a diagnosis. For example, it's quite intriguing that the group of autistic children that participated in the Surian et al. (1996) study showed an impaired understanding of Gricean conversational norms, and yet, the adults with ASD in the Pijnacker et al. (2009) study showed no difficulties with scalar implicatures. This contrasting set of results seems to point to the conclusion that an understanding of conversational norms isn't necessary to understand implicatures. But we can't be sure that Pijnacker's subjects, who showed an intact understanding of scalar implicature, were *also* impaired in their understanding of conversational norms, as this sample of subjects was a completely different one from the group that was tested by Surian and colleagues. We'd be able to draw a more confident conclusion if the Pijnacker subjects themselves had been tested on their grasp of conversational norms, and showed some difficulties.

Another (and related) concern is that mind-reading deficits may not be the only cause of the symptoms that are associated with ASD. There's evidence that difficulties with cognitive control may also play a role. This complicates our understanding of any differences in behavior between typical and ASD populations; do the differences reveal that mind reading is an important component of that particular linguistic behavior? Or does that behavior rely critically on good cognitive control? Or both? To tease this apart, it's important to compare the results of ASD populations with those of other groups who have difficulties with cognitive control, but who don't show the mind-reading deficits that come with ASD (a group of individuals with ADHD might be an appropriate comparison).

Finally, what does it mean if we test typical and ASD populations and find no differences in their performance on a specific task, as we've seen here in the case of scalar implicature and pronoun production? Does this mean that typical subjects don't engage the mind-reading system in interpreting scalar expressions or uttering pronouns? Not necessarily. It could be that the two groups are achieving the same linguistic task with different underlying machinery, and more fine-grained experimentation might be needed to pick up on the differences. For example, in Chapter 5, we saw some evidence that in learning new words, typical children often look for direct evidence of the speaker's intent to refer to a specific object; autistic kids can learn new words too, but seem to rely more on general associative learning than on trying to figure out what the speaker meant in using that word. Competing explanations could amount to more than just providing alternative explanations for the same phenomena—it could well be that both factors sometimes actually jointly contribute to explaining a phenomenon. And, it's entirely possible that in some of the disagreements between researchers who advocate for highly social mechanisms, versus those who argue for largely egocentric ones, both sides may be right—it just depends on who you're testing.

PROJECT

The last section described research with ASD populations dealing with conversational implicature and pronouns—both areas in which there are questions about the extent to which mind-reading capabilities need to be involved. Now consider another aspect of language for which these same issues might arise (either taken from discussions in this chapter, or from your own ideas and observations). Propose and design a research program that includes testing of participants diagnosed with ASD. In designing your study, take into consideration the various methodological tools and concerns that have been addressed in this chapter. Pay special attention to considerations that come up in comparing different populations, as discussed in "Digging Deeper."

12 Language Diversity

Throughout this book, I've based discussions about psycholinguistics almost entirely on English, with only the occasional dip into other languages. All along, however, I've assumed that experiments with English speakers tell us something about how *language* works in the mind, not just how *English* works. This is an assumption that's widely shared by language researchers. Since humans seem to be able to learn any language that they're immersed in as children (rather than being genetically programmed for, say, only Germanic languages), we can reasonably conclude that the basic cognitive machinery for language is the same for all humans, regardless of the language environment they happen to be born into.

Still, it's worth noting that English represents just one grain of sand on a vast beach of linguistic possibilities. There are about 7,000 living languages spoken around the world today. To put this into perspective, researcher Mark Pagel (2000) reminds us that we humans—members of a single mammalian species—speak more different languages than there are species of fellow mammals. And the number of current languages reflects just the bare remnants of an even greater *past* diversity of languages. Pagel suggests that linguistic diversity may have peaked about 10,000 years ago, just before the development and spread of agriculture. Until then, there would likely have been many languages spoken by small, geographically isolated social groups. The invention of farming caused populations to fan out into larger geographic areas, with the likely result that local indigenous languages were often replaced by the language spoken by the spreading agricultural society. (This would be similar to how languages

LANGUAGE AT LARGE 12.1

The great language extinction

Raise your hand if either of your parents is a native speaker of a language other than English. Raise your other hand if English is your only language. If both your hands are in the air, you may be one small example illustrating the large-scale language loss that typically afflicts families who immigrate to an English-speaking country from elsewhere. In the United States, for example, reality runs contrary to anti-immigrant rhetoric: far from becoming isolated in linguistic ghettos and failing to learn English, most immigrant families generally *lose* their heritage language within a couple of generations, as reported by Robert Lane Greene (2011). Even among Mexican immigrants, currently the slowest group in the United States to shed their ancestral language, fewer than 10% of fourth-generation immigrants speak Spanish very fluently. As Greene points out, who needs further *disincentives* to speak the heritage language when the economic and cultural imperatives to speak English are already so great?

The language loss that's experienced by immigrants is just one part of a larger phenomenon of global language loss, in which many local languages are being abandoned in favor of larger, more economically powerful languages. Currently, about half of the world's languages are considered to be endangered. Where English is spoken, it's not uncommon to find that between 80% and 90% of the languages native to that area have been lost. To many researchers, this accelerating rate of language death parallels the mass extinction of biological species due to industrialization and habitat loss.

For biologist Mark Pagel (2000), this parallel is more than mere analogy. He's found that similar geographic conditions seem to foster biological and linguistic diversity. In North America the highest concentration of different mammals *and* different languages occurs at about 40 degrees north latitude. The most linguistically diverse place on the entire planet is Papua New Guinea, which teems with languages as well as with wildlife; more than 800 languages are currently spoken there. Pagel argues that the key factor supporting the profusion of both languages and wildlife is the wide array of different and distinctive habitats contained within this relatively limited geographic area. Among animals, habitat diversity encourages the evolution of highly specialized traits for a particular, distinctive habitat, without intrusion by invasive species. In the same way, if people are able to find sources of food without having to travel great distances, they may be more likely to develop a highly specialized culture within their habitat, and not interact a great deal with other groups. Modern society, with its rapid industrialization and globalization, poses a threat to these diverse, self-contained habitats, and to the species and languages that they harbor.

Language extinction looms large in Australia, where an estimated 80% of native languages will not survive the current generation (**Figure 12.1**). As Pagel notes, "languages are suffering a mass extinction comparable to that of biological species, and the linguistic landscape is, like parts of Australia itself, rapidly coming to resemble a desert."

Of course, not everyone would mourn the loss of language diversity—after all, idealists have long dreamed that a universal language might unite the world's peoples and promote harmony. Still, the extinction of a language is an undeniable severing of a people's ties to its cultural heritage. And for language scientists, the mass extinction of languages is devastating—it's through the sheer variety of human languages and their possible and probable shapes that we come to learn so much about what a human language is and can be.

It's unrealistic to think that language extinction can be stopped by having people retreat back into self-contained habitats—the world is likely to require increased cross-cultural interactions for the foreseeable future. But fortunately, the human capacity for language encompasses more than just one, and there's a growing awareness in many countries that nurturing multilingualism may be the way to keep languages alive. Languages like Maori (New Zealand), Basque (Spain/France), Irish, and many native North American languages, once stigmatized, have become the focus of energetic language revitalization efforts.

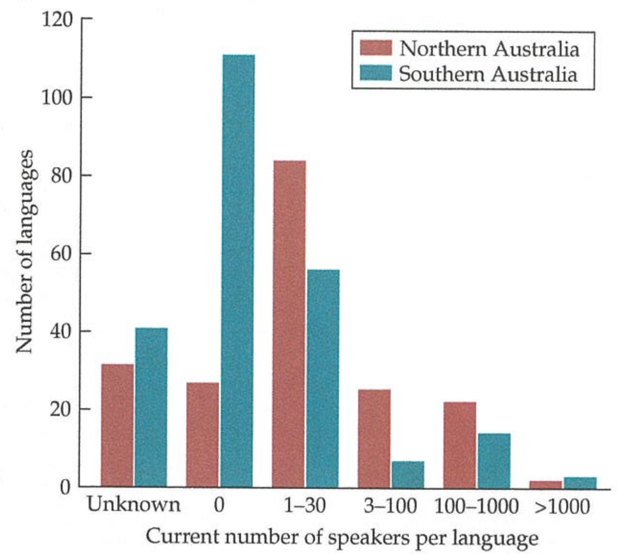


Figure 12.1 The number of speakers per language for aboriginal languages of Australia. The languages of southern Australia, which had greater numbers of European settlers, are even more endangered than languages in the north. (Adapted from Pagel, 2000.)

spoken by smaller numbers of individuals today are being gobbled up by a few of the world's dominant languages.) By calculating the rates at which new languages may have emerged and old ones become extinct, Pagel estimates that as many as half a million languages have been spoken since humans first started talking.

A glance at the world's languages shows that language diversity isn't just a matter of different languages having different words, or of languages having slight tweaks in sound inventories or the ways in which elements are ordered in a sentence. Very different *systems* emerge from among the world's languages:

- The observed number of distinct phonemes (units of sound) in a language ranges from 11 to 164 (Maddieson, 1984).
- Languages differ enormously in the number of morphemes (units of meaning) that they allow to be stuck together to form a single word. Some languages, like Mandarin and Vietnamese, are essentially limited to a single morpheme per word—you wouldn't even tack on a separate morpheme to mark past tense or plural on a word. Other languages allow so many morphemes to be attached to a verb that a single word can capture as much information as a full sentence of English.
- Languages vary dramatically in the *kind* of information that has to be present in the linguistic code. For instance, Mandarin doesn't require that tense be encoded anywhere in a sentence—it can be left entirely up to the context to indicate whether an action is situated in the past, present, or future. On the other hand, Turkish requires a grammatical marker that encodes whether the event described by a sentence was witnessed directly by the speaker, or whether it reflects secondhand knowledge.
- Languages make use of different categories of words as their syntactic building blocks. Not all languages have adjectives or adverbs, for example, or even distinguish between nouns and verbs. On the other hand, some languages have word categories that English lacks. Japanese has a set of words called "classifiers," which you use when specifying the number of objects. The classifier captures some information about the type of object that's being counted. When expressing the concept of three pencils, you'd have to include in the phrase the classifier *hon*, which is used with long, cylindrical objects; other classifiers accompany large animals, small animals, drinks, and so on.
- Some languages, including English and French, have a very fixed word order. Some other languages, like Czech and Russian, permit a certain amount of scrambling of elements. Still others have what appears to be a completely free word order.

The tremendous variety of human languages points up a paradox: How can there be such stunning differences across languages, given that humans all share the same underlying equipment? Clearly, the human mind is able to stretch itself in various directions to accommodate a dazzling assortment of language systems. By closely examining the ways in which languages differ and the various cognitive challenges they pose, we can get an intriguing perspective on the breadth and elasticity of the human linguistic capacity.

At the same time, we also have the sense that people who speak different languages often display profound differences. To learn a new language, we're often told, is to fling open a window onto a different culture. When people make statements like this, they often mean that the language *itself* embodies important aspects of the culture of its speakers and their way of thinking—and by extension, that if you know their language, you too will be able to think like they do. For example, a web page from Bowdoin College suggests that students consider learning Japanese because:

> **Learning Japanese allows you to see the world in a different way.** The philosopher Ludwig Wittgenstein is known to have written this famous aphorism: "The limits of my language are the limits of my world." As he suggests, our ability to imagine possible worlds [is] extended by learning a new language. Companies and graduate schools are interested in cultivating talented people who can function in an increasingly diverse and interconnected world. Learning a new language is the first step towards "thinking outside the box."

Adopting a similar approach, the website of the French consulate general in Houston lists ten "good reasons to learn French," including these two:

> First and foremost, learning French is the pleasure of learning a beautiful, rich, melodious language, often called the language of love. French is also an analytical language that structures thought and develops critical thinking, which is a valuable skill for discussions and negotiations.

It's "common knowledge" that some languages are more romantic—or more orderly, harsh, delicate, simple, sloppy, nuanced, precise, or flowery—than others. (You can conduct a little experiment: survey your friends, and see how much consensus there is about which languages embody each of the above adjectives.) Like the sixteenth-century Holy Roman Emperor Charles V, one might conclude that it's best to acquire a number of different languages for different purposes. Charles is alleged to have asserted that he spoke "Spanish to God, Italian to women, French to men, and German to my horse."

If you share the intuition that languages are infused with certain cultural qualities, let me ask you this: Exactly *which* properties of, say, Italian make the language so romantic? Or what is it *exactly* about the grammar of French that helps the language to structure thought with a pristine clarity? When confronted with questions like these, many people draw a blank. But in order to evaluate the claim that learning a language will open up a different way of thinking, some serious theory-building needs to be done. In what ways could cultural values possibly be stamped into a language, and how might the resulting linguistic features shape the thinking of its new learners? In this chapter, we'll wrestle with these fascinating questions:

- How do languages differ from each other, and what do they have in common?
- Does the human mind impose limits on possible languages?
- Can a lifetime (or even a short time) of speaking a particular language in turn impose constraints on the way we think?

12.1 What Do Languages Have in Common?

How universal are language universals?

Even at a brawny half million, the set of past and present human languages represents a minuscule portion of the languages that could possibly exist. As we've discussed at various points in this book, certain basic properties seem to be critical starting points for human languages. For instance, languages combine meaningless units of sound to create an inventory of meaningful words, they combine words into complex groupings, and their grammatical patterns rely on abstract notions of structure rather than simple linear order. None of these elements are *required* by a system of communication. This becomes obvious when people deliberately set their minds to inventing an artificial language—the results aren't always constrained by even these most basic features of organic human languages (recall the example of the language invented by John Wilkins in the seventeenth century, discussed in Language at Large 2.1).

Even if you stay within the core properties of language, it's not hard to dream up highly deviant grammatical rules that are unlike anything natural languages seem to do. Here's my attempt at generating a few such weird "non-rules":

- If the subject of the sentence begins with the sound /d/, reverse the usual order of the verb and direct object.
- When nouns that refer to objects that are typically red are combined with verbs, the verb is never marked for the past tense.
- When speaking to a person of higher social rank than yourself, omit all prepositions.
- The first noun in a sentence is pronounced in such a way that its last consonant is dropped.

No one has ever compiled an exhaustive list of "outlandish rules that have yet to turn up in language" (and it would be pointless to try). But there have been serious attempts to define the boundaries of natural human languages, and to document patterns that *do* regularly crop up across many languages.

Among these attempts, the work of Joseph Greenberg (1963) has been especially influential. Like other **language typologists** who study the ways in which languages vary, rather than focusing on idiosyncratic rules or specific words, Greenberg tried to capture the broader patterns within languages, patterns that provide some insight into the "bones" of a language. Based on a survey of about 30 languages, he created a list of 45 generalizations, commonly referred to as **Greenberg's linguistic universals** (see Table 12.1). Greenberg's list contains some statements that were true for all of the languages that he studied. Some of these are simple, such as "All languages have categories of pronouns involving at least three persons and two numbers" (Universal 42). Others are more complex **implicational universals** of the form "If A, then B." For example, "If a language has gender categories in the noun, it has gender categories in the pronoun."

Greenberg found very few truly exceptionless patterns among the languages he sampled. And in recent years, linguists have turned up more and more cases of outlier languages that seem to violate what had been previously thought to be universal properties of language. In Box 6.3, you read about Pirahã, a language

language typologists Researchers who study the ways in which languages vary with the aim of describing and explaining crosslinguistic variation.

Greenberg's linguistic universals A set of observations about common or universal structural patterns across a sample of 30 languages by Joseph Greenberg. Published in 1963, Greenberg's observations are still used as the basis of a great deal of inquiry in language typology.

implicational universals Crosslinguistic generalizations that are formulated as conditional statements ("If a language has A, then it has B").

TABLE 12.1 Examples of Greenberg's linguistic universals[a]

Universal 1	In declarative sentences with normal subject and object, the dominant order is almost always one in which the subject precedes the object.
Universal 14	In conditional statements, the conditional clause (the *if* clause) precedes the conclusion (the *then* clause) as the normal order in all languages.
Universal 17	With overwhelmingly more than chance frequency, languages with dominant order VSO (verb-subject-object) have the adjective after the noun.
Universal 18	When the descriptive adjective precedes the noun, the demonstrative and the numeral, with overwhelmingly more than chance frequency, do so also.
Universal 19	When the general rule is that the descriptive adjective follows, there may be a minority of adjectives which usually precede, but when the general rule is that descriptive adjectives precede, there are no exceptions.
Universal 31	If either the subject or object noun agrees with the verb in gender, then the adjective always agrees with the noun in gender.
Universal 36	If a language has the category of gender, it always has the category of number.
Universal 37	A language never has more gender categories in nonsingular numbers than in the singular.
Universal 38	Where there is a case system, the only case which ever has only zero allomorphs is the one which includes among its meanings that of the subject of the intransitive verb.
Universal 42	All languages have pronoun categories involving at least three persons and two numbers.

[a]These are only a few of the language universals proposed by Joseph Greenberg (1963). Greenberg's original numbers are preserved here, since language researchers frequently refer to these universals by their original numbers.

that may lack syntactic recursion. And there's even a language that fails to show duality of patterning: in the emerging language known as Al-Sayyid Bedouin Sign Language (ABSL; see Figure 2.6), signs representing morphemes and words are holistic rather than being made up of decomposable parts, as is the case for all known spoken languages and other sign languages such as American Sign Language (ASL) (Sandler et al., 2011). It may be that there are very few generalizations (if any) that are true for all of the world's languages. Nevertheless, Greenberg also drew attention to patterns that were statistical rather than absolute, that is, patterns that stood out as very common across languages, but for which exceptions could be found, such as "In declarative sentences with subject and object noun phrases, the dominant word order is almost always one in which the subject precedes the object."

Greenberg's crosslinguistic studies suggested that human languages prefer to set up camp in a very small corner of the space of possible languages. Languages seem to shun not only the bizarre kinds of "non-rules" that I invented earlier, but also perfectly reasonable language choices, like ordering objects before subjects. What could explain the tendency for languages to be so like-minded? An intriguing explanation might be that the human brain imposes certain biases and limitations on what a language should look like. These could come through genetically based learning biases (that is, a universal grammar), or could result from various cognitive factors that make some linguistic forms easier to learn or to process than other forms.

How cognitively meaningful are language universals?

It's certainly tantalizing to think that we might learn something profound about the human mind by looking at the ways in which languages are alike. But as many researchers have pointed out, similarities across languages can crop up for reasons that have little to do with the inner workings of the human

BOX 12.1
Language change through language contact

While some elements are more likely to be borrowed than others, almost any aspect of one language can become folded into another as a result of language contact. Here are some examples, drawn from Thomason (2001).

Sound changes

"Click" consonants, present in the Khoisan languages of Africa, are among the most unusual sounds across languages. Rare as these sounds are, they've proven to be amenable to adoption by other languages. Several Bantu languages, such as Zulu, have integrated click consonants into their sound systems through language contact.

Before the Norman conquest of England in the eleventh century, the English sounds [f] and [v] were variants of the same phoneme; the voiced sound [v] occurred in the middles of words, while its unvoiced counterpart [f] appeared at word beginnings. However, after the Norman invasion, many French words were absorbed into the English language, including words that began with the [v] sound. Rather than conforming to the original English phonology, which would have resulted in their being pronounced with a word-initial [f] and, in many cases, sounding very similar to existing English words, these new words were uttered with word-initial [v]. Over time, this led to the two sounds splitting apart into completely separate phonemic categories /f/ and /v/.

Changes to morphology

In many languages, the pronoun for the first person plural is ambiguous between an "inclusive" meaning that includes the person(s) being addressed, and an "exclusive" meaning that does not. (Note that the English pronoun *we* contains just this kind of ambiguity. If someone tells you "We should repaint the house," does this include you or not?) The Dravidian languages spoken in southern India (including, Tamil, Kannada, and Malayalam) use distinct pronouns for these two meanings. Other Indian languages such as Marathi and Gujarati came to adopt the practice of using two separate pronouns to disambiguate the inclusive and exclusive meanings.

Mednyj Aleut, a language spoken on Bering Island, represents a blend of Russian and Aleut, and the language has imported the entire complex system of Russian verb endings.

In English, the morpheme *-able* on words like *readable* and *unstoppable* originally came into the language as an accessory on borrowed French words. Eventually, the morpheme came to be attachable to any English verb—including newly minted words like *Googleable* and *e-mailable*.

Changes to syntax

Finnish is genealogically distinct from most of the European languages, but has been shaped by close contact with its linguistic neighbors. Its basic sentence structure shifted from a subject-object-verb (SOV) word order to subject-verb-object (SVO), bringing the language more in line with nearby Indo-European languages.

Many of the languages spoken in India show evidence of having influenced each other's grammatical structures. For example, Dravidian inherited relative clauses and passive constructions from Sanskrit.

Predictors of change

Researchers who study change through language contact have noticed a few patterns. For one thing, languages that start off being quite similar to each other are more likely to borrow many elements from each other. Second, the intensity of the contact can determine the nature of the borrowing: Words are easily borrowed from one language into another even if the contact between them is very casual. But in order to mix grammatical elements together, a community usually needs to have quite a few bilingual speakers who know both languages well—after all, it takes a fairly deep knowledge of another language to incorporate its grammatical structures, as opposed to the odd word here and there.

Finally, the extent of one language's influence on another depends partly on the speakers' attitudes about incorporating linguistic changes. In some cases, speakers of a minority language may be very resistant to pressures from a dominant language. For example, current speakers of Montana Salish also know English, but the language community has resisted lexical borrowings, even when a concept has no corresponding word in Montana Salish. When speakers of Montana Salish were first acquainted with objects such as cars or televisions, they invented novel words using Montana Salish rather than importing the English words. Thus the Montana Salish word for a car is *p̓íp̓úyšn*—literally, "wrinkled feet," referring to the appearance of tire tracks.

mind. Recurring patterns across languages might simply reflect events of history. Many different languages can evolve out of a common ancestral tongue, and it wouldn't be surprising to find striking similarities among them merely due to their shared origins. Or, similarities could arise because speakers of two different languages came into regular contact with each other and influenced each other (see **Box 12.1** on the previous page). For instance, modern English offers a captivating record of the many imprints that other languages have left upon it. Such imprints are most obvious in the form of "borrowed" words like *lingerie* (French), *maximum* (Latin), *democracy* (Greek), *marijuana* (Spanish), *kindergarten* (German), or *futon* (Japanese). According to some estimates, as much as 75% of the vocabulary of English originated in other languages, largely Latin and French (Thomason, 2001). But language contact doesn't just affect word borrowings; more structural elements of phonology or syntax can also creep into languages from other tongues.

These historical accounts can be fascinating in their own right. But they don't necessarily tell us that much about how language works in the human mind, or where the boundaries of human language lie. To get a sense of the deeper constraints underlying language, it's important to survey a very large number of languages, placing special emphasis on those patterns that come up again and again even for languages that aren't historically related to each other, and whose speakers have not had contact with each other.

In one ambitious line of work, linguist Matthew Dryer (1992) tested Greenberg's universals against a much larger set of 625 languages, and found that, while some of Greenberg's proposed universals failed to hold up in the larger sample, others did quite well. There seemed to be particularly strong evidence for certain word order correlations. For example, if a language places the verb before the direct object in a sentence, you can bet on the fact that it also has prepositions, which occur *before* their associated noun phrases (also called prepositional objects). English is one such language (let's call it a "Type A" language):

Dimitri	swept	the porch	with	a broom.
	verb	object	preposition	prepositional object

On the other hand, if a language places the verb *after* its object, as does Japanese, the chances are very good that it makes use of *post*positions, which occur *after* the associated noun phrase (let's call this a "Type B" language):

Taroo-	*waboo-*	*deinu*	*-obutta*
Taroo	stick-with	dog	hit

Strong correlations like these have been taken as evidence of a deep cognitive bias. Researchers who argue for the existence of an innate universal grammar have suggested that we have inborn "settings" that constrain the possible word orders that we learn as children—we innately "know" that a language is either Type A or Type B. Other researchers have argued that the strong word order correlations reflect a general processing preference to keep word order rules consistent across different kinds of phrases.

More recently, Michael Dunn and his colleagues (2011) have argued that the word order correlations that characterize the Type A versus Type B languages don't reveal anything especially deep about the nature of human languages; instead, they simply reflect the historical development of languages. To support this argument, Dunn and his colleagues built a statistical model to take into account the lineage histories of languages. If word order correlations are indeed "true" universals that constrain the preferred word orders of human language, we'd expect them to appear independent of language lineage

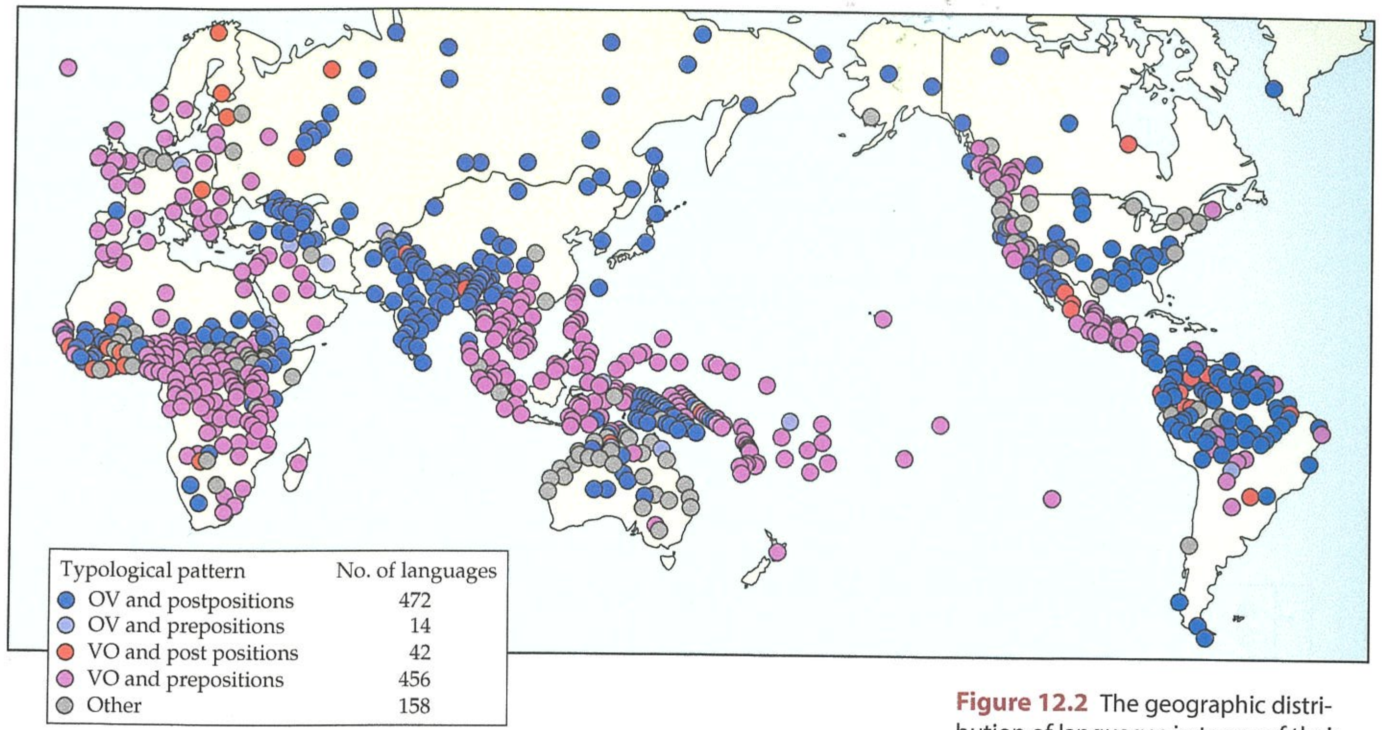


Figure 12.2 The geographic distribution of languages in terms of their ordering of the verb (V) and object (O) and presence of pre- and postpositions. Note how languages cluster geographically by type, suggesting that typological patterns may be due to linguistic lineage or language contact. (From Dryer, 2013c; see http://wals.info/feature/95A#2/14.9/152.8.)

or geography. But Type A and Type B languages divide up rather neatly by geography, as shown in **Figure 12.2**. Furthermore, Type A and Type B word order patterns tend to cluster in languages that are very closely related to each other, as shown in **Figure 12.3**.

The fact that word order type can be largely predicted by a language's historical roots and/or its geographic context makes it hard to rule out the possibility that either a common ancestry or contact between languages is responsible for the close connection between verb-object order and whether a language has prepositions or postpositions. Under this account, there's nothing "deep" about the fact that languages that have the verb-object order also tend to have prepositions rather than postpositions; the correlation simply reflects the fact that languages with a shared history are likely to be similar in a *number* of different ways. If you looked closely enough, you might find many other correlations, none of them particularly deep or meaningful.

Here's an analogy: Suppose you find a correlation between a language's use of tone to distinguish lexical meanings and the likelihood that its speakers use chopsticks as eating utensils. Does this point to some intrinsic connection between lexical tone and chopstick use? That seems rather unlikely. A more plausible interpretation is simply that the speakers of various tone languages share a common cultural and linguistic heritage, which included, among other things, the wielding of chopsticks during dinner.

Are there meaningful constraints on the shape of human languages, constraints that truly reflect something about the human mind? Looking at recurring crosslinguistic patterns and universals can provide some hints (and nowadays researchers can begin to explore interesting patterns with the help of research tools such as the online database of languages found in the World

WEB ACTIVITY 12.1

Variation across languages In this activity, you'll explore the online World Atlas of Language Structures (WALS). This resource identifies many ways in which languages vary from one another, and allows you to see the geographic distributions of various linguistic features.

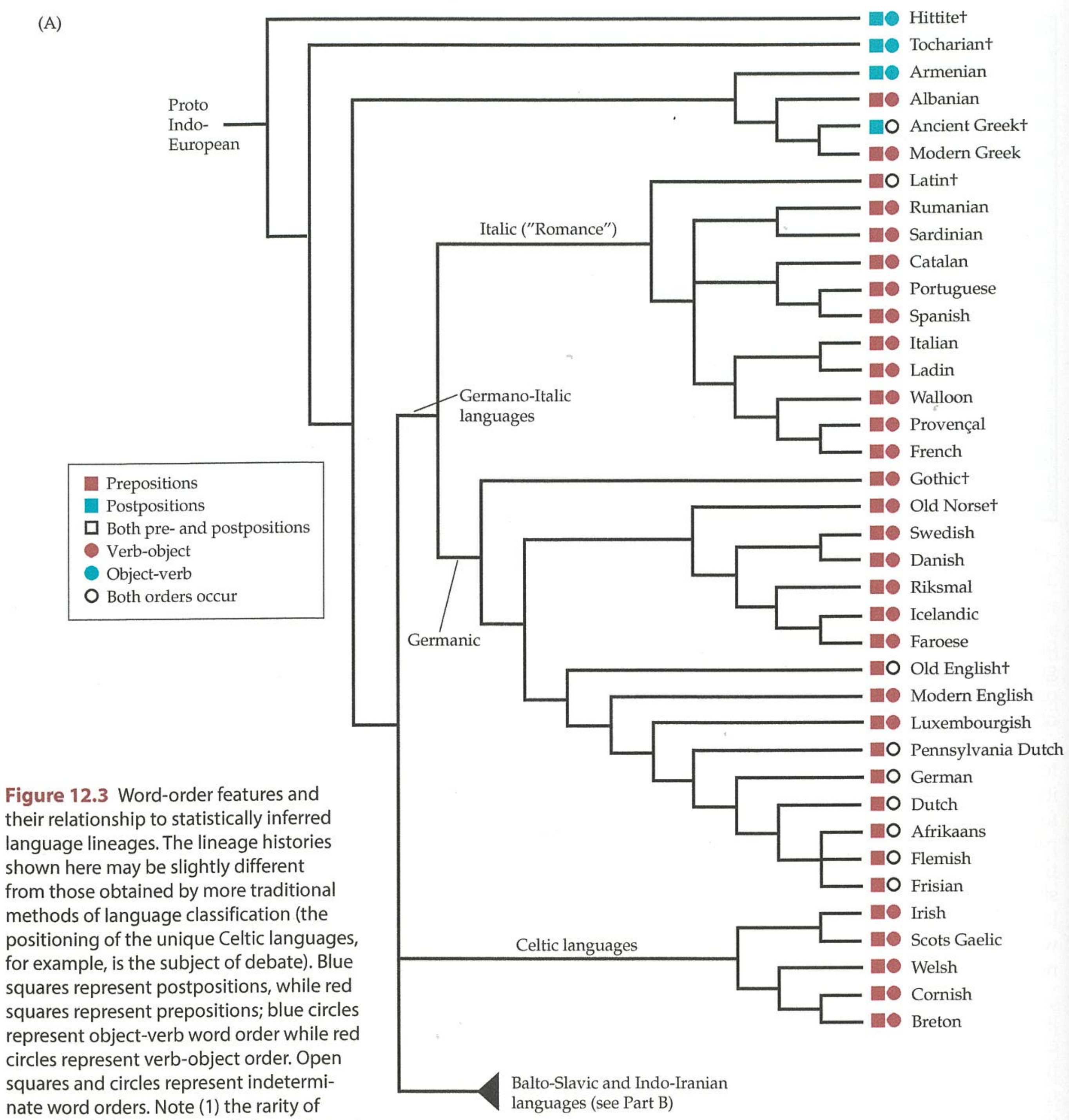
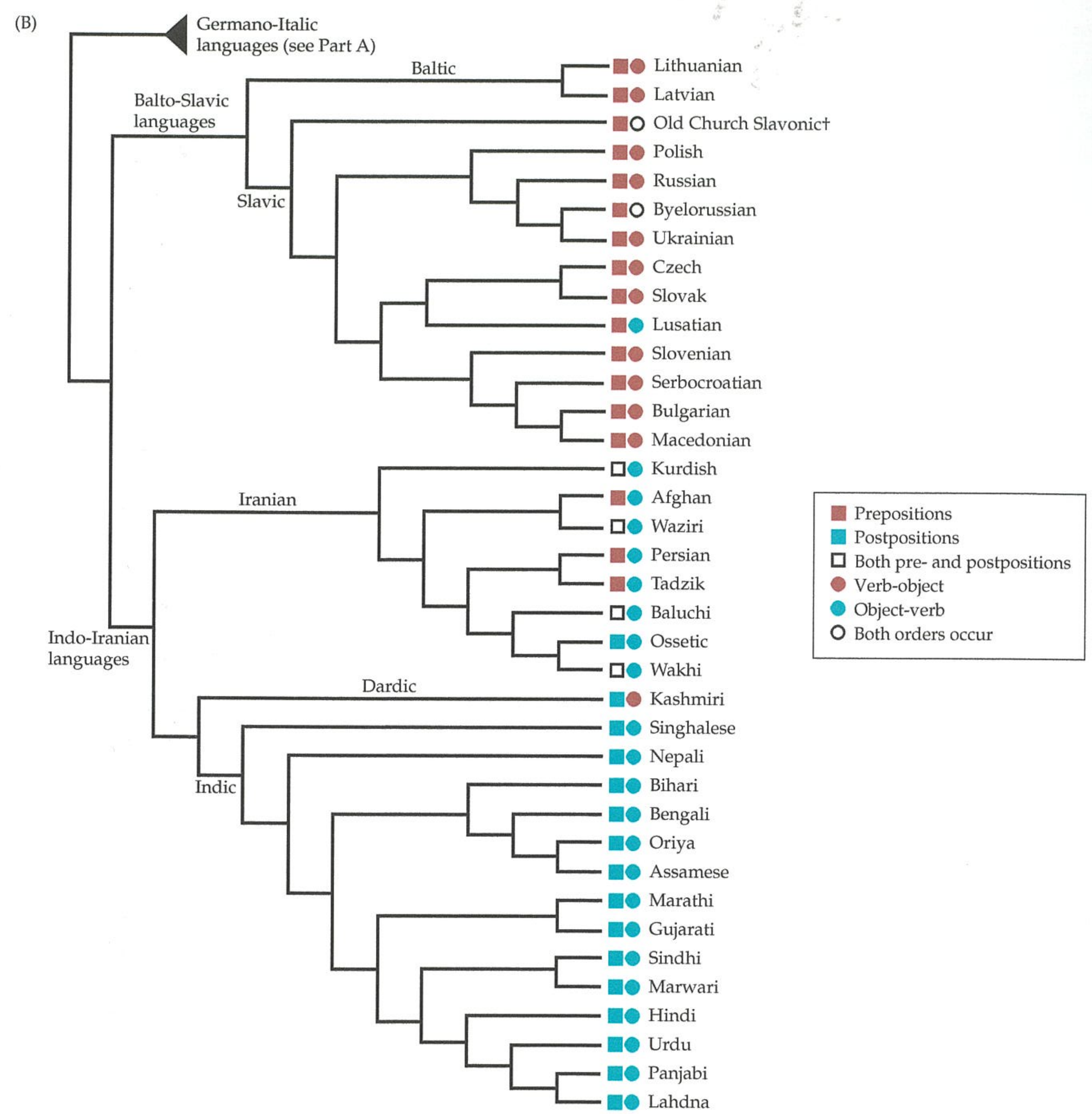


Figure 12.3 Word-order features and their relationship to statistically inferred language lineages. The lineage histories shown here may be slightly different from those obtained by more traditional methods of language classification (the positioning of the unique Celtic languages, for example, is the subject of debate). Blue squares represent postpositions, while red squares represent prepositions; blue circles represent object-verb word order while red circles represent verb-object order. Open squares and circles represent indeterminate word orders. Note (1) the rarity of mixed (red-blue or blue-red) elements; and (2) the degree to which linguistic elements are similar across closely related languages. (Adapted from Dunn et al., 2011.)

Atlas of Language Structures, or WALS). But as we've just seen, there are some remaining challenges in sorting out exactly where these similarities come from and how extensive they are. Luckily, we can work toward this question from another angle: we can start by looking at whether certain patterns are generally easier for people to learn or to produce than others, and by seeing whether there's a connection between the easy patterns and those that are commonly found across languages.



12.2 Explaining Similarities across Languages

Learning biases

Many researchers have argued that the existence of universal or recurring language patterns is evidence that the human mind plays favorites, finding some linguistic forms more "learnable" than others. But as you just saw, it's not always clear whether these crosslinguistic similarities have a cognitive basis as opposed to a historical one. Hence, a number of language scientists are finding more direct ways to search for learning biases.

If you were interested in pursuing this line of research, you'd have a few options. One strategy might be to gather detailed profiles of children's language development across a variety of languages, to see how well the learning data lined up with typological patterns. Would it take kids longer to master a weird feature that very few languages have, and would they make many mistakes on their way to learning it? This would be a perfectly logical approach, but it would come with some thorny practical and methodological challenges. For example, you could only study those "universals" for which you could find exceptions, since the whole approach would be to see whether the exceptional linguistic features were harder to learn than the common ones. This would eliminate the possibility of looking at some of the most robust—and possibly most interesting—universals. Another challenge would be that if you wanted to directly compare the learning trajectories for rare versus common structures, you'd need to make sure that the children in your studies got comparable doses of the two structures in their input. This would be a real concern, since structures that are rare across languages also tend to be rare even *within* those languages that allow them. As a result, these structures could take a long time for children to control, not because of their inherent difficulty, but simply because kids wouldn't come across them that often. None of these challenges would be insurmountable ones, but they'd constrain the questions you could ask, and make your research time-consuming and expensive.

Or, you could take the easy way out: invent a pair of miniature artificial languages with just the properties you're interested in, and compare how easily they're learned in a lab setting where you can carefully control the learners' input. This approach allows you to neatly test whether certain patterns are more readily learned than others. It even allows you to invent a language with structures that don't exist in any known human language.

We've already spent a bit of time in this book exploring the potential of experiments based on learning artificial language. In Chapter 4, you read how people (ranging from babies to adults) learned to use the statistical patterns in an artificial language to segment words from speech, even after a very short exposure to the language. In Chapter 6, you read how children learned syntactic categories based purely on statistical input, and that they showed sensitivity to statistical variation in deciding whether to generalize certain patterns. But language learners don't always faithfully reproduce the statistical patterns in the input they get, and this is where things get really interesting. For example, in a learning study with adults and 5- to 7-year-old children, Carla Hudson Kam and Elissa Newport (2005) created two versions of an artificial language. In one of these, any "noun" was always partnered with a word that functioned as a determiner (like *the* or *a* in English). In the other version, determiners showed up inconsistently, accompanying "nouns" only 60% of the time. The children and adults were later tested on their knowledge of the language (for example, by completing a partial sentence, or by judging whether certain sentences were correct). Most of the children coped with the inconsistent input by regularizing it—the majority of them produced the determiners all the time, while a few others always left them off. The adults, on the other hand, were more likely to preserve the original inconsistent patterns and produce determiners in about the same proportion as they'd heard them in the input. But a later study by the same authors showed that adults too could be nudged in the direction of over-regularization if the language was more complex or the learning task was made more difficult (Hudson Kam & Newport, 2009). So, it seems that when language learning is challenging, both children and adults are prone to changing a language by making it more systematic and regular.

TABLE 12.2 Examples of adjective/noun/numeral orders

Language	Order	Standard English order
Cherokee	*u-wo'-du a-ge-hyu'-tsa* pretty girl	pretty girl
	tso-i gu:-gu three bottles	three bottles
Yoruba	*bata titun* shoes new	new shoes
	awo meje dishes seven	seven dishes
Basque	*etxe zuri* house white	white house
	bi zuhaitz two trees	two trees
Sinhala	*loku pot* big books	big books
	geval tunak houses three	three houses

Source: Culbertson et al., 2012.

It's precisely in these gaps between the patterns in the input and the learners' output that we might find evidence of learning biases. Will certain language structures—those that are more "natural" or common for language—act as stronger magnets for generalization, causing language learners to settle into them on the basis of fairly sparse evidence in the input? If so, we might find that people are more eager to generalize patterns that are very common across languages, but don't readily generalize patterns that languages seem to avoid. This would be very compelling evidence that not all linguistic patterns have the same status in the mind.

Jennifer Culbertson and her colleagues (2012) used exactly this kind of reasoning to study a common crosslinguistic pattern known as "Greenberg's Universal 18," which states that,

> When the adjective precedes the noun, the numeral, with overwhelmingly more than chance frequency, does likewise.

The four logically possible ways of ordering adjective/noun pairs and numeral/noun pairs are:

1. adjective-noun & numeral-noun (as in English: *prickly socks* & *three socks*)
2. noun-adjective & noun-numeral (*socks prickly* & *socks three*)
3. noun-adjective & numeral-noun (*socks prickly* & *three socks*)
4. adjective-noun & noun-numeral (*prickly socks* & *socks three*)

All four word orders occur in real languages (see Table 12.2). But as you can see in Table 12.3, some of these word orders are more common across languages than others; in particular, the fourth order (adjective-noun & noun-numeral), as Greenberg originally noted, is exceedingly rare, representing just 4% of the 851 languages in the WALS sample.

Culbertson and her colleagues created an artificial learning task in which subjects had to learn the "alien" names for ten novel objects, as well as the names for five properties and five numerals. During a two-phase learning process, each

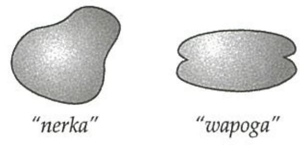
TABLE 12.3 Noun/adjective/numeral orders across 851 world languages

	Noun-adjective	Adjective-noun
Numeral-noun	149 (17%)	227 (27%)
Noun-numeral	443 (52%)	32 (4%)

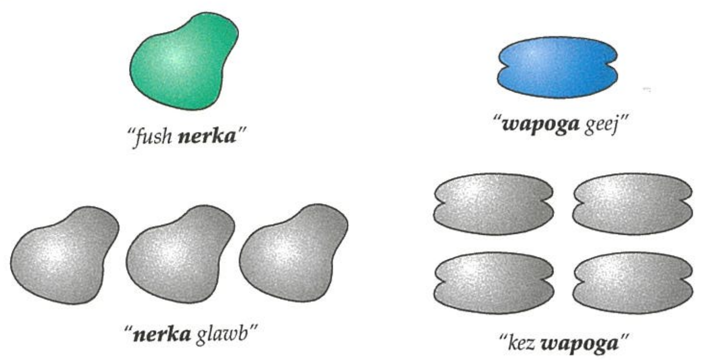
Source: Culbertson et al., 2012.

(A) Learning

Phase 1: Learning nouns
Participants saw strings of single "objects" and heard and repeated made-up "nouns" that named the objects. After successfully testing on their knowledge of the nouns, they moved on to Phase 2.



Phase 2: Learning numerals and adjectives
Participants saw novel word pairs in which adjectives designating colors or numerals combined with the previously learned nouns (nouns are in bold):

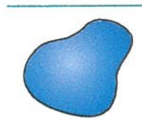


Word pairs were presented in different proportions of the various possible word orders (distribution conditions). Participants were randomly assigned to one of the five distribution conditions in the table, and after 80 exposures, performed on a test (Phase 3) requiring them to correctly match each complex phrase to one of four images.

Assigned condition	Adjective-noun	Noun-adjective	Numeral-noun	Noun-numeral
0	50%	50%	50%	50%
1	70%	30%	70%	30%
2	30%	70%	30%	70%
3	30%	70%	70%	30%
4	70%	30%	30%	70%

(B) Testing

Phase 3: Subjects were required to produce a complex phrase containing adjectives or numerals as well as nouns. They were awarded 10 points for correct vocabulary, and an additional 5 points if they produced the same word order as their "alien informant." Participant responses were statistically matched to the word-order probabilities from Phase 2.

(1) Picture	(2) Participant responds	(3) Vocabulary points	(4) Informant responds	(5) Order points
	"nerka geej"	10	"nerka geej"	5

(C) Results

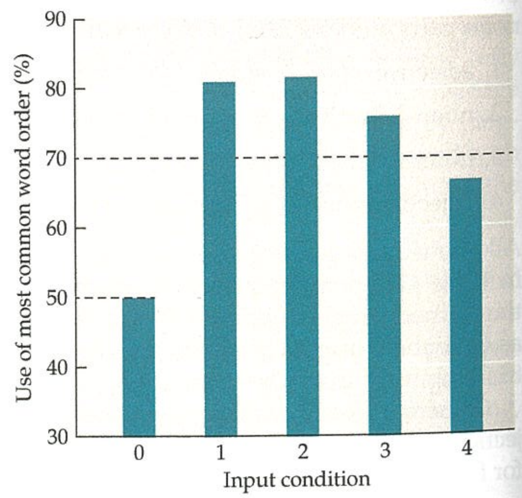


Figure 12.4 Experimental design and results from Culbertson et al. (2012). (A) Participants underwent a two-phase learning process, with evaluation for mastery at each phase. (B) The test phase was designed to assess whether participants would regularize inconsistent input in a way that lined up with the distribution of word orders across languages. (C) The graph shows the proportion of trials in which subjects produced the most common word for each input condition ("use of most common word order"). The dotted line indicates the frequency with which that order appeared in the target input. Note that participants tended to produce the most common word order with even greater regularity than it appeared in the input, except in Condition 4, which corresponded to the typologically rare distribution. (Adapted from Culbertson et al., 2012.)

participant heard a version of the alien language in which all four of the above word orders occurred, but in varying proportions (see **Figure 12.4A**). Hence, they all received some data that was inconsistent with respect to the word order for that language. They then moved onto a test phase, in which the question of interest was whether they'd regularize the inconsistent input and (more to the point) whether they'd regularize inconsistent input in a way that lined up with the distribution of word orders across languages. If so, this would suggest it's no accident that some word orders are more common across languages than others.

In the test phase of the study (see **Figure 12.4B**), the subjects had to describe either a single alien object or a set of them, using the correct noun, adjective, and numeral words. They were awarded ten points for producing the right vocabulary items. They were given five bonus points if they generated the same word order as their computerized "alien informant." The responses of the "informant" were statistically generated in order to match the version of the language that each subject was working with—that is, if the noun-adjective order had appeared in the subject's target language 70% of the time, the "alien" was 70% likely to produce a noun-adjective order on the test items.

The researchers found that learners tended to regularize word order for some but not all of the languages (see **Figure 12.4C**). They were most likely to regularize patterns in which the noun occupied the same slot relative to both the adjective and the numeral—that is, the adjective-noun & numeral-noun pattern (Order 1 in Table 12.2), and the noun-adjective & noun-numeral word pattern (Order 2 in Table 12.2). They were *least* likely to regularize the pattern that was identified as rare by the language typologists (the adjective-noun & noun-numeral word order; Order 4 in Table 12.2). Even in a learning task that introduced some pressure to regularize, learners resisted producing a general pattern that went against the typological grain.

This study is only one of a number of experiments with artificial languages that show a close fit between learning biases and some attested crosslinguistic patterns. Findings like these make it hard to dismiss all typological universals as the results of historical forces, revealing no deeper truths about the ways in which human minds and human languages are made for each other.

Naturally, finding clear evidence of learning biases has the effect of fueling researchers' desire to explore their underlying nature. Throughout this book, we've explored two possible explanations for learning biases. One explanation, offered by researchers of a nativist bent, is that learning biases are evolutionary "gifts" that help us navigate our way through the enormous space of logically possibly languages; they are part of the innate language-specific knowledge we've inherited from our articulate ancestors. The other perspective is that learning biases reflect patterns that are especially easy to notice or remember; they come from general aspects of our cognition, and aren't necessarily language-specific.

Of course, it's possible that *both* kinds of learning biases exist, and each learning bias, as it's discovered, will need to be evaluated on a case-by-case basis. To do this, we might create non-linguistic analogues to artificial languages, and see if certain patterns are preferred over others even in non-linguistic domains. If they're found to be, that will suggest that the learning bias is grounded in general cognition. Here's one example of this approach in action, from a study by Julie Hupp and her colleagues (2009), who wanted to look at what's responsible for the fact that more languages prefer to mark inflectional information, such as plural or tense, by tagging the end of a word with a suffix (as in *cat-s*, or *walk-ed*), rather than by using a prefix at the beginning of a word (see **Table 12.4**).

TABLE 12.4 Preference for suffixes versus prefixes across 869 world languages

Preference	Number (percent) of languages
Equal prefixing and suffixing	147 (15%)
Prefer suffixing	529 (55%)
Prefer prefixing	52 (16%)
Little or no prefixing/suffixing	141 (14%)

Source: Dryer, 2013c; see http://wals.info/.

The researchers first set out to test whether this suffix-preference pattern reflects a general bias for language. They had subjects listen to a two-syllable sequence, such as *ta-tee*, and then asked them to decide whether this sequence was more similar to *bee-ta-tee*, or to *ta-tee-bee*. To make sense of this task, think about how you'd answer the question, "Is *cat* more similar to *cats* or to *scat*?" Presumably, your answer would be *cats*, because *cats* is just *cat* with an extra element attached, whereas *scat* and *cat* are two completely different words. By analogy, if people think that *ta-tee* is more similar to *ta-tee-bee* than to *bee-ta-tee*, it suggests that they find it easier to think of adding an extra piece at the end of a word (as a suffix) than at its beginning (as a prefix). When sounds are added at the beginnings of words, they might be more likely to be perceived as completely changing the identity of the original words, much as adding /s/ to the beginning of *cat* creates a completely different word. Sure enough, people generally found the sequences to be more similar if the syllable was attached to the end of the original sequence rather than to the front. The experimental results of this little study line up neatly with the crosslinguistic preference for using inflectional suffixes rather than prefixes.

The next step was to see whether this bias in similarity perceptions would extend into *non*-linguistic domains. Hupp and her colleagues concocted parallel tasks using sequences of musical notes or sequences of visual symbols as stimuli (see **Figure 12.5**). The results echoed the data for syllable strings; people thought that adding material to the front of a sequence changed it more deeply than adding material to the back end. The authors concluded that a domain-general bias nudges languages to attach inflectional information to words as suffixes rather than as prefixes.

Experiments like the ones I've just sketched represent very early explorations of learning biases in artificial languages, and more data need to be gathered. (For example, the study by Hupp and colleagues, while intriguing, was limited to English-speaking American subjects, and we'd certainly want to know how the learning task would play out for speakers of a language that handles affixes in a completely different way, or lacks them entirely.) But these studies no doubt mark the beginnings of a productive exchange between experimental psycholinguists and language typologists who meticulously pick through the languages of the world to find crosslinguistic patterns and correlations. In the lab, language scientists can test how learning biases might apply pressure on languages to settle into "friendly" grammars rather than difficult ones. They can also study how very subtle initial biases might change the shape of a language as it's transmitted from one set of users to another, mirroring historical processes of language change. And, in principle, studies of artificial language learning could also be used to address the question of whether learning biases truly *are* universal across all of the world's populations, or whether some populations might have evolved slightly different linguistic biases than others (see **Box 12.2**).

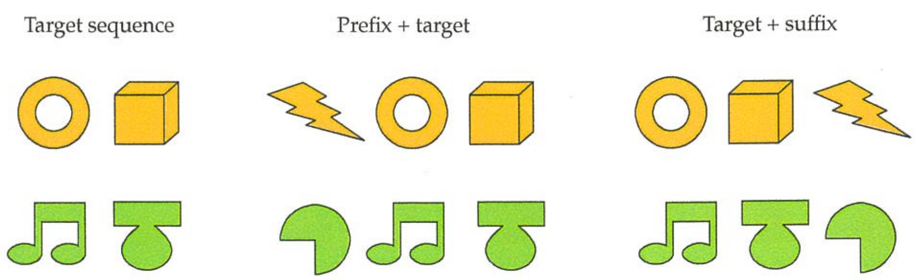


Figure 12.5 Two examples of the visual sequences devised by Hupp et al. (2009). Results of the study revealed that people judged the target strings to be more similar to the target + "suffix" sequences than they were to the "prefix" + target sequences.

BOX 12.2
Do genes contribute to language diversity?

Throughout this book, I've adopted the common assumption that whatever genetic predisposition people have for learning language is a universal *human* predisposition. In broad strokes, this is surely right, in contrast to many more noticeable physical traits—people of, say, Japanese heritage may not usually grow to be as tall as people of Dutch heritage, but toss both of them into the same linguistic environment, and there's no noticeable difference in their mastery of that particular language, whatever it is.

At the same time, I've acknowledged that at the *individual* level, there are likely to be some genetically based differences among people. For example, disorders of language and speech (such as specific language impairment or stuttering) do tend to have a strongly heritable component. What's more, even individual differences that fall within the "typical" range of language abilities may have some genetic basis—for instance, Karin Stromswold (2001) found that heritability is a factor in the size of young children's expressive vocabularies, presumably due to a heightened capacity for word learning.

Is it possible, then, that some of the diversity found among the world's languages could be traced back to subtle genetic differences that shape the learning process of individuals within a language community? Recently, several researchers have begun to explore this possibility, as summarized by Dan Dediu (2011).

Any genetic variations that bias the learning process are likely to be very slight. This is apparent in the flexibility with which people can learn a variety of different languages, and the readiness of entire languages to absorb the traits of other languages they come into contact with. But even very slight individual differences in learning biases might be amplified within a larger population through the process of cultural transmission, as has been suggested by several computational models.

Some suggestive evidence for a possible link between genes and language structure comes from the study of tone languages, in which pitch is used to distinguish words or to convey grammatical information. Linguistic tone *is* a feature that can hop from one language to another through borrowing, suggesting that you don't need to have to be born into a specific linguistic population to be able to learn it—but Dediu has argued that the trait is more stable and less subject to language change than you'd expect if its transmission were purely culturally determined. More specifically, Dediu and his colleague Robert Ladd (2007) have argued that the geographic distribution of tone languages is connected to the prevalence of two human genes involved in brain growth and development, *ASPM* and *Microcephalin*.

It's still very, very early days for the genetic study of linguistic diversity, but in principle, the hypothesis could be systematically explored from a number of different angles. One possible approach might be to design artificial-language studies to test whether learning biases vary across geographic populations in a systematic way that is related to structural properties of the languages spoken in a geographic region.

Speakers' choices

As I've just discussed, the contours of human language may take some of their shape from learning biases that result in some structures being easier to learn than others. But this isn't the only way in which our cognitive makeup might tug language in one direction rather than another. Human language may also be influenced by factors that make some structures easier to *produce* than others.

Ultimately, the shape of a language is the result of millions of sentences uttered by its many speakers. Structures that are frequently produced are most likely to survive transmission across many generations; they are the most robustly learned, and hence the least likely to shape-shift or erode from a language over time. But what determines whether a certain structure is likely to be produced in the first place?

As you saw in Chapter 9, the act of speaking involves making constant choices from among the varied menu of linguistic options for expressing a particular idea. In that chapter, you read about how speakers' choices can be driven by cognitive pressures; in any particular instance, people will often choose a linguistic form that makes the arduous task of speaking just a little bit easier. Now, what if we zoom out and look at the production of language through the lens of linguistic universals or tendencies? Some provocative questions and predictions quickly arise.

Are some linguistic forms systematically easier to produce than others? If so, we'd expect that, wherever their language permitted, speakers would tend to use the easier forms more often than the more difficult options for expressing the same idea. This simple asymmetry could set in motion a cross-generational language shift: the input to new language learners would be riddled with the easier-to-say forms, while the harder-to-produce structures would be more sparse. We've already seen that new learners of a language have a tendency to over-regularize the input they hear; therefore, the next generation of language users would be prone to exaggerate these statistical differences. Over time, the harder structures might drop out of the language entirely, while the easier structures would be preserved. So we'd predict that if were to look across languages, we'd be more likely to find structures that ease the demands of speaking rather than corresponding structures that put more stress on cognitive resources during speaking.

Let's take a more concrete look at how production pressures could lead to common crosslinguistic patterns. I'll draw on an example from Maryellen MacDonald (2013), one researcher who has argued that production pressures are likely to play an important role in explaining crosslinguistic tendencies. In English, if we want to describe an event that involves two participants—the subject (S) and the object (O) of a verb (V)—we normally use the word order: subject-verb-object (SVO). The SVO order is just one of six possible ways in which these three linguistic units could be combined. But if we look across many of the world's languages, some of the options are wildly more popular than others. The vast majority of languages embrace the solution of placing the subject first; and languages seem to be almost allergic to ordering the object before the subject (see **Figure 12.6**).

Is it possible to tell a story about how production pressures might lead to a bias for placing the subject toward the beginning of a sentence? As discussed in Chapter 9, we know that when a word or phrase is highly accessible, speakers tend to utter it as quickly as possible. The general idea is that as soon as speakers are mentally prepared to utter a word or phrase, they spit it out in order to avoid clogging up working memory while planning the rest of the sentence. We saw that various factors could affect the accessibility of a linguistic unit. For example, shorter phrases were more likely than longer phrases to be uttered early in the sentence. Words could be made more accessible through previous mention, or through priming with a semantically related word. And in some experimental manipulations, characters or objects in a visual scene were made more visually salient through the use of flashing markers—with the result that drawing visual attention to an entity made it more likely that speakers would mention it first.

If, upon planning a sentence, speakers find that it's especially easy to bring to mind the instigators of the action they want to describe, then they should prefer to order subjects first, wherever possible. Some researchers (e.g., Bock et al., 1992) have suggested that animate concepts, such as *dentist* or *woman*, are retrieved from memory more easily than inanimate ones, such as *flower* or *car*. Animacy is very closely linked to subjecthood; notice that whenever an event involves an animate participant and an inanimate entity, the animate participant is almost always the subject—just try coming up with sentences involving the word pairs *dentist/flowers*, *steak/panther*, and *boy/book*.

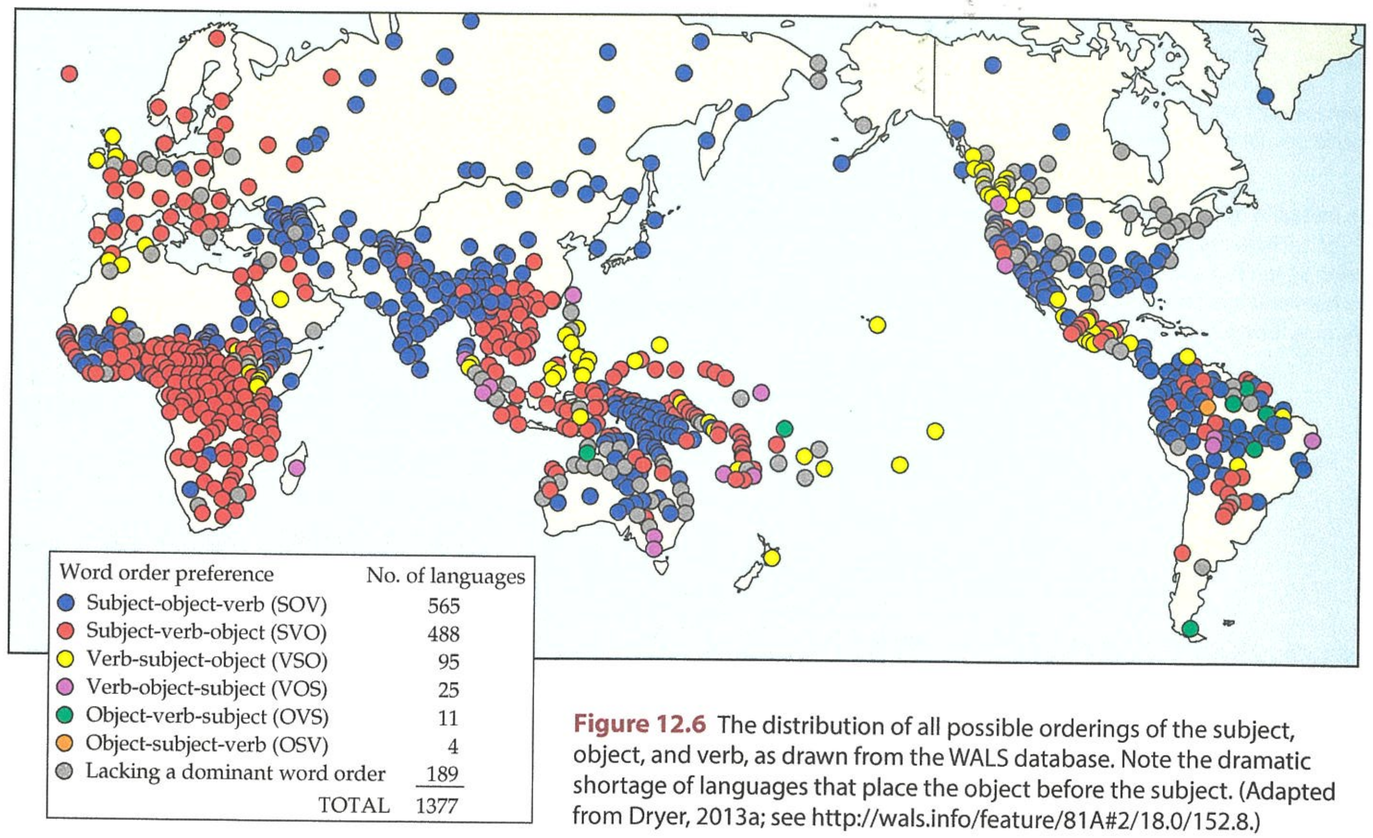


Figure 12.6 The distribution of all possible orderings of the subject, object, and verb, as drawn from the WALS database. Note the dramatic shortage of languages that place the object before the subject. (Adapted from Dryer, 2013a; see http://wals.info/feature/81A#2/18.0/152.8.)

Indeed, a number of studies have shown that speakers exercise their grammatical options in such a way as to order animate nouns earlier in the sentence than inanimate nouns. In the study by Kay Bock and her colleagues, subjects were more likely to use the passive voice if it resulted in the animate participant being ordered before the inanimate object in the sentence: *The boy was hit by the truck*, rather than *The truck hit the boy*.

Let's imagine that the English language evolved from a previous state in which word order was quite fluid and multiple word orders were allowed, as is the case for a number of existing languages. So, for example, all of the following sentences could have meant the same thing:

The students devoured the free pizza. (SVO)

The students the free pizza devoured. (SOV)

Devoured the students the free pizza. (VSO)

Devoured the free pizza the students. (VOS)

The free pizza the students devoured. (OSV)

The free pizza devoured the students. (OVS)

Because of the pressures on language production just described, the subject-first orders would be easier to produce than others, and hence would be the most likely to be uttered. Let's suppose that the SVO order was the most commonly produced order. (In Web Activity 12.2, you'll have the opportunity to explore why this particular order might be preferred over the other subject-first solutions.) The next generation of learners, who heard the SVO order more often than any of the others, would magnify this bias in their own language usage. The subsequent generation would see an even greater proportion of SVO sentences, and so on, until this word order became "fixed" in the grammar of the language as the only allowable word order for that language.

The above scenario is somewhat speculative, and not all researchers agree that it's the best way to explain the crosslinguistic preference for ordering subjects first—careful testing of competing theories has yet to take place. But the general research program behind this specific account is a promising one. As you saw in Chapter 9, the cognitive demands of producing language on the fly can lead speakers to systematically favor some linguistic forms over others that have the same meaning. This imbalance in the frequencies of alternative forms could put the language on a path to crystallizing the original bias with the result that only the favored alternative would be deemed acceptable. Naturally, any such production biases would have to interact with the learning biases of new learners—learning biases could either accelerate or put the brakes on language change that was triggered by speakers' tendencies to use easier forms.

Communicative efficiency

There's clear evidence that speakers often act as cognitive misers, exploiting various linguistic options to reduce their own processing costs. But avoiding extra work is obviously not the only thing that motivates speakers—after all, the whole point of talking in the first place is to get your message across. Presumably, the desire to avoid communicative breakdown also guides the linguistic choices of speakers.

One of the most common ways in which communication can break down is through ambiguity—the speaker intends one meaning, but the language he's chosen doesn't allow the hearer to figure out which one of several possible meanings he has in mind. Throughout the last few chapters, we've seen that speakers have a mixed record when it comes to detecting and avoiding possible ambiguities that could disrupt the hearer's interpretation. For instance, we've seen that speakers have a hard time noticing that words like *bat* can mean either of two very different things, or eliminating the garden path effect for sentences like *I hear you are leaving*, where the hearer might initially be lured into thinking that the word *you* is the direct object of hear (as in *I hear you loud and clear*). We've also seen that speakers don't always take into account the hearer's perspective or knowledge, showing a tendency to fall back on more egocentric strategies when cognitive resources are scarce. It's hard to make the case that avoiding processing difficulties for the hearer is uppermost in the speaker's mind as he makes rapid-fire decisions about how to linguistically encode his intended message. Still, for the most part, speakers and hearers do seem to avoid communicative meltdowns. The ambiguity that's inherent in everyday language can usually be resolved by the hearer —for instance, pronouns, which are *exceedingly* ambiguous, are usually interpreted without any angst, and only some garden path sentences cause anything more than a minor hiccup in language processing. Some balance seems to have been struck between the processing demands on the speaker, and potential for confusion by the hearer.

One intriguing idea is that speakers are freed up from having to consider potential ambiguity on a sentence-by-sentence basis because their language won't allow them to produce truly problematic ambiguities—the idea is that language quickly evolves in ways that limit the rampant possibility of gross misinterpretations. The remaining ambiguities that the language does allow can be handled quite easily by the hearer.

WEB ACTIVITY 12.2

Mumbling, male nurses, and SVO order In this activity, you'll explore some of the ways in which speakers may adapt their language in order to maximize communicative efficiency.

Many languages have flexible word orders, in which case word order doesn't provide any information about whether a noun phrase plays the role of subject, direct object, or indirect object in a sentence. Such languages often have **case markers** to do the job—these are morphemes that are attached to nouns (or some element within noun phrases) to identify who did what to whom. In English sentences like *The man bites the dog* versus *The dog bites the man*, we depend on word order to decipher meaning. But in German, you can unambiguously describe a man-bites-dog incident using either one of the following word orders:

Der Mann beißt den Hund.

Den Hund beißt der Mann.

This is because the article *der* in *der Mann* ("the man") signals that this noun phrase is a subject, while the article in the noun phrase *den Hund* ("the dog") unambiguously marks it as a direct object. If you wanted to describe a dog-bites-man occurrence, you could avail yourself of these two options, again, neither of them ambiguous:

Der Hund beißt den Mann.

Den Mann beißt der Hund.

In German, articles are always helpfully marked for case. But some languages are a bit stingier with their case markers, and use them for certain types of noun phrases but not others. What's interesting is that in such languages—referred to as languages with **differential case marking**—the case marker tends to appear in only those situations where it would be most helpful for avoiding ambiguity. For example, in Sinhalese, word order is flexible, and the case marker *-wa* often appears as a suffix on direct objects, but only if they are animate:

Amara Lalani-wa edda	or	*Lalani-wa Amara edda*
Amara Lalani pulled	or	Lalani Amara pulled

In fact, both of the above sentences would translate into English as "Amara pulled Lalani," and the case marker is crucial for letting the hearer know who did the pulling. But if the object is inanimate, the danger of ambiguity is slight, because inanimate objects rarely instigate events. In these cases, Sinhalese prohibits the object from wearing an accusative case marker, so if you want to tell someone that Amara bought a book, you can't say:

**Amara pota-wa gatta*	or	**Pota-wa Amara gatta*
Amara book bought	or	Book Amara bought

Instead, the word for the inanimate object, a book, has to appear bare:

Amara pota gatta	or	*Pota Amara gatta*
Amara book bought	or	Book Amara bought

Across languages, case markers crop up where the potential for ambiguity is greatest. If other aspects of the language (such as word order) step up to take on a disambiguating function, the case system is vulnerable to erosion over time, and may eventually be wiped entirely from the language. This is what happened in the transition from Latin, which had case marking and relatively free word order, to Italian, Portuguese, and Spanish, which have more rigid word order. Although remnants of case marking stayed in the pronoun system and prepositions, most of the case markers from Latin are gone, because systematic word order is used instead. The selective presence of such markers suggests that languages settle into a "sweet spot" where the speaker is freed from unnecessary complexity, but is steered by the grammar of the language to provide enough information to avoid serious confusion in hearers.

case markers Morphemes that occur within a noun phrase to signal its grammatical function (e.g., subject, direct object, indirect object). Case markers may occur on nouns, articles, adjectives, or on any or all of these.

differential case marking A system of case marking in which case markers appear selectively on some but not all noun phrases. For example, object case marking may be limited to appearing with animate nouns.

Once again, experimenting with artificial languages provides an opportunity to test ideas about how languages might evolve to balance the demands of language production with the need for clear communication. In one such study, Maryia Fedzechkina and her colleagues (2012) created an artificial language with flexible word order, in which both subject-object-verb (SOV) and object-subject-verb (OSV) orders were allowed. This meant that speakers couldn't assume that hearers would be able to understand the meaning of a sentence based solely on its word order. The language also contained optional case markers, and hence the opportunity to disambiguate sentence meanings. But unlike languages with differential case marking, this miniature language didn't distribute its case markers in a selective way to prevent the worst ambiguities. For example, case marking appeared just as often on inanimate objects as on animate ones, despite the fact that it's much more useful for disambiguating meaning when it appears on an animate object. Adult learners of this language came into the lab for four sessions of training and testing. Even though they were exposed to input in which case markers were evenly doled out between animate and inanimate nouns, this wasn't the pattern that they reproduced in their own speech. Instead, they shifted the language toward differential case marking: they were more likely to produce case markers on animate objects than inanimate ones. In other words, case marking appeared more often in sentences whose meanings were especially hard to infer based on animacy.

This simple experiment suggests that when the grammar of a language leaves room for systematic confusion about meaning, new learners are apt to "fix" it by providing extra information where it's most likely to be needed. It's not clear exactly how this happens. For example, it could be that people are somehow biased to learn linguistic systems that maximize communicative efficiency, so informative case marking is easier to learn than less informative case marking. Or, it could be that the speakers in this study were able to anticipate the potential for serious misunderstandings, and avert communicative disaster through their judicious use of case markers. It will take some careful experimentation to tease apart the various competing explanations.

12.3 Words, Concepts, and Cultures

Different words, different thoughts?

In the previous sections, I focused on the tantalizing similarities that turn up across languages, and delved into some possible explanations for why it is that languages often look so much alike. But human nature being what it is, we're often at least as fascinated by the *differences* among groups of people as we are by the things that all humans seem to have in common. Language offers endless opportunities for creative speculation about the nature of these differences. It's worth keeping in mind, though, that while it's easy to *assert* that speakers of a different language have deeply different ways of thinking, it's much harder to come up with a plausible theory for how this might happen.

A promising place to start is with the mapping of words onto concepts. As you saw in Chapter 5 (Language at Large 5.1), languages carve up conceptual space in quite different ways. The presence or absence of specific words in a language often serves as a magnet for cultural commentary. For example, in writing about the sexual shenanigans of former Italian prime minister Silvio Berlusconi, journalist Rachel Donadio (2011) couldn't help pointing out:

> It is not always easy to translate between Italian and American sensibilities. There is no good English word for "veline," the scantily clad Vanna White-like showgirls who smile and prance on television, doing dance numbers even in the middle of talk shows. And there is no word in Italian for accountability. The closest is "responsibilità"—responsibility—which lacks the concept that actions can carry consequences.

The Italian word ***responsibilità*** has a broad meaning, roughly lumping together meanings that English slices up into different words such as *responsibility, accountability, guilt,* and *liability.* In drawing attention to this fact, the writer implies that there is some connection between Berlusconi's antics and the Italian lexicon. The same connection has been drawn by other commentators in less subtle ways, as in the following from a blog post by economist Frederic Sautet (2006):

> Last week Graham Scott gave a lecture on public sector management and governance at the Mercatus Center. Dr. Scott was the Secretary of the New Zealand Treasury between 1986 and 1993, which was a very important position at the time of the NZ reform process. … During his lecture Graham Scott remarked that the word "accountability" has no translation in many languages. For instance, it has no direct translation in French and Spanish. I presume it is the same with other Latin-based languages, such as Italian or Portuguese. While the word "responsibility" is Latin in its origin (and thus has equivalents in French and Spanish and other languages), it encompasses more than just accountability and, for that reason, is much less precise. In Scott's view, the concept of accountability is at the core of the public management reforms in New Zealand. But its absence in many other languages may limit (and perhaps has already limited) the adoption of similar reforms elsewhere. Or it may lower the quality of their results. This would show the power of language in shaping institutions.

If you unpack such commentaries, you'll find that they're wrapped around at least one, and often both, of the following assumptions:

1. If a language doesn't have a word to convey a specific concept or to distinguish that concept from other similar ones, it's because its speakers don't care enough about that concept or particular distinction to devote a separate word to it.
2. If a language doesn't have a separate word for a specific concept, speakers of that language will have a hard time understanding the concept.

Let's hold these assumptions up to the bright glare of scientific sunlight, and evaluate how likely it is that they're correct.

Do words reflect culturally important concepts?

The first assumption—that the lack of a word reflects the lack of importance speakers place on the concept—is based on the premise that the relationship between concepts and the lexicon is not arbitrary. Presumably, if a concept is salient or important enough, speakers will invent a word for it. There has to be some truth to this notion. In Chapter 5, I discussed how children make reasonable guesses about the meanings of words by relying on natural ways to form categories. They have expectations about which aspects of the world are most likely to be talked about—talk about rabbits is more likely than talk about rabbit parts, or rabbit textures, or about the category that includes things that taste good in a stew.

It's likely that some conceptual distinctions are universally more salient than others, leading to some predictable ways of structuring the lexicon of any language. For example, in his book *Through the Language Glass*, linguist Guy Deutscher (2010) invites you to imagine coming across an old manuscript that describes a language called "Ziftish." In Ziftish, it turns out, there is a word *bose*, which is used to refer to white roses and all birds except those with red chests. Another word, *rird*, is used for red-chested birds and all roses except white ones. Deutscher wants to know: Do you, the discoverer of this manuscript, take it to be a factual diary of an early explorer? Or a fictional account—perhaps a long-lost sequel to *Gulliver's Travels*? The manuscript reeks of fiction, because it seems deeply implausible that a language would confer words on such unnatural categories.

This example suggests that salient or natural concepts attract words more readily than less natural ones. So what about when languages diverge in how they map concepts onto language? A reasonable hypothesis is that the divergence reflects how important these concepts are for particular communities of speakers.

It's not hard to come up with examples where certain lexical distinctions align neatly with the cultural importance of their corresponding concepts. Surely, a culture where all food is cooked on a spit over the fire or warmed up in the microwave has no need for a specialized vocabulary that distinguishes between words like *sauté, braise, grill, boil, bake, blanch, poach, broil, simmer, fricassee, flambé, steam, fry, caramelize, stew, sear,* and so on. When it comes to such words, necessity is a plausible mother of invention.

Perhaps the most striking example of how culture can drive the invention of words is in the domain of color vocabulary. Even within a single language like English, some speakers feel compelled to specify that a color is *magenta* or *chartreuse*, or to distinguish between *crimson* and *scarlet*, while others are perfectly satisfied with basic color terms like *blue, yellow,* and *red*. But if you think that all languages at least distinguish between the basic colors—much as they'd likely have separate words for general categories like birds or flowers—you'd be wrong. For example, it may surprise you to hear that color terms are virtually absent from the ancient Greek epics of Homer, as noted in 1858 by classics scholar William Gladstone (who is better known for having served four terms as the prime minister of Great Britain). This observation led Gladstone to speculate that the ancient Greeks were color-blind, and that humans have only very recently developed color vision as we have it today. This notion was taken seriously by many scientists of the time, even though some questioned whether the evolution of color vision could have developed in the short span of several thousand years of human history.

But in the twentieth century, as linguists and anthropologists began to comb through the existing languages of the world, they found many languages that didn't have separate words for green versus blue, for example, or even red versus yellow (see **Box 12.3**). The speakers of these languages weren't color-blind. They could *see* that green and blue were different colors, they just thought it would be odd to call them by different names—just as you might be able to see that two slightly different shades of red are different from each other, but wonder why anyone would need to have different words for them. Speakers of all languages, it turns out, can detect color differences that they don't bother to mark with different words. But language communities can differ widely in the number of distinct color words they feel are necessary. Many of the languages with very small color vocabularies are spoken in non-industrial societies that don't do much manufacturing of objects involving artificial color. In such a society, the color of an object is largely predictable from its inherent nature, so why would you bother to specify it?

To help you imagine what it might be like to live in a culture where a detailed color vocabulary seems unimportant, Guy Deutscher invites you to imagine a

BOX 12.3
Variations in color vocabulary

What could be more basic than color concepts like brown, green, yellow, and red? A survey of color terms across the world's languages shows surprising diversity in the number of color terms that are used in a language. In English, if we consider just those color terms that correspond to single, commonly used words, we have a total of eleven: *black, white, red, yellow, green, blue, purple, brown, orange, pink,* and *gray*. This represents the upper end of the vocabulary size for basic color terms across languages, as shown in **Figure 12.7**. Many languages make do with as few as three terms to express color.

Some languages use simple color words to make distinctions that we don't make in English. For example, Russian and Greek have separate words for dark and light shades of blue. But many other languages use a single term to name colors that English refers to by different names. For example:

Yupik (as spoken in Siberia) and Pirahã (Brazil; see Box 6.3) use a single word for green and blue.

Lele (Chad) and Javaé (Brazil) use a single word for yellow, green, and blue.

Gunu (Cameroon) and Tacana (Bolivia) refer to red and yellow with a single word.

Some researchers (most notably Brent Berlin and Paul Kay, 1969) have argued that there are certain important color-naming universals reflecting underlying perceptual constraints. As with linguistic structures, some color terms appear to be more common than others. For instance, if a language has only three color terms, the terms tend to separate into these categories: (1) white; (2) red and yellow; and (3) black, green, and blue. And, while it's common for languages to use a single word for green and blue, basic word distinctions between dark and light blue are rare.

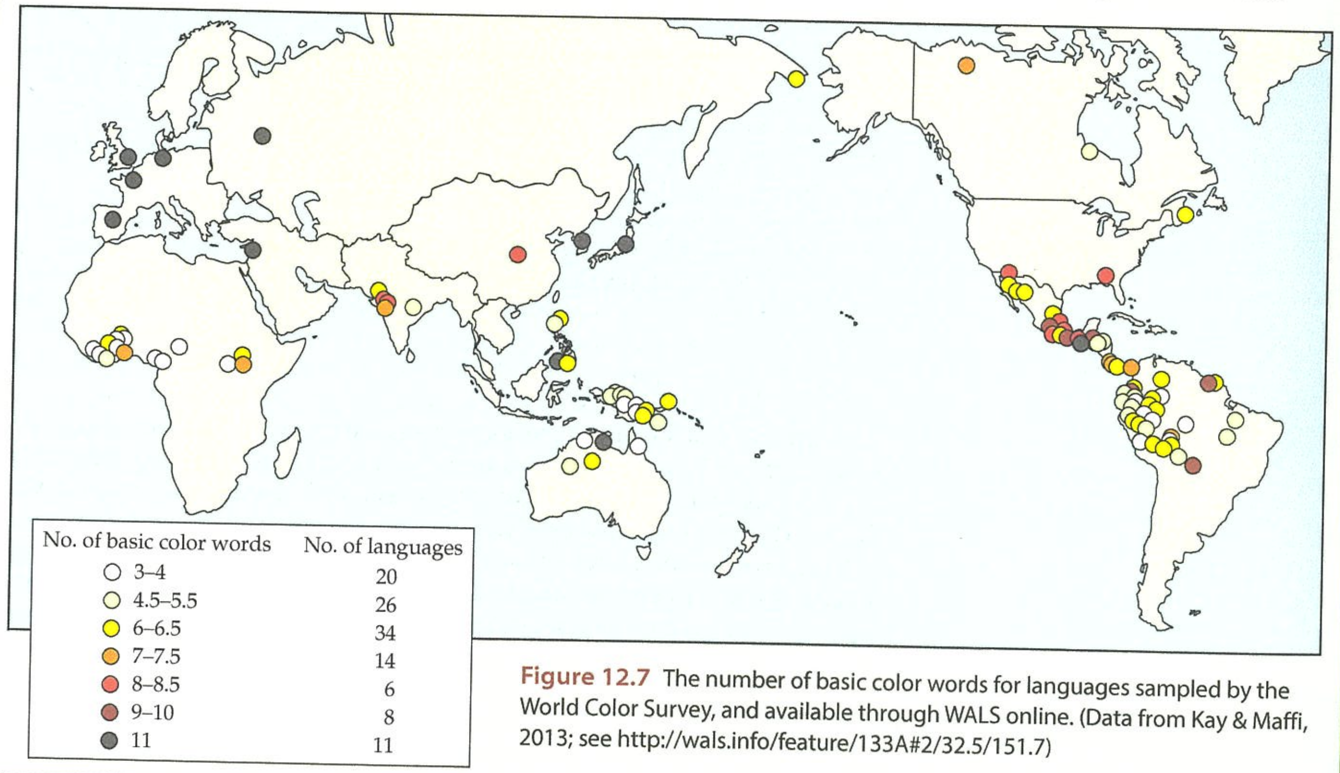


Figure 12.7 The number of basic color words for languages sampled by the World Color Survey, and available through WALS online. (Data from Kay & Maffi, 2013; see http://wals.info/feature/133A#2/32.5/151.7)