

Tools for Analyzing the World's Languages

Morphology and Syntax

J. Albert Bickford

Based on earlier work by John Daly, Larry Lyons, and Mary Rhodes

©1998 by the Summer Institute of Linguistics, Inc.

07 06 05 04 03 02 01 00 99 98 10 9 8 7 6 5 4 3 2 1

Library of Congress Catalog No: 98-60950
ISBN: 1-55671-047-x

Printed in the United States of America
All Rights Reserved

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—electronic, mechanical, photocopy, recording, or otherwise—without the express permission of the Summer Institute of Linguistics, with the exception of brief excerpts in journal articles or reviews.

Copies of this and other publications of the Summer Institute of Linguistics may be obtained from

International Academic Bookstore
Summer Institute of Linguistics
7500 W Camp Wisdom Road
Dallas, TX 75236-5699 USA

Voice: 972-708-7404
Fax: 972-708-7433
E-mail: academic_books@sil.org
Internet: <http://www.sil.org>

Contents

Preface.....	vii
Preliminaries	
1 Introduction	3
2 Standard Grammatical Terminology	7
Morphology Group 1	
3 Morphemes and Hypotheses	19
4 Introduction to Morphology	25
Syntax Group 1	
5 Introduction to Syntax: Constituent Structure, Syntactic Categories, and Grammatical Relations	37
6 The Base: Phrase Structure Rules and the Lexicon	47
7 Embedding and Noun Phrase Structure	61
8 Verbal Valence: Subcategorization and Selectional Restrictions	75
9 Obliques	93
Morphology Group 2	
10 Inflectional Morphology	113
11 Derivational Morphology	135
12 Suppletion and Morphophonemics	151
13 Nonlinear Affixation	167
Syntax Group 2	
14 Nonactive Complements	183
15 Variable Orders of Constituents	201
16 Questions	219
17 Commands	235
18 Overall Structure of a Grammar	247
Morphology Group 3	
19 Case and Agreement	251
20 Word Division and Clitics	273
21 Passive and Voice	293
Syntax Group 3	
22 Embedded Clauses	317
23 Relative Clauses	331

Practical Helps

24	Hints for Linguistic Writing	351
25	Guide to Writing a Grammar Sketch	355
26	Grammar Filing.....	363
27	Lexical Filing.....	371

References

	Index.....	389
--	------------	-----

	381
--	-----

Noun

A NOUN refers to a person, place, thing, idea, or abstract concept. Within a clause, a noun typically functions as any one of the following: subject of the verb, direct object of the verb, indirect object of the verb, or object of a preposition.

Nouns can be subdivided into smaller groups in several ways. One subdivision is into proper and common nouns. A PROPER noun is the name of a particular individual, location, or product, e.g., *Becky, Lyndon Johnson, Canada, New York, Toyota, Nestlé Quik*. A COMMON noun is a generic name for any member of a group that shares certain characteristic properties: e.g., *professor, building, nomination*.

Another subdivision is into count and mass nouns. A COUNT noun refers to items which can be counted (i.e., which can co-occur with a numeral), e.g., *boy, pencil, carrot, car*. A MASS noun refers to items which cannot be counted, because they are conceived of as an undifferentiated mass, e.g., *water, sand, corn, oxygen*. Mass nouns are always singular in form.

Pronoun

PRONOUNS are used as replacements for nouns and noun phrases. Like nouns, pronouns refer to people, places, things, ideas, or concepts. They occur in the same positions in the clause as noun phrases. The basic difference between nouns and pronouns is that pronouns refer to someone/something in terms of their role in the speech situation. For instance, *I* is the designation that a speaker uses for himself, and *you* is the designation the speaker uses for the person(s) being talked to. When someone new starts speaking, these pronouns suddenly refer to a new set of individuals; their interpretation depends on a fact of the speech situation (who is speaking to whom). English personal pronouns can be classified according to person, number, gender, and how they are used within a sentence.

		Subject		Object		Possessive		Reflexive	
		Sg.	Pl.	Sg.	Pl.	Sg.	Pl.	Sg.	Pl.
first person		<i>I</i>	<i>we</i>	<i>me</i>	<i>us</i>	<i>my</i>	<i>our</i>	<i>myself</i>	<i>ourselves</i>
second person				<i>you</i>		<i>your</i>		<i>yourself</i>	<i>yourselves</i>
third person	masculine	<i>he</i>		<i>him</i>		<i>his</i>		<i>himself</i>	
	feminine	<i>she</i>		<i>her</i>		<i>their</i>		<i>herself</i>	
	neuter	<i>it</i>		<i>it</i>		<i>its</i>		<i>itself</i>	

There are also INDEFINITE PRONOUNS, e.g., *one, ones, some, someone, something*, and what could be called POSSESSED PRONOUNS (referring to the thing possessed), e.g., *mine, ours, yours, his, hers, theirs*.

Verb

A VERB is the core of a clause, the word that links the different parts of the clause together. Verbs generally indicate an action, a relation, or an experience: e.g., *walk, punch, undergird, own, realize*.

Adjective

An ADJECTIVE describes a property or characteristic of a noun, e.g., *beautiful, large, ordinary, overwhelming*. Adjectives in some sentence positions can also modify pronouns, as in 'He was greatly embarrassed'. Note that some things traditionally called adjectives really belong in separate classes: quantifiers, articles, or demonstratives (see below).

Quantifier

A QUANTIFIER tells how many or how much there is of an entity, e.g., *one, two, some, all, many*. A NUMERAL is a special type of quantifier which indicates a precise number: e.g., *zero, one, two, three, first, second, third*.

Adverb

According to traditional grammar, an ADVERB modifies an adjective, another adverb, a verb, or a whole clause or sentence. Adverbs are used for a variety of purposes.

(2)	Function	Example
degree	<i>very, completely</i>	
manner	<i>quickly, well</i>	
certainty	<i>possibly, certainly</i>	
circumstance	<i>early, later, there, then</i>	
negation	<i>not</i>	

In other words, adverbs are a mixed bag and probably shouldn't be grouped together. (See chapter 9 "Obliques," p. 98, for a more detailed explanation and a better analysis.)

Article

An ARTICLE helps pick out individuals designated by common nouns. In English, the DEFINITE article, *the*, indicates that the speaker assumes that the person spoken to knows which individual the noun refers to. The INDEFINITE article, *a*, indicates that the speaker does not assume that the one spoken to can identify the individual(s) the noun refers to.²

Demonstrative

DEMONSTRATIVES point out an object (person or thing) in time or space. Demonstratives may either function in place of or alongside a noun (i.e., either as a pronoun or as a modifier of a noun). English demonstratives can be classified by number and according to whether they locate the object near the speaker (PROXIMAL) or away from the speaker (DISTAL).

(3)		Singular	Plural
Proximal	<i>this</i>	<i>these</i>	
Distal	<i>that</i>	<i>those</i>	

For further discussion, see chapter 7 "Embedding and Noun Phrase Structure," p. 71.

Preposition

A PREPOSITION indicates the relation of a following noun phrase to the rest of the sentence, e.g., *to, with, from, by, for, of*. The following noun phrase is called the OBJECT of the preposition. (For further discussion, see chapter 9 "Obliques," p. 95.)

²In this book, the term ARTICLE is seldom used. Articles are usually analyzed as a special type of DETERMINER; see chapter 5 "Introduction to Syntax" (p. 39).

Indirect Object

The INDIRECT OBJECT is a phrase which typically refers to someone who receives an object or a message. It may be helpful to contrast indirect object with direct object.

- | (14) | Direct object | Indirect object |
|------|---|-----------------|
| | <i>Mike sent the package to his girlfriend.</i> | |
| | <i>Lorraine threw the Frisbee to Daniel.</i> | |
| | <i>Manuela said many kind things to me.</i> | |

In many traditional grammars, INDIRECT OBJECT does not refer to the same class of phrases that it does in this book. In this book, indirect objects in English are always prepositional phrases; elsewhere, you may find *me* in '*Manuela told me many kind things*' classified as an indirect object, but not here. Even among linguists, there is some disagreement over the analysis of clauses like these, but the analysis adopted here is probably the most widely used. To those familiar with other analyses, I apologize for the potential confusion, but in fact very little in this book depends on the issue. For further discussion and justification, see chapter 8 "Verbal Valence" (p. 77).

Transitivity

Clauses containing a direct object are TRANSITIVE.

- (15) *He appreciated her contribution.*
She needed our help.
Repeat the sentence again, please.

Clauses not containing a direct object are INTRANSITIVE.

- (16) *He moved to Arizona permanently.*
His wife died after a long illness.
He screamed at the top of his lungs.

2.5. Grammatical categories relevant primarily to nouns and pronouns

GRAMMATICAL CATEGORY refers to a way of classifying nouns, verbs, or other words, based on certain systematic differences between them, such as PERSON, NUMBER, and TENSE. The ones discussed in this section relate most directly to nouns and pronouns; those that relate primarily to verbs are discussed in the next section.

Person

PERSON indicates whether a person or thing is the speaker or the one(s) spoken to.

- (17) FIRST PERSON the person referred to is the speaker (or a group that includes him/her)
- SECOND PERSON the person referred to is the one(s) spoken to (or a group that includes the one(s) spoken to but does not include the speaker)
- THIRD PERSON the person(s) or thing(s) referred to is/are neither the speaker nor the one(s) spoken to

First person and second person are relevant only for pronouns and for prefixes and suffixes with pronoun-like meanings. Nouns in most languages are always third person.

Number

NUMBER indicates whether a noun is SINGULAR (referring to one item) or PLURAL (more than one).

- (18) Singular Plural
woman *women*
backhoe *backhoes*
revival *revivals*

Gender

GENDER refers to a semi-arbitrary classification of nouns based in part on the sex or animacy of the thing described.

- (19) MASCULINE males
 females
 NEUTER things that are neither male nor female
 ANIMATE animals
 INANIMATE plants and nonliving things

Note, however, that such classifications in languages are somewhat arbitrary; not all feminine nouns may refer to things that are inherently female and some things that are inherently female may be in some other category. For example, *Mädchen* ‘maiden’ in German is neuter.

In English, gender is relevant only to third person singular pronouns (*he*, *she*, *it*). In many other languages, nouns, too, are divided into different groups based on gender. The gender of the noun may make requirements about the form of its modifiers. For example, in French the definite article (*the*) has different forms (*le* versus *la*) depending on the noun’s gender.

- (20) *Le roi* ‘the king’ (masculine)
 la reine ‘the queen’ (feminine)

For further discussion, see chapter 19 “Case and Agreement” (p. 262).

Case

CASE refers to a way of marking a noun or pronoun to indicate its grammatical relation or other function within a clause. In English, case is relevant only to pronouns (*I*, *me*, *my*). In other languages, case is also important for nouns. (For further discussion, see chapter 19 “Case and Agreement,” p. 252.)

- (21) NOMINATIVE used for subjects
 ACCUSATIVE used for direct objects
 DATIVE used for indirect objects
 OBJECTIVE used for more than one type of object
 GENITIVE used for possessors

In any one language, cases with these names may have more uses than what are listed here.

2.6. Grammatical categories relevant primarily to verbs

Most commonly, the following terms are used to describe the morphology of verbs, but may also be applied to whole clauses.

languages can vary unpredictably and without limit. We know now that this is not true, but it did seem like it at the time.

There is a traditional way of classifying this variety, that is, a TYPOLOGY, which recognizes three main types of morphological systems, called ISOLATING, AGGLUTINATIVE, and FUSIONAL.

On one extreme are ISOLATING languages, which have the simplest possible morphology: practically every word consists of a single morpheme. Isolating languages are especially common in southeast Asia. Consider the following Vietnamese example:⁶

- (11) *khi tôi đến nhà ban tôi, chúng tôi bắt đầu làm bài.*
 when I come house friend I Plural I begin do lesson
 When I came to my friend's house, we began to do lessons.

There is no indication on the verb for person, number, or tense; *dén* could be translated either ‘come’, ‘comes’, or ‘came’, depending on context. There is similarly no distinction like that in English between ‘I’ and ‘my’; the same word in Vietnamese is used in both contexts. The grammatical meaning ‘Plural’ is expressed by a separate word, *chúng*, not an affix.

In contrast, some languages have large numbers of morphemes per word, as in the following example from Chukchi (Paleosiberian, northeastern Siberia):

- (12) *ta-meyŋ-a-levtə-payt-ärkən*
 1Sg-great-head-ache-Imperfective
 I have a fierce headache.

This one word in Chukchi corresponds to a whole sentence in many other languages. In languages like Chukchi, many sentences consist simply of a single long word. There are so many long words that it may be impractical to print literature in double-column format; too many words would be split across two lines, and some are even too long to fit in a single column!

In Chukchi, morphemes are easy to find; they are clearly separable from each other, and moreover, a given morpheme does not change very much when it appears in different contexts. Such languages are traditionally known as AGGLUTINATIVE (or AGGLUTINATING).

Other languages have several morphemes per word, but finding them is fraught with many and varied difficulties. There are not clear boundaries between morphemes, and such languages are called FUSIONAL. Most European languages are fusional. Consider, for example, the different forms of two Russian nouns. Try to make morpheme cuts before you proceed.

(13) lime-tree table

	Singular	Plural	Singular	Plural
Nominative	<i>stol</i>	<i>stoly</i>	<i>lipa</i>	<i>lipy</i>
Accusative	<i>stol</i>	<i>stoly</i>	<i>lipu</i>	<i>lipy</i>
Genitive	<i>stola</i>	<i>stolov</i>	<i>lipy</i>	<i>lip</i>
Dative	<i>stolu</i>	<i>stolam</i>	<i>lipe</i>	<i>lipam</i>
Instrumental	<i>stolom</i>	<i>stolami</i>	<i>lipoj</i>	<i>lipami</i>
Prepositional	<i>stole</i>	<i>stolax</i>	<i>lipo</i>	<i>lipax</i>

Notice that you cannot separate off one set of morphemes that indicates only number (singular and plural) and another set of morphemes that indicates only case (nominative, accusative, etc.), as you would probably be able to do in an agglutinative language. Instead, both number and case are combined into one large set of case/number suffixes, so that each suffix indicates both number and case. Further, there are different suffixes depending on the stem used; the genitive singular suffix is *-a* on *stol* but *-y* on *lip*. Again, this is unlike what normally happens in agglutinative languages, in which a single meaning tends to be represented the same way wherever it occurs.

⁶All examples in this section are from Comrie (1998:139-49).

4.7. Position classes

With this background, we're ready to take our analysis beyond identifying morphemes and their meanings, to look at the way they combine into words. We start with agglutinative languages; isolating languages don't have enough morphology to be interesting, and fusional languages involve complexities that we deal with later. Here's some data from a typical agglutinative language, Yagua (Peba-Yaguan, Peru).⁷

- (14) a. *tsanta* He plants.
- b. *tsantʃa* He weaves.
- c. *tsantariüy* He wants to plant.
- d. *tsantsariüymáā* He already wants to weave.
- e. *náāntʃatsí* She *wove* a week ago.
- f. *tsantahay* He planted yesterday.
- g. *náāntarúüyháymáā* She was already wanting to plant yesterday.
- h. *tsantfariüytsimáā* He was already wanting to weave a week ago.
- i. *náāntʃahay* She *wove* yesterday.
- j. *náānta* She plants.
- k. *tsatúmuruüy* He wants to tie together.
- l. *náātunuháymáā* She already tied together yesterday.
- m. *tsatúmuruüytsimáā* He was already wanting to tie together a week ago.
- n. *tsatímu* He ties together.

Make morpheme cuts and gloss morphemes now, before reading any further. (Remember to use pencil.)

As you can see, as many as one prefix and three suffixes occur together in the same word (e.g., sentences (14g, h, m)). And, they seem to occur in a fixed order; *-ruüy* (whenever it occurs) immediately follows the stem and *-máā* (whenever it occurs) always comes last. This ordering can be represented succinctly in a type of chart called a POSITION CLASS CHART.

	-1	0	+1	+2	+3
Person	STEM	Desiderative	Tense		Aspect
<i>náā-</i>	she	<i>-ruüy</i> want	<i>-háy</i> recent past (yesterday)	<i>-máā</i> perfect	
<i>tsa-</i>	he		<i>-tsí</i> distal past (last week)		

Each column in the chart represents a POSITION CLASS, a class of morphemes that occur in the same position in a word. Any two morphemes that occur in the same position class will never occur in the same word; they are MUTUALLY EXCLUSIVE. (If they did occur together, one would have to

⁷Data from Elson and Pickett (1983:10-11), originally from Tom Payne. Transcription converted to IPA. Acute accent represents high tone; the slight variations in tone marking are not explained in the source, and I ignore them here.

precede the other, so they would be in different position classes.) When two affixes from different classes do appear together, they appear in the order given in the chart. It is convenient to assign a number to each position class. One way of doing this starts with the stem as class 0 and numbers out from there, with prefix classes receiving negative numbers and suffix classes receiving positive numbers.⁸ (Thus, if there are 3 prefix classes, the first (leftmost) one is numbered -3.)

The labels on the columns give some indication of the meanings of the affixes in that position class. The label ‘Desiderative’ refers to the meaning of *-ruūy*, ‘want’. The label ‘Tense’ is based on the assumption that there are two past tenses, one for recent past (yesterday) and one for distant past (a week ago or more). The label ‘Aspect’ is used because ‘Perfect’ (the gloss of *-mää*) is often classified as a type of aspect. Note there is also a column for the stem, but no stems are actually listed; being an open class, there may be hundreds of them, so we don’t clutter the chart with them. How do you make a chart like this? Start out with a small number of columns. (Hint: start with the same number of prefix columns as the *maximum* number of prefixes on any one word in the data, and do likewise for suffixes. For Yagua, you would start with one prefix class and three suffix classes. Often it turns out that this is an accurate estimate.) Then, ask yourself questions like these:

- (15) a. Which affixes (whenever they occur) are always first or last in the word, or always immediately precede or follow the stem?
- b. Which affixes are next in line in these positions, after the affixes found in question 1?
- c. More generally, for each pair of affixes that can occur together in a word, which one comes first?

Make hypotheses about which affix belongs in which column. Add or subtract columns, or move affixes to different columns, as needed, until all affixes have been included. Then check your chart once again against the data to make sure it covers all the facts correctly.

Position class charts are most useful when a language’s morphology is of a certain specific type: it should be agglutinative (with several morphemes per word and clear divisions between them) and the morphemes must come in a fixed order. Chapter 10 “Inflectional Morphology” shows a way of representing morphological structure which works for all languages, including those that are not suitable for position class charts.

4.8. Review of key terms

MORPHEMES are unanalyzable (MINIMAL) clusters of SEMANTIC, PHONOLOGICAL, and GRAMMATICAL properties, often defined with the phrase ‘minimal meaningful unit’. Their meanings may be either LEXICAL (easily glossed with TRANSLATION EQUIVALENTS) or GRAMMATICAL (defined primarily in REFERENCE to larger grammatical structure). Grammatical meanings commonly found in verbs include TENSE (PAST, PRESENT, FUTURE, NONPAST, NONFUTURE, RECENT/REMOTE), ASPECT (PERFECTIVE, IMPERFECTIVE, HABITUAL, PROGRESSIVE, as well as PERFECT ‘ASPECT’), MOOD (INDICATIVE, IMPERATIVE), and AGREEMENT. Morphemes may be classified as STEMS or AFFIXES (e.g., PREFIXES or SUFFIXES), as BOUND or FREE, and as belonging to OPEN or CLOSED classes. The study of how morphemes combine to form words is called MORPHOLOGY.

The morphological systems of languages can be classified with the traditional TYPOLOGY that divides languages into three types: ISOLATING, AGGLUTINATIVE, and FUSIONAL. Word structure in agglutinative languages can profitably be presented using POSITION CLASS CHARTS, which show how affixes are grouped into POSITION CLASSES, each of which contains a set of MUTUALLY EXCLUSIVE affixes.

⁸Following Joseph Grimes 1983.

4.9. Questions for analysis

In scientific research, it is often more important to know what questions to ask than to know what the answers are. To help you get a sense of important questions in linguistics, most chapters include a section called 'questions for analysis'. The questions in these sections summarize the main things you should consider when applying that chapter to the analysis of real language data. Use them as a guide as you are learning how to do linguistic analysis.

1. What morphemes are found in the data?
2. What are their semantic properties (i.e., what do they mean)?
3. What are their phonological properties? (In the simplest cases, this question reduces to 'How are they pronounced?')
4. What are their grammatical properties? What types of words do they occur in? Do affixes group into position classes? How many classes are there, and what affixes are in each class?
5. Is the language isolating, fusional, agglutinative, or some combination of these?

4.10. Sample description

Beginning with this chapter, I include concise factual descriptions of the data covered in each chapter, which you can use as models in your own writing.⁹ The first ones are quite short, like your first attempts probably will be. See chapter 24 "Hints for Linguistic Writing" for suggestions about how to write well for linguists; linguistic description differs from other writing you may have done.

Yagua verb structure

The verbal morphology of Yagua, an agglutinative language of the Amazon basin in Peru, can be represented in the following position class chart:

Person	-1	0	+1	+2	+3	Aspect
	STEM	Desiderative	Tense			
náā-	she	-ruuy	want	-háy	recent past (yesterday)	-máā perfect
tsá-	he			-tsí	distal past (last week)	

4.11. For further reading

A more refined view of morphological typology can be found in chapter 2 of Comrie 1989. Comrie distinguishes two parameters along which morphological systems can vary; the index of synthesis (how many morphemes per word) and the index of fusion (how clear are the divisions between morphemes). The traditional isolating/agglutinative/fusional typology can be defined in terms of these two parameters.

Usually, position class charts are easy to construct, but when there are large numbers of morphemes and position classes, things can be complicated enough that a highly systematic method is needed. Such a method was developed by Joseph Grimes (1967) based on a suggestion by Frank Lister. (Lister, incidentally, was a student in one of Grimes' courses. One of the exciting aspects of linguistics is that there is relatively little distinction between students and teachers; it is not uncommon for people to make significant contributions to the field while they are still students.) This method is also described, somewhat more understandably, in Elson and Pickett (1983:12–16) and Daly, Lyman, and Rhodes (1981:171–77). It has been incorporated into a computer program called Paradigm (Joseph Grimes 1983), which also provides information about morpheme co-occurrence (which morphemes can occur with which others).

⁹I have found it helpful to ask students always to provide an informal description of any data they analyze for homework in my classes, using terminology that would be readily understood by linguists from all theoretical backgrounds.

As a challenge, see if you can add brackets to (43) and then draw a tree for it, following the same pattern as in (42).

Some languages, like Brôu (discussed earlier), use PPs as the primary means of indicating possession.¹⁸

- (44) *alik phən kɪ?*
pig of I
my pigs

Languages vary as to whether they use NPs or PPs to indicate possession; some use only one or the other, some (like English) use both.

In closing, don't confuse POSSESSION (a syntactic concept) with OWNERSHIP (a semantic concept having to do with culturally-defined control of resources). Possession is often used to express ownership, but it can also express many other relationships.

- | | |
|--|--------------------------------|
| (45) <i>the artichoke's leaves</i> | part-whole relationship |
| <i>the artichoke's seedlings</i> | kinship |
| <i>the artichoke's garden-mate</i> | social relationship |
| <i>the artichoke's tremendous size</i> | abstract quality |
| <i>the artichoke's startling achievement</i> | initiator of an action |
| <i>the artichoke's ultimate destruction</i> | the item affected by an action |

7.6. Demonstratives

Most languages have a set of words with meanings like *this*, *that*, *these*, and *those*. These are called DEMONSTRATIVES because they 'point' at their referent. Demonstratives frequently occur as determiners within a noun phrase.

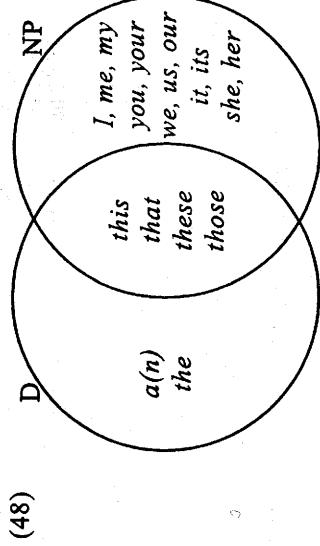
- (46) *that artichoke*
these silly, repetitive examples

In English and many other languages, the same words can also be used in place of a whole noun phrase.

- (47) *That is an artichoke.*
I'd like these.

Apparently, demonstratives in such languages are both determiners and pronouns. Thus, in our formal grammar, they should be listed in the lexicon both in the category D and in the category NP. In other words, these two groups of words overlap.

¹⁸Data from Miller 1964:74. Since the phrase structure rule posited earlier already includes a PP after the noun, it already accounts for such phrases. It does not distinguish possessors from other types of PPs; however, this could be handled by introducing the feature [Poss] in the lexical entry for *phən* 'of'. This feature would be added to the tree when *phən* is inserted and would then be copied to the PP node by means of feature percolation (see chapter 19 "Case and Agreement," p. 256). So, it is unnecessary to include [Poss] in the NP phrase structure rule for the data shown here, although this might be necessary when other examples from Miller 1964 are considered. Further, the preposition is optionally omitted if the possessor immediately follows the head noun, as shown in (28). This could be accounted for by positing a silent variant of *phən* in the lexicon, with a restricted distribution requiring it to occur immediately following a head noun.



In our formal grammars, we can represent this as follows. List the demonstratives apart from the other Ds and NPs, like a separate category, and use braces in its title to show that these words can be inserted either at a D node or an NP node in a tree.¹⁹

- (49) {D/NP}
this/these
that/those

Warning: don't follow this advice without examining the data. In some languages, there are two different types of demonstratives. One type belongs to the category D, the other type belongs to the category NP. In these languages, the two categories D and NP do not overlap, and you should not analyze demonstratives as in (49). Just include the demonstrative determiners with the other Ds and the demonstrative pronouns with the other NPs.

7.7. Review of key terms

The essence of phrase structure is the EMBEDDING of one phrase inside another. Each PHRASAL CATEGORY contains one and only one HEAD. The set of all categories that are heads or have heads is called the set of MAJOR CATEGORIES; those word-level categories which do not serve as heads are called MINOR CATEGORIES. According to the CONSTRAINTS ON PHRASE STRUCTURE RULES presented in (30), there is a one-to-one relationship between phrases and heads, and within a phrase all constituents except the head must be either MINOR CATEGORIES or must themselves be PHRASAL CATEGORIES. In essence, these express our understanding that embedding is the norm for phrase structure in language.

Within noun phrases, languages typically have one or more of the following types of constituents: DETERMINERS (D), QUANTIFIER PHRASES (QP), ADJECTIVE PHRASES (AP), PREPOSITIONAL PHRASES (PP), or other noun phrases. Adjective phrases and quantifier phrases may in turn contain DEGREE PHRASES (DegP), which are headed by DEGREE WORDS (Deg). DEMONSTRATIVES generally are either determiners and/or pronouns. POSSESSION, not to be confused with OWNERSHIP, is typically indicated by the embedding of a POSSESSOR, which is either an NP or a PP, inside a larger POSSESSED NP.

7.8. Questions for analysis

1. Consider the following questions for each distinct type of phrase:
 - a. What constituents can it have besides its head?
 - b. Which of these constituents can themselves be phrases? That is, what are the possibilities for embedding?

¹⁹An alternative analysis of demonstrative pronouns is that they are simply determiners modifying silent noun heads.

	Pattern A (no pro as DO)	Pattern B (DO can be <i>pro</i>)
intransitive verbs	never have overt DOs	never have overt DOs
$V_1 \text{ SUBCAT} \langle \text{NP}[\text{Su}] \rangle$	always have overt DOs	optionally have overt DOs; any ‘missing’ DO always has a pronominal meaning
$V_1 \text{ SUBCAT} \langle \text{NP}[\text{Su}], \text{NP}[\text{DO}] \rangle$	may or may not have overt DOs	optionally have overt DOs; in some cases a ‘missing’ DO has a pronominal meaning while in other cases it has no meaning at all
verbs that can be either $V_1 \text{ SUBCAT} \langle \text{NP}[\text{Su}], (\text{NP}[\text{DO}]) \rangle$		

If you find three classes of verbs as in pattern A, assume that the language does not allow *pro* to be a direct object (although it might still be possible as subject). If you find three classes as in pattern B and there are no verbs that always have a direct object, assume that *pro* can be a DO in this language. By allowing the possibility of having silent pronouns in our analysis, we can distinguish easily between the two types of languages.

One final variation on these patterns is that some languages do not have verbs in the third class. For example, in Turkish, even the verb ‘eat’ must always have a direct object. If necessary, a generic word is used as the direct object, one that adds little meaning other than that provided by the verb itself.²⁵

- (43) *Adam yemek yedi.*
man food he ate
The man ate.

- (44) *Orhan yazi yazdi.*
Orhan written.thing wrote
Orhan wrote.

8.7. Typological notes: Marking of subjects and objects

There are several strategies that languages use to distinguish subjects, direct objects, and indirect objects from each other. We just saw one: sometimes languages can omit subject pronouns but not direct object pronouns. There are several other common ways that subjects and objects differ.

Some languages, like Choapan Zapotec and Kalkatungu (below), rely almost entirely on word order and meaning to indicate which NP in the clause is subject, which is direct object, etc.

- (45) *janiku ntia aña wampa*
white.man money gave girl
The white man gave money to the girl. (Kalkatungu, Australia)²⁶

Some languages use special morphological forms, called CASES, for nouns and pronouns, to indicate their grammatical relation. For example, in (46), the suffix glossed ‘Dative case’, shows that *ba* ‘you’ is the indirect object. In (47), the suffix glossed ‘Accusative case’ shows that *filia* ‘daughter’ is the direct object.

²⁵ Data from Underhill 1976:52.

²⁶ Data from Blansitt 1986:33, citing Blake 1969:35. The transcription has been standardized.

- (46) *cha kle bla kete ba-ge*
I be banana give you-Dat
I am giving bananas to you.

- (47) *mater filia-m amat*
mother daughter-Acc loves
The mother loves the daughter.

(Latin)

English shows another type of case-marking, involving special forms of pronouns. For example, the first person singular pronoun has three forms: *I* is in nominative case (used for subjects), *me* is in objective case (used for direct and indirect objects), and *my* is in genitive case (used for possessors). (We come back to case marking again in chapter 10 “Inflectional Morphology” (p. 128) and especially chapter 19 “Case and Agreement” (p. 253), so don’t bother to learn the names for the different cases now. The important point is knowing what case is and the role it plays in marking grammatical relations.)

Other languages (including English) mark the indirect object with a preposition, but leave the subject and direct object as bare noun phrases. (Another way to say this is that the indirect object in languages like English is a prepositional phrase.)

- (48) *Lakilaki itu mengirim surat kepada wanita itu* (Indonesian, Austronesian)²⁸
man the sent letter to woman the
The man sent a letter to the woman.

- (49) *i³-xi² piy³-bit² o³-ñim² je³-it² pe³* (Munduruku, Brazil)²⁹
mother food she.gave her.son to
The mother gave her son food.

(When a ‘preposition’ follows its NP, as in (49), it is called a POSTPOSITION, and the PP is called a POSTPOSITIONAL PHRASE.)

Some languages even use PPs for direct objects, at least in some circumstances. (Here, I’ve used the gloss ‘P’ for prepositions that have no equivalent in English.)

- (50) *Marta ve /_{PP} a Jorge J.* (Spanish)³⁰
Marta sees P Jorge
Marta sees Jorge.

- (51) *Ng madakt /_{PP} ar a dərumk J.* (Palauan, Austronesian, Belau)³¹
He fears P thunder
He is afraid of the thunder.

We return to this topic in chapter 21 “Passive and Voice,” where it becomes crucially important to know what devices a language uses to mark different grammatical relations.

²⁷Data from Blansitt 1986:32, citing Gunn 1975:95.

²⁸Data from Chung 1983.

²⁹Data from Blansitt 1986:32, citing Crofts 1973:50.

³⁰Actually, the situation in Spanish is a bit complicated. PPs are used for direct objects in Spanish that are definite persons; otherwise NPs are used (i.e., the preposition *a* is omitted). I won’t try to accommodate this type of fact in grammars in this book; the point here is to highlight some of the ways in which languages differ from each other.

³¹Palauan data from Josephs 1975:48. The rule governing the preposition *ar* is similar to Spanish; it is required (but only with imperfective verbs) when the object is specific. The word *a* in Palauan has no discernible meaning; it simply occurs at the beginning of most noun phrases.

Obliques usually express a wider range of meanings than do subjects and objects. Some express secondary characters.

- (1) **BENEFACTIVE**
He opened the door [for the small children].
Many nations provided funding [for the Persian Gulf War].
- (2) **ACCOMPANIMENT**
He ate dinner [with his guests].
Why don't you come [with me]?
He mixed the dirt [with the manure].
- (3) **INSTRUMENT**
Many people have learned to eat [with their fingers].
The mill will not grind [with water that is past].
We make a living [by what we get], but we make a life [by what we give].
Let's go [on foot].
- (4) **TIME**
He awoke [late the next morning].
It has been too hot [the whole week].
Most successful sales come [after the fifth call].
- (5) **LOCATION (of an object or an entire event)**
He saw the dog [over there].
Little Miss Muffet sat [on a tuffet].
- (6) **SOURCE (starting point)**
He came [from Alabama].
Please take that silly hat [off your head].
- (7) **PATH**
They passed [through many twisted corridors].
One if [by land], two if [by sea]...
- (8) **GOAL (endpoint)**
He ran [to the back of the room].
I'll put the letter [on your desk].

Others express locations of part of an event (with verbs of motion).

- (9) **MANNER**
[Very slowly], she backed out of the cage.
Sergeant Pepper dismantled the bomb [with great care].
[Unfortunately], we cannot retain your services.

This is not in any way a complete list, but it covers the most common types. (A few more are introduced in chapter 22 "Embedded Clauses.")

9.3. Syntactic structures used for obliques

The meanings of obliques are comparable to semantic roles like agent, patient, experiencer, and recipient (see chapter 8 “Verbal Valence,” pp. 80ff.). All languages have some way of expressing ideas such as the oblique meanings illustrated above, but as you can see just from these examples, there can be a variety of structures used to express any one idea. When you analyze a particular language, you need to examine its grammatical structures and then determine for each structure what range of meanings it can convey.

So, let’s take a look at the variety of structures used for obliques, in English, Swahili (Bantu, East Africa),² and a few other languages. You will find that the internal structure of obliques is far more important than their distribution for establishing their syntactic category. This is because many obliques have the same distribution, but quite different internal structures.

Prepositional and postpositional phrases

The structure most used for obliques is a noun phrase combined with a PREPOSITION (such as *under*) to form a PREPOSITIONAL PHRASE (PP), such as *under the table*. Conversely, the most common use of PPs is as obliques, although we have also seen them used as possessors, indirect objects, and even direct objects. Languages tend to package noun phrases with attached prepositions (i.e., inside PPs) according to the following ranking:

- (10) Su DO IO Oblique
 NP ←————→ PP

That is, PPs are used commonly for obliques, somewhat less often for indirect objects, quite rarely for direct objects, and almost never for subjects.³

The preposition plays a pivotal role in a PP, since it signals the relationship of the NP to the rest of the clause. For example, it is the preposition in the following examples that indicates the grammatical relation or meaning of the phrase.

- (11) *to Arthur* Indirect Object or Goal
for Arthur Benefactive
with Arthur Accompaniment
on Arthur Location
through Arthur Path

Of course, the meaning of the noun phrase itself also plays a factor, since many prepositions can be used to express more than one meaning:

- (12) *in three minutes* Time
in the candy dish Location
in a huff Manner

Swahili likewise uses prepositions for some obliques.

²Data from Vitale 1981 and Wilson 1985.

³This ranking, sometimes called an obliqueness hierarchy or relational hierarchy, has figured very importantly in certain theoretical frameworks, such as Relational Grammar, Lexical-functional Grammar, and Head-driven Phrase Structure Grammar. Pollard and Sag (1987:118–20) point out its empirical importance, which is far greater than the rather fuzzy correlation noted here. A variant of it comes back in chapter 23 “Relative Clauses,” pp. 340 as the Noun Phrase Accessibility Hierarchy.

- (13) *Fatuma alisimama kwa muda mrefu*
 Fatuma she.stood for period long
 Fatuma stood for a long time.

- (14) *Goal* *Time*
aliingia katika chumba kikubwa kile
 he.entered into room large that
 He entered that large room.

- (15) *Source*
kitabu kilianguka kutoka rafu-ni
 book it.fell from shelf-Loc
 The book fell from the shelf.

- (16) *Instrument*
chakula kiliwa kwa kijiko
 food it.was.eaten with spoon
 The food was eaten with a spoon.

- (17) *Manner*
alikwenda kwa haraka
 he.went with haste
 He went quickly.

One Swahili preposition, *kwa*, is used to express several different meanings. Besides what are illustrated here ('for' and 'with'), it can mean 'to', 'towards', 'by', and 'means of'. Structures used for obliques, such as particular prepositions, typically do not line up in a neat one-to-one fashion with the list of possible oblique meanings. That list is designed for use with all languages, but each language packages these meanings in its own way.

So, when you are analyzing prepositions, or more generally, when you are analyzing any obliques, you have to examine each structure's range of uses and compile a list of its different meanings. One consequence of this is that it is often impossible to translate obliques in a literal word-for-word fashion. You have to determine what the oblique means, then find a suitable structure in the target language that expresses the same idea.

In some languages (especially SOV languages), 'prepositions' follow the NP, in which case they are called POSTPOSITIONS (also P) and the phrases are called POSTPOSITIONAL PHRASES (also PP). For example, Turkish uses postpositions, not prepositions.⁴

- (18) *Ahmet / etmek için / diktkâna gitti.*
 Ahmet bread for to.store went
 Ahmet went to the store for bread.

- (19) *Cevabi / yarına kadar / bulurum.*
 answer tomorrow until I.will.find
 I'll find the answer by (lit., until) tomorrow.

But, aside from word order, postpositions are identical in function to prepositions; they serve as a link between the NP and the rest of the clause.

In a PP, the preposition or postposition is the head and the noun phrase is called its OBJECT. The structure that is generally assumed for PPs is as follows:

⁴Data from Underhill 1976:157, 171.

- (46) a. If it looks like an NP, it's probably an NP.
 b. If it looks like an NP, with the addition of some extra word at the beginning or end, that extra word is likely a P and the whole phrase a PP. This is especially likely if the extra word expresses the relationship of the NP to the larger context.
 c. If it contains nothing that is likely to be a noun or pronoun, it's probably not an NP or PP.
 d. If it contains a modifier that could be a degree word, it's likely an AdvP and the word that is modified is likely an Adv.
 e. If it is a single word, it is probably not a preposition, since prepositions normally have objects. (It might, however, be an unmodifiable single word that could be classified as a PP.)
 f. If it is more than one word, it is not an unmodifiable single word. (Okay, that may seem obvious now, but sometimes it's hard to keep track of the obvious when you're doing analysis in a new language.)
 g. If it looks a great deal like what you've come to recognize as an NP or AdvP, but there are some minor differences from the norm, consider listing it in the lexicon as an idiom.

As we've seen, there is no neat one-to-one relationship between the different structures used for obligues in a language and their meanings. This means that you won't find neat correspondences between two languages either, so make sure you analyze the language, not the glosses! Just because one language uses a PP to express a certain idea, doesn't mean other languages will.

As you make hypotheses about the different obligues you find, you may find it useful to construct a chart something like the following to keep track of which structures are used for which oblique meanings. This one summarizes the facts we've seen in Swahili:

	PP	NP	AdvP	unmodifiable word
Ben	X			
Acc				
Instr	kwa			
Time	kwa	X?		X?
Loc		-ni		
Src	kutoka			
Path		-ni		
Goal	katika	X (-ni)		X?
Manner	kwa		X	

The X's mean that an example of a structure with a particular meaning has been found. If there is a specific morpheme (such as a preposition or suffix) that seems important to the meaning, you can put it in the chart too. If there is more than one such morpheme for a given meaning, list them all. Use a question mark for examples that you're less sure about. (Note that idiomatic phrases are left out of the chart; they need to be treated individually.)

This will help you to spot cases where the same structure (or the same morpheme) is used with more than one meaning. It will also help you find the different ways available for expressing the same idea (which could come in very handy when translating materials into the language). Note, however, that it is primarily for your own use in developing an analysis; it may not be very useful to your readers in a final write-up.

Be sure to compare the structures used for obligues with those used for subjects and objects too. For example, in some languages the notion benefactive may be expressed in the same way as the notion recipient. That is, there may be no difference in form between such phrases as *to the teacher* and *for the teacher*; the same preposition will be used for both or both will be bare noun phrases.

The third difference between inflection and derivation is that different inflected forms of a word can usually be usefully organized into a type of chart called a PARADIGM. For an example of this, we need to go to Spanish, a language that has more inflectional morphology than English. Most books about Spanish present the different forms of verbs in paradigms, like the following chart of the present tense of *andar* 'to walk'.

	Singular	Plural
first person	<i>ando</i>	<i>andamos</i>
second person	<i>andas</i>	<i>andáis</i>
third person	<i>anda</i>	<i>andan</i>

There are two GRAMMATICAL CATEGORIES shown in this paradigm: person and number. Each form is specified by some unique combination of some person and some number. For example, we identify *andas* as the second person singular present form of *andar* and *andan* as its third person plural present form. The two grammatical categories allow us to classify the six forms in a systematic way. This is the nature of paradigms: they consist of a set of forms which are cross-classified by means of a set of grammatical categories.

This cross-classifying of forms in paradigms is characteristic of inflectional morphology, but not of derivational morphology. Derivational morphology groups words together into pairs, like *march* and *marcher*, *farm* and *farmer*, but never into larger sets like the one in (4).³ Usually, the paradigms of different verbs in a language are similar or identical.⁴ For example, person and number are relevant to all verbs in Spanish, not just *andar*. This means we can uniquely identify every form of every verb by specifying the verb involved and some combination of these grammatical categories.

Often the affixes used on different verbs are very similar to each other. Thus, we can generalize (4) to represent the inflection of all verbs like *andar* by only showing the affixes involved.⁵

	Singular	Plural
first person	-o	-mos
second person	-s	-ís
third person	-Ø	-n

A chart like this is an important step toward writing rules that account for the inflection of this class of verbs.

Actually, we need to distinguish two senses of the word PARADIGM. In one sense, a paradigm is a chart like (4) or (5), an informal means of presenting data, a way of using ink on paper. In the other sense, a paradigm is the set of inflected forms of some word, a set which is cross-classified by means of grammatical categories. A paradigm in this sense is part of the structure of the language itself. The two senses are related; a paradigm in the first sense is a way of representing a paradigm in the second sense.

The differences between inflection and derivation that we have discussed are summarized in the following chart:

³Of course, some inflectional paradigms may have only two forms, such as the singular and plural forms of nouns in English. So, just because forms come only in pairs does not mean that they are derivational. However, when you find any larger set of related forms, especially if more than one dimension (grammatical category) is involved, the type of morphology is almost certainly inflectional.

⁴The phonological material used to express the grammatical categories may differ for different verbs (see chapter 12 "Suppletion and Morphophonemics"), but generally the same grammatical categories will be relevant to all words in a class.

⁵I am assuming that the final /a/ of the stem deletes before the /o/ in the first singular and have adopted an informal notation for the stress shift to the final syllable of the stem in the second plural forms.

system, we need an organized representation of the grammatical categories involved and their relationships. That is, the different endings should be arranged into paradigms, like the one in (5). So, one advantage of using paradigms is that they help make sense of complex morphological systems with irregular verbs or in fusional languages, when position class charts aren't very helpful. But even in agglutinative languages, paradigms can be helpful; you have probably already discovered that it is easier to identify inflectional affixes and set up position class charts if the different forms of each word have first been arranged in some systematic way, in other words, organized in paradigms.

Thus, the first thing you should do when studying inflectional morphology is to arrange the data in paradigms and determine the grammatical categories involved. Only then make morpheme cuts. In some cases, you can carry the analysis further and construct position class charts; in other cases, this is impossible or unrevealing.

We have been contrasting two different perspectives on morphology.⁶ One, called ITEM AND ARRANGEMENT, looks at morphology by focusing on individual morphemes, like beads on a string. Position class charts are examples of this perspective. The other perspective, called WORD AND PARADIGM, looks at morphology by focusing on grammatical categories and paradigms. It is useful for inflectional morphology in all types of languages and especially in fusional languages or where there is a lot of irregularity.

This book uses both perspectives. Item and arrangement can be very useful for presenting data informally, in position class charts, in some types of languages. But, since word and paradigm is useful for all languages, it is the basis for our formal grammars. Specifically, our grammars describe morphology using a theoretical framework that has been variously called Extended Word and Paradigm or A-morphous Morphology (Anderson 1982, 1992). The rest of this chapter explains how to handle inflectional morphology within such a system.⁷

10.4. Grammatical categories and inflectional features

The essence of the word and paradigm perspective is its focus on GRAMMATICAL CATEGORIES, like person, number, and tense. What exactly is a grammatical category? Grammatical categories are sets of abstract elements (like singular and plural) which are MUTUALLY EXCLUSIVE. Each grammatical category has only a small and fixed number of elements, that is, grammatical categories are closed classes. Thus, a grammatical category is a small, closed class of mutually exclusive grammatical properties.

How do we add this concept to our formal grammars? We do so with FEATURES. For example, to represent that a word is singular, we might assign it the feature [-plural]. To represent that a word is third person, we might assign it the feature [3 person].⁸ In these examples, 'plural' and 'person' are the names of the features, '-' and '3' are their VALUES.

We've encountered features occasionally so far in this book. For example, in chapter 6 “The Base” (p. 54), we used a feature [\pm common] to distinguish proper and common nouns. We have also used a feature [GR] to represent grammatical relations, with values [Su], [DO], etc. You may also have encountered features in phonology to represent the similarities and differences of sounds; for example, all sounds made with rounded lips have the feature [+round]. In the same way, features are used in morphology to represent the similarities and differences of words within a paradigm. Just

⁶The terminology comes from Hockett 1954, which is primarily concerned with contrasting item and arrangement with item and process (which in the discussion in this book has been subsumed under the heading item and arrangement), and which mentions word and paradigm only in passing.

⁷I have found that Anderson's approach to morphology is useful pedagogically because it provides formal expression for several important traditional concepts, such as paradigms, grammatical categories, inflection versus derivation, etc., and so reinforces them. As a bonus, it prepares people to understand transformational rules by showing how a sentence is built in several steps, even while students are still absorbing the basics of phrase structure. Although not without its critics (see for example Scalise 1984:191–97), it does reasonably well on descriptive adequacy, since it handles irregular verbs, morphology in fusional languages, and complex types of affixation more easily than item-and-arrangement approaches.

⁸I use the term ‘feature’ variously to refer either to the name of a feature ([plural]), to a specific feature value specification ([+plural]), or to the set of possible values for a feature ([\pm plural]). Context should always clarify which is meant.

as different combinations of phonological features represent different sounds, different combinations of morphological features represent different forms in a paradigm.

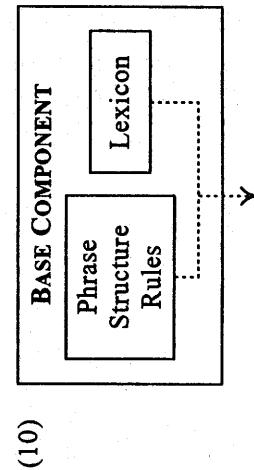
Many features are BINARY; they have only two values, usually '+' and '-'. Often this is because they represent grammatical categories that contain only two elements. For example, number (singular versus plural) is often represented with the binary feature [±plural]. At other times we deal with grammatical categories that have more than two elements, so a single binary feature will not do. The inflectional category of person is like this; it normally has three choices: first, second, and third person. One way to represent this formally is with a single feature that has three values: [1 person], [2 person], and [3 person].⁹ Another is to use two binary features, such as [±me] and [±you].¹⁰ These work as follows:

- (9) first person: [+me, -you]
- second person: [-me, +you]
- third person: [-me, -you]

One of the tasks in the morphological analysis of a language is making a precise hypothesis about the grammatical categories involved by choosing an appropriate set of features. For suggested features to use to represent different grammatical categories, see the appendix to this chapter (p. 133).

10.5. Inflectional morphology in a formal grammar

Now let's see how inflectional morphology can be fit together with syntax in a formal grammar. Recall (from chapter 6 "The Base," p. 48) that the base component builds trees with phrase structure rules, then fills in the terminal nodes from the lexicon.



Morphology fits into this model by providing details about the terminal nodes.

We assume that the grammar builds a sentence in two phases: first syntax, then inflectional morphology. In the syntactic phase, the phonological material of inflectional affixes is not inserted from the lexicon. Instead, inflectional morphology is represented in trees produced by the base solely by inflectional features on preterminal nodes. For a sentence like *The boys cried*, the base would produce a tree that looks like this (using phonetic transcription to represent actual pronunciation more accurately):¹¹

⁹I follow Generalized Phrase Structure Grammar (Gazdar, Klein, Pullum, and Sag 1985:22ff) in allowing non-binary features, although my notation for them is slightly different.

¹⁰For an example of the usefulness of binary features representing person, see Matthews 1972a:105–10, where these two features, with the addition of [±plural], are used to distinguish eight person/number combinations in Huave (Mexico), including such things as first plural inclusive/exclusive, and first dual inclusive.

¹¹The broad phonetic (IPA) transcription of English in this book is based on upper Midwestern dialects in the United States, in a way that hopefully provides minimum distraction from the main point of the passage. Nothing depends on the exact choice of symbols used or the level of analysis represented.

(37) NP	
<i>aj</i>	[Nom case]
<i>mi</i>	[Obj case] $\begin{cases} \text{1 person} \\ \text{– plural} \end{cases}$
<i>maj</i>	[Gen case]

	<i>ju</i>	[Gen case]	$\begin{bmatrix} 2 \text{ person} \end{bmatrix}$

	<i>hi</i>	[Nom case]	$\begin{bmatrix} 3 \text{ person} \end{bmatrix}$
		<i>hm</i>	[Obj case] $\begin{cases} \text{– plural} \\ \text{[Gen case]} \end{cases}$
		<i>hz</i>	[m gender]

	<i>fi</i>	[Nom case]	$\begin{bmatrix} 3 \text{ person} \end{bmatrix}$
		<i>hu</i>	[Obj case] $\begin{cases} \text{– plural} \\ \text{[Gen case]} \end{cases}$
		<i>hu</i>	[f gender]

Notice that *ju* ‘you’ has been left unlabeled; this is because it is a regular form; just like nouns, it is unmarked for nominative and objective case.³⁰

To complete the analysis, it is necessary to have some way of introducing the features for case on the proper nodes in trees. This matter is covered in chapter 19 “Case and Agreement” (p. 254), where there are rules that introduce [Nom case] on a subject NP, [Gen case] on a possessor NP, etc. For now, just assume that the features are there and don’t worry about how they get there.

10.9. Summary of English grammar so far

Since we’ve added quite a few new elements to our grammar in this chapter, let’s expand the summary of English grammar presented in chapter 9 “Obliques” (p. 107) to see how everything fits together. This grammar also handles some aspects of subject-verb agreement (in light of the similar analysis of Fore).³¹

I. Base

A. Phrase structure rules

$$S \rightarrow NP[S_u] VP$$

$$VP \rightarrow V (NP[D_O]) (VP) (PP[I_O]) \left(\begin{array}{c} NP \\ PP \\ Advp \end{array} \right)^*$$

$$PP \rightarrow P NP$$

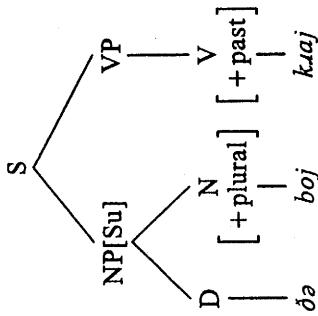
$$NP \rightarrow \left(\begin{array}{c} D \\ NP[Poss] \end{array} \right) (OP) (AP)^* N (PP)$$

$$AP \rightarrow (DegP) A$$

³⁰ Although the third singular feminine pronoun uses *her* for both objective and genitive case, it would unfortunately not work to list the form just once and leave it unlabelled for case. This would allow the regular genitive case suffix -’s to attach to it, producing an incorrect genitive form. See chapter 20 “Word Division and Clitics” (p. 285).

³¹This is not ultimately a very satisfactory way of handling agreement. For a better approach, see chapter 19 “Case and Agreement.”

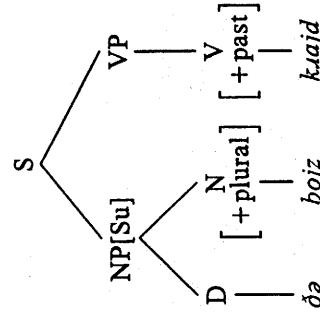
- (11) Output of the base component for *The boys cried.*



The [+ plural] on the N represents the plurality of *boys* and the [+ past] on the verb represents the past tense of *cried*.¹² The only phonological material at the terminal nodes belongs to the stems, exactly as they are listed in the lexicon. A tree structure like (11), which is produced entirely by the base component, is called its DEEP STRUCTURE.

This, of course, is incomplete. Somehow, the inflectional features need to be ‘spelled out’ by adding phonological material to the stems, such as /z/ and /d/.

- (12) Structure of *The boys cried* after spelling out inflectional affixes



This completes the process of generating the sentence. A tree structure like (12), which matches the actual sentence we are trying to produce, is called its SURFACE STRUCTURE.¹³

The rules which convert the terminal nodes from one form to the other are described in the next section.

Inflectional spellout rules

To get from (11) to (12), we need one rule that adds -z to the end of any noun that is [+ plural] and another that adds -d to the end of any verb that is [+ past]. We call these rules INFLECTIONAL SPELLOUT RULES¹⁴ and write them as follows:

¹² A binary feature for tense is adequate in English, because the morphology only distinguishes two forms: past and nonpast. Other so-called ‘tenses’ are handled by auxiliary verbs like *have* and *will*, so morphological features do not need to reflect them.

¹³ Though the term SURFACE STRUCTURE is borrowed from Transformational Grammar and Anderson’s (1982, 1992) work is couched in transformational terms, this approach to morphology could just as easily be combined with a nontransformational syntax.

¹⁴ Inflectional spellout rules are half syntactic and half phonological. Like syntactic rules, they refer to syntactic information (like ‘N’ and ‘[+ plural]’ in (13)). Phonological rules generally do not refer to syntactic information, only phonological information. However, inflectional spellout rules do not change syntactic information, only phonological information (like adding -z), and normally they act on only one word at a time. In this, they are like phonological rules and unlike syntactic rules. Inflectional spellout rules serve as a link between syntax and phonology, since they apply to the output of syntactic rules and their output becomes the input for phonological rules.

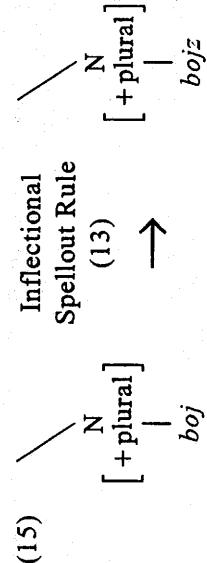
(13) Inflectional spellout rule for noun plurals

$$\begin{array}{c} \text{N} \\ | \\ [\text{X}] \end{array} \rightarrow [\text{Xz}]$$

(14) Inflectional spellout rule for past tense on verbs

$$\begin{array}{c} \text{V} \\ | \\ [\text{+ past}] \\ | \\ [\text{X}] \end{array} \rightarrow [\text{Xd}]$$

In this notation,¹⁵ the arrow represents the change made by the rule: the difference between its input (the left side) and its output (the right side). In (13), the rule takes as its input any N node in any tree which is [+ plural]. The X in brackets on the bottom line stands for whatever phonological material happens to be under that node. In the surface structure tree, this is changed slightly by adding the suffix -z.¹⁶ Look at the partial trees below.



Given this way of handling inflectional morphology, we can now refine our understanding of the lexicon. The lexicon does not need to contain the whole paradigm of a word (all its inflected forms), but only its stem, since this is what must be inserted in trees produced by the base.¹⁷ Nor does it need to list any of the inflectional affixes, since these are spelled out later. The lexicon, then, is primarily a list of just *stems*, not of all morphemes in a language.

Feature assignment rules

We've seen how inflectional morphology is represented by inflectional features in trees, how those features are 'spelled out' with phonological material, and have noted what this means for the lexicon. But, we've skipped a step: we first need to make sure the right features get in the right places in the trees produced by the base. What puts them there?

There are some features, like [±past], which apply to all members of a particular word class. All English verbs have both Present and Past forms and every verb in every tree must be labeled either [+past] or [-past].¹⁸ So we need a special type of phrase structure rule, a FEATURE ASSIGNMENT RULE, which puts the feature [past] on every verb.¹⁹

(16) V → [±past]

¹⁵This notation differs in a few details from that used by Anderson (1982), mostly in the use of brackets around feature bundles and for showing the constituent structure of phonological representations. Also, Anderson calls these rules 'Word Formation Rules', but I reserve this term for the derivational rules introduced in chapter 11 "Derivational Morphology."

¹⁶This suffix is variously pronounced [z], [s], and [iz], depending on the stem. See chapter 12 "Suppletion and Morphophonemics" (p. 153) for how to handle this variation and why only [z] is included in the inflectional spellout rule.

¹⁷This comment refers to regular inflection. For irregular inflection, see below.

¹⁸This is an oversimplification, since in some contexts, English verbs are not inflected for tense.

¹⁹The term 'feature assignment rule' is my own invention. Anderson attributes the idea for this type of rule to Chomsky's (1965:82–83) analysis of noun subcategorization. I depart from Anderson and Chomsky, though, in not positing feature assignment rules for features that are fully specified in the lexicon, as in (17). Feature assignment rules are similar to the feature co-occurrence restrictions of Generalized Phrase Structure Grammar (Gazdar, Klein, Pullum, and Sag 1985), with one difference: they build structure by adding features to trees, rather than just stating what combinations of features are possible.

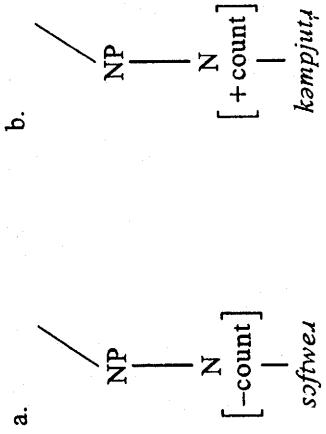
Feature assignment rules are part of the base component and work together with the other phrase structure rules and the lexicon to build the trees. This rule says to add either [+ past] or [- past] to a V node when building a tree.²⁰

Other features are more limited in their distribution. For example, [+ plural] can only be allowed to appear on certain nouns, called count nouns. Other nouns, like *software* do not have plural forms (i.e., **softwares* is not a grammatical English word).²¹ This latter group is called mass nouns. This means, before we can talk about plural forms of nouns, we must first divide the class of nouns into two subcategories in the lexicon, using the features [+ count] and [-count].

(17) N [+ count]	N [-count]
<i>səndwɪʃ</i> sandwich	<i>sənd</i> sand
<i>kəmpjʊtə</i> computer	<i>sɔfweə</i> software

When a noun is inserted in a tree, we assume it carries with it all the features that it has associated with it in the lexicon (although we don't usually write all of them).²² So, every N node in every tree ends up with a feature for [count].

(18) Partial trees, just after lexical insertion



Then, we need feature assignment rules to add the feature [\pm plural]. But, this must be done selectively; mass ([−count]) nouns must always be [−plural], while count ([+ count]) nouns can be either [+ plural] or [−plural].²³

(19) Feature assignment rules for English nouns

N[−count]	\rightarrow	[−plural]
N[+ count]	\rightarrow	[±plural]

²⁰There is a possible ambiguity in the notation here; in this formulation, feature assignment rules, unlike other phrase structure rules, do not add new nodes underneath old ones, but only add new features to existing nodes. In practice, this is not a problem, since any time the right-hand side of the arrow consists of a single set of brackets (and what they enclose), the interpretation is to add features to an existing node; in all other cases the interpretation is to add new nodes to the tree. In Chomsky's (1965:82–83) original formulation, feature assignment rules originally added new nodes, but this resulted in a ‘complex symbol’ which in later generative work has been replaced by the notion of a bundle of features at a single node. The original notation is retained, rather than some novel notation, since in practice this is only a minor problem.

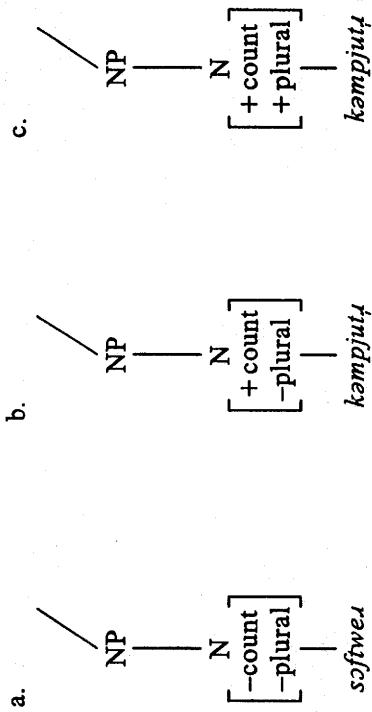
²¹Actually, this is oversimplified, since some mass nouns like *sand* do have corresponding count nouns with specialized senses. (‘The soldiers trudged over the shifting *sands*.’) The facts of the English count/mass distinction are complex, and I will not attempt to account for all of them.

²²Anderson, in his presentation of A-morphous Morphology (1982:594, 1992:90–91), assumes that lexical insertion takes place at S-structure, not deep structure. However, the discussion here is not yet assuming a transformational framework, so the approach in the text is more typical of Chomsky (1977:10) in assuming that structures produced by the base contain phonological information.

²³An alternate analysis would be not to assign any value for [plural] to nouns that are [−count]. This would correctly prevent inflectional spellout rules like (13) from applying to mass nouns. However, this would cause problems for the analysis of verb agreement, since mass nouns as subjects always require singular present tense forms of the verb. To leave mass nouns unspecified for number would allow them to co-occur with either singular or plural agreement, at least, given the approach to agreement adopted in chapter 19 “Case and Agreement.” A third alternative would be to make [−plural] be the default value for [plural], but this may simply be a notational variant of the analysis given in the text.

When these rules are applied to the partial trees in (18), the choices in them allow any of the following results:

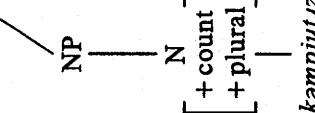
- (20) Completed deep structure trees, after all feature assignment rules have applied



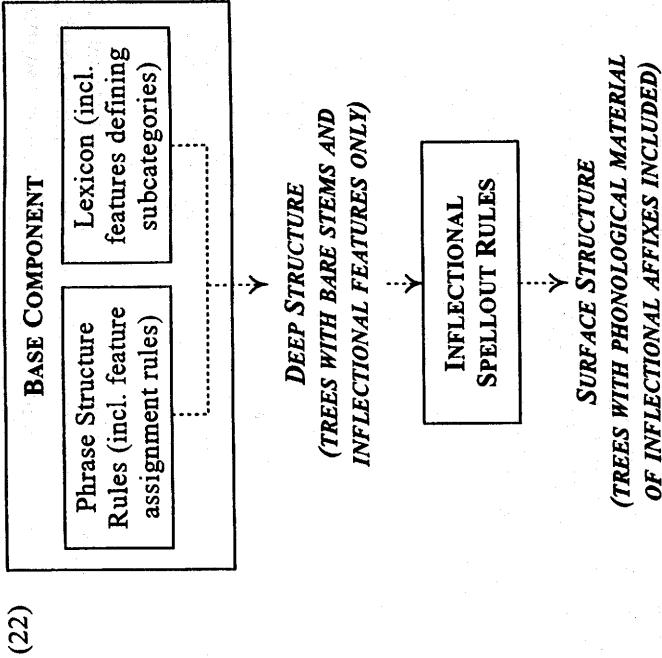
Note that (18a) could never have been assigned the feature [+ plural]; it had to become (20a).

After the inflectional spellout rule (13) applies, (20c) becomes (21); the other two, since they are marked [- plural], are left unchanged.

- (21) Surface structure for (20c)



To summarize, we can add some detail to our diagram of a formal grammar.



Some features relevant to inflectional morphology come from the lexicon along with the stems, while others are assigned by feature assignment rules. Wherever they come from, they end up as part of the initial trees built by the base. Thus, these trees contain all the information needed to specify the inflectional morphology, expressed as inflectional features. The actual phonological material of the affixes is not added until later by the inflectional spellout rules, which modify the terminal nodes produced by the base into surface structure based on the features that they find in the tree.

10.6. Building a formal analysis from a position class chart

Although position class charts are very useful in agglutinative languages, they do not tell us everything we need to know. For example, a language may have tense suffixes that are omitted in some contexts. Position class charts do not tell when to omit them, only where they occur when they are used. Position class charts also do not say anything about which affixes can co-occur with which others, only their relative order when they do co-occur. But, this type of information *should be* part of a formal grammar. So, to do a complete analysis, we should know how to take the information in a position class chart and construct a formal grammar, incorporating the extra information as we go.²⁴

Consider the following data from Fore (Trans-New Guinean, Papua New Guinea):²⁴

- | | | | |
|--------------------|--------------------|---------------------|-------------------------|
| 1. <i>natuwi</i> | I ate yesterday. | 8. <i>naturi</i> | We ate yesterday. |
| 2. <i>nagasuwı</i> | I ate today. | 9. <i>nagasuni</i> | We ate today. |
| 3. <i>nakuwi</i> | I will eat. | 10. <i>nakuni</i> | We will eat. |
| 4. <i>nata'ni</i> | You ate yesterday. | 11. <i>nagatusi</i> | We two ate today. |
| 5. <i>nata'naw</i> | You ate yesterday? | 12. <i>nakusi</i> | We two will eat. |
| 6. <i>nakiyi</i> | He will eat. | 13. <i>naturi</i> | They ate yesterday. |
| 7. <i>nakiyaw</i> | He will eat? | 14. <i>natsi</i> | They two ate yesterday. |

Notice some new variations on familiar grammatical categories. There appear to be two different past tenses, one used for events earlier today, the other for events yesterday (and possibly earlier).

²⁴Data from Merrifield, et al., 1987, #26.

This is not uncommon: to divide past time into two tenses, often called recent and remote past. Also, the category of number has three values: singular, dual (two), and plural (three or more).

After making morpheme cuts, we can summarize the morphology in the following position class chart:²⁵

(23)	Stem	+1: Tense	+2: Subject Agreement	+3: Mood
	-gas	recent past	-u ^w	first singular
	-t	remote past	-us	first dual
	-k	future	-un	first plural
			-a n	second singular
			-iy	third singular
			-a s	third dual
			-a w	third plural

How do we construct a formal analysis of this data, building on what we know from the position class chart?

First, we decide on a set of features: both what features to use and how many values there are for each. For tense, we posit two features. One is [tense], with the three values [past tense], [pres tense], and [fut tense]. We assume that near and far past are distinguished by a feature of [\pm recent] and that this feature is used only on verbs that are [+ past]. (Note that already we are clarifying something that is left unclear in (23), we are assuming that there is a present tense in Fore, even if we don't know how it is indicated.)²⁶ Similarly, for subject agreement we assume a feature of person, with three values: [1 person], [2 person], and [3 person]; and a feature of number, with three values: [sg number], [dl number], and [pl number].²⁷ Finally, for mood, we assume one feature with two values: [\pm interrogative].

Second, we incorporate these assumptions into feature assignment rules, as a formal hypothesis about Fore grammatical categories. To do so, we must consider which features can occur with which other features. Since the position class chart does not provide this information, we look back at the data. It appears that person, number, tense, and mood can freely occur with each other in any combination, except that [recent] is relevant only for [past tense]. (We don't have examples of all the combinations, but it is reasonable to assume that they can all occur, so we make that prediction in our analysis.)

(24) Feature assignment rules for Fore

$$V \rightarrow \left[\begin{array}{l} \{1, 2, 3\} \text{ person} \\ \{\text{past, pres, fut}\} \text{ tense} \\ \{\text{sg, dl, pl}\} \text{ number} \\ \pm \text{ interrogative} \end{array} \right]$$

$\text{V[past tense]} \rightarrow [\pm \text{ recent}]$

Notice the use of braces to list the possible values of features like person which have more than two values. The braces mean the same thing here as they do in regular phrase structure rules; they enclose a set of choices. The choices can either be arranged horizontally (as above, with commas separating them) or vertically.

²⁵Although the agreement markers might conceivably split into separate morphemes for person and number, further division does not yield a simpler or more general analysis. The partial regularities are probably due to the history of the system, not to its current structure.

²⁶Although we may have made the assumption that there is a present tense in Fore when we constructed (23), the chart itself does not make it explicit.

²⁷This approach to features for agreement is adequate for the moment, but requires refinement if a verb agrees with two distinct nominals or if a noun shows agreement with a possessor. For now, if data like this must be analyzed in class, I suggest that you use makeshift features like $\{\{1, 2, 3\} \text{ personSg}, \{\pm \text{ pluralDO}, \{\pm \text{ personPl}\} \text{ personPoss}\}\}$. However, this is only slightly better; the best solution is in chapter 19 "Case and Agreement" (p. 259ff.).

$$(25) \quad V \rightarrow \left[\begin{array}{c} \left[\begin{array}{c} 1 \\ 2 \\ 3 \end{array} \right] \text{ person} \\ \left[\begin{array}{c} \text{sg} \\ \text{dl} \\ \text{pl} \end{array} \right] \text{ number} \\ \left[\begin{array}{c} \text{past} \\ \text{pres} \\ \text{fut} \end{array} \right] \text{ tense} \\ \pm \text{ interrogative} \end{array} \right]$$

The first notation is more compact; that's the only difference.

Third, we write inflectional spellout rules to add the phonological material for each affix. In simple cases like this, there is one rule for each affix. The rules for the first position class are as follows:

- (26) a. $\begin{matrix} V \\ [+] \text{ recent} \end{matrix} \rightarrow [X_{gas}]$ (for recent past tense)
- b. $\begin{matrix} V \\ [-] \text{ recent} \end{matrix} \rightarrow [X_t]$ (for remote past tense)
- c. $\begin{matrix} V \\ [\text{fut tense}] \end{matrix} \rightarrow [X^k]$ (for future tense)

Notice that we did not need to refer to the feature [past tense] in (26a, b), because the feature [recent] only appears on verbs that are [past tense]. Thus (26a, b) can apply only to past tense forms.

The rules for the second suffix class are similar. Following is a partial list.

- (27) a. $\begin{matrix} V \\ \left[\begin{array}{c} 1 \text{ person} \\ \text{sg number} \end{array} \right] \end{matrix} \rightarrow [X_{uw}]$ (for first person singular subject agreement)
- b. $\begin{matrix} V \\ \left[\begin{array}{c} 1 \text{ person} \\ \text{dl number} \end{array} \right] \end{matrix} \rightarrow [X_{us}]$ (for first person dual subject agreement)
- c. $\begin{matrix} V \\ \left[\begin{array}{c} 2 \text{ person} \\ \text{sg number} \end{array} \right] \end{matrix} \rightarrow [X_{a \cdot n}]$ (for second person singular subject agreement)
- d. ... (similarly for other agreement markers)

And now, the rules for the third class of suffixes.

- (28) a. V
 $[+ \text{interrogative}] \rightarrow [Xaw]$ (for interrogative mood)
- b. V
 $[- \text{interrogative}] \rightarrow [Xi]$ (for indicative mood)

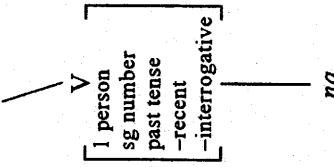
Fourth and finally, we state the order in which these rules apply, so we get the suffixes in the proper order.

- (29) The rules in (26) apply first, then those in (27), and finally those in (28).

In other words, the ordering of the rules mirrors the ordering of the position classes.

Let's see how this analysis works in a particular case, by generating the word *natuwi* 'I ate (yesterday)'. The initial tree produced by the base contains only the stem, *na*, plus a bundle of features provided by the feature assignment rules (24). These features indicate the specific form in the paradigm of *na* that will eventually be generated.

- (30) Structure for *natuwi* produced by the base

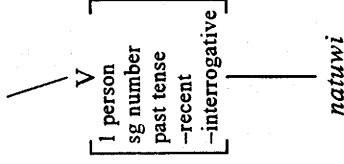


The three sets of inflectional spellout rules apply in order and add one affix at a time. The phonological form at the terminal node changes like this:

- (31) a. Output of the base: *na*
 b. Output of rule (26b): *nat*
 c. Output of rule (27a): *natu*
 d. Output of rule (28b): *natuwi*

The resulting surface structure is thus:

(32) Surface structure of *natuwi*



To recap, we need to do four things to build a formal analysis from a position class chart:

- (33) a. characterize the grammatical categories with a set of inflectional features
- b. write feature assignment rules to insert those features into trees
- c. write inflectional spellout rules to add the phonological material for each affix
- d. state the order in which the inflectional spellout rules apply

10.7. Irregular inflection in formal grammars

In every language, there are words that are INFLECTED IRREGULARLY, that is, which have some forms that do not follow the regular inflectional rules. These irregular, or SUPPLETIVE, forms must be listed in the lexicon, since they are not predictable by rule and must be learned individually. For example, the lexical entry of *go* might look like this:

(34) V

$$\begin{array}{c} go \\ \quad \left\{ \begin{array}{l} \text{went [+ past]} \\ \text{go} \end{array} \right. \end{array}$$

The part of the lexical entry that provides phonological information is more complex than we have seen previously, since it includes both the stem *go*, which is the base for the regular forms *goes* and *going*, and also the irregular past tense form *went*. How do we make sure the right form gets in the tree? There are two parts to the answer.

First, if a tree contains a V[+ past] node, the phonological form *went* must be inserted, since it is the one which carries the feature [+ past]. If, on the other hand, the tree contains a V[-past] node, the form *go* must be inserted, since the [+ past] feature of *went* is incompatible with the [-past] feature already in the tree. In other words, if there is more than one phonological form listed in a lexical entry, you insert the one that is most specific while still being compatible with the features already in the tree. This gets the right form of the verb into the tree in the first place.

Second, we also have to prevent the inflectional spellout rules from tacking on the regular past suffix to *went*, wrongly producing **wentied*. To do so, we can rely on a general convention in Generative Grammar: whenever one rule is more specific than another and both could logically be applied in a given situation, the more specific one is applied and the less specific one is skipped. A lexical entry that specifies *went* as the past tense form of *go* is more specific than the inflectional spellout rule that adds -(e)d to any verb. Thus, inserting an irregular form in a tree automatically forces the regular inflectional spellout rule to be skipped. The more specific lexical entry wins and the inflectional spellout rule is prevented from interfering. On the other hand, if a tree has a V[- past] node, so that the form *go* is inserted, the inflectional spellout rule that adds -(e)s to form

the third singular present is still allowed to apply, because the lexical entry for *go* says nothing about these other categories.

Nothing needs to be said in an individual grammar about the way suppletive stems are handled; the above ‘rules’ are assumptions about how all grammars work and thus only need to be stated once, in the general linguistic theory.²⁸

10.8. Case marking on pronouns

In many languages, nouns and pronouns are marked for CASE, which means they have a different morphological form depending on where they occur in the clause. Case occurs minimally in English. It is most clearly seen in pronouns, each of which has three different forms.²⁹ Which form you use in a particular context depends on whether it is a subject (nominative case), an object (objective case), or a possessor (genitive case).

- (35) a. *I see a tiger.*
- b. *The tiger sees me.*
- c. *It's time for my exit.*

In most languages whose pronouns are marked for case, it is impossible to identify distinct morphemes. Instead, there is one fused form that represents both the pronoun and its case, just as in irregular verbs. This is certainly true of English.

	Nominative	Objective	Genitive
first person singular	<i>I</i>	<i>me</i>	<i>my</i>
first person plural	<i>we</i>	<i>us</i>	<i>our</i>
second person	<i>you</i>	<i>you</i>	<i>your</i>
third person singular masculine	<i>he</i>	<i>him</i>	<i>his</i>
third person singular feminine	<i>she</i>	<i>her</i>	<i>her</i>
third person singular neuter	<i>it</i>	<i>it</i>	<i>its</i>
third person plural	<i>they</i>	<i>them</i>	<i>their</i>

No reasonable morpheme cutting is possible.

To account for case marking on pronouns, we can use the same mechanism as for irregularly inflected verbs. We list each form of a pronoun in the lexical entry and label it with a feature of [case].

²⁸Anderson (1986) presents the two principles relied on here: the precedence of irregular lexical entries over regular inflectional spellout rules and the precedence of the more specific stem in lexical insertion over a less specific one. Both are instances of the disjunctive ordering principle first described by the ancient Sanskrit grammarian Panini (Anderson 1982:293) and commonly referred to today as the ‘Elsewhere Condition’.

²⁹The possessive case suffix -'s used on nouns is discussed in chapter 20 “Word Division and Clitics.”

(37) NP	
<i>aj</i>	[Nom case]
<i>mi</i>	[Obj case] { [1 person [Gen case]] [- plural]
<i>maj</i>	

<i>ju</i>	[Gen case] { [2 person]
-----------	-------------------------

<i>hi</i>	[Nom case] { [3 person]
<i>hum</i>	[Obj case] { [- plural]
<i>hiz</i>	[Gen case] { [m gender]

<i>fi</i>	[Nom case] { [3 person]
<i>hu</i>	[Obj case] { [- plural]
<i>h<u>f</u></i>	[Gen case] { [f gender]

Notice that *ju* ‘you’ has been left unlabeled; this is because it is a regular form; just like nouns, it is unmarked for nominative and objective case.³⁰

To complete the analysis, it is necessary to have some way of introducing the features for case on the proper nodes in trees. This matter is covered in chapter 19 “Case and Agreement” (p. 254), where there are rules that introduce [Nom case] on a subject NP, [Gen case] on a possessor NP, etc. For now, just assume that the features are there and don’t worry about how they get there.

10.9. Summary of English grammar so far

Since we’ve added quite a few new elements to our grammar in this chapter, let’s expand the summary of English grammar presented in chapter 9 “Obliques” (p. 107) to see how everything fits together. This grammar also handles some aspects of subject-verb agreement (in light of the similar analysis of Fore).³¹

I. Base

A. Phrase structure rules

$$S \rightarrow NP[S_u] VP$$

$$VP \rightarrow V (NP[DO]) (VP) (PP[IO]) \left(\begin{array}{c} NP \\ PP \\ AdvP \end{array} \right)^*$$

$$PP \rightarrow P NP$$

$$NP \rightarrow \left(\begin{array}{c} D \\ NP[Poss] \end{array} \right) (QP) (AP)^* N (PP)$$

$$AP \rightarrow (DegP) A$$

³⁰ Although the third singular feminine pronoun uses *her* for both objective and genitive case, it would unfortunately not work to list the form just once and leave it unlabelled for case. This would allow the regular genitive case suffix -’s to attach to it, producing an incorrect genitive form. See chapter 20 “Word Division and Clitics” (p. 285).

³¹This is not ultimately a very satisfactory way of handling agreement. For a better approach, see chapter 19 “Case and Agreement.”

$\text{AdvP} \rightarrow (\text{DegP}) \text{ Adv}$
 $\text{QP} \rightarrow (\text{DegP}) \text{ Q}$
 $\text{DegP} \rightarrow \dots \text{ Deg}$
 $\text{N}[-\text{count}] \rightarrow [-\text{plural}]$
 $\text{N}[\text{+ count}] \rightarrow [\pm\text{plural}]$
 $\quad \quad \quad \{1,2,3\} \text{ person}$
 $\text{V} \rightarrow \begin{bmatrix} \pm \text{ plural} \\ \pm \text{ past} \end{bmatrix}$

B. Lexicon (sample lexical entries)

VI[SUBCAT < NP[Su] >]	
wɔk	walk
go	{ go
wɛnt [+ past]	}
N[+ count]	
boj	boy
tʃajld	{ child
tʃildren [+ plural]	}

NP	
aj	[Nom case]
mi	[Obj case]
maj	[Gen case]
ju	[Gen case]
juu	[2 person]

II. Inflectional Spellout Rules³²

N	
	[+ plural]
	[X] → [Xz]
V	
	[+ past]
	[X] → [Xd]
V	
	[3 person]
	- plural
	- past
	[X] → [Xz]

³² As discussed in chapter 12 “Suppletion and Morphophonemics” (p. 154), the output of these rules becomes the input of phonological rules which modify the phonological material to produce the variant forms of the suffixes, such as [-s] and [-iz] for the noun plural.

10.10. Review of key terms

INFLECTIONAL MORPHOLOGY is distinguished from DERIVATIONAL MORPHOLOGY by producing different forms of the same word, by being more PRODUCTIVE, and by being organized according to GRAMMATICAL CATEGORIES into PARADIGMS. A WORD AND PARADIGM grammar takes this paradigmatic structure into account, while an ITEM AND ARRANGEMENT grammar looks primarily at morphemes like beads on a string.

Grammatical categories can be represented by FEATURES. Inflectional features may have just two VALUES (i.e., be BINARY) or more than two, but in all cases the different values for a feature are MUTUALLY EXCLUSIVE.

Features are placed in DEEP STRUCTURE trees by FEATURE ASSIGNMENT RULES, which are a type of phrase structure rule. For REGULAR inflection, the affixes corresponding to each combination of features are added by INFLECTIONAL SPELLOUT RULES, which result in a tree structure called SURFACE STRUCTURE. For IRREGULAR INFLECTION (i.e., SUPPLETIVE STEMS), the forms are provided in the lexicon. For example, the different CASE forms of English pronouns can be handled this way.

10.11. Questions for analysis

1. What types of words are inflected?
2. What is the paradigm for each type of word?
 - a. What grammatical categories are indicated on that class of word (person, number, tense, etc.)?
 - b. What are the choices within each grammatical category? (E.g., How many tenses are there and what are they? Does number include dual in addition to singular and plural?)
 - c. How are each of these choices indicated? In other words, what is the phonological material for each feature value (or for specific combinations of feature values) and where does it occur in the string of inflectional affixes?
3. Are there any cases of irregular inflection? On what types of words? What grammatical categories are involved? In what way(s) are they irregular?

10.12. Sample description

(This description covers a broader range of facts than you would probably be writing about after reading only this far in the book. It is included to give you lots of examples of different ways of talking about morphology.)

English count nouns are inflected for number; singular is unmarked, plural is generally marked with a suffix that is spelled *-s*.

	Singular	Plural
(1)	<i>boy</i>	<i>boy-s</i>
	<i>computer</i>	<i>computer-s</i>
	<i>hope</i>	<i>hope-s</i>
(2)	<i>child</i>	<i>children</i>
	<i>sheep</i>	<i>sheep</i>
	<i>man</i>	<i>men</i>

There are, however, a few nouns with irregular plurals.

(2)

Singular

Plural

<i>child</i>	<i>children</i>
<i>sheep</i>	<i>sheep</i>
<i>man</i>	<i>men</i>

English verbs are inflected for past versus nonpast tense and for the person and number of the subject. However, very few of the inflectional possibilities allowed by this system are actually

realized. In nonpast forms (used in what is traditionally called the present tense), third person singular is indicated with a suffix *-s*. Other person/number combinations are all unmarked.

- (3) *I/we/you/they* *see*
he/she/it *see-s*

The irregular verb *to be* shows slightly richer inflection.

- (4) *I* *am*
we/you/they *are*
he/she/it *is*

Past tense (for all persons and numbers) is indicated by the suffix *-ed* for many verbs (the so-called ‘weak’ verbs).

- (5) Nonpast Past
walk *walk-ed*
exit *exit-ed*
shout *shout-ed*

However, many verbs (the ‘strong’ verbs) have irregular past forms.

- (6) Nonpast Past
see *saw*
sing *sang*
think *thought*

Only one verb, *to be*, shows agreement with the subject in the past tense. It has two past tense forms, *was* for first and third person singular, and *were* for other person/number combinations.

English pronouns are marked for person, number, and case. Third person singular forms are also inflected for gender. The marking is largely suppletive; little meaningful morpheme-cutting is possible.

	Nominative	Objective	Genitive
first person singular	<i>I</i>	<i>me</i>	<i>my</i>
first person plural	<i>we</i>	<i>us</i>	<i>our</i>
second person	<i>you</i>	<i>you</i>	<i>your</i>
third person singular masculine	<i>he</i>	<i>him</i>	<i>his</i>
third person singular feminine	<i>she</i>	<i>her</i>	<i>her</i>
third person singular neuter	<i>it</i>	<i>it</i>	<i>its</i>
third person plural	<i>they</i>	<i>them</i>	<i>their</i>

10.13. For further reading

For more on the inflectional/derivational distinction, see Anderson (1985a) and Matthews 1974, chapter 3.

Matthews (chapter 12) also discusses the theoretical foundation on which this approach to inflectional morphology is based. In particular, it discusses the distinction between item and arrangement and word and paradigm approaches to morphology and how they relate to generative grammars. Anderson’s work is an elaboration of the foundation that Matthews laid. Anderson (1982) provides a brief introduction to the Extended Word and Paradigm framework, while his later book

entitled *A-Morphous Morphology* (1992) provides a thorough exposition of this approach and its rationale.

10.14. Appendix: Features for inflectional morphology

This appendix lists the most common distinctions which are used in languages as the basis for inflectional morphology, together with typical values and some suggestions for formalizing them in terms of features. It is not necessary (and probably not a good use of your time) to memorize these lists. Rather, use them to reinforce your understanding of the text and as a source of ideas when analyzing particular languages. Glance through them now, then refer back to them later as needed. These are only suggestions, since the facts of the language will often dictate the type of feature system that must be used. Use them as a starting point for your analysis and adapt them as needed to fit the facts of the language.

Also, just because a distinction is listed here doesn't mean that it is always inflectional; in fact, some of these like gender, number, and aspect can be derivational (see chapter 11 "Derivational Morphology," especially p. 141ff.), though it is more common for them to be inflectional.

Inflectional categories commonly associated with nouns

2-way number	[± pl] OR [± sg]
3-way number	[Sg number] for singular, [D1 number] for dual, [Pl number] for plural; OR something like [-pl, -dl] for singular, [-pl, + dl] for dual, [+ pl] for plural
2-way gender	[± masc] OR [± fem]
3-way gender	[{masc, fem, neut} gender]
Animacy	[± anim] OR [± human]
Honorifics	[± respect]
Person	[{1, 2, 3} person] OR [± me, ± you]
Inclusivity (in first person plural)	[± incl] (when using [{1, 2, 3} person]), OR [+ me, + you] (for first person inclusive plural, when using [± me, ± you])
Definiteness	[± definite]
Distance	[± proximal] OR [1 distal], [2 distal], etc. (as many values as there are degrees of distance in the system)
Case	[Nom case], [Gen case], [Acc case], [Dat case], [Obj case], etc.

Inflectional categories commonly associated with verbs

Tense	[± past] OR [± future] OR (less likely) [{past, pres, fut} tense]
Remoteness	[± recent] (used in combination with tense features)
Aspect	[± perfective], [± habitual], [± progressive], OR some combination of these
Active versus Nonactive	[± dynamic] OR [± stative]
Mood	[± realis] AND/OR [± imperative]
Negation	[± positive]
Yes-No Question	[± YNQ] (OR just [±Q] if the same morphology is used in content questions too)
Transitivity	[SUBCAT⟨ ... ⟩] (see chapters 8 “Verbal Valence” and 21 “Passive and Voice”)
Voice	[± passive], [± reflexive], [± causative], [± impersonal], if these are relevant to syntactic or inflectional rules (often they aren't).
Agreement	For agreement in person, number, etc., with the subject or direct object, use noun features, positioned in either of two ways: see this chapter (p. 124) or chapter 19 “Case and Agreement” (p. 260ff.).

12.6. Review of key terms

ALLOMORPHY (ALLOMIC VARIATION) refers to variations in the form of a particular morpheme depending on its context; each of these variants is called one of its ALLOMORPHS. There are two types of allomorphy: MORPHOPHONEMICS and SUPPLETION. Suppletion can occur either in the form of IRREGULAR STEMS (STEM SUPPLETION) or in SUPPLETIVE AFFIXATION, which is the use of different SETS of affixes with different INFLECTIONAL CLASSES of stems. Suppletive affixation can be LEXICALLY CONDITIONED or PHONOLOGICALLY CONDITIONED.

If the allomorphy is morphophonemic, then only the UNDERLYING FORM is included in the lexicon or morphological rule; the allomorphy itself is handled by phonological rules which produce the correct SURFACE FORM. If it is suppletive, then the underlying form of each suppletive allomorph is included in the lexicon and/or morphological rules.

It may be helpful to refer to chapter 18 “Overall Structure of a Grammar” to help visualize how morphophonemic rules relate to the morphological rules in the grammar.

12.7. Questions for analysis

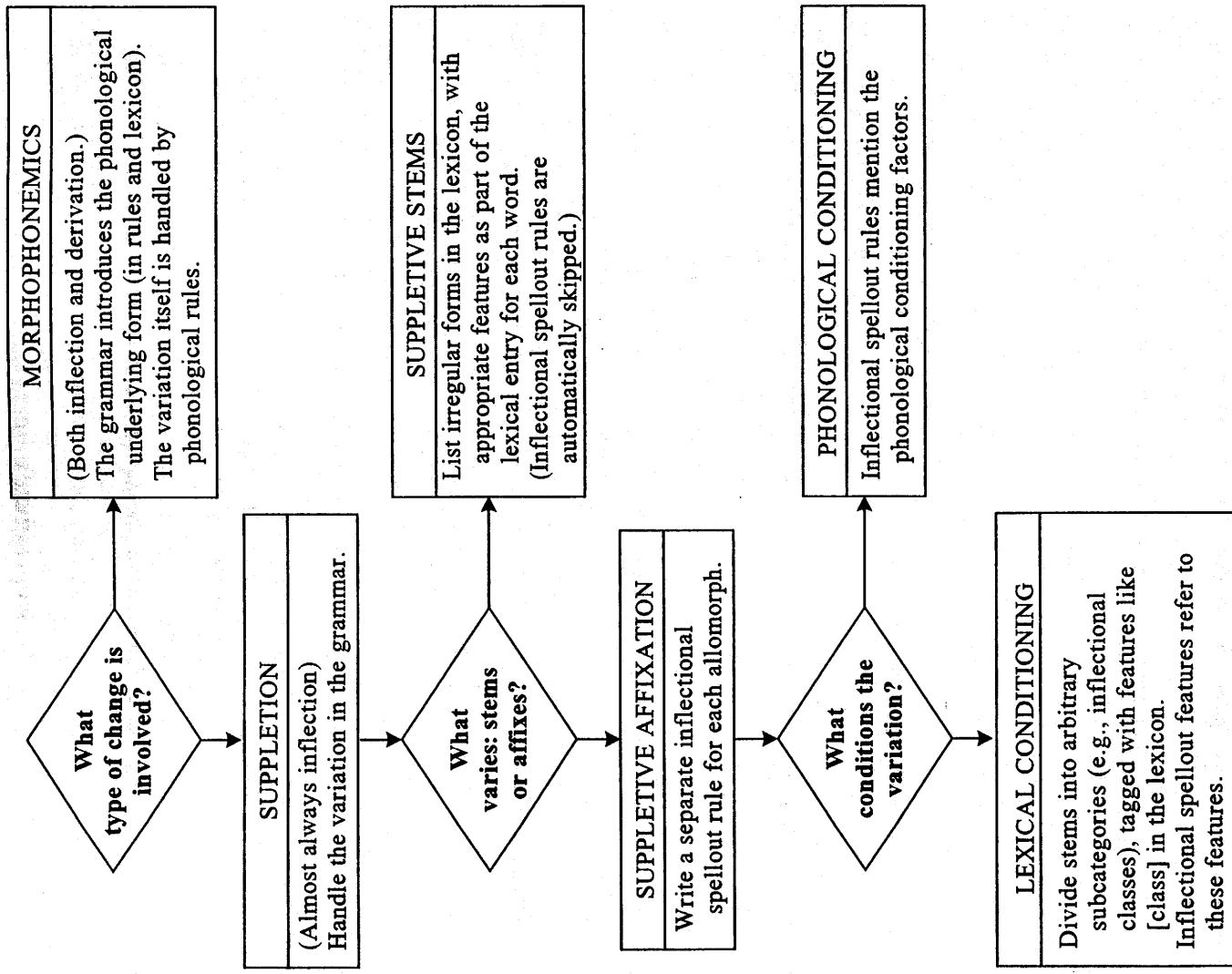
The considerations in §12.5 (p. 160) can be incorporated into a step-by-step procedure that you can use when analyzing allomorphy. This procedure also incorporates the questions for analysis (in italics) for this chapter, which identify the information to include in a description of the data for others.

1. If the morphology is inflectional, first organize your data into paradigms and make hypotheses about the grammatical categories involved. If it is derivational, list pairs of related words in parallel columns. Make tentative morpheme cuts and glosses. Make clear cuts first. Rely on contrast more than recurring partials.
2. If you find allomorphic variation, list the allomorphs of each morpheme. *Which grammatical categories or derivational affixes are involved in the variation? What classes of words do they occur on? What aspects of the variation are morphophonemic, and what aspects are suppletive?*
3. If you find morphophonemics: *What phonological processes are responsible for the variation? What is the underlying form of each morpheme?* Include only the underlying form in the grammar rules or lexicon. Do as much with phonology as you reasonably can.
4. If, after doing this, there is still some allomorphy unaccounted for, treat it as suppletive.¹⁴ *Does this suppletion primarily affect stems or affixes?*
5. If the allomorphy is in stems:
 - a. *What parts of the paradigm are irregular?*
 - b. List the irregular forms in the lexicon, with the appropriate inflectional features.
6. If the allomorphy is in affixes:
 - a. Group the allomorphs of several related affixes into sets, one set for each class of stems. *How many inflectional classes are there? Make a list of the stems in each class.*
 - b. *What is the conditioning factor for the variation?* Check each class of stems for phonological or other similarities that might define them as a class. If you find them, refer to these similarities in the inflectional spellout rules. If you find no similarities, you will have to treat the class as an arbitrary lexical subcategory. This means using a feature like [class] to identify each subcategory in the lexicon, then referring to this feature in the inflectional spellout rules.

This procedure is summarized in the following chart:

¹⁴Since most suppletion is inflectional, the following questions presuppose this. But to handle derivational suppletion of affixes (see footnote 12), the grammar would require a separate word formation rule for each variant. If there is a clear conditioning factor, this can be mentioned in the word formation rule. If the choice of variant is arbitrary, then nothing more needs to be done by rule since the lexical entries of the derived forms will show which variant is used in each case.

(39)



(33)	Constituent order	Examples
Statement	S V O Obl	<i>She stuck the needle in his leg yesterday.</i>
Topicalization	O S V Obl	<i>The needle she stuck in his leg.</i>
Change of setting (see p. 105)	S V O Obl	<i>Yesterday, she stuck it in his leg.</i>
Yes-no question	Aux S V O Obl	<i>Did she stick it in his leg?</i>
Content questions	S V O Obl	<i>Who stuck the needle in his leg?</i>
	O Aux S V Obl	<i>What did she stick in his leg yesterday?</i>
	Obl Aux S V O	<i>Where did she stick the needle?</i>
	Obl Aux S V O	<i>When did she stick the needle in his leg?</i>
Command	V O Obl	<i>Stick it in his leg.</i>
Relative clauses	S V O Obl	<i>... (the nurse) who stuck the needle in his leg...</i>
	O S V Obl	<i>... (the needle) which she stuck in his leg...</i>
Complement clause	S V O Obl	<i>... that she stuck the needle in his leg.</i>
Oblique clause	C S V O Obl	<i>... when she stuck the needle in his leg.</i>

Is English SVO, as we have supposed, or is it rather OSV as in the case of topicalization? Since the auxiliary, a type of verb, can occur before the subject in some clauses, does this mean that English is VSO? Where do obliques belong—at the beginning or the end of the clause?

Surveying this variety, there is still quite a bit of regularity, which is obvious when we line up the constituents in columns as in the chart above. In most clauses, the subject precedes the verb, the verb precedes the objects, and the obliques occur at the end. Putting this all together, we get SVO Obl; this then is the best hypothesis for basic order. This does not mean that most clause types have exactly SVO Obl order, but this order does recur as a theme throughout all clause types. Most clauses in fact have some minor variation on the SVO theme, but it is clearly a variation on SVO order rather than some completely different order.

Recall that one goal in constructing an analysis is to be as simple and general as possible. The simplest analysis overall is that English is SVO: all sentences have SVO Obl order in deep structure and movement transformations generate the other orderings. The derivation of each alternative ordering is fairly simple, since only one or two constituents need to be moved and only for clearly defined reasons. If we had chosen OSV as the basic order, for instance, then most clauses would require a transformation to move the object after the verb, in addition to any other transformations that apply to it, and there would be no apparent reason for the movement. Choosing SVO order as basic rather than OSV allows us to move the object only under very specific circumstances: when it is topicalized.

In summary, then, the earlier advice still stands: look for the orderings of constituents which occur most often. However, there is a subtle twist: 'most often' does not refer to statistical frequency in texts, but to the order that occurs in the greatest *variety* of constructions.¹³ Once you find it, search for what special circumstances (focus, questioning, being the main clause, etc.) are causing some clauses to deviate from this order.

17.8. Review of key terms

The area of MOOD reflects the relationship of the meaning of a clause to the real world and includes such meanings as INDICATIVE versus IMPERATIVE and IRREALIS versus REALIS. Specifically, we have looked at the differences between COMMANDS and STATEMENTS. Syntactically, commands almost always lack overt subjects, yet the presence of a second person subject can be demonstrated by evidence from REFLEXIVE PRONOUNS. (Since they must be COREFERRENTIAL WITH the subject, a second person reflexive pronoun in a command shows that the subject is second person.)

¹³ See footnote 22 (p. 214) for discussion of different ways the term BASIC WORD ORDER is used in different theoretical frameworks.

In some languages commands can be derived by a transformation that DELETES the subject of an imperative clause. Morphologically, verbs in commands tend to have simpler morphology, as in English's use of the bare stem or INFINITIVE. Typically, imperative verbs omit overt reference to person (which is understood to be second person) or time (which is understood to be future); they may also have a special IMPERATIVE affix.

17.9. Questions for analysis

1. Is the subject in a command always omitted or may it be optionally expressed? Is this different from what happens in statements? Is there any difference of meaning in a command when the subject is expressed or not?
2. What grammatical categories are marked on indicative verbs which are not marked on imperatives, and vice versa?
 - a. Is there a special imperative affix or a specialized sense of some other grammatical category (such as future or irrealis)?
 - b. If indicative verbs are marked for tense, aspect, or the person of the subject, are these features omitted on imperative verbs?
3. How are these categories marked?
 - a. What are the specific affixes?
 - b. Are they the same set as in indicative verbs, or is there a suppletive set used only in imperatives? If a special set, what is it?

17.10. Sample description

The verb form in Hebrew commands is very similar to the form in second person imperfect statements, but with omission of the *ti-* prefix and some vowel changes. Like the imperfect, it inflects for the person and gender of the subject.

	Imperfect	Imperative
2 Sg. Masc.	<i>tifmor</i>	<i>fmor</i>
2 Sg. Fem.	<i>tifmariy</i>	<i>fimriy</i>
2 Pl. Masc.	<i>tifməru</i>	<i>fmru</i>
2 Pl. Fem.	<i>tifmornah</i>	<i>famornah</i>

Subject pronouns are normally omitted.

17.11. For further reading

Perlmutter and Soames (1979:8–23) discuss reflexive pronouns and imperatives in English from an early transformational perspective. They include several exercises that focus on the evidence for including *you* as the underlying subject of imperatives in English. The emphasis is on helping the reader develop the ability to construct arguments to decide between competing hypotheses. This theme is carried through the whole book, which covers many of the major syntactic structures of English and issues that were important in generative grammar during the 1960s and 1970s.

Although positive second person commands are in some sense the most important, there are many other ways in which we may use language to influence the world around us. Languages often have a variety of other ‘imperative’ structures for these other purposes. Sadock and Zwicky (1985) provide a brief survey of a number of them, as well as providing examples from a number of different languages. Their article would be a good next step if you want a broader survey of this subject.

19

Case and Agreement

19.1. Goals and prerequisites

This chapter will help you do the following:

- ⑥ identify case and agreement phenomena in data and distinguish them from each other
- ⑥ describe data involving case marking on nouns and pronouns
- ⑥ describe data involving agreement of verbs with subjects and objects, nouns with possessors, and modifiers with head nouns
- ⑥ construct formal analyses that account for case marking of noun phrases in nominative-accusative systems
- ⑥ construct formal analyses that account for agreement with complements in nominative-accusative systems
- ⑥ explain the function of feature percolation in the formal analysis of case and agreement
- ⑥ describe data involving ergative-absolutive patterns of case and agreement

It assumes that you are familiar with the following material:

- ✓ bound and free morphemes, agreement marking on verbs and nouns (chapter 4 “Introduction to Morphology”)
- ✓ phrases and their heads, possession (chapter 7 “Embedding and Noun Phrase Structure”)
- ✓ semantic roles, grammatical relations, the SUBCAT feature, and case marking (chapter 8 “Verbal Valence”)
- ✓ complements vs. adjuncts (chapter 9 “Obliques”)
- ✓ the formal analysis of inflectional morphology (feature assignment rules, inflectional spellout rules, etc.), especially the partial analysis of agreement on verbs and case on English pronouns (chapter 10 “Inflectional Morphology”)
- ✓ why agreement and case are inflectional, not derivational (chapter 11 “Derivational Morphology”)
- ✓ allomorphy, especially suppletion with lexical conditioning (chapter 12 “Suppletion and Morphophonemics”)
- ✓ nominal and adjectival complements (chapter 14 “Nonactive Complements”)

19.2. The morphology of grammatical relations

In this chapter, we discuss two types of inflectional morphology that are sensitive to grammatical relations: case and agreement. We have touched on them several times in previous chapters, but there is more to them than what we have seen so far. Especially, we need to examine their formal analysis in more detail.

They are important, because together with word order, they can provide important evidence about grammatical relations as you do your analysis of a language. This is particularly important when analyzing clauses in which semantic roles do not line up with grammatical relations in the ‘normal’ way (see chapter 21 “Passive and Voice”). In such clauses, the only things we can rely on for making hypotheses about grammatical relations are facts about the form of the utterance, such as word order, case, and agreement.

19.3. Case

Recall what we assumed early in the book (chapter 6 “The Base,” p. 57)—that all noun phrases are alike. That’s why we do not write separate phrase structure rules for subject and object NPs. This assumption is correct as far as syntactic structure is concerned. But, in many languages there are morphological differences in noun phrases, depending on what role they play in the larger structure; this type of morphology is called CASE. In such languages, we say that the noun phrases are MARKED FOR CASE.

Case helps identify the grammatical relations in a clause. For example, case marking on pronouns in English helps speakers determine the grammatical relations, and thus the meaning, of poetic sentences like the following:

- (1)
 - a. *She me loved.* (SOV)
 - b. *Her I did not love.* (OSV)

We have already seen an analysis of case marking on English pronouns in chapter 10 “Inflectional Morphology” (p. 129), which lists the different case forms of pronouns in the lexicon. Each form is tagged with a feature for case, and this feature forces the form to be inserted in a tree at an NP that has an identical case feature. We did not, however, write the rules that determine how the case features got on the NPs in the tree in the first place. That is one of the topics for this chapter.

Case on nouns in Latin

Typically case is indicated in one of two ways. One, there may be suppletive forms of pronouns, as in English. Two, there may be affixes (called CASE MARKERS) on the head noun of a noun phrase. Case markers are like little flags on a noun which help identify its grammatical relation.

To show examples, we must look at a language that marks case on its nouns. Latin serves this purpose well, since it has a well-developed case system. In the following examples, notice how the same noun has different suffixes, depending on its function in the clause.

- (2) *puella columba-m ama-t*
girl dove love-3sg
The girl loves the dove.

- (3) *columba puella-m ama-t*
dove girl love-3sg
The dove loves the girl.

- (4) *puella columba-m libera-t*
 girl dove free-3sg
 The girl is letting the dove go.
- (5) *filia femina-e aqua-m columba-e da-t*
 daughter woman water dove give-3sg
 The woman's daughter is giving water to the dove.

- (6) *femina grammatica-m puella-e doce-t*
 woman grammar girl teach-3sg
 The woman teaches grammar to the girl.

The suffixes indicate something about the grammatical relation of the noun. What appears to be the direct object (the patient) is marked with *-m*, what appear to be indirect objects (recipients and addressees) or possessors are marked with *-e*, and what appears to be the subject (the agent) has no suffix.¹ We can thus make two interdependent hypotheses:

- (7) a. The grammatical relations are what we would expect them to be, based on semantic roles.
 (Thus we assume agents and experiencers are subjects, patients and themes are direct objects, etc.)
- b. There is case marking on the nouns:
 -*θ* for subjects
 -*m* for direct objects
 -*e* for indirect objects and possessors

Traditionally, a case which is used for subjects is called NOMINATIVE, a case for direct objects ACCUSATIVE, a case for indirect objects DATIVE, and a case for possessors GENITIVE. All these different cases are required for the analysis of Latin. (Although the nouns above have the same form for indirect objects and possessors, other nouns do have distinct forms for these two functions, so we must recognize a distinction between dative and genitive case.)²

The reason that we need terminology for case in addition to terms for grammatical relations is that usually one case is used for more than one purpose. For example, in Latin, nominative case is used for nominal and adjectival complements as well as for subjects, and accusative case is used for the objects of many prepositions.³

- (8) *Lydia-θ puella-θ est*
 Lydia-Nom girl-Nom is
 Lydia is a girl.
- (9) *Lydia-θ pulcra-θ est*
 Lydia-Nom pretty-Nom is
 Lydia is pretty.

¹The *a* at the end of all the nouns is traditionally analyzed as part of the case suffix. I present it here as part of the stem, because this is what it appears to be on the basis of this limited data. More generally, I attempt to sidestep the issues that must be raised in full analysis of Latin case morphology. Thus, the analysis is very incomplete; the aim is to provide only enough detail to illustrate the main points about case systems generally. For a much more complete analysis of Latin morphology in a word-and-paradigm model, see Matthews 1972b.

²If Latin really did use the same forms for indirect objects and possessors for all nouns, we would assume that there was only one case that was used for both. We might call it either GENITIVE or DATIVE (but not both), since either label would work equally well in characterizing how the case is used.

³Other prepositions require a fifth case, called ablative. So, even if we know a noun phrase's general role in the whole clause (e.g., as the object of a preposition), we can't predict its case, since it sometimes depends on properties of individual lexical items. This is another argument for the point made in the main text.

- (10) *puella-θ ad villa-m ambulat*
 girl-Nom to country.house-Acc walks
 The girl is walking to the country house.

Another reason is that different classes of nouns take different suffixes for the same cases. Just to give you some flavor of the system, here are some of the suffixes used for various classes of nouns.

	Noun class:	I	II	III	IV
Nominative	-θ	-us	-θ	-us	
Genitive	-e	-ī	-is	-īs	
Dative	-e	-ō	-ī	-ūī	
Accusative	-m	-um	-em	-um	

In short, there is not a simple, direct mapping from function (grammatical relation or other position in the clause) to the individual suffixes. Grammarians have long recognized that in order to compute the morphology for a noun correctly we must use a two-step process with case-marking as the intermediate step. First we state what functions should be represented by each case, and then we state how to spell out each case for each class of nouns.⁴

Formal analysis of case marking

Let's consider how to formalize this analysis in a grammar of Latin. The phrase structure rules are straightforward. (These rules ignore verbal morphology and assume a flat clause structure.)

- (12) S → NPI[Su] (NP[DO]) (NP[IO]) V
 NP → N (NP[Poss])

Next, we need to get case features on the NPs. When case is determined directly by grammatical relation,⁵ this can be done in the feature assignment rules, part of the phrase structure rules. (This chapter shows the features for grammatical relations [GR] in full to emphasize that they are features; elsewhere, for example, [Su GR] is abbreviated as [Su].)

- (13)
$$\begin{bmatrix} \text{[Su]} & \left\{ \begin{bmatrix} \text{[NC]} & \text{[GR]} \end{bmatrix} \right\} \end{bmatrix} \rightarrow \begin{bmatrix} \text{[Nom case]} \\ \text{[Acc case]} \\ \text{[Dat case]} \\ \text{[Gen case]} \end{bmatrix}$$
- [DO GR] → [Acc case]
 [IO GR] → [Dat case]
 [Poss GR] → [Gen case]

These add the case features to the NP nodes, alongside the [GR] features.

Finally, we need inflectional spellout rules, to get the proper case markers on the nouns.⁶

⁴Compare Pollard and Sag's (1987:128–29) concern to maintain the traditional distinction between form (case) and function (grammatical relations).

⁵When case is assigned by specific prepositions, a formal grammar can represent the case assignment with a SUBCAT feature on the preposition. For example, *ad* 'to' can take the feature [SUBCAT < NP[Acc case] >]; this states that the object of *ad* must be in accusative case.

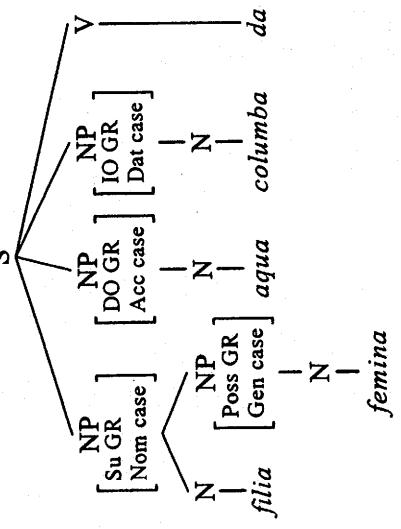
⁶These rules account only for the nouns discussed here. A separate set of rules is needed for each inflectional class of nouns. That is why each rule here has a feature to limit its application to nouns from the first inflectional class (traditionally called FIRST DECLENSION). See chapter 12 "Suppletion and Morphophonemics" for more discussion about suppletive allomorphy of affixes.

$$(14) \quad N \quad \left[\begin{array}{c} \text{Acc case} \\ \text{I class} \end{array} \right] \rightarrow [X_m]$$

$$(15) \quad N \quad \left[\begin{array}{c} \left[\begin{array}{c} \text{Dat} \\ \text{Gen} \end{array} \right] \text{case} \\ \text{I class} \end{array} \right] \rightarrow [X_e]$$

But, wait a minute—there's a glitch. To see what it is, look at the tree that the base rules build for sentence (5).

(16) Output of base component for (5)

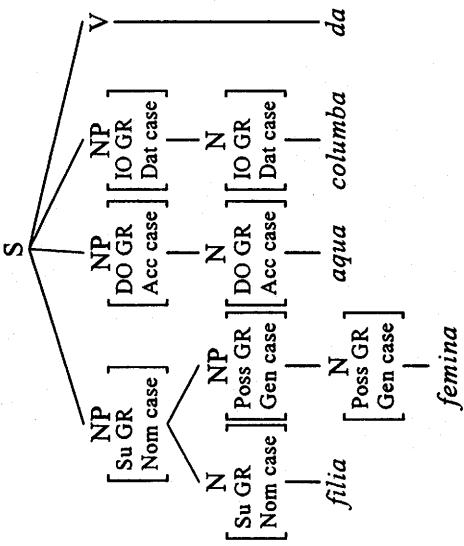


The feature assignment rules (13) put case features on NP nodes, but the inflectional spellout rules (14)–(15) need them on N nodes.⁷ How do we get the case features from the NP nodes to the N nodes?

The answer goes back to what was discussed in chapter 7 “Embedding and Noun Phrase Structure.” Recall that an N is the head of its NP. One characteristic that most linguists assume about phrases and their heads is that they share many important features, including features for case. One way of thinking about this is that anytime an inflectional feature is assigned to a phrasal node, it is automatically copied down to the head. So, the structure of (5) should really be shown as follows:

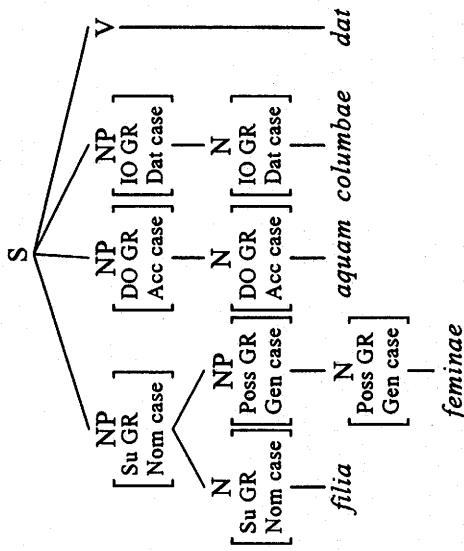
⁷Note that the case marking is definitely on the head noun, not on the final word in the noun phrase. So, we do not have the option of treating case marking as a phrasal affix, as with English possessive -s, and rewriting the inflectional spellout rules to refer to NPs instead of Ns.

(17) Output of base component for (5), with copied features displayed



With this understanding, the inflectional spellout rules apply properly to the nouns, producing the following surface structure:

(18) Surface structure for (5) after application of inflectional spellout rules



The automatic copying of features from a phrase to its head is sometimes called FEATURE PERCOLATION. Our theory assumes it happens automatically, whenever it is needed. There is no need to write rules for it; we simply take advantage of it in our analysis.⁸ Feature percolation is also important in our analysis of agreement, later in this chapter (p. 260).

Case marking versus prepositions and postpositions

In many languages, prepositions or postpositions function in many ways like case markers. In Japanese (for example) the subject, objects, etc., are all postpositional phrases; the particular postposition used is dependent on the grammatical relation of the PP. In such a language, the gloss of the P is simply the appropriate feature for the grammatical relation of the PP ([Su], [DO], etc.). Likewise, in English, indirect objects are PPs, with the preposition *to* being identified in the lexicon as [IO].⁹

⁸Strictly speaking, we would need to specify which features are HEAD FEATURES, i.e., the ones that are automatically copied between a phrasal node and its head. For our purposes, let's assume that any features needed for case and agreement are head features.

⁹Any time a tree contains a PP[IO], the feature [IO] is copied down to the P, forcing us to choose *to* from the lexicon, since it is the P with the most specific set of features which are consistent with the features already in the tree.

Conversely, in some languages there are special case markers that occur on obliques and which serve the same function as a preposition or postposition. Consider the following examples from Dámena (Chibchan, Colombia).¹⁰

- (19) *íngwi kən kaygəmám-ba gágə kóma uyá*
one wood earth-Obl placed Infw did
He laid a stick on the ground.

- (20) *pwebrú-ba naka awín*
village-Obl came did
He came to the village.

- (21) *será-ba nákakó awín*
machete-Obl cut.me did
He cut me with a machete.

In this case, an affix /-ba/ ‘Obl’ is functioning in a role that is usually assumed by prepositions and postpositions.

Despite the similar functions of prepositions, postpositions, and case markers, it is important to distinguish them because their structure is different. This book uses ‘case marker’ for affixes and ‘preposition/postposition’ for syntactically separate words. One reason for maintaining this distinction is that prepositions or postpositions may occur in the same language with case marking. For example, Latin has both case markers as suffixes on nouns and prepositions as separate words in front of the noun phrase. (See example (10).)

19.4. Agreement

We have encountered AGREEMENT several times already in this book, especially in the morphology of verbs. Any affix on a verb which indicates the person, number, or some other characteristic of the subject is an AGREEMENT AFFIX (or AGREEMENT MARKER). In agreement, the morphology of some word ‘points’ to some noun phrase in the clause and redundantly indicates one or more of its features. The number suffix on English present tense verbs is an example. If the subject noun phrase is third person singular, the verb ends in -s. Otherwise, the verb has no ending.

- (22) a. *The boy runs/*run.*
b. *I/we/you/they *runs/run.*

We say that in English the verb AGREES WITH the subject in person and number. By this, we mean that there are different forms of every verb; to determine the correct form in a particular clause, we must consider the person and number of the subject. The morphology of the verb is sensitive to features that are not logically part of the verb itself, but which are logically part of the subject. This is what we mean when we say that the morphology of the verb points to the subject and redundantly indicates its person and number.

This is important, because it can help us identify the subject in unclear cases. Consider, for instance, the following two (highly poetic) sentences:

- (23) a. *God the years controls.*
b. *God the years control.*

¹⁰ Examples are from Cindy Williams 1993:34–35, whose transcription follows Americanist practice. Only the relevant morphological structure is shown. The oblique case marker *-ba* is an enclitic, probably a phrasal affix (see chapter 20 “Word Division and Clitics,” p. 282, footnote 22). It is used on addressees and recipients as well as obliques.

Sentences (23a) and (23b) make very different theological claims. What is there about their almost identical structures that gives them such different meanings? It is the agreement marking on the verb. In the first sentence, the verb has an -s suffix, indicating that the subject is third person singular. The only third person singular noun phrase is *God*, so that we know that it is the subject, despite the unusual word order. In the second sentence, the verb tells us that the subject is not third singular. Again, there is only one noun phrase, *the years*, that could be subject. The agreement has allowed us to identify the grammatical relations of the noun phrases, despite the use of a non-basic word order.

In informal descriptions of languages, agreement may be described using one of several formulas. The first is the easiest to use.

- (24) The subject and verb agree in person and number

The others are a little trickier, because they require that you identify the *source* of the features, i.e., the constituent to which they logically belong. For example:

- (25) The verb agrees with the subject in person and number.

One would *not* say that ‘The subject agrees with the verb in person and number’ because the features of person and number are inherent characteristics of things (nouns), not actions and states of affairs (verbs). The thing that ‘does the agreeing’ is always the category which has the agreement marking (e.g., the verb), and the thing it ‘agrees with’ is always the logical source of the features, i.e., the noun phrase that the agreement morphology points to (e.g., the subject).

Finally, a third way to describe agreement is like this:

- (26) The person and number of the subject are marked (or registered) on the verb.

Formal analysis of verb agreement

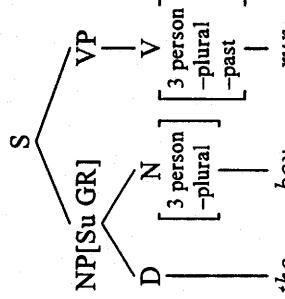
To provide a formal analysis of agreement, we need to improve a bit on what we've done so far. Until now, we've written feature assignment rules to make verbs carry certain features of the subject, such as person and number in English. (Compare chapter 10 “Inflectional Morphology,” p. 124.)

- (27) Feature assignment rule for English verb agreement (until this chapter)

$$V \rightarrow \left[\begin{array}{l} \{1,2,3\} \text{ person} \\ \pm \text{plural} \\ \pm \text{past} \end{array} \right]$$

This results in a tree like the following:¹¹

- (28) Output of base component (until this chapter)



¹¹For this chapter only, I make explicit the assumption that all nouns carry the feature [3 person].

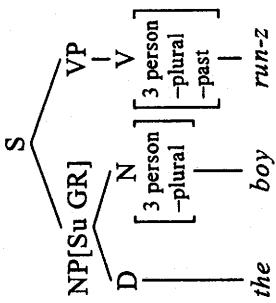
Then we've written inflectional spellout rules to add the agreement affixes themselves.

- (29) Inflectional spellout rule for English verb agreement (until this chapter)

$$\begin{array}{c} V \\ \left[\begin{array}{l} [3 \text{ person}] \\ [-\text{plural}] \\ [-\text{past}] \end{array} \right] \\ [X] \end{array} \rightarrow [Xz]$$

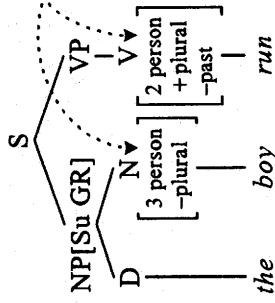
These operate on the tree to produce the correct surface structure.

- (30) Surface structure (until this chapter)



This is all very good, as far as it goes. However, what part of our grammar guarantees that the right combination of features appears on the V? What forces the features for person and number on the V to match those on the subject? What prevents the grammar from generating a tree like the following, in which the verb has the wrong features and, thus, fails to get the right agreement suffix?

- (31) Incorrect surface structure (wrongly permitted by the analysis in previous chapters)



Somehow, we have to prevent this from happening.

There's a second problem too. The approach we've used so far doesn't work very well in languages whose verbs agree with both subjects and direct objects. For example, in Chichewa (Bantu, Malawi),¹² verbs carry agreement prefixes that indicate (among other things) the person and number of both the subject and direct object. The following two examples are identical, except for the object agreement prefix.

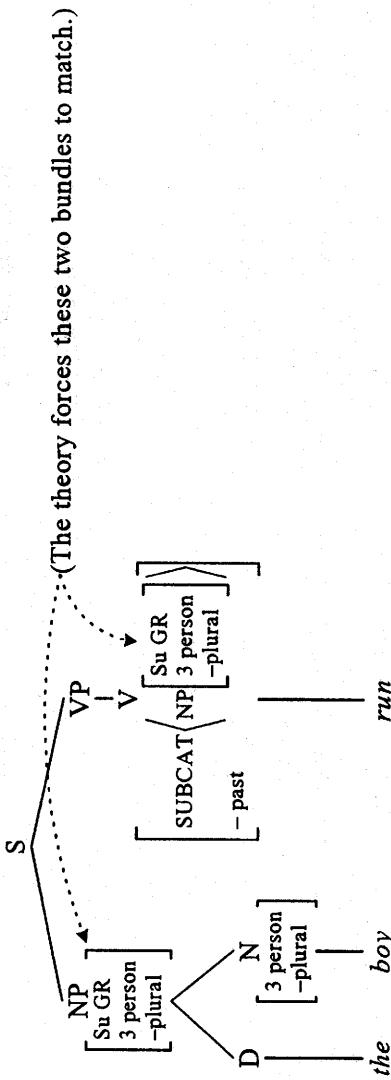
- (32) a. *ndi-ku-mu-ona*
1sgSu-Pres-3sgDO-see
I see him/her.
b. *ndi-ku-wa-ona*
1sgSu-Pres-3pDO-see
I see them.

¹²Chichewa data is from Corbett and Mienje 1987.

In a formal analysis of Chichewa, the verb must be marked with two sets of features for person and number, one set for the subject and one for the direct object. How is a grammar supposed to keep the two sets of features from getting confused with each other?

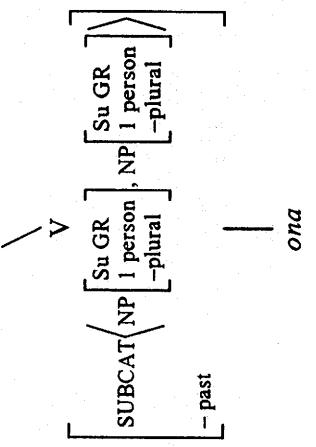
Fortunately, there is an easy solution to both problems. We can take advantage of the SUBCAT feature introduced in chapter 8 “Verbal Valence” (p. 82) and put it to a new use. For subject agreement, we can position the agreement features in that part of the SUBCAT list that requires a subject. In other words, we want to have the phrase structure rules and lexicon produce a tree like the following:

- (33) Output of base component (using agreement features in SUBCAT list)



This solves both problems at once. For the first problem, it automatically guarantees that the subject agreement features on the verb will match the features on the subject, since our theory requires the SUBCAT list on a verb to match the actual nodes in the tree.¹³ Note that, to make this work properly, we have again relied on feature percolation, but this time going upward. In building (33), the person and number features from the subject noun were copied up to the NP node, where they could be checked against the SUBCAT list. For the second problem, in languages that have verbal agreement with both subjects and direct objects, the SUBCAT list provides a way of keeping subject and object inflectional features separate from each other on the verb, because it refers separately to the subject and object of a transitive verb.

- (34) Partial deep structure for (32b): Chichewa *ndikuwaona* ‘I see them’



Now we just need to revise our rules to accommodate the new location for the agreement features. We rewrite the English feature assignment rule (27) so that it can help build (33). In order

¹³To be precise, any feature that is specified for the subject in the SUBCAT list must match a feature that is on the subject NP of the clause. If necessary, a feature that is assigned by rule to the SUBCAT list is copied to the appropriate node in the tree. On the other hand, there are many features that are part of the subject NP in the tree that are not mirrored in the SUBCAT list. Thus, the feature assignment rules have the effect of specifying exactly what features are the basis for agreement, by assigning only those features to the SUBCAT list.

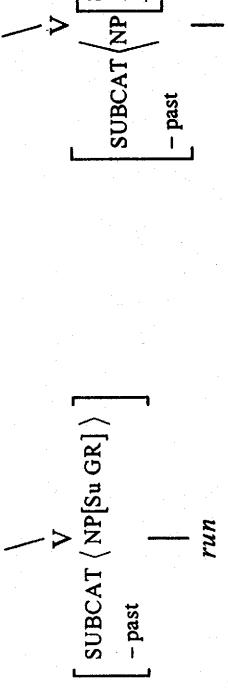
to simplify the notation, we split it into two rules, writing the one that adds the agreement features to the SUBCAT list in an abbreviated form.¹⁴

- (35) a. $V \rightarrow [\pm\text{past}]$
 b. in $V[\text{SUBCAT}]: [\text{Su GR}] \rightarrow \begin{bmatrix} \{1,2,3\} \text{ person} \\ \pm \text{plural} \end{bmatrix}$

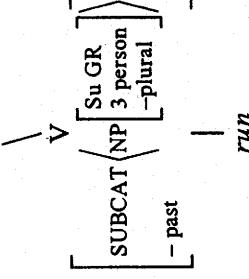
To see how this works, recall that the SUBCAT feature comes originally from the lexicon with the verb. Once it is in the tree, (35b) adds features for person and number alongside the feature for grammatical relation ([Su GR]) that is already there in the list. To illustrate, the diagrams in (36) show these steps in building the tree (33).

- (36) Steps in assigning features to the verb in (33): *runs* (3sg.)

Just after lexical insertion and (35a)



After addition of agreement features by (35b)



We also revise the inflectional spellout rule (29) which adds the /-z/ suffix so it can refer to the new, improved SUBCAT feature.

- (37) $\begin{array}{c} V \\ \backslash \\ \text{SUBCAT} \left\langle \begin{bmatrix} \text{Su GR} \\ \text{3 person} \\ \dots \\ \text{-plural} \end{bmatrix} \right\rangle \\ | \\ \text{-past} \\ | \\ [\text{X}] \end{array} \rightarrow [\text{Xz}]$

This will attach /-z/ to the verb in (33), producing the proper surface sentence ‘The boy runs’. The ellipsis (3 dots) inside the SUBCAT list in the rule allow it to apply equally well to transitive verbs. For languages like Chichewa that have object agreement, we need to do something similar. The rules we wrote for English in (35) also work for Chichewa, but we need an additional feature assignment rule to position the object agreement features in the correct place in (34). As an exercise, write it. Also, write an inflectional spellout rule for the third person plural object prefix that would

¹⁴Without the abbreviation, the rule would have to be written like this:

$V[\text{SUBCAT} \left\langle \text{NP}[\text{Su GR}] \dots \right\rangle] \rightarrow \begin{bmatrix} \text{SUBCAT} \left\langle \begin{bmatrix} \text{Su GR} \\ \{1,2,3\} \text{ person} \\ \dots \end{bmatrix} \right\rangle \\ | \\ \pm \text{plural} \end{bmatrix}$

need to be added to this tree to help produce the surface structure of (32b). How would this rule be ordered with respect to the rules for the other affixes in (32b)? For the answers, see the footnote.¹⁵

In summary, by refining the way we represent agreement features on the verb, we've been able to account for agreement fully without any new rules, just refinements of old ones. In the rest of this chapter, we look at other types of agreement and the way the SUBCAT feature can be used to account for them.

Gender agreement in Imyan Tehit

In English and most European languages you may be familiar with, the verb agrees in PERSON and NUMBER with the subject. These features are often used in agreement systems. Some languages, however, base their agreement systems on other features of noun phrases, such as GENDER. GENDER is an arbitrary or semi-arbitrary system for classifying nouns and pronouns, which is often based on sex, animacy, or shape. In English a distinction is made in personal pronouns on the basis of sex, i.e., *he* (masculine), *she* (feminine), and *it* (neuter). In Spanish there are only two classes, ‘masculine’ and ‘feminine’, and all nouns are either one or the other, so that both classes include nouns which English speakers would probably regard as neuter. Other languages have gender classifications based upon other distinctions, such as animate versus inanimate or the size and shape of an object.

Gender plays a relatively minor role in English, since it is relevant only to the choice of pronouns. In many languages, however, gender is also an integral part of the agreement system. For example, the verb may agree in gender with the subject as well as person and/or number. One such language is Imyan Tehit (West Papuan, Irian Jaya).¹⁶ The following examples illustrate the possibilities for a single verb:

- (38) a. *tet* *toso* *wale* *qyet*
1SgPro hear already news
I heard the news already.
- b. *nən* *noso* *wale* *qyet*
2SgPro hear
You (sg.) heard the news already.
- c. *Maria* *moso* *wale* *qyet*
Mary hear already news
- d. *Yoel* *woso* *wale* *qyet*
Joel hear already news
- e. *faf* *foso* *wale* *qyet*
1PIInclPro hear already news
- f. *mam* *moso* *wale* *qyet*
1PIExclPro hear already news

¹⁵The following feature assignment rule for Chichewa states that verbs agree with their direct objects in person and number:

$$\text{in } V[\text{SUBCAT}]: [\text{DO GR}] \rightarrow \boxed{\begin{array}{l} \{1,2,3\} \text{ person} \\ \pm \text{ plural} \end{array}}$$

The inflectional spellout rule for the third plural object prefix is below. Two sets of ellipses are included, the first one so it will ignore the subject and the last one to allow ditransitive verbs to show agreement.

$$V \left[\begin{array}{c} \text{DO GR} \\ \text{SUBCAT} \left(\dots \begin{array}{c} \text{NP} \left[\begin{array}{c} 3 \text{ person} \\ + \text{ plural} \end{array} \right] \dots \end{array} \right) \end{array} \right] \rightarrow [waX]$$

This rule must apply before the inflectional spellout rules for the other prefixes, since *wa-* is in the innermost position class.

¹⁶Data from Ron Hesse (personal communication). Transcription follows the practical orthography, in which /y/ is an approximant (IPA /j/); /f/ is a bilabial fricative (IPA /ɸ/); /q/ is a uvular stop (IPA /q/), which is pronounced voiced and/or velar in most positions in these examples.

- g. *nan* *woso* *wale* *qyet*
 2PIPro hear already news
 h. *na-i* *woso* *wale* *qyet*
 person-PI hear already news

Note the different forms of the verb 'hear' (identify the agreement affixes if you wish). In order to pick the right form, we need to know several things about the subject. Its person and number are always important. First person plural is subdivided into INCLUSIVE ('we including you') and EXCLUSIVE ('we but not you'). Third person singular is subdivided by gender: masculine and feminine. To summarize in a paradigm:

	Singular		Plural	
	Masculine	Feminine	Inclusive	Exclusive
1	<i>toso</i>		<i>foso</i>	<i>moso</i>
2		<i>noso</i>		
3	<i>woso</i>	<i>moso</i>		<i>yoso</i>

Note that number is irrelevant for agreement in second person, even though there are distinct pronouns for second person singular and plural. Also, the third singular feminine and first plural exclusive forms are both *moso*; this however appears to be a coincidence. This ambiguity normally causes no problems, because the full meaning is available from the context, normally from the subject noun phrase itself.¹⁷

We can express these facts about agreement in the following feature assignment rules:¹⁸

- (40) in V[SUBCAT]:

$$\begin{aligned} [\text{Su GR}] &\rightarrow \left[\begin{array}{l} \{1,2,3\} \text{ person} \\ \pm \text{plural} \end{array} \right] \\ \left[\begin{array}{l} 1 \text{ person} \\ + \text{plural} \end{array} \right] &\rightarrow \left[\pm \text{inclusive} \right] \\ \left[\begin{array}{l} 3 \text{ person} \\ - \text{plural} \end{array} \right] &\rightarrow \left[\begin{array}{l} \{\text{m}\} \text{ gender} \\ \{\text{f}\} \end{array} \right] \end{aligned}$$

These three rules, all of which apply within the verb's SUBCAT list, require the verb first to agree with the subject in [person] and [plural], then if the subject is first plural, also to agree in [inclusive], and if the subject is third person singular, also to agree in [gender].

This takes care of the agreement features on the verb. But, this is just half of the story; we also need to be sure the proper features get on the subject NP in the first place. We need a feature assignment rule (41) to give every noun a feature for singular or plural.

- (41) N → [±plural]

Then, all nouns must be marked in the lexicon as third person and each noun as either masculine or feminine, as in (42).¹⁹

¹⁷Subject pronouns may be omitted from the clause when the meaning is clear from context.

¹⁸For gender, I use a feature with two values (m or f), rather than ones like [+feminine] and [+masculine]. Partly this is to avoid irrelevant discussions about gender bias in the analysis (discussions which are irrelevant because the choice of feature names is, from the point of view of the formal grammar, completely arbitrary). More significantly, this approach provides an easy way to add more genders for languages that require them.

¹⁹Normally, nouns are always third person. However, in Spanish, nouns can apparently be used as the subjects of first person plural verbs: *Los Hijos del Sol no aguantamos eso*. 'We, the children of the sun, don't tolerate that.' (Example from Eugene Loos, personal communication.) This suggests that they may be unmarked for number. However, there is a possible alternative analysis, pointed out to me by Colleen Ahland, in which the noun phrase *Los Hijos del Sol* is an appositive to *pro*.

- (42) N[3 person]
Maria [f gender]
Joel [m gender]
qyet [f gender]
 news²⁰

Finally, every pronoun must carry an appropriate combination of features, as in (43).

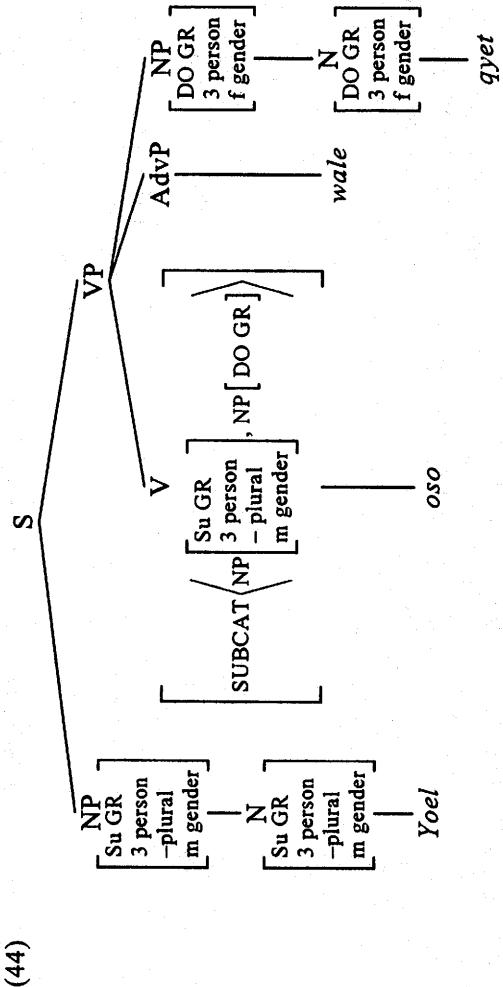
- (43) NP
faf $\left[\begin{array}{l} 1 \text{ person} \\ + plural \\ + inclusive \end{array} \right]$ we (inclusive)

nen $\left[\begin{array}{l} 2 \text{ person} \\ - plural \end{array} \right]$ you (sing.)

nan $\left[\begin{array}{l} 2 \text{ person} \\ + plural \end{array} \right]$ you (pl.)

 ...

When words are inserted into a tree, they carry with them any features that they have in the lexicon. Any features on nouns are then automatically copied to the NP nodes. Thus, the structure of (38d) looks something like this:



Finally, we need inflectional spellout rules to produce the correct verb forms. But, this is largely routine, and I leave it to you as an exercise, if you wish.

Possessor agreement in Tzeltal

Agreement between the verb and its subject is probably the most common type of agreement, but there are several others. In many languages, the head noun of a noun phrase often has affixes indicating the person and/or number of its possessor; this is another type of agreement. For example, consider the following paradigms from Tzeltal (Mayan, Mexico).²¹

- (45)
- | | | |
|------------------------------|--------------|-------------|
| first person ('my') | 'hand/arm' | 'house' |
| <i>h-kap</i> | <i>h-na</i> | |
| second person ('your') | <i>a-kap</i> | <i>a-na</i> |
| third person ('his/her/its') | <i>s-kap</i> | <i>s-na</i> |

²⁰Nothing in the data shows that *qyet* is feminine; this information is directly from Ron Hesse.

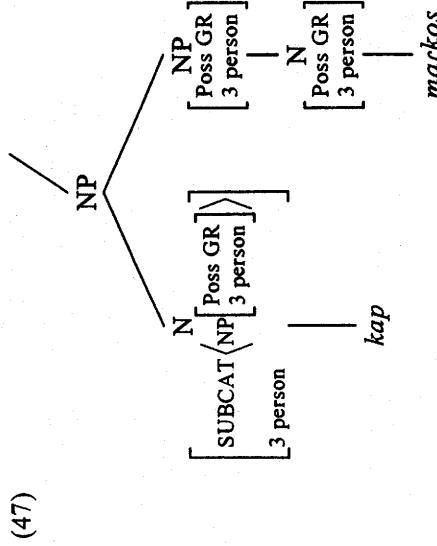
²¹See footnote 5, p. 29.

Viewed in isolation, it may not be obvious that the prefixes are agreement markers. But look at how the third person forms are used in context.

- (46) *s-kap* *markos*
 3-hand Mark
 Mark's hand

We see the same thing as in a verb agreeing with its subject; the head noun *kap* has a prefix *s-* that redundantly indicates the person of its possessor *markos*. (In fact, exactly the same prefixes are used in Tzeltal to show verb agreement with subjects of transitive clauses.) So, we can say informally that nouns agree with the person of their possessor (if any).

The formal analysis here is like the one for verb agreement. We use the SUBCAT feature on the head noun to keep track of the agreement features of the possessor; in this case it states that the possessor is third person.



Notice that *kap* has its own feature for third person, too; the SUBCAT feature enables the head noun to have a record both of its own person and the person of its possessor, without confusing the two.

This works especially well, because we need the SUBCAT feature on nouns for another purpose anyway. Some nouns in Tzeltal, like *kap* ‘hand’, require possessors, just as many verbs require direct objects. Other nouns, like *na* ‘house’, permit possessors but do not require them. In other words, the possessor in a noun phrase is a complement of the head noun. So, we need to subcategorize nouns with the SUBCAT feature just like we subcategorize verbs.

- (48) N[SUBCAT < NP[Poss GR] >] N[SUBCAT < (NP[Poss GR]) >]

Any NP that contains a possessor can have a head noun from either class; any NP that doesn't have a possessor can only have a head noun from the second (optionally possessed) class.²²

What do we need to generate (46)? First, we need a feature assignment rule (49) to add the agreement feature of person to the possessor in the SUBCAT list of a possessed noun.

- (18) in N[SI|BCAT]: [Poss GB] \rightarrow [{1 2 3} person]

This, together with other parts of the base component, will produce the tree (47). The theory will make sure that (49) assigns [3 person] to the SUBJECT list, since the possessor is [3 person].

²²As with verbs that take optional complements, this approach assumes that when an optionally possessed noun is inserted in a tree, its SUBCAT feature is adjusted so that it exactly matches the tree. The parentheses around the possessor in the SUBCAT list are removed if the noun has a possessor, while the possessor is removed from the SUBCAT list if the noun does not have a possessor.

Then, to produce the proper surface structure, we need an inflectional spellout rule like the one in (50).

$$(50) \quad \begin{array}{c} \text{N} \\ \left[\text{SUBCAT} \left(\text{NP} \left[\text{Poss GR} \right] \right) \right] \\ \left[\text{X} \right] \end{array} \rightarrow [sX]$$

For practice, try writing the inflectional spellout rules for first and second person.²³

Modifier agreement within a noun phrase

The types of agreement we have seen so far all involve a head (a verb or noun) agreeing with its complements. There is another type of agreement, however, in which the head noun of a noun phrase controls the form of other words in the phrase, such as adjectives and determiners.²⁴ For example, in Spanish, articles and adjectives change form depending on the gender of the head noun.

- (51) a. *un hombre alto*
a man tall
a tall man
- b. *una mujer alta*
a woman tall
a tall woman

Some of the most striking examples of modifier agreement come from Bantu languages (Africa), in which the same prefixes occur both on the noun and on its modifiers, indicating the gender class of the noun.

- (52) a. *ki-su* *ki-refu*
knife long
a long knife
(Swahili)²⁵
- b. *wa-tu* *wa-refu*
people tall
tall (lit., long) people

Indeed, these same prefixes are also used on verbs to indicate agreement with the subject.

- (53) a. *ki-su* *ki-meanguka*
knife has.fallen
The knife has fallen.
- b. *wa-tu* *wa-meanguka*
people have.fallen
The people have fallen.

²³ As shown in chapter 12 "Suppletion and Morphophonemics" (p. 159), the actual rules in Tzeltal must also be sensitive to the phonological shape of the noun. This particular set of affixes is used only on consonant-initial stems. The rule should probably be generalized to apply to subjects of transitive verbs also, but it is not immediately obvious how to do this.

²⁴The SUBCAT feature cannot be used to account for modifier agreement. For a formal analysis within Head-driven Phrase Structure Grammar, see Pollard and Sag 1994:83–84, 88–91.

²⁵Data from Gregersen 1967:1.

The term CONCORD is often used to refer to this type of agreement system, in which the same prefixes are used repeatedly on different types of words.

Pronouns versus agreement markers

This book carefully distinguishes two types of morphemes, agreement markers and pronouns, which are sometimes confused. It reserves the term PRONOUN for free morphemes which are mutually-substitutable for noun phrases. Affixes on the verb which indicate the person and number of some noun phrase are not pronouns but AGREEMENT MARKERS. Confusion can arise because they communicate the same information; pronouns in one language often correspond to agreement markers in another. For example, compare the following pairs of Spanish and English sentences:

- (54) a. *Juana corre.* *Jane is running.*
 b. *Corr-e.* *She is running.*
 c. *Corr-es.* *You are running.*
 d. *Corr-o.* *I am running.*

In the Spanish sentences (54b)–(54d), there are no overt subjects, while the corresponding English sentences have pronouns as subjects. The agreement suffixes on the Spanish verbs communicate the same information about the subject that the English pronouns communicate.

Despite this similarity in meaning and use, there are two clear syntactic differences between pronouns and agreement markers. First, pronouns are separate words, while agreement markers are affixes. Second, agreement markers can co-occur with a full noun phrase, while pronouns cannot.

- (55) *Juana corre.* **Jane she is running.*

This contrast is most striking when the subject is a pronoun, co-occurring with the agreement markers on the verb.²⁶

- (56) *Ella corre.* **She she is running.*

So, just as case markers are distinguished from postpositions (see p. 256), agreement markers are distinguished from pronouns. Though the functions and meanings of two morphemes may be quite similar, grammatical analysis is concerned with identifying their form.

The following chart may help you sort all this out:

Function	Separate word	Affix
Indicates the relation of a constituent to its larger context (grammatical relation or semantic role)	preposition or postposition	case
Indicate features (like person, number, and gender) of other constituents	pronoun	agreement

For further suggestions about how to tell these things apart in unclear cases, see the discussion in chapter 20 “Word Division and Clitics” (especially p. 279ff., p. 282, and footnote 22).

While we’re clarifying terminology, let’s tackle SUBJECT MARKER. Sometimes it is used to mean ‘verbal affix indicating agreement with the subject’ and sometimes to mean ‘case marker used for subjects’. For this reason, it’s probably best to avoid terms like SUBJECT MARKER and OBJECT MARKER. When necessary, say ‘subject agreement marker’, ‘nominative case marker’, etc.

²⁶In Spanish, this usage generally corresponds to an English pronoun with heavy stress: *She is running.*

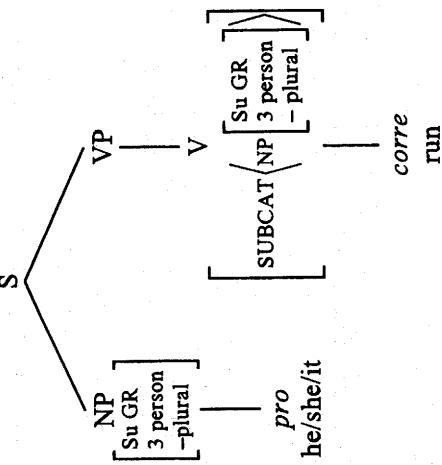
Agreement and pro-drop

Recall (from chapter 8 “Verbal Valence,” p. 86) that many languages permit subject (or other) pronouns to be omitted, and that they are often called pro-drop languages. There are two further issues to explore about pro-drop in light of the information in this chapter.

Roughly speaking, pro-drop languages tend to have richer sets of agreement markers than do non-pro-drop languages like English. For example, Spanish has separate verb forms for all six combinations of person and number of the subject. Almost all English verbs, on the other hand, have only two forms: one for third singular and the other for all other person/number combinations. English agreement morphology does not tell much about the subject, so pronominal subjects cannot be dropped without significant loss of information about the subject. Thus, it is understandable why English does not allow pro-drop. Spanish, with its much richer agreement system, can safely allow pro-drop without losing the ability to communicate the person and number of the subject. So, the richer the system of agreement in a language, the more likely it is to allow pro-drop. However, this is not a hard-and-fast rule; some pro-drop languages do not have any verb agreement, and some languages with fairly rich verbal agreement cannot freely omit the subject.²⁷

In the formal analysis of pro-drop phenomena, we accounted for the absence of the subject by assuming that the language had a silent pronoun *pro* which could appear as subject. So, for example, the base for Spanish might generate a tree like the following:

(57) Deep structure for *Corre ‘he/she/it runs’*



Notice that there are person and agreement features on *pro*. These get copied there automatically when the features for agreement are placed in the SUBCAT list. This is an aspect of the analysis using *pro* that hasn’t been explicit yet, but we have assumed it (and represented it informally in trees with glosses). That is, even though *pro* does not have features for person, number, etc., in the lexicon, it can eventually get them in a specific tree as a result of agreement features in the SUBCAT feature.²⁸

19.5. Ergativity

In the examples considered so far, both case and agreement treat subjects of intransitive clauses the same as subjects of transitive clauses. Both types of subject had the same case markers and triggered the same pattern of agreement, regardless of transitivity. However, many languages treat subjects of intransitives in the same way as the direct objects of transitives for the purposes of case

²⁷Haegeman 1991:418 provides useful discussion and references on this point.

²⁸Indeed, it may be necessary to see that pronominal features get assigned to *pro* even when it is not involved with agreement, in order for semantic interpretation rules to work properly, but we don’t need to resolve that issue here.

or agreement. For example, in Dámena (Chibchan, Colombia),²⁹ the subject of a transitive clause has a case suffix *-ga*, but the subject of an intransitive clause is unmarked, like a direct object.

- (58) *ranžáde-θ naya*
my.father went
My father left.

- (59) *ranžadé-ga dumágə-θ gwagá awin*
my.father lion killed did
My father killed a lion.

The same thing can happen with agreement. In Tzotzil (Mayan, Mexico),³⁰ one set (Set A) of verbal affixes indicates agreement with the subject of a transitive clause, while another set (Set B) indicates agreement with either the direct object of a transitive or the subject of an intransitive.

- (60) *Ch-Incomplete 2A mil -on. 1sgB*
You're going to kill me.

- (61) *Tal -em -on.*
come Perfect 1sgB
I have come.

This phenomenon is called ERGATIVITY. In ergative case and agreement systems, it is convenient to have a special term ABSOLUTIVE that means ‘subject of an intransitive clause or direct object of a transitive clause’. Similarly, the term ERGATIVE means ‘subject of a transitive clause’. These terms make it easier to describe the facts of the language.

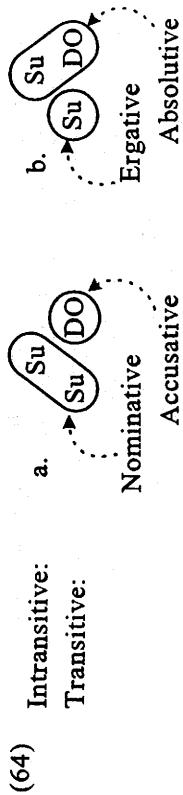
- (62) Damana case marking
a. ergative noun phrases are case-marked with *-ga* (ERGATIVE case)
b. absolutive noun phrases are unmarked (ABSOLUTIVE case)

- (63) Tzotzil verb agreement
a. Set A affixes indicate agreement with the ergative of the clause (ERGATIVE agreement)
b. Set B affixes indicate agreement with the absolute of the clause (ABSOLUTIVE agreement)

Although it may seem strange for a language to treat intransitive subjects as if they were direct objects, it really isn't if you look at it in terms of the needs of communication. Languages often treat subjects and direct objects differently with respect to case or agreement; this helps distinguish them for the listener. In intransitive clauses, however, the subject doesn't need to be distinguished from a direct object, so it doesn't matter how it is treated. Sometimes languages treat it the same as the subject of a transitive clause, according to what is called a NOMINATIVE-ACCUSATIVE PATTERN, as in (64a). Sometimes they treat it the same as a direct object, according to an ERGATIVE-ABSOLUTIVE PATTERN, as in (64b).

²⁹Data from Cindy Williams 1993:29, whose transcription is based on Americanist practice. Only the relevant morphological structure is shown. The ergative marker *-ga* is an enclitic, probably a phrasal affix (see chapter 20 “Word Division and Clitics,” p. 282, footnote 22.). It is also used on certain nominals that trigger indirect or direct object agreement (Cindy Williams 1993:30–31) but which could conceivably be analyzed as subjects at some level.

³⁰Data from Aissen 1987:44–45.



In most theoretical frameworks, the formal analysis of ergative case and agreement systems is more complicated than the analysis of nominative-accusative systems. For this reason, I will not attempt to provide a formal analysis for ergativity here.³¹ However, you should be able to recognize ergative case and agreement systems when you find them and to describe them informally.

19.6. Review of key terms

CASE and AGREEMENT are two types of inflectional morphology that are sensitive to grammatical relations. Thus they help us take our analysis beyond just guessing at the grammatical relations based on the semantic roles involved. In both, FEATURE PERCOLATION is important in making our formal grammars work correctly.

Case represents the grammatical relation of a noun phrase. Both nouns and pronouns may be MARKED FOR CASE. Affixes on nouns that indicate case are called CASE MARKERS. Possible cases include NOMINATIVE, ACCUSATIVE, DATIVE, GENITIVE, ERGATIVE, and ABSOLUTIVE.

Agreement, which is typically marked on verbs but can also be marked on other types of words, indicates something about the features of some other constituent in the clause; grammatical relations are used to state which other constituent is important. Quite frequently, for example, the verb AGREES WITH the subject and/or direct object, a noun agrees with a possessor, or a modifier agrees with its head. Features of GENDER, NUMBER, and PERSON (including INCLUSIVE versus EXCLUSIVE in first person plural) typically are involved in agreement.

Both verb agreement and case marking can manifest ERGATIVITY. That is, they can follow an ERGATIVE-ABSOLUTIVE pattern instead of the more familiar NOMINATIVE-ACCUSATIVE pattern. AGREEMENT MARKERS (= AGREEMENT AFFIXES) should not be confused with PRONOUNS, despite their similarity of meaning. Pronouns, for example, may be optional (in pro-drop languages), but agreement markers usually are not. Similarly, CASE MARKERS should not be confused with prepositions and postpositions. Finally, I suggest avoiding terms like SUBJECT MARKER and OBJECT MARKER, even though they occur in the literature, since they may refer to either case or agreement.

19.7. Questions for analysis

For case:

1. On what words in a noun phrase is case marked? (head noun only? modifiers, too? only pronouns?)
2. How is case marked? That is, what are the forms of the case markers?
3. Does case-marking follow a nominative-accusative or ergative pattern?

For agreement:

1. On what word is the agreement marked? (the verb? the head noun of a noun phrase? an adjective or other noun phrase modifier?)
2. In what features does it agree? (person? number? gender? case?)
3. What is the source of these features? (subject? direct object? possessor? head of the NP?)

³¹Pollard and Sag 1994 do not offer any suggestion on how to handle ergativity within a Head-driven Phrase Structure Grammar. One possibility might be to add ergative and absolute case features in appropriate places in the SUBCAT feature on the verb, which would then get automatically copied to the subject and direct object NPs. Inflectional spellout rules for case and agreement could refer to these features rather than to grammatical relations.

4. How is the agreement marked? (What are the morphological forms of the agreement markers?)
 5. Does verb agreement follow a nominative-accusative or ergative pattern?

19.8. Sample descriptions

Subject-verb agreement in Imyan Tehit

In Imyan Tehit, the verb agrees with the subject in person, number, and (in third person singular) gender.

- (1) *tet t-oso wale qyet*
 1SgPro 1Sg-hear already news

I heard the news already.

- (2) *saf f-oso wale qyet*
 1PIInclIPro 1PIIncl-hear already news

We (incl.) heard the news already.

- (3) *Maria m-oso wale qyet*
 Mary 3SgFem-hear already news

- (4) *Yoel w-oso wale qyet*
 Joel 3SgMasc-hear already news

The complete paradigm of the agreement prefixes is as follows:

(5)	Plural			
	Singular	Masculine	Feminine	
	Inclusive		Exclusive	
1	<i>t-</i>		<i>f-</i>	<i>m-</i>
2			<i>n-</i>	
3	<i>w-</i>	<i>m-</i>		<i>y-</i>

Latin case

Latin (an SOV language with relatively free word order) exhibits case on nouns, by means of case/number suffixes. There are at least four cases: nominative (for subjects and nominal complements), accusative (for direct objects and the objects of certain prepositions), dative (for indirect objects), and genitive (for possessors).

- (1) *filia-∅ semina-e aqua-m columba-e dat*
 daughter-NomSg woman-GenSg water-AccSg dove-DatSg gives
 The woman's daughter is giving water to the dove.

19.9. For further reading

This chapter relies heavily on features to do the work necessary to describe case and agreement. This approach is largely based on work in Generalized Phrase Structure Grammar (Gazdar, Klein, Pullum, and Sag 1985) and Head-driven Phrase Structure Grammar (Pollard and Sag 1987). For a basic introduction to GPSG, including a discussion of feature percolation, see Sells (1985:77–134).

For more extensive discussion of such matters as case, agreement, gender, and other inflectional categories, see Anderson 1985a.

20

Word Division and Clitics

20.1. Goals and prerequisites

This chapter will help you do the following:

- Ⓐ state syntactic guidelines and a few basic phonological guidelines for developing hypotheses about word boundaries
- Ⓐ use these guidelines to identify clitics in written data, regardless of whether they are written attached to adjacent words or separately
- Ⓐ classify clitics, using standard terminology
- Ⓐ describe clitics informally, including stating the reasons for considering a morpheme to be a clitic and for classifying it as a bound word or phrasal affix
- Ⓐ construct a formal analysis that accounts for a clitic's distribution (except clitic doubling)

This chapter assumes that you are familiar with the following material:

- ✓ bound versus free morphemes (chapter 4 “Introduction to Morphology”)
- ✓ embedding in noun phrases (chapter 7 “Embedding and Noun Phrase Structure”)
- ✓ different structures used for obliques (chapter 9 “Obliques”)
- ✓ the formal analysis of inflectional morphology, especially the analysis of case-marking on English pronouns (chapter 10 “Inflectional Morphology”)
- ✓ the distinction between inflection and derivation (chapter 11 “Derivational Morphology”)
- ✓ morphophonemics, especially for English -(e)s, and suppletion (chapter 12 “Suppletion and Morphophonemics”)
- ✓ portmanteau forms (chapter 13 “Nonlinear Affixation”)
- ✓ movement transformations (chapter 15 “Variable Orders of Constituents”)
- ✓ the formal analysis of agreement and case, the distinction between case-marking and prepositions, and the distinction between agreement-marking and pronouns (chapter 19 “Case and Agreement”)

(Is it beginning to be clear that the different parts of a grammar can be extremely interrelated?)

20.2. Word boundaries

So far, we have maintained a clear distinction between morphology (the structure of words) and syntax (the structure of sentences). We have assumed we knew what a WORD was and where the WORD BOUNDARIES (or WORD BREAKS) were in a sentence. Now it's time to question those assumptions. It is one thing to divide a sentence into words when there is an established written tradition for a language; people simply learn to write spaces where everyone else writes them. It is quite another thing to identify word boundaries in an unwritten language and to help speakers develop their own conventions for writing them.

So, what exactly is a word, and how do we make reasonable hypotheses about word boundaries? First off, we need to recognize at least three different senses of the word 'word'.

- (1) a. 'word in a dictionary' OR
 'lexical entry'
 b. 'word form' OR
 'fully inflected word'
 c. 'word in a sentence'
- a stem, together with all its inflected forms (i.e.,
 its entire paradigm)
 a particular inflected form from the paradigm of
 some word (in sense a)
 a string of material in a sentence that forms one
 of the basic building blocks of the sentence

Earlier chapters (especially chapter 10 "Inflectional Morphology") have focused on senses a and b, which are concerned with words as objects within a grammar. This chapter is concerned with clarifying sense c, that is, with finding word boundaries in specific sentences. We will be asking the following types of questions:

- (2) a. How do we make an analysis of boundaries between words in a sentence?
 b. How do we decide if a morpheme is a word or an affix, if it is bound or free?
 c. How many terminal nodes do we need in a tree? How do we distribute the
 phonological material of a sentence among them?
 d. What do we treat as syntax and what as morphology?
 e. Where do we write spaces in a practical orthography?

There are quite a few guidelines, both syntactic and phonological, to use in answering these questions. In most cases analyzing word boundaries is straightforward, but sometimes the syntactic guidelines suggest different word breaks than the phonological ones. Thus, we must distinguish PHONOLOGICAL WORD BOUNDARIES and SYNTACTIC WORD BOUNDARIES; these are discussed in the next two sections.

Phonological word boundaries

The traditional conception of word boundaries is phonological—a word in sense (1c) is a minimal utterance.¹

- (3) **Phonological guideline #1**
No utterance can be shorter than a single phonological word.

If native speakers do not recognize a morpheme out of context or would never pronounce it by itself, then it is not a word. If they can, it probably is.² For example, the minimal reply to a question is a single word, not part of a word.

¹Notice the similarity to the definition of MORPHEME in chapter 3 "Morphemes and Hypotheses" (p. 19) as a minimal meaningful unit. The same considerations about 'minimal' apply here as are discussed in chapter 4 "Introduction to Morphology" (p. 25).

²This guideline has to be used with caution when speakers have an established written or grammatical tradition. In the course of learning to read or analyze their own language, they may also have learned how to pronounce bound forms in isolation. In these cases, looking at normal responses to questions is a much more reliable test than just discussing words as words to see if they can be spoken in isolation.

(4) *How many children do you have?*

- (4) *How many children do you have?*

Possible answers: Impossible answers:

*-teen *-s

When you learn a word in isolation (which is normally what happens), this guideline alone can give you a pretty good idea about word boundaries. When you hear the same word in a sentence, a reasonable hypothesis is that there are word boundaries on either side of it (allowing for any affixes, of course).

A second phonological guideline is that pauses are generally impossible inside words.

(5) Phonological guideline #2

Pauses are only possible at word boundaries.

If there is a pause at a certain point in a sentence or if a pause is possible there, then that point is probably a phonological word boundary. For example, consider how an irritated elementary school teacher might speak to the class, pausing for effect between each word, but not within words.³

- (6) a. [bojz... ɿənd... gɪlz]
 b. * [boj... z... ɿənd... gɪl... z]

The problem with (6b) is that the plural suffixes, which are bound morphemes, have been pronounced as separate words.

Once you begin to apply these guidelines in a language, you may find phonological behavior is affected by word boundaries. You can use them as additional diagnostic tests.

[View Google guideline #3](#)

Look for phonological rules that provide information about how to pronounce words that help define word boundaries in unclear cases.

For example, the rules in English which account for the variation in voicing of the plural suffix -s do not operate across word boundaries. There is good reason to believe that the underlying form of this suffix is /-Z/. There is a phonological rule that causes /z/ to devoice after a voiceless segment (as in *bats*), but a /z/ which begins a separate word (as in *the bat zoomed*) does not devoice in that environment.

- (8) Input to phonological rules:
 Output of phonological rules:
 ‘the bat zoomed’
 bæt-z
 bæt-s

So, if you find a /z/ at the beginning of some morpheme which sometimes becomes [s] because of this rule, then the rule is treating the morpheme as a suffix, not a separate word. That is, the rule provides evidence that the morpheme is a suffix. As another example, the rules of English require full words to have at least one vowel, but this requirement does not apply to morphemes like the plural, which often occur without vowels. This is evidence that it is a suffix, not a separate word.

Specific rules like this can be very useful in identifying phonological word breaks. To find them, work from clear cases to unclear ones. First develop a hypothesis about a phonological rule which is sensitive to obvious word breaks, and refine it until it works well in clear cases. Then, use this hypothesis to help you make further hypotheses about word breaks in other cases that are less obvious. Use what you know to discover what you don't know.

³The phonetic transcription of English here is an approximate broad phonetic transcription of one upper Midwestern U.S. dialect, not based on any particular phonological analysis. It is used at several points to focus on the phonological system of the language rather than irrelevant details of the practical orthography.

Exercise: Phonological word boundaries

Consider the following facts about Tohono O'odham (Uto-Aztecan, Arizona):⁴

- (9)
 - a. All stressed syllables are word-initial, but not all words are stressed.
 - b. All words begin with one or more consonants.
 - c. Within a word, many vowel sequences are forbidden, such as /ao/, /ou/, and /ai/. They are forbidden even when the vowels are separated by a glottal stop /ʔ/ or /h/.

We can use the guidelines to identify the phonological word boundaries in the following clause, even without identifying morphemes.

- (10) čihiaʔohíhuʔiħígaimístol

The girl is chasing the cat.

Before reading on, try using the guidelines in (9) to insert phonological word boundaries in (10). If you need help, consult the footnote.⁵

Syntactic word boundaries

When we turn to syntactic word boundaries, we are essentially trying to determine what to put on separate terminal nodes in a tree. Our first hypotheses about syntactic word boundaries are based on what we know about the phonology.

- (11) **Syntactic guideline #1**

Any phonological word break is generally also a syntactic word break.

So, once we have decided that there is a phonological word boundary before *zoomed* in *the bat zoomed*, we are largely committed to recognizing a syntactic word boundary there, too, and putting *zoomed* on a separate terminal node. (Warning: the opposite is not true, as discussed below; some syntactic word breaks are not phonological ones!)

But, there are purely syntactic reasons for making syntactic word breaks too. One of the strongest is based on the overall structure of the tree.

- (12) **Syntactic guideline #2**

Any major constituent break (e.g., the beginning or end of a phrase) is also a syntactic word break.

This, together with our understanding of noun phrase structure (chapter 7 “Embedding and Noun Phrase Structure,” p. 65), explains why there are word boundaries between the different morphemes in NPs like the following. Between almost every pair of morphemes is a phrase boundary, hence also a syntactic word boundary.

- (13) /NP the /QP two / /AP /DegP very / little / dogs /

Other guidelines may also be helpful in deciding whether a morpheme is an affix or a separate syntactic word.

⁴Information about Tohono O'odham (formerly known as Papago) from Dean Saxton; based on an illustration by Langacker (1972). Transcription based on Americanist practice.

⁵By (9a), we need a word break somewhere before each stressed vowel. By (9b), we must locate each word break just before a consonant. This gives us čihiaʔo híhuʔiħígaimístol. (Placing the second word break between /d/ and /h/ is a guess, although a reasonable one.) By (9c), our first ‘word’ needs to be divided further, since /aʔo/ is not permitted within words. By (9b), the glottal stop belongs with the /o/: čihiaʔo híhuʔiħígai mísstol.

(14) Syntactic guideline #3

Affixes tend to occur next to only a single type of word (their stems) and in a fixed order; words occur more freely in various combination with each other.

This means words can often be moved with respect to each other; this is usually not possible for morphemes within a word.

- (15) a. *I see the tiny, little people down there on the ground.*
 b. *I see the little, tiny people down there on the ground.*

- (16) a. *teach-er-s*
 b. **teach-s-er*

It also means a syntactic word or phrase can usually be inserted between two other morphemes only at syntactic word breaks; trying to insert a word between a stem and an affix usually results in the affix being attached to the wrong type of word.⁶ So, for example, to find out if the English plural marker -(e)s is a word or affix, we could try inserting adjectives (which we already know to be separate syntactic words) and PPs between the noun and the plural marker.

- (17) a. *the dog-s*
 b. **the dog-big-s*
 c. **the dog [in the manger]-s*⁷

This doesn't work, so according to both phonological and syntactic guidelines, -(e)s is bound.

Finally, one can often tell the difference between words and affixes by sniffing out the irregular forms in the language.

(18) Syntactic guideline #4

Morphology often shows great irregularities, while combinations of separate syntactic words do not.

We have seen this morphological irregularity often (especially in chapter 12 "Suppletion and Morphophonemics"). In contrast, irregular combinations of separate syntactic words (which I have called TOTAL FUSION, see chapter 13 "Nonlinear Affixation," p. 176) hardly ever occur. For example, there are irregular noun plurals like *oxen* in place of *oxes, but there are not any irregular combinations of the definite article *the* with particular nouns. So, by this guideline, the plural -(e)s is an affix and *the* is probably not (although we can't tell for sure, because it might be a perfectly regular affix).

Exercise: Syntactic word boundaries

Justify as many of the spaces in the following examples as you can, by citing specific syntactic evidence that each represents a syntactic word boundary. Where there are morpheme breaks inside words, cite the evidence that they are not syntactic word boundaries. See the footnote for a partial answer.⁸

⁶This criterion needs to be used with caution, because of compounding (see chapter 11 "Derivational Morphology," p. 137) and noun incorporation (chapter 21 "Passive and Voice," p. 302). So, before concluding that a particular boundary is a word boundary just because a word can be inserted at that point, one must eliminate these two possibilities. That is, this is ungrammatical when the -s is interpreted as the plural marking on *dog*, not the possessive 's or the plural of *manger*.

⁷First, construct a tree for each sentence. Every phrase boundary is also a word boundary; this takes care of many of the spaces. However, this is somewhat circular (until we become sure of our hypothesis about phrase structure), since we may have used our hypotheses about word boundaries to help justify our phrase structure rules. So, we should consider other criteria. Several substrings of the sentences can be used as answers to questions, such as 'Arthur', '(a) cactus', 'all', 'over', 'his', 'please', and 'choose', and are phonological words by other guidelines; this provides evidence for syntactic word

- (19) a. *Arthur grabbed a cactus, and he has prickers all over his hand!*
 b. *Hand me the calculator, please.*
 c. *Which 'un did ya' choose?*

20.3. Clitics: When the guidelines disagree...

Word breaks are not always obvious. Sometimes the different guidelines don't even agree; some guidelines may suggest that a certain morpheme is bound, while others may suggest that it is free. If it is unclear whether a morpheme is a word or an affix, it is generally called a CLITIC. In some ways (especially by phonological guidelines), clitics are like affixes; in other ways (especially by syntactic guidelines), they are like words.

As examples, consider two homophonous morphemes in English, both of which are spelled 's':

- (20) Contracted 's (from *is*)

- a. *hu-z g̩n* Who's gone?
- b. *hwat-s ðæt* What's that?
- c. *hwitf-iz jorz* Which's yours?

Possessive 's

- d. *ðə mæn-z opnjan* the man's opinion
- e. *ði kæt m ðə hæt-s opnjan* the cat in the hat's opinion
- f. *ðə pitf-iz pit* the peach's pit

Should we analyze them as affixes or words? Phonological guidelines suggest that both are affixes.

- (21) a. They are never pronounced in isolation.
 b. They cannot be preceded by pause.
 c. They do not contain vowels (at least in some environments), unlike clear cases of words.
 d. They undergo devoicing, like the plural suffix in (8).

Syntactic guidelines suggest that they are separate words.

- (22) a. Contracted 's functions as the verb in a clause (as a contraction of *is* or *has*).
 b. Possessive 's always attaches to the end of a noun phrase, not always to the head noun (look again at (20e)).
 c. Both can attach to a variety of word types (see (24) below).
 d. Contracted 's is completely regular; there is no stem suppletion.

On the other hand, at least one syntactic guideline suggests that possessive 's is an affix.

- (23) There are irregular possessive forms of pronouns, like *my*, which are found in place of regular combinations like *'/s.

breaks at their edges. Most of the spaces mark places where large chunks of material can be inserted which are clearly word-sized or larger, as seen in (1).

- (i) *Hand over to me, won't you, the electronic calculator in the corner, please?*

Many pieces between spaces can be rearranged freely.

- (ii) *Ya' did choose 'un, didn't ya?*

Finally, for every item between spaces which has lexical meaning, we can substitute other items, without running across any irregular changes or combinations with neighboring items with grammatical meaning (the ones that could conceivably be affixes). Compare (iii) to (19a).

- (iii) *My daughter bumped a bucket of paint, and she has dark brown splotches all over her new clothes!*

These morphemes are not clearly words or affixes, but have some characteristics of each. They are PHONOLOGICALLY BOUND like affixes and (at least partially) SYNTACTICALLY FREE like words. So, in order to have some label for them until we figure out exactly what they are, we call them clitics.

20.4. Finding clitics

Cross-linguistically, clitics generally have grammatical meaning, rather than lexical meaning. Most belong to closed classes like pronouns, prepositions, auxiliary verbs, and conjunctions. They usually attach to the edges of words, outside of derivational and inflectional affixes. They almost always are clearly-defined strings of segments, like prefixes and suffixes (rather than infixes, morpheme processes, or other nonlinear affixes discussed in chapter 13 “Nonlinear Affixation”).⁹ Still, clitics are not always obvious, because analysts tend to jump to conclusions about whether something is a word or an affix. When you get data in written form, decisions about where to put spaces have already been made. Clitics may be written either as words or affixes, perhaps inconsistently. Sometimes even clear cases of words or affixes are written the ‘wrong’ way. You can’t assume that written words always correspond to phonological or syntactic words, although most of the time they will.

When you collect your own data, be equally suspicious. Be on the lookout especially for apparent affixes that have some word-like properties and, thus, might need special analysis as clitics. Just because something is phonologically bound does not mean that it should always be analyzed as an affix. On the other hand, clitics are relatively uncommon. Don’t assume something is anything other than a straightforward affix or word unless you have good evidence to the contrary. If it looks like an affix and sounds like an affix, then assume it is an affix, unless you know that it also has some characteristics of a word. Do the same for apparent words. So, what is good evidence that something is a clitic rather than an affix?

One type of clitic that is easy to spot is one that attaches to a variety of different types of words. An example is the possessive -'s, as it is used in colloquial spoken English.¹⁰ (Linguists sometimes use an equals sign rather than a hyphen on clitics, to distinguish them from affixes.¹¹)

- (24) a. *[the woman's tennis racket*
- b. *[anyone who likes children]'s ideas about child-rearing*
- c. *[the one with red on]'s atrocious behavior*
- d. *[people who hurry]'s ideas about politeness*
- e. *[someone who types quickly]'s job prospects*

Here, -'s attaches to a head noun, a nonhead noun, a preposition, a verb, and an adverb. This property of many clitics is sometimes called PROMISCUOUS ATTACHMENT. Affixes, on the other hand, usually attach to only one type of word. But, there is a generalization about possessive -'s; it always attaches to the last word in the noun phrase (whatever it may be). A better way to say this is that it attaches to the whole phrase.

Promiscuous attachment can also be illustrated by a morpheme which occurs in Huichol (Uto-Aztecan, Mexico).¹²

⁹One exception is noted by Zwicky 1987:138, in Tongan, for a phrasal affix. Barbara Hollenbach (personal communication) and I have even encountered bound words which are realized as morpheme processes in a few Otomanguean languages. These are clearly subject pronouns in terms of their syntax, with the same distribution as the subject pronouns in other Otomanguean languages (of which the Petíoles Mixtec data in (26) is typical). However, they are apparently realized on verbs as processes, such as the laryngealization of all vowels from the stem vowel to the end of the word. (It is possible that an autosegmental representation of the bound word, with appropriate spreading rules, might avoid the need to posit that a bound word can be a process.)

¹⁰Because this construction is not sanctioned in standard English, many of these seem odd, if not ungrammatical, when encountered in writing. However, such constructions commonly occur in natural speech and for that reason should be considered grammatical, in at least some dialects.

¹¹See, for example, Klavens 1985 and Spencer 1991:379ff.

¹²Data based on Merrifield et. al. 1987, #186, with clarifications (chiefly tone and vowel length) supplied by Joseph Grimes (personal communication). Transcription follows Americanist conventions. According to Grimes, “*cic*, like certain other

- (25) a. *zákí* canyon
 b. *zákí ?amípáa* big canyon
 c. *zákicé*
 d. *zákí ?amípacé* in the big canyon

The morpheme *-cfe* ‘in’ seems to be an affix, since it is phonologically bound. But, it attaches to either a noun or an adjective, again whichever is last in the noun phrase. Its location is defined with respect to the whole phrase, rather than with respect to a single word. This is word-like behavior, like a postposition. Thus, *-cfe* can be called a clitic.

Another type of clitic that is easy to spot appears at first glance to be an agreement marker, but disappears whenever a full noun phrase appears, as seen in the following examples from Peñoles Mixtec.¹³ In this case a morpheme that looks like an agreement affix, *-ši* ‘3f’, is mutually-substitutable for a full noun phrase.

- (26) *ní-šinu-ši*
 Cmp-run-3f
 She ran.

- (27) *ní-šinu šeči-á'*
 Cmp-run girl-that
 That girl ran.

However, clear cases of agreement markers do not disappear when there is a full noun phrase as subject. (See the various examples in chapter 19 “Case and Agreement,” pp. 257ff, and the discussion on p. 267.) In other words, *-ši* ‘3f’ has the syntactic properties of a pronoun (a separate word), even though it is phonologically bound. Thus, it is better to analyze *-ši* not as an agreement marker but as a pronoun, because it is mutually-exclusive with some category of free forms (for example, noun phrases).¹⁴

Often, clitics are variant forms of morphemes that are clearly separate words, but the clitics are phonologically reduced (missing one or more segments, vowels changed to schwa, stressless, etc.). For example, English [‘ænd] ‘and’ has a shortened clitic form [n̩] (usually spelled ‘n’). The full and reduced forms are sometimes simply variants of the same lexical entry; the two forms are identical in meaning and grammar; they differ only in their phonological properties. The reduced form acts phonologically like an affix, but in all other respects is a separate word.¹⁵

Again, if it looks like an affix and sounds like an affix, analyze it as an affix, unless you have good reason to suspect otherwise. Some good reasons to analyze a morpheme as a clitic are as follows:

(28)		Evidence	Probable type of clitic
Promiscuous attachment		bound word or phrasal affix	
Mutually-substitutable with a category of free forms		bound word	
Phonologically-reduced variant of some free form		bound word	

(This chart also suggests which of two types a clitic might be; these are discussed below.)

enclitic postpositions in Huichol, forms a separate phonological foot, while other enclitic postpositions become part of the final foot of the item to which they attach.

¹³Data from Peñoles Mixtec provided by John Daly. Transcription follows Americanist conventions. Superscript *n* represents nasalization.

¹⁴For more on distinguishing agreement markers from pronouns, see chapter 19 “Case and Agreement” (p. 267).

¹⁵Be cautious though. The reduced form may be a completely separate lexical entry, with only a historical relationship to the full form. As a result, there may some differences in semantics or syntax. Also, some clear cases of affixes may bear similarities to free words that are their historical ancestors, so not all phonologically-reduced forms are clitics. See, for example, Zwicky’s (1987) analysis of English *n’t* as a suffix.

20.5. Types of clitics

Clitics are often classified based on two dimensions:

- where they attach
- how word-like versus affix-like they are

Let's consider each of these classifications.

Where does it attach? Proclitics and enclitics

Clitics may be classified based on whether they attach at the beginning or end of another word. A PROCLITIC is a clitic that is phonologically attached to the beginning of some other item, like a prefix. Many phonologically reduced forms in English are proclitics.

- (29) ... *ðɪz-* *pens!* ...
(Where did) the pencil (go?)

- (30) ... *n= æpl*, *n= æ tʰuək*, *n= ǣ bɔl*
(Santa brought me) 'n apple, 'n a truck, 'n a ball...

- (31) *gɪt n= ðɪz- kəɹ*
Git 'n th' car.

An ENCLITIC is a clitic that is phonologically attached to the end of some other item, like a suffix. The two clitics 's in (20) are enclitics, as is 'cie 'in' in (25) and -sí '3f in (26).

How word-like or affix-like is it? Unstressed words, bound words, and phrasal affixes

However, it is far more significant to classify clitics based on how word-like or affix-like they are. Some clitics are essentially words with the phonological properties of affixes, others are essentially affixes with a few syntactic properties of words.¹⁶

Of those clitics that are essentially words, we can distinguish two subtypes. Some are simply words that are never stressed or can occur without stress, such as the reduced forms of *the* and *in* mentioned in (29)–(31). These have all the properties of words, except that they (the reduced forms) are never pronounced in isolation, since they are stressless. Although these are sometimes considered to be clitics, most recent discussions assume that a morpheme should not be considered phonologically bound simply because it is stressless.¹⁷ We call such morphemes UNSTRESSED WORDS, and consider them to be clitics only marginally.

Rather, we should concentrate on syntactic words that are phonologically bound in more interesting ways than unstressed words. For example, English has two homophonous contractions 's (from *is* and *has*); both are clearly verbs and thus are syntactically free. But, both undergo the same morphophonemic alternation [s/z/iz] as the noun plural -(e)s and the verb third person singular -(e)s. Since these rules only apply within phonological words, these contractions are phonologically

¹⁶Following a proposal by Nevis (1986) with later refinements by Zwicky (1987), I assume that ultimately all clitics can either be analyzed as words with some affix-like properties (bound words) or affixes with some word-like properties (phrasal affixes). That is, the term CLITIC does not refer to a principled class of linguistic objects, but rather a pretheoretic amalgamation of two very different types of morphemes. Nevertheless, it can be a useful concept, if defined as 'a morpheme which is not clearly either a word or an affix'. This is particularly true in early stages of analysis.

¹⁷For example, Zwicky (1987:133) states, "I propose here to reserve the term [clitic] for elements whose description requires more than the stipulation that they may or must be prosodically dependent," and again (1985:287), "[the accentual test of clitichood] is most unreliable; it should never, I think, be used as the sole (or even major) criterion for a classification." Nevis (1986:68–70) echoes, "It is true that a large number of grammatical function words do not bear stress or accent as a matter of principle. But these can be distinguished from true clitics, which have additional properties." Besides the existence of unstressed words that are not clitic, both point out the existence of stressed clitics.

bound. Thus they are clitics, for more reasons than just being unstressed. They are called BOUND WORDS, since they are essentially words which happen to be phonologically bound.

As another example, Telugu (Dravidian, south India) has two types of postpositions. Some are free; they are clearly separate words.¹⁸

- (32) *kaaru rooddu miida unnadi*
car road on it.is
The car is on the road.

- (33) *kaayitaalu peeti kinda unnaay*
papers box under they.are
The papers are under the box.

Others are bound, but otherwise have essentially the same syntactic traits as the free postpositions.¹⁹

- (34) *atanu madraasu-nunci waccæədhu*
he Madras-from he.came
He came from Madras.

- (35) *waadu kukkamu karra-toot koṭṭaaḍḍu*
he dog stick-with he.beat
He beat the dog with the stick.

For this reason, the simplest analysis seems to be to consider both the bound and free forms to be postpositions. Those that are phonologically bound are examples of bound words.

In contrast to bound words, some clitics are essentially affixes, but they have the odd characteristic that they attach to a whole phrase, rather than a specific word. The English possessive '-s' discussed above in (24) is the standard example of this type of clitic.²⁰ Such clitics are called PHRASAL AFFIXES.

Another example of phrasal affixes is provided by case markers in Demeña²¹ mentioned in chapter 19 “Case and Agreement” (pp. 257 and 269). These are enclitics, not suffixes, since they display promiscuous attachment. They always occur at the end of a noun phrase, attaching either to a pronoun, the head noun, or a following adjective.²²

¹⁸Data from Krishnamurti and Gwynn 1985:93–95. Underdot on alveolars represents retroflexion.

¹⁹The analysis here is tentative. Krishnamurti and Gwynn 1985:95 do not specify in what sense the free postpositions are free, except that they can “occur as independent words.” In other words, the bound postpositions appear to be clitics, but their exact phonological properties are not specified. When I had the opportunity to consult a native speaker from Hyderabad, it appeared that free postpositions began with a rise in pitch (stress?), while bound ones did not. At least one bound postposition, ‘*too* ‘with’, has a corresponding free form *tooti*. My consultant also noted that bound postpositions are written attached in the practical orthography and that these examples represent very informal usage. Krishnamurti and Gwynn also mention that there are some syntactic and morphological differences between the free and bound postpositions, but their exact nature is unclear.

²⁰Compare also Radford’s (1988:65) description of ‘s as a “phrasal inflection.”

²¹Data from Cindy Williams 1993:29, 33–35, whose transcription follows Americanist conventions. Only the relevant morphological structure is shown.

²²I assume that these case markers are best analyzed as phrasal affixes rather than clitic postpositions for the following reasons: (1) If the ergative marker is a suffix, then the phrase structure rules can be stated more simply, since subjects will always be NPs regardless of transitivity. (2) The ergative marker can be omitted from transitive subjects in some circumstances (Cindy Williams 1993:33–34), which is not typical behavior for a postposition. (3) If the genitive enclitic was a postposition, it would be more difficult to formalize agreement of the head noun with the possessor (which occurs with inalienable possession, Cindy Williams 1993:36–38), since the SUBCAT feature for agreement on the head noun would then have to refer to the object of a PP rather than to a bare NP. To put this another way: from a typological perspective, the object of a PP usually does not control agreement outside the PP. Hence, possessors are probably better analyzed as NPs, and the genitive enclitic as a phrasal suffix. (4) There is a distinct set of locative postpositions (Cindy Williams 1993:80) which have a different distribution from the oblique case marker. Two of these are free forms; the other, ‘*ka* ‘in, on, at’, is apparently a clitic. As evidence that they are distinct from the case markers, (a) the oblique case marker can co-occur with one of the free postpositions (Larry Williams 1993:109); (b) the ergative and oblique case markers precede the Topic clitic ‘*-ru* (see examples in Cindy Williams 1993:152 and Larry Williams 1995:7, 44, and elsewhere), which can attach to a variety

- (36) a. *íngwi kən [kaygəmám]-ba gágə káma uyá*
one wood earth=Obl placed InfW did
He laid a stick on the ground.
- b. *íñgúna Zukwegáj]-ba nángwi*
trail good=Obl he.stands
He is standing on the good road.

- (37) a. *[ranžadé]-ga dumágə gwagá awín*
my.father=Erg lion killed did
My father killed a lion.²³
- b. */súži bənsʃ]-ga zey mákənakúki káma ketwá*
bird white=Erg fear don't.let.come.for.you InfW said
The white bird said to her, "Don't let fear come for you."

- (38) a. *[nawin]-že urága*
we=Gen house
our house
- b. *fména tšukwegán]-že mäyķasára*
woman big=Gen clothes
a big woman's clothes

We have seen that the term CLITIC refers to two quite different types of morphemes: bound words and phrasal affixes. In fact, these two types probably shouldn't both be called by the same label. When we use the term CLITIC for both, it implies that they have significant characteristics in common. In fact, about all they have in common is that they tend to get confused with each other. So, the term CLITIC is useful only as a temporary way of talking about a morpheme which is not obviously either a word or an affix; it does not ultimately shed much light on what it actually is. It is better to use terms like BOUND WORD and PHRASAL AFFIX whenever possible.

20.6. How to handle clitics in a formal grammar

The formal analysis of different types of clitics is straightforward, with only a couple of slight twists on what we've seen before. Unstressed words and bound words are handled like other words, while phrasal affixes are handled like inflectional affixes.

Clitics that are essentially words (bound words)

Since bound words (and unstressed words) are just a special type of syntactic word, they are listed in the lexicon with all the other words. You may want to include an equals sign on bound words as a way of showing that they are phonologically bound.

- (39) P (Huichol)
*cίe to

(The equals sign is just an informal notation, not an official part of a formal grammar. To be complete, we would need to explain how bound words 'change' from being free to bound as they move from the syntactic to phonological parts of the grammar.²⁴ The equals sign is only a reminder that our grammar is incomplete in this respect.)

²³The different phrases and thus is probably a bound word; (c) the clitic locative postposition generally follows the Topic clitic (see examples in Larry Williams 1995:20, 28, and elsewhere) and, thus, is probably also a bound word.

²⁴Presumably a large wildcat such as a mountain lion or jaguar.

24 Nevis (1986), for example, discusses a process of 'liaison', which attaches a clitic phonologically to a neighboring word between the syntactic and phonological components, and discusses similar proposals by others.

It's probably best not to include the equals sign for unstressed words, since these are not phonologically bound in any significant sense. Thus, for the English preposition *in*, we list the reduced form /n/ next to the full form in the lexicon. When it comes time to insert some form of this word in a tree, we are free to choose either the full or reduced form.

- (40) P (English)
 m, n location in a place

Similarly, if bound words are alternate forms of ordinary (free) words, we list both forms, as in the verb *be*.

- (41) V [SUBCAT < NP[S₁], [NC] >]

<i>bi</i>	<i>æm, -m</i>	[1 person, -plural, -past]	(copula)
<i>iz, -z</i>		[3 person, -plural, -past]	
<i>aI, -I</i>		[-past]	

As always, we list '*s*', the contracted form of *is*, in its phonological underlying form /-z/, since phonological rules will account for its allomorphs [-s] and [-iz]. (Compare the way we handled the plural suffix -(e)s in chapter 12 "Suppletion and Morphophonemics," p. 153.)

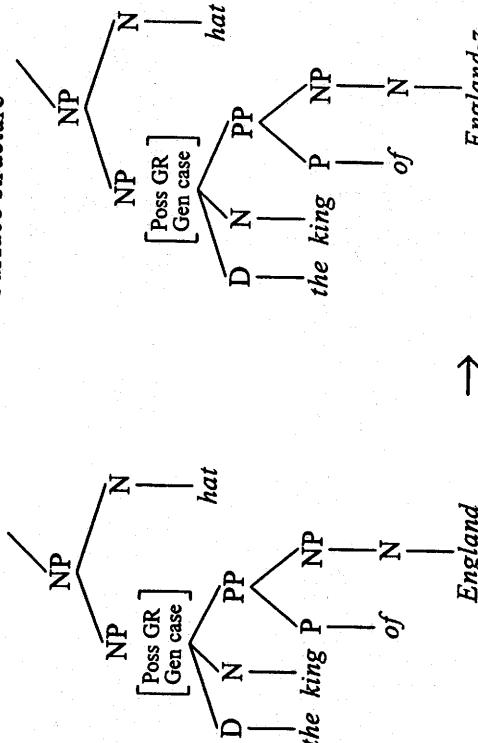
Clitics that are essentially affixes (phrasal affixes)

Phrasal affixes are handled in the grammar like any other inflectional affix; they are added with an inflectional spellout rule. The only unusual thing about the rule is that it refers to a *phrasal category*—in this case, NP.²⁶

- (42) NP
 [Gen case]
 [X] → [Xz]

This rule applies in the following way to add the possessive suffix to the possessor NP *the king of England*:

- ### 43) Deep structure



STUDIES IN LITERATURE

there are some restrictions on when the bound forms can occur, which are not accounted for here.

I assume that all possessors carry "Case and Agreement" (D 254).

Like all inflectional spellout rules, (42) does not apply to possessive pronouns, whose forms are listed explicitly in the lexicon like all irregular stems. (See chapter 10 “Inflectional Morphology,” p. 129.)

(44)	NP		
	aj [Nom case]	[1 person]	
	maj [Gen case]	{ [- plural]	
	mi [Obj case]		
	hi [Nom case]	[3 person]	
	hs [Gen case]	{ [- plural]	
	hm [Obj case]		[m gender]
	...		

When the irregular forms are inserted in the tree, this automatically blocks the regular inflectional spellout rule (42) from applying and thus prevents ungrammatical forms like the following:

- (45) *my's
*his's
*her's
*their's

On the other hand, one English pronoun is completely regular: *it*. Only one form must be listed in the lexicon. When *it* is inserted as a possessor, the inflectional spellout rule (42) adds /-z/ to it, just as it does to all ordinary NPs. After the application of regular phonological rules, the surface form *its* results. The only thing ‘irregular’ about this form is that its spelling does not include an apostrophe, but this is a fact about the writing system, not about the spoken language, which is the primary focus of our analysis.

Distinguishing bound words from phrasal affixes

Phrasal affixes may appear at first glance to be syntactically free, like bound words, because they attach to a variety of different types of words. However, on closer examination, it may become apparent that the simplest analysis is to treat them as inflectional affixes, not as bound words on a separate terminal node. Compare, for example, the English possessive suffix in (24) to the Huichol postposition in (25). Both exhibit promiscuous attachment. But, we have analyzed them differently: English 's as a phrasal affix (42), while Huichol -cfe as a bound word (39). Why?

The only way to be sure in cases like this is to try both possible analyses (bound word versus phrasal affix) and see which one works best. Suppose we analyzed English 's as a bound word. Then we would have to put it in some category in the lexicon, the best choice might be to consider it a postposition. (See chapter 19 “Case and Agreement,” p. 256, for the parallel between case markers and pre/postpositions.) But, this would force us to complicate our PP rule to allow the NP to occur either before or after the head P.

- (46) PP → (NP) P (NP)

Further, we would need to mark each P in the lexicon as to whether it was a preposition (most of them) or postposition ('s), rather than letting the PP rule take care of this detail for everything at once. Finally, possessive pronouns (*my*, *your*, etc.) would have to be assigned to the category PP, rather than NP, since under this analysis possessors are PPs; this is weird, since usually pronouns are NPs, not PPs.²⁷ Clearly, it is simpler and more natural to treat 's as a phrasal affix.

²⁷Many of these difficulties are not insurmountable, given other theoretic assumptions. It might be the case, under other assumptions, that the analysis would work out better by treating possessive 's as a bound word. So, like most analytical decisions, this one is dependent on the theoretical assumptions being used. We can't say absolutely that 's is a phrasal affix, only that under certain assumptions, this is the best analysis.

On the other hand, in Huichol, it works just fine to analyze *-cfe* as a postposition (a bound word), as far as we know. Further, bound words are more common than phrasal affixes, so our first hypothesis about a clitic should always be that it is a bound word; we analyze it as a phrasal affix only if doing so simplifies the analysis.

There's another reason, too, to treat '*'s* as a phrasal affix. Recall (from chapter 9 "Obliques," p. 95) that there is a tendency for PPs to be used mostly for obliques and rather rarely for subjects and objects.

- (47) Su DO IO Oblique
 NP ←————→ PP

Case markers, on the other hand, are more commonly used for subjects and objects and quite rarely used for obliques. Possessors tend to be like subjects and objects in this regard. So, naturalness suggests that it is better to treat English '*'s* as a case marker (a phrasal affix) and Huichol *-cfe* as a postposition (a bound word).²⁸

These, then, are some of the considerations that can be relevant when you try to classify a clitic as a bound word or a phrasal affix, specifically, as a pre/postposition or a case marker. Of course, the arguments based on simplicity of the analysis are stronger than the ones based on naturalness. But, since all these arguments point in the same direction, we can be reasonably sure of the conclusion.

20.7. Special properties of some clitics

Some clitics, specifically bound words, have special grammatical properties that need to be accounted for in addition to their status as clitics.

Some bound words don't appear in the position that you'd expect, but rather move someplace else. For example, the direct object pronouns in Spanish are bound words. They are attached to the left of a verb (as proclitics), instead of following the verb like objects that are full noun phrases.

- (48) a. *Manuela tiró el voleibol.*
 Manuela threw the volleyball
 Manuela threw the volleyball.
 b. *Manuela lo- tiró.*
 Manuela 3sgDO= threw
 Manuela threw it.

Even though they appear in a different part of the clause, they do not co-occur with full direct object noun phrases; as in Mixtec (26), this is evidence that they are pronouns, not agreement markers.

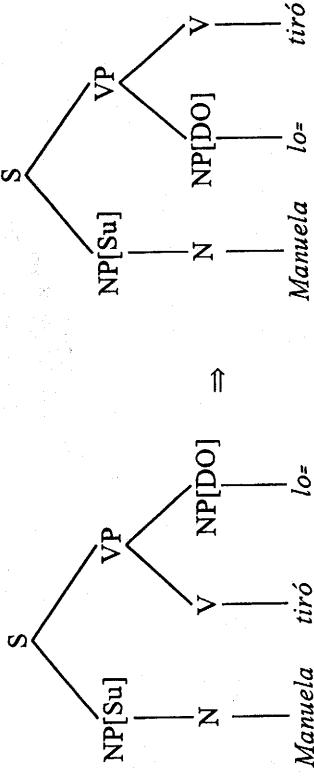
- (49) **Manuela lo- tiró el voleibol.*
 Manuela 3sgDO= threw the volleyball
 (Manuela threw the volleyball.)

These facts can be handled easily by generating the object pronouns in deep structure in the normal place after the verb, then obligatorily moving them with a transformation to precede the verb.²⁹

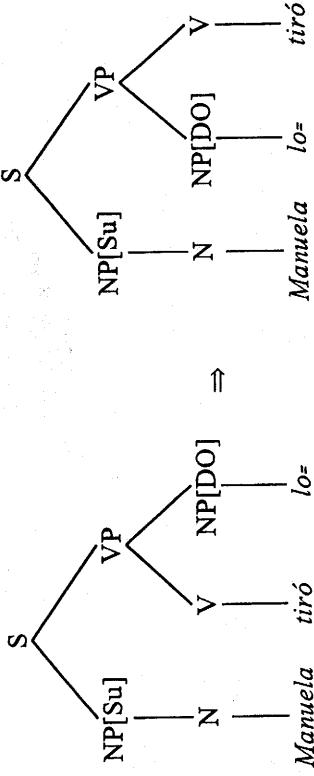
²⁸Naturalness is relevant in another way, too. Huichol is a Uto-Aztecan language, a family which is predominantly head-final. Although Huichol may not be head-final (noun phrases are head-initial), postpositions are not out of place in the family as a whole. In European languages like English, however, postpositions are quite rare.

²⁹Such a transformation would need some way to distinguish pronouns, inserted from the lexicon, from full NPs developed by the phrase structure rules. The most natural way to do this is to mark all pronouns in the lexicon as NP[+pro], while NPs with noun heads are [-pro]. This feature is needed anyway for the binding conditions (Chomsky 1981:184–85). The facts of

(50) a. Deep structure



b. Surface structure



In many languages, clitics tend to cluster just after the first word in a constituent. These are commonly called SECOND POSITION CLITICS. In Latin, for example, the conjunction *-que* 'and' is an enclitic. Unlike most conjunctions, it does not appear between the two conjoined constituents (in this case, two nouns), but one word later.

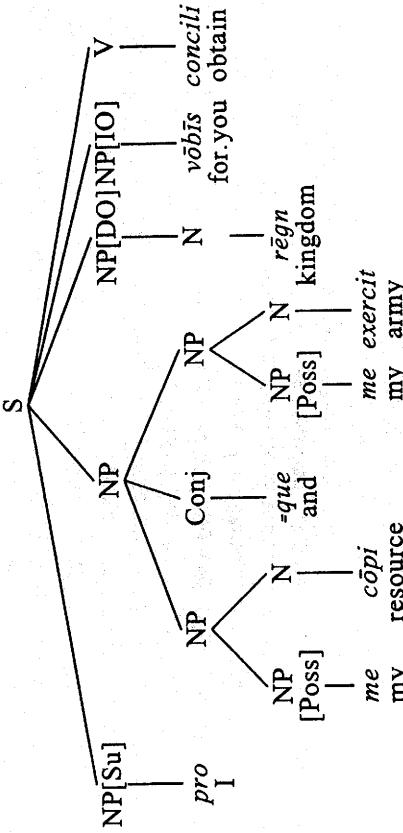
- (51) *senātus populus -que rōmānus*
 senate people and of.Rome
 the senate and people of Rome

If the second conjoined constituent contains more than one word, *-que* is still placed after the first word and thus ends up inside the constituent.³⁰

- (52) */ Meīs cōpīis / I meō -que exercitū / rēgna / vōbīs conciliābō*
 my resources my and army kingdoms for.you I.will.obtain
 (With) my resources and my army, I will obtain kingdoms for you.

Such examples can be handled by generating the clitic between the two conjoined constituents and then moving it one word to the right with an obligatory transformation.³¹

(53) Deep structure

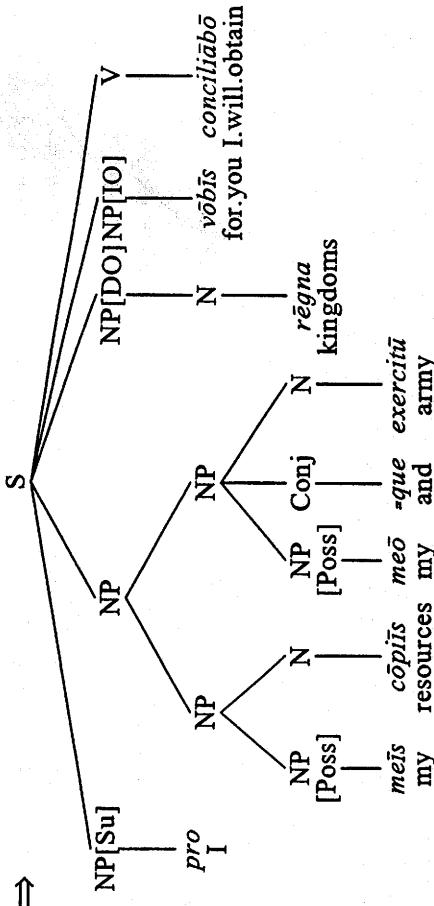


Spanish are a bit more complex than discussed here, however, since object pronouns are proclitics only when the verb is finite; on nonfinite verbs they are enclitics.

³⁰From Hines, Welch, and Hopkinson (1966:207), which is an adaptation from Caesar's Gallic War, Book 1, chapter 3. The meaning 'with, by means of' comes from the ablative case marking on the nouns; this structure is normal for instrument obliques in Latin.

³¹The discussion here makes the standard assumption that conjunctions join two constituents of the same type to form a larger constituent of that same type. In addition to the transformation that moves the clitic, there are inflectional spellout rules; for simplicity of presentation, I omit the inflectional features from the trees. The stem forms in the deep structure are based on the traditional analysis of the morphology. The tree assumes a flat clause structure for Latin, but this is not crucial.

(54) Surface structure



Another type of complication sometimes occurs with pronominal clitics, which may sometimes act just a bit like agreement markers. For example, indirect object clitics in Spanish, unlike the direct objects, sometimes do co-occur with an overt indirect object PP.

- (55) a. Full indirect object PP only

Martín dió la rana a su madre.
 Martin give(3sgPast) the frog to his mother
 Martin gave the frog to his mother.

b. Clitic indirect object pronoun only

Martín le- dió la rana.
 Martin 3sgIO- give(3sgPast) the frog
 Martin gave the frog to her.

c. Clitic pronoun together with a full PP

Martín le- dió la rana a su madre.
 Martin 3sgIO- give(3sgPast) the frog to his mother
 Martin gave the frog to his mother.

This makes them look more like agreement markers (affixes) and not pronouns (separate syntactic words). Yet, they are not fully like agreement affixes, either, since they are optionally absent when a full PP is present; agreement markers are present at all times.³² When a language allows a prounoun clitic to co-occur with a full NP or PP, we say the language allows CLITIC DOUBLING.

Formal treatment of clitic doubling is too complicated for this book. Don't try to write formal rules to account for it; just describe informally what word-like and affix-like traits each clitic has and list it in some appropriate category in the lexicon. The same is true for the many other special behaviors that clitics exhibit, which are too numerous to discuss here. Describe the full range of facts informally, take the formal analysis as far as you can, and start reading in the literature for ideas on how to proceed further.

20.8. Writing clichés in practical orthographies

One important practical consideration that comes up when analyzing clitics is how to write them in a practical orthography: as separate words or affixes. (There's more to designing a practical orthography than inventing an alphabet; you must also consider how to use spaces, hyphens, and other punctuation marks.) When you are making a first attempt at designing a practical orthography

³²Full noun phrases can occur without a doubling clitic only when they have the semantic roles of recipient or addressee.

a good rule of thumb is to use spaces to reflect syntactic word boundaries. That is, phrasal affixes should probably be written attached and bound words written as separate words.

However, the reaction of native speakers is usually a very important factor in orthography design, not just the linguistic facts. In many languages speakers prefer to write bound words attached, since that is the way they sound. (You may want to try using a hyphen or an equals sign with bound words in this case.)

As with all orthography design, this is more than just asking people about their preferences. You should also consider ease of learning (for new readers) and ease of use (for experienced readers and writers). Evaluating these factors may require careful testing, helping people learn to read and monitoring their progress, to see which of several options gives the best results overall. The final decisions on practical orthographies should normally be made not by linguists but by the people who are going to have to live with them: those who depend on the language as their primary means of communication.³³

20.9. Review of key terms

When dividing a sentence into WORDS by finding WORD BOUNDARIES (or WORD BREAKS), we can identify either PHONOLOGICAL WORD BOUNDARIES or SYNTACTIC WORD BOUNDARIES, depending on whether we are concerned with phonological or syntactic characteristics. When a morpheme has some characteristics of words and some of affixes, it is commonly called a CLITIC; clitics are typically PHONOLOGICALLY BOUND but SYNTACTICALLY FREE. Clitics often display other traits such as being members of closed classes with grammatical meaning, occurring at the edges of words, PROMISCUOUS ATTACHMENT, and (for pronoun clitics) mutual-substitutability for full noun phrases. Clitics may be classified as PROCLITICS or ENCLITICS and as BOUND WORDS or PHRASAL AFFIXES. (UNSTRESSED WORDS are sometimes also considered clitics, but we have not done so.) Sometimes clitics have special grammatical properties, such as appearing in an unexpected position (especially SECOND POSITION CLITICS) or CLITIC DOUBLING.

20.10. Questions for analysis

For each morpheme that is phonologically bound:

1. Is there any evidence that it is syntactically free? (Is it a bound word?)
2. If not a bound word, is it positioned with respect to a phrase instead of a word? (Is it a phrasal affix?)
3. What syntactic and phonological properties does this morpheme share with clear cases of words?
4. What ones does it share with clear cases of affixes?
4. Does it have any special grammatical properties, such as occurrence in an unexpected position or clitic doubling?

20.11. Sample descriptions

Latin -que

Latin has a conjunction, **que*, which is an enclitic. Rather than occurring between the two conjoined phrases, it attaches to the end of the first word in the second phrase.

³³Exactly how to go about making such a decision in a particular society (or encouraging and enabling others to make it) is not something that I pretend to give advice on, nor do I mention here all the relevant factors to be considered in designing a useful practical orthography.

- (1) *senātus populus =que rōmānus*
 senate people and of.Rome
 the senate and people of Rome
- (2) *Meis cōpīs meō =que exercitū régna vōbīs conciliābō*
 my resources my and army kingdoms for.you I.will.obtain
 With my resources and my army, I will obtain kingdoms for you.

Spanish object pronouns

Spanish direct and indirect object pronouns are proclitics attached to the verb, rather than occurring in the normal object position after the verb. Indirect object clitics precede direct object ones.

- (1) a. *Manuela lo- tiró.*
 Manuela 3sgDO= threw
 Manuela threw it.
- b. *Le- dió la rana.*
 3sgIO= gave the frog
 He gave the frog to her.
- c. *Me- lo- dió.*
 1sgIO= 3sgDO= gave
 He gave it to me.

Indirect object clitics sometimes exhibit clitic doubling, but this does not occur (in most dialects) with direct objects.

- (2) a. *Le- dió la rana a su madre.*
 3sgIO= gave the frog to his mother
 He gave the frog to his mother.
- b. **Manuela lo- tiró el voleibol.*
 Manuela 3sgDO= threw the volleyball
 (Manuela threw the volleyball.)

20.12. For further reading

Matthews 1974:20–36 has an extensive discussion of the three different senses of WORD and the importance of distinguishing among them.

For more discussion of cliticization and several recent theoretical approaches to understanding it, see Spencer 1991:350–94. This chapter includes detailed descriptions of clitic systems in three Slavic languages, which illustrate many of the characteristics commonly associated with clitics around the world (some of which have not been introduced here).

The way of classifying clitics introduced in this chapter is based largely on work by Arnold Zwicky in collaboration with various others. Three of his articles might be especially interesting at this point. Zwicky and Pullum 1983 provides half a dozen criteria that can be used for deciding whether a word is a clitic or an affix. It illustrates how these criteria can be applied to data by arguing that the English contraction *n't* is not a clitic (and thus not like the contracted form 'of is'), but rather an ordinary suffix that occurs only on auxiliary verbs. Zwicky 1985 provides a number of other such guidelines; perhaps more importantly it discusses the proper role of guidelines like these in linguistic analysis. It also discusses (and dismisses) the related notion of 'participle' as a useful linguistic concept. Zwicky 1987 provides an analysis of the fact that the English possessive 's is

absent after the plural suffix *-s*. These articles are somewhat more technical than works recommended in other chapters, and two use classifications of clitics that differ from what is used here (this chapter follows Zwicky 1987). But you do have the background now to understand much of what they contain, and they are well worth reading as examples of how to apply linguistic theories to analytic problems in individual languages, as well as providing helpful perspectives on larger issues within the field.

Although most of the factors that affect the design of practical orthographies come from other disciplines like phonology, sociolinguistics, and education, some aspects of morphology and syntax may also be relevant, as discussed above. There are several books available that provide a reasonably comprehensive treatment of the subject, including a number of case studies in specific languages showing the practical application of the basic principles of orthography design. Venezky 1970 provides a compact summary of principles of orthographic design, especially letter choice and spelling conventions. For more details, there are some book-length treatments. Smalley 1963 is a collection of previously-published articles which includes a couple of articles specifically on word division. Though it discusses many linguistic issues in terms of older theoretical frameworks, the issues themselves are still relevant. Fishman 1977 builds on the earlier book, including perspectives from work in Generative Phonology in the 1960s and 1970s. It examines especially the role of social, political, and other nonlinguistic factors in design, revision, and acceptance of writing systems. Finally, Weber (1993, to appear), especially chapter 2, provides an extensive case study of how one might apply such principles to specific languages and illustrates some of the sharp differences of opinion that can arise over practical writing systems, which are often very controversial and politically-sensitive issues.

for. Think, too, of what type of information you need to write formal rules; let your theoretical framework guide you.

You will want to be selective. Concentrate on getting *useful* examples; rather than many of them. Having too many randomly-collected examples leads to a disorganized mess that is difficult to sort through when constructing an analysis. Be purposeful in the data you select. Quality is much more important than quantity; each example should ideally provide some new piece of information to move your analysis forward.

Two types of data are especially useful.

- Pairs or sets of examples that are virtually identical, but which differ in only one small detail.

For example, examples that mean ‘Tomorrow I will go to my sister's house,’ and ‘Yesterday I went to my sister's house’ could provide helpful information about tense marking on the verb.

- Sets of examples that show the full range of variation possible for a particular construction.

For example, if you're trying to write phrase structure rules for clauses, you'd want examples of intransitive, transitive, and ditransitive clauses, both with and without various combinations of obliques.

In most files that focus on the internal structure of some constituent, the primary goal is a phrase structure rule. Recall what types of subconstituents typically occur in constituents of this type, and look specifically for them. For example, in an analysis of noun phrases, look for ways of expressing definiteness, quantity, quality, etc. Look for maximal or near-maximal expansions (to find the relative order of elements) and minimal expansions (to find out which elements are optional).

For files that focus on closed classes, such as prepositions or tense affixes, try to compile a complete list of such elements. That is, you want at least one example of each different preposition (but not every example of every preposition that occurs in your data).

26.5. Organizing your filing system

Keep your files in some organized sequence, so you can find them easily. For example, you might adopt an overall organization similar to the following:

1. Sentence structure
 - a. Commands
 - b. Yes-no questions
 - c. Content questions
 - d. Focus
- e. Topicalization
2. Clause structure (in statements)
 - a. Basic (active) clauses (subject, objects, obliques)
 - b. Nonactive complements
 1. nominal complements
 2. adjectival complements
 3. locative complements
 4. existential clauses
- c. Variable orders of clausal constituents (other than what is covered under sentence structure)
- d. Relative clauses
- e. Complement clauses
- f. Oblique clauses

3. Phrase structure
 - a. Noun phrases
 - b. Verb phrases (if not handled under clause structure)
 - c. Adjective phrases
 - d. Quantifier phrases
 - e. Adverb phrases
 - f. Prepositional or postpositional phrases
4. Word-level categories and subcategories
 - a. Verbs
 - b. Noun
 - c. Adjectives
 - d. Numerals
 - e. Other quantifiers
 - f. Determiners
 - g. Pronouns
 - h. Prepositions or postpositions
 - i. Adverbs
 - j. Unmodifiable single-word obliques
 - k. Interrogative (WH) words
1. Conjunctions
- m. Others, especially closed classes, i.e., particles
5. Clitics (one file for each suspected clitic or each class of similar clitics, including reasons for considering it an ordinary word, bound word, phrasal affix, or ordinary affix)
6. Inflectional morphology (one file for each category of word or each position class, including paradigms, position class charts, complete lists of affixes and their glosses, formal analysis, etc.)
7. Derivational morphology (one file for each derivational affix or set of related affixes)

Alternately, you can use the table of contents from a published grammar in a related language. Whatever model you follow, don't follow it slavishly; adapt it to the structure of the language, your analysis, and your own personal needs. As you work, add new files and rearrange them as your understanding of the language grows. At any one time, though, you should keep things in some logical sequence. It helps to make a 'table of contents' similar to the list above for your filing system and to keep things arranged in that order.

On a computer, it can be helpful to include numbers or letters at the beginning of the file name so the files will be displayed in order. If necessary, use leading zeros so all the numbers have the same number of digits. So, for example, the above outline might be packaged into computer files with names like the following:

- (1) 01a commands
- 01b yes-no questions
- 01c content questions
- 01d focus
- 01e topicalization
- 02a active clauses
- 02b1 nominal complements
- 02b2 adjectival complements
- (etc.)

(On some operating systems, you'll have to abbreviate the file names to 8 characters and omit spaces.)

In a manual system, by all means, *keep all the pages of the same file together and in order!* (I wish I didn't have to mention that; my apologies to those who wouldn't dream of doing otherwise.)

26.6. Making the system work for you (and not vice versa)

There are two opposite but equally dangerous traps that people fall into with regards to filing. The usual mistake is to file too little and do most of the analysis on the original copy of the data. This places a tremendous burden on your memory, as you try to remember where all the relevant facts are for a particular topic, what hypotheses you have considered so far, etc. This leads to inefficiency at best and incomplete or incorrect understanding of the language at worst. Do enough filing to move your analysis along towards your goals.

The other trap is to file aimlessly and tediously, without a clear idea of what you are trying to do. You must keep your goals clearly in mind and stay aware of whether you are making any concrete progress towards achieving them. Don't just pile up mountains of data. Often, people do this out of confusion and anxiety; they keep themselves busy for hours, but don't make any progress. If you tend towards doing this, make a point of stepping away from your work every hour, quieting your mind, and putting things back in perspective. Establish habits now that will serve you well both at home and on the field.

26.7. Characteristics of a good grammar filing system

Good grammar files have the following characteristics:

- They help you make progress in your analysis.
 - They are neat enough that you and your close colleagues (e.g., teachers) can read them without being distracted by their appearance.
 - They have adequate organization:
 - They are arranged in some logical order, so you can find information quickly.
 - They are divided into small enough sections so that each section provides a focused and thorough treatment of just one topic.
 - They are easy to expand and modify.
 - They treat data with respect:
 - Data and analysis are clearly separated.
- Each example can be traced back to its source (where it was first transcribed or recorded).
- The analysis includes informal observations, such as:
 - brief descriptions of the data
 - hypotheses about rules
 - plans to obtain further data
 - predictions made by your rules
 - problems that you haven't covered
 - The analysis includes elements of a formal analysis to make your hypotheses precise, plus explanations (when appropriate) about how this analysis works and why it is better than others:
 - rules
 - lexical entries
 - trees
 - etc.

26.8. For further reading

Healey (1975:427–51) and Samarin (1967:151–74) include information about various types of manual filing systems. Although such systems are gradually being replaced by computers, much of this information is still useful as it relates to principles applicable to any filing system. Further, some tasks (especially relatively small ones) will always be better suited to manual filing techniques

than computerized ones, since there is usually considerable overhead involved in setting up a computerized filing system.

Although people have been using computers to file grammatical data for many years, systems that allow one to do so without much technical knowledge are fairly recent. In the Macintosh world, the program called HyperCard has proven itself quite useful, as Valentine (1990) demonstrates. For MS-DOS, Davis and Wimbish (1993) describe the Shoebox computer program and a grammar file outline that is packaged with it, which provides capabilities for grammar and lexical filing and interlinear text glossing. Shoebox is also available in a Windows/Macintosh version from the Summer Institute of Linguistics (<http://www.sil.org> on the Internet). For Windows (and eventually Macintosh), SIL has developed a software system called LinguaLinks which includes some grammar filing capabilities as well as lexical filing, interlinear text glossing, structured editors for producing grammar sketches, and ready access to a wealth of reference material.

Payne (1997) provides a great deal of helpful advice for preparing a grammar sketch, including information that would be useful in setting up a grammar filing system: an outline of topics to cover, questions to consider about each, and many examples from different languages illustrating the possible answers to those questions. For a much more detailed outline of the grammar of a language, see the Lingua Descripta questionnaire (Comrie and Smith 1977), which also provides a wealth of questions to consider, much like the questions for analysis in this book. Be warned, however, that it is very detailed (72 pages!), so you probably want to save it for use on the field, and even then be selective in your use of it.