

Syntactic Theory: A Formal Introduction

Second Edition

Ivan A. Sag Thomas Wasow
Emily M. Bender

August 10, 2006

CENTER FOR THE STUDY
OF LANGUAGE
AND INFORMATION

Copyright © 2003
CSLI Publications
Center for the Study of Language and Information
Leland Stanford Junior University
Printed in the United States
10 09 08 07 06 2 3 4 5 6

Library of Congress Cataloging-in-Publication Data

Sag, Ivan A., 1949–
Syntactic theory / by Ivan A. Sag, Thomas Wasow, and Emily M. Bender.— 2nd ed.

p. cm. — (CSLI lecture notes ; no. 152)

Includes bibliographical references and index.

ISBN-13 978-1-57586-399-3 (alk. paper)

ISBN-10 1-57586-399-5 (alk. paper)

ISBN-13 978-1-57586-400-6 (pbk. : alk. paper)

ISBN-10 1-57586-400-2 (pbk. : alk. paper)

1. Grammar, Comparative and general—Syntax.

I. Wasow, Thomas. II. Bender, Emily M., 1973– III. Title. IV. Series.

P291.S25 2003

415–dc21 2003010792

CIP

∞ The acid-free paper used in this book meets the minimum requirements of the American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

CSLI was founded in 1983 by researchers from Stanford University, SRI International, and Xerox PARC to further the research and development of integrated theories of language, information, and computation. CSLI headquarters and CSLI Publications are located on the campus of Stanford University.

CSLI Publications reports new developments in the study of language, information, and computation. Please visit our web site at

<http://csli-publications.stanford.edu/>

for comments on this and other titles, as well as for changes and corrections by the authors and publisher.

Glossary

This glossary contains linguistic terms that either play a direct role in or are presupposed by this book. For further information, there are a number of dictionaries of linguistics, including Crystal 1985 and Trask 1993. (In the following definitions, ‘q.v.’ stands for Latin *quod vide* ‘which see’.)

AAVE In informal speech, many African Americans use a distinctive variety of English known as African American Vernacular English, or AAVE for short. Also known as Black English, African American English, and Ebonics, both the phonology and aspects of the syntax of AAVE have been extensively studied.

absolute *See* case.

accusative *See* case.

active A verb form or clause that is not in the passive is referred to as active. *See also* passive; voice.

affix An affix is a morphological element added to a stem to form another stem or a word. Two common types of affix are prefixes (e.g. *re-*, as in *reread*; *out-*, as in *outperform*) and suffixes (*-ed*, as in *visited*; *-s*, as in *visits*). Many languages exhibit other types of affix as well, including infixes (an element inserted into a stem) and circumfixes (e.g. a pair of elements wrapped around a stem).

agreement In many languages, the forms of certain elements can vary to indicate such properties as person [q.v.], number [q.v.], gender [q.v.], etc. Often, these variations are marked with affixes. Some grammatical relationships between pairs of linguistic elements require that they agree on these properties. In English, for example, present tense verbs are marked to indicate whether their subjects are third-person singular (with the suffix *-s*), and nouns indicate plurality (also with a suffix *-s*). The systematic covariation of the forms of the subject and verb is called ‘subject-verb agreement’. Similarly, pronouns must agree with their antecedents in person, number, and (if third-person) gender. *See also* inflection.

anaphor *See* anaphora.

anaphora Certain expressions depend for their interpretation on their association with some other element in the discourse (usually earlier). Paradigm examples are pronouns like *he*, *her*, and *itself*; other examples include *do so* and verb phrase ellipsis. ‘Anaphora’ is the term for the relationship between such elements and their antecedents. The term ‘anaphor’ is sometimes used for all anaphoric elements and is

sometimes reserved for only certain kinds (primarily reflexives [q.v.] and reciprocals [q.v.]). *See also* antecedent; binding; ellipsis.

antecedent This term is used for a linguistic element that is the basis for interpreting some anaphoric element occurring (typically later) in the sentence or discourse. In particular, pronouns are often described as referring back to their antecedents (or to the referents of their antecedents). *See also* anaphora.

argument (or syntactic argument) This is a general term for any phrase that is selected for by a lexical head, such as a complement or specifier. This usage derives from the semantic term ‘argument’, which refers to a component of a proposition that must occur with a given predicate or relation. For example, the meaning of the verb *wash* has two semantic arguments (the washer and the washed) that correspond to the two syntactic arguments of the verb *wash* (the subject and the object) – as in *Alex washed the car*. The simplest examples of (syntactic) arguments are noun phrases, but prepositional phrases and subordinate clauses can also function as arguments. *See also* complement; specifier; subcategorization; valence; argument structure.

argument-marking preposition English prepositions serve two distinct functions. In some cases, a preposition is used to indicate the role its object NP plays within the semantic structure of some predicate. In other cases, the preposition itself functions as a predicate, and its object is one of its arguments. In this text, the first kind of use is called an ‘argument-marking preposition’. An example is the preposition *on* in *They rely on us*. We call the second kind of preposition ‘predicational’, illustrated by the use of *on* in *They are on the porch*.

argument structure In the theory developed in this text, the phrases that serve as arguments of a given head are listed in the value of a feature called ARGUMENT-STRUCTURE (ARG-ST). This term is also sometimes used in a less technical sense to designate the semantic relations between a head [q.v.] and its arguments. *See also* argument.

aspect Many languages have special grammatical elements for locating in time the situations referred to. Among the temporal notions often expressed are whether situations are in process or completed and whether they occur repeatedly. These notions are often called ‘aspect’, and words or affixes whose function is to express aspect are called ‘aspectual markers’. *See also* perfective, progressive.

aspectual marker *See* aspect.

auxiliary This term refers to elements found in many languages that share the following semantic and syntactic characteristics: (i) they express such notions as time (past, present, future; continuation, completion), necessity, possibility, obligation, permission, negation, or questioning; and (ii) they occur in fixed positions in sentences, usually at or near the beginning or end. English auxiliaries are a special kind of verb. It is the auxiliary verb that is inverted with the subject in yes/no questions (e.g. *Did she fall?*) and that carries the negative suffix in contractions [q.v.] (e.g. *can't*, *won't*).

base form Almost all English verbs have an uninflected form that occurs after *to* in infinitives [q.v.] and after modals [q.v.], which we refer to as the ‘base form’.

- binding** Pronouns are often said to be ‘bound’ by their antecedents [q.v.], and the term ‘binding’ is used to refer to the relationship between pronouns and their antecedents. The study of the principles governing pronominal anaphora [q.v.] is called ‘binding theory’. *See also* reciprocal; reflexive.
- case** Certain words – particularly nouns and pronouns – can appear in different forms depending on their relationship to other elements in the sentence in which they appear. In English, for example, personal pronouns exhibit nominative case (e.g. *I, they*) or accusative case (e.g. *me, them*), depending on whether they are subjects or objects. In many languages, case is the primary way of indicating the roles of the noun phrases in the clause – that is, who did what to whom. Among the names of cases commonly used across languages are ‘nominative’, ‘accusative’, ‘dative’, ‘genitive’, ‘ergative’, and ‘absolute’. *See also* inflection.
- clause** A clause is a phrase that includes a predicate and all of its arguments and modifiers. The term is sometimes limited to phrases headed by a verb.
- common noun** Nouns are often divided into two kinds: proper and common. Proper nouns are names, denoting individual things by themselves, and do not normally take determiners, complements, or modifiers. Common nouns stand for kinds of things and take determiners [q.v.], modifiers [q.v.], and (sometimes) complements [q.v.]. In English orthography, proper nouns are conventionally capitalized, but common nouns are not.
- competence** In a number of works, Chomsky has distinguished between the (largely unconscious) knowledge of language that people have and how they put that knowledge to use in speech and writing. The former is called ‘competence’; the latter ‘performance’. The study of linguistic competence abstracts away from such factors as memory limitations, disfluencies, and speech errors. Work in generative grammar [q.v.] has concentrated largely on developing models of competence, though there has been much discussion of how such models relate to what is known about performance.
- complement** The lexical head [q.v.] of a phrase characteristically selects which arguments co-occur with it, and some of these are referred to as ‘complements’. When the phrase’s head is a verb, the complements include what are traditionally called direct and indirect objects, as well as some prepositional phrases and subordinate clauses [q.v.]. Subjects – and determiners of NPs – are arguments that are not complements, but specifiers. Complements occur as sisters to the lexical head in syntactic structure and, in English, follow the head. For example, a verb such as *hit* takes one complement, namely, an NP (e.g. *hit the ball*); *rely* takes a PP complement (e.g. *rely on Sandy*). A preposition such as *in* also takes a single NP complement (e.g. *in the box*). Some nouns can also take complements, such as *picture* which takes an optional PP complement (e.g. *picture of Kim*). *See also* argument; specifier.
- complementizer** The term ‘complementizer’ is close in meaning to the traditional term ‘subordinate conjunction’. It is normally reserved for elements introducing clausal complements headed by a verb. In English, the use of *that* to introduce subordinate

clauses [q.v.] (as in *It is remarkable that ice floats*) is the clearest example of a complementizer. *See also* complement.

conjunction (conjunct) Traditional grammarians use the term ‘conjunction’ to refer to words that connect two linguistic units in some way. In this text, we use it only for what are traditionally called ‘coordinate conjunctions’, that is, words connecting elements that are, intuitively, of equal status. In English, the paradigmatic coordinate conjunctions are *and* and *or*, though *but* and *nor* can also function in this way. The individual words or phrases that have been conjoined can be referred to as conjuncts. *See also* coordination.

constituent The term ‘constituent’ is used by linguists as a near synonym for ‘phrase’, meaning a part of a sentence that functions syntactically as a single unit. The difference is that ‘constituent’ is usually limited to phrases that are proper parts of larger expressions.

construct We refer to the MOTHER value of a construction instantiation [q.v] as a ‘construct’.

construction This term has a traditional informal use, designating any sequence of words or types of words that pattern alike in some way. Thus, grammarians may talk of an ‘imperative construction’, a ‘passive construction’, or a ‘filler-gap construction’. In Chapter 16 of this book, we introduce a more precise meaning: constructions are a type of feature structure containing the features MOTHER and DAUGHTERS. The grammar rules and lexical rules of the theory developed in Chapters 1-15 are seen as two kinds of constructions in this sense.

construction instantiation Constructions [q.v] are descriptions, specifying the values of some features and leaving others unspecified. A construction instantiation is a fully resolved feature structure that is consistent with the specifications of some construction.

context-free grammar (CFG) A context-free grammar is a particular type of formal system that has proved very useful in the precise characterization of computer languages and also serves as the starting point for much work in syntactic theory. CFGs consist of an initial symbol [q.v.], a finite lexicon with words classified into grammatical categories [q.v.], and a finite collection of rules of the form $A \rightarrow \omega$, where A is a single symbol (representing a type of phrase), and ω is a finite string of lexical and/or phrasal categories.

contraction Reduced forms of words are sometimes combined with other words (that would typically occur adjacent to the reduced words) to form a new word; these are referred to as ‘contractions’. English examples include combinations of a finite auxiliary [q.v.] verb with a reduced form of *not* to produce such words as *isn’t* and *can’t*, as well as simple contraction of finite auxiliaries, e.g. *They’re arriving tomorrow* and *Kim’s here*.

control Some complements have no overt specifier, but are interpreted as if they had subjects with the same reference as (i.e. coindexed with) another complement to the same predicate. For example, in both *Pat tries to be on time* and *We urged Pat to be on time* the individual Pat is understood as the person who is meant

to be on time. This relationship (between two noun phrases, the second typically an unexpressed subject) is referred to as ‘control’; in this case with the NP *Pat* being the ‘controller’ of the unexpressed subject of the infinitival phrase. Control predicates are not to be confused with raising [q.v.] predicates (like *continue* and *expect*), one of whose arguments actually is simultaneously an argument of another complement. A number of syntactic and semantic diagnostics distinguish these two types of predicate. *See also* raising.

coordination This term refers to the linking of two words or phrases of equivalent syntactic status (i.e. neither coordinate (or conjoined) element is subordinate to the other). An example of a coordinated clause is *Kim cooked the food and Sandy washed the dishes*. *See also* conjunction.

copula This term is used by traditional grammarians to refer to verbs with little semantic content, which serve to link a subject with a predicate. In English, the copula is *be*.

coreference/coreferential Two referring expressions that refer to the same entity are called ‘coreferential’, and the relationship between them is called ‘coreference’. *See also* anaphora.

count noun Common nouns are divided into two subclasses known as ‘count’ and ‘mass’ according to which determiners they can combine with. Count nouns co-occur with *a(n)*, *several*, *few*, etc.; mass nouns co-occur with *much* and can occur in the singular with no determiner. This distinction is correlated with a semantic distinction: mass nouns usually are used to refer to substances and count nouns to (sets of) entities. A portion of a substance (e.g. helium) is still the same substance, whereas a portion of an entity (e.g. a bicycle) is not usually an entity of the same kind. This correlation is not perfect, however, as evidenced by the mass noun *furniture* and by minimal pairs like *cabbage* (which can be either count or mass) vs. *lettuce* (which, for many speakers, must be mass).

dative Many languages employ a case called dative to grammatically encode the participation of some argument (q.v.) in a given situation as recipient, goal, or beneficiary. *See case*.

declarative/interrogative/imperative These are terms used in the classification of sentence types. Declarative sentences are used to make a statement (or – equivalently for our purposes – to assert the truth of a proposition [q.v.]), as in *The mayor is reading a book*. Interrogative sentences are used to ask questions [q.v.], as in *What are they doing?* Imperative sentences are used to give orders (or to issue ‘directives’ [q.v.]), as in *Read a book!*

defeasible A constraint is said to be ‘defeasible’ if it can be overridden – that is, if it allows for the existence of exceptions. *See also* inviolable.

demonstrative Expressions used for referring through direct indication (often accompanied by pointing) are called ‘demonstratives’. The best examples in English are *this*, *that*, *these*, and *those*.

description We model elements of language using feature structures, which are either atoms or else functions (in the mathematical sense). Such functions map features

into feature structures (atoms or functions). They are moreover total functions, in the sense that they map every feature in a relevant domain into a value. Often, rather than specifying a full feature structure, it is convenient to describe a class of feature structures by specifying only the values of some features. Many of the constructs of our grammar – notably lexical types, lexical entries, and lexical rules – are descriptions in this sense.

descriptive grammar *See* prescriptive grammar.

determiner The sorts of specifiers [q.v.] that nouns take are called ‘determiners’. These include articles (*a, the, etc.*), quantifiers [q.v.] (*some, every, many, two, etc.*), and possessives [q.v.] (*my, Kim’s, etc.*). *See also* specifier.

directive A directive is a particular kind of semantic object, characteristically associated with imperative [q.v.] sentences. It is the kind of object that can be issued by uttering such a sentence, and fulfilled by causing the conditions associated with the sentence to be met. The grammar in this text puts [MODE dir] into the semantics of imperative sentences.

discourse This term refers to units of language longer than a sentence – for example, dialogues or paragraphs.

distribution Linguists use this term to refer to the set of total environments – or contexts – in which some linguistic unit can occur.

ditransitive verb In this book, verbs that take two NP objects are called ‘ditransitive’. The standard example is *give*, in examples like *The teacher gave the students an exam*. *See also* intransitive verb; transitive verb; valence.

dummy Words that evidently have no meaning and serve only to fill some grammatical function are sometimes called ‘dummies’. The paradigm examples in English are the *there* that occurs in existential sentences (e.g. *There is a seat available* and the *it* of extraposition [q.v.] (e.g. *It is fortunate that you have a seat*). Other terms used for these are ‘expletives’ and ‘pleonastic’ elements.

ellipsis Ellipsis means ‘leaving out’ or ‘omitting’: in certain contexts, parts of a sentence can be omitted if their interpretation is reconstructable. An example is the following case of verb phrase ellipsis, where the bracketed material may be left out:

- (i) Pat won’t taste the soup, but Chris will [taste the soup].

See also anaphora.

ergative *See* case.

existential *be*/existential *there* English has a special construction for expressing existence, involving the dummy *there* as subject and forms of the verb *be*. These are called ‘existential’. *See also* dummy.

expletive *See* dummy.

extraction Some grammatical theories deal with long-distance dependencies [q.v.] by means of rules that move the filler from the gap position to where it actually appears. Since the position of the filler is always less deeply embedded in the tree than the position of the gap, this is sometimes referred to as ‘extraction’ of the filler. This terminology is carried over into the present text in the Subject Extraction Lexical Rule.

extraposition Predicates that can take complementizer [q.v.] phrases (i.e. *that*-clauses) as subjects can also occur with a dummy *it* as subject and the CP as the last complement. The latter construction is called ‘extraposition’, and is exemplified by the following:

- (i) It bothers Alex that Dana left.

The term is also sometimes used for expressions in which a complement or modifier is separated from its head by intervening material, as in *A review appeared of Lee’s latest book*. See also *dummy*.

feature structure A standard way of representing linguistic information is in terms of complexes of features and values. A feature can be thought of as a dimension along which different linguistic entities (such as words, phrases, or sentences) may differ, and values identify locations on those dimensions. A feature-value pair models a property of a linguistic entity that distinguishes it in a linguistically interesting way from some other entities. For example, the feature PERSON (PER) in English has three possible values, namely ‘1st’, ‘2nd’, and ‘3rd’. It is a property of the word *you* that it is second-person, and we represent that with the feature-value pair [PER 2nd]. A feature structure can thus be treated as set of feature-value pairs in which no feature is paired with more than one value. (Feature structures are functions and **types** are used to organize the domains and ranges of such functions.) Values of features in our theory may themselves be feature structures, or even lists of feature structures. Feature structure descriptions are standardly given in terms of matrices, listing feature names paired with their values, also known as ‘feature specifications’. See also *inheritance hierarchy*; *type*.

filler See *long-distance dependency*.

finite-state grammar Finite-state grammars are a type of formal system sometimes used to describe certain rather simple artificial languages. They are mathematically equivalent to regular expressions. See also *context-free grammar*; *regular expression*.

finite verb A finite verb is one that is marked for tense [q.v.] (present or past, in English).

gap See *long-distance dependency*.

gender The nouns in many languages divide into classes, differing in their patterns of inflection and agreement. In a number of languages (e.g. French and German), these noun classes are referred to as ‘genders’, because nouns used to refer to males or females (of any species) are generally (though not invariably) grammatically masculine or feminine, respectively. In English, gender is marked grammatically only on third-person singular pronouns (*he*, *she*, and *it*) and is virtually always predictable from the sex of the referent.

generative grammar Chomsky introduced this term based on the idea that a grammar is a formal system for generating the sentences of a language. The term is now used in at least three distinct senses, to denote: (i) work in the Chomskyan tradition (fairly broadly conceived); (ii) an explicit system of rules, principles, and/or constraints that characterizes all and only the well-formed sentences of a language; or (iii) the system in the mind or brain of a speaker that makes language use possible.

genitive *See case.*

grammatical category Words and phrases can be classified in various ways, any of which can be called a ‘grammatical category’. The term is usually used to refer to parts of speech [q.v.], such as noun, verb, etc., as well as types of phrases, such as noun phrase, verb phrase, and sentence.

head The constituent [q.v.] of a phrase that is grammatically the most important constituent of that phrase is called the ‘head’ of the phrase. The head usually determines the category of the phrase, as well as many of its other properties. Thus noun phrases have nouns as heads, verb phrases have verbs as heads, etc. The term is used ambiguously to refer to the word that functions as head of the phrase and any subphrase containing that word. For example, in *the destruction of the city*, both *destruction* and *destruction of the city* can be called heads of the phrase.

idiom Some combinations of words have interpretations that are not fully predictable from the meanings that those same words have in other contexts. These are known as ‘idioms’. English examples include *take advantage* to mean (roughly) ‘exploit’, *keep tabs on* for ‘monitor’, and *kick the bucket* for ‘die’. Parts of an idiom are sometimes called ‘idiom chunks’, e.g. *advantage* in *take advantage*. Idiom chunks play a central role in one of the diagnostics for distinguishing raising [q.v.] predicates from control [q.v.] predicates.

imperative *See declarative.*

infinitive Certain kinds of nonfinite verbs are referred to as ‘infinitives’. English infinitives are preceded by *to*, which we analyze as a verb with the feature [INF +], taking a [FORM base] verb phrase as its complement. In many other languages, the infinitive verb form is marked with special affixes.

inflection Languages often add affixes to words to mark the syntactic function or relationships of the word in the sentence. For example, present tense verbs in English are usually inflected with the suffix *-s* when the subject is third-person singular, and past tense verbs are inflected with *-ed*. The term may also be used to refer to the affix itself. Among the common uses of inflectional affixes are to indicate tense [q.v.], agreement [q.v.], number [q.v.] (singular or plural), and case [q.v.]. The theory in this book employs lexical rules [q.v.] to account for the relationships among different forms of a lexeme. Note, however, that we restrict the term ‘inflectional lexical rule’ to those that map lexemes to words; other instances of what some might call ‘inflection’ (e.g. the participial forms of verbs) are handled by means of other types of lexical rule.

inheritance hierarchy The elements of some domains of study can naturally be organized into classes, based on shared properties. Some classes can be further subdivided into subclasses, with additional shared properties. The organization of such domains can be thought of as a hierarchy, with the most inclusive class (encompassing the entire domain) at the top, and the most restricted classes at the bottom. In between are various classes of interest. The properties associated with particular classes are inherited by their subclasses, and ultimately by their individual members. Domains organized in this way are referred to as ‘inheritance hierarchies’. In

linguistics, inheritance hierarchies have been used to organize lexical information, among other things. *See also* type.

initial symbol Grammars characterize languages. But languages can be conceived of in a variety of ways: as consisting of sentences, of phrases, of any expressions that can serve as stand-alone utterances, etc. A formal theory of grammar must include a specification of which of the expressions it characterizes are to be regarded as those that constitute the language. The initial symbols of a formal theory state what is to count as an element of the language. In this book, the initial symbol definition specifies conditions that phrases must satisfy if they can stand alone, i.e. be used in isolation to communicate a message.

interrogative *See* declarative.

intonation This term is used to refer to the patterns of pitch in speech.

intransitive verb A verb that does not take any NP objects is referred to as ‘intransitive’. A standard example in English is *die*. *See also* ditransitive verb; transitive verb; valence.

inversion Grammarians use this term fairly generally to refer to any construction in which two elements appear with their typical ordering reversed. In this text, it is used in particular for sentences (mostly questions) in which a finite auxiliary [q.v.] verb precedes the subject, as in *Are you sleeping?*

inviolable A constraint is said to be ‘inviolable’ if the grammar never allows it to be violated. Constraints that are not inviolable are said to be ‘defeasible’ [q.v.].

island constraint While long-distance dependencies can, in principle, stretch over arbitrary distances, there are some pairings of filler and gap positions that are not possible. For example, a gap inside a CP subject cannot, in general, be paired with a filler outside that CP, as in **Which candidate did [that I voted for ___] bother you*. The constraints on possible filler-gap pairings are known as ‘island constraints’. *See also* long-distance dependency.

Kleene star It is useful in the formal representation of languages (both natural and artificial) to allow certain patterns to be repeated any finite number of times (including zero). The standard notation for this is a superscripted asterisk, known as the ‘Kleene star’ (after the mathematician Stephen Kleene). For example, ab^*c is shorthand for the infinite set of strings: *ac, abc, abbc, abbbc,* ‘Kleene plus’, denoted by a superscripted plus sign, means any nonzero number of repetitions. *See also* regular expression.

lexeme The term ‘word’ is used ambiguously to mean either a particular form, such as *sees*, or a set of related forms such as *see, sees, saw, seen, and seeing*. To avoid this ambiguity, linguists sometimes posit an abstract entity called a ‘lexeme’ that gives rise to a family of related words. *See also* word.

lexical entry Information about individual words [q.v.] that must be stipulated is put into the lexicon [q.v.] in the form of descriptions that we call lexical entries. They are ordered pairs, consisting of a phonological form (description) and a partial feature structure description. Fully resolved lexical sequences [q.v.] consistent with lexical entries can serve as the INPUT values of lexical rules [q.v.].

lexical rule Lexical rules are one of the mechanisms (along with the type hierarchy [q.v.]) used to capture generalizations within the lexicon. Families of related words – such as the different inflectional forms of a verb – can be derived from a single lexical entry [q.v.] by means of lexical rules. We formalize lexical rules as a type of feature structure with features INPUT and OUTPUT. There are three subtypes of lexical rules: derivational (relating lexemes [q.v.] to lexemes), inflectional (relating lexemes to words [q.v.]), and post-inflectional (relating words to words).

lexical rule instantiation Our lexical rules [q.v.] are descriptions, specifying the values of some features and leaving others unspecified. A lexical rule instantiation is a fully resolved feature structure that is consistent with the specifications of some lexical rule.

lexical sequence Ordered pairs that can serve as the INPUT and OUTPUT values of lexical rules [q.v.] are called lexical sequences. They consist of a phonological form and a fully resolved feature structure.

lexicalism Lexicalism often refers to the doctrine that (1) the internal structure of words is independent of how words are put together to make sentences, and (2) words are the atoms of syntactic combination. For example, in a lexicalist theory, bound morphemes (inflectional affixes that must be attached to a word) are not treated as independent syntactic elements, as they are in most (nonlexicalist) versions of Transformational Grammar (see Appendix B). Theories of grammar also differ in their organization and in where they locate syntactic information. Some theories (e.g. Generalized Phrase Structure Grammar) have rich systems of rules and relatively impoverished lexical entries. Others (e.g. Categorical Grammar or Lexical Functional Grammar) have highly structured lexical entries and a small number of very general rule schemata. ‘Lexicalism’ is sometimes also used to distinguish the latter sort of theory.

lexicon The list of all words [q.v.] (or lexemes [q.v.]) of a language is called its ‘lexicon’.

The lexicon is the repository of all idiosyncratic information about particular words, including syntactic, semantic, and phonological information. In some theories of grammar, the lexicon can also contain a great deal more systematic information, organized by a type hierarchy [q.v.] and/or lexical rules.

long-distance dependency Certain constructions, including *wh*-questions, topicalization, and relative clauses, permit an element in one position to fill the grammatical role associated with another position. The two positions can be arbitrarily far apart. For example, in *Which student did the principal say that the teacher thought was responsible?* the NP *which student* functions as the subject of *was responsible*, although they are separated by most of the sentence. Such constructions are called ‘long-distance dependencies’ (LDDs). Elements like *which student* in the above example are called ‘fillers’, and the position normally associated with the filler’s role (in this case, immediately preceding *was responsible*) is called the ‘gap’. *See also* island constraints.

main clause *See* root sentence.

modal The English verbs *can*, *could*, *may*, *might*, *must*, *shall*, *should*, *will*, and *would*, along with their negated forms (*can’t*, etc.) are referred to as ‘modals’ or ‘modal

verbs'. They share the following properties: they function only as finite verbs [q.v.]; they exhibit auxiliary behavior (negation, inversion, contraction, and ellipsis); they take base form VP complements; and they show no agreement [q.v.] (i.e. no third-person singular *-s* suffix). Some other languages have similar syntactically distinctive classes of words expressing necessity, possibility, obligation, and permission; these are also known as modals. *See also* auxiliary.

mass noun *See* count noun.

model Understanding real world phenomena may be enhanced by investigating mathematical entities that share certain properties with the phenomena in question. Such mathematical entities are called 'models' of the real-world phenomena. Models are useful because they make it possible to abstract away from incidental properties and focus on those that are of theoretical interest. With respect to the grammar we develop in this text, we use the word 'model' in two different senses. On the one hand, the grammar as a whole is a model of (a fragment of) the English language, or of speakers' knowledge of English. On the other hand, fully resolved feature structures are models of linguistic entities. In this sense, 'model' contrasts with 'description' [q.v.].

modifier Most phrases consist of a head [q.v.], together with that head's arguments [q.v.]. Semantically, the head typically denotes either a situation or an individual, and the arguments denote essential associated entities. In addition, phrases may contain modifiers, which serve to place further restrictions on the situation or individual picked out by the phrase as a whole. Modifiers can take a wide variety of syntactic forms, including adjectives and adjective phrases, adverbs and adverbial phrases, prepositional phrases, and modifying clauses (such as relative clauses). *See also* argument structure.

morphology This term refers ambiguously to the study of word structure – how words are put together out of stems and affixes – or to word structure itself.

negation Languages include devices for reversing or contradicting the meaning or truth conditions of expressions, a semantic effect known as 'negation'. In English, the most common element expressing negation is the word *not*.

nominalization Nominalizations are nouns constructed out of words of other categories, usually through affixation. An example is *destruction*, derived from the verb *destroy* through the affixation of *-tion* (together with some other modifications). The term 'nominalization' is also used to refer to a process of turning verbs and adjectives into nouns.

nominative *See* case.

number Most English nouns take different forms depending on whether they can head NPs that refer to single entities or multiple entities, e.g. *some dog/dogs, some man/men*. Similarly, present tense [q.v.] verbs with third-person subjects have different forms depending on whether the subjects are singular or plural. The term 'number' is used for such distinctions. Some languages also mark number on other types of words, e.g. adjectives may be marked for the number of the noun they modify. There are also languages that make finer number distinctions than just

singular vs. plural, notably languages that have special ‘dual’ forms for expressions referring to sets with exactly two members.

orthography This term refers to written representations of language. For example, the plural of the noun *doe* and the present tense form of the verb *do* that goes with a third-person singular subject share the same orthography (namely, ‘does’), although their pronunciations (and almost everything else about them) are different.

overgenerate A grammar that licenses sentences that are not part of the language the grammar writer is trying to analyze is said to ‘overgenerate’. This term is usually used when a proposal for part of a grammar of a natural language licenses strings that native speakers of that language say are not well-formed sentences of the language. *See also* undergenerate.

paradigm Certain words have multiple inflectional forms. For example, verbs in English typically change their form depending on whether they are past or present tense, and their present-tense forms depend on the person and number of the subject. They also have a variety of nonfinite forms. The full array of inflectional forms of a word is known as its ‘paradigm’. *See also* inflection.

paraphrase Two sentences are said to be paraphrases of one another if they differ in form but convey the same meaning.

parsing This term refers to the process of assigning a tree structure to a sentence. Many computer systems designed to process natural language include components for parsing, and much psycholinguistic research is concerned with discovering what parsing mechanisms humans use in language comprehension.

part of speech This is the traditional term for lexical categories (i.e. categories of words), based on a combination of semantic and distributional criteria. Among the standard parts of speech are noun, verb, adjective, preposition, adverb, and conjunction. *See also* grammatical category.

participle Certain nonfinite verbs – usually ones that share some properties with adjectives – are referred to as ‘participles’. English has three types of participles: present participles, which end in *-ing* and usually follow some form of *be*; past participles, which usually end in *-ed* or *-en* and follow some form of *have*; and passive participles, which look exactly like past participles but indicate the passive voice [q.v.]. The three participles of *eat* are illustrated in the following sentences:

- (i) Termites are eating the house.
- (ii) Termites have eaten the house.
- (iii) The house was eaten by termites.

passive Many languages have a construction in which the grammatical subject of a verb plays the same semantic role that the object plays when the verb in question appears elsewhere (in active [q.v.] forms). The term ‘passive’ is used to refer both to this construction, and to the verb whose arguments’ roles are at issue. In English, the passive form of the verb looks exactly like the past participle and is usually preceded by a form of *be*; a prepositional phrase headed by *by* is also common, and is used for marking what would be the subject if the verb were not passive. An example is *The dog was attacked (by wombats)*. *See also* participle; voice.

perfective Many languages have special verb forms or constructions used to indicate that the event denoted by the verb is completed. These are referred to as ‘perfective’ (or just ‘perfect’) in aspect. The English perfective involves the combination of *have* with a past participle [q.v.], as in *The dog has eaten the cake*. *See also* aspect.

performance *See* competence.

person Many languages distinguish grammatically among expressions referring to the speaker, to the hearer, and to third parties. This is called the expression of ‘person’. Reference to the speaker or a set including the speaker is called ‘first person’; reference to (sets including) the addressee(s) is called ‘second person’; and everything else is called ‘third person’. Person distinctions are clearest with pronouns, since these are the most common forms used to refer to the speaker and hearer. But in some languages nouns also show person marking, and verbs and adjectives may agree with nouns and pronouns in person.

phonetics Phonetics is the study of the acoustic or articulatory properties of speech sounds.

phonology Phonology is the study of the sound systems of languages, i.e. the systematic grammatical patterns in the distribution [q.v.] of speech sounds.

possessive Many languages have grammatical mechanisms for indicating a relation of possession between the referents of two NPs. When one noun or NP is marked as the possessor of another, this marking is referred to as the ‘possessive’. In English, the possessive is marked by *'s* attached at the end of the noun phrase functioning as the ‘possessor’. There is also a set of determiners that express possession (*my, our, your, etc.*). These are called ‘possessive pronouns’.

pragmatics The information conveyed by a linguistic expression in a particular instance of use is typically much more than just its literal (or ‘linguistic’) meaning. The study of how linguistic meaning contributes more generally to communication is called (linguistic) ‘pragmatics’. *See also* semantics.

predicational preposition *See* argument-marking preposition.

prefix *See* affix.

prescriptive grammar Much of traditional work in grammar is concerned with setting norms – that is, dictating that some usages are ‘incorrect’ and hence to be avoided. Modern linguists refer to this as ‘prescriptive grammar’ (or just ‘prescriptivism’). Most scientific work on grammar purports instead to be ‘descriptive’, seeking systematic explanations for the way the language is actually used.

productive A relationship between two linguistic forms is said to be ‘productive’ if it generalizes to novel forms. For example, the use of the suffix *-ing* to mark the present participle form of a verb is productive, since it gets applied to new coinages (as in *faxing*). Productivity is usually thought of as a matter of degree, with exceptionless relationships counting as more productive than those with exceptions.

progressive Special verb forms or constructions used to indicate that the event denoted by the verb is in progress are referred to as ‘progressive’ aspect. The English progressive involves combination of *be* with a present participle [q.v.], as in *The dog is eating the cake*. *See also* aspect.

proper noun *See* common noun.

proposition A proposition is a particular kind of abstract object, specifically, the sort of thing that can be true or false. Propositions are also what one asserts, believes, denies, etc. Declarative sentences characteristically express propositions, a fact represented in this text by putting [MODE prop] in the semantics of such sentences.

quantifier Words or phrases used to restrict the number or amount of some referent are called ‘quantifiers’. In English, these include such expressions as *all*, *each*, *some*, *many*, *few*, *two*, *more than half*, etc.

question A question is a particular kind of semantic object, specifically, the sort of thing that can be asked and answered. Interrogative sentences are characteristically associated with questions, a fact represented in this text by the presence of [MODE ques] in the semantics of such sentences.

raising Some predicates take one more syntactic argument than semantic argument. In these cases, the extra syntactic argument functions as the subject of another complement and must obey any special co-occurrence restrictions imposed by that complement. These predicates are called ‘raising’ predicates. Raising is exemplified by the sentences *Pat continues to be on time* and *We expected Pat to be on time*. In these examples, *Pat*, though a syntactic argument of *seem* and *expect*, is semantically an argument only of *be on time*. A semantically empty dummy [q.v.] is possible with raising predicates, where it would not be possible in the corresponding positions with control predicates: *There continued/*tried to be demonstrations on campus*. *See also* control.

reciprocal A reciprocal pronoun is one that expresses a mutual relationship, such as the English pronoun *each other*. *See also* anaphora.

referent This term is used for the entity (e.g. a person, object, notion, or situation) that is denoted by (a use of) a linguistic expression.

reflexive Many languages use special forms of pronouns when the subject and object refer to the same individual or individuals, e.g. the English forms ending in *-self* or *-selves*. These are called ‘reflexives’ or ‘reflexive pronouns’. It is common for these pronouns also to be acceptable in some other environments, but those environments differ from language to language. *See also* anaphora; binding.

regular expression It is possible to characterize the well-formed expressions of some simple formal languages by means of a few abbreviatory devices. One system that has proved very useful in some contexts involves templates, made up of words and/or categories of words, together with parentheses (to indicate optionality), a disjunction symbol (to indicate alternatives), and Kleene star [q.v.] (and/or Kleene plus), to indicate arbitrary numbers of repetitions of a sequence. Such templates are called ‘regular expressions’. *See also* finite-state grammar.

relative clause These are clauses that are used to modify nouns or noun phrases. A relative clause characteristically contains either a gap or a pronoun understood to be coreferential with the noun it modifies.

root sentence The traditional distinction between main clause and subordinate clause is motivated in part by the fact that certain phenomena seem to be restricted to main clauses, e.g. the inversion of finite auxiliaries [q.v.] in English interrogatives

(compare: *Can I do it?* vs. *I wonder whether I can do it*). Consequently, some version of this distinction has been maintained in most formal theories of grammar. The term ‘root sentence’ is sometimes used for main clauses, or, more technically, a phrase of category S that is not dominated by anything else. *See also* subordinate clause.

saturated In the system of grammar developed in this book, a saturated phrase is one that is specified as [SPR < >] and [COMPS < >]. The intuition behind this is that headed phrases can be thought of as being generated bottom-up, starting from the lexical head, via a regime of cancelling elements from the head’s valence specifications. For example, a verb combines first with however many complements are on its COMPS list to build a VP (a verbal phrase that is [COMPS < >] but [SPR < NP >]); the resulting (SPR-)unsaturated phrase then combines with the subject NP to build a saturated phrase, i.e. an S.

semantics Semantics is the branch of linguistics concerned with the study of linguistic meaning. Linguists also use the locution ‘the semantics of’ some expression as a way of talking about the literal interpretation of that expression. Not all information that is conveyed by the utterance of an expression is part of its semantics, but the line between literal meaning and what is conveyed can be hard to draw. At a minimum, the semantics of a (declarative) sentence is usually taken to include a specification of the conditions under which it would be true. *See also* pragmatics.

situation Situations are what natural language sentences are about: events or states (real or imaginary), involving entities, their properties, and their relations to one another.

specifier We use the term ‘specifier’ to cover subjects of clauses, determiners of noun phrases, and certain other constituents that are neither heads of the phrases they appear in nor complements to the heads. In English, the specifier of a phrase precedes its head [q.v.] and complements [q.v.]. *See also* determiner; complement.

subcategorization Lexical heads differ according to how many and what types of things they must combine with in order to make complete phrases. Each grammatical category [q.v.] (that is, part of speech [q.v.]) can be divided into subcategories, based on the valence, or combinatoric potential, of the particular words. When we talk of the subcategorization of a verb (or other type of head), we mean the restrictions on which sorts of phrases it can combine with. Another common locution is to say that a given verb ‘subcategorizes for’ a certain phrase, meaning that it combines with such a phrase. *See also* valence.

subordinate clause A subordinate clause is one that is dependent on, and usually a constituent [q.v.] of, another clause [q.v.]. An example of a subordinate clause is *when Kim went* in *Sandy came when Kim went*. *See also* root sentence.

suffix *See* affix.

tense Finite verbs come in different forms depending on the time they denote; these forms are called ‘tenses’. English has present and past tense, exemplified by the present tense forms *walk* and *walks*, and by the past tense form *walked*. Some languages also have future tenses, but English uses other means (e.g. the modal [q.v.] *will*) to express future time. *See also* aspect; finite verb.

transitive verb Verbs that take an NP object are called ‘transitive’. The term can also be used for other parts of speech that can take objects, e.g. prepositions. It is sometimes taken to encompass both simple transitive verbs (that is, those taking a single object) and ditransitive verbs. A standard example of a transitive verb is *hit*. *See also* ditransitive verb; intransitive verb; valence.

type Elements of any collection can be sorted into types, based on properties they have in common. In the theory presented in this text, linguistic entities (e.g. words and phrases) are described by means of feature structures [q.v.]. Particular features are appropriate only for certain types of entity, and constraints on possible feature-value pairings are also associated with particular types. *See also* feature structure; inheritance hierarchy; lexicon.

type hierarchy Types (q.v.) are organized into a hierarchy that determines the properties of linguistic entities through the mechanism of constraint inheritance. The type hierarchy is especially important for capturing regularities in the lexicon. *See* inheritance hierarchy.

undergenerate A grammar that fails to license sentences that the grammar writer wants it to generate is said to ‘undergenerate’. *See also* overgenerate.

unification The operation of unification merges two feature structure descriptions into a description that contains all the information in both. Two feature structure descriptions can unify so long as the information in them is consistent – that is, so long as there is no feature for which they have conflicting values. The unification simply consists of all of the features and values specified in the two feature structure descriptions. Unification is an operation for merging descriptions of a certain form whose effect is equivalent to conjunction of constraints.

universal grammar Many linguists claim that there is a great deal in common among the grammars of the world’s languages. Most advocates of this position believe that the commonalities exist because linguistic structure is largely determined by human biology. The term ‘universal grammar’ is used to mean three subtly different things: (i) what is common to the world’s languages; (ii) linguists’ representations of these commonalities; and (iii) the biological structures that are claimed to underlie the common features.

unsaturated *See* saturated.

valence This term is used (by analogy with the chemical term) to refer to the combinatoric potential of words and phrases. In this text, the VALENCE (VAL) features are those that specify this kind of information. Specifically, the VAL features SPR and COMPS for the verb *put* specify that it requires a subject NP, an object NP, and a PP in order to form a clause. *See also* argument; argument structure; ditransitive verb; intransitive verb; transitive verb.

voice This term refers to the way the semantic arguments of the verb are expressed grammatically. The term is used in English primarily to distinguish active voice and passive voice, but some other languages have far richer systems of voice. *See also* active; passive.

word This term is used in many different ways. In this text, a word is a particular form derived from a lexeme by some inflectional rule. *See also* lexeme, lexicon.