Analyzing Syntax

A Lexical-functional Approach

Paul R. Kroeger

CAMBRIDGE

Analyzing Syntax

Analyzing syntax: a lexical-functional approach is a comprehensive and accessible textbook on syntactic analysis, designed for students of linguistics at advanced undergraduate or graduate level. Working within the 'Lexical Functional Grammar' (LFG) approach, it provides students with a framework for analyzing and describing grammatical structure, using extensive examples from both European and non-European languages.

As well as building on what linguists have learned about language in general, particular attention is paid to the unique features of individual languages. While its primary focus is on syntactic structure, the book also deals with aspects of meaning, function, and word structure that are directly relevant to syntax.

Clearly organized into topics, this textbook is ideal for one-semester courses in syntax and grammatical analysis.

PAUL R. KROEGER is Associate Professor and Head of Department at the Graduate Institute of Applied Linguistics, Dallas. He has previously published *Phrase structure and grammatical relations in Tagalog* (1993). He has also carried out linguistic fieldwork in East Malaysia, and has written for many journals including *Pacific Linguistics, Oceanic Linguistics*, and the *Philippine Journal of Linguistics*.

Analyzing Syntax

A Lexical-Functional Approach

PAUL R. KROEGER

Graduate Institute of Applied Linguistics, Dallas

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo, Delhi

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521016544

© Paul R. Kroeger 2004

This publication is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 2004
Third printing 2008

Printed in the United Kingdom at the University Press, Cambridge

A catalogue record for this publication is available from the British Library

Library of Congress Cataloging in Publication data
Kroeger, Paul, 1952–
Analyzing syntax : a lexical-functional approach / Paul R. Kroeger.
p. cm.
Includes bibliographical references and index.
ISBN 0 521 81623 8 ISBN 0 521 01654 1 (pb.)
1. Grammar, Comparative and general – Syntax. 2. Lexical-functional grammar.
I. Title.
P291.K76 2004
415 – dc21 2003055382

ISBN 978-0-521-81623-6 hardback
ISBN 978-0-521-01654-4 paperback

Cambridge University Press has no responsibility for the persistence or accuracy
of URLs for external or third-party internet websites referred to in this publication,
and does not guarantee that any content on such websites is, or will remain,
accurate or appropriate.

For my parents, Dick and Cathie

Contents

Preface and acknowledgments

This book provides a framework for analyzing and describing grammatical structure, building on what linguists have learned about language in general while paying careful attention to the unique features of each particular language. Its primary focus is on syntax (sentence structure), but it also deals with aspects of meaning, function, and word structure that are directly relevant to syntax.

This is a book about syntactic analysis, rather than syntactic theory. I have adopted a simplified version of Lexical Functional Grammar (LFG) as the analytical framework for the book, but I have tried in each chapter to emphasize linguistic phenomena over formal notation. The analyses presented here are very much in the spirit of LFG, but the notation employed is modified and simplified compared to that of standard LFG. Those readers who want a more complete introduction to LFG as a formal system are encouraged to consult Bresnan (2001), Dalrymple (2001), or Falk (2001).

This book is written at a level which should be appropriate for advanced undergraduate or beginning graduate students. It presupposes some familiarity with basic linguistic concepts and terminology, but no previous background in formal syntax. The contents can be covered fairly easily in a typical semester-length course. The book does not assume that its readers have native-speaker intuitions about English. So, for example, in some places the meanings of English idioms are explained, alternative possible interpretations of certain constructions are explicitly spelled out, etc.

Only a few exercises are included in this volume, and they are clustered in the first five chapters of the book. Most of these exercises focus more on the interpretation of linguistic evidence than on the primary analysis of unfamiliar language data. Many teachers will want to supplement these with other, more analytical, data problems, depending on the needs and background of the students. In teaching this material, I have also found it extremely helpful to have students write a short research paper about some aspect of the syntax of their own language, or another language that they know well. Exploring one particular issue in greater depth helps to solidify their grasp of the framework as a whole. This project also gives students a small taste of what it feels like to do original research, and gets them reading some of the relevant linguistic literature.

It would be only a slight exaggeration to say that every linguist I know has helped me with this project in one way or another, and it is impossible for me to list them all by name. But, in addition to specific contributions mentioned in

the footnotes, special thanks are due to the following people for linguistic and editorial advice: Wayan Arka, Dorothee Beerman, Joan Bresnan, Don Burquest, Mary Dalrymple, Yo Matsumoto, Sam Mchombo, Brian O'Herin, Jane Simpson, Fu Tan, and Janet Watkins. Thanks also to the faculty, staff, and students of the Linguistics Department at Stanford University, who did so much to make my sabbatical there in 2000–01 enjoyable and productive. Finally, thanks to my wife, Chaw-Nen, for her patience with this very time-consuming process.

Abbreviations

1	1st person
2	2nd person
3	3rd person
A	agent (trans. subject)
ABS	absolutive
ACC	accusative
ACCOMP	accompaniment
ADNOM	adnominal
ADVBL	adverbializer
ANTIPASS	antipassive
APPL	applicative
ART	article
ASP	aspect
ASSOC	associative
AUX	auxiliary
AV	active voice (Austronesian: Actor voice)
BEN	benefactive
CAUS	causative
CLASS	classifier
COMP	complementizer
COP	copula
DAT	dative
DEF	definite
DS	different subject
DU(AL)	dual
DV	dative voice (Austronesian)
EMPH	emphatic
ERG	ergative
EX(CL)	exclusive
F(EM)	feminine
FOC	focus
FUT	future tense
GEN	genitive
HABIT	habitual
HON	honorific

IMPERF	imperfective
INCHOAT	inchoative
INCL	inclusive
INDIC	indicative
INF	infinitive
INSTR	instrumental
INTERROG	interrogative
intr.	intransitive
INVOL	involuntary/non-volitional
IRR	irrealis
IV	instrumental voice (Austronesian)
LNK	linker
LOC	locative
M(ASC)	masculine
MOD	modality
N(EUT)	neuter
NEG	negative
NMLZR	nominalizer
NOM	nominative
NON.FIN	non-finite
NONPAST	nonpast tense
O.agr	object agreement
OBJ	primary object
OBJ_2	secondary object
OBL	oblique argument
OV	objective voice (Austronesian)
P	patient (trans. object)
p.c.	personal communication
PASS	passive
PAST	past tense
PERF	perfective
PFCT	perfect
PERS	person; personal name
pl / PL	plural
POL	politeness
POSS	possessor
POTENT	potential
PRES	present tense
PRO	pronoun
PROG	progressive
PRTCPL	participle
PTCLE	particle
PURP	purpose
Q(UES)	question

REAL	realis
REC.PAST	recent past tense
REL	relativizer
REM.PAST	remote past tense
S(UBJ)	subject
S.agr	subject agreement
S-ADJ	sentential adjunct
S-COMP	sentential complement
SBJNCT	subjunctive
SEQ	sequential
sg /SG	singular
SIMUL	simultaneous
SS	same subject
TOP	topic
tr./TRANS	transitive
UV	Undergoer voice (Austronesian)
XADJ	open adjunct
XCOMP	open complement (or predicate complement)

1 Three aspects of syntactic structure

Probably no one has ever before said or heard the following sentence, yet any normal adult speaker of English will understand it:

(1) John Adams could have been elected to a fourth term as President, if his step-sister had not been so ugly.

In the same way, a speaker of any language will say and hear many sentences during the course of a normal day which he has never said or heard before. Moreover, other speakers of the same language will not only recognize these original creations as being well-formed sentences but will also (usually) understand what they mean.

These observations tell us something important about the nature of language. A person who knows how to speak a language does not have to memorize every possible sentence in that language. Rather, speakers produce sentences CREATIVELY. Some common phrases and sentences may be repeated so often that they are memorized as a single unit, e.g., idioms and proverbs. But, for the most part, we do not memorize sentences; rather, we construct them when we need them, to express a particular idea.

This creative use of language is possible because the patterns of a language are determined by a set of RULES. A speaker who (unconsciously) "knows" these rules can use them to create and understand any number of new sentences. Linguists use the term GRAMMAR to refer to the set of rules needed to produce well-formed utterances in a particular language. The chief concern of this book will be to help you analyze and describe the kinds of grammatical patterns commonly found in human languages, in particular those aspects of the grammar that are relevant to SYNTAX, i.e., the arrangement of words in a sentence.

1.1 Grammar and grammaticality

Our ultimate goal, as linguists, is to understand the speaker's "internal grammar," i.e., the system of rules which a speaker unconsciously uses in speaking and listening. Since speakers are not aware of using these rules, they cannot simply tell us what the rules are. (As noted below, we are not speaking here of PRESCRIPTIVE rules, like those commonly taught in school. Those would be easier for speakers to describe, because they are consciously learned.) We begin by making observations about what people say, what they do not say, how they

interpret what other speakers say, etc. Our description of the grammar, based on careful analysis of these observations, will involve formulating explicit rules that can model (i.e., produce the same grammatical patterns as) the speaker's internal grammar.

Of course, in order to speak and understand a language we must know not only rules but also words. Linguists use the term LEXICON to refer to the set of all the words (or, more generally, meaningful elements) in the language. We can think of the lexicon as the speaker's "mental dictionary." Much of the information in the lexicon is unpredictable, such as the fact that the pronunciation /kæt/ refers (in English) to a small carnivorous mammal with whiskers. But the lexicon contains a number of regularities as well. Precisely how much regularity is a hotly debated question; we will return to this issue in chapter 3. In this chapter, we introduce some basic concepts needed for discussing grammatical systems.

1.1.1 "Grammaticality" and variation

In claiming that sentence (1) is a well-formed English sentence we do not imply that it is either true or sensible. In fact, it is stunningly illogical and historically inaccurate on several counts. (That is why no one else is likely ever to have said it before.) This distinction between the form of a sentence and its meaning is an important one. In some cases we might consider a sentence to be well-formed even when it has no sensible meaning at all. Chomsky (1957) used the following pair of sentences, which have become famous through countless repetitions, to illustrate this point:

(2) a Colorless green ideas sleep furiously.
b Furiously sleep ideas green colorless.

Chomsky claimed: "Sentences [2a] and [2b] are equally nonsensical, but any speaker of English will recognize that only the former is grammatical." The second part of this claim has sometimes been disputed, but the essential point is that speakers have an intuitive sense of "grammaticality," or correctness of form, which does not depend on our ability to interpret the meaning of the sentence.

Chomsky argued that any speaker of English would consider sentence (2a) to be "grammatical," even though it makes no sense. (More specifically, the meanings of the individual words in this sentence are not compatible with each other, and so this combination of word-meanings produces contradictory information.) Conversely, as the following sentences show, we can often understand a sentence perfectly well even if it is not "grammatical":

(3) a Me Tarzan, you Jane.
b Those guys was trying to kill me.
c When he came here?

When speakers reject the sentences in (3) as being "ungrammatical" or "bad English," they mean that one (or more) of the rules of their internal grammar

has been violated. One way in which a linguist tries to understand this internal grammar is to formulate a set of rules which will model the judgments of a native speaker: a set of rules which will produce all of the sentences that speakers consider to be grammatical but none of the sentences that speakers consider to be ungrammatical.

This may sound like a fairly straightforward goal, but there are a number of factors which complicate the process. For one thing, speakers of a given language do not always agree about the grammaticality of particular sentences. Each of the sentences in (4a, b, c), for example, would be perfectly acceptable to English speakers from certain geographical areas, but sound quite odd to English speakers from other areas. This kind of variation among regional dialects is found at least to some extent in most of the world's languages. Sentence (4d) involves a different kind of variation, namely variation among individual speakers. It is not associated with a particular region; even speakers from the same dialect area may differ as to whether they would say such sentences, or accept them as being grammatical.[1]

(4) a I might could be persuaded to try that.
b My back door needs fixed.
c The ship is arriving Monday week.
d They have come visited us every day this month.

Of course language is a means of communication, and if two speakers have radically different internal grammars, communication will be extremely difficult if not impossible. For many purposes, we can speak of "the grammar" of a language as if it were a body of knowledge which all speakers of that language must share; and to a very large extent this is true. But we must always be aware of the variation among dialects and individuals which this over-simplified view ignores. Indeed, the variation itself is also something that linguists seek to document and explain.

Another challenge to the linguist who wants to describe and model the rule system of a language is that languages are always changing. Moreover, community attitudes about language often change more slowly than the actual practice of the community in speaking the language. Consider the examples in (5).

(5) a With two things hath God man's soul endowed.
b I know not what course others may take, but as for me . . .
c The problem is, is that no one wants this job.
d The mission of the *Enterprise* is to boldly go where no man has gone before.

The word order of (5a) was perfectly normal in Old English (before 1100 AD), but most speakers of modern English will probably consider it extremely odd, if not actually ungrammatical. Example (5b) is taken from a famous speech by Patrick Henry (1775 AD). Similar examples are very common in Shakespeare and the King James Bible (1611 AD), but today they are regarded as archaic or (in modern usage) unnatural.

The construction in (5c) is a much more recent innovation. It is now very widely used in informal speech, at least in American English, but most educated

speakers would probably reject it in formal written styles. Moreover, standard reference books on English grammar, school textbooks, etc. do not recognize it as an acceptable way of speaking.

Sentence (5d) is an example of a "split infinitive," because the adverb *boldly* appears in the middle of the infinitive *to go*. This pattern emerged in the fourteenth century due to a cluster of morphological and syntactic changes in Middle English, and has been used quite commonly ever since (Hall, 1882; Kiparsky, 1997). However in the nineteenth century, as part of a growing concern for defining "correct" usage in English, influential authorities began to assert that the split infinitive was "bad English" (i.e., ungrammatical), apparently because no such pattern exists in Latin. This judgment is maintained in school textbooks to the present day.

The case of the split infinitive is a notorious example of a PRESCRIPTIVE approach to grammar: grammarians telling other people how to talk. Of course, there are many contexts where a prescriptive approach is appropriate and indeed necessary: e.g., in planning and developing the standard form of a new national language, in teaching adult language learners to speak correctly, in teaching high school students to write acceptable essays, etc. But these areas of "applied linguistics" will not be our primary focus.

The approach we adopt in this book will be DESCRIPTIVE rather than prescriptive. This means that we take it as our goal to observe, describe, and analyze what speakers of a language actually say, rather than trying to tell them what they should or should not say. Most of our examples will come from published sources. As a result, these examples often represent a standardized or high-prestige variety of the language.[2] But the same approach can be applied to data from non-standard dialects, or even languages which have no written form at all. Indeed, one of the goals of this book is to equip you to undertake this kind of research.

Our approach will be primarily SYNCHRONIC; that is, we will be primarily interested in describing the structure of a particular language or dialect at a particular time (normally the present), rather than comparing it with related dialects or investigating how the language has changed over time.

As we noted in the introduction to this chapter, SYNTAX is the branch of linguistics which seeks to describe and account for the arrangement of words in a sentence. In order to do this we will often need to look at the structure of the words themselves, i.e., at certain aspects of the MORPHOLOGY. For example, in some languages the presence of a certain prefix or suffix on the verb determines where other words in the sentence may or may not occur. Moreover, even though form and meaning are partially independent of each other, as we have seen, they are also intimately connected. So in studying the form of sentences we will often be led to consider their SEMANTICS (i.e., meaning) as well.

Finally, in addition to the form and meaning of individual sentences, we will sometimes need to consider connections between one sentence and another, or the function of a particular sentence in a specific context. (The study of these aspects of language use is called PRAGMATICS.) And we may occasionally mention stylistic or social factors which influence the kind of sentence patterns a speaker

might use; but our primary focus will be on the sentence patterns themselves. Morphology, semantics, pragmatics, and sociolinguistics are major fields of study in their own right, but in this book we will touch on them only where they are directly relevant to the syntactic issues that we consider.

1.1.2 Sentence structure

A sentence is not simply a string of words, one after another. An overwhelming body of linguistic and psycho-linguistic evidence shows that speakers think of sentences as having a fairly complex structure, with certain words grouped together to function as units, larger groups formed from smaller groups, and important relationships defined between one group and another.

Of course, these structural relationships are "invisible," because all we hear is the string of words. In some ways the task of the linguist is similar to that of early chemists. By observing changes in the physical properties of substances when they were combined in various ways, these chemists were able to develop theories about the unseen structures of atoms and molecules which could account for their observations. In the same way, the linguist tries to understand unseen linguistic structures based on observations about what can be heard.

One reason for thinking that sentences do in fact have this kind of abstract structure is that a given string of words may be AMBIGUOUS, i.e., allow two different interpretations, even when none of the individual words in the string is itself ambiguous. Consider the following examples (adapted from Huddleston, 1984).

(6) a Liz attacked the man with a knife.
b Ed likes Sue more than Jill.
c The proposal that Hitler was advancing seemed preposterous.

The meaning of example (6a) depends on the structural relationships of the phrase *with a knife*. This sentence could be used to answer two different questions: *Who did Liz attack? The man with a knife* or *What did Liz attack the man with? With a knife*.

The meaning of example (6b) depends on the relationship of the word *Jill* to the verb *like*: is *Jill* to be understood as the subject of *like* (as in *Ed likes Sue more than Jill likes Sue*), or as the object of *like* (as in *Ed likes Sue more than Ed likes Jill*)? The meaning of example (6c) depends on the relationship between the word *proposal* and the phrase *that Hitler was advancing*. Is the proposal something which is being advanced by Hitler? Or is Hitler himself supposedly advancing, and the proposal merely a report (by some other person) of this event?[3] Such examples of STRUCTURAL AMBIGUITY can provide important evidence about the structures which speakers assign to sentences.

Languages show great variety in terms of the strings of words that they use; that is one of the reasons why word-by-word translation fails so miserably. But when we compare abstract structural properties, languages turn out to be much more similar to each other than we might have guessed. In some respects the

variation among languages is surprisingly limited. For this reason, studying data from a wide range of languages can be a great help to us in knowing what to look for when we tackle a new language, particularly if that language has not been analyzed before.

1.2 Outline of a framework

The job of the syntactician can be divided into two main steps: first, determine the correct structure(s) for each grammatical sentence; and second, formulate a set of rules which will distinguish between grammatical and ungrammatical structures for that language (i.e., allow us to predict which structures will be grammatical and which ungrammatical). Of course, the two tasks are closely related, and we will be concerned with both. But we will devote most of our attention in this book to the first of these issues, in particular to the kinds of evidence that are relevant for determining linguistic structure.

In order to discuss the details of syntactic structure with any kind of precision we will need to develop (i) a technical vocabulary; (ii) a system for representing structural relations; and (iii) a set of concepts which are relevant to this task. Such an inventory of vocabulary, notation and concepts is called a syntactic FRAMEWORK.

A good framework must do at least three things. First, it should make it easy to describe the syntactic patterns found in any particular language. Second, it should make it easy to compare syntactic patterns between languages. Third, it should allow us to make generalizations about human language in general, i.e., to state THEORIES (factual claims about how language works).

In the remainder of this chapter, we will sketch out the beginnings of a framework for syntactic analysis. As a way of introducing some of the concepts which we will need to use in talking about syntactic structure, let us first think about structural complexity. What makes a sentence "complex"? Which of the following Malay sentences is the most complex? Try to rank them, from simplest to most complex.

(7) a Dia mandi.
3sg bathe
'He is bathing.'

b Saya makan nasi.
1sg eat cooked.rice
'I eat/am eating rice.'

c Orang tua itu makan nasi goreng setiap hari.
person old that eat rice fry each day
'That old person eats fried rice every day.'

d Dia belajar untuk menjadi pensyarah.[4]
3sg study in.order become lecturer
'He is studying to become a lecturer.'

One way to measure complexity, though perhaps not the most revealing way, is by absolute length (number of words). On this basis, (7a) is clearly the simplest and (7c) the most complex. Of course, there are other, and more persuasive, reasons for considering (7a) to be the simplest. Semantically it names a single event, 'bathe', which involves just one participant. Sentence (7b), on the other hand, names an event which involves two participants, the "eater" and the "eaten."

Sentence (7c) names an event of exactly the same type as (7b). It uses the same verb (*makan* 'eat') and involves the same number of participants playing the same roles. So in this respect we might evaluate them as being equally simple. But (7c) also contains an additional piece of information, namely the fact that this event occurs 'every day.' This time phrase is, in a sense, added on to the basic description of the event.

In discussing the meaning of a clause, we will use the term PREDICATE to refer to the word which names the action, event, or state described by that clause. Typically this word will be a verb.

Now any event named by the predicate 'eat' must involve at least two participants, the "eater" and the "eaten." (This is true even though one or the other of these participants may not be mentioned in a particular description of the event, e.g., *John is still eating* or *The fish was eaten*.) For this reason we say that the predicate 'eat' takes two ARGUMENTS. But a time phrase like 'every day' is not an inherent part of the meaning of 'eat.' This kind of phrase can be added freely to virtually any clause that describes an event. An extra piece of information of this kind, something that is not an argument of the predicate, is called an ADJUNCT.

The ARGUMENT STRUCTURE of a predicate is a representation of the number and type of arguments associated with that predicate, as illustrated in (8). We will use the general term AGENT to represent the participant who performs a certain action, and the term PATIENT for the participant that something happens to; see section 1.2.1 for further discussion and examples.

(8) *bathe* < agent >
eat < agent, patient >

Argument structure is important to the syntax, because it determines many of the basic grammatical properties of the clause in which the predicate occurs. Argument structure is closely related to meaning, but it is obviously not intended to represent the full meaning of a sentence, or even of a predicate. For example, it is true that (7b) and (7c) have the same argument structure; but that does not mean that the meaning of the two sentences is equivalent. One reason for this is that adjuncts are not a part of the argument structure. Another reason is that the argument structure indicates only the role which each argument plays in the event, but does not give any information about the inherent properties of the arguments themselves, e.g., the fact that the agent in (7c) is an old person.

This last point brings us back to another important difference between (7b) and (7c). Sentence (7b), repeated below as (9a), consists of three basic elements: a subject, a verb, and a direct object. (What we mean by "subject" and "direct object" will be discussed in section 1.2.3.) Each of these elements is named by

a single word. Sentence (7c) (ignoring the adjunct phrase) has the same three basic elements; but here the subject and object are each expressed by a phrase containing more than one word, as shown by the brackets in (9b).

(9) a [Saya] makan [nasi].
1sg eat cooked.rice
'I eat/am eating rice.'

b [Orang tua itu] makan [nasi goreng].
person old that eat rice fry
'That old person eats fried rice.'

The fact that groups of words can function as units (or CONSTITUENTS) within sentences is an important aspect of the grammar of every human language. In analyzing the structure of a sentence, it is very important to identify the constituent boundaries (i.e., to determine which words group together as units), to specify the linear order of constituents in the sentence, and to specify the ordering of the words within each constituent. The aspect of syntactic structure which represents these kinds of information is called CONSTITUENT STRUCTURE (or PHRASE STRUCTURE).

Now let us return to sentence (7d), repeated below as (10). This sentence is shorter than (7c), in that it contains fewer words; and it involves only a single participant (the one who is studying). So at first glance it may look simpler than (7c). But the meaning is more complex. This sentence actually describes two events, 'studying' and 'becoming a lecturer.' The two events have a certain logical relationship to each other: the agent does X in order to achieve Y. Each of these events is named by a distinct predicate, *belajar* 'study' and *menjadi* 'become,' and each predicate has its own argument structure.

(10) Dia belajar untuk menjadi pensyarah. (= 7d)
3sg study in.order become lecturer
'He is studying to become a lecturer.'

Corresponding to this semantic complexity, the grammatical structure of (7d) is considerably more complex than that of (7c). This kind of sentence structure will be discussed in some detail in chapter 5. For now we will just observe that in (7d) we find one CLAUSE (or simple sentence) embedded within another. That is, one clause functions as a constituent of another, specifically in this case as an adjunct. Moreover, the subject of the embedded clause is understood to be identical with the subject of the main clause. The function of the embedded clause within the larger sentence, and the relationship between the subject of the embedded clause and the subject of the main clause, are part of the FUNCTIONAL STRUCTURE of the sentence.

We have mentioned three aspects of the structural complexity of sentence (7d): argument structure, constituent structure, and functional structure. These three aspects of syntactic structure will be important for our analysis of a wide variety

of constructions; so in the remainder of this chapter we will briefly review some of the basic features of each.

1.2.1 Argument structure

As stated above, argument structure is a representation of the number and type of arguments associated with a particular predicate. Determining the number of arguments is not always as easy as one might expect; but identifying the "type" of these arguments may seem even more difficult. What exactly do we mean by this?

In fact this question has been a hotly debated issue among linguists. The approach which we will adopt here is to assign participants to broad semantic or conceptual categories according to the role they play in the described event or situation: "agent" for participants that do something; "patient" for participants to whom something is done; "experiencer" for participants who think or feel something, etc. Unfortunately (but not surprisingly) there is no one set of semantic role labels which all linguists agree on, and different linguists sometimes use the same labels in different ways. But in this book we will refer to (at least) the following semantic roles:

(11) **INVENTORY OF SEMANTIC ROLES:**

AGENT: causer or initiator of events

EXPERIENCER: animate entity which perceives a stimulus or registers a particular mental or emotional process or state

RECIPIENT: animate entity which receives or acquires something

BENEFICIARY: entity (usually animate) for whose benefit an action is performed

INSTRUMENT: inanimate entity used by an agent to perform some action

THEME: entity which undergoes a change of location or possession, or whose location is being specified

PATIENT: entity which is acted upon, affected, or created; or of which a state or change of state is predicated

STIMULUS: object of perception, cognition, or emotion; entity which is seen, heard, known, remembered, loved, hated, etc.

LOCATION: spatial reference point of the event. The LOCATION role includes the sub-types SOURCE, GOAL, and PATH, which respectively describe the origin (or beginning-point), destination (or end-point), and pathway of a motion

ACCOMPANIMENT (or COMITATIVE): entity which accompanies or is associated with the performance of an action

Some examples of the most common of these roles are given in (12):

(12) a *John gave Mary a bouquet of roses.*
AGENT RECIPIENT THEME

b *John baked Mary a chocolate cake.*
AGENT BENEFICIARY PATIENT

c *John opened the lock with a key.*
AGENT PATIENT INSTRUMENT

d *The key opened the lock.*
INSTRUMENT PATIENT

e *Sherlock Holmes heard a piercing scream.*
EXPERIENCER STIMULUS

For many verbs, it is not too difficult to identify the arguments with one or another of these roles; some examples are given in (13). But in other cases a particular argument may not seem to fit naturally into any of these categories; for example, what is the semantic role of *mother* in *Susan resembles her mother*? Or an argument may appear to bear two roles; for example, *John* seems to be both an agent and a theme in *John jumped into the well*.

(13) *dance* <agent>
eat <agent, patient>
love <experiencer, stimulus>
give <agent, theme, recipient>

These kinds of issues have been discussed at considerable length, but for our purposes they do not represent a major problem. The primary function of these role labels is to allow us to distinguish among the arguments of a particular predicate.[5] Thus some linguists prefer to use unique labels for each predicate, as we did above in referring to the arguments of *eat* as the "eater" and the "eaten." But it is also true that the categories in (11) are quite useful for describing a wide variety of grammatical patterns. That is why the use of these labels remains so popular, in spite of certain well-known problems.

Notice that the list in (11) is restricted to ARGUMENT roles. Some other commonly expressed types of semantic information, e.g., time, manner, purpose, etc. are not included here, because the elements which express these concepts are almost always ADJUNCTS rather than arguments. This distinction between arguments and adjuncts is important, but not always easy to make. The basic difference is that arguments are closely associated with the meaning of the predicate itself, while adjuncts are not.

Adjuncts contribute to the meaning of the sentence as a whole, but are never necessary to complete the meaning of the predicate. Thus adjuncts are always optional, whereas arguments are frequently obligatory. Sentence (14a) shows that the object of *use* is obligatory, and therefore an argument. As (14b) illustrates, even when an argument is grammatically optional it may be semantically obligatory; for example, even when the patient of *eat* is not expressed, we know that something gets eaten. But adjuncts can always be omitted without creating this kind of implication, as seen in (15).

(14) ARGUMENTS
a Mary used my shirt for a hand towel.
*Mary used for a hand towel.

b John ate an apple.
John ate. (implies that John ate something)

(15) ADJUNCTS
a George fell down the stairs last night.
George fell down the stairs. (no implied time reference)

b My daughter intentionally swallowed a penny.
My daughter swallowed a penny. (neutral with respect to volitionality)

A second important difference is that arguments must be unique within their clause, but adjuncts may be freely multiplied. For example, the third sentence in (16a) is ungrammatical because it contains two recipient phrases; this shows that the recipient of *give* is an argument. On the other hand, the grammaticality of (16b–c) shows that time and manner phrases are adjuncts, rather than arguments.

(16) a John gave Susan a bouquet of roses.
John gave a bouquet of roses to his mother.
*John gave Susan a bouquet of roses to his mother.

b TIME ADJUNCTS
George fell down the stairs last night at 3:00 AM during the typhoon.

c MANNER ADJUNCTS
My daughter suddenly, impulsively, without thinking, swallowed a penny.

A third difference is that verbs may place SELECTIONAL RESTRICTIONS on their arguments, but not on adjuncts. For example, the patient of *drink* must be a liquid. The patient of *assassinate* must be an important political figure. German has two words for 'eat': *essen*, which requires a human agent; and *fressen*, which requires a non-human agent. Restrictions of this kind, which are associated with a specific verb, are never imposed on adjuncts.

Notice that selectional restrictions of this kind must be stated in terms of semantic roles (agent, patient, etc.) rather than grammatical relations (subject, object, etc.). This is illustrated in (17–19). The examples in (17) show that the patient of *drink* must be a liquid, whether it appears as object or subject. The examples in (18) show that the verb *love* requires an animate experiencer, not an animate subject; (18b), which has an animate subject, is extremely odd, whereas (18a), which has an inanimate subject, is perfectly sensible. And (19) shows that the experiencer of *teach* must be animate, whether it appears as a direct object (19a) or the object of a preposition (19b).

(17) a #John drank his sandwich.
b #The sandwich was drunk by John.

(18) a That book is loved by children around the world.
b #Children around the world are loved by that book.

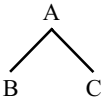
(19) a #Mary taught her motorcycle classical Chinese.
b #Mary taught classical Chinese to her motorcycle.

A number of other criteria for distinguishing arguments from adjuncts have been proposed, but this is a complex issue and we will not pursue it further here.

1.2.2 Constituent structure

The CONSTITUENT STRUCTURE of a sentence contains information about constituent boundaries, linear order and syntactic categories (parts of speech). The most commonly used notation for representing constituent structure is the phrase structure tree diagram. A very simple tree diagram is shown in (20):

(20)



This tree contains three NODES. The top-most node, *A*, is the MOTHER of the two lower nodes, *B* and *C*. *B* and *C* are DAUGHTERS of the same mother, and so we refer to them as SISTER nodes. The simple tree in (20) represents a constituent of category *A* which is composed of two parts, one of category *B* and the other of category *C*, occurring in that order.

When a tree diagram is used to represent the constituent structure of a grammatical unit (e.g., a phrase or sentence), syntactic categories are used to label the nodes; the most common of these are listed in (21).

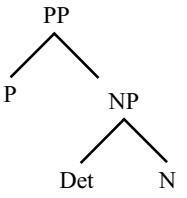
(21)

Word-level		**Phrasal**	
N	= noun	NP	= noun phrase
A	= adjective	AP	= adjective phrase
V	= verb	VP	= verb phrase
P	= preposition	PP	= preposition phrase
Det	= determiner	S	= sentence or clause
Adv	= adverb		
Conj	= conjunction		

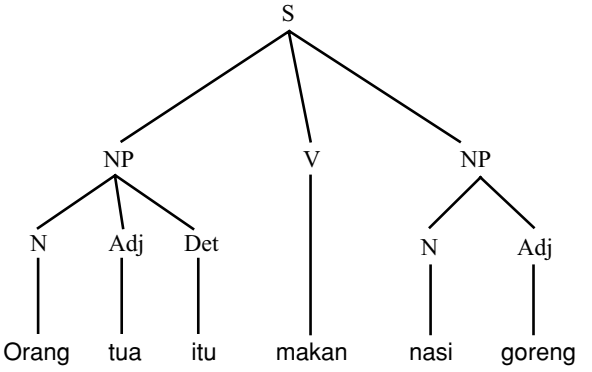
A CONSTITUENT within a sentence corresponds to all and only the material which is dominated by a single node (i.e., all the daughters, granddaughters, great-granddaughters, etc. of that node). The linear order of constituents is shown by the left-to-right order of the corresponding nodes, and the linear order of words in the sentence is shown by the left-to-right order of the TERMINAL NODES (i.e., the lowest nodes in the tree).

To give a concrete example, the constituent structure for a simple preposition phrase (PP) in English (omitting the terminal nodes) is shown in (22a). This diagram shows that the preposition P must precede its object NP; and that within NP, the determiner Det will precede the head noun N. This tree represents the structure of phrases such as *on the beach*, *under the table*, etc. Figure (22b) shows the structure of the Malay sentence in (9b), which is repeated below. Note that the tree in (22b) assumes a "flat" clause structure, lacking any VP node, because we have not presented any evidence for the existence of a VP constituent in Malay.

(22) a



b



(9) b [Orang tua itu] makan [nasi goreng].
person old that eat rice fry
'That old person eats fried rice.'

In order to specify which tree structures are grammatical in a particular language, the grammar of the language must contain (i) a set of PHRASE STRUCTURE RULES, and (ii) a LEXICON. A phrase structure (PS) rule is a simple, context-free rule (i.e., no conditioning environment is stated) which defines a possible combination of mother and daughter nodes. The simple PS rule in (23) says that the tree fragment in (20) is a valid combination, i.e., a grammatical structure.

(23) A → B C

In order for a larger tree to be considered grammatical, each node in the tree must be LICENSED in this way by some phrase structure rule in the grammar of the language. To license (or "generate") the PP structure in (22a), the grammar of English must contain the rules in (24a). To generate the tree structure in (22b), the grammar of Malay must contain rules like those in (24b); the parentheses in this rule are used to mark optional elements.[6]

(24) a **English**
PP → P NP
NP → Det N

b **Malay**
S → NP V NP
NP → N (Adj) (Det)

Phrase sructure rules of this kind determine possible tree structures, but the actual words of a sentence (terminal nodes of the tree) are provided by the lexicon. The lexicon can be thought of as the speaker's "mental dictionary." It is more than just a list of all the words in the language. Each word must have a LEXICAL ENTRY which contains information about the meaning, pronunciation, and grammatical features of that particular word. The grammatical information contained in the lexical entry will determine the contexts in which this word may occur. An important part of this information is the word's syntactic category (part of speech). While general phrase structure rules like those in (24) license possible combinations of mother and daughter nodes in the tree, a special rule of LEXICAL INSERTION licenses a node whose label is a word-level category (e.g., N, A, V, etc.) to have as its only daughter a word belonging to that same category.

In the next chapter we will discuss linguistic criteria for identifying constituents and categories in a particular language; but first we need to review a few basic facts about functional structure.

1.2.3 Functional structure

GRAMMATICAL RELATIONS like subject and object are an important part of the grammatical structure of a sentence, in particular of that aspect which we referred to above as FUNCTIONAL STRUCTURE. The terms "subject" and "object" are very familiar, of course, but it may be helpful to clarify what they actually mean. Some traditional grammars define the subject as the "doer of the action", and the object as the "person or thing acted upon by the doer." This definition seems to work for sentences like (25a–b), but is clearly wrong in examples like (25c–d).

(25) a Mary slapped John.
b A dog bit John.
c John was bitten by a dog.
d John underwent major heart surgery.

John is "acted upon" in all four of these sentences; but the NP *John* bears the object relation in (25a–b) and the subject relation in (25c–d). Phrases like "the doer of the action" or "the person or thing acted upon" identify particular semantic roles, namely agent and patient. But, as these examples illustrate, the subject is not always an agent, and the patient is not always an object.

Another traditional definition of the subject is "what the sentence is about." Again, this definition seems to work for many sentences, but fails in others. All three of the sentences in (26) seem to be "about" Bill (i.e., *Bill* is the topic); but *Bill* is the subject in (26a), the object in (26b), and neither subject nor object in (26c). These examples show that we cannot reliably identify the subject of a sentence with either the agent or the topic. Rather, we must define subjects

and other grammatical relations on the basis of their syntactic properties. We will return to this issue frequently in later chapters; see especially chapters 10 and 11.

(26) a Bill is a very crafty fellow.
b (Jack is pretty reliable, but) Bill I don't trust.
c As for Bill, I wouldn't take his promises very seriously.

Subjects and objects are often referred to as TERMS, or DIRECT ARGUMENTS. Arguments which are not subjects or objects are called INDIRECT or OBLIQUE ARGUMENTS. These labels reflect the idea that the relationship between a verb and its subject or object is "closer," or has greater syntactic significance, than the relationship with other elements in the clause. One indicator of the special status of subjects and objects in English is that subjects and objects are expressed by bare noun phrases, whereas oblique arguments (as well as many adjuncts) are marked with prepositions. Some examples of oblique argument phrases are presented in the following sentences:

(27) a Michael Jackson donated his sunglasses to the National Museum. (RECIPIENT)
b Samson killed the Philistines with a jawbone. (INSTRUMENT)
c The Raja constructed a beautiful palace for his wife. (BENEFICIARY)
d The Prime Minister deposited his money in a Swiss bank. (LOCATION)

We will use the abbreviations SUBJ, OBJ, and OBL to refer to subjects, objects and oblique arguments, respectively. Some sentences, like those in (28), seem to have more than one object NP. In this case we will refer to the object which is closest to the verb (marked with a single underline in these examples) as the DIRECT or PRIMARY OBJECT (abbreviated OBJ). We will refer to the object which is farther from the verb (marked here with a double underline) as the SECONDARY OBJECT, using the abbreviation OBJ_2.

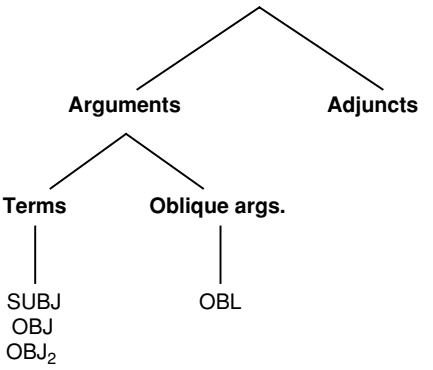
(28) a Mary gave her son a new bicycle.
b Reluctantly, Henry showed Susan his manuscript.
c Uncle George told the children a story.

Notice that the secondary object relation does not correspond to the traditional notion "indirect object."[7] In traditional grammar, *Mary* would be called the "indirect object" of both sentences (29a) and (29b). But in this usage the term "indirect object" actually refers to a semantic role (recipient, goal, or beneficiary) rather than to a grammatical relation. The grammatical relation of *Mary* is different in these two sentences: *Mary* is the primary object in (29a) but an oblique argument in (29b).

(29) a John gave Mary his old radio.
b John gave his old radio to Mary.

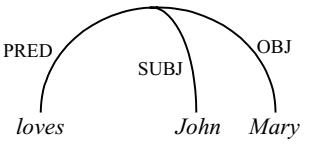
To summarize, then, we can classify the elements of a simple clause (aside from the predicate itself) as either arguments or adjuncts. In order to be expressed grammatically, arguments must be assigned a grammatical relation within the clause. We have identified two basic classes of argument relations, obliques vs. terms (or direct arguments). Terms (i.e., SUBJ, OBJ, OBJ_2) play an active role in a wide variety of syntactic constructions, while oblique arguments are relatively "inert." This classification of clausal elements is summarized in the following diagram:

(30)



It is sometimes helpful to ignore (temporarily) details of word order and constituent structure within a sentence, and to focus exclusively on the assignment of grammatical relations. On such occasions we will use a simple network diagram to represent this information.[8] For example, the relational structure of the simple sentence *John loves Mary* could be represented as follows:

(31)



Each arc of the diagram represents a major element of the clause. The PRED(icate), i.e., the verb in this case, is the head of the clause; the other arcs represent either arguments or (in more complex examples) adjuncts. Note that the order of elements here is arbitrary; the diagram is an abstract representation of grammatical relations, and carries no information about word order and constituency.

1.2.4 Correspondences

We have introduced three different representations, each corresponding to a different aspect of syntactic structure: argument structure, constituent structure, and functional structure. Each representation provides a partial description of the structure of a sentence. In one sense, these three representations

are independent of each other. None of the three is derived from the others. Rather, all three descriptions are true simultaneously, each containing a different type of information.

On the other hand, it is important for us to be able to specify the mappings, or correspondences, between elements of these three structures. For example, we need to be able to show who did what to whom; that is, which NP in the constituent structure corresponds to which semantic role in argument structure. Grammatical relations play a crucial role in defining these correspondences.

The mapping between semantic roles and grammatical relations is often referred to as LINKING. One way of representing this mapping is shown in (32). As discussed in chapter 3, section 3.2.1, a number of different theories have been proposed concerning how these correspondences are determined. But for now we will simply assume that all of this information is contained in the lexical entry of the verb.

(32) a *sing* <agent>
|
SUBJ

b *hit* <agent, patient>
| |
SUBJ OBJ

c *put* <agent, theme, goal>
| | |
SUBJ OBJ OBL

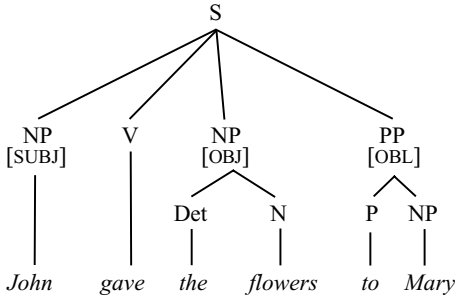
We will used the term SUBCATEGORIZATION to refer to the set of grammatical relations which are specified in a verb's lexical entry. As the name implies, we can use this information to divide the lexical category Verb into a number of subcategories: verbs which take only a subject, verbs which take a subject and an object, etc. The subcategorization of a verb plays a major role in determining what syntactic environments that verb may occur in. To see why, let us consider the mapping between constituent structure and functional structure.

We can use an annotated tree structure like that in (33a) to show the grammatical relation which is assigned to each phrasal constituent in the tree. (Most syntacticians posit a VP constituent for English which includes the verb, its objects, and oblique arguments but not the subject. The evidence supporting this hypothesis will be discussed in chapter 2, but for the purposes of this chapter we will temporarily assume the "flat" structure shown in [33a].)

How is the assignment of grammatical relations determined? In many languages, this mapping is largely determined on the basis of case and/or agreement morphology. But in English, which has very little of that kind of morphology but relatively fixed word order, the mapping is determined by position: each grammatical relation can be associated with a particular position in the phrase structure.

These associations can be stated using annotated phrase structure rules like the one in (33b).

(33) a



b S → NP V (NP) (PP)
[SUBJ] [OBJ] [OBL]

The rule in (33b) will correctly generate the tree in (33a). Unfortunately, it would also allow the ungrammatical sentences in (34) to be generated:

(34) a *John likes.
b ??Mary gives the young boy.
c *The girl yawns Mary.

These sentences are bad because the number of NPs in the clause does not match the number of arguments which the verb requires. More precisely, the subcategorization requirements of the verbs are not satisfied: the set of grammatical relations which the verb must assign does not match the number of NPs available to bear those relations. *Like* is a transitive verb which requires an object; *yawn* is an intransitive verb which does not take an object; and *give* requires three arguments, while only two NPs are present in (34b). Obviously the grammar of the language must include some mechanism for rejecting such sentences as being ungrammatical.

1.2.5 Well-formedness conditions

The mismatch between the structure of the clause and the subcategorization requirements of the verb becomes obvious when we compare the functional structure of the clause with the verb's argument structure. The functional structure of (34b) is shown in (35a). As this diagram indicates, the clause is INCOMPLETE: the verb *give* requires an oblique argument (the recipient) but the clause does not contain any obliques. The problem with (34c) is just the opposite: the clause contains too many arguments, rather than too few. Specifically, as shown in (35b), the clause contains an OBJ; but no OBJ appears in the subcategorization of the verb *yawn*. In such a case we say that the clause is INCOHERENT.