phonological Input

As an example of these three categories, let us consider the extent to which phonological features are specified in the input. Such degree of specification is a direct result of two

standard assumptions: first, a grammar characterizes what a speaker knows of his or her language; and second, lexical entries contain all specific knowledge about individual morphemes. If, for instance, in some language (like American English) all low vowels are unround vowels (which allows the [+low, –round] vowels [æ] of *cat* and [a] of *hot*, but not a [+low, +round] vowel, such as British English [ɒ] of *caught*), then one effect of the grammar is to ensure that all [+low] vowels occur in the output as [–round] vowels.

Let us consider our questions about the input in light of American English low vowels.

- **What *must* be present?** Whether or not a vowel is [+low] is an idiosyncratic fact about a particular vowel: [+low], then, is an essential part of the lexical entry of a low vowel.
- **What *cannot* be present?** Whether a low vowel is [+back] ([a]) or [–back] ([æ]) is an idiosyncratic property of low vowels. Thus, [–back] *cannot* be present in the lexical entry of the vowel [a] in *hot* [hat]. Were [–back] present, the morpheme would be *hat* ([hæt] with [–back] [æ]), not the desired *hot*.
- **What *might* be present?** Whether a vowel is round or not is predictable if it is a low vowel, so neither [+round] nor [–round] are *necessary* in the representation of low vowels in English. If [–round] is present, no harm is done; if [+round] is present in the input, it must be absent in the output: the requirement (constraint) that low vowels be unrounded overrides any specifications in lexical entries.

In OT, where constraints reside only in the constraint hierarchy, the grammar of American English must have the ability to assign [–round] to [+low] vowels no matter how they are specified for [round] in the input. This is readily achieved by ranking LOWROUND, a constraint requiring that [+low] vowels be [–round], and FAITHLOW, requiring faithfulness to input [low] values, above FAITHROUND, a constraint requiring faithfulness to input [round] values, as summarized in (1).

(1) **Constraints relating [low] and [round] in American English and their ranking**
  a. **LOWROUND**: [+low] vowels are [+round]
  b. **FAITHLOW**: input [+low] vowels are [+low] in the output
  c. **FAITHROUND**: input [+round] vowels are [+round] in the output
  d. **LOWROUND, FAITHLOW » FAITHROUND**

Regardless of the representation for the [round] feature in the input, any input [+low] vowel must surface as [–round], as demonstrated in the **tableau des tableaux** in (2). Three inputs are shown: /a/, which is [+low, –round]; /ɒ/, which is [+low, +round]; and /A/, which is simply [+low]. Regardless of the input, the constraint hierarchy invariably selects [a] as the optimal output. Furthermore, in each case, the decision is made by the two higher-ranked constraints. The subordinate FAITHROUND plays no decisive role.

Thus, in English, for an input corresponding to the vowel [a], [+low] *must* be present. Conversely, [–back] *cannot* be present, for if it were present, [æ] would surface, not [a]. Finally, either [+round] or [–round] *might* be present, but it really doesn't matter whether either is present, because FAITHROUND is subordinate to LOWROUND and FAITHLOW.[1]

[1]What about [+back]? Is [+back] necessary in the input for the vowel [a], or is [+low] alone sufficient? There is motivation for a constraint LOWBACK: [+low] vowels are [+back]. Low vowels tend very strongly to be [+back], since tongue body lowering and backing are sympathetic gestures

(2) **Tableau des tableaux for output [a] in American English**

input		*output*	LowRound	FaithLow	FaithRound
/a/	☞	[a]			
		[ɒ]	*!		
		[ə]		*!	
		[A]	*!		*
/ɒ/	☞	[a]			*
		[ɒ]	*!		
		[e]		*!	
		[A]	*!		*
/A/	☞	[a]			*
		[ɒ]	*!		
		[e/ə]		*!	*
		[A]	*!		

Choosing the Correct Input

How, then, do we choose the correct input? There are two basic approaches to this question. The first is to propose some additional mechanism that makes the selection, a mechanism outside of the general principles of OT. A variety of such mechanisms are available: the best known from the OT literature is **lexicon optimization** (see also Chapters 1 and 3). Lexicon optimization examines tableaux des tableaux, such as (2), and determines the optimal input–output pairs by finding which one has the fewest highly ranked constraint violations. In our example, it is the completely faithful /a/ ↔ [a], which

physiologically. With such a constraint outranked by FaithBack in English, [æ] is able to surface (FaithBack ensures that an input [–back] will surface). More to the point, whether surface [a] corresponds to input [a] or input [A] (i.e. [+low], with no [back] specification) is moot: in both cases, LowBack ensures that a back vowel [a] surfaces.

input		*output*	FaithBack	LowBack
/a/	☞	[a]		
		[æ]		*!
		[A]		*!
/A/	☞	[a]		
		[æ]		*!
		[A]		*!
/æ/		[a]	*!	
	☞	[æ]		
		[A]		*!

If LowBack and LowRound outrank FaithBack and FaithRound, the result is a language in which there is only one low vowel, the back, unround [a] (e.g. Spanish).

has no violations at all. The optimal input, like all of the candidate inputs, are fully (i.e., redundantly) specified for phonological features. In other words, feature specifications which *might* be present *are* present in the input.

Another reasonable mechanism is to retain whatever is common to all inputs which correspond to a given output, and to remove what is not shared, the **minimal specification** approach. In our example, every input vowel is [+low], but values for [round] vary. Accordingly, the optimal input–output pair is /A/ ↔ [a]. With minimal specification, inputs are like underlying representations in classical generative phonology in that there is both a single correct input for each distinct output, and the optimal input is partially (i.e., nonredundantly) specified for phonological features. Feature specifications which *might* be present *are not* present in the input.

Both lexicon optimization and minimal specification follow the intuitive approach to the relation between input and output — in any domain, not just linguistics — as a **mapping** from input to output. We normally consider inputs to be underlying, or basic, and the outputs to be derived, or secondary. For example, when we add up a bunch of numbers, say 5, 6, and 7, we think of those numbers as input, and their sum, 18, as output. We don't know what the sum is until we have added up the numbers.

However, the relation between input and output does not have to be thought of that way. One can equally legitimately think of there being a mapping from output to input. For example, given the number 18, we can ask what bunches (analogous to **numerations** as defined in Chapter 6) of numbers add up to that number. Answers include {5, 6, 7}, {2, 4, 6, 6}, and {3, 3, 3, 3, 3, 3}. OT, as an output-oriented model, suggests considering the nature of the input in this same light: it does not matter which input is selected as long as the selected input results in the desired output. Consequently, any winning input–output pairs can be considered viable: /a/ ↔ [a], /ɒ/ ↔ [a], and /A/ ↔ [a]. According to this view, there need not be a unique input for a given output. Moreover, the input may be redundantly or nonredundantly specified.

This theory of the input represents more of a break with classical theories of underlying phonological representations than do minimal specification and lexicon optimization. Not only is it not necessary that there be a unique input for a given output, the input may contain elements, such as [ɒ], which appear in no output. The input is simply a set of forms which is associated with a given output by a particular constraint ranking. We call this approach the **OT perspective** since it is available under OT, but runs counter to traditional generative analysis.

Proponents of the OT perspective need not assume that every speaker of American English represents the vowel of, say, *hot* in all of the ways that the approach permits. It is consistent with this approach that different speakers represent it differently, even though they pronounce it the same. Suppose then that speaker A represents that vowel as /a/, whereas speaker B represents it as /ɒ/, and suppose also that both change their constraint rankings, so that input /ɒ/ results in output [ɒ], presumably as in British English. Then speaker A would continue to pronounce *hot* as [hat], whereas speaker B would now pronounce it as [hɒt]. Thus variability in inputs is supported if constraint reranking results in speakers' producing distinct outputs from apparently identical inputs. On the other hand, if all speakers actually posit identical inputs for a particular output, then reranking of constraints could never result in differences of this sort.

lexicon optimization	The optimal input is selected from all inputs corresponding to a single output by being the one which incurs the fewest highest–ranked constraint violations.
minimal specification	The optimal input contains all information that is common to all inputs corresponding to a single output but contains no information beyond that.
the OT perspective	Any input which results in the correct input-output pairing is a viable input.

The Role of Alternations in Constraining the Input

Morphophonemic alternation does, however, impose strict limits on how widely input representations can vary under the OT perspective. In most dialects of English, the words *petal* and *pedal* are pronounced identically: [pʰéɾəl], where [ɾ] is a "flapped *r*", the raised *h* indicates aspiration, and the acute accent marks main stress. Given that there is no other evidence bearing on the form of the input for these words, the second consonant of both can be analyzed as corresponding to any of the inputs /t/, /tʰ/, /d/, and /ɾ/, despite the distinct orthographic symbols.

The proviso that *there is no other evidence bearing on the form of the input* is crucial, as we can see when we consider the words *metal* and *medal*, which are also pronounced identically, as [méɾəl]. Both words (more precisely, both morphemes), however, have different pronunciations when certain suffixes are attached. When the suffix *–ic* is added to *metal*, the resulting word *metallic* is pronounced [mətʰǽlək]; but when the suffix *–ion* is added to *medal*, the resulting word *medallion* is pronounced [mədǽlyən]. There is no reason to believe that the difference in the output corresponding to the second consonant of these words is related to the difference between the *–ic* and *–ion* suffixes. Rather, the difference has to do with how the second consonants of the morphemes *metal* and *medal* may be represented in the input. In the case of *metal* it is /t/ or /tʰ/, and not /d/ or /ɾ/; whereas in the case of *medal* it is /d/, and not /t/, /tʰ/ or /ɾ/. The assumption that the input corresponding to [méɾəl] can contain /d/, /t/, /tʰ/ or /ɾ/ can only be sustained if one is prepared to deny that the morphemes *metal* and *medal* do not occur in *metallic* and *medallion*.

The fact that *petal* and *pedal* do not have alternate forms in different morphological contexts whereas *metal* and *medal* do also has an effect on the input form of the second vowels of each of the latter pair of morphemes. In the former pair, we can assume that that vowel can be /ə/ or any short vowel such as /i/, /ɛ/, or /æ/, since any unstressed short vowel in English surfaces as [ə]. In the latter pair, however, we must assume that it is /æ/, since that is the form the vowel takes in the outputs for *metallic* and *medallion*, where this vowel is stressed. This means that the input forms for the morphemes *metal* and *medal* are restricted to /mɛtæl/ (or /mɛtʰæl/) and /mɛdæl/ respectively.

Another type of information typically included in inputs is the ordering of phonological segments. For example, the words *tack* and *cat* differ solely in the order of the two consonants, hence that order must be specified in the input. On the other hand, if

we know that the vowel is first, then the order of the two consonants need not be included in the input, for only one sequencing of these consonants results in a possible English word: *act* [ækt] is an actual word of English, whereas *[ætk] is impossible, due to English constraints on admissible coda sequences. According to the OT perspective, inputs in which the consonants are ordered *k–t* and *t–k*, as well as inputs in which they are not ordered with respect to each other, are legitimate, and the question of how native speakers actually represent the sequence of segments of the word *act* is left open.[2]

Prosodic Structures

Given that phonological segments and aspects of their order must in general be specified in the input, what about other sorts of phonological structure? Consider the prosodic structure of morphemes, for example, the fact that *petal*, *pedal*, *metal*, and *medal* each have two syllables, the first of which is stressed; i.e., that it has the form of a binary trochaic foot (see Chapter 2). Should that structure be represented in the input? Assuming that this prosodic structure can be predicted from the constraint hierarchy applied to the segmental structure of each of these morphemes, this structure can be, but does not have to be, present in the input. Strict adherence to lexical optimization would require that this structure be present, maximizing redundancy; the minimal specification perspective would require that it not be present, minimizing redundancy. The OT perspective permits inputs both with and without the prosodic structure specified. If the prosodic structure is unspecified in the input, the constraint hierarchy will select only those outputs in which the correct structure appears. If the correct prosodic structure is specified in the input, no harm is done: the output is completely faithful to the input at least as far as prosodic structure is concerned. What happens if an incorrect prosodic structure is specified in the input, for example, an iambic structure?

The answer to this question is not straightforward, for it depends on the answer to a different question: are there cases in which the prosodic structure of a morpheme *must* be present in the input? This question is reminiscent of the long–debated question of whether or not stress patterns are phonemic (contrastive) in English.[3]

A clear case for stress patterns being at least marginally phonemic in English is presented by pairs of words like *Pascal* [pæ̀skʰǽl] and *rascal* [rǽskəl]. Given that *rascal* combines with the suffix *ity* resulting in *rascality* [ræ̀skʰǽləɾi], we conclude that the inputs for the two words contain the segment sequences /pæskæl/ and /ræskæl/ respectively. However, from those sequences alone, one cannot determine that the output associated with the former contains two stresses, whereas the output associated with the

[2]On the other hand, given that both *ask* and *ax* are well–formed in standard English, we see that the order of coda consonants *s–k* and *k–s* must be specified in the input. Those speakers whose grammar contains a constraint which disallows output *sk* coda clusters (including those who pronounce *ask* and *ax* alike as [æks]) may nevertheless represent the word *ask* as containing the input *s–k* sequence.

[3]Given pairs of words like *ínsùlt* (noun) and *ìnsúlt* (verb), it is sometimes concluded that the prosodic structure for at least certain words in English must be considered phonemic. On the other hand, it can also be maintained that the prosodic structures in these cases are predictable on the basis of other information also present in the input, such as the syntactic categorization of the words, and their morphemic composition (prefix and stem).

latter contains a single stress. Clearly, these inputs must be distinguished, so that each is correctly associated with the appropriate output. Probably the simplest way to do so is to require that the prosodic pattern of the former, at least, be present in the input, and to require that faithfulness to that pattern outrank the constraints that favor binary trochaic feet.

English must allow faithfulness to certain input stress patterns to outrank constraints favoring binary trochaic feet. What happens, then, if *metal* or *pedal* is lexically represented with an iambic foot? If left intact, the lexical iamb would produce stress only on the second syllable as in *balloon* [bəlún]: *[mətʰǽl]. Since faithfulness to an iamb is necessary for a word like *balloon*, any word which is lexically represented as an iamb would be so pronounced. Thus *metal*, *medal*, etc., cannot include iambic foot structure in the input. No prosodic structure is necessary, but if any is included, it must be the correct trochaic structure.

The Morphological Input

Classical generative morphology has made the same assumption that classical generative phonology made: that there is a single input for each distinct output. The OT perspective, however, leads us to expect that morphological inputs can have the same freedom of input representation that we found with phonological inputs; that is, a given output word (understood as a complex consisting of phonological, semantic, categorical, and structural information) might be associated with more than one input representation, which differ in how they represent information that might be present in the input.[4]

Blocking

Perhaps the most convincing evidence that output words can be associated with more than one input representation is provided by the phenomenon of **blocking**, which occurs when the existence of a particular word with a particular meaning prevents the expression of the same meaning by another word, typically one which is formed by a regular morphological process. For example, the existence of the word *went* 'go:Past' is said to block the existence of the word **goed* 'go:Past' in English; see Kiparsky (1982), Pinker (1995).

In many instances of blocking such as this one, the blocking word contains fewer morphemes than the blocked word. This suggests a constraint such as **MONOMORPH**, which favors the expression of particular meanings by words of one morpheme.

[4]Our concern here is the nature of the input to GEN for evaluation by a specific constraint hierarchy. Sometimes this input is a single morpheme, e.g. $[go]_V$ 'go'; sometimes it is a numeration of several morphemes, e.g. {$[cat]_N$ 'cat', *s* 'more than one'}. Our working assumption is that such numerations may be created randomly, and that EVAL selects the best possible output for each input; EVAL selects the null parse in the case of input numerations that are too ill–formed, e.g. {*s* 'more than one', *d* 'Past'}. See also note 9.

(3) MONOMORPH: Words consist of one morpheme.

Blocking of **goed* results if MONOMORPH outranks the constraints that require that the phonological and categorical content of the input numeration {$[go]_V$ 'go', *ed* 'Past'} be preserved in the output, as shown in the tableau in (4).

(4) A tableau for *went* 'go:Past'

$[go]_V$ 'go', *ed* 'Past'	MAX('PAST')	MONOMORPH	MAX(GO)	MAX(ED)
$[[go]_V\ ed]_V$ 'go:Past'		*!		
☞ $[went]_V$ 'go:Past'			*	*
$[go]_V$ 'go'	*!			*

Then, since *went* is also the optimal output for the input consisting of the lexical item *went*, it follows that output *went* is associated with two distinct inputs, namely the numeration {*go*, *ed*} and the lexical item *went*.[5]

Zero Affixation

Our next example involves pairs of words which are identical phonologically, but which differ both semantically and categorically, such as the noun *nail* 'nail' and the verb *nail* 'attach with a nail'; henceforth $[nail]_N$ and $[nail]_V$, respectively. One familiar analyis of such pairs is to consider the noun form to be basic and the verb form to be derived by a **zero–affix**, which effects the appropriate categorical and semantic changes but adds no phonological content. This type of analysis parallels the minimal specification perspective in phonology. An alternative is one which does not postulate zero–affixes, but rather treats both the noun and the verb forms as separate inputs, paralleling the lexicon optimization perspective. Both types of analysis assume that there is exactly one input for each output. For the output consisting of $[nail]_V$, the minimal specification perspective postulates the input consisting of $[nail]_N$ plus a zero–affix with certain properties, whereas the lexicon optimization perspective postulates the input consisting of $[nail]_V$ itself.

According to the OT perspective, given the appropriate constraints, both inputs are possible. Thus, we may have speakers of English all of whom use the words $[nail]_N$ and $[nail]_V$ exactly alike, but who represent the input for $[nail]_V$ differently. For example, according to minimal specification, the lexicon might contain $[nail]_N$ and a zero–affix which derives certain verbs, including $[nail]_V$, from nouns; while according to lexicon optimization, it might contain $[nail]_N$ and $[nail]_V$, and no affix. From the OT perspective, however, it might contain all three items: $[nail]_N$, $[nail]_V$, and a zero–affix capable of deriving $[nail]_V$ from $[nail]_N$.

[5]For convenience, we may omit the categorical or semantic information associated with particular morphemes.

Ordinary Affixation and Back–Formation

Next we consider cases of ordinary affixation, in which the affix has both phonological and semantic content, as in the word *cats* 'more than one cat', which is transparently related to the morphemes *cat* 'cat' and *–s* 'more than one' (see Chapter 4). The output structures of this word and others like it are generally assumed without argument to correspond to unique inputs, such as the numeration {[*cat*]$_N$, *s*}. There are, however, two alternative inputs to consider, one structured, [[*cat*]$_N$, *s*]$_N$,[6] and the other unstructured (monomorphemic), [cats]$_N$. We argue that all three types of inputs are empirically necessary, but that each has its own particular distribution. For instance, we show that the monomorphemic input is inappropirate for transparent plurals.

Consider first an appropriate use of an unstructured monomorphemic input. This is exactly like the input [went]$_V$ 'go:Past' discussed above, and provides a means of representing irregular plurals, both ones for which there is an unrelated singular form (e.g. *people/person*) or no singular at all (e.g. *odds/*odd*$_N$). In each of these cases, we suggest that the input form is unstructured, e.g. [people]$_N$ 'more than one person'.

Is the unstructured form a possible input form for a word like *cats*, which is transparently related to the singular form *cat*? We believe not, for an input like [cats]$_N$ cannot characterize the relation between this and the singular. Thus, it simply does not characterize the speaker's knowledge of the language, that the plural *cats* is related to both the singular *cat* and the plural *s*. The constraint hierarchy must require that this knowledge be expressed, thereby requiring that the output *cats* correspond to a polymorphemic input. This leaves two possibilities: the structurally complex lexical item [[cat]$_N$, s]$_N$, and the numeration {[cat]$_N$, s}. From the OT perspective, it could be either. The numeration option is necessary for it reflects the ability to create a plural from a singular; without this option we would be incapable of creating plurals but could only reiterate plurals we had heard before.

The argument for structurally complex inputs is itself more complex. Consider a word like *player* 'one who plays', which, like *cats*, is transparently related to two morphemes, [*play*]$_V$ 'play' and *–er* 'one who *V–s*', where *V* is the verb stem to which the suffix attaches. As before, we can consider two possible inputs for the output *player*: the numeration {[*play*]$_V$, *–er*}, and the structured lexical item [[*play*]$_V$, *er*]$_N$, which incorporates the suffix.

The case for there being a lexically structured input [[*play*]$_V$, *er*]$_N$ related to the output [[*play*]$_V$ *er*]$_N$ is quite strong for all speakers of English. The word *player* has several meanings in addition to 'one who plays', all of which are related to the meanings of [*play*]$_V$ and *–er*, but none of which can be fully predicted from those meanings, including 'actor', 'gambler', 'participant in an organized activity', 'member of a sports team', 'one who plays a musical instrument', and 'device for producing musical sounds'. Correlating these meanings with the lexically structured input [[*play*]$_V$, *er*]$_N$ gives a means of

[6]The notation [[*cat*]$_N$, *s*]$_N$ (equivalently [*s*, [*cat*]$_N$]) is intended to indicate that both morphemes [*cat*]$_N$ and *s* occur as constituents of the presumed lexical item *cats*, but not necessarily in that order. We assume that the correct output order is determined by constraints; see Chapter 4, and the subsection "Morpheme Order" below.

characterizing the systematic relation between *player* and the morphemes [*play*]$_V$ and *–er* while simultaneously characterizing the idiosyncratic semantic properties.[7]

We conclude, then, that inputs can include single unstructured lexical items made up of a single morpheme, including those, like [*went*]$_V$, whose meanings are normally expressed polymorphemically; single structured lexical items, like [[*play*]$_V$, *er*]$_N$; and numerations, like {[*cat*]$_N$, *s*}, but each has its specific role. Numerations express completely transparent morphology; structured lexical items express morphology whose form is transparent but whose semantics is not; and unstructured lexical items express irregular morphology.

Further confirmation that the existence of multiple inputs for a single output is a natural state of affairs is provided by the phenomenon of **back–formation**. Consider a dialect of English in which the word *burglar* 'one who steals' exists, but not *burgle* 'steal'. Some speakers of this dialect, we may presume, represent both the input and the output simply as [*burglar*]$_N$ 'one who steals'. Others recognize the occurrence of the *–er* affix in that word (the spelling is irrelevant; it has the right pronunciation), and represent the input as [[*burgle*]$_V$, *er*]$_N$ and the output as [[*burgle*]$_V$ *er*]$_N$ 'one who steals'. However, they do not recognize [*burgle*]$_V$ as an independent lexical item. It is precisely these speakers whom we would expect to "discover" that word, and associate with it the meaning obtained by subtracting, as it were, the meaning of *–er* from the meaning of *burglar*. That is, they are able to extract the word *burgle* from the lexical representation of *burglar*. Having done so, they would have developed a grammar in which *burglar* is now associated with two inputs: the lexical item [[*burgle*]$_V$, *er*]$_N$, and the numeration {[*burgle*]$_V$, *er*}.[8]

We turn now to consider the categorical and phonological properties of morphemes, and consider which of these properties must be, might be, and cannot be present.

Categorical and Phonological Properties of Morphemes

The categorical properties of a morpheme tell the syntactic and/or semantic role(s) of the morpheme. For instance, a stem may be a noun or a verb, etc. Affixes "do" more: an affix may change the root's category (e.g. the adjective *happy* is a noun when paired with the suffix -ness, *happiness*); it may add grammatical information like tense, mood, number, case, (e.g. the addition of *-ed* to an English verb creates a past tense: *wash* vs. *washed*); it may add other types of information, such as reflexivity, reciprocity, reversive

[7]There is no reason to exclude the meaning 'one who plays' from also being associated with the complex lexical item. If this is done, then the output [[*play*]$_V$ *er*]$_N$ 'one who plays' might be associated with two distinct inputs: the numeration {[*play*]$_V$, *er*} and the lexical item [[*play*]$_V$, *er*]$_N$.

[8]Back–formation can occur whenever a form which is monomorphemic for some speakers is bimorphemic for others, and one of those morphemes does not occur as an independent word, but has the potential to do so. For those English speakers for whom the noun *destruction* is related to the input numeration {[*destroy*]$_V$, *ion*}, or to the bimorphemic lexical item [[*destroy*]$_V$, *ion*], no back–formation is possible, because *destroy* already exists as an independent word. Similarly for those for whom only the monomorphemic input [*destruction*]$_N$ exists, no back–formation is possible. However, for those for whom the input is {[*destruct*]$_V$, *ion*], back–formation is possible, resulting in the creation of the intransitive verb *destruct* 'break apart'.

action, diminutive, comparative (*happy* vs. *happier*), etc. (e.g. *phonology*, a noun, 'the study of language sound patterns' vs. *phonologist*, also a noun, 'one who does phonology').

This information is largely idiosyncratic, and thus is information which *must* be present and *cannot* be absent, for the most part. Questions arise about the role such information may play in constraints, but not about whether such information must be present.

Next, we consider whether phonological information is present at all in the input of affixes, arguing that at least sometimes such information cannot be omitted. Above, we discussed the so-called "zero–affixes", which give rise to morphological effects despite having no phonological content. Such effects may simply be the result of constraint interaction; they may also be a response to inputs of the type discussed above. Such morphemes, along with ones such as the Paamese reduplicative morpheme *RED* discussed in chapter 4, raise the issue of whether phonological content is necessary at all for morphemes.

We hypothesize that phonological content is necessary for at least some of the phonologically expressed morphemes, because the segments of affixes can be subject to the same constraints that hold of root segments. We would expect quite divergent behavior between affix segments and root segments were their sources distinct, i.e. if affixal segments were selected through constraint interaction, rather than being inherent in the input. This empirical result is consistent with the substantive claim that constraints are universal, not language–particular statements such as "the plural is expressed by a coronal fricative", which might account for English, but which would not be relevant for most other languages, such as Yawelmani and Paamese.

Morpheme Order

Once we accept that the phonological properties of morphemes are present in the input, we must face the question of whether the morphemes are linearly ordered in the input. As shown in Chapter 4, morpheme order can generally be determined by the definition and ranking of specific ALIGNMENT constraints, provided that certain categorial properties of morphemes are represented in the input, such as that a particular morpheme is a stem or an affix. Consequently, morpheme order generally need not be specified in the input.

Empirical support for this view comes from McCarthy's (1995) observation that it is not uncommon for phonological alternations to be sensitive to the order of segments in the input. If morphemes are unordered in the input, then the segments of one morpheme are not ordered with respect to segments of another morpheme and so the conditions for order-sensitive alternations are satisfied within a morpheme, predicting that such alternations will not take place when the relevant segments are in different morphemes.[9]

There are other implications of not ordering morphemes in the input. For example, this assumption results in a simpler account of the ungrammaticality of certain morphologically complex words. By definition, a word is ungrammatical in a language if it fails to be the output for any morphological input. Consider for example the Turkish word *ellerimizden* 'from our hands', consisting of the noun stem *el* 'hand', the suffix *–ler* 'plural', the possessive suffix *–imiz* 'our' (actually analyzed as two morphemes, *–im* 'first person' and *–iz* 'plural for personal proforms'), and the directional suffix *–den* 'from'. Given an input consisting of the numeration {*den, el, im, iz, ler*} and their categorizations, they can only be combined in the order indicated; other combinations, such as **denelimizler*, are ungrammatical in Turkish. This results simply from the interaction of particular alignment constraints for Turkish morphology without regard to the input order of the morphemes.

Now suppose that morpheme order is part of the input. Then given the input *el* + *ler* + *im* + *iz* + *den*, where each '+' indicates the concatenation of the morphemes it connects (thus fixing their order), the output *ellerimizden* is selected as before. However, *den* + *el* + *im* + *iz* + *ler* is also a possible input, to which the ungrammatical word **denelimizler* is maximally faithful. To rule it out, we must suppose that the alignment constraints which determine the correct morpheme order in Turkish outrank those faithfulness constraints, resulting in an output either in which the morphemes are reordered to *ellerimizden*, or to which morphemes are deleted or added so as to result in some other grammatical Turkish word (e.g. *elimiz* 'our hand') or the null parse. On the other hand, if we assume that the input to the morphological component is unordered, then the problem of dealing with inputs like *den* + *el* + *im* +*iz* + *ler* simply does not arise.[10]

In the subsection "Ordinary Affixation and Back–Formation" above, we pointed out that morphological inputs could be structured, while leaving information about the order of morphemes unspecified. However, there are examples, such as the English compound nouns *housework* 'work of housekeeping', and *workhouse* 'house of correction for minor offenders', which suggest that at least in some cases, morpheme order *must* be specified in the input in order to account for the difference in meaning between the two outputs. If the morphemes contained in the inputs associated with these words as outputs are unordered, then those inputs would have the same structure, namely $[[house]_N, [work]_N]_N$, one instance associated with the meaning of *housework* and the other with the meaning of *workhouse*.

However, it is not the *order* of the morphemes which accounts for the difference in meaning between the two compounds, but rather the *structural relationship* between them. In these, like in most other two–word compounds in English, the second word is the **head** and the first is its **complement**. In Chapters 5 and 6, it was pointed out that the head–complement relation in syntax can be expressed structurally, in terms of X–bar theory. Suppose we extend this idea to morphology, specifically for the analysis of compounds. If the structure of the compounds is represented as in (5), in which *NP* is the complement and the inner *N* is the head, the order of the morphemes in the output can be

[9]On the other hand, if morphemes are ordered in the input, then those conditions can be satisfied if the input order gives rise to the appropriate segment sequence for the alternation. This does not appear to be the case.

[10]Eliminating morpheme order from the input, however, does not eliminate all problematic inputs. For example, in Turkish, the suffix *in* is a possessive suffix meaning 'your', which occurs in the same position in a Turkish word as the suffix *im* 'my' does, as in *elin* 'your hand'. Now consider the input numeration {*el, im, in*}. The maximally faithful outputs **elimin* and **elinim* are both ungrammatical (as are the corresponding English phrases **my your hand* and **your my hand*). To eliminate them, we must suppose that Turkish grammar contains a highly ranked constraint that effectively forbids combination of possessive suffixes associated with a single noun stem. Violating that constraint results either in the deletion of all but one of those suffixes, or the null parse, depending on how the constraint is stated.

determined by an alignment constraint that requires heads to be final in English compound words.

(5) **Presumed input structures for the compound nouns *housework* and *workhouse***
   a. $[[house]_{NP}, [work]_N]_N$ 'work of housekeeping'
   b. $[[house]_N, [work]_{NP}]_N$ 'house of correction for minor offenders'

The input structures in (5), however, cannot be correct as they stand, because they contain the category NP, which is a phrasal category, and phrasal categories do not belong to the morphological component of the grammar. Phrasal categories belong to syntax; moreover, they are assigned to output structures in the syntax, not to inputs, the inputs being numerations of words such as $[house]_N$ and $[work]_N$, which are themselves the output of morphology. That is, representations such as $[house]_{NP}$ in (5a) should be replaced by the syntactic mappings of input numerations, whose members all belong to categories of the morphological component. Such a mapping is expressed by syntactic tableaux, which we can represent schematically as $T(\{w_1, \ldots, w_n\})$, where $\{w_1, \ldots, w_n\}$ is a numeration of input words. For example, the structures in (5) may be revised as in (6).

(6) **Revised input structures for the compound nouns *housework* and *workhouse***
   a. $[T([house]_N), [work]_N]_N$ 'work of housekeeping'
   b. $[[house]_N, T([work]_N)]_N$ 'house of correction for minor offenders'

In this way, only morphologically–defined categories are present in the inputs of compound words; the appearance of syntactically–defined categories is the result of the application of syntactic evaluation to those inputs. The resulting outputs, the compound words themselves, are in turn re–input to syntax, as in the input numeration {*hate*, *housework*, *I*} which results in the syntactic output *I hate housework*.

A structure such as (6b) can also be part of an input numeration in morphology, for example $\{[[house]_N, T([work]_N)]_N, s\}$, which is uniquely associated with the output *workhouses*. There are two constraints in English which determine where the plural affix appears in plural compound nouns: one requires it to be suffixed to the head of the compound, and the other to the last word of the compound. In the case of *workhouses* both constraints are satisfied, whereas neither is satisfied by the candidate **workshouse*. In case the head is not the last word of the compound, for example *jack–in–the–box* 'kind of toy', where the plural affix appears depends on how those constraints are ranked. For those speakers for whom both *jacks–in–the–box* and *jack–in–the–boxes* are well formed, the constraints are tied.

When the head of a compound is itself morphologically complex, as in *matchmaker*, the potential for multiple inputs arises, as in (7).

(7) **Inputs that may be associated with output *matchmaker***
   a. $[[[make]_V, er]_N, T([match]_N)]$
   b. $[[[make]_V, T([match]_N)]_V, er]$

Presumably, the output structure associated with (7a) is $[[[match]_{NP}\ [[make]_V\ er]_N]_N$, whereas that associated with (7b) is $[[[match]_{NP}\ [make]_V]_V\ er]_N$. The latter case is analogous to *burglar*, discussed above, in which the verb stem *burgle* may be recognized as part of the input, but not as an independent word. As in that case, back–formation is possible, which may lead to the eventual creation of the compound verb *matchmake*, just as the compound verb *air–condition* has already been back–formed from the compound noun *air–conditioner*.

In summary, we have shown the following concerning input–output relations in morphology. First, the same multiplicity of potential inputs for a given output that is possible in phonology is also found in morphology. Second, the constraints which handle the multiplicity of inputs in morphology account for the phenomena of blocking and back–formation. Third, categorical and phonological information must generally be represented in morphological inputs. Fourth, morpheme order generally must not be represented in morphological inputs. Fifth, certain syntactic tableaux must be represented in some morphological inputs, at least those involving compounding.

The Syntactic Input

There have been several different conceptions of the relation between input and output in syntax in the history of generative syntax, but in all of them, it has been assumed that each output is uniquely associated with a specific input. For example, in the Minimalist Program (see Chapter 6), in which the input is considered to be simply a numeration of words, any given input corresponds to a multiplicity of outputs, each essentially a structured arrangement of that input numeration. The outputs which correspond to a given input may or may not differ in meaning. For example, putting aside the effect of Case–marking (see below), the input numeration {*loves*, *Marina*, *Mico*} is associated with the outputs *Marina loves Mico* and *Mico loves Marina* (more precisely, the structures which those strings conventionally represent), and these obviously differ in meaning. The converse, however, does not hold. Given an output, say the structure of *Marina loves Mico*, there presumably is a unique input, namely the numeration {*loves*, *Marina*, *Mico*}. Structurally ambiguous strings such as *Flying planes can be dangerous* actually represent distinct outputs, and each of these is uniquely related to a single input (possibly the same input).

In Generative Semantics on the other hand (see Huck & Goldsmith 1995 for recent discussion), syntactic inputs are considered to be representations of pure meaning, each of which is related to the various outputs that express that meaning. Thus given the input '*love*'('*Marina*', '*Mico*'), a purely semantic representation of some sort, we may have as output both *Marina loves Mico* and *Mico is loved by Marina*, assuming that the latter is synonymous with the former. However, as in the Miminalist Program, only one input is associated with a given output.

Partially Structured Inputs in Syntax

OT syntax, as it has been developed in Chapters 5 and 6 and elsewhere, is also committed to the view that the input consists minimally of a numeration of words, each of which includes phonological, morphosyntactic, and semantic information, thus excluding the possibility of a treatment of the input as "pure semantics". Nevertheless, the idea that the input–output relation in syntax is constrained by semantics is an attractive one, so one would like to know whether it is possible within OT to develop a theory of that relation in which semantically divergent outputs such as *Marina loves Mico* and *Mico loves Marina* are associated with distinct inputs.

For these particular sentences, a mechanism is available, namely the marking of noun phrases for Case, which may be accomplished by requiring that at least their heads be specified for Case in the input, so that, for example, *Marina* marked for nominative case ($Marina_{NOM}$) is a distinct lexical item from *Marina* marked for accusative case ($Marina_{ACC}$). Accordingly, the input numeration for *Marina loves Mico* is {*loves*, $Marina_{NOM}$, $Mico_{ACC}$}, whereas the input numeration for *Mico loves Marina* is {*loves*, $Marina_{ACC}$, $Mico_{NOM}$}. However, this mechanism does not succeed in distinguishing inputs for another class of sentences, those whose outputs differ in the location of noun phrases with the same case. For example, both *Marina* and *Mico* are specified for nominative case in the sentences *Marina said that Mico left* and *Mico said that Marina left*.

To distinguish between these outputs in input numerations, we would have to add information to the lexical items *Marina* and *Mico* about whether they can occur in a main or in a subordinate clause, e.g. $Marina_{MAIN}$ vs. $Marina_{SUB}$. But even this desperate expedient would fail to distinguish, for example, *Dante denied that Marina said that Mico left* from *Dante denied that Mico said that Marina left*, in which *Marina* and *Mico* occur in different subordinate clauses.

What to do? More information is needed about inputs than is provided by numerations of words in order to insure that semantically distinct outputs correspond to distinct inputs. The obvious answer is to structure the input. However, this obvious answer has an equally obvious objection: it is the role of constraints to evaluate the structures that can be associated with inputs. By structuring the input, the constraints are left with nothing to do except to check for faithfulness violations of input structures. A subtler answer is based on the observation that syntactic structure is one of the things that *may* be in the syntactic input. Given the OT perspective, we are free to structure the input if doing so has desirable consequences, such as providing a unique output for a particular input. One side–effect is that a given output may be associated with more than one input, but as we have seen, that already occurs under the OT perspective in both phonology and morphology.

Let us consider again the output sentences *Marina said that Mico left* and *Mico said that Marina left*. How can we structure their inputs so that they are distinctive, yet leave room for syntactic constraints to perform nontrivial evaluations of candidate outputs? One way to do it is to adopt the mechanism proposed in the subsection "Morpheme Order" above, namely to incorporate syntactic subtableaux into syntactic inputs. More precisely, suppose that the evaluation of the subordinate clauses may be included in the input numerations for those sentences, so that one possible input for the output *Marina said that Mico left* is {*said*, $Marina_{NOM}$, *T*(*that*, *left*, $Mico_{NOM}$)}. Let us call such an input a **partially–structured input**. Suppose also that any change to the optimal structure of a subtableau in a partially–structured input results in a violation of a faithfulness constraint we call **FAITH(SUBTAB)**. Then the optimal output for the partially–structured input just given is *Marina said that Mico left*, as shown in (9).

(8) **FAITH(SUBTAB):** The optimal structure of a subtableau is selected in the output.

(9) **A tableau for *Marina said that Mico left* containing partially–structured input**

said, $Marina_{NOM}$, *T*(*that*, *left*, $Mico_{NOM}$)	FAITHSUBTAB
☞ Marina said that Mico left	
Mico said that Marina left	*!

Thus, the OT perspective enables us to formulate an input which is uniquely related to a particular output (or perhaps to all synonymous outputs that make use of the same vocabulary), without trivializing the role of the constraints that evaluate syntactic structure. Of course, the output sentence *Marina said that Mico left* is still optimally (along with other outputs, such as *Mico said that Marina left*) related to the unstructured input numeration {*left*, *said*, *that*, $Marina_{NOM}$, $Mico_{NOM}$}, but that simply leads to the now–familiar situation of a given output being related to more than one input.

Given that FAITHSUBTAB is itself a constraint, the question naturally arises whether it can be violated. The answer, not surprisingly, is yes. Consider the output structure for the sentence *Marina asked when Mico left*, which is uniquely related to the partially–structured input {*asked*, *Marina*, *T*({*left*, *Q*, *when*, *Mico*})}, where *Q* is an interrogative complementizer (see Chapter 5). The highest–ranked candidate selected by the subtableau *T*({*left*, *Q*, *when*, *Mico*}) is *when did Mico leave*. However, the value of *T*({*left*, *Q*, *when*, *Mico*}) in the main numeration is the lower–ranked candidate *when Mico left*. This and similar examples show that entire subtableaux, not just substructures, appear in partially–structured inputs in syntax.

The Problem of Too Many Inputs

The arithmetic analogy given above in the subsection "Choosing the Correct Input" carries over to linguistics in another way: the "larger" the output, the larger the number of inputs that can be related to it. In syntax, and in some languages also in morphology, the size of the output is potentially unbounded. As we consider increasingly larger outputs in syntax (measured, say, by the number of phrases for which subtableaux may be appear in the input), the number of inputs that may be optimally associated with those outputs grows extremely rapidly.[11] This suggests that people are highly selective in the kinds of input–output relations they assign to sentences beyond a certain size. We conjecture that this process is also controlled by optimization, with the input "chunked" into subtableaux so that each part is the largest possible while still insuring that semantically distinct outputs are associated with distinct inputs. But this question, like many of the others raised in this volume, can only be answered with further research.

[11]The number of potential inputs grows exponentially with the size of the output, as measured by the number of phrases for which subtableaux may be constructed.