1 Introduction to morphology and syntax

If you were to ask anyone the question "What is language?" you would probably receive an answer that includes the word "communication." Most of us, if we think about our language at all, have the common-sense notion that language exists for the purpose of communication. This way of thinking views language as a "tool" that people use to accomplish the "job" of communication. It may not be the only tool that people use for this job, and it may help accomplish other jobs as well. However, many people, both linguists and non-linguists, have the idea that the *main* purpose of human language is communication.

Viewing language as a tool has profound consequences for all kinds of applications. Whether you are planning to contribute to linguistic theory, document one of the many unwritten languages of the world, prepare educational materials, or simply learn to speak a second language, you will profit greatly from a perspective that considers language as a tool for communication. In this introductory section, we will explore this perspective in some detail, after which we will discuss some fundamental concepts of linguistic analysis.

Every tool has two components: a **FUNCTION** and a **FORM**. The function is the job the tool is designed to accomplish, and the form is the tangible structure that accomplishes that job. For example, the main function of the kind of hammer pictured here is to pound nails into wood and to remove them. The form is the shape of the iron head attached to a handle, as in this picture. Though individual hammers may differ from one another in many ways, they also have a lot in common. This particular form is specially adapted to the function of pounding nails. If it had a form that was very different from this, it would not serve this purpose. Imagine a hammer with a paper head, or one lacking a handle. Such poor excuses for hammers would not be very useful for pounding nails (though they might serve some other purpose). So the function "motivates" (provides a reason for) the form of this very useful device. Without a function, the form would be simply an odd-shaped lump of iron and wood.

Of course, you don't *have* to use a hammer to pound nails – a hard rock or the heel of your shoe might do. Furthermore, because the hammer has its particular form, it also may be used to accomplish other functions, perhaps straightening metal, or breaking up concrete. But its main function has the greatest influence on its basic form.

Language also consists of a function and a form. Common sense tells us that the main function of language is to help people communicate. The form consists of sounds, gestures, or other physical variations in the environment capable of being perceived by other people. Furthermore, as in the case of the hammer, the form of language makes sense in terms of its basic function, as we will see throughout this book. Without the function of communication, language would be no more than random noises or other physical variations in the environment.

While the hammer analogy may be helpful in understanding the relation between function and form, in fact language is a much more complex tool than a hammer in a number of ways. First of all, the function of language is more complex. While there are many kinds of nails, and several ways you may want to pound them in or pull them out, the ways of using a hammer are rather limited. On the other hand, there is an infinite number of ideas that people want to communicate every day, and many subtle kinds and shades of meaning that people feel a need to express. Second, the form of language is more complex than that of a hammer. The form of most languages consists of a small number of sounds, organized into **WORDS, PHRASES, CLAUSES, SENTENCES**, and **DISCOURSES**, including conversations, sermons, speeches, arguments, and other highly complex communicative structures.

As with any tool, the forms of a language "make sense" in terms of their functions, though they are not precisely determined (or mathematically "predicted") by those functions. Indeed, what we first notice about a new language is how different it is from our own. If all languages are tools to accomplish the job of communication, why are they so different from one another? To begin to answer this question, let's consider another cultural tool that varies greatly around the world – the structure of houses. The vast differences among houses from one part of the world to another reflect different solutions to similar problems – the needs for shelter, warmth, space for food preparation, rest, etc. The different solutions are motivated by many factors, including the local ecology, but the structure of a particular house is not *inevitable* given the various motivating factors. Even in my own town, some houses have flat roofs, and others have sloping roofs. The different forms of roofs all fulfill the same function of providing shelter. In a similar way, different languages may use very different forms to express the same concept.

Linguists have found that, in spite of the many superficial differences among languages, there is a core of basic similarities. Can you imagine a language without words?[1] Without sentences? Such ways of communicating do exist, e.g., facial expressions, and styles of dress. These systems do help people understand one another to a certain extent, but we would hardly want to call them languages. They compare to languages as rocks and shoes may compare to hammers – capable of being used to pound nails, but not uniquely adapted or designed for that purpose. A language, however, is a highly complex system of interrelated parts uniquely adapted for the purpose of human communication. Though individual languages

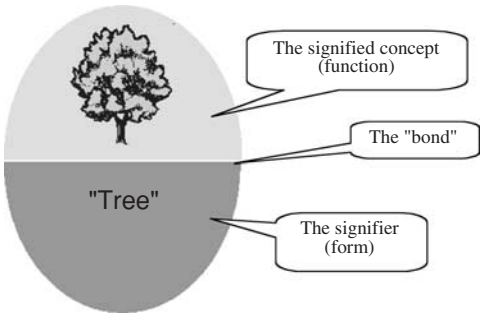

Figure 1.1 *The form–function composite*

do differ greatly in many respects, the functions of language provide a motivation for the many basic similarities in form.

In the following sections we will discuss some of the terms and concepts that linguists use to explore the structure of languages.

The form–function composite

Linguists usually assume that language consists of elements of form that people employ to "mean," "express," "represent," or "refer to" other things. Although linguists often imply that the linguistic forms themselves express concepts, this must be taken as a shorthand way of saying that speakers *use* linguistic forms (among other tools) to accomplish acts of expressing, referring, meaning, etc. (Brown and Yule 1983:27ff.). For example, a word is a linguistic form. In and of itself it is just a noise made by someone's vocal apparatus. What makes it a *word* rather than just a random noise is that it is produced intentionally in order to express some idea. When used by a skilled speaker, words can combine into larger structures to express very complex ideas. While linguistic forms help people formulate ideas, and may constrain the concepts that can be entertained, the linguistic forms themselves are logically distinct from the ideas that might be expressed, in the same way that the form of a hammer is distinct from the job of pounding nails.

Langacker (1987), building on Saussure (1915), describes linguistic units as consisting of **FORM–FUNCTION COMPOSITES**, as illustrated in figure 1.1.

The upper half of the diagram in figure 1.1 represents the meanings, concepts, or ideas expressed in language, while the bottom half represents the linguistic units themselves. The line across the center represents the relationship, or the "bond" between the two. Various terms have been used to refer to the parts of this composite. Terms associated with the top half include "signified," "meaning," "semantics," "function," "conceptual domain," and "content." Terms

associated with the bottom half include "sign," "signifier," "symbol," "structure," and "form."

In ancient times, philosophers who thought about language often considered words to be inherently connected to their meanings. Invariably, the language the philosopher spoke (Sanskrit, Greek, or Latin) was considered to be the language that expressed the "true" meanings of words. In more recent times, linguists have tended to emphasize the **ARBITRARINESS** of linguistic signs. That is to say, there is not necessarily an inherent connection between the form of a sign and its meaning. The noise spelled *tree* in English certainly has no inherent connection to the range of concepts that it can express. Indeed, even in related languages, such as German and French, very different noises (spelled *baum* and *arbre* respectively) express roughly the same idea. Even more recently, linguists are beginning to notice that linguistic signs are arbitrary to a certain extent, but that they are also **MOTIVATED** by factors such as understandability, **ICONICITY** (including **SOUND SYMBOLISM**), and economy.[2]

Why is the bond between sign and signified concept, form and function, motivated? Linguists assume that the bond between symbol and signified concept is intentional. That is, language users *intend* to establish a link between form and meaning – they consciously *want* their utterances to be understood. From this it follows that the forms used to represent concepts will be structured so as to make the link obvious, within limits of cognitive ability, memory, etc. This is not to deny the possibility that certain aspects of language may actually have no relation to the concepts expressed or may even serve to *conceal* concepts. However, we make it a working assumption that in general language users want and expect linguistic forms to represent concepts to be communicated.

In any symbolic system, there must be consistency in the relationship between the symbols and categories or dimensions in the symbolized realm. We do not live in a "Humpty Dumpty world" where words mean anything we want them to mean (Carroll 1872). In order to communicate with others, we rely on the probability that words in our language mean approximately the same thing to other people as they do to us. Ideal symbolic systems (e.g., computer "languages") maximize this principle by establishing a direct, invariant coding relationship between every form and its meaning or meanings. However, real languages are not ideal symbolic systems in this sense. They exist in an environment where variation and change are normal rather than exceptional. New functions appear every day as new situations, concepts, and perspectives speakers wish to express. Vocal and auditory limitations cause inexact pronunciation and incomplete perception of messages. These and many other factors lead to variation in the form of language, even in the speech of a single speaker. The bond between form and meaning in real language, then, is neither rigid nor random; it is direct enough to allow communication, but flexible enough to allow for creativity, variation, and change.

Creativity and recursion

As discussed above, any language is a highly structured symbolic system consisting of many interrelated parts. It is also a very human phenomenon, used by people every day in new and creative ways to accomplish an infinite number of communicative tasks. Let's discuss some examples of how people can creatively mold and shape their language in response to specific needs.

Lewis Carroll's famous poem *Jabberwocky* (1872) starts out with the following verse:

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

Even though many of the words in this verse are nonsense, in context we can infer a lot about the linguistic structure, and even develop a rough image of the scene being described. For example, we know that *brillig* probably refers to a time, because it is preceded by *'twas*. We also know that *toves* refers to something that can perform actions (probably persons or animals of some sort), because they *did gyre and gimble*, and these words obviously refer to actions. We also know that *wabe* must describe a place where *gyring* and *gimbling* may occur. *Slithy* and *mimsy* must be modifiers (**ADJECTIVES**) that describe properties of the *toves* and *borogoves* respectively.

The overall impression one gets from this verse is probably something like a forest setting involving strange, mythical creatures in some kind of special state or condition. We wait expectantly for the second verse to help fill in the gaps in our mental scene.

This example is from a famous author, but we don't even have to study great literature to see how language is used creatively to accomplish communicative work. Everyday conversation will easily suffice. For example, I recently heard the following sentence in an actual conversation:

(1) My dog just snerdled under the fence.

I don't find the word *snerdle* in any of my dictionaries. Yet, this sentence is immediately understandable, in the right context, to anyone who is a fluent speaker of English. We know *snerdle* must be a **VERB**, because it has a **SUBJECT** (*my dog*) and takes the **PAST TENSE** ending *-ed*. These are structural facts about this sentence. Because the sentence has these structural features, we can make a very good guess about what the function, i.e., the *meaning*, of the sentence might be. Because we know something about dogs and fences, and we know about verbs that start with *sn-* (*snort, sniff, sneeze, snore*, etc.), and verbs that end in a **PLOSIVE CONSONANT** plus *-le* (*wiggle*, *waddle*, *fiddle*, *jiggle*, *sidle*, *giggle*, etc.), we can

develop a very specific mental image based on this sentence. You may even say that the speaker provides a meaning for the verb *snerdle* by using it in exactly this context. It would be quite difficult to guess what this word "means" apart from its use in a specific communicative context. If this new verb fills a gap in the vocabulary of English, it may catch on to the point where it may even begin to appear in dictionaries. This kind of inventiveness characterizes every language on earth and is one way that new words are added to the vocabulary of any language.

Throughout this book we will see examples of how the forms of language arise in response to communicative needs. Here is one more important example. As mentioned above, there is an infinite number of ideas and **NUANCES** that people may care to express using language. However, the human mind is finite. It is not possible for one person to store or to learn an infinite amount of information. How is it, then, that a speaker of a human language can potentially express an infinite number of ideas, using a finite mind? Any system that is charged with this task must exhibit what linguists call **RECURSION**. In other words, any system that takes a limited input and produces an unlimited output must be able to combine elements in the input recursively – over and over again – with enough complexity that the appropriate infinite range of outputs is possible. Here is a simple example. Take a phrase like:

(2) The cat

We all know that there are many cats in the world. If I need to distinguish among them, I can "modify" this phrase:

(3) The cat in the hat

There are also many hats in the world. If I need to distinguish which hat I am talking about, I can modify the **NOUN** *hat* in the same way that I modified the noun *cat* earlier:

(4) The cat in the hat with a yellow ribbon

There are also many yellow ribbons in the world . . .

I think you can see where this is going. Since I can use a noun to modify another noun, I can potentially express an infinite number of ideas, starting with just a few basic words. The above examples illustrate **EMBEDDING**, which is just one of many respects in which all languages are recursive. Words, such as nouns, can be embedded within larger structures which can in turn be embedded within others, up to infinity. Any system that did not provide for such recursion would not qualify as a language. Why? Because it wouldn't be able to do the job of a language. So the forms of the language, in this case the way speakers construct noun phrases, are determined by the function, in this case, the need to express a potentially infinite number of ideas. Recursion is another respect in which every language is creative. It allows everyone who is a fluent speaker to formulate and

express an infinite number of ideas. The only limitations are the communicative needs and imagination of the speaker.

Grammar

What image comes to mind when you hear the word **GRAMMAR**? For many people this word brings back painful childhood memories involving lists of "do's" and "don'ts" in speech and writing: "never say 'ain't'," "never split an infinitive," "never say 'him and me'," etc.

To a linguist, the word "grammar" has a very different meaning. Grammar in the broadest linguistic sense is simply everything a person needs to know in order to be a fluent speaker of a language. For example, the way of forming a noun phrase discussed above is part of the grammar of English – it is something that all English speakers unconsciously "know." Sometimes the word **TACIT** is used to describe a person's linguistic knowledge (as well as other culturally conditioned behavioral patterns). What this means is that people are not normally aware of their internalized grammar. They can become aware of it, for example by taking a linguistics class. However, most people simply use their grammar without thinking about it, just as they use their tacit knowledge of other aspects of social behavior, like facial expressions, ways of eating, walking, expressing emotions, and many others. Grammar, to a linguist, is something to be discovered, described, and explained, rather than something to be invented and enforced. It includes a good portion (some would say all) of the mental habit patterns and categories that allow people in a community to communicate with one another. Grammar is internal to the human mind, but allows the mind to "connect" to other minds that have similar grammatical patterns.

Under the heading of "Grammar" there are traditionally several subheadings, including **PHONETICS**, **PHONOLOGY**, **MORPHOLOGY**, **SYNTAX**, and **SEMANTICS**. In the rest of this chapter we will discuss some of these subheadings.

Morphology and syntax

In this section we will briefly discuss how the subject matter of this book, sometimes referred to as **MORPHOSYNTAX**, relates to the other subheadings within the domain of Grammar.

Phonetics and phonology have to do with how the sounds of language are produced in the human vocal organs (lungs, larynx, mouth, nasal cavity), and how sounds are systematically organized in particular languages. Morphosyntax has to do with how these sounds combine to form words and sentences. Semantics has to do with the meanings of individual elements of linguistic structure and their combinations. **DISCOURSE ANALYSIS** is a term that describes the study of how

sentences combine to form conversations, stories, lectures, and other extended forms of speech.

Actually, the term "morphosyntax" is a hybrid word that comes from two other words – morphology and syntax. Since "morphosyntax" sounds better than "syntophology," the former is the word that linguists prefer to use.

Morphology is simply the study of shapes. For example, zoologists may study the morphology of camels – how their bodies are shaped. Different species of camels have different body shapes. Some have one hump and others have two. Morphology in linguistics has to do with how words are shaped, and how the shapes of words may be systematically adjusted in order to accomplish communicative tasks. You can also think of morphology as the study of how meaningful units combine to shape words.

Syntax, on the other hand, is how words combine to form sentences. One reason many linguists like to talk about morphology and syntax together is that sometimes a communicative job that is performed by word shapes (morphology) in one language is performed by combinations of words (syntax) in another. So if linguists want to compare different languages, it helps to be able to refer to "morphosyntax." For example, look carefully at the following sentences from Naga, a Tibeto-Burman language of Northern India, with their English equivalents:

(5) a. ngama ate hethoang 'I will teach him.'
I him will.teach

b. ate hethoang ngama 'I will teach him.'
c. atema nganang hethohang 'He will teach me.'
d. nganang hethohang atema 'He will teach me.'

In example 5a, meanings are given in English directly under the Naga words. In Naga, the main way in which a speaker communicates who is teaching and who is being taught is by the shapes of the words. In all of these sentences, the word that mentions the person who is teaching ends with *-ma*, no matter where this word appears in the sentence. It can appear at the beginning (examples 5a and 5c) or at the end (examples 5b and 5d). In all these sentences, the word that mentions the primary *actor* (in this case the person who teaches) ends in *-ma*. Therefore we say that the job of expressing who the actor is in a sentence is accomplished morphologically, i.e., by the shapes of words, in Naga.

In English the situation is quite different. In English, the way a speaker communicates who is acting and who is being acted upon is mostly word order. Consider these examples:

(6) a. Zarina taught Aileron.
b. Aileron taught Zarina.

These sentences do not mean the same thing, even though the shapes of all the words are identical. The difference in meaning is expressed only by the order

of the words. Therefore we say that the job of identifying the actor in English is accomplished *syntactically*.

The first part of this book (chapters 1 through 5) deals mostly with morphology. The second part (chapters 6 through 10) deals mostly with syntax. However, it should be kept in mind that these are not necessarily two completely distinct domains. Syntactic structure certainly affects morphology, and morphology is one very important way that syntactic structure is revealed. The main ideas to keep in mind to this point are:

- Language is a tool for communication; therefore structural similarities among unrelated languages can, in most cases, be attributed to common communicational functions.
- Languages can accomplish the same or similar communicative tasks by changing the shapes of words (morphologically) or by changing how words are arranged (syntactically).

Lexicon

So far we have described two subheadings within the general domain of Grammar in any language – the morphology and the syntax. We have seen that communicational jobs that are accomplished morphologically in one language can be accomplished syntactically in another. There is one other subheading that perhaps should be considered alongside these two. This is the **LEXICON**. Different linguistic theories have vastly different ideas of what constitutes the lexicon of a language. The characterization presented here is flexible enough to encompass most of the theoretical variation, while remaining true to a common understanding of what linguists mean when they talk about the lexicon of a language.

In the broadest sense, the lexicon of a language consists of a list of all the **UNITS** in that language. Units in the lexicon are **IDEALIZED** mental constructs, or images. They are not actual words, phrases, or sentences, but rather mental "pictures" that can be called up from memory when needed for the purpose of producing actual words, phrases, and sentences. Sometimes these pictures are referred to as "representations" or "templates." Such units are called **LEXICAL ENTRIES**. For example, *cat* is an entry in my internal mental lexicon of English. As such, it is no more than an idealized representation – a memory, so to speak, of a noise that has served a certain range of functions in previous conversations I have been involved in. Because I can depend on the probability that other English speakers share a similar memory, that representation is available in English conversations as the need arises. In the lexicon, however, it is no more than a potentiality, an abstract representation of the possibility of some specific linguistic behavior.

The lexical entry for a linguistic unit consists of a cluster (conceived sometimes as a list and sometimes as an image) of all its characteristics. The term "entry" is based on the metaphor of the lexicon as a dictionary. We talk about the "dictionary entry" of a word as consisting of information about its spelling, pronunciation, meanings, and usages. Lexical entries are something like that, except they are conceived of as unconscious mental pictures stored in individual speakers' minds, rather than in published books or computer disks.

In addition to whole words, like *cat*, parts of words can also be units in the lexicon. For example, the *-ed* part of a word like *walked* means **PAST TENSE**. This is part of what one has to know in order to know English, therefore *-ed* is in the lexicon of English. It may be more accurate to say that the *pattern* of a verb followed by *-ed* is in the lexicon of English. This may be represented in a formula as:

(7) VERB + -ed = $[\text{VERB}]_{\text{past tense}}$

In other words, it is not just any *-ed* that means "past tense," but only those instances of *-ed* that are attached to verbs. The formula in 7 is one way of representing on paper the unconscious pattern in the minds of all English speakers that allows them to express the past tense of many verbs.

In this broad notion of the lexicon, **SYNTACTIC STRUCTURES** may also be located there. Actual phrases and sentences are not part of the lexicon, but abstract, idealized patterns are. For example, 8 is a syntactic pattern of English:

(8) PREPOSITION + NOUN PHRASE

This pattern specifies that any member of a class of things called **PREPOSITIONS** and any member of a class of things called **NOUN PHRASES** can combine to form a unit. This idealized pattern gives rise to a whole range of possible linguistic structures in use, for example:

(9) a. in the house
b. under the bed
c. with a hammer
d. on the mat
e. down the rabbit hole
f. through the mystical forest inhabited by strange beings and fraught with unfathomable dangers, none of which were apparent to Alice when she first began following the White Rabbit

The phrases in 9 are not in the lexicon. Rather, they are composed of other elements that are in the lexicon. The pattern in 8 is one of those elements, under a broad view of the lexicon.