An Introduction to Syntax

ROBERT D. VAN VALIN, JR.

Department of Linguistics, University at Buffalo, The State University of New York

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK www.cup.cam.ac.uk
40 West 20th Street, New York, NY 10011–4211, USA www.cup.org
10 Stamford Road, Oakleigh, Melbourne 3166, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain

First published 2001

Printed in the United Kingdom at the University Press, Cambridge

Typeface 10/12pt Times [GC]

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

Library of Congress Cataloging-in-Publication Data

Van Valin, Robert D., Jr.
An Introduction to Syntax / Robert D. Van Valin, Jr.
p. cm.
Includes bibliographical references and index.
ISBN 0 521 63199 8 (hdbk.) — ISBN 0 521 63566 7 (pbk.)
1. Grammar, Comparative and general—Syntax. I. Title.

P291.V357 2001
415—dc21 00–062129

ISBN 0 521 63199 8 hardback
ISBN 0 521 63566 7 paperback

Contents

List of figures

CHAPTER 1

Syntax, lexical categories, and morphology

1.0 Introduction

This book is an introduction to the basic concepts of syntax and syntactic analysis. Syntax is a central component of human language. Language has often been characterized as a systematic correlation between certain types of gestures and meaning, as represented simplistically in Figure 1.1. For spoken language, the gestures are oral, and for signed language, they are manual.

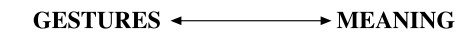

Figure 1.1. *Language as a correlation between gestures and meaning*

It is not the case that every possible meaning that can be expressed is correlated with a unique, unanalyzable gesture, be it oral or manual. Rather, each language has a stock of meaning-bearing elements and different ways of combining them to express different meanings, and these ways of combining them are themselves meaningful. The two English sentences *Chris gave the notebook to Dana* and *Dana gave the notebook to Chris* contain exactly the same meaning-bearing elements, i.e. words, but they have different meanings because the words are combined differently in them. These different combinations fall into the realm of syntax; the two sentences differ not in terms of the words in them but rather in terms of their syntax. Syntax can thus be given the following characterization, taken from Matthews (1982:1):

> The term 'syntax' is from the Ancient Greek *sýntaxis*, a verbal noun which literally means 'arrangement' or 'setting out together'. Traditionally, it refers to the branch of grammar dealing with the ways in which words, with or without appropriate inflections, are arranged to show connections of meaning within the sentence.

First and foremost, syntax deals with how **sentences** are constructed, and users of human languages employ a striking variety of possible arrangements of the elements in sentences. One of the most obvious yet important ways in which languages differ is the order of the main elements in a sentence. In English, for example, the **subject** comes before the **verb** and the **direct object** follows the verb. In Lakhota (a Siouan language of North America), on the other hand, the subject and direct object both precede the

verb, while in Toba Batak (an Austronesian language of Indonesia; Schachter 1984b), they both follow the verb. This is illustrated in (1.1), in which *the teacher*, *waǫ́spekhiye ki* and *guru i* function as subjects, and *a book*, *wówapi wą* and *buku* function as direct objects.

(1.1) a. The teacher is reading a book. English
b. Waǫ́spekhiye ki wówapi wą yawá. Lakhota
teacher the book a read
c. Manjaha buku guru i. Toba Batak
read book teacher the

The Lakhota and Toba Batak sentences also mean 'the teacher is reading the book', and in the Lakhota example the subject comes first followed by the direct object, whereas in the Toba Batak example the subject comes last in the sentence, with the direct object following the verb and preceding the subject. The **basic word order** in Toba Batak is thus the opposite of that in Lakhota. There are also languages in which the order of words is normally irrelevant to the interpretation of which element is subject and which is object. This is the case in the following Russian sentences.

(1.2) a. Učitel'nica čitaet knigu. Russian
teacher read book
b. Knigu čitaet učitel'nica.
book read teacher
c. Čitaet učitel'nica knigu.
read teacher book

Again, all three of these sentences mean 'the teacher is reading the book', and in these Russian examples the order of the words is not the key to their interpretation, as it is in the sentences from the other three languages. Rather, it is the form of the words that is crucial. The *-a* on the end of *učitel'nica* 'teacher' signals that it is the subject, and the *-u* on the end of *knigu* 'book' indicates that it is the direct object. If the word for 'teacher' were the direct object in a sentence, then it would end in *-u*, as in (1.3).

(1.3) a. Ženščina videla učitel'nic**u**. Russian
woman saw teacher
b. Učitel'nic**u** videla ženščina.
teacher saw woman
'The woman saw the teacher.'

These changes in the form of the words to indicate their function in the sentence are what Matthews referred to as 'inflections', and the study of the formation of words and how they may change their form is called **morphology**. These examples illustrate the important relationship between syntax and morphology: something which may be expressed syntactically in some languages may be expressed morphologically in others. Which element is subject and which is object is signalled syntactically in the examples from English, Lakhota and Toba Batak, while it is expressed morphologically in the Russian examples. Syntax and morphology make up what is traditionally referred to as '**grammar**'; an alternative term for it is **morphosyntax**, which explicitly recognizes the important relationship between syntax and morphology. Even though

this book is focussed on syntax, morphology will nevertheless be an important part of the discussion.

Thus a more complex picture of the nature of language emerges than that given in Figure 1.1; it is summarized in Figure 1.2.

Figure 1.2. *Language as a correlation between gestures and meaning (revised)*

All of the examples looked at so far involve simple sentences, but one of the most important syntactic properties of language is that simple sentences can be combined in various ways to form complex sentences. In terms of Figure 1.2, one could say that syntax makes possible the formulation of expressions with complex meanings out of elements with simple meanings. One of the defining features of human language is its unlimited nature; that is, the number of meaningful expressions that can be produced by users of a human language is potentially infinite, and this expressive potential comes from the combination of the basic meaningful elements with syntactic principles. Much of the interest in language in psychology and cognitive science comes from what the study of the cognitive mechanisms underlying language use and acquisition can reveal about the human mind.

This book has three goals: first, to introduce the basic concepts of syntax; second, to elucidate the principles and tools of syntactic analysis, which make it possible for linguists to analyze the grammatical systems of human languages; and third, to give an overview of the typological range of phenomena found in human languages which syntacticians seek to describe. The content of this book is presupposed by more advanced courses in syntactic theory, and hence it is intended to prepare the reader for such courses. The perspective of the book is primarily **descriptive**, and theoretical issues will be raised only in chapter 6. To many people the term 'grammar' evokes bad memories of **prescriptive** rules learned in school, e.g. 'don't split infinitives!' Since the early part of the twentieth century, linguistics has rejected the prescriptive tradition which underlies school grammars and focusses instead on describing what users of human language actually do, not on prescribing what they should do.

A central part of the description of what speakers do is characterizing the **grammatical** (or **well-formed**) sentences of a language and distinguishing them from **ungrammatical** or (**ill-formed**) sentences. Grammatical sentences are those that are in accord with the rules and principles of the syntax of a particular language, while ungrammatical sentences violate one or more syntactic rules or principles. For example, (1.1a) is a grammatical sentence of English, while *Teacher the book a reading is* would not be. Ungrammatical sentences are marked with an asterisk, hence **Teacher the book a reading is.* This sentence is ungrammatical because it violates some of the word order rules for English, that is (i) basic word order in English clauses is subject–verb–object, (ii) **articles** like *the* and *a* precede the **noun** they modify, and (iii) **auxiliary verbs** like *is* precede the **main verb**, in this case *reading*. It is important to note that these are English-specific syntactic rules; this word order is perfectly grammatical in Lakhota,

as (1.1b) shows, and if the Lakhota words were arranged in the English order, e.g. **Ki waǫ́spekhiye yawá wą wówapi* [the teacher reads a book], the result is thoroughly ungrammatical. Well-formed sentences are those that are in accord with the syntactic rules of the language; this does not entail that they always make sense semantically. For example, the sentence *the book is reading the teacher* is nonsensical in terms of its meaning, but it violates no syntactic rules or principles of English; indeed, it has exactly the same syntactic structure as (1.1a). Hence it is grammatical (well-formed), despite being semantically odd.

The organization of the book is as follows. In this chapter a number of distinctions that are relevant to the discussion in the remainder of the book are introduced. First, two aspects of syntactic structure are distinguished, one of which will be the main topic of chapters 2 and 3, and the other will be the main topic of chapter 4. Second, the traditional notion of parts of speech are reviewed, as these categories will be important throughout the book. Finally, a brief introduction to some of the basic concepts of morphology and morphological analysis is presented, with emphasis on those notions that will be especially pertinent to the discussion in the succeeding chapters.

The next three chapters present basic syntactic phenomena from two different analytic perspectives and introduce the concepts and analytic tools used in each. Many of the same grammatical phenomena will be analyzed from each perspective. In chapter 5 the basics of writing a grammar to describe syntactic phenomena will be presented; the formulation of rules to express the generalizations arising from syntactic analysis and the role of the lexicon in a grammar will be discussed. Different linguistic theories make different sets of assumptions about the nature of syntactic structure and accordingly employ different analytic principles and tools. In chapter 6 the basic ideas of four linguistic theories will be summarized, and their approaches to important grammatical phenomena, including the formation of information questions (e.g. *What did you see?*) and the **passive voice** (e.g. *The bread was eaten by the mouse*), will be compared and contrasted. These two phenomena are especially revealing for a comparison of theories, because the accounts given by the various theories highlight the conceptual and analytic differences among them.

1.1 Aspects of syntactic structure

In the syntactic structure of sentences, two distinct yet interrelated aspects must be distinguished. The first one has already been mentioned: the function of elements as subject and direct object in a sentence. 'Subject' and 'direct object' have traditionally been referred to as **grammatical relations**. Hence this kind of syntax will be referred to as '**relational structure**'. It includes more than just grammatical relations like subject and direct object; it also encompasses relationships like **modifier–modified**, e.g. *tall building* or *walk slowly* (*tall*, *slowly* = modifier, *building*, *walk* = modified) and **possessor–possessed**, e.g. *Pat's car* (*Pat's* = possessor, *car* = possessed). Relational structure will be the primary focus of chapters 2 and 3.

The second aspect concerns the organization of the units which constitute sentences. A sentence does not consist simply of a **string** of words; that is, in a sentence like *The teacher read a book in the library*, it is not the case that each word is equally related to the words adjacent to it in the string. There is no direct relationship between *read* and *a* or between *in* and *the*; *a* is related to *book*, which it modifies, just as *the* is related

to *library*, which it modifies. *A* is related to *read* only through *a book* being the direct object of *read*, and similarly, *the* is related to *in* only through *the library* being the object of the **preposition** *in*. The words are organized into units which are then organized into larger units. These units are called **constituents**, and the hierarchical organization of the units in a sentence is called its **constituent structure**. This term will be used to refer to this second aspect of syntactic structure. Consider the eight words in the sentence *The teacher read a book in the library*. What units are these words organized into? Intuitively, it seems clear that the article *the* or *a* goes with, or forms a unit with, the noun following it. Is there any kind of evidence beyond a native speaker's intuitions that this is the case? Determining the constituent structure of sentences is the major topic of chapter 4, but a brief preliminary look at the kind of evidence needed follows.

If the article forms a unit with the noun that follows it, we would expect that in an alternative form of the same sentence the two would have to be found together and could not be split up. Thus in the passive version of this sentence, *A book was read by the teacher in the library*, the unit *a book* serves as subject, and the unit *the teacher* is the object of the preposition *by*. The constituent composed of a noun and an article is called a **noun phrase** [NP]; as will be shown later, NPs can be very complex. The preposition *in* and the NP following it also form a constituent in this sentence (*in the library*); it is called a **prepositional phrase** [PP]. The fact that the PP is a constituent can be seen by looking at another alternative form, *In the library the teacher read a book*. Finally, the verb plus the NP following it form a unit as well, as shown by a sentence like *I expected to find someone reading the book, and reading the book was a teacher*. The constituent composed of a verb plus following NP is called a **verb phrase** [VP]. As with NPs, VPs can be quite complex. In each of these alternative forms, a combination of words from the original sentence which one might intuitively put together in a single unit also occurs together as a unit, and this can be taken as evidence that they are in fact constituents. Using square brackets to group the words in constituents together, the constituent structure of *The teacher read a book in the library* may be represented as in (1.4). ('S' stands for 'sentence'.)

(1.4) $[_{S}\,[_{NP}$ The $[_{N}$ teacher]] $[_{VP}\,[_{V}$ read] $[_{NP}$ a $[_{N}$ book]] $[_{PP}\,[_{P}$ in] $[_{NP}$ the $[_{N}$ library]] $_{PP}]$ $_{VP}]$ $_{S}]$

Note the nesting of constituents within constituents in this sentence, e.g. the NP *the library* is a constituent of the PP *in the library* which is a constituent of the VP *read a book in the library*. In chapter 4 constituent structure will be explored in detail.

At the beginning of this section it was noted that the two aspects of syntactic structure, relational structure and constituent structure, are 'distinct yet interrelated', and it is possible now to see how this is the case. For example, a VP was described as being composed of a verb and the following NP, but it could alternatively be characterized as involving the verb and its direct object. Similarly, a PP is composed of a preposition and its object. NPs, on the other hand, involve modifiers, and accordingly the relation between *the* and *teacher* could be described as one of modifier–modified. Thus, these two aspects of syntactic structure are always present in a sentence, and when one or the other is emphasized, the sentence is being described from one of the two perspectives. It will be seen later that different grammatical phenomena seem to be more easily analyzed from one perspective rather than the other.

1.2 Lexical categories

In the discussion of the constituents of sentences, reference has been made to nouns and noun phrases, verbs and verb phrases, and prepositions and prepositional phrases. Nouns, verbs and prepositions are traditionally referred to as 'parts of speech' or 'word classes'; in contemporary linguistics they are termed **lexical categories**. The most important lexical categories are noun, verb, **adjective**, **adverb** and **adposition**, which subsumes prepositions and **postpositions**. In **traditional grammar**, lexical categories are given **notional definitions**, i.e. they are characterized in terms of their semantic content. For example, *noun* is defined as 'the name of a person, place or thing', *verb* is defined as an 'action word', and *adjective* is defined as 'a word expressing a property or attribute'. In modern linguistics, however, they are defined morpho-syntactically in terms of their grammatical properties.

Nouns may be classified in a number of ways. There is a fundamental contrast between nouns that refer uniquely to particular entities or individuals and those that do not; the best example of the first kind of noun is a proper name, e.g. *Sam*, *Elizabeth*, *Paris* or *London*, and nouns of this type are referred to as **proper nouns**. Nouns which do not refer to unique individuals or entities are called **common nouns**, e.g. *dog*, *table*, *fish*, *car*, *pencil*, *water*. One of the important differences between proper and common nouns in a language like English is that common nouns normally take an article, while proper nouns do not, e.g. *The boy left* versus **The Sam left* (cf. **Boy left* versus *Sam left*). Common nouns may be divided into **mass nouns** and **count nouns**. Count nouns, as the name implies, denote countable entities, e.g. *seven chairs*, *six pencils*, *three dogs*, *many cars*. Mass nouns, on the other hand, are not readily countable in their primary senses, e.g. **two waters*, **four butters*, **six snows*. In order to make them countable, it is necessary to add what is sometimes called a 'measure word', which delimits a specific amount of the substance, e.g. *two glasses/bottles/drops of water, four pats/sticks of butter, six shovelfuls of snow*. Measure words can be used with count nouns only when they are plural, e.g. **six boxes of pencil* versus *six boxes of pencils*, **two cups of peanut* versus *three jars of peanuts*. **Pronouns** are closely related to nouns, as they both function as NPs. Pronouns are traditionally characterized as 'substitutes' for nouns or as 'standing for' nouns, e.g. *John went to the store, and he bought some milk*, in which *he* substitutes or stands for *John* in the second clause. This, however, is true only of **third-person** pronouns like *he*, *she*, *it*, or *they*; it is not true of **first-person** pronouns like *I* or **second-person** pronouns like *you*. First- and second-person pronouns refer to or index the speaker and addressee in a speech event and do not replace or stand for a noun.

Verbs can likewise be categorized along a number of dimensions. One very important dimension which will be discussed in detail in chapters 2 and 3 is whether a verb takes just a subject (an **intransitive** verb), or a subject and a direct object (a **transitive** verb), or a subject, direct object and **indirect object** (a **ditransitive** verb). This will be referred to as the 'valence' of the verb. Another dimension concerns the kind of situation it represents. Some verbs represent static situations which do not involve anyone actually doing anything, e.g. *know* as in *Chris knows the answer*, or *see* as in *Pat sees Dana over by the bookcase*. Some symbolize actions, e.g. *run* as in *Kim ran around the track*, or *sing* as in *Leslie sang a beautiful aria*. Others refer to a change of state, e.g. *freeze* as in *The water froze* (the change in the state of the water is from liquid to solid), or *dry* as in *The clothes dried quickly* (the change in the state of the

clothes is from wet to dry). Some represent complex situations involving an action plus a change of state, e.g. *break* as in *Larry broke the window with a rock* (Larry does something with a rock [action] which causes the window to break [change of state]). This classification of verbs is quite complex and is more appropriately in the domain of semantics rather than syntax. However, some syntactically relevant aspects of the meaning of verbs will be investigated in chapter 2.

Some examples of adjectives in English include *red*, *happy*, *tall*, *sick*, *interesting*, *beautiful*, and many others. Adjectives typically express properties of entities, e.g. *a* red *apple*, *a* tall *woman*, *a* beautiful *sunset*. Some properties are inherent attributes of an entity; for example, some apples are red because they are naturally so, whereas some barns are red because they have been painted red, not because they are inherently red. Hence color is an inherent property of apples but not of barns. Some languages signal this distinction overtly. In Spanish, for example, the adjective *feliz* means 'happy', and whether it is an inherent or permanent property of the person referred to is signaled by the verb it is used with, i.e. *Maria es feliz* 'Maria is happy (a happy person)' versus *Maria está feliz* 'Maria is happy (now, at this moment but not necessarily always)'. Spanish has two verbs meaning 'be', *ser* and *estar*, and one of the differences between them is that *ser* plus adjective (*es* in this example) is used to signify inherent or permanent attributes, while *estar* plus adjective (*está* in this example) serves to indicate non-permanent, transitory attributes.

English adverbs typically, but not always, end in *-ly*, e.g. *quickly*, *happily*, *beautifully*, *rapidly* and *carefully*. *Fast* and *friendly* are exceptions; *fast* is an adverb without *-ly* (it can also be an adjective), and *friendly*, despite the admonitions of road signs in Texas to 'drive friendly', is an adjective, e.g. *a* friendly *waiter*. Adverbs modify verbs, adjectives and even other adverbs, and they can be classified in terms of the nature of this modification; manner adverbs, for example, indicate the manner in which something is done, e.g. *The detective examined the crime scene carefully*, or *The ballerina danced beautifully*, while temporal adverbs, as the name implies, express when something happened, e.g. *Kim talked to Chris yesterday*, or *Dana will see Pat tomorrow*. *Yesterday* and *tomorrow* do not end in *-ly* and have the same form when functioning as an adverb that they have when functioning as a noun, e.g. *Yesterday was a nice day*, *Tomorrow will be very special*. The most common adverbial modifiers of adjectives and adverbs are words like *very*, *extremely*, *rather*, e.g. *a very tall tree*, *the extremely clever student*, *rather quickly*. This class of adverbs is referred to as **degree modifiers**.

Prepositions are adpositions that occur before their object, while postpositions occur after their object. English and Spanish have only prepositions, e.g. English *in*, *on*, *under*, *to*, Spanish *en*, *a*, *con*, whereas Japanese and Korean have only postpositions. German has both: *in dem Haus* 'in the house' (preposition *in*) versus *dem Haus gegenüber* 'over across from the house' (postposition *gegenüber*).

There are a number of minor categories. The category of **determiners** includes articles like *a* and *the*, and **demonstratives** like *this* and *that*. Determiners modify nouns in relation to their referential properties. Articles indicate roughly whether the speaker believes her interlocutor(s) can identify the referent of the NP or not; an **indefinite article** like *a(n)* signals that the speaker does not assume the interlocutor(s) can identify the referent of the NP, while a **definite article** like *the* indicates that the speaker does assume that the interlocutor(s) can identify it. Demonstratives, on the other hand, refer to entities in terms of their spatial proximity to the speaker; English

this refers to an entity close to the speaker, while *that* refers to one farther away. (*Which book do you mean? This one here or that one over there?* versus **This one over there or that one here?*) Many languages make a three-way distinction: close to the speaker (English *this*, Spanish *esta* [FEM]), away from the speaker but not far (English *that*, Spanish *esa* [FEM]), and farther away from the speaker (archaic English *yon*, Spanish *aquella* [FEM]). These distinctions are also expressed by locative demonstratives, e.g. English *here*, German *hier*, Spanish *aqui* versus English *there*, German *da*, Spanish *ahí* versus English *yonder*, German *dort*, Spanish *allí*. **Quantifiers**, as the label implies, express quantity-related concepts. English quantifiers include *every*, *each*, *all*, *many*, and *few*, as well as the numerals *one*, *two*, *three*, etc., e.g. *every boy*, *many books*, *the seven sisters*. **Classifiers** serve to classify the nouns they modify in terms of shape, material, function, social status and other properties. They are found in many East and Southeast Asian and Mayan languages, among others. They are similar in many respect to the measure words that occur with English mass nouns, but they occur with all nouns regardless of the count–mass distinctions, e.g. Cantonese *yāt būi séui* [one CL water] 'one cup of water versus *yāt jā séui* 'a jug of water', versus *yāt jēun séui* 'a bottle of water' with a mass noun, *nī ga dihnlóuh* [this CL computer] 'this computer' (classified as machine) versus *nī bouh dihnlóuh* 'this computer' (classified as model) versus *nī go dihnlóuh* 'this computer' (classified as object) with a count noun (Matthews and Yip 1994). **Conjunctions**, like *and*, *but* and *or*, serve to link the elements in a conjoined expression. There are conjoined NPs, e.g. *a boy and his dog*, conjoined verbs, e.g. *Leslie danced and sang*, and conjoined adjectives, e.g. *Lisa is tall and slender*. All major lexical categories can be linked by conjunctions to form conjoined expressions; this will be discussed in more detail in chapters 3 and 4. **Complementizers** mark the dependent clause is a complex sentence, e.g. English *that* as in *Sally knows that Bill ate the last piece of pizza*. The final category is **particles**, which is a classification often given to elements which do not fall into any of the other categories. Many particles have primarily **discourse** functions, e.g. English *indeed*, German *doch*, Spanish *entonces*.

There is an important opposition that divides lexical categories into two general classes, based on whether the membership of the class can readily be increased or not. Languages can usually increase their stock of nouns, for example, by borrowing nouns from other languages or creating new ones through **compounding** (e.g. *black* + *board* yields *blackboard*) or other morphological means (e.g. *rapid* + *-ly* = *rapidly*), but they do not normally create or borrow new adpositions, conjunctions or determiners. Lexical categories such as noun and verb whose membership can be enlarged are termed **open class** categories, whereas categories such as adposition, determiner or conjunction, which have small, fixed membership, are called **closed class** categories.

The definitions of lexical categories given so far are primarily the notional ones from traditional grammar. These definitions seem intuitively quite reasonable to speakers of Indo-European languages, and they seem to correlate nicely with the syntactic functions of the different parts of speech. Let us define three very general syntactic functions: **argument**, modifier and **predicate**. In a sentence like *the teacher read an interesting book*, *the teacher* and *an interesting book* are the arguments, *read* is the predicate, and *the*, *an* and *interesting* are modifiers. Similarly, in *Kim is tall*, *Kim* is the argument and *is tall* is the predicate. The term 'argument' here includes NPs and PPs functioning as subject, direct object or indirect object. The notions of predicate and

argument will be discussed in more detail in the next chapter, but for now one can say simply that in a sentence the predicate expresses the state of affairs that the referents of the arguments are involved in. (The terms 'predicate' and 'argument' are also used in semantics with a different meaning; they are being used here and elsewhere to refer to syntactic notions, unless otherwise noted.) It is usual to distinguish 1-place, 2-place and 3-place predicates, depending on how many participants there are in the state of affairs depicted by the predicate. Being sick is a state of affairs involving only one participant, hence *be sick* is a 1-place predicate which takes one argument, e.g. *Kim is sick.* In *the teacher destroyed the note*, there is an action of destroying involving a teacher and a note. Destroying involves a destroyer and something destroyed; hence *destroy* is a 2-place predicate and takes two arguments. Finally, giving involves a giver, something given and a recipient, and therefore *give* is a 3-place predicate and takes three arguments, e.g. *The teacher gave an interesting book to Kim.* Given these distinctions, it seems intuitively clear that nouns would be arguments, verbs would be predicates and adjectives would be modifiers, and this is in fact the case very often.

But not always. Nouns and adjectives can function as part of a predicate, as in *Dana is a phonologist* and *Chris was sick*. Even though they are part of the predicate, they are still formally distinct from verbs; they do not take **tense suffixes** like verbs do, i.e. **Dana phonologists* or **Chris sicked.* The **copula** *be*, a kind of verb, carries these verbal inflections. Contrast this with the situation in Lakhota, in which nouns and adjective-like words do bear verbal inflections when functioning as predicates, in this instance **agreement** in **number** with the subject.

(1.5) a. Wičháša ki hená lową́-pi. Lakhota
man the those sing-PL
'Those men are singing.'[1]

a′. Lakhóta ki hená lową́-pi.
Sioux the those sing-PL
'Those Siouxs (Indians) are singing.'

b. Wičháša ki hená lakhóta-pi.
man the those Sioux-PL
'Those men are Siouxs (Indians).'

b′. Lakhóta ki hená wičháša-pi.
Sioux the those man-PL
'Those Siouxs (Indians) are men.'

c. Wičháša ki hená khúža-pi.
man the those sick-PL
'Those men are sick.'

[1] In most examples from languages other than English, there will be an interlinear **gloss** with a translation for each meaningful element in the sentence directly under it as well as a free translation into English in the third line. In the interlinear gloss, the translation will be lined up directly under the element being translated. Complex words will be broken up into their meaningful parts (see section 1.3 below) separated by hyphens, and the translation for each part will be joined to the translations for the other parts by hyphens and placed below the whole word. Thus in (1.5a), for example, *wičháša* means 'man', *ki* means 'the' and *hená* means 'those'; the last word, *lową́pi*, is broken up into two parts, *lową́* and *pi*, which are linked by a hyphen, and each part is translated (*lową́* means 'sing' and *pi* means 'plural subject'), with the translations linked by a corresponding hyphen and placed below the Lakhota word. If an element requires a translation involving more than one English word, the words will be joined by a '.', e.g. 'was.washed' in (2.4a). Finally, grammatical notions like tense and number are glossed using abbreviations which are listed at the beginning of the book.

Nouns in Lakhota do not normally carry any indication of number; the only way to tell that the NP containing *wičháša* 'man' is **plural** in (1.5a) is by means of the plural demonstrative *hená* 'those' (cf. *hé* 'that'). In particular, the plural suffix *-pi* is impossible on the noun *wičháša* in (1.5a); based on (1.5a), (1.5a′), one could conclude that it occurs only on verbs. But this would be incorrect, as the sentences in (1.5b, c) show. Nouns like *wičháša* 'man' do take *-pi* when they function as a predicate, rather than as an argument. Hence nouns in Lakhota seem to function readily as predicates, something their English counterparts do not do. Adjective-like words also function directly as predicates, as (c) illustrates; there is no copular element analogous to English *be* in either of these sentence types. Verbs and adjective-like words can also serve as arguments in Lakhota, as in (1.6).

(1.6) a. Hokšíla ki hená čhéya-pi. Lakhota
boy the those cry-PL
'Those boys are crying.'
b. Čhéya ki hená hokšíla-pi.
cry the those boy-PL
'The ones crying are boys.'
c. Khúža ki hená wičháša-pi. (cf. (1.5c))
sick the those man-PL
'The sick ones are men.'

The verb *čhéya* 'cry' serves as the predicate in (1.6a) and the argument in (1.6b). Note that in the English translation the verb *cry* cannot simply function as the subject; it must, rather, occur in a complex expression *the ones crying*. In Lakhota, by contrast, *hokšíla* 'boy' and *čhéya* 'cry' simply exchange positions in the sentence without any formal modification. The same is true of the noun *wičháša* 'man' and the adjective-like word *khúža* 'sick' in (1.6c). Thus, the expected correlations between noun and argument, verb and predicate and adjective and modifier are not as strong in Lakhota as they are in English.

An even more striking example of this lack of correlation between lexical class and syntactic function can be seen in Nootka, a Wakashan language spoken on Vancouver Island in British Columbia, Canada (Swadesh 1939).

(1.7) a. Wała:k-ma qo:ʔas-ʔi. Nootka
go-3sgPRES man-the
'The man is going.'
a′. Qo:ʔas-ma wała:k-ʔi.
man-3sgPRES go-the
'The one going is a man.'
a″. Qo:ʔas-ma.
man-3sgPRES
'He is a man.'
b. ʔi:ḥ-ma qo:ʔas-ʔi.
large-3sgPRES man-the
'The man is large.'
b′. Qo:ʔas-ma ʔi:ḥ-ʔi.
man-3sgPRES big-the
'The large one is a man.'

c. Ciqšƛ-ma ʔo:kwiɬ qo:ʔas-ʔi.
speak-3sgPRES to man-the
'He speaks to the man.'
c′. ʔi:ḷ-ma ciqšƛ-ʔi.
large-3sgPRES speak-the
'The one speaking is large.'
c″. ʔo:kwiɬ-ma qo:ʔas-ʔi.
to-3sgPRES man-the
'He is [in relation] to the man.'
d. Wałaːk-ma ʔatḥiya.
go-3sgPRES night
'He is going at night.'
d′. ʔatḥiya-ma wałaːk-ʔi.
night-3sgPRES go-the
'His going is at night.'

The basic pattern in these Nootka sentences is PREDICATE-*ma* ARGUMENT-*ʔi*, with *-ma* signalling both tense and subject agreement in terms of **person** and number. The sentences in (1.7a), (1.7a′) and (1.7b), (1.7b′) look like their Lakhota counterparts in (1.5) and (1.6). The striking examples are in (1.7c) and (1.7d). (1.7c) contains a preposition, *ʔo:kwiɬ* 'to', and in (1.7c″) it functions as the predicate, as indicated by its occurring first in the sentence with the suffix *-ma*. In (1.7d) there is an adverb, *ʔatḥiya* 'at night', and in (1.7d′) it is the predicate. Thus in Nootka, the expected correlations between lexical category and syntactic function appear to be even weaker than in Lakhota.

This has important implications for the traditional view of lexical categories. This view assumes that the semantics of words predict their category and hence their function. The examples from Lakhota and Nootka call this seriously into question, since functioning syntactically as predicate, argument or modifier does not follow from the meaning of words as expected in many instances. Moreover, the first link in the chain of inference, from meaning to category, does not even hold up from a cross-linguistic perspective. The notional account assumes, at least implicitly, that the major lexical categories are universal, but this turns out not to be true for all of them. Every language has noun and verb as lexical categories, even Nootka. This reflects the fundamental role of reference and predication in communication. One of the most important functions of language is to allow speakers to depict states of affairs in the world, and in order for them to do this, there must be linguistic devices which refer to the participant(s) in a state of affairs and other devices which denote the action, event or situation in a state of affairs. Lexical items specialized for the first task are nouns, those specialized for the second are verbs. Even though in examples from Lakhota and Nootka verbs function as arguments (and hence as referring expressions) and nouns as predicates, it is nevertheless the case that the basic use of words like *wičháša* 'man' in Lakhota and *qo:ʔas* 'man' in Nootka is as an argument; similarly, the basic use of words like *čhéya* 'cry' in Lakhota and *wałaːk* 'go' in Nootka is as a predicate.

What about the other major lexical categories? There are languages which lack adpositions altogether; they express the semantic content of prepositions and post-positions by means of the kind of suffixes on nouns in the Russian examples in (1.2)

and (1.3). The concepts expressed by these endings are called '**case**', and the endings are called 'case markers'. Case will be discussed a great deal throughout this book. Russian has both case suffixes and prepositions, but Dyirbal, an Australian Aboriginal language (Dixon 1972), has only case suffixes and no adpositions at all. Hence the lexical category 'adposition' is not universal. It also appears that adjective is not universal. In Lakhota, for example, the words expressing properties like 'red', 'tall', 'big', etc., are formally verbs and have basically the same morphosyntactic properties as verbs, as the examples in (1.6) showed in part. Hence there is no reason to posit a category 'adjective' distinct from that of 'verb'; the words corresponding to adjectives in a language like English, e.g. *khúža* 'sick' in (1.5) and (1.6), are really a subtype of verb in Lakhota. In Dyirbal and Quechua, spoken in the northern Andes mountains in South America (Wölck 1987), on the other hand, words of this type have the same morphosyntactic properties as nouns, and therefore they should be analyzed as a subtype of nouns. Finally, there has been much less research done on adverbs cross-linguistically than the other major categories, and therefore it is difficult to draw any conclusions about their universality.

Thus, it appears that noun and verb are universal lexical categories, but adposition and adjective are not. It is crucial to keep in mind that when it is claimed that adjective is not a universally valid lexical category, it does not mean that there are languages which lack words expressing properties like 'red', 'big', 'happy', etc. Rather, it means that the words expressing these notions behave morphosyntactically like members of one of the other classes (verb in Lakhota, noun in Dyirbal and Quechua).

In modern linguistics, the determination of the category of a word is not based on its meaning but rather on its morphosyntactic behavior, i.e. the elements it cooccurs with and the morphosyntactic environment(s) it occurs in. Meaning is not irrelevant to the function of a word, but it does not reliably predict it either. The term which is used to refer to classes based on their morphosyntactic properties is **form class**. Consider the similarities and differences between common and proper nouns in English, which was initially characterized semantically. They are both a type of noun, because they both occur in the major morphosyntactic environments which nouns (and NPs) occur in, e.g. as the subject or direct object of a verb, as the object of a preposition in a PP, and with *be* as a **predicate nominal** (*The girl gave a book to the teacher*, *Pat introduced Kim to Dana*; *Max is my lawyer*, *My lawyer is Max*). Other form classes cannot occur in these positions, e.g. **The yellow put a clumsily on the receive.* However, they differ in that common nouns can be modified by determiners and adjectives, while proper nouns cannot, e.g. *a tall girl* versus **a tall Dana*. Furthermore, common nouns, if they are count nouns, can take plural inflection, while proper nouns cannot, e.g. *the tall girls* versus **Danas*. Thus there are both syntactic and morphological differences between common and proper nouns which can be used to distinguish them as belonging to two distinct subclasses of the category noun.

English verbs can be differentiated from the other major classes by both morphological and syntactic criteria. Morphologically, only verbs take the suffixes *-ing* 'progressive', *-ed* 'past tense', or 'past participle', *-s* 'third-person singular subject–present tense' and *-en* 'past participle'. Syntactically, they occupy a unique position in a clause, and they may be modified by adverbs but not by adjectives or demonstratives. There are no consistent morphological properties that characterize English adjectives; there are distinctive endings that some adjectives carry, e.g. *-y* as in *slimy* (related to the noun

slime) or *tricky* (related to the noun *trick*), and *-ic* as in *toxic* (related to the noun *toxin*) or *metric* (related to the noun *meter*). Many adjectives take *-er* for their **comparative** forms, e.g. *taller*, *faster*, and *-est* for their **superlative** forms, e.g. *tallest*, *fastest.* However, many do not, e.g. **beautifuler*, **beautifulest*; these adjectives take *more* and *most* to indicate their comparative (*more beautiful*) and superlative (*most beautiful*) forms. English adjectives occupy a specific position within NPs, i.e. DEM-QNT-ADJ-N, as in *the seven tall trees* (**tall the seven trees*, **the tall seven trees*), and they may function predicatively only in combination with the copula *be*, e.g. *The tree is tall*, **The tree talls*). Finally, English adverbs, as noted earlier, often (but not always) end in *-ly*; they function only as modifiers (but never of nouns), e.g. *the* extremely *quick rabbit*, *the rabbit ran* very quickly, **the* quickly *rabbit*, and never as predicates, e.g. **The rabbit is* quickly.

This brief discussion of the morphosyntactic properties of the major English classes has not been exhaustive, but it does illustrate how morphological and syntactic criteria can be used to characterize the form classes in a language. Even though the criteria for the classes are ultimately morphosyntactic, the labels for the classes reflect the traditional notional distinctions. That is, after having established the existence of a form class based on the morphosyntactic properties of its members, the semantic properties of the prototypical members of the class determine the name of the class. Hence if the prototypical members of a class include elements that function as the name of a person, place or thing, then the class will be given the label 'noun'.

1.3 **Morphology**

Even though this book is about syntax and syntactic analysis, it is not possible to get very far without some basic knowledge about morphology. It was already shown in section 1.0 that some languages use morphology to express what other languages express syntactically, and in the previous section it was noted that the inflectional properties of words are relevant to determining their category. In this section some basic concepts of morphology and the basic techniques of morphological analysis will be introduced; both will play a role in the syntactic analyses in later chapters. The discussion will be limited to those aspects of morphology which are relevant to syntactic analysis; this is not intended to be a general introduction to this complex and important part of linguistics.

Morphology is concerned with the structure of words, and morphological analysis is the process by which linguists break complex words down into their component parts. Consider the Lakhota word *wahi*, which means 'I arrive'. Just looking at it by itself, it is not possible to determine whether it is a simple or complex form. If it is compared with another form, *yahi* 'you [**singular**] arrive', it can be seen immediately that there is a common part to each of the words, *-i*, *-hi* or *-ahi*, and a different part in each, *wah-*, *wa-* or *w-* and *yah-*, *ya-* or *y-*. It is clear that these are complex forms made up of more than one component. There are two parts to each: one meaning 'arrive', which is *-i*, *-hi* or *-ahi*, and the other meaning first-person singular ('I') or second-person singular ('you') subject. It is not possible to tell from just these two forms, however, exactly what the two components are, since there are at least three ways to divide up these forms. Is the form for 'arrive' *-i*, *hi* or *ahi*? Is the form for first-person singular subject *wah-*, *wa-* or *w-*? The answer to this question becomes somewhat clearer when the form

ų́hi ‘we [**dual inclusive**] arrive’ (i.e. ‘you [sg] and I arrive’) is examined. The common parts to the three forms are *-i* or *hi*, and this would seem to eliminate the *-ahi* possibility. But we are still left with two possibilities for ‘arrive’ and for ‘I’ (*wah-* or *wa-*), ‘you’ (*yah-* or *ya-*) and ‘we [dl incl]’ (*ų́h-* or *ų́-*). In order to resolve the issue, it would useful to look at some other verbs, as in (1.8).

(1.8)	a. *wa*lową	‘I sing’	Lakhota
	a′. *ya*lową	‘you [sg] sing’	
	b. *wa*čhį	‘I want’	
	b′. *ya*čhį	‘you [sg] want’	
	c. *waʔ*u	‘I come’	
	c′. *yaʔ*u	‘you [sg] come’	
	d. na*wa*žį	‘I stand’	
	d′. na*ya*ži	‘you [sg] stand’	
	e. a*wa*phe	‘I wait’	
	e′. a*ya*phe	‘you [sg] wait’	

In the first example in each pair the only common semantic element is first-person singular subject, and the only form common to all of them is *wa-*. Similarly, in the primed examples in the set the only common semantic element is second-person singular subject, and the only form common to all of them is *ya-*. Therefore *wa-* must mean ‘I’ and *ya-* must mean ‘you [singular]’. If ‘I’ and ‘you’ are *wa-* and *ya-*, respectively, then ‘arrive’ must be *-hi*, not *-i*. Moreover, this means that ‘we [dual inclusive]’ must be *ų́-*, not *ų́h-*. Thus, the correct analysis of the first three forms is *wa-* ‘I’, *ya-* ‘you [sg]’, *ų́-* ‘we [dl incl]’ and *-hi* ‘arrive’. It may further be concluded that *lową* means ‘sing’, *čhį* means ‘want’, *ʔu* means ‘come’, *nažį* means ‘stand’ and *aphe* means ‘wait’. Each of these words *wahi*, *yahi* and *ų́hi*, as well as those in (1.8), is composed of two meaningful parts, which are called **morphemes**. Morphemes are the smallest meaningful units in language. *Hi* cannot be broken down into *h* + *i*, nor can *wa-* or *ya-* be broken down into *w-* or *y-* + *a*; these smaller forms have no meaning. Hence, *wa-*, *ya-* and *hi*, as well as *ų́-*, are all morphemes in Lakhota, as are *lową*, *čhį*, *ʔu*, *nažį* and *aphe*.

The part of the complex form to which a morpheme is added is called the **stem**. In these examples, the stems are *hi*, *lową*, *čhį*, *ʔu*, *nažį*, and *aphe*. Morphemes like *wa-* and *ya-* which occur before the stem are called **prefixes**. Morphemes which occur after the stem are called suffixes; examples of suffixes can be found in the Russian sentences in (1.2) and (1.3), i.e. the markers *-a* and *-u*, and in the Lakhota examples in (1.5) and (1.6), i.e. *-pi* ‘plural’. In (1.8d), (1.8e) *wa-* occurs within the stem itself; in these forms it is an **infix**. The general term which covers prefixes, infixes and suffixes is **affix**.

This example has been very simple, but it illustrates the basic principle used in breaking words down into their component morphemes: *look for recurring forms that correlate with consistent meanings*. In all of the Lakhota words examined above, the form *wa-* correlates with the meaning ‘first-person singular subject’, the form *ya-* correlates with the meaning ‘second-person singular subject’, and the form correlates *ų́-* with the meaning ‘first dual inclusive subject’. Similarly, in comparing (1.8a) and (1.8a′), the form *lową* consistently correlates with the meaning ‘sing’.

The basic principle of morphological analysis stated above (‘look for recurring forms that correlate with consistent meanings’), requires four very important qualifications.

The first is that a single meaning, e.g. plural, may be expressed by several different forms. Consider the following simple example from English.

(1.9) a. dog /dɔg/ a′. dogs /dɔgz/
b. cat /kæt/ b′. cats /kæts/
c. rose /roz/ c′. roses /rozəz/

The forms in the second column have a consistent semantic difference from the forms in the first column, but there does not appear to be a consistent formal distinction correlating with the semantic difference. Rather, there is /z/, /s/ and /əz/. However, if one looked at a large number of English nouns, one would find that nouns ending in a voiced sound take /z/ to indicate plural, that nouns ending in a voiceless, non-sibilant sound take /s/, and that nouns ending in a sibilant take /əz/. These forms are clearly related phonologically, and the form of the plural suffix is predictable from the phonological shape of the end of the word to which it is added. Hence these three forms may be viewed as conditioned variants of a single morpheme; they are referred to as its **allomorphs**. They are phonologically conditioned allomorphs, since the choice of allomorph is determined by the phonological shape of the stem. There are other allomorphs of the plural morpheme which are not phonologically conditioned. For example, the plural of *box* is *boxes*, but the plural of *ox* is *oxen*, not **oxes.* The plural of *ox* is said to be morphologically conditioned, because it is not phonologically predictable and is an idiosyncratic property of the word *ox.* Other examples of morphologically conditioned plurals in English include *mice* for *mouse* and *teeth* for *tooth.*

Words like *mice* and *teeth* illustrate the second qualification: meanings need not be represented by segmentable parts of words. With words like *dogs* (/dɔgz/) and *cats* (/kæts/), it is easy to break them up into two parts, one meaning ‘dog’ or ‘cat’ and the other meaning ‘plural’. But it is not obvious that *mice* (/mays/) and *teeth* (/tiθ/) can be broken down into two comparable parts, one meaning ‘mouse’ or ‘tooth’ and the other meaning ‘plural’. Forms like these have posed profound problems for morphological analysts and theorists alike. For the purposes of this book, it is enough to state that such a form is morphologically conditioned but unsegmentable (cf. *oxen*, which is morphologically conditioned but segmentable) and to state the meanings expressed by the form.

When there is a group of phonologically conditioned allomorphs like those of the English plural in (1.9), it is customary to select one of them as the basic allomorph to represent the morpheme. Typically, the allomorph occurring in the widest range of environments is taken as basic, although other factors may come into play. With respect to the three allomorphs in (1.9), /z/ occurs in the greatest number of environments, and therefore it is a good candidate for the basic allomorph.

As noted in the initial discussion in this section, there is a choice as to how to divide up a complex form. With respect to Lakhota *wahi* ‘I arrive’ and *yahi* ‘you [sg] arrive’, there were three initial hypotheses: *w-* ‘I’, *y-* ‘you’ and *-ahi* ‘arrive’, versus *wa-* ‘I’, *ya-* ‘you’ and *-hi* ‘arrive’, versus *wah-* ‘I’, *yah-* ‘you’ and *-i* ‘arrive’. Given just these two forms, there is no reason to choose one analysis over the others, but when the form *ų́hi* ‘we [dl incl] arrive’ is considered, the possibilities are reduced to two, *-hi* or *-i*, etc. Consideration of the data in (1.8) leads to the conclusion that the simplest analysis of all three forms is *wa-* ‘I’, *ya-* ‘you’, *ų́-* ‘we [dl incl]’ and *-hi* ‘arrive’. Why is this the simplest analysis? Because it avoids positing allomorphs for any of the morphemes.

If the analysis of 'arrive' as being *-i* were maintained, then one would have to claim that *wa-* 'I' has an allomorph *wah-* before the verb *-i*. While such an alternation is not impossible, positing it nevertheless results in a more complex account than the alternative analysis, which does not postulate any allomorphic variation for any of the morphemes. Hence, all things being equal, the simplest analysis is to be preferred, and one criterion for simplicity is positing the least amount of allomorphic variation compatible with the facts.

Not only can a single meaning be expressed by multiple forms, but a single form can express multiple meanings. This is the third qualification to the basic principle of morphological analysis. For example, in Russian the *-a* suffix on *učitel'nica* 'teacher' in (1.2) expresses three distinct concepts: **nominative** case, singular number, and **feminine gender**. The *-s* suffix on verbs in English likewise expresses three concepts: third-person subject, singular subject, present tense. Thus, even though there are simple instances in which a single form consistently pairs with a single meaning, the kinds of complexities involving allomorphic variation and multiple concepts in a single form are very common.

The fourth qualification is that structural patterns in a language may require the analyst to posit that a meaning is expressed by the *absence* of a form. Consider the following **paradigm** for the verb *hi* 'arrive' in Lakhota.

(1.10)	a. wahi	'I arrive'	Lakhota
	b. yahi	'you [sg] arrive'	
	c. hi	'he/she arrives'	
	d. ų̨hi	'we [dl incl] arrive'	
	e. ų̨hipi	'we [pl] arrive'	
	f. yahipi	'you [pl] arrive'	
	g. hipi	'they arrive'	

It has already been established that *wa-* signals first-person singular subject, *ya-* second-person subject and *ų̨-* first-person dual inclusive subject, and in (1.5) and (1.6) it was shown that *-pi* indicates plural, which is the case here as well. What, then, signals third person singular subject? The absence of *-pi* in (1.10a)–(1.10d) indicates that the subject is non-plural, and it is the absence of *wa-*, *ya-* and *ų̨-* that signals that the forms in (1.10c) and (1.10g) are third person. Thus, third person is marked by the absence of a prefix, and it is customary to represent this by '∅'. Hence the form for 'he/she arrives' would be *∅-hi*. This meaningful absence of a phonological form is called a **zero** morpheme. Zero morphemes are normally posited only within paradigms such as that in (1.10) in which one form is distinctive by virtue of the absence of an affix.

Morphemes may be divided into two general classes: **lexical morphemes**, which have substantive semantic content, e.g. English *dog*, *rose*, *cat*, or Lakhota *hi* 'arrive', and **grammatical morphemes**, which lack substantive semantic content and express grammatical notions like person, number, gender, tense or case, e.g. English *-s* on verbs, Lakhota *wa-*, or Russian *-a*. The lexical versus grammatical opposition correlates with the earlier distinction made between open and closed classes: typically, lexical morphemes are open-class items, while grammatical morphemes are closed-class items. This opposition also relates to another important contrast, that between **free morphemes** and **bound morphemes**. Free morphemes are elements that can stand alone as independent

words, e.g. *dog*, *car*, *to*, *the* and *or*, whereas bound morphemes cannot occur by themselves as independent words, e.g. *-s*, *-ing*, *-ed* in English. Bound morphemes are usually grammatical morphemes, while free morphemes may be both lexical and grammatical, as the above examples from English illustrate. There are languages in which lexical morphemes can be considered bound morphemes, in that they cannot occur without an accompanying grammatical morpheme. In Russian, for example, *vide-* means 'see', which makes it a lexical morpheme, but it is not a free morpheme, since it cannot occur as a complete word without the addition of a suffix indicating its tense and subject agreement, e.g. *-l-a* 'PAST-FEMsg', yielding *videla* 'saw'. These distinctions are summarized in Table 1.1.

Table 1.1. *Types of morphemes*

	Free	Bound
Lexical	*dog*, *sing*; Lakhota *hokšíla*, *čhéya*	Russian *vide-*
Grammatical	*the*, *a*; Lakhota *ki*, *hená*	*-s*, *-ing*, *-ed*; Lakhota *wa-*, *-pi*

Since the focus of this book is on syntax, primary concern will be given to grammatical morphemes which express syntactic notions, e.g. subject, or syntactically relevant notions, e.g. person agreement on a verb.

In the next chapter, the notion of subject and grammatical relations in general will be examined.

Notes and suggested readings

For an excellent overview of systems of lexical categories across languages, see Schachter (1985). Hopper and Thompson (1984), Langacker (1991) and Croft (1991) argue for the universality of noun and verb as lexical categories based on the fundamental role of reference and predication in language. Jacobsen (1979) takes a detailed look at Nootkan languages and argues that there is evidence in favor of postulating noun and verb as lexical categories in these languages. Dixon (1977a) investigates the category 'adjective' cross-linguistically.

An excellent introduction to morphological analysis is Nida (1946); two more recent texts, which include discussion of morphological theory, are Bauer (1988) and Spencer (1991).

Exercises

1. Pretend the italicized nonsense words in the following sentences are real words of English. Identify the form class of each one, and state the morphosyntactic properties of each that lead you to assign it to a particular category. [section 1.2]

   (1) a. The dog *wugged* the ball.
   b. The dog is *wugging* the ball.
   c. The dog likes to *wug* the ball.
   d. The dog gently *wugged* the ball.
   e. *The *wug* kicked the ball.
   f. *The dog chased the *wug* cat.

(2) a. The tall *blick* sat by the river.
b. The *blicks* played in the park.
c. Mary sent a present to her favorite *blick*.
d. Sam is not a *blick*.
e. *Max *blicked* the cat.
f. *The *blick* animal ran away.

(3) a. A *nork* person walked by the car.
b. Mary is very *nork*.
c. *Sam *norks*.
d. *The *nork* called me yesterday.

(4) a. *Li* cat slept by the fire.
b. I bought *li* three interesting books.
c. Mary didn't like *li* one.
d. I don't care for *li*.
e. *Two *li* dogs barked at the cat.
f. *Sam *lis* every day.

(5) a. Max walked *blishly* down the corridor.
b. Max walked down the corridor *blishly*.
c. *Blishly*, Max walked down the corridor.
d. Sam did so extremely *blishly*.
e. *Pat is *blishly*.
f. *The *blishly* woman looked unhappy.

(6) a. Larry placed the book *za* the table.
b. *Za* the table Sam found his glasses.
c. **Za* green book fell on the floor.
d. *I don't like *za*.
e. *Sam *zas* every day.
f. *Sam found his gloves *za*.

(7) a. Anna bought *nace* rare books.
b. I liked *nace* of them.
c. *Nace* left the party early.
d. I thought she bought too *nace*.
e. *Anna bought rare *nace* books.
f. *Sam *naces* every morning.
g. *The tall red *nace* fell off the shelf.

2. In the following verse from Lewis Carroll's famous poem *Jabberwocky*, identify the form class of each of the italicized words. State the morphosyntactic properties that lead you to assign it to a particular category. Give two different analyses of the words in the last line. [section 1.2]

'Twas *brillig*, and the *slithy toves*
Did *gyre* and *gimble* in the *wabe*;
All *mimsy* were the *borogroves*,
And the *mome raths outgrabe*.

3. Consider the following examples from English:
(1) The **break** is just above the knee.
(2) Please **maple syrup** your pancakes from the dispenser on the table. (sign in a cafe in Adelaide, South Australia)
(3) The **climb** up the north face is very difficult and dangerous.
(4) Kinko's, the new way to **office**. (advertisement)

(5) The referee **book**ed the player for the foul.
(6) A good **cry** often makes you feel better.

What are the implications of examples like these for the discussion of English lexical categories in section 1.2? Compare these with the examples from Nootka and Lakhota; is English as flexible in terms of the functions of its lexical categories as these languages? [section 1.2]

4. In the following Italian sentences, what formal properties do the nouns, articles, adjectives, adverbs and verbs have? Do they have any special morphological marking indicating their category? Do they have any specific or special syntactic properties? With respect to the latter, assume simple **declarative** utterances and ignore **elliptical** answers to questions, e.g. Q: *How does Fred run?* A: *Slowly*, and **vocative** expressions, e.g. *'Hey man!'*. [section 1.2]

(1)	La bella ragazza parla rapidamente.	'The pretty girl speaks rapidly.'
(2)	Il giovane ragazzo parla chiaramente.	'The young boy speaks clearly.'
(3)	Le belle ragazze parlano chiaramente.	'The pretty girls speak clearly.'
(4)	I giovani ragazzi parlano rapidamente.	'The young boys speak rapidly.'
(5)	Parlano.	'They speak.'
(6)	*I parlano.	
(7)	Parlano rapidamente.	'They speak rapidly.'
(8)	*Bella ragazza parla.	
(9)	Una bella ragazza parla.	'A pretty girl speaks.'
(10)	*Bella parla.	
(11)	La macchina rapida arriva.	'The fast car arrives.'
(12)	Le macchine rapide arrivano.	'The fast cars arrive.'
(13)	Il ragazzo povero arriva.	'The poor boy arrives.'
(14)	Chiaramente parla.	'He/she speaks clearly.'
(15)	La bella arriva.	'The pretty one arrives.'
(16)	Il giovane parla.	'The young one speaks.'
(17)	*Bella.	
(18)	*I.	
(19)	*Chiaramente.	
(20)	Parla.	'He/she speaks.'
(21)	Arrivano.	'They arrive.'
(22)	*Ragazzo.	
(23)	*Macchina.	

5. Break down the following words from Ayacucho Quechua (Parker 1969) into their constituent morphemes and state their meaning. [section 1.3]

(1)	Rimani.	'I speak.'
(2)	Warmita rikun.	'He/she sees the woman.'
(3)	Runtuta mikurqani.	'I ate an egg.'
(4)	Runa rimarqa.	'The man spoke.'
(5)	Warmi rikun.	'The woman sees him/her.'
(6)	Runakuna rimarqaku.	'The men spoke.'
(7)	Runata rikuni.	'I see the man.'
(8)	Warmikuna rimanku.	'The women speak.'
(9)	Runtuta mikuni.	'I eat an egg.'
(10)	Runa hatun.	'The man is big.'
(11)	Warmikunata rikurqani.	'I saw the women.'
(12)	Runa daliwan.	'The man hits me.'
(13)	Rimarqani.	'I spoke.'
(14)	Runtu hatun.	'The egg is big.'

(15)	Warmi rikurqawa.	‘The woman saw me.’
(16)	Rikurqa.	‘He/she saw him/her.’
(17)	Riman.	‘He/she speaks.’
(18)	Warmikuna rikurqawaku.	‘The women saw me.’
(19)	Warmi runtuta dalirqa.	‘The woman hit an egg.’
(20)	Mikurqaku.	‘They ate.’

What would the following forms mean:

(21) Mikun.
(22) Warmi runata rikurqa.
(23) Runata dalirqani.
(24) Runa rikurqawa.

How would you say the following in Ayacucho Quechua?

(25) The men saw an egg.
(26) The woman ate an egg.
(27) The women hit [PAST] the man.
(28) The men see the woman.

6. Break down the words in the following sentences from Hungarian (de Groot 1989) into their constituent morphemes and state their meaning. If a morpheme has more than one allomorph, give all of them. [section 1.3]

(1)	János a könyvet Marinak adta.	‘John gave the book to Mary.’
(2)	A könyv az asztal alatt van.	‘The book is under the table.’
(3)	Mari Jánosnak adta az órát.	‘Mary gave the clock to John.’
(4)	A gyerek látta Marit.	‘The child saw Mary.’
(5)	Az óra az asztal fölött van.	‘The clock is above the table.’
(6)	Mari látja Jánost.	‘Mary sees John.’
(7)	A gyerek tette az órát az asztalhoz.	‘The child put the clock near the table.’
(8)	Mari Jánosnak adja a könyvet.	‘Mary gives the book to John.’
(9)	A level az asztal alatt van.	‘The letter is under the table.’
(10)	János a könyvet az asztalra tette.	‘John put the book on the table.’
(11)	Mari látta a gyereket.	‘Mary saw the child.’
(12)	A könyv van az asztal mögött.	‘The book is behind the table.’
(13)	Mari írja a levelet.	‘Mary writes the letter.’
(14)	Mari a gyereknek adta az órát.	‘Mary gave the clock to the child.’
(15)	János írta a levelet.	‘John wrote the letter.’