

How Language Works

Mike Gasser
Indiana University

1 Introduction

Students studying linguistics and other language sciences for the first time often have misconceptions about what they are about and what they can offer them. They may think that linguists are authorities on what is correct and what is incorrect in a given language. But linguistics is the *science* of language; it treats language and the ways people use it as phenomena to be studied much as a geologist treats the earth. Linguists want to figure out *how language works*. They are no more in the business of making value judgments about people's language than geologists are in the business of making value judgments about the behavior of the earth.

But language is a cultural phenomenon and we all have deep-seated, cultural ideas about what it is and how we ought to use it, so knowing where to begin in studying it scientifically is not a trivial matter at all. Issues arise that would not if we were geologists figuring out how to study earthquakes or the structure of the earth's crust. For this reason, before we dive into the study of language, we will need to examine some of the biases that we all have concerning language and to set some ground rules for how we are going to proceed. Because there is more than one way to begin, it will also be useful to establish a basic stance to guide us. Finally, because human language is an enormously complex subject, the book will focus on a narrow range of topics and themes; there will be no pretense of covering the field in anything like a complete fashion. This first chapter is designed to deal with these preliminary issues. But first, you will need to know about the various conventions that I will be using in the book.

1.1 The book: organization and conventions

Organization

The organization of this book is based on the idea that human language has a small set of basic properties, each of which plays a role in the workings of language as an instrument for communication and thought. Each chapter in the book (after this one) introduces a new property. Chapter 2 discusses words and word meaning. Chapter 3 discusses phonological categories, the units that are combined to make word forms. Chapter 4 discusses phonological processes, the ways in which the units of word form interact with one another. Chapter 5 discusses compositionality, the principle that allows complex meanings to be expressed by combinations of words. Chapter 6 discusses how words are organized into larger units and how these allow us to refer to states and events in the world. Chapter 7 discusses how the grammars of languages divide the world into abstract conceptual categories. Chapter 8 discusses the productivity and flexibility of language and how grammar makes this possible.

Languages

Most linguistics texts draw their examples from an unconstrained set of

languages. This has the disadvantage that students are left with little sense of how the different aspects of each language fit together. It also invites the kind of errors that may crop up when linguists rely on examples from a wide variety of other linguists. For these reasons, almost all of the examples in this book are limited to a set of nine languages. You can see the word for 'language' in each of these nine languages in the upper-left corner of the [Table of Contents](#) page, and, together with words in eight other languages, at the top of each page. If you're interested in knowing more about these languages, each is described briefly in [this appendix](#).

Other references

Throughout the book I will include links to other references that are available online. In particular I will often link to articles within the English edition of the collaborative encyclopedia, [Wikipedia](#). Wikipedia includes some articles that have not been written by knowledgeable people or have suffered from disagreements among the editors, but for important topics on language, the articles have been edited many times and have stabilized into relatively useful and reliable overviews of these topics. Because I will link to Wikipedia so often, I will use this special symbol with the Wikipedia "W" icon for these links: [W](#).

Conventions

General

Because this book is on the World-Wide Web and there is no paper version of it, you must have a Web browser [W](#) to read it. Be sure to use an up-to-date version of the browser software; otherwise, the pages will not display properly, and you may not be able to listen to the sound files. The best-known browsers that are usable across different platforms (in particular both computers running Windows and running Macintosh OS) are [Netscape](#) (use version 7 or later), [Firefox](#) (use version 1 or later), and [Opera](#) (use version 8 or later). For Windows users, another option is [Internet Explorer](#). For Macintosh users, another option is [Safari](#).

There are many links to sound files in the book. To see if your browser is set up properly for playing these files, click on this link: [this is a recording](#). There are also some links to movie files. If your browser is set up for these, you should see a picture of a woman below and should be able to play the movie by clicking on the controls below it. (The movie shows the sign in [American Sign Language](#) meaning 'movie'. For this and other ASL movies in the book, I am indebted to the [Communications Technology Laboratory](#) for their [Sign Language Browser](#).)

Terms usually appear highlighted like this. When important terms are introduced for the first time, they appear [like this](#). When such terms appear later in the book, there is often a link back to their first appearance. All of these important terms are also listed in the [glossary](#). [Concepts](#) are sometimes displayed LIKE THIS. Emphasized words appear *this way*.

The book is divided into chapters and sections, with one webpage for each

section. Sections are divided into subsections. Many subsections begin with an example and one or more questions to get you started thinking about the topic; these examples and questions appear in

boxes like this.

This version of the book (3.0) includes some discussion of computational issues, that is, how it might be possible to program a computer to handle some aspects of language. These parts assume basic knowledge of computation and computer programming. You can safely skip them if you prefer. Computational parts appear in

boxes like this.

Some sections may also contain less important portions that can be skipped. These appear in an indented chunk of text in a smaller font like this.

Most sections also include some comments on the text that appear in the margin on the left (or flush with the right in the printer-friendly version).

(something in the margin)

next to the part of the text that they refer to.

At the end of each chapter is a section containing problems on the material in the chapter. There is a link to the problems covering a given section at the bottom of that section page.

Linguistic examples

In the book, linguistic examples from languages other than English usually include a representation of the pronunciation of the word(s) and the meaning of the word(s). The meaning usually appears in the form of a **gloss**, that is, a word or brief phrase in English designed to give a general sense of what the expression means. Glosses appear between single quotes (' ').

For some linguistic examples the precise pronunciation is important; for others it isn't. When the pronunciation matters, the example is shown using phonetic symbols. A list of all of the symbols used in the book appears in this appendix. The symbols for sounds appear between slashes (/) or between angle brackets ([]); the difference between these two notational conventions will become clear in Chapter 3.

When precise pronunciation is not important, linguistic examples appear in *italics*. For languages like English that use the Roman alphabet, the standard orthography (spelling) is used. For languages that do not use the Roman alphabet (Chinese, Japanese, Amharic, and Inuktitut among our main group of nine languages), the examples are transliterated into the Roman alphabet. When an example consists of a complex word, a phrase, or a sentence in a language other than English, it will often appear in a standard three-line format. The first line is for the expression itself. In this line, words will sometimes be broken into constituent **morphemes**, that is, units of meaning, separated by hyphens (-). The second line is for the meanings of individual words and morphemes. Here hyphens separate the meanings of morphemes that are separated by hyphens on the first line. Meanings for **grammatical morphemes** appear in SMALL CAPITALS, often abbreviated. When more than one word is used to indicate the meaning of a single morpheme, these words are joined by a colon (:). The

third line is for a gloss for the whole expression, enclosed, as elsewhere, in single quotes. Here is a Spanish example.

Juan	y	Ana	habla-ba-n	con	el	piloto
Juan	and	Ana	talk-IMPF-3P:PL	with	the:MAS	pilot
'Juan and Ana were talking with the pilot.'						

In this example the word *hablaban* has been broken into three separate morphemes, and a meaning is given for each of these below the word in the second line. There are two aspects to the meaning of the third morpheme (-n), 3P (third person) and PL (plural), so these are joined by a colon. Don't worry if you don't understand what these morphemes are doing (or even what a morpheme is) at this point; all of this will be explained later.

English usage

The English used in this book is meant to be a relatively informal variety of standard written American English. As you will see in the [section on what linguists study](#) and the [section on what linguists do not study](#), there is sometimes disagreement about what counts as standard usage. One advantage of writing a book on the Web is that I get to decide the conventions myself rather than being forced to conform to the standards imposed by an editor. Because this is a book about language and because particular usages can sometimes lead to negative, kneejerk responses, it may be worth mentioning two features of the English used in this book that do not count as standard for everybody. First, I will often use *they*, *them*, and *their* to refer to an unspecified *singular person*^W, as has been common in English speech and writing since [at least the time of Shakespeare](#)^W. An example is the sentence *what language a child learns depends on what language **they** are exposed to*. Second, I will avoid the word *whom*^W altogether. This word is very infrequent in any modern English spoken dialect and is not used in a consistent fashion in formal writing either.

Readers unfortunate enough to have been taught rigidly traditional English teachers will also notice split infinitives^W (*to overwhelmingly reject this proposal*), prepositions at the ends of clauses (as in the example with "singular they" above), and sentences beginning with *and* or *but*.

1.2 What we study

A science^W takes phenomena of one kind or another as its subject matter and attempts to describe and explain them objectively. Scientists gather particular kinds of data,^W analyze them, and create theories that account for the data. Here are some sciences and informal descriptions of the phenomena they are concerned with.

Science	Subject matter
chemistry	how substances combine to form other substances
psychology	how individuals behave
sociology	how people behave in groups

cultural anthropology	how human cultures resemble and differ from each other
linguistics	how language works

How would linguists and other language scientists go about objectively describing and explaining **how language works**? What kind of data would they examine? How would they analyze the data? What would it mean to "account for" the data with a theory? We will look at these questions in this section.

Social science and objectivity

You're a female American anthropologist specializing in the study of family structure and the roles that different family members play. You're studying a remote ethnic group, and you discover that the people believe that baby girls need to be brought up more harshly than boys to prepare them for a harsh life and that fathers have no particular role to play in the bringing up of their children. How might it be difficult for you to be objective in your study of these beliefs and behaviors?

Of the sciences listed in the table above, all but chemistry are concerned with *human* behavior. For all of these sciences, the work of the scientist is complicated by the attitudes that we have toward the behavior; that is, it may be difficult to be objective in our study. It is probably cultural anthropologists who face this difficulty most often. The anthropologist described in the box above, having grown up among highly educated people in modern US society, almost certainly believes that boys and girls should be treated equally and that fathers should play an important role in the raising of their children. But she must somehow put these beliefs aside in her work. Instead of labeling the people's practices as "wrong", she must attempt to see how they fit in with the other practices in the society, whether she notices consistent patterns in their behavior. If she is to make any value judgments about the behaviors, they should be based on whether the behaviors contribute to the stability of the society, not on her own feelings about them. Her job is to describe and explain the society, not to judge it.

As another example, consider the case of another anthropologist specializing in the scientific study of religious beliefs and practices. To do this objectively, he has to put aside whatever beliefs he himself has on questions such as the existence of God and life after death. There is no way he can pretend not to have such beliefs; he just has to try to keep them from getting in the way. Note that this is not a problem for chemists; in their work they don't have to worry about their prior attitudes toward particular chemical reactions.

When we attempt to study human phenomena such as religion or family roles or language scientifically, we do not deny the inevitability or even the value of our attitudes toward the phenomena. It's just that *evaluating* the phenomena and trying to convince people to behave or believe in a particular way is not the business of scientists (at least not the main business); it's somebody else's. A priest or minister has a very different purpose from an anthropologist specializing in the study of religion (although the priest or minister can probably benefit from the insights of anthropologists who study religion). A family counselor who advises a husband to allow his wife more freedom is not doing science (although the

counselor can probably benefit from the insights of anthropologists who study the family).

It is important to come to grips with some of our preconceptions about language before we begin to approach language as the object of scientific study. The next section is about some of those preconceptions and where they come from.

Attitudes toward the speech of others

You overhear the following conversation.

1. A: *Did you hear those two girls talking? "He don't mean nothin'." "I seen it." "Me and him fought." Can't they learn to speak English?*
B: *I know what you mean. They're just lazy, if you ask me.*

What do you think of comments like this? Do speakers from some regions or speakers belonging to some social or ethnic groups tend to be lazier than others in their speech?

As you certainly know, people are quite conscious of how they differ from people from other regions, social groups, or ethnic groups. They notice differences in dress, in food, in patterns of social interaction, in which qualities are valued or attract attention. And it is natural to evaluate these features of other groups, to think of their dress as fashionable or weird, to think of their food as tasteless or gross, to think of their social behaviors as friendly or offensive. The same is true for language. People hear speech that differs from their own and they may find it sloppy, elegant, or monotonous. These impressions may also be associated with the languages of particular groups rather than (or in addition to) the people themselves: we may find a certain language more expressive, more logical, even more masculine. What's the source of these impressions? Are they accurate?

Differences between languages

Undeniably communities of people do tend to differ. To take an obvious example, food preparation is more important in some cultures than others; some cultures are famous the world over for their cuisine. For language, the differences are again obvious to anyone. It's not just that languages sound different. Some languages make distinctions — in sounds, in words, in grammar — that others don't; in fact most of this book is about just this topic. And people learning a second language often have trouble making the distinctions that aren't part of their first language. What we naturally notice, as speakers of a particular language, is what is "missing" in other languages and what kinds of mistakes second-language learners make in trying to speak our language. This may lead us, consciously or unconsciously, to think there is something deficient about the other language or even about the speakers of the other language. It is very difficult for us to see it from the other perspective, to see that we also fail to make distinctions that matter in the other language and have trouble making them when we try to learn that language.

For example, as speakers of English, we may be surprised to find that Japanese has no words corresponding to English *a* and *the*, words that are so basic to English we may almost take them for granted. And we may be struck by the errors that Japanese learners of English make in trying to master these words. Similarly, we are struck by the confusions Japanese learners may have in pronouncing English words with the sounds that we

write with / and r, a distinction not made in Japanese. But these same Japanese speakers may be surprised when they first learn that English has only one word for 'you' (Japanese has at least six possibilities) and struck by the tendency of English-speaking learners of Japanese to always use the same word for 'you'. And they are similarly struck by the difficulty English-speaking learners of Japanese have with distinctions in vowel length and pitch change, distinctions that don't exist in English.

In fact there is no evidence that people in some cultures speak in sloppier or more elegant or more monotonous ways than people in other cultures. And while languages do differ in striking ways, these different features seem to balance each other out. As far as we know, all languages are equally expressive, equally logical. If you're not already convinced of this, I hope you will be after you have read this book.

But the example in the box above doesn't concern two different languages; it concerns a single language, English, and its speakers. The fact is that there is also considerable variation within English (or any other major language); that is, English has dialects. I'll have a lot more to say about dialects in [the section on dialects and languages](#). For now, the main point to be made is that what linguists have learned about the essential equality of languages applies to dialects as well. Though it is often even harder for people to accept this fact for dialects than for languages, as far as anyone knows, there is nothing inherently inferior or superior about any dialect of any language.

Linguistic chauvinism and intolerance

So if impressions like those of the speakers in the box above have no basis in fact, where do they come from? There are three possibilities. First, these people may have been told by an authority, for example, an English teacher, that certain usages are just plain wrong. Clearly, the reasoning would go, anyone who knows this should not be using those forms. We'll return to this issue in [the next section](#). The fact is that what is "wrong" is all relative. The girls quoted in the conversation in the box would almost certainly find it wrong to say *it doesn't mean anything* when speaking to each other. If they wrote *it don't mean nothin'* in a school essay, on the other hand, that would be another matter (though it would still not be reason to call them lazy; it would just be evidence that they had not learned the rules of the variety of English that is appropriate in school). Second, these people may have a stereotype concerning the group in question, and they may be transferring that stereotype to the speech of that group. Third, what they hear differs from the English they speak, and people may be quite intolerant when it comes to speech. Especially if they belong (or believe they belong) to a political, economic, or intellectual elite, their view may be something like the following: "the way I speak the language is the right way; any other way is wrong".

Whatever the reason for the impressions of A and B in the box, a linguist would respond to them by saying that the two girls were simply speaking a different dialect of English, a dialect with its own grammar differing from the grammar of the dialect of A and B.

Describing and explaining language

You're a linguist who wants to study the English spoken in the Caribbean island nation of Grenada. How do you go about this? What sort of data do you gather?

Data for research on language

Linguists and other language scientists, unlike A and B in the box in the last section, are interested in what people do, not what somebody thinks they should do. The English of the girls overheard by A and B is just as legitimate an object of study as the speech of any other group. To carry on their study, clearly researchers need to gather examples of language. There are two sorts of ways to get these.

- By collecting naturally occurring language, either written texts or spoken language.

Linguists usually study spoken \overline{W} (or signed \overline{W}) language because it is more basic than written language. \overline{W} Most of the human languages that have existed have not been written at all, and among those that are written, many people do not read or write them. In addition, though language learning continues throughout life, most of the basic patterns of a language are probably mastered by the time a child is six years old. So the written form of the language has little or nothing to do with this fundamental early learning of language.

- By eliciting language by asking people particular questions or by doing experiments that call for language.

Linguists use both kinds of data. For example, once you'd arrived in Grenada, you might get permission to record phone conversations, then transcribe the conversations, perhaps using a special notation that shows the speakers' pronunciation. Or you might recruit one or more willing speakers to help you in your study by translating words or sentences from your English into theirs or by telling you whether certain sentences are possible in their English.

Experiments on language

Rather than using words and sentences produced by speakers (or writers), linguists and (even more often) other language scientists sometimes gather other kinds of data. For example, they might record the acoustic properties of speech or the movements of the tongue, lips, and jaw during speech. Or they might present people with words, sentences, pictures, or movies (the stimulus) and see how they respond or how long it takes them to respond to them. The responses in experiments like these could involve

- saying something, for example, the first word that comes to mind or a sentence that describes a visual stimulus.
- pressing a button, for example, to make a choice concerning the stimulus.

Given some data, a linguist or other language scientist has to do something with it. Much of this book will be concerned with what they might do, so what follows will be just an introduction to this topic. For every kind of research on language, there are two things to be considered: what aspect of language is being studied and what the research is supposed to accomplish (and how it does this). We'll start with the first and look at the second in the next section.

The content of research on language

Language, even a particular individual language, is far too complex a subject to be studied in its entirety by any one researcher. As already mentioned in the [overview of the book](#), there is a higher-level distinction we can make concerning what research is supposed to be about, between the

study of *language as system* and the study of *language behavior*. In either of these two cases, the language scientist is normally studying only some aspect of the phenomenon, one or more of four sorts of things about a language (or dialect): its sounds, its words, its grammar, and its use in context. A researcher interested in the sounds of the language (or dialect) might try to figure out what the basic sounds of the language are, how they combine to form words, or how speakers produce the sounds. Many researchers in this area believe that it is possible to study the sounds of a language more or less independently from the other aspects of the language.

A second area of research is the words of a language, usually thought of as organized in some sort of abstract dictionary, referred to as the **lexicon**. A researcher interested in words might study how speakers find words when they are formulating sentences or how abstract meanings build on simpler meanings (how is *over* in *get over a problem* related to *over* in *jump over the puddle*?).

Another possible kind of research would try to characterize what counts as a possible sentence in the language, that is, what's **grammatical** in the language. It is not as easy as it might seem to define this concept. We must be careful to avoid any bias on the speaker's part based on what they have heard from teachers of their language in school because what we care about is what people actually say, not what someone tells them they should say.

But we also cannot just treat any sentence that occurs as grammatical because people make **speech errors**. By "errors", we do not mean that they break rules that apply to dialects other than theirs (for example, by saying *ain't* or *he don't*). Instead we mean slips of the tongue, false starts, and hesitations. For example, the following example includes several speech errors.

2. *Well, I think you ... I mean, the ... the ... um, this isn't coming out right at all.*

People produce such "sentences" all the time, but they clearly also know that there is something wrong with them. That is, linguists probably do not want their descriptions of a language to include such sequences. So grammatical sentences are possible sentences that do not contain speech errors.

This is not to say that there is nothing interesting about speech errors. In fact, like human errors more generally, they can give us lots of insights about the underlying mechanisms. There is a whole community of researchers that take speech errors of one kind or another as the data they try to explain.

Degrees of grammaticality

But there is another complexity; grammaticality doesn't seem to be an all-or-none matter. That is, while some forms may be completely acceptable to all of the speakers of a given dialect all of the time and other forms may be completely unacceptable to all of the speakers all of the time, there may also be intermediate cases that are not so clear. For example, some English speakers use *a chalk* for a stick (or piece) of chalk; others would be less comfortable with this (though they would not find it as unacceptable as, say, *a clay* for a lump of clay). There may even be variation within a single speaker. For example, an English speaker may say *for my wife and me* on some occasions and *for my wife and I* on others.

Returning to the examples ridiculed by A and B in the box above, we see

that by the definition of grammaticality that linguists work with, such sentences as this one may be perfectly grammatical for the girl who said it.

3. *He don't mean nothin'.*

A complete description of the grammar of these girls (and the community of speakers that they belong to) would have to specify just what counts as a grammatical sentence (for example, sentence 3) and what doesn't (for example, *he don't nothin' mean* or *he don't meant nothin'*), possibly singling out areas of grammar where there is disagreement and variation among the speakers. This description would obviously have to say something about word order and about which forms can go with which other forms (*meant* is a perfectly good word in their dialect, but not following *don't*).

But some linguists are not satisfied with just describing the grammatical sentences because this says nothing about what those sentences are for. Instead these linguists are concerned with describing how meanings and functions of language relate to words and grammatical sentences (for more on this idea, see [this section](#)). So an account that includes sentence 3 above as a grammatical sentence (for some English speakers) doesn't help us understand how this sentence conveys information about some person familiar to the hearer (*he*) and about the speaker's belief about that person's intentions. This book follows this second position on what we should be describing, that is, that we should be saying how language accomplishes things for speakers and hearers.

Finally, those same linguists who are interested in how sentences convey meaning may also be interested in describing a fourth sort of aspect of the language, a sort of "correctness" that is different from grammaticality. A sentence can be grammatical and meaningful — that is, the words and grammatical patterns in the sentence can sound right and correctly describe some possible situation in the world — but the sentence can still be inappropriate. Consider the following sentence.

4. *There is no life on the moon.*

This sentence makes perfect sense and describes a true state of affairs. But if you walked up to a stranger on the street and said it, they'd think you were crazy. It would not be an appropriate way to begin (or end, for that matter) a conversation with a stranger. Just as speakers of a language have knowledge about what is grammatical in their language, they also have knowledge about what is **appropriate**.

Learners' errors

Of course not everyone who uses a language or dialect (see [this section](#) for the difference between dialects and languages) knows how to do so grammatically and appropriately in all situations. In particular, language learners have only imperfect knowledge of the language or dialect they are learning, and they can be expected to make errors. Children learning English as a first language may say "doos" for *juice* or *me up* when they want to be picked up. Teenaged speakers of English as a first language may still commit errors of appropriateness, using informal expressions such as *bigtime* in formal contexts. And adults learning English as a second language may say "diss" or "dees" for *this* or *I make the homework* for *I am doing the homework*. We consider these to be errors, but there are only errors from the perspective of the system defined by the behavior of the adult native speakers of the language or dialect that is being learned. Such examples can also be seen relative to the learner's own linguistic system, which has its own pronunciation, vocabulary, grammar, and patterns of

usage. Researchers studying second-language learning often find it useful to treat the learner's knowledge of the second language as a sort of language in its own right, what they call "interlanguage".

This section has looked at the kinds of topics that interest linguists and other language scientists and the kinds of data that might look for to help them in their research. But we haven't thought much about what the outcome of the research is. What would it mean to describe or explain language? We'll look at these aspects of research on language in the next section.

Problems

1.3 How we study language

Accounts, generalizations, and theories

In addition to content, we can look at research on language from the perspective of what it is trying to accomplish and how it does that. First, we need to go back and consider again what science is all about. As we've seen, a scientist starts with a phenomenon of interest, gathers some data on the phenomenon, and attempts to come up with a description or explanation of it. This may take the form of a discussion in some language (such as English or Chinese), a set of equations, or an algorithm, that is, a precise description of a set of processes to be carried (by a person or a computer). Whatever form it takes, this result of the scientist's research may be referred to as an "account" or an "analysis" of the phenomenon. A scientific account is expected to include some sort of generalization about the phenomenon, that is, to go beyond simply listing the data.

A generalization can be relatively specific, applying only to a very narrow range of data. For example, an anthropologist studying kinship relations in some ethnic group might conclude that the relation between a mother-in-law and daughter-in-law matters more than other in-law relations in the culture. However, most often, an anthropologist wants a particular culture to be seen only as an example of a more general phenomenon. That is, the goal is not just to describe or explain one culture but to say something general about human culture. So the anthropologist might want to state a generalization about in-law relations across all ethnic groups belonging to some type or to all ethnic groups. In the most general cases, the generalization is usually stated in terms of a particular theory, which is a general set of principles for understanding phenomena of a particular type. A theory is supposed to offer an *explanation* of the phenomena, not just a description. For example, kinship theories start with a set of basic categories that are supposed to be sufficient both for describing and explaining the role of the different possible kinship relations in all societies. A theory is like an "account" or an "analysis", only more general.

How scientists describe and explain: generalizations and theories

To take a linguistic example, a researcher might be interested in describing the way present-tense verbs are negated in English dialects like the one referred to by A and B in [the box in the last section](#):

- *I/you/he/she/it/we/they don't go*

The linguist could make the low-level generalization that for all but a few verbs in these dialects, the present tense negative is formed by putting *don't* before the base form of the verb (*go, mean, eat, etc.*). This generalization would apply only to the dialects being investigated.

Or the researcher could note that there is a distinction in the affirmative in these dialects that is not made in the negative, the same distinction that is made in standard English dialects:

- *I/you/we/they go*
- *he/she/it goes*

Explaining languages, explaining language

Going beyond these English dialects, the linguist might then discover that something similar happens in many languages, for example, among the languages that this book focuses on, in Japanese and Amharic, making the generalization that languages tend to make more distinctions in the affirmative than the negative. The researcher could then go even further and try to place negation within a set of other forms in terms of how likely they are to favor distinctions; for example, the researcher might discover that fewer distinctions are made in verbs in subordinate clauses^W (for example, in *the places they go*) than in independent clauses^W (for example, in *they go many places*) in many languages. (Don't worry if you don't understand what "subordinate", "independent", or "clause" means.) Even more abstractly, the linguist could try to explain why there would be such tendencies across many languages. For example, they might propose that it is more difficult to produce and understand negative forms than affirmative forms and that this puts pressure on the users of the language, and hence on the language itself, to compensate by making some other aspect of the grammar simpler. This kind of proposal would be a theory that is designed to describe and explain a set of grammatical phenomena across many languages.

To summarize this section so far, we see that a scientist trying to understand language makes generalizations about data. These generalizations can apply only to one language, or they can apply to language in general. A linguist or other language scientist often works within the context of a particular theory of language. I'll have more to say about the role of theories in this section. But we've said nothing so far about what the goal of a description or explanation is. There are two kinds of possibilities, related to the two fundamental ways of looking at language that are mentioned in the overview of the book: focus on the product (sounds, words, sentences, etc.) and focus on the process (speaking, reading, understanding, etc.).

Product and process

Consider this line from the English comedy show "The Two Ronnies":

1. *She left her husband for the garbageman.*

The joke, in case you missed it, is based on the ambiguity,^W that is, the multiple possible interpretations, of the verb *left*. The sentence could mean that the woman abandoned her husband in favor of the garbageman, or it could mean that she put him outside for the garbageman to carry away. It turns out that ambiguity is quite common in language and is the basis of puns^W such as this one. What sort of a problem does this present for a description or explanation of how language works? What aspects of language would an explanation of ambiguity need to refer to?

All instances of language are obviously the result of processes: speakers, writers, and signers (in the case of sign language) produce something that

we call language and hearers, readers, and sign observers attempt to understand something that has been produced. Over a longer time scale, what a person knows about a language (that is, how to produce and understand it) changes; this is the process that we call learning, development, or acquisition. Over an even longer time scale, every language also changes; the English of today is not the same as the English of 1900. This slower process is called language change. Finally, over the longest time scale that is relevant for language, we know that at some point in the distant past, for example, 200,000 years ago, the ancestors of modern humans did not have anything like what we call language; the process that resulted in the kind system we have now is called language evolution.

Most linguists choose to ignore all of these processes and focus instead on the products, the words, sentences, and entire discourses that are produced and understood by people at a particular point in historical time. In their research they attempt to describe and explain these products, to generalize about what must appear, what may appear, and what may not appear. Other language scientists, especially psycholinguists and computational linguists, focus instead on the processes themselves. In their research they attempt to describe and explain these processes, to generalize about what is going on during language behavior, language change, or language evolution, for example, when a speaker pronounces a word, when a child learns how to combine words into sentences, when a population of agents "invents" grammar in the process of evolving a communication system. Their accounts and theories often take the form of algorithms and are often implemented in the form of computer programs. They may be called processing accounts or computational models.

There are at least two important differences between these two ways of looking at language. First, the product-oriented perspective deals with *static* objects; even though it took time for the words or sentences to be produced or understood, the things being studied have no real time in them, except in that certain parts come before other parts. In the process-oriented perspective, on the other hand, time cannot be ignored. Processing happens in real time, and the implementations of processing accounts as computer programs obviously run in real time. These accounts differ a lot in terms of how serious they are about the time course of human language processing, but they are all in some sense *dynamic*.

Production and comprehension

Second, processing accounts are *directional*. At its most basic, language processing is either production or comprehension. Even processing accounts that are concerned with the slower processes of learning or evolution are based on some idea of how production or comprehension takes place. I'll have more to say about production and comprehension later in this chapter. For now the important idea is that these processes occur in opposite directions.

In production, a speaker, writer, or signer starts with something to be communicated (perhaps to themselves) and then goes through a process that results finally in an instance of language (a spoken, written, or signed utterance of some kind). When I produce the sentence *take the garbage out*, I start with something that isn't language at all, something that doesn't include the words *take* and *garbage* but is more like my mental representation of some situation in the world, either one I'm experiencing (seeing the garbage piling up in the house) or one I'm imagining (seeing the hearer taking the garbage out). Then somehow I get from this thought

to the utterance itself. In comprehension, a hearer, reader, or sign observer starts with an instance of language and then goes through a process that results in some sort of approximation to what the speaker, writer, or signer wanted to communicate. When I hear somebody say the sentence *take the garbage out*, I start with some sounds and from these eventually figure out what the speaker wants and what I'm supposed to do, and of course what I'm supposed to do isn't language at all; it's an action. The product-oriented perspective ignores this directionality, treating it as irrelevant to what makes language work.

So who is right? Does the product-oriented perspective or the process-oriented perspective give us more insight into how language works? To some extent, the answer depends on what we're after. If we want a way to describe languages in as efficient and understandable a way as possible, then it may be that we can confine ourselves to product-oriented research. The outcome of this research could be an archive of many languages in a form that would allow researchers from different fields and people who want to teach or learn the languages to consult the archive. Of course if our goal is to understand what people are doing when they are actually using language or if we want to write programs that allow computers to use language, then we will need to rely on process-oriented research.

But what if we are interested in the more abstract and theoretical question of *why* language is the way it is? Which perspective is the right one, or do we need both? There is a lot of controversy within the language sciences on this point, with one camp, associated especially with the famous linguist Noam Chomsky^W, claiming that we can learn what makes language special by studying products alone. The idea is that processing is something separate, something to be understood in its terms, but not something we need to refer to to understand how language works. The opposing camp, associated with the theories in linguistics known as cognitive linguistics^W and functionalist linguistics and with many language researchers in fields outside of linguistics proper, takes the view that the nature of language is intimately tied to the way it is used, that if we want to understand *why* language is the way it is, we need to refer to processing. As I discuss more in [this section](#), this book belongs more in the second camp. I will assume that both product-oriented and process-oriented perspectives can help us understand how language works.

Let's return to the issue of ambiguity, illustrated in the box above, to see how thinking about process as well as product can help us figure out what is going on. For the sentence in question, a product-oriented approach might simply include the information that *left* (or *leave*) has at least two meanings, though of course it would have to be more precise about what is meant by "have two meanings". A process-oriented approach would look at the processing of the sentence from both directions. This would make clear that ambiguity is a "problem", that is, a potential challenge for a person or a computer, in the comprehension direction but not necessarily in the production direction. It is such a problem in fact that a very large body of research from the processing perspective has looked at what is called disambiguation. The problem is that when a word is ambiguous, a listener (reader, sign observer) has to figure out which meaning is intended (or, in the case of a joke like the one in the box, to see that the sentence has a possible interpretation for each of the meanings). It is easy to imagine how a person hearing a sentence like the one in the box would know that the word *left* has more than one meaning. What is hard to explain is how the person knows which meaning is the right one, or, in this case knows that both are right. Theories of disambiguation are designed to explain the

process. We'll return to disambiguation later; the important point for now is that it only becomes an issue when we look at language from a processing perspective.

In this section and the last, we've seen that the business of linguists and other language scientists (and the subject of this book) is trying to describe and possibly explain particular languages, language in general, and how people use languages. But this is not the only way we might treat language. Some people who are not scientists and may not be particularly interested in the scientific study of language do talk about language as a part of their work. And one of their concerns may be deciding what people should and should not say and trying to enforce these decisions. This kind of work, some of its consequences, and how linguists sometimes comment on it are discussed in the next section.

1.4 What we don't do: prescribing and evaluating language

Standards

You're in a creative writing class discussing short story writing, and you say

1. *You shouldn't introduce a new character unless they have an important part to play in the story.*

The instructor of the course corrects you: "'He or she has', not 'they have'." How do you feel about being corrected in this way? Do you think the instructor was justified?

Just as it is normal to evaluate the religious beliefs and practices and the family behavior of particular individuals within our culture, it is also normal to evaluate the linguistic behavior of people. We may treat some speech and writing patterns as acceptable or unacceptable, superior or inferior, appropriate or inappropriate. Attitudes such as these must be based on a standard, some idea of what counts as desirable behavior. People who are concerned with defining and maintaining linguistic standards are **prescribing** (rather than describing) language. Usually they prescribe aspects of grammar, and in this role they are referred to as prescriptive grammarians. There are three kinds of questions we can ask of linguistic standards and language prescription.

1. What is the purpose of the standard? What use is it put to?
2. Where does the standard come from?
3. How is the standard enforced?

Why a standard

One reason for a linguistic standard is that people within a community (for example, a nation) will be better able to understand one another if they agree on a set of words and a set of rules for pronunciation, spelling, and grammar. The process of defining the rules (and sometimes the set of words as well) is called **standardization**. Once a standard has been agreed on, it can be used in the media and taught in the schools.

Japan provides a good example of this process of standardization and promulgation of the standard. Children all over Japan, speaking widely divergent dialects of Japanese, are all taught to speak and write Standard

Japanese in school, and radio and television announcers are all expected to be familiar with the standard vocabulary, pronunciation, and grammar. If these announcers were to speak in the southern dialect of Kagoshima or the northern dialect of Aomori, people in most of Japan would have difficulty understanding them. Something like this happens in the United States, though more informally. Children across the country are taught the same written standard vocabulary and grammar, though they may not be taught a single standard pronunciation, and radio and television announcers tend to speak in a single standard.

A related, though often unstated, goal of a standard may be to eliminate diversity, which may be perceived as a threat to national unity. Sometimes the diversity is reflected in different, though closely related, dialects, but sometimes it is reflected in completely different languages. It is one thing to teach a standard in the schools across a country. It is another to discourage people from speaking their local dialects or languages (for more on dialects and languages, see [the next section](#)). Sometimes the native speakers of the non-standard dialects or languages become willing participants in this process in their desire to be integrated into the society.

In Japan, for example, standardization sometimes had the effect of eliminating diversity. In the case of the Ryukyu Islands, the southernmost region of Japan, the dialects that are spoken there are so different from the standard that they may be considered a separate language (or languages). Today, following the repressive policies of school administrators in the first half of the twentieth century as well as the economic and social pressures for Ryukyans to conform, this language is on the verge of extinction. This is a familiar phenomenon in many countries with minority populations. In the US and Canada, immigrant and American Indian students have often been prevented from speaking their home languages in schools. These educational policies were most notorious in the case of American Indian students, who were sometimes sent to schools which kept them separated from their communities and were punished for speaking in their native languages. These policies were one of the major factors leading to the extinction and near extinction of the majority of North American Indian languages. For more on the topic of [language death](#) and "endangered languages", that is, languages in danger of dying, see the websites of [Terralingua](#) and [the Foundation for Endangered Languages](#), two organizations dedicated to linguistic diversity.

Finally a (never stated and perhaps often unconscious) purpose of a standard may be to exclude certain groups from power. If the standard is based on the speech of one group, either from a particular region or a particular class, then this gives people in that group an advantage when it comes to jobs and ultimately power.

Standard languages and politics

Clearly language standardization is a political issue, and as such it is not really the business of linguists (though it is studied by sociologists interested in the social and political aspects of language). However, as linguists often become closely involved with the people whose languages they study, they may become advocates for these groups when their languages or their well-being are threatened because of the language policies of governments.

Other reasons are often stated for prescribing language, though these may mask the ultimate political reasons. These reasons include the supposed illogical or ambiguous nature of some constructions used by people, and linguists have also sometimes gotten involved in the debate in these cases

because of their expertise. An example is the use of English *they*, *them*, and *their* to refer to a single person, usually of unknown or unspecified gender, as in the example in the box above.^W The idea is that since *they* is supposed to be plural, it should not be allowed to refer to one person. In cases like this, prescriptive grammarians are trying to actually improve the language or perhaps to preserve what they suppose to be an earlier, purer form of the language. The problem is that linguists and other language scientists, experts on the "logic" and the degree of ambiguity in language, can find nothing inherently wrong with any of these constructions, at least not with the usual examples of "bad" English constructions. In fact languages seem to have their own built-in "prescriptive" mechanism which weeds out whatever patterns don't work. It's a survival-of-the-fittest sort of arrangement, and it implies that those common patterns such as "singular *they*" (which has been around in the spoken and written language for at least 500 years) are quite fit linguistically.

Where standards come from

If we are to define a set of standards for a community, they have to come from somewhere. One possibility for the source is from other languages, and, strange though this may seem, standardizers have sometimes resorted to it. Take the "split infinitive",^W which prescriptive English grammarians sometimes argue against. An English infinitive is an expression consisting of *to* followed by a verb stem (don't worry now if you don't know what this is), for example, *to go* or *to sing*. So according to this rule, it is incorrect to say or write *to boldly go* or *to not sing*. The main argument seems to have been based on the fact that Latin and German and other familiar European languages cannot split their infinitives. But at least nowadays it seems just silly to try to force one language into the patterns of another.

Only slightly less extreme is the attempt to impose patterns from earlier in the history of the language onto present-day speakers. An example of this is the distinction between the first consonant in *which* and the first consonant in *witch*. For most speakers of American English and the English of England, there is no distinction, but there used to be (and there still is in some dialects, for example in Irish English and Southern (US) English).^W There are still school teachers in the United States who tell their students that it is "wrong" to pronounce *which* like *witch* (my daughter had such a teacher). As we will see at various points throughout this book, language is always changing, and there is little we can do to stop this. In any case, if the loss of the distinction between these two consonants interfered with comprehension, it wouldn't have happened.

It ain't about good and bad language; it's about class and education

If the standards come instead from among the speakers of the modern language, the question is which speakers. Language usage varies both with region and with social group, so if we want to standardize, we need to choose a region or group. As noted above, there are sometimes political considerations involved. In many countries, standard usage comes from the speech and writing of well-educated or upper-class people from a particular region, though the regional origin of the standard is not so clear in the case of the United States. For example, the current English standard prefers *you aren't* to *you ain't* because the latter form, although common in regional dialects all over the English-speaking world, is associated with uneducated speakers.

Actually this argument only holds today. The history is more complicated. *Ain't* was once used by upper-class speakers, but prescriptive grammarians who felt

that it was somehow "lazy" or "illogical" succeeded in mostly obliterating it from upper-class speech in the English-speaking world. See [this short history of *ain't*](#) from the *American Heritage Dictionary of the English Language* (2000).

Another example is the pronunciation of pairs of words like *pin/pen* and *since/sense* as the same. ^W This is a feature of the speech of a large region of the United States, including the South and some areas bordering on it, but it would not be considered a feature of standard American English.

How standards are enforced

Given reasons for creating a standard and agreement on what constitutes the standard, there is the question of how the standard patterns are to be spread through the population. The obvious venue for this is the schools, and in many countries considerable attention is devoted to making sure students are familiar with the standard language of their community. Much of this involves simply exposing students to examples of (mostly written) texts in the standard, and this often has the desired effect, at least in the writing of the students. Sometimes the teaching of the standard involves attempts to prevent or undo frequent non-standard patterns of usage. Here we are sometimes dealing with standards that don't correspond well to the usage of *any* native speakers of the language, including educated adults. For example, teachers may try to get their children to stop using "singular *they*". Not surprisingly, efforts like this are mostly futile; children are likely to find it impossible to change their grammar in a way that doesn't match what they hear around them.

Trying to teach grammar to native speakers

Although their efforts rarely seem to have an effect on children's speaking, in a limited way they may affect the children's written language. This is at least possible when the "rule" amounts to the prohibition of a particular form, for example, *ain't*. When it is more complicated, the prescribing may backfire, leading to behavior that is not what the prescribers had in mind. An interesting example is the use of "subjective" pronouns (*I, he, we, etc.*) vs. "objective" pronouns (*me, him, us, etc.*) in English. In an attempt to prevent usages such as *him and me are friends*, teachers have created a situation in which many speakers now say *with he and I*, instead of *with him and me*, not at all in accordance with the original intention of the teachers. This is an example of hypercorrection ^W, which occurs when a prescriptive rule is applied in too many cases. For more on the issue of hypercorrection with English pronouns, see [this interview](#) with English professor Jack Lynch, which explains the usage but also takes a prescriptivist approach.

In summary, we've seen two sorts of usages that people try to enforce, usually attempting to prevent an "incorrect" form. In one situation, there is a form perceived as incorrect, such as *ain't*, that is characteristic of speakers of some dialects but not normally part of the speech or writing of speakers of the standard dialect. In another sort of situation, there is a form perceived as incorrect, such as "singular *they*", which *is* used by many or most of the speakers of the standard dialect as well as by speakers of other dialects. In this second sort of situation, there is often disagreement about what should count as the standard, and it is this kind of case where linguists sometimes become involved because of their interest and expertise in what people actually say or write. (My advice is not to get involved in one of these arguments. For some reason it often seems hard for people, at least for English speakers, to be rational about what counts as "correct" or "incorrect".)

I have focused on English. For a brief history of English "usage", that is, the concern for what is right and what is wrong in standard English, see [this](#)

passage by E. Ward Gilman from the 1989 edition of *Webster's Dictionary of English Usage*. Not surprisingly, there are similar discussions and debates concerning usage in some other languages, though this sort of concern does not appear to be at all universal.

Other evaluation of language

So far we have looked at attempts to enforce linguistic standards on people. But people, especially composition and creative writing teachers, are also involved in evaluations of language based on other kinds of criteria. One is appropriateness, discussed in the last section. Appropriateness is one aspect of language that we continue to learn as adolescents and adults, and it is often explicitly taught. For example, a student might be corrected for using a word or phrase perceived as too casual in certain contexts, say, *be into* in the sense of 'be interested in' in an essay. Or an employee might be corrected by a colleague for calling his boss "dude".

Another important kind of evaluation concerns the effectiveness of language. A usage may be grammatical and appropriate to the context but still not accomplish the speaker's or writer's goal. For example, language may be needlessly ambiguous, as in the following example.

2. *Clark broke open the coconut with the screwdriver and then put it back in his pocket.*

A composition teacher might criticize the student's use of *it* in this sentence; does *it* refer to the coconut or the screwdriver? Or a use of language may fail to accomplish a higher-level goal of the speaker or writer. Consider the following example at the beginning of an argument.

3. *Income tax should be abolished. For one thing, people could use the money they pay as taxes for food or housing.*

A teacher might find the writer's point trivial and obvious and suggest leaving it out.

These are all legitimate reasons for evaluating language of course. While linguists and other language scientists are not in the business of evaluating people's language, they can be of help by studying what appropriateness is, what makes an expression interpretable by a hearer or reader, and how the parts of a text relate to one another.

Problems

1.5 Dialects and languages

Idiolects and dialects


Two Americans are talking about a couple they have just met.

- *She sounded English to me, but he doesn't seem to have any accent at all.*

Two English people are talking about the same couple.

- *He sounded American to me, but she doesn't seem to have any accent at all.*

What's going on here? Who has the accent?

What I know about my language and how to use it is called my idiolect.  It

almost certainly varies in minor ways from the idiolects of all other speakers. But what *is* an idiolect? That is, what kinds of things do I know? In one sense, this whole book is an answer to that question, but we need to have a first cut at the answer here to help us get started.

- I know words. I have a vocabulary, a set of words which I know how to pronounce and use appropriately. For example, I know how to say the word *apple*, I know that it refers to a particular type of fruit, I come up with this word when I want to refer to a particular apple, and I understand it when I hear it.
- I know how to pronounce words and combinations of words more generally. That is, there are aspects of pronunciation that go beyond individual words. For example, I know to pronounce the ending that we spell *-ed* like /t/ in words like *picked* and *watched* but to pronounce it like /d/ in words like *signed* and *burned*.
- I know how to put words together into sentences in meaningful ways. For example, I know that if I want to ask when a particular train leaves I can say *when does the train leave?*, but not *when leaves the train?*.
- I know how to use language appropriately to achieve my goals. I know that if I want a friend to lend me \$100, it is better to say *I was wondering if you could lend me some money* than to say *give me \$100*.

I'll be much more careful later on about how each of these types of knowledge is described, but for now I'll say (informally) that my idiolect involves knowledge about vocabulary, pronunciation, grammar, and usage.

Of course no one is really interested in describing idiolects. Linguists and other language scientists study the speech of communities of people, not of individuals. More specifically, they study the knowledge of vocabulary, pronunciation, grammar, and usage that is *shared* by the members of a speech community. Because the members of the community agree on this knowledge, because it differs (at least in some ways) from the knowledge shared by other communities, and because it is mostly arbitrary, I will refer to the knowledge as **linguistic conventions**.^W

But what is a speech community? I will use this term to refer to any group of people that shares a set of linguistic conventions differing in some noticeable way from the conventions found elsewhere. You may know that in the United States people in some cities have some characteristic features in their pronunciation, although they are easily understood by people elsewhere in the United States. For example, people native to Pittsburgh are known for using *you uns* (or *yinz*) to mean 'you plural'. Here's an example from the (partly tongue-in-cheek) "Pittsburghese" website: *if yinz wants served, raise your hands*. The number of conventions that distinguish Pittsburghers from other English speakers in the northeastern United States is actually pretty small, but because there is such a set of conventions, we can consider these people to be a speech community. The speech patterns, that is, conventions of vocabulary, pronunciation, grammar, and usage, of a speech community are called a **dialect**,^W so we can speak of a "Pittsburgh dialect".^W

Note that a dialect may not be defined entirely on the basis of its physical location. Cities often contain a variety of ethnic and social groups with different speech patterns. For example, the African-American population of many US cities (for example, Pittsburgh) often has a quite different dialect from the Euro-American population of the same cities.

Which dialect do you speak? There may be a number of possible answers.

What about larger communities? Pittsburghers share some speech conventions with speakers in other cities of the northeast and north midwest, for example, their pronunciation of the *a* in a word like *hands*, as in the example above (more on this pronunciation [later on](#)). And people in that larger region share some conventions with people in an even larger region encompassing speakers in most of the northern and western United States, for example, their pronunciation of the long English vowels (*bite*, *beat*, *bait*, *boat*, etc.). And people in that even larger region share many conventions with English speakers all over North America, including most of their grammar and usage conventions, as well as a number of pronunciation conventions, for example, the tendency to pronounce the words *latter* and *ladder* in roughly the same way.

This idea of larger and larger communities, each sharing fewer and fewer conventions, is an over-simplification in one sense. The fact is that the boundaries of the communities overlap in many ways. If we look at particular vocabulary, we may find a region with one boundary, whereas if we look at other vocabulary or at some pronunciation convention, we may find another boundary. For example, Pittsburghers tend to say *pop* (as opposed to *soda* or some other word) for carbonated drinks, and they share this convention with many speakers in the northern midwestern cities who also share their pronunciation of the vowel in *hands*, but *not* with speakers to the east of them, in New York City, for example, who share the pronunciation but not the word. (New Yorkers tend to say *soda* rather than *pop*.) Thus where we draw the boundaries around a dialect depends on which convention or set of conventions we're looking at. For more about *soda* vs. *pop*, see [this interesting website](#).

Another way what I've said so far is an over-simplification is that there is great variation within any of these regions. Some of this variation has to do with the constant contact between dialects that is a fact of life in most communities. Some of the variation also has to do with the fact that people often know a range of ways to say things and they may sometimes avoid their local dialect in favor of a standard (see below) in certain situations.

Each of these shared sets of conventions, whether at the level of a small village, a subculture within a city, or a larger region, is a dialect. And a linguist can be interested in describing any level and any aspect of the dialect at any level (pronunciation, vocabulary, grammar, usage). The pronunciation associated with a dialect is called an **accent**.^W

Languages

What is a language? How would you tell someone (say, an alien with no knowledge of human culture) what English is, without using the word *language*?

We can of course extend the boundaries in our example even further, beyond North America to include England, Scotland, Wales, Ireland, Australia, New Zealand, a number of Caribbean countries, and communities within many other countries. This large speech "community" is not really a community in the usual sense of the word, but it does share many conventions. For example, in all of these places, speakers make a question from a sentence like *he ate potatoes* by inserting the word *did* and changing the form of the verb *ate*: *did he eat potatoes?*, and of course speakers in all of these places share the word *potato* for referring a class of tuberous vegetables. The conventions of this large "community" are of course what we refer to as "English", which we consider a **language**.^W Thus in one sense a language is a set of dialects. In another sense it is (like a dialect) a set of conventions shared by a speech community.

Two dialects of one language or two separate languages?

But how do we decide when a collection of dialects is a language and not just another, more general dialect? As we've already seen, a dialect can also be a set of dialects (the North American English dialect consists of Southern dialect, New England dialect, Canadian dialect, etc.). What makes English a language and not just another very general dialect? What makes Canadian English a dialect of English and not a language in its own right?

The answer to this question is complicated. In fact there is no clear answer because the words *dialect* and *language* are used in different ways for different purposes. There are two completely different kinds of criteria related to the distinction between dialect and language, linguistic criteria and social or political criteria.

Linguistic criteria

Given two overlapping sets of linguistic conventions associated with two different speech communities, for example, Mexican Spanish and Argentine Spanish, how do we decide whether they should count as two dialects or two separate languages? One criterion is the degree of overlap: how similar are the vocabulary, the pronunciation, the grammar, and the usage? Unfortunately there's no simple way to measure this overlap, at least no way that researchers would agree on. One way to have a sense of the overlap, though, is **mutual intelligibility**, the extent to which speakers from the two or more speech communities can understand each other. Mutual intelligibility is not easy to measure, and it is often based on the impressions of speakers and hearers, how much they understand when they encounter members of the other group or how long it takes them to get accustomed to the speech of the other group. We also need to establish some sort of intelligibility threshold; no two speakers can be expected to understand each other all of the time. So none of this is precise at all. The idea is simply that if two sets of linguistic conventions are similar enough so that their speakers can usually understand each other, then the two sets of conventions should count as dialects of the same language rather than separate languages. On these grounds, we call Mexican Spanish^W and Argentine Spanish dialects of the same language (Spanish^W) because speakers of these dialects normally have little trouble understanding each other.

To find out what should count as a separate language on grounds of mutual intelligibility, a good resource is Ethnologue, an online database of all of the world's known languages, 6,912 according to their current listing. The Ethnologue compilers attempt to use mutual intelligibility to decide what should count as a language. While English is listed as a single language, both German and Italian are listed as multiple languages. Each of these languages, for example, the variety of Italian called Sicilian, is usually referred to as a "dialect", but, according to the *Ethnologue* compilers, these are distinct enough to be considered separate languages. Again, the criterion of mutual intelligibility is a rough one, and some of *Ethnologue's* claims are controversial.

^W

Social and political criteria

Another sort of criterion for what counts as a dialect is the social or political unity of the group in question. In Bavaria, a state in southern Germany, and in parts of Austria most people speak a dialect called Bavarian or Austro-Bavarian,^W which on grounds of mutual intelligibility could be considered a language distinct from the speech of Germans and Austrians

in other regions. Ethnologue calls Bavarian a language. But Bavarian is clearly closely *related* to those other dialects and not more closely related to dialects of some other language, and so for mainly political reasons, it is convenient to consider it a dialect of the German "language", rather than a language in its own right. Something similar can be said about the speaking conventions of the older generation in the Ryukyu Islands in southern Japan (because these dialects are dying out, most young people do not speak them). On the basis of mutual intelligibility, we could divide the island dialects into several separate languages, each distinct from the Japanese language (as is done in Ethnologue and in the Wikipedia article on these languages). But the Ryukyu Islands are politically part of Japan, and these dialects are clearly *related* to Japanese and not related at all to any other known language (unless we consider each of them to be a language). So for political reasons, it is convenient to consider them dialects of Japanese, just as the dialect of Osaka is considered a dialect of Japanese.

Mutually intelligible "languages"

At the other extreme are examples like the languages spoken in the northern European countries Sweden, Norway, Denmark, and Iceland. These "languages" are all related to one another, and speakers from some pairs of countries within these have little difficulty understanding one another when they are speaking the standard dialects of their languages, despite the obvious differences, especially in pronunciation. Thus on grounds of mutual intelligibility, we might consider some of these "languages" to be dialects of a single language. But Icelandic, Swedish, Danish, and Norwegian are official "languages" of separate countries, and there are separate spelling conventions for some of the sounds in the languages.

Actually the situation is even more complex than this because Norway has two official dialects, and a fifth related language, Faroese, is spoken in the Faroe Islands, which are administered by Denmark.

So for mainly political reasons, they are considered separate languages rather than dialects of a single language.

To summarize, the line between dialects of one language and separate languages is somewhat arbitrary. However, wherever we draw the line, three points should be clear.

- Every language has multiple dialects.
- Every speaker of every language is also a speaker of at least one dialect of that language.
- Since the pronunciation conventions of a dialect constitute an accent, every speaker of every language speaks with some accent. There is no such thing as "speaking without an accent".

Standard dialects

The following appears on the website of a person who spent some time in Pittsburgh: "probably relating to the rest of Pittsburgh's terrible dialect, which I, fortunately, did not pick up". Why would some dialects be thought of as "terrible"?

Some dialects within a language may be singled out for special status. When we're dealing with a political unit, such as a nation, in which related dialects are spoken by most people, one dialect is often treated as the

standard dialect.^W You know something about this already from [the last section of the book](#). The standard dialect is often the only dialect that is written, and it is the one that is taught in schools and (with some exceptions) used in the media. Thus in Germany, Austria, and the German-speaking part of Switzerland, it is Standard German^W that is taught in the schools and used in broadcasting, even though most people in this region are not native speakers of the Standard German dialect. This means that most people in the German-speaking countries end up bidialectal. The same situation holds in Japan, where it is Standard Japanese, based on Tokyo dialect, that is taught in the schools and used in the media. Note how this makes it possible to speak of a German or Japanese speech community, even when the native dialects of people in these communities are very different from one another, because all educated speakers in these communities share the standard dialect, often as a second dialect.

So what do we mean when we say "German" or "Japanese"? There are two possibilities. "German" could mean Standard German, that is, one of the set of dialects spoken in Germany and also the basis of written German. Or it could mean the collection of related dialects, some mutually unintelligible, which are spoken in Germany and other countries where Standard German is the official language (Austria) or one of the official languages (Switzerland). When linguists refer to "German" or "Japanese", without specifying the dialect, they normally mean the standard dialect.

Dialects of English in the US and England

In the United States, the situation is somewhat simpler than in Germany or Japan because the differences among most of the dialects are not nearly as great; native speakers of English in the United States have little trouble understanding each other. (An important exception is African-American Vernacular English (AAVE),^W spoken mainly by many African-Americans.) As in Germany and Japan, we have an (informal) standard dialect, for vocabulary, grammar, and usage, if not for pronunciation. Thus children in Pittsburgh learn in school to write sentences like *the school needs to be renovated* rather than *the school needs renovated*, which would be grammatical in their local dialect. Americans tend to be relatively tolerant of differences in *accent*, however. Teachers in schools throughout the country teach the standard grammar but use their own local pronunciation. If we have a standard accent, it is the one people associate with television announcers, the accent characteristic of much of the Midwest and the West. This accent is called General American;^W I will have more to say about it [later](#).

The situation in England is similar to that in the United States, and the standard vocabulary, grammar, and usage that children learn to write in English schools are very similar to the American standard. However, in England, there is a stronger idea of a standard accent than in the United States and more pressure for children to learn this accent if it differs from their home accent. This accent is referred to as Received Pronunciation (RP);^W it is based on the speech of educated speakers in southern England. (Note that RP is standard *English* English pronunciation, not *British* English; in Scotland, there is a quite different standard accent.^W) I'll have more to say about RP and how it differs from General American and other English accents in [this section](#).

"Just" a "dialect" or a full-blown "language"

The existence of a single standard dialect among a set of non-standard

dialects has important social implications. The non-standard dialects have less prestige, and their use may be discouraged in formal situations, not just situations in which writing is called for. Sometimes, as in the Ryukyu Islands in Japan or in some regions of France and Spain, this leads to the decline and possible death of the non-standard "dialects" (which would be considered languages by the mutual intelligibility criterion). In other situations, speakers of non-standard dialects retain pride in their local speech patterns, while recognizing that they are not appropriate in certain situations. Finally, this pride, along with other cultural differences separating the speakers of the non-standard dialect from the speakers of other dialects (non-standard or standard), may lead to pressure to have non-standard dialects given official status, especially if they differ significantly from the standard. At this point the words *dialect* and *language* become politically charged terms because the supporters of official status for the non-standard dialect may feel the need to argue that it is not "just" a dialect of the larger language but rather a language in its own right. This has happened in the United States with AAVE ([here](#) is an essay on this topic by the sociolinguist John Rickford) and in Europe with many languages that are normally considered "dialects" of other languages ([this website](#) includes many of them as well as links to other sites concerned with the "minority language" question in Europe and elsewhere).

Language families

Why some languages resemble each other

We've seen how as we extend the boundaries of speech communities, we get fewer and fewer shared conventions. When we reach the level of a language such as English, Spanish, or Mandarin Chinese, we have a speech community which shares a set of conventions (in some cases a standard dialect) which allows people in the community to communicate with one another despite dialect differences. But we can go beyond a language. So for English, we could extend the boundaries to include the Netherlands, Germany, Scandinavia, and some other regions in western Europe. We'd now find a much smaller set of shared conventions. All speakers in this large "community", for example, share a word meaning 'all' which is similar in pronunciation to the English word *all*. But there would be no reason to call this set of conventions a "language" since the speakers obviously do not understand each other and do not belong to a single political unit with a single standard dialect. Instead we refer to this set of conventions, or set of languages, as a **language family**.^W In this case, the Germanic languages.^W The members of a language family resemble each other because they are genetically related; that is, historically they derived from a common ancestor language. (Note that this use of the word *genetic* differs somewhat from its use in biology; the speakers of Germanic languages are not necessarily genetically closer to one another than they are to the speakers of other languages.) The ancestor of the modern Germanic languages was not a written language, so we can only infer what it was like.

In most cases we can go even further back; the ancestor languages of two or more families themselves may have had a common ancestor language. Thus the modern Romance languages,^W including Spanish, French, Italian, Portuguese, Catalan, and Romanian; the modern Germanic languages; and many other languages spoken today in Europe, the Middle East, and South Asia, apparently descended from a much older (and also unwritten) language. This means we can group all of these languages into a single family, in this case the one we call Indo-European.^W Sometimes, to distinguish the lower from the higher levels within a family tree of

languages, we use "language family" only for the largest grouping (for example, Indo-European) and "branch" to refer to groupings within this (for example, Germanic and Romance). Note that there may be many intermediate levels in the family tree of languages. Within Germanic, for example, there is North, including the Scandinavian languages, and West, including English, Dutch, and German.

Note also that languages may resemble each other in one way or another for reasons other than a genetic relationship. The main non-genetic source of similarity is language contact; when the speech communities for two language are in close cultural contact, their languages often influence one another. So modern Japanese vocabulary includes thousands of words borrowed from Chinese and uses the Chinese writing system (as well as writing systems specific to Japanese). But, except in the sense that all human languages may be ultimately related to one another, there is no evidence that Japanese is genetically related to Chinese. A more complicated situation occurred in Western Asia with the complicated cultural influences among people speaking Arabic, Persian, and Turkish. These three languages belong to separate language families (Afro-Asiatic, Indo-European, and Altaic, respectively), which are either unrelated to one another or only very distantly related, but Turkish and Persian have borrowed many words from Arabic, Turkish has also borrowed many words from Persian, and Persian borrowed its writing system from Arabic.

1.6 A stand and two themes

A stand

Linguistics is a relatively young science, and psycholinguistics and computational linguistics are even younger, so it's not surprising that these fields are still fraught with controversy. Though there is general agreement on some core topics, some quite basic issues are still up in the air. It is not the place of an introductory textbook to go into all of the controversies; I will only try to do this for one topic, language learning, perhaps the most hotly debated of all. But it is impossible to discuss any topic related to language without taking some sort of stand on the questions that divide language scientists. All linguistics textbooks take such a stand, though, sadly, few of them tell you that they are doing this. The stand taken in this book is that language and language behavior are not phenomena separate from the rest of human perceiving, acting, and reasoning; that we can only understand how language works by understanding how it fits into the rest of human behavior. Another way to think of this position is that we will be treating the language sciences as cognitive sciences.

This position is related to two of the ways of looking at and studying language that we've already discussed. As noted in the [overview of the book](#), we can study language as a system independent of the people who use it, or we can study language behavior, focusing on use. As noted in the section on how language is studied, we can be concerned with [product or process](#). The stand that this book takes implies that it cannot be enough to study language as a system in its own right (though it might be useful some of the time to do this) or to study product rather than process (though again it may be useful some of the time). It also implies that the border between language and non-language is not necessarily a clear one and not something we should spend much time worrying about. So for every topic covered in this book, I will try to look at it both ways, to see how the system/product perspective and behavior/process perspective both contribute to an understanding of how language works.

I didn't invent this position of course. A number of other people have defended one or another aspect of it. On the need to consider cognitive processes outside of language in order to understand language, cognitive linguists like Ronald Langacker^W and George Lakoff^W have made this case in many of their writings, in which they invoke general psychological notions such as attention, categorization, and memory. On the need to treat language as process, an very convincing argument was made in 1983 by the computer scientist Terry Winograd^W in his important book *Language as a Cognitive Process* (1983).

In the rest of this section, I discuss two themes that are consistent with this stand and that will guide the rest of the book.

Two themes

Meaning, function, and form

The Main Theme of the book is that language associates **meaning** and **function** with **form** and that understanding how language works requires that we focus on these associations. All three of these terms will become clearer later in this book. For now we will just try to get the gist of what's intended by them.

Linguistic form concerns the way a linguistic expression sounds (for spoken language) or looks (for written or signed language) or how a linguistic expression is produced. There are two ways of looking at form. The usual way within linguistics is to think of form as a sequence of elements. For written language this is a natural way to treat form because written sentences *are* sequences of elements (characters). For spoken or signed language, it is less so. Treating a stretch of spoken or signed language as a sequence of elements means in effect focusing on the written transcription of the language in terms of units of sound or primitive gestures of some sort. This approach tends to play down the role of time, to look at a stretch of language as a static object. An alternative and more radical way of looking at form (at least for spoken or signed language) is to treat a stretch of language as something inherently continuous and dynamic, something that cannot (or should not) be transcribed as a written sequence of elements. In this book, I will usually discuss linguistic form in the first way, mostly because this is the way it is usually done and because it seems to simplify the analysis. But I will also insert periodic warnings about the biases that are built into this approach. One particular bias we should be aware of is one based on literacy; that is, as skilled users of alphabets, we may tend to think of spoken and signed language in this way (for an argument of this sort, see [this paper by linguist Robert F. Port](#)).

I will use meaning to refer to what language is *about*, the concepts that words and linguistic patterns refer to. By **concept** I will mean a unit of cognitive experience, a way people have of abstracting over their experiences in the world. For example, in a baby's experience the same face keeps appearing over it, and the baby abstracts over these different occurrences of face-appearing the concept of DADDY. In this book labels for concepts will often appear LIKE THIS. If the nature of form in the study of language is controversial, the nature of meaning is even more so. Some people deny that there is such a thing as meaning or that talking about meaning is helpful or that there is a consistent way to define "meaning" or "concept" or "about." I will try to take these positions seriously in this book, but the fact is that I don't know how to even get started without looking at the pole of language that is opposed to form.

Concepts without words

Whatever we mean by concepts, it is clear that not all of them are associated with words. Words are linguistic; concepts need not be. We all have concepts that we have no words or grammatical patterns for. For example, one concept I have is the little depressed region bounded by two vertical ridges that is found between the nose and the middle of the upper lip. I have no idea what to call that place (though I'm pretty sure it has a name in at least some languages); that concept does not represent the meaning of a word for me.

Another point to note here is that meanings do not need to be seen as things that are "out there" in the world. Since the stand taken in this book is that language is basically a cognitive phenomenon, this would be a strange way to think about meaning. Instead I will be treating meanings as things that are "in here", cognitive entities realized in our brains, depending on our *interpretation* of what's out there and including imagined entities that aren't "out there" at all (though they are inevitably based on things that are).

I will use function in two senses, first, for the uses that people put language to and second, for the uses that particular words or patterns have within stretches of language. It is the former sense that concerns us here. People use language to refer, to assert, to command, to convince, to get information, to entertain, to deceive, and much more, and these uses of language obviously have something to do with the forms that they choose. (If you want somebody to lend you their computer, you don't say *you will lend me your computer*, you say something like *could I borrow your computer?*.)

Many people who study language make a distinction between function in this sense (part of what they call "pragmatics"**W**) and meaning (what they may call "semantics"**W**). I would argue instead that the difference is a matter of degree, but that there's not much point in worrying about where the line between pragmatics and semantics is.

Constraints

A second theme of the book is that language is the way it is in part because of **constraints** coming from the nature of human biology and human cognition. A constraint is a kind of limitation on what is possible. Consider constraints that come from the nature of the human body itself. For spoken language, the physical and physiological properties of the vocal tract constrain what can be produced, and the auditory system constrains what can be perceived. For sign language, the physical and physiological properties of the hands, arms, upper body, and face constrain what can be produced, and the visual system constrains what can be perceived.

Just as important are cognitive constraints, constraints arising from the nature of the human mind. First, there are limitations on human memory. Cognitive scientists divide human memory into short-term**W** and long-term memory.**W** Short-term memory is used to temporarily store the information that is needed as we interpret the world around us. When we are faced with a complicated visual scene, we can't look at everything at once, and we have to scan the scene in order to come up with some kind of interpretation. But we need to temporarily store some sort of record of what we've already seen in the scene as we do this. That information is kept in short-term memory. Likewise, when we are listening to a sentence, we need to remember the parts that we've already heard as we are trying figure out the meaning of the whole. Short-term memory is used for this as well. The same applies to reading and to the visual processing of sign language. Cognitive science research has shown that our short-term

memory has a very restricted capacity, and this places strong constraints on how sentences and discourses are organized.

We use our long-term memory to store information that we learn through experience. Information may remain in long-term memory for a very long time, even indefinitely, and human long-term memory is very large, larger than the memory of any current computer. But long-term memory is still limited. We obviously do not have an infinite memory capacity, and just because we have stored an item in memory doesn't mean we can retrieve it when we need it. We will see later how the finiteness of long-term memory matters for how language works.

Since languages have to be learned, limitations on human learning^W are obviously also relevant. This is perhaps the most complex and controversial topic of all. Everyone at least agrees that language must be learnable because children obviously learn it. That is, it must be possible to figure out the forms, the functions, the meanings, and the associations between them on the basis of the examples of language that are available to young children. It turns out that coming up with an explanation for how this might happen has proven to be very challenging though.

Speaker orientation and Hearer orientation

Finally there are constraints that are specific to the two ways in which language is processed, language production, that is, speaking, writing, and signing, and language comprehension, that is, listening, reading, and interpreting linguistic signs. To simplify matters, I will be referring to a language producer as the **Speaker**, even though writers and signers as well as speakers proper are intended. And I will be referring to a language comprehender as the **Hearer**, even though readers and sign interpreters as well as hearers proper are included. (To remind ourselves of these distinctions, I will capitalize Speaker and Hearer when they are used in this way.)

For the Speaker, the main constraint is that the production of linguistic forms be easy. It is easier for the Speaker to make relatively few distinctions because the Speaker has to remember what the distinctions are and to make the extra effort to keep things distinct. For example, maintaining the agreement between subject and verb in the present tense in English (*the girl **sings**, the girls **sing***) requires an effort on the part of the Speaker. What is easy for the Speaker depends in turn on constraints from the body. For example, large movements of the tongue tip are more difficult to execute than short movements, so short movements should be preferred from the Speaker's perspective.

For the Hearer, the main constraint is that linguistic forms that need to be distinguished can be easily distinguished in comprehension. This constraint also depends on the body, specifically the parts of the nervous system that are responsible for hearing (for spoken language) and vision (for written and signed language). For example, if a language contains a large number of homophones, that is, different words which sound the same (such as *two, too, and to* in English), this may put a burden on the Hearer.

Speaker-oriented constraints and Hearer-oriented constraints often oppose each other; what simplifies things for the Speaker (for example, not making many distinctions) complicates things for the Hearer. We will see many examples in the book of these opposing tendencies.

The opposition of Speaker orientation and Hearer orientation is particularly clear as languages change. Languages change for a variety of reasons —

contact with other languages, imperfect learning by children, random fluctuation — but it appears that all languages are always changing. Most of the changes can be seen as either Speaker-oriented or Hearer-oriented. For example, the grammar of a language may become simplified as some suffixes are dropped, a change that seems to result in less work for the Speaker. But the two kinds of trends always balance each other out in the end, and the simplification of the grammar in one way will probably be compensated for by an increase in complexity (from the perspective of the Speaker) somewhere else in the grammar of the language. Otherwise language would fail as a communicative device. These built-in pressures in favor of the Speaker and the Hearer apparently prevent the world's languages from moving in some general overall direction. That is, at least in recent history, it does not appear that languages have generally been getting simpler or more complicated (in any sense of these words) as they evolve.

Problems

1.7 Why study language

What good is the scientific study of language? Why does anyone do it? Why should you care about it? These are the sorts of questions you have a right to ask about any university course. The answer to the last question depends a lot, of course, on how you happened to end up in a course using this book in the first place and on what your interests and long-term goals are. Language is a part of everyone's life, but it is more central to some people than to others. But I happen to believe that a scientific look at language should be a part of the basic curriculum, like mathematics and history are.

Second-language learning

Many of you have already studied one or more languages other than your first, and more of you will later on. A few of you may teach a foreign language. In either case, you are not likely to find the learning process an easy one. Some of the difficulties faced by second-language learners have to do with differences between their first and second languages, differences in pronunciation, vocabulary, grammar, and usage. You do not have to know linguistics to learn a second language; after all, people all over the world who have never heard of linguistics do this successfully all the time. However, knowing what pronunciation, vocabulary, grammar, and usage are; how they can differ between languages; and how people seem to learn them (as first- or second-language learners) can help you be aware of and understand your problems and possibly correct some of them. A second-language teacher needs to be able to focus on problem areas, for example, by giving lots of practice or by simplifying other aspects of the language being learned. It is difficult, if not impossible, both to understand the source of the problem and to come up with ways of addressing it without understanding the nature of the material being learned, that is, what linguists and other language scientists study.

First-language learning

Some students, when beginning a linguistics course, believe it will help them improve their knowledge of their first language. In fact this is *not* something you can expect from a linguistics course. You already know the great majority of the words, the grammatical patterns, and the usage conventions that you will need to survive. Of course you can improve; you can learn new words and expressions, become more proficient with the

grammatical patterns that are part of formal language, and get better at using language to accomplish your goals. Most of the improvement should come naturally as you are exposed to the complex language of academia and the workplace. But you will also face evaluations of your language by other people — teachers, colleagues, supervisors, even family and friends — throughout your life. You will be in a better position to make use of this criticism and advice if you understand what sort of problem is involved (if there really is a problem) and how it fits into the larger scheme of things. This is where linguistics can help.

Another way in which most people encounter first-language learning is in raising children. Of course you don't have to know linguistics to know how to "teach" a baby a language; babies aren't really taught language anyway. But knowing what it is that babies learn when they learn a language can make the process more enjoyable. You'll be able to better appreciate what an amazing process learning a language is, why so many people are fascinated with how this process happens, how your baby does the same kinds of things that others do and, at the same time, how your baby's learning steps are unique.

You may end up in a job that involves first-language learning more directly, as a teacher of your first language to native speakers. In this capacity, part of your job will probably be to make sure your students are competent in the standard dialect of the language. As we saw in [the section on what linguists study](#) and [the section on prescription](#), sorting out what belongs to the standard is not a trivial matter, and a knowledge of what linguistic conventions are seems essential. As I've already said, there is a lot of confusion and controversy about what should be emphasized or even taught at all. You will also need to teach your students what counts as appropriate and effective language. Again linguistics and other language sciences can help; some language scientists devote their efforts to figuring out what makes different expressions appropriate in different situations, while others are concerned with how words and expressions are interpreted by hearers and readers.

Or you might work as a speech therapist, dealing with people with speech disabilities of one kind or another. Here the relevance of the scientific study of language is obvious; you first need to know what the norms of a language are before you can hope to address the ways in which your clients or patients deviate from these norms.

Cognition

Language is probably the best window we have on the workings of the human mind. Language gives us the extraordinary ability to describe the contents of our thoughts, an ability that no other animal has. Of course there are many unconscious aspects to cognition that we cannot talk about, but these properties are apparently also reflected in what we say. The units of language — elements of form, words, grammatical patterns, conventions of usage — are in some sense also units of cognition. The implication is that the study of what all languages share is also the study of what it is to be human, something that is certainly an important topic for any educated person.

Linguistics and tolerance

What we can learn about the human mind by studying how languages *differ* from one another is more controversial. Linguists and other cognitive scientists disagree on how deeply the nature of a person's first language influences how the person thinks and views the world around them. I'll

have more to say on this topic at various points throughout the book. But language is such an important part of our lives that learning about the languages of other people, including how those languages differ from ours, is in a very real sense learning about those people. As with any other aspect of culture, lack of knowledge can lead to intolerance. It is easy to believe that other languages are inferior in one way or another to ours, to think that some languages, especially the languages of relatively small ethnic groups in the Third World, are more primitive than others. So the realization that a language like Tzeltal, Lingala, Amharic, or Inuktitut has a set of grammatical categories and communicative options not even found in English is an eye-opening experience. Looking more closely at languages, and in particular at languages that might seem exotic to us, can make us more tolerant.

Finally, language is what we use to influence the beliefs of one another. This happens in everyday conversation, as we argue about who forgot to put the mayonnaise jar back in the refrigerator or what the results of the latest election mean. It happens in advertising, as companies do their best to get us to buy their toothpaste, shoes, and cars. It happens in education, as educators provide us with what they say are truths and try to convince us that knowing how to speak Spanish or how to do a t-test will be useful to us later in life. It happens in politics, as politicians and political activists try to get us to vote for them or to support their program. None of this is new, but the enormous quantity of information that is now available to most of us *is* new, and most of this information is designed to change our beliefs in some way. Obviously an educated person needs to be able to navigate their way through all of this, to sort out the nonsense, to see how bias and ideology are behind what is being claimed, to be a critical reader and listener, to make informed decisions. I don't believe that any of this is possible without understanding the role that language plays in knowledge, belief, and persuasion.

1.8 Problems

1.8.1 Describing and explaining, Prescribing and evaluating

1.8.1.1

Note: This problem set assumes some basic knowledge of English dialects, so if you are not a native speaker of English, you will probably want to collaborate with someone who is.

Each of the following includes two ways to say (at least roughly) the same thing. For at least some people, the first way (A) could be seen as "wrong" in some sense. We saw in these two sections of the book that there are various ways in which a particular linguistic form can be thought of as "wrong" (though not necessarily by linguists).

- a. The use is not grammatical or otherwise acceptable (in the case of the pronunciation of a word) in any dialect of the language.
- b. The use is grammatical or acceptable in some non-standard dialects, but not grammatical in the standard dialect.
- c. The use is grammatical or otherwise acceptable for most speakers of the standard dialect, but some speakers may find it "wrong" for one reason or another.
- d. The use is grammatical or acceptable but not appropriate in the context.
- e. The use is grammatical and appropriate but not effective.

For each of the following, decide which type of "mistake" is involved in A (as compared to B), and explain your answer in a sentence or two. There may not be a single right answer for some of the problems.

1. A: *nuclear* pronounced noo-kyuh-lur
B: *nuclear* pronounced noo-klee-ur

This pronunciation is used by many speakers of standard American English (including President George W. Bush, though it could be argued that his accent is not standard), but some speakers feel that it is not standard, probably because it does not reflect the relationship between the word and the noun *nucleus*.

2. A: *yellow* pronounced yell-uh
B: *yellow* pronounced yell-oh

This pronunciation is acceptable in some non-standard accents but not in standard American English.

3. A: *Mr. President, what you're saying sucks.*
B: *Mr. President, I can't accept what you're saying.*

This use of *sucks* is acceptable in informal American English, but, like slang more generally, it is inappropriate in a formal context like this.

4. A: *I come back last night.*
B: *I came back last night.*

This use of *come* as the past tense of *come* is grammatical in some non-standard English dialects but not in standard English.

5. A: *Next shake the test tube that you dissolved the salt in.*
B: *Next shake the test tube in which you dissolved the salt.*

This use of a preposition like *in* at the end of a clause is perfectly grammatical in standard English, but there are a small number of diehards that prefer the use in B. B probably sounds very formal to most speakers of standard English.

6. A: *Whoever it belongs to should claim it.*
B: *Whomever it belongs to should claim it.*

Both A and B are used by speakers of standard English, though A is probably more common, and it follows the "rule" that is taught in schools for the use of *who* and *whom* because *whoever* is the subject of *should claim*. You sometimes see uses like in B in standard English, though.

7. A: *He left glass water on table.*
B: *He left a glass of water on the table.*

A is not grammatical in any dialect of English. It could have been produced only by a learner of English, either a child learning English as a first language or someone learning English as a second language.

8. A: *When I exited the bed this morning, my hair looked like rabbit ears.*
(an eight-year-old speaking to his mother; this is a real example)
B: *When I got out of bed this morning, my hair looked like rabbit ears.*

This use of *exit* is grammatical but not appropriate in this relatively informal context.

9. A: *John ordered his meal from the waiter, and he brought it to him right away.*
B: *John ordered his meal from the waiter, and the waiter brought it to him right away.*

In A the word *he* could refer to either John or the waiter, though in the context most people would realize that it refers to the waiter. Some

people might argue that *he* here is not as effective as *the waiter*, as in B.

10. A: *How did y'all like the party?*
B: *How did you like the party?*

In some non-standard dialects of American English, *y'all* means 'you plural'. It is perfectly grammatical in those dialects but not in the standard dialect.

11. A: *I would be thankful if it could be let known to me whether taking a student under your fold and tutelage for a long and hopefully eventful partnership would be acceptable to you.* (a real example from an e-mail to me)
B: *Please let me know if you can take on a new student.*

The writer of A is using grammatical English, but the elaborate and indirect language is neither appropriate (in any context?) nor an effective way to convey what they want.

1.8.1.2

You are a linguist studying the grammar of a particular dialect of English. You have recorded a stretch of speech between two men who are speakers of the dialect, and now you are taking examples from the recording to use as data. Included in the recording is the following statement by one of the speakers.

- *You don't even realizing it.*

You are trying to decide whether your account of the grammar of this dialect should include this sentence. Unfortunately you no longer have access to the two speakers, so you can't ask them questions. For each of the following, say whether it would be a relevant thing for you to do in order to help make your decision. Explain each answer.

1. You consult a book written by another linguist about the grammar of a closely related dialect.

Since you can expect the grammatical conventions of the two dialects to be similar and the description by the linguist of the grammar of the other dialect to be concerned with what people really say (as opposed to what somebody says they should say), the book is relevant to your decision. In particular if you discover that a form like this is grammatical in the other dialect, then there's a good chance it is in this dialect too.

2. You listen for similar sentences in the rest of your recording.

This is relevant to your decision because if you find similar sentences, for example, *I don't believing it*, then the original sentence was probably not a speech error.

3. You ask more educated speakers from the same region as the two men whether they could say something like this.

This is not relevant to your decision. Dialects often vary with education, so if the form is ungrammatical for more educated speakers (likely since it's ungrammatical in the standard dialects of English that you're familiar with), this doesn't tell you whether it's grammatical for your speakers.

4. You listen to what immediately follows the example on the recording to see if the speaker corrected himself.

This is relevant to your decision because a correction would be clear evidence that the sentence was a speech error and hence not an example of a grammatical convention that you need to account for in

your grammar.

1.8.2 How we study language

In this section, we saw that ambiguity is an example of a phenomenon in language that can best be appreciated from the perspective of comprehension. Consider the English words that are written *bank*, meaning roughly 'side of a river,' 'kind of financial institution,' and 'tilt while making a turn' (there are other meanings as well). Because all of these are written (and pronounced) the same, a reader (or listener) has to disambiguate the form when it appears. But this can be done in different ways, using different kinds of information. For each of the following sentences, say what kind of information a reader could use to disambiguate the word *bank*, and rank the three for how difficult they would probably be for a computer program. Imagine that the person or program knows only the three meanings of *bank* given above.

1. *I don't have much my money in my account at the **bank**.*

If our reader or program had a list of other words associated with each meaning of *bank*, the list for the 'financial institution' would probably include *money* and *account*. So if we just look for the meaning that is most likely given then other words in the sentence, we'd select the right one in this case. This sentence wouldn't require a very deep understanding to disambiguate *bank*.

2. *Before landing, the plane will have to **bank** to the right.*

You may have noticed that two of the meanings are noun meanings and one (this one) a verb meaning. The phrase *will have to* must be followed by a verb, so only that meaning is compatible with the context. Since this appears to be an absolute constraint (neither of the other two meanings could conceivably fit here), the disambiguation in this case could be seen as easier than in the first sentence.

3. *When I climbed out of the boat, all of my money fell onto the **bank**.*

Like the first sentence, this one uses of the noun meanings. If we try to disambiguate on the basis of co-occurrences, *boat* might lead to one of the meanings (apparently the correct one), while *money* might lead to the other. So it seems that in this case, something more sophisticated is required, some deeper "understanding" about how money (that is, actual physical money: coins and bills) can only fall onto a physical surface like a river bank and not, for example, onto an institution. That is, this is the most difficult of the three sentences to disambiguate.

1.8.3 Two themes

For each of the following facts about human language, say whether it derives from **Speaker (or Learner) orientation** or **Hearer orientation**, that is, whether it makes things easier for the Speaker, the Hearer, or the Learner. Explain your answer in a sentence. There may be more than one possible answer.

1. *Fos* is a more likely word in a human language than *zkt* is.

This is a Speaker-oriented fact. It is simply easier to pronounce the sequence *fos* than the sequence *zkt* because of the vocal organs (especially the tongue) work.

2. Languages tend not to have many homophones, that is, words that sound the same but have completely different meanings such as *bear* meaning the animal and *bear* meaning 'tolerate'.

Homophones are difficult to understand so the rarity is a Hearer-oriented fact.

3. Languages do not have specific names for every object; instead they have names for categories of objects.

Specific names for every possible object would be far too many to remember; this is Learner-oriented. (It could also be Speaker- or Hearer-oriented because even if all of the words could be learned, it could be hard to retrieve them from memory during language production or understanding.)

4. In sign languages, signs are normally not made with unrelated movements of the two hands, for example, with one hand moving up and down while the other moves sideways.

It's very difficult to coordinate unrelated movements of the hands; this is Speaker-oriented.

5. Apparently all human languages have a way of distinguishing statements from questions.

Hearers need to be able to distinguish between statements and questions to understand what Speakers intend; this is a Hearer-oriented property of language.

6. Some languages have only three vowels. If so, these vowels tend to be pronounced quite differently from one another, usually similar to the vowels in *hot*, *heat*, and *hoot*.

Hearers have to be able to distinguish the vowels in the language in order to distinguish the words, so the more different they are from each other, the easier it will be for the Hearer. This is a Hearer-oriented property of language.

| [contact author](#) | [table of contents](#) | [printer-unfriendly version](#) |

© 2006. Indiana University and Michael Gasser.

Permission is granted to copy, distribute and/or modify this document under the terms of the [GNU Free Documentation License, Version 1.2](#) or any later version published by the [Free Software Foundation](#).

URL: www.indiana.edu/~hlw/Introduction/pf.html
Edition 3.0; 2006-08-25