

# **When Languages Die**

The Extinction of the World's Languages and  
the Erosion of Human Knowledge

**K. David Harrison**

**OXFORD**  
UNIVERSITY PRESS

2007

“Do you find it easy to get drunk on words?”

“So easy that, to tell you the truth, I am seldom perfectly sober. Which accounts for my talking so much.”

—Harriet Vane and Lord Peter Wimsey in *Gaudy Night*, by Dorothy Sayers

“Studying the form of a linguistic expression without studying the meaning is like sipping a fine wine, swishing it around in your mouth, and spitting it out—it can be fun, but not intoxicating.”

—Randall Eggert, Linguist

All linguists share a fascination with words, and we are trained to seek out and describe intricate patterns within human languages. As languages rapidly vanish into the vortex of cultural assimilation, linguists justifiably fear they will never see the full range of complexity and structures human minds can produce.<sup>1</sup>

When Noam Chomsky proclaimed language ‘a window on the mind’, an entire research program for the discipline of linguistics was launched. In the fifty years since, this research has already yielded many important insights into human cognition. With his famous sentence “Colorless green ideas sleep furiously,” Chomsky demonstrated how linguists can explore complex structures (sounds, phrases, sentences, etc.) even when there is no meaningful content at all. The lack of meaning does not hinder linguists in our investigation of mental structures, who have come to focus mainly on the structures themselves, not their cultural meanings. This has been the conventional wisdom in linguistics for at least four decades.

But although languages certainly contain abstract structures, they evolve and exist to convey information within a specific cultural matrix, and that function permeates and influences every level of language. To its critics, including this author, the Chomskyan program has been unduly

narrow, overly focused on large, globally dominant languages, and preoccupied with structure at the expense of content.<sup>2</sup>

Linguists' preoccupation with these abstract structures (collectively termed 'grammar') has led to a microscopic approach that treats languages like laboratory specimens, utterly divorced from their natural environments, the people who speak them, and the content of those peoples' thoughts. As linguist Mary Haas pointed out, this approach hinders us in seeing the larger picture: "In their search for universal tendencies . . . some scholars have taken an atomistic approach. In other words, they have obtained examples of relative clauses, auxiliary verbs, the copula, and so on, from speakers (or grammars) of as many languages as possible without regard to anything else in the language." Endorsing a sensible alternative, Haas continues, "In the present climate of interest in the problem of language universals, we must not overlook the importance of the holistic approach. . . . A language must be understood and described as a whole. It is not a thing of bits and pieces, haphazardly strung together."<sup>3</sup>

Linguists who do field work on languages find it hard to ignore the rich cultural matrix or to examine things like sentence structure in isolation from the rest of the language. As soon as one looks at the content of language—what people care to talk about—it is obvious that this is also richly structured and a worthy object of study for any science investigating the mind. And it becomes obvious that structure may be grossly misunderstood if meaning is ignored. One example of this is the complexities surrounding how to say 'go' in Tuvan, as discussed in chapter 4. Without awareness of how speakers attend to ground slope underfoot, and to river current, it is hard to imagine even understanding that this system exists, let alone how it works. This small portion of Tuvan grammar depends on the human body's interaction with the local environment, as interpreted through Tuvan cultural norms. Such examples may be found for every language mentioned in this book, if we take care to look deeply enough.

One goal of this book is to advocate a restored balance between studying the structure of language and its meaningful content. We linguists have perhaps only a few decades left to document the lion's share of linguistic diversity before it vanishes forever. Endangered languages stand to play an increasingly central role in the study of the mind.<sup>4</sup> Any language, no matter how obscure or well-known, how large or how small, provides challenging patterns and complexities for linguists to describe. Even English, studied by hundreds of linguists for hundreds of years, has yet to

yield all its secrets to science. As with word order in Urarina (discussed in chapter 1), an obscure fact from a language spoken by just a few dozen people can take a well-established scientific theory and turn it completely upside-down. Since we cannot know the goals and tools science will have fifty or one hundred years hence, we must aim for the fullest description possible of each language now.

### Language Change Just Happens

Languages are highly complex, self-organizing systems in constant flux. The English spoken by our great-great-great grandparents, who might have used a word like ‘hither’, is very different from how we ‘converse’ nowadays. Geoffrey Chaucer could not chat with Bill Gates. We all participate in constant change, but no individual speaker controls the speed, trajectory, or character of change. A process of emerging complexity—not yet well understood—gives a language its constantly changing and characteristic shape.

Individual speakers of any language can and do make up new structures on a whim, by slip-of-the-tongue, or through creativity. Rap singers’ terms ‘b-iz-itch’ (or ‘biznitch’ or ‘biznatch’) or cartoon character Homer Simpson’s ‘saxa-ma-phone’ and ‘platy-ma-pus’ are examples of recently invented speech play.<sup>5</sup> These innovations only become part of the language by a mysterious process of social learning and consensus. Other speakers must adopt and use (and perhaps revise or expand upon) these new ways of talking. At first, purists may denounce such changes as ‘bad English’. But if the changes endure, dictionary writers and grammar teachers eventually catch up and acknowledge such innovations.

Besides consciously creative innovations, many changes take place of which speakers are unaware. Californians whose grandparents pronounced the words ‘cot’ and ‘caught’ differently now pronounce these words the same. Somewhere along the line they lost an entire vowel. Nobody decided to jettison it, it just happened. Eastern U.S. speakers who maintain the ‘cot’/‘caught’ distinction may find this vexing, leading to misunderstandings. (When I listen to people who lack that vowel, I often wonder, did they mean ‘sot’ or ‘sought’, ‘hottie’ or ‘haughty’, ‘body’ or ‘bawdy’? For me, and speakers who share my set of vowels, these paired words all sound unambiguously distinct.)

People also unconsciously change their own speech habits even over the course of a lifetime. We adopt new words like ‘phat’, ‘metrosexual’, ‘pizzled’, new expressions like ‘twenty-four seven’, and we may even shift our pronunciation. Queen Elizabeth II’s speech has changed noticeably in the fifty years since she ascended the throne. Measurements of her vowels in her annual Christmas radio speeches showed that from the 1950s to the 1980s she shifted noticeably away from the “Queen’s English” and towards pronunciations favored by the lower social classes.<sup>6</sup>

Nobody directs this intricate process of language change, on its individual or group levels—it is an orchestra without a conductor or even a musical score. There is no central decision-maker or authority, but orderly change happens nonetheless. Like complex termite mounds that get built with no blueprint, architect, or foreman, language is a self-organizing system. It has many distinct parts that interact in complex and often unpredictable ways, resulting in surprising and unplanned patterns.

No schoolteacher, committee, or lexicographer has authority to decide whether ‘biatch’<sup>7</sup> or ‘puhleeze’<sup>8</sup> counts as a word of English or not. If English speakers use such words widely enough, they become part of English. This is true of new meanings for old words (‘spam’ used to mean canned meat, now it means unsolicited e-mail), new coinages (‘e-commerce’, ‘conversate’), borrowings (*jihad* from Arabic, *perestroika* from Russian), and even new grammatical constructions.

### **Are All Languages Equally Complex?**

It has become almost dogma in linguistics to affirm that “all languages are equally complex.” This statement is usually followed by “and capable of expressing any idea.”<sup>9</sup> The second idea is logically separate from the first. Any language can indeed express any concept or idea that its speakers care to talk about—this is a testable hypothesis. So while it is uncontroversial that all languages possess equal expressive potential, at the level of structures languages do differ widely. Once the equal complexity model is adopted, a number of further assumptions follow, for example: “A language which appears simple in some respects is likely to be more complex in others.”<sup>10</sup> This often popularly construed as the notion that if a language simplifies one part of its grammar it necessarily gains some complexity elsewhere, as if regulated by a thermostat.

Such claims are problematic, if only because they remain hard to test empirically. Most of the world's languages remain undescribed or underdescribed. We lack any agreed-upon unit for measuring complexity,<sup>11</sup> especially *across* distinct domains such as vowel pronunciation and sentence building. And complexity arises from many disparate factors, starting certainly with the innate ability of the human brain, but also including the size of the speech community, the level of contact among speakers, the range of uses of a language, the modality (spoken or signed), and intricate historical processes of language change.

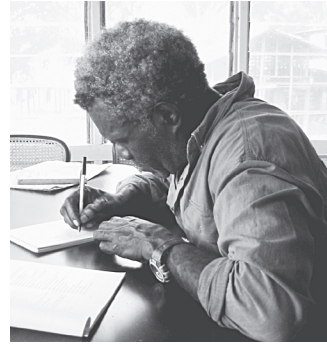
Yet one finds the 'equal' complexity idea in textbooks, blogs, introductory linguistics classes, and the like.<sup>12</sup> As evidence, it is noted that any neurologically normal human child can learn any human language when raised among people speaking that language. An Icelandic child raised by Swahili parents will come to speak flawless Swahili, and vice versa. Studies comparing acquisition rates of children learning different languages show slight differences for certain kinds of structures, but all kids still all turn out to be fluent speakers of their native tongue by age 7 or so.<sup>13</sup>

The sentiment behind this argument is noble: of course, we should not regard any other people or culture as primitive or any more or less intelligent than ourselves. Ultimately, statements about the equal complexity of languages may owe more to political correctness than they do to any empirical evidence. However, a fundamental quantitative problem with the claim remains: we have no established way to measure complexity within a single language or across multiple languages. Further, if the scope of our investigation is narrowed to certain parts of a language (say, only



**Figure 7.1**

Timothy Taureviri, a speaker of Central Rotokas who has worked with linguists like Stuart Robinson for years, transcribing his language. Courtesy of Stuart Robinson



**Figure 7.2**

Rotokas villagers from Togarao taking a rest from performing *singsing kaur*, a traditional song and dance performance with bamboo flutes. Courtesy of Stuart Robinson

sounds or only word-structure), certain languages appear vastly more complex in specific areas than do others.

One example comes from Phonology, or the organization of sounds in language. Rotokas (spoken in New Guinea by 4,320 people) reportedly gets by with a mere six consonants: *p*, *t*, *k*, *v*, *r*, and *g*, while Ingush, a language of the Caucasus (230,000 speakers) boasts a whopping 40 consonants.<sup>14</sup> Besides many common sounds like ‘p’, ‘b’, and ‘f’, Ingush uses a special series of ejective consonants that are produced by closing and raising the vocal chords to compress air inside the pharynx, then releasing the pressure suddenly to create a popping sound to accompany the consonant. Ejectives are moderately rare, occurring in only about 20 percent of the world’s languages. To employ *seven* distinct kinds of ejectives, as does Ingush, is exceedingly rare.<sup>15</sup> But even Ingush is not the upper limit: Ubykh, which reportedly had 70 consonants, lost its last speaker in 1992.<sup>16</sup>

Rotokas, which may have as few as six consonants, is by no means a simple language. On the contrary, Rotokas crams entire utterances into single words. The following 13 syllables comprise just a single word, with hyphens inserted here for readability (notice the reduplicated form of the verb form *rugo* ‘to think’, in boldface).<sup>17</sup>

*ora-rugorugo-pie-pa-a-veira*  
‘They were always thinking back.’

As evidenced in these words, Rotokas has simple syllable structure, allowing only one vowel and a maximum of one consonant per syllable. Ingush appears more complex, allowing multiple consonants to sit next to each other, for example, *bw*, *hw*, *ljg*, and *rjg*.<sup>18</sup>

*bwarjg*      ‘eye’  
*hwazaljg*    ‘bird’

So Rotokas and Ingush show two very different kinds of complexity in their sound systems. Ingush has 40 consonants, somewhat shorter words, and allows multiple consonants per syllable. Rotokas has few consonants, permits only one consonant per syllable, and builds very long words. Each language has complexity of a different type and in a different area of its grammar.<sup>19</sup> These are apples and oranges; we cannot yet sensibly pose or answer the question of which system is more complex. Further, it suggests that such an egalitarian position would be meaningless.

Setting aside the controversy over equal complexity, professional linguists would probably all agree on the following. If we took a survey of *only* the world’s 100 biggest languages, we would not only miss many unique complexities found in smaller languages and thus present in human language in general—but our very notion of what human language is would be severely skewed.

Imagine a zoologist describing mammals by looking only at the top hundred most common ones. It would be easier to examine dogs and cats and cows and rabbits, all of which are composed of the same building blocks as other mammals. But if we did, we would never know that a mammal could swim (whales), fly (bats), lay eggs (echidna), use tools (sea otters and orangutans), or have an inflatable balloon growing from its head (male hooded seal).<sup>20</sup> Ignorance of unusual mammals would impoverish our notion of what mammals can be. It is precisely the weird and wonderful exceptions that afford us a full view of the possibilities.<sup>21</sup>

### Complexity Run Amok

Small languages whose grammars seem otherwise average or run-of-the-mill often contain islands of astonishing complexity. While all languages may look more or less complex from a distant, bird’s eye view, upon closer inspection we find particular areas of some languages’ grammars that seem to have run amok, stretching the very limits of complexity. This does not demonstrate that some languages are *on the whole* more complex than others, but it certainly opens the door for us to pose the question. Since grammars are shaped by culture and environment, as well as by human



brains, and are constantly changing, they might plausibly vary within the limits of what intelligent human brains require of them.

In this chapter, we present some impressively complex sub-systems of the grammars of small and endangered languages. And we argue that, were these systems to vanish undocumented, we might never imagine or suspect their possible existence. We would thus remain ignorant of some types of linguistic complexity that can arise. Because they arose over long periods of time in unique conditions (and owe something to random chance), such systems would not likely reappear in the subsequent future course of linguistic change. Lacking knowledge of tongues like Tabasaran, Rotokas, Sora, Gros Ventre, or Yanyuwa, we are deprived of unique insights into human cognition and the upper bounds of linguistic complexity.

In this chapter, we look at some linguistic complexities rich enough to intoxicate any language lover. For non-linguists, all examples are explained clearly and compared to English or other widely spoken languages. We will consider what they may tell us about human cognition and the self-organizing system known as language that has colonized our brains. How far can it go? What kinds of fantastic structures does it build?

Answering this question has long been the prime directive of linguistics. As Noam Chomsky eloquently put it:

Language is a mirror of mind in a deep and significant sense. It is a product of human intelligence. . . . By studying the properties of natural languages, their structure, organization, and use, we may hope to learn something about human nature; something significant, if it is true that human cognitive capacity is the truly distinctive and most remarkable characteristic of the species.<sup>22</sup>

Many of Chomsky's intellectual heirs have interpreted the directive narrowly. They investigate some small sub-part of language structure, often paying scant attention to the intellectual and cultural content of what people are actually saying. I have tried to demonstrate in this book that many kinds of linguistic knowledge, such as when to say 'go' in Tuvan (chapter 4), cannot be properly understood or described if divorced from their social and physical environment. In this chapter, I will show that many of the kinds of structures Chomsky and his followers have been interested in are to be found only in small, obscure, and endangered languages.



**Figure 7.3**

Galina Innokentovna Adamova (1924–2001), shown here on her funeral bier in June 2001. Among the last fluent speakers of Tofa, she worked with me to record Tofa songs and narratives. Photograph by Thomas Hegenbart, courtesy of Contact Press Images

## Smelly Talk

Starting with a very simple example, we will look at a single morpheme in Tofa, the language of Siberian reindeer herders discussed in chapter 2. A morpheme is the smallest meaningful building block in language that may be used to build a complex word. In English, the word ‘sing’ is one morpheme, and the suffix ‘-able’ is another kind of morpheme that attaches to it to build the word ‘singable’. Many morphemes, like ‘-able’, never stand alone as words, but can be added to other words to change their meaning.

Tofa has a morpheme that speakers can add to any noun. It changes that noun into an adjective meaning ‘smelling of’ or ‘smelling like’. So if we take the word *ivi* ‘reindeer’ and add the olfactory suffix *-sig*, we get a new word *ivisig* that means ‘smelling like a reindeer’. The smell suffix has not been reported for other languages, though it certainly might exist elsewhere. And we can only guess as to why smelliness was regarded as important enough to Tofa culture that their language evolved a unique morpheme to signal it.

## Sound Talk

Moving on to a more complex case, let us take a look at Tuvan, the language of nomadic yak herders discussed in chapters 3, 4, 5, and 7. Tuvans spend a great deal of their time hunting and herding animals in the mountain landscape. They seem to have a heightened sensitivity to sounds, especially animal sounds, nature sounds, and the natural acoustic properties of outdoor spaces (types of echoes). Their sound aesthetic is partly reflected in the art of “throat-singing” or “overtone singing” that has made them world famous.<sup>23</sup> But on a more day-to-day basis, Tuvans who hunt and herd animals show superb abilities to mimic natural sounds. They use this ability to sing to the yaks to calm them, to call wild boars while hunting, to imitate bird and marmot sounds, and to tell playful stories involving animals.

The Tuvan language, not surprisingly, has evolved a very rich vocabulary to describe and imitate natural sounds. Of course, all languages have onomatopoeia: English has words like ‘sizzle’, ‘bang’, and ‘rustle’, all giving an imitative sense of the actual sound. But English speakers cannot really make up a new onomatopoeic word on the fly and be understood. If I want to describe the sound of a cow chewing its cud, I might say ‘munching’ or ‘chomping’, but I cannot just invent a whole new word, say, ‘flarping’, and expect to be understood. Tuvan speakers *can* do this. Their language allows them to describe a very wide range of natural sounds, using both ready-made words and newly coined ones. Tuvan provides means for speakers to creatively make up brand new words to represent sounds and be immediately understood by others.

It works like this. Pairs of consonants in Tuvan represent classes of sounds. For example, a word with a *k* and *ng* (as in English ‘king’) would represent a metallic ringing of impact sound. The speaker can fill in different vowels: high vowels to represent high-pitched or rapid sounds, low vowels to represent low-pitched or slow sounds, and so on. *Kongur* is the sound of a big bell ringing or a large metal pipe striking an object. *Kingir* or *küngür* would be jangling stirrups or clanging keys, while *kangyr* might be a giant empty metal barrel rolling along. With eight vowels, Tuvan provides many possible combinations, and speakers can use and understand most of these combinations, even if they have never heard them used before.

For example, if you hear someone blowing their nose or the sound of water in a babbling brook, you might use or create a word with the consonants *sh* and *l* combined with various possible vowels:<sup>23</sup>

<i>šülür</i>	sound of a nearly dried up river, or sound of mucous (snot) being forcefully blown out of the nose.
<i>šölür</i>	sound of a bundle of wood falling loudly, or sound of loud slurping
<i>šalyr</i>	sound of dry leaves or grass rustling
<i>šölür</i>	sound of water in a babbling brook
<i>šylyr</i>	sound of rustling (e.g., paper in the wind)
<i>šulur</i>	to chatter or blab
<i>šilir</i>	(this word does not exist, but when asked, native speakers reliably report it has something to do with water sounds)

Tuvan thus equips its speakers with an unusually complex, combinatory system for expressing and representing sounds.<sup>24</sup>

We do not know the full extent of sound symbolic words in the world's languages. A similarly rich and expressive system was documented by linguist Martha Ratliffe in White Hmong (about 500,000 speakers), where *mis mos* is the “sound of cows or horses pulling up grass,” *mig mog* denotes “dogs fighting over a bone,” *plij plawj* “pigeons flying or dry husks falling off bamboo,” *nphis nphoos* the sound of a “drip from a pipe into a tank,” *mlij mloj* the sound of “two separated cats meowing before fight[ing],” and *rhiv rhuav* imitates “people shuffling through dry leaves with force.”<sup>25</sup> While onomatopoeia is known to exist in all languages, few documented ones have shown such rich possibilities as Tuvan and White Hmong.

### **Willy-nilly Talk**

Nearly all known languages have processes for building doubled words like ‘flim-flam’, ‘helter-skelter’, or ‘money-schmoney’. Often in such paired words, the individual parts have no meaning (what is a ‘flim’ anyway, much less a ‘flam’?), but they take on meaning as part of a whole. Sometimes only one half of a doubled word is a real word, like ‘fiddle-faddle’, and sometimes both have meaning, as in ‘flip-flop’. The words may differ depending on the language, but this process—which linguists dub ‘reduplication’—pops up predictably and in subtly different forms in languages all over the world.

Scientists have collected examples of reduplication from over a thousand languages, seeking large-scale patterns and similarities in form or meaning.<sup>26</sup> Many languages use reduplication much more often than English does, but we are far from knowing the full range of possible patterns.

Not surprisingly, reduplication often signals repetition of an action or event. In Tuvan, the verb *halyr* means ‘to run’; the doubled form ***halyr-halyr*** means to run repeatedly, over and over. Reduplication can also add emphasis or intensity to a word. The Tuvan word *kyzyl* means ‘red’ and *borbak* means ‘round’. These same words, when partially reduplicated as ***kyp-kyzyl*** and ***bop-borbak***, mean ‘intensely red’ and ‘completely round’.

Rotokas, the language with so few consonants, takes simple noun or verb and doubles part of it to express greater quantity or frequency (reduplicated portions of words are boldfaced below).<sup>27</sup>

<i>tapa</i>	‘to hit’	>	<b><i>tapatapa</i></b>	‘to hit repeatedly’
<i>kopi</i>	‘a dot’	>	<b><i>kopikopi</i></b>	‘spotted’
<i>kavau</i>	‘to bear a child’	>	<b><i>kavakavau</i></b>	‘to bear many children’

There are many more ways languages build reduplicated forms, but they tend to add the same kinds of meanings: repetition, intensity and emphasis. But a most unusual and unexpected use of reduplication is found in Eleme, a language spoken by 58,000 people in Nigeria. Eleme speakers double part of a verb in order to *negate* it.<sup>28</sup>

<i>moro</i>	‘He saw you.’	>	<b><i>momoro</i></b>	‘He didn’t see you.’
<i>rekaju</i>	‘We are coming’	>	<b><i>rekakaju</i></b>	‘We are <b>not</b> coming.’

Without knowing Eleme, linguists might never have guessed that the fairly common mechanism of reduplication could take on such an unusual function—one in which more quite literally means less.

### Touchy-feely Talk

If asked what clams, buttons, and frisbees had in common, you might say they are all basically flat and round, even though they differ in so many important ways (one is alive, one has holes, one flies, etc.). At some abstract level of thought, it may make sense to lump them together. Many languages—called classifier languages—do just that, by assigning every noun to one of several abstract categories. Of course, millions of sub-

stances, colors, smells, and tastes exist, and they combine in infinite ways in natural objects. So how many categories do we sensibly need and what are they?

The Carrier language, spoken by 1,500 people in Canada's British Columbia, employs a classifier system that forces speakers to pay attention to tactile and other qualities of objects. In Carrier, you cannot usually just say 'I gave.' What is the object given? Is it small and granular? Liquid in an open container? Mushy? Fluffy? Two-dimensional and flexible? Long and rigid? Depending on these shape and tactile qualities, which determine how your hand would grasp the object, an entirely different verb form must be used.<sup>29</sup>

Of course, English speakers experience objects in a tactile way, too, and we are aware of their physical properties. But English does not force us to pay attention to these qualities each time we refer to an object. They are there to be described if we choose, but most often we simply say 'give'. Languages can force their speakers to pay attention to certain aspects of the world, thus shaping how people think. Clearly, there is no universal set of categories or ways to divide up the natural world; just try to get people to agree whether a tomato is a more like a vegetable or a fruit, or a dolphin more like a fish or a cow. For speakers of classifier languages, certain

'he gives me'

*sgatodzih*

*sgantadzih*

*sgadutel*

*sgatikal*

*sgatilchus*

*sgantaldo*

an object like

  
(sugar)

  
(blueberries)

  
(stick)

  
(tea in a cup)

  
(shirt)

  
(fluff)

**Figure 7.4**

Speakers of Carrier use very different forms of the word "give" depending on the tactile properties of the object being given.

subtle similarities among objects may be made more readily apparent because they are built into in the very grammar and talked about on a daily basis.

Cantonese, with 52 million speakers, sorts all objects and entities into classes. Native speakers sometimes disagree about what falls into which category. They even debate on Internet discussion boards which classifiers apply to certain objects and proffer examples to clarify usage:

“Use JX *faai* (with the 3rd tone), for large, flat, sometimes hard objects (cookies, boards, individual leaves).”

“Use , *gau* (with the 6th tone), for small pieces of edible food and other chunks of smallish things (cake, buns, many of the dim sum foods, rocks and boulders, unidentifiable chunks of things).”<sup>30</sup>

Cantonese classifiers *must* be used whenever you use numbers, for example, you cannot say ‘five rocks’ in Cantonese without inserting the appropriate classifier word between the number and the object name: ‘five *gau* rock’ or ‘two *faai* leaf’. If the language were to vanish—an unlikely scenario for Cantonese, but a looming threat for almost every other language I’ve discussed—our understanding of how the human brain can categorize objects would be impoverished. We might never know a way of viewing the world in which cookies and leaves fall together or dump-lings group with boulders.

English sometimes uses a type of classifier for nouns that are not countable: a ‘pile’ of sand, a ‘glass’ of milk, an ‘expanse’ of water. Your choice of what classifier to use is flexible but constrained. You may say a ‘cup’ of sand or a ‘pool’ of water, though you cannot sensibly say a ‘pile’ of water. But classifiers make up a limited system in English and we can get by without them. We do not yet know how many of the world’s languages employ classifiers and how complex these systems may be. Some very small and endangered languages have classifier systems of great complexity, but that divide up the natural world in very different ways than does Cantonese. Yupno, for example, spoken in Papua New Guinea, rigidly classifies everything in the world into one of three states: ‘hot’, ‘cold’ and ‘cool’.<sup>31</sup>

Nivkh, a Siberian language with under 300 speakers, has a highly complex classifier system that applies only in Nivkh numbers. Nivkh may once

have resembled Cantonese in that the number word came first, then the classifier word, then the object. But Nivkh classifier words no longer stand alone: they have become a part of the number word itself. In what may be the most elaborate counting system yet known in any language, Nivkh uses 26 distinct number series. Each series is limited to a special object or class of objects. Nineteen of the classifiers apply only to very specific objects, such as boats, sleds, fishing nets, skis, finger widths used to measure the thickness of animal fat, and batches of dried fish. Six other classifiers apply to classes of objects united by some common, abstract property:

<i>common quality</i>	<i>examples of objects</i>
come in pairs	eyes, ears, hands, legs, boots, mittens
small and roundish	nuts, bullets, berries, teeth
thin and flat	leaves, blankets, shirts

The twenty-seventh Nivkh classifier is for odd objects that do not fit into any class.

Another special use of numbers to classify things is found in Native American languages belonging to the Salish family, spoken in the Pacific Northwest. These languages adjust their words in special ways to signal what is being counted. The adjustment involves taking a part of the word and repeating it, using ‘reduplication’ as discussed earlier in this chapter. By analogy, if I said ‘twenty’ or ‘fifty’ while counting objects, imagine that in Salish I would have to say ‘**twe**-twenty’ or ‘**fi**-fifty’ for counting people. Salishan languages have special number forms for animals, so a three-way classification pits objects vs. people vs. animals. Again, we could imagine a special form in English, where ‘fifty-**ifty**’ and ‘twenty-**enty**’ might indicate that animals are what is being counted.

Salish languages also adjust the words for ‘what?’ and ‘how many?’, so you can always tell if a Salish speaker is asking about numbers of animals, things, or people. Imagine if in English ‘**ho**-how’ meant ‘How many people?’ while ‘how-**ow**’ meant ‘how many animals?’ The chart table 7.1 shows some counting words from Squamish (15 speakers), a language of the Salish family. We use special phonetic characters to represent some unusual Squamish sounds, but what is important here is to notice the boldfaced parts of the words that have been reduplicated. There are three patterns. For counting objects, just the basic form is used. For people, a rather large chunk (or all) of the basic form is reduplicated, while for animals a smaller chunk is doubled, plus the original word may lose a vowel.<sup>32</sup>





**Figure 7.5**

On the left, a speaker of Nivkh, photographed in Siberia in 1898–1899. On the right, contemporary Nivkh speakers Sergei and Natasha Firun with their two children in the town of Liugi, on the northwest coast of Sakhalin Island, June 1990.. Courtesy of the American Museum of Natural History (left), and courtesy of Bruce Grant (right)

Looking at just five languages that classify objects (English, Cantonese, Carrier, Squamish, and Nivkh), we can see how such systems divide up the world in radically different ways. They impose categories of shape, dimensionality, and animacy onto objects and force speakers to attend to these properties of the world, for example, whenever they wish to count. English imposes a minimal burden; we can simply use the number ‘two’ for any pair of objects. If we want to signal some kind of special unit, we might say ‘pair’, ‘twosome’, ‘couple’, or ‘duo’, but such uses are rare. And we can always just say ‘two shoes’.

Speakers of Nivkh or Squamish, by contrast, must know the proper class of objects in order to count them.<sup>33</sup> Counting is an area where dif-

**Table 7.1** Examples of reduplicated numbers in Squamish

	<i>Counting objects</i>	<i>Counting people</i>	<i>Counting animals</i>
9	<b>c'əs</b>	<b>c'əs-c'əs</b>	c'ɪ-c's
10	<b>ʔupn</b>	<b>ʔəp-ʔúpɪn</b>	<b>ʔú-ʔpɪn</b>
How many?	<b>kʷin</b>	<b>kʷin-kʷin</b>	kʷɪ-kʷin







ferent languages can impose complex classification systems and thus increase both the cognitive task (sorting things into categories) and the amount of information hidden in simple counting.

Languages like Carrier, Nivkh, and Squamish each force a speaker to pay attention to some particular aspect of the world around them and then encode this information in the grammar of everyday talk. Scientists still do not know the possible range of such systems, and this limits our understanding of the interface between grammar, the human body, and the environment. To what extent can a language encode and make mandatory in its grammar information about physical objects in the world? Urgent studies of small and endangered languages will be needed to complete the picture.

### Information Packaging

Languages contain and package ideas in a way that few other media can. Of course, you can express lots of ideas without language: ever play charades? See a stone cross in a cemetery? Hear a Chopin sonata? All these symbolic media express ideas without language. But language is so much more efficient. That is why we have textbooks in schools rather than teaching biology through the medium of dance, song, and charades, nor (usually) by sending students out to observe animals (though observation and dissection can complement biology lessons). Humans rely first and foremost on language because it is the most compact and efficient channel for transmitting ideas.

If you think this is just a matter of names or labels for things, you have underestimated the vast efficiency of information packaging that goes on in language. We take this entirely for granted! If I say ‘my nephew’ in English, what information is encoded? You know I am talking about a male

	Nivkh	Squamish	English
 people	men 'two (people)'	ʔnʔanʔus 'two people'	two (of anything)
 leaves	merakh 'two thin flat things'	ʔanʔus 'two things'	
 skis	mirsh 'two skis'		
 batches of dried fish	mer 'two batches of dried fish'		
 boats	mim 'boats'		
 animals including fish	mor 'two (animals)'		

**Figure 7.6**

A comparison of six of Nivkh’s twenty-six distinct number categories, all of Squamish’s three number categories, and the one English category.

person, and you know that he is related to me by blood. But here it gets more vague. Is he older or younger than me? Unclear. Is he the son of my sister or my brother? Unclear. Is he the son of an older sibling of mine or a younger sibling? Unclear. Is he a boy or a man? Unclear.

The English word ‘nephew’ reflects a set of tools (kinship terms) we use to define social relations. They also reflect our society’s decisions about what information to include and what to leave out. These decisions are

made not by individuals with executive powers, but by tacit consensus within a speech community about what is worth labeling and noting. Evolving over centuries, and by a mysterious process we do not yet understand, word labels are no less real or effective because of how they came about.

Different societies have traversed very different decision paths in constructing their social reality, and maintaining or changing their kinship terms. It is no surprise then, that corresponding to the single English word ‘nephew’, many languages have a much larger repertoire of more specific terms. Rotuman (9,000 speakers) has a highly complex set of kinship terms, with unique words, for example, denoting ‘elder son of elder brother’, ‘younger son of elder brother’, ‘elder son of younger brother’, or ‘younger son of younger brother’. These terms help reinforce a legal framework for enforcing inheritance and land tenure in Rotuman society.<sup>34</sup> Linguistically, the result is a highly compact, highly efficient system of knowledge that packs multiple bits of information into small spaces. The more information there is in a label, the less inductive reasoning or context-based inference is required.

The less information there is in a label, the more the brain must work to construct general categories. Highly abstract terms can be harder to learn: for example, ‘vegetable’ in English includes a very wide range of things, ranging from spherical and purple (beets) to long and green (scallions). You might be well into adulthood before you learn that a previously unknown item (okra) falls into this class, or that a long-familiar one (tomato) does not. A large and diverse class of objects grouped under one label can be hard to learn. Similarly, a very narrow and specific label can be hard to learn. Tuvan has a special kinship term that means ‘the two wives of my two brothers’. If you have three brothers, or only one brother has a wife, the term never applies. The word applies only to a specific sibling and spouse scenario, but is also quite abstract and rarely used.

Linguistic labeling systems do seem to follow a certain logic. For example, there are no known systems that call all yellow objects ‘blue’ on Tuesdays but ‘yellow’ the rest of the week. And there are no known labels that denote an animal and its tail (though many languages use the same word for an animal and its edible meat, or an animal and its pelt, or a tree and its fruit, or younger brothers and sisters collectively). The logic of information packaging in linguistic labels only vaguely mirrors natural categories out there in the world. More often, it imposes socially con-

structured and culturally specific categories. The possible range of such systems provide important insights into how language mediates and shapes or is itself shaped by human perception of and interaction with the world.

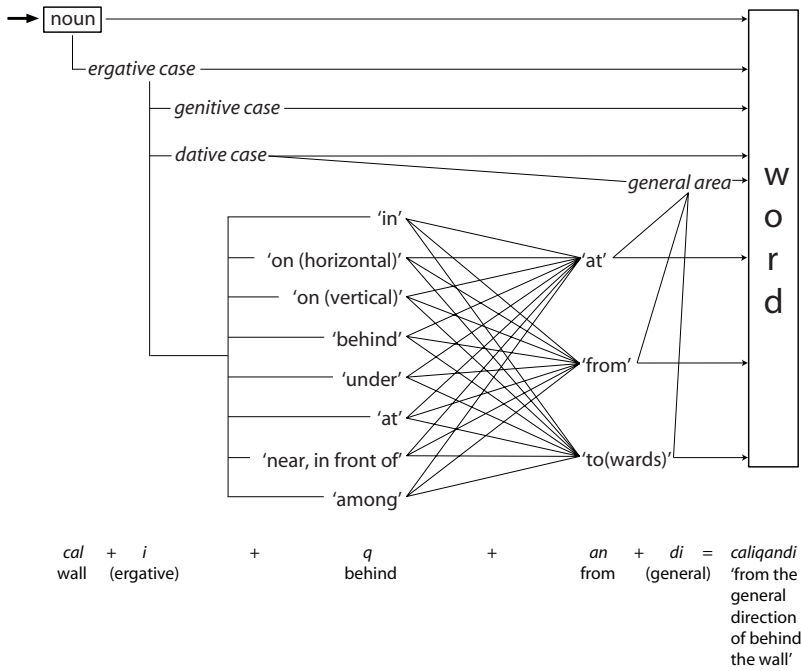
### World Record Languages?

Anyone who has tried to learn classical Greek or Latin will be familiar with the so-called case system as manifested in endings that are added onto nouns, pronouns, or adjectives. Latin has six cases: nominative, genitive, dative, accusative, ablative, and vocative, plus remnants of a locative case. Cases indicate relations among words. Latin *puella* means ‘girl’, but *puellae* means ‘for the girl’ or ‘of the girl’ (the ending *-e* signals the word is in the dative case). We know many languages get by with no case at all, while others have very complex systems. Mandarin has no cases, English has only a residue of earlier cases, apparent in differences in pronouns like ‘him’ vs. ‘he’ vs. ‘his’. Russian has six cases, while Finnish has at least 14. But it is not yet known how much complexity is possible in a case system, or the full range of word-to-word relationships that may be signaled by case endings.

Two languages spoken in the Caucasus mountains of southern Russia show very rich case systems, perhaps far in excess of other languages. Tabasaran, spoken by 95,000 people, even got listed in the 1997 *Guinness Book of Records* as having the most (52) cases. It turns out this number may have been a bit inflated by enthusiastic linguists. Nonetheless, Tabasaran and the nearby Tsez (spoken by 7,000 people) both have case systems of astonishing complexity.<sup>35</sup>

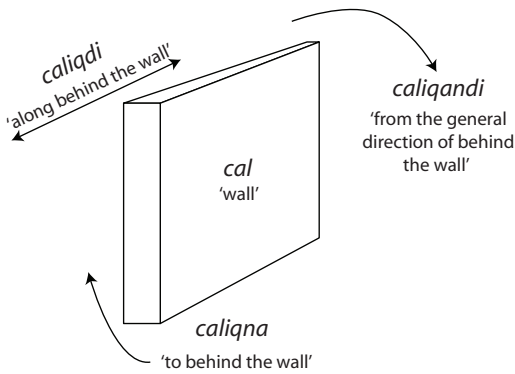
The question of exactly how many is one we will leave to the experts. Linguist Bernard Comrie points out that a basic distinction needs to be made in Tabasaran between ‘core’ cases (which can attach directly to a noun) and ‘non-core’, which can only attach after another case suffix is already present. He notes that while Tabasaran cases have probably been overestimated, there is still a large number of possible combinations of multiple suffixes, each with a unique meaning. A Tabasaran noun may have up to 53 distinct forms, once you add case suffixes specifying location and movement of objects in relation to that noun.

The following examples show a Tabasaran word with multiple possible case suffixes:



**Figure 7.7**

A flow-chart showing how to build a complex word using the Tabasaran case system. Starting with a noun in the upper left, follow one of many possible paths to add one or more suffixes that encode spatial and other meanings.



**Figure 7.8**

Location, direction, and motion expressed by complex Tabasaran nouns with case suffixes.

<i>cal</i>	wall
<i>cal-i</i>	wall (+ ergative case)
<i>cal-i-k</i>	on the vertical surface of the wall (+ ergative + spatial)
<i>cal-i-q</i>	behind the wall (+ ergative + spatial)
<i>cal-i-q-na</i>	to behind the wall (+ ergative + spatial + motion)
<i>cal-i-q-an</i>	from behind the wall (+ ergative + spatial + motion)
<i>cal-i-q-an-di</i>	from the direction of behind the wall (+ ergative + spatial + motion + general)
<i>cal-i-q-di</i>	along/across behind the wall (+ ergative + spatial + general)

### Word, Interrupted

Most languages (except for signed languages, as we shall see) require their words to appear as distinct sequential units. It is rare in spoken language to pronounce part of a word, then stop and insert another word, then go back to finish the first word. English has a few special cases like ‘fan-*fuckin*-tastic’ or ‘whoop-dee-*damn*-doo’, but speakers cannot just insert any old words wherever they please.

In Eastern Arrernte (2,000 speakers in Australia), many words can appear inside of other words. The Arrernte word for ‘sitting down’ is made up of three parts, a verb and two suffixes:

verb	<i>arrern</i>	‘to place’
suffix 1	<i>-elh</i>	(indicates an action done to oneself)
suffix 2	<i>-eme</i>	(indicates present tense)

Stinging these together yields a long word *arrernelheme* meaning ‘(he or she) is sitting down’. If you want to say, ‘She is *supposedly* sitting down’, you can insert the word *akwele* (‘supposedly’) inside the verb, producing *arrerneakwelelheme*. Notice that the word *akwele* inserts itself not only inside the word, but right in the middle of suffix 1 (*-elh*), not respecting any neat boundaries between morphemes. Optionally, you can also leave the word *akwele* outside the verb, but Arrernte provides the unusual possibility of nesting words within words.<sup>36</sup>

In Sora (288,000 speakers in eastern India) many words canglom together into a single one. Sora produces astonishing words like *kung-kung-*

*deduu-boob-mar* (this is all one word, with hyphens inserted for readability) meaning ‘a man with a clean-shaven head’ (notice that by doubling *kung*, the word for ‘shave’, we get the meaning ‘clean-shaven’). Breaking this word into parts we get ‘shave + shave + remove hairs + head + man’. This looks at first glance like compounding, a process of stringing words together to form a single word, which happens in many well-known languages. German is famed for unwieldy long words like *Nasenspitzenwurzelentzündung*, meaning an ‘inflammation of the root of the tip of the nose’. But Sora is doing something quite different. It is not merely stringing words together. In Sora, verbs literally swallow other words like pythons, sucking them in by a process linguists call incorporation.

English has a limited form of word incorporation, as in ‘We bungee-jumped’ where a noun ‘bungee’ becomes part of (but is not inside of) a verb ‘jump’. Sora goes much further, allowing verbs to suck in direct objects, indirect objects, and instruments from elsewhere in the same sentence. If an angry Sora speaker says “I will stab you in the belly with a knife,” it comes out as *poo-pung-koon-t-am*. The nouns ‘belly’ and ‘knife’ both get sucked up inside the verb *poo-t* ‘will stab’ like so many rats inside a well-fed python. The result is not a string of words, but a single giant verb.

<i>poo</i>	<i>-pung</i>	<i>-koon</i>	<i>-t</i>	<i>-am</i>
[STAB	+belly+knife	+will	+thee]	

But since Sora, unlike German, has no written form, how do we know *popungkoontam* is actually one word and not just several spoken rapidly or strung together? When a python swallows a rat, both change in appearance. Rat gets balled up inside, python bulges on the outside. Sora words, sucked up inside of verbs, also contract, morphing into smaller versions of themselves. For instance, *koondin*, ‘knife’, is squished into *koon*. Even *am*, still hanging out of the python’s mouth, as it were, is a compressed form of the full *pronoun*, ‘you’.<sup>37</sup>

An even stranger case is found in the Gta’ language, spoken by 3,055 hunter-gatherers in the hills of eastern India. Gta’, like Sora, allows verbs to swallow multiple nouns. But an adjective modifying a swallowed noun may remain stranded on the outside, modifying from a distance, as the word ‘sharp’ in the example “I will stab you in the belly with a *sharp* knife” modifies ‘knife’:



sharp [STAB +belly +knife +will +thee]

Syntacticians who build tree models of languages find it difficult to accommodate such “exotic” structures because they go against common notions of how we think sentences and words are built.<sup>38</sup> As glaring exceptions, Arrernte, Sora, and Gta’ are essential in helping scientists formulate universal rules about how words interact. Without these languages, our understanding of fundamental processes of word-building would be limited.

### Man-Talk, Woman-Talk

In some languages, men and women talk very differently, or a speaker of either sex will talk differently depending on the sex of the interlocutor or the person being talked about. Sex matters a great deal in many languages in ways it barely matters (if at all) in English. Of course, we might say ‘sir’ to a man and ‘sister’ for female sibling. In colloquial English, a recent study found that the word ‘dude’ is three times more likely to be uttered in conversations between men than those involving women.<sup>39</sup> But it is hard to find in English any examples of how the sex of the speaker or addressee affects the actual grammar (not just the pronunciation) of the language.

Small and endangered languages offer many more examples of how sex interacts directly with grammar. In Arapaho (1,038 speakers in Oklahoma), even expressions like ‘hello’, ‘yes’, and ‘wait’ are totally different when said by a man than by a woman.<sup>40</sup> In Arapesh (spoken by 30,000 speakers in three dialects in New Guinea), if I say the word *mehinen* to you and you are a man, I am talking about your sister’s son, but if you are a woman, then I am referring to your brother’s daughter.<sup>41</sup> In other words, the sex of the person being talked *about* can only be known if the sex of the person being talked *to* is known. If you were eavesdropping on my Arapesh conversation but could not see my addressee, you would not know if I was gossiping about a man or a woman. Also, it is impossible to translate ‘nephew’ or ‘niece’ into Arapesh unless you know who the aunt or uncle is.

In Gros Ventre<sup>42</sup> (10 or fewer speakers left in Montana), men and women once used different sounds, words, and exclamations. For ‘bread’, men say *jatsa* and women *kyatsa*; for ‘hello’ boys would say *wei* and girls *ao*.<sup>43</sup> The sound ‘ch’ was spoken only by adult, fluent male speakers, while ‘k’ was used in its place by women, children, and non-fluent adult males

including visiting linguists. Words like ‘teepee’, ‘porcupine’, ‘buffalo’, and ‘boy’ had distinctly different pronunciations. All these distinctions began to merge as the number of speakers dwindled. But in the past, the community was keenly aware of sex differences in speech. If a male child entering his teen years continued to pronounce ‘k’, he would be admonished sternly to use ‘ch’ instead. A linguist from outside the tribe would be told to use ‘k’ instead.<sup>44</sup>

Yanyuwa (70 speakers in Australia) women and men talk so differently that their speech is really two different dialects.<sup>45</sup> Differences go beyond sounds or words, encompassing grammatical affixes, pronouns, and other parts of speech. Women’s talk is reportedly more complex, and men imitate it only imperfectly. The Yanyuwa rigidly enforce speech-sex differences by scolding mistakes, especially those made by newly-initiated adult men expected to adopt fully male speech. One young Yanyuwa man recounted: “When I spoke like a woman my father said to me, ‘Where are your breasts and woman’s parts [vagina]?’ I was really ashamed. I was very careful for a while after that to speak men’s words.”<sup>46</sup> Use of opposite sex speech is only tolerated in risqué acts, such as a man impersonating a woman in dance, or in myth songs recounting the female creators’ voices.

Similar to the gender restrictions discussed above, many languages require speakers to use different words or speech styles or even different grammar rules when talking to people of higher social status. Formality patterns, found in very large languages like Japanese (125 million speakers) or Javanese<sup>47</sup> (75 million speakers), also pop up in smaller languages. Sasak (2.1 million speakers), spoken on Lombok Island in Indonesia, is said to have at least three distinct levels of formality: low, high, and very polite. Depending on your own social status relative to your addressee, you must utter one of three very different sentences to say exactly the same thing.<sup>48</sup>

**Table 7.2** The sentence “What did you just say?” at three distinct formality levels in Sasak

<i>Level of formality</i>	<i>‘What’</i>	<i>‘say’</i>	<i>‘you’</i>	<i>‘now’</i>	<i>‘this’</i>
Very Polite	Napi	basen	dekaji	baruq	nike?
High	Napi	basen	pelinggih	baruq	nike?
Low	Ape	inin	side	baruq	no?

*Source:* Data from Syahdan (1996:89), cited in Austin 1998.

Of course, we do this in English too. Speaking to a younger brother you might say “gimme a buck!” whereas speaking to the president you might say “Would you be so kind as to lend me a dollar?” The intent is the same, the style and vocabulary radically different. In English, speech register (or formality level) is encoded mostly in word choice and intonation. In other languages, it is encoded not only in words but in sounds, parts of words, syntax, and other grammatical levels. How extreme can sex-based or status-based differences be within a single language? We do not yet know to what extent such social conventions may influence grammar. With the demise of languages like Yanyuwa, Arapaho, and Gros Ventre, we may never know.

### Handy Talk, Talking Hands

Most of the world’s signed languages—spoken natively by deaf people—have never been properly counted or documented. A common myth is that these are just versions of English or Spanish or another local language, with a hand sign for each word. Nothing could be further from the truth. American Sign Language (ASL, used by up to 500,000 deaf people as their primary language) is no closer to English in its words, structures, and grammar rules than is Japanese. Another common myth is that signed languages use mostly *iconic* gestures, meaning hand shapes that look like or mimic the things they refer to, and thus signs can be universal to all deaf people. This is also false. A speaker of ASL and a speaker of Japanese Sign Language have no common language. Signed words are overwhelmingly abstract, not imitative, which is why speakers of one sign language cannot understand what speakers of another one are saying.

Debunking these myths has been a major accomplishment of linguistic science.<sup>49</sup> Researchers have demonstrated that sign languages are fully complex, fully functioning human languages, not simplistic gesture systems or in any way inferior to any other human language. But to get a full picture of human language ability, scientists must include in their research all known sign languages. So far, there are 121 identified and named sign languages used in deaf communities around the world, but potentially a great many more remain completely undocumented.<sup>50</sup>

Many sign languages are now rapidly vanishing. This is in part because many deaf communities possessing unique sign languages are small, in-

digenous, and rural. As countries spend more resources on their deaf citizens, deaf children are sent to urban boarding schools where they are taught only the standard national sign language used in the country. Many original sign languages are now endangered and will vanish before their existence is ever known to science.

Signed communication systems arise spontaneously wherever deaf people live.<sup>51</sup> These may start out as simple systems of gestures with a limited range of uses. But as soon as there is a community of deaf people, and often within just one generation, these systems develop into full-fledged languages, rapidly becoming as complex as spoken languages.

In fact, in some ways sign languages can even be more complex. Because they speak with the hands, signers have a unique possibility not available to spoken languages or even written texts. Many signs require only one hand, so it is possible to use the other hand to make another sign, uttering two words at exactly the same time. Scientists have documented simultaneous use of two one-handed signs in sign languages of Italy, Ireland, and Quebec. But how often do speakers actually make use of this possibility, and if so, what do they use it for?

In Italian Sign Language (ISL, number of speakers unknown), a speaker can say something like ‘A car stops at a traffic light’ or ‘A newspaper lies on the table and one of its pages turns over’ by using two one-handed signs simultaneously. What is interesting is that each of the two signs seems to affect the other, changing its shape in some basic way, but still allowing it to be recognized as a distinct gesture.<sup>52</sup> CAR, for example, is supposed to be a two-handed sign, imitating both hands gripping a steering wheel. And TRAFFIC LIGHT flexes the fingers of the right hand to denote a blinking light.<sup>53</sup> In combining the two signs, only the right hand makes the sign for ‘car’, while the left hand signs ‘traffic light’. Of course, there is no spoken language in which you can simultaneously say ‘car’ and ‘traffic light’.

### **Unusual Hand Shapes as Words**

Signed languages are poorly documented and may have many more surprises in store for us. Because they are not written down, they can only be observed in the moment of speech or recorded on videotape for later analysis. Anthropologist Angela Nonaka studies endangered signed languages



**Figure 7.9**

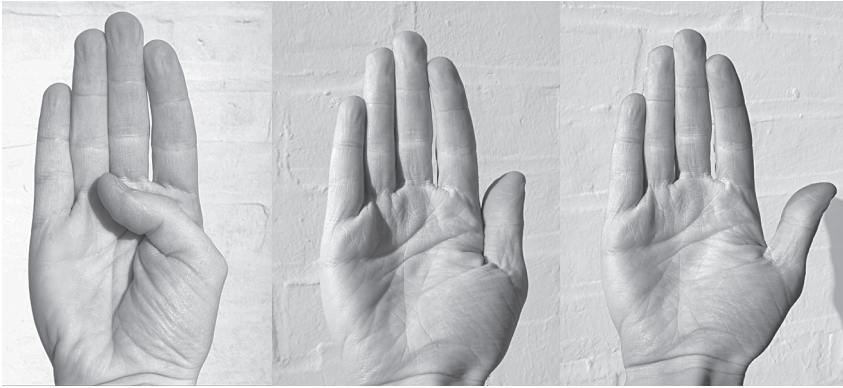
The two-handed sign in Italian Sign Language (ISL) meaning CAR (left), the right-handed sign TRAFFIC LIGHT (center), and a simultaneous combination of both signs (right), meaning ‘A car stops at a traffic light’. Demonstrated here by linguist Donna Jo Napoli, a non-native speaker of ISL and ASL. Courtesy of Robbie Hart

of Thailand and reports that they possess some unusual and scientifically interesting features.

There are at least six signed languages native to Thailand, plus the national standard Thai Sign Language taught to deaf children in schools.<sup>54</sup> The national variety was introduced by educators in the 1950s and is based on American Sign Language (ASL). Two sign languages spoken before ASL arrived, Old Bangkok Sign Language and Old Chiangmai, are now endangered, having no fluent speakers under the age of 45, and no longer being used on a daily basis.

A third, Ban Khor Sign Language, is spoken in a remote village in northern Thailand by fewer than 1,000 deaf people and their relatives. Ban Khor has a hand shape that is one of the most universal ones, found thus far in all known sign languages. In ASL, it is the hand shape used for the letter ‘b’, and it looks like this.

According to the grammars of signed languages, each hand shape has a number of possible orientations and contact points. For example, once I make the ‘b’ sign, I can turn my hand in various directions and make contact with various body parts, but all these are strictly limited. Not all possible orientations and hand shapes are allowed by the grammar. This



**Figure 7.10**

Three varieties of the 'b' hand sign found in the world's sign languages.

is analogous to spoken languages, where not all sound combinations that can be pronounced by the mouth are allowed. For example, in English, the word *spap* or *smam* are certainly possible combinations of sounds, but to most speakers they sound odd, if not impossible as words. Likewise, English forbids, but Italian allows words to begin with 'sb', such as *sbaffo* ('a smudge'); whereas Italian forbids but English allows words to end with 'sp', such as 'clasp'.

Anthropologist Angela Nonaka has discovered a highly unusual use of the 'b' hand shape in Ban Khor sign language.

So far, no other known sign language takes the 'b' hand shape and places it in this particular turned orientation with respect to the body. It is also unusually (for a stationary sign) positioned so that it obscures the face. Without Ban Khor Sign, we would not know that this placement of the 'b' hand shape was even possible in a signed language. Signed languages are poorly documented and may hold many more surprises in store for us. But many will vanish even before people outside the speech community become aware of their existence.

### ***Languages and Prehistory***

Languages contain buried clues that can help us trace the prehistory of humans and their migrations around the globe. Because language change



**Figure 7.11**

Signs for “name” and “foreigner” demonstrated by two native speakers of Ban Khor Sign Language. Photograph by Angela Nonaka courtesy of Cambridge University Press

happens so rapidly, when a population splits, the two resulting groups can end up after some time speaking two separate, mutually incomprehensible tongues. Each language is thus one piece in the puzzle to tracing ancient human migrations that led people to the Americas, Polynesia, and so on. Linguistic evidence from shared vocabulary may also reveal prehistoric contacts among unrelated peoples. Two native languages of southern California have in their vocabularies some special words referring to canoes and canoe-making technology. These appear to have been borrowed from ancient Polynesians who must have sailed to California in prehistoric times.<sup>55</sup> Often linguistic evidence is needed to supplement archeological and genetic data in understanding the history of human habitation and contact patterns around the globe.

For example, genetic evidence clearly points to links between natives of Central Siberia and North America, as the two groups share unique traits not found elsewhere.<sup>56</sup> But linguistic links between Siberians and Native

Americans have proved elusive. Languages change so rapidly that after only 1,000 or so years of divergence, what were once close dialects may change beyond recognition, even though the peoples themselves retain cultural or genetic similarities. Linguist Edward Vajda has found intriguing parallels in verb structure and sound correspondences in basic vocabulary that link the Siberian language Ket (990 speakers) to native Alaskan languages like Tlingit (700 speakers) and Eyak (1 speaker), and to the more geographically distant Navajo (148,000 speakers).<sup>57</sup> Though controversial and awaiting further research, Vajda's initial results are tantalizing and may provide elusive clues about the prehistoric peopling of the Americas. They may provide the first solid linguistic link between the populations of North Asia and North America, revealing something about Ice-Age migrations of human populations.

Another intriguing puzzle of human prehistory, and one that linguistics may help solve, is cultural evolution. Humans made the transition from hunting and gathering to agriculture in different places and times. The Mlabri people living in the hills of Thailand and Laos practice a very different way of life than do other peoples in the area, who are all settled agriculturalists. The Mlabri roam the forests, building temporary houses of leaves, and surviving by hunting and gathering. It was assumed that the Mlabri must therefore be descendants of an original hunter-gatherer people who had never adopted agriculture. But when genetic tests were done, the Mlabri showed surprisingly little genetic diversity, indicating that their entire population must have sprung from a common ancestor (perhaps a single woman and from one to four males) as recently as 500 to 1,000 years ago.<sup>58</sup>

Linguistic studies revealed that the Mlabri tongue is related to Tin (46,000 speakers), also spoken in the hills of Thailand. In diverging from Tin, Mlabri underwent a series of well-defined sound and grammar changes over a millennium to bring it to its present form. But since the Tin are known to have been practicing agriculture for well over 1,000 years, the Mlabri would seem to present a rare case of recent *reversion* from a once agriculturalist society to a hunter-gatherer one.

Support for the reversion hypothesis is found in many Mlabri words and myths that refer to agriculture.<sup>59</sup> Scientists do not know what founding event led the Mlabri to go off on their own, abandon agriculture, and become roving forest dwellers. By looking at both genes and languages, it is possible to peer deeper into the past of the Mlabri and thus reconstruct one small part of human prehistory.



## Discoveries Await Us

Throughout this book, I have argued that small and endangered languages will be important to humanity and to science for the kinds of cultural knowledge they contain—technologies for interacting with animals, plants, countable objects, time, and topography. For each of these domains, I also suggested ways in which cultural knowledge is uniquely packaged in any given language and ingeniously encoded in its words and grammatical structures.

In this final chapter, I have departed from cultural knowledge to talk about pure structure of the kind that interest most professional linguists: grammar—the invisible building blocks of cognition. Grammar deservedly preoccupies most linguists, and it is a realm of the mind where many astounding discoveries remain to be made. Any single discovery, even a *eureka* moment, may seem modest or inconsequential on its own. Ban Khor sign language takes a familiar hand shape and places it in an unusual position. Nivkh has a unique classifier for dried fish, Tofa a special morpheme for smell. Mlabri has ancient farming-related words even though its speakers are hunter-gatherers. Rotokas may have as few as six consonants. Eleme doubles part of a word to negate it. Carrier forces its speakers to pay attention to tactile qualities of objects. Sora allows a verb to swallow multiple nouns. Tuvan and White Hmong have unusually rich inventories of words to imitate sounds.

But when we sum up all these discoveries, both across many languages and within a single one, we achieve a slightly clearer insight into the grand realm of human cognition. Language may by its very structure force speakers to attend to certain qualities of the world (shape, size, gender, countability). Languages are complex self-organizing systems that evolve complex nested structures and rules for how to put the parts of words or sentences together. No two languages do this in the same way. We do not yet have a grasp of what the limits to such complexity are or where the boundaries lie. Endangered languages enormously widen and deepen our view of what is possible within the human mind. As strenuously as I have argued in previous chapters for their importance to humanity and to the planet, I argue here for their deep relevance to pure scientific inquiry. As we delve into languages, many revelatory discoveries await us.

57. Ventureño Chumash: Beeler 1964, 1967, 1988. Additional Ventureño data may be found in Harrington 1981. A survey of native Californian numeral-base systems is found in R. B. Dixon and Kroeber 1907.
58. Ainu: Batchelor (1905: 96) comments on the ‘cumbrous’ nature of the Ainu system. There are some very recent efforts at cultural and language revitalization among the Ainu, but very few, if any, speakers remain.
59. The Thulung numbers of 1944 come from Rai 1944 qtd. in Allen 1975, and although whether they were still widely used at that time is in question, they are closer to the Tibeto-Burman proto-forms of the numbers. Thulung numbers of 2000 come from Lahaussais 2003. The orthography has been slightly simplified in this data set.
60. Comrie 2005. This chapter was inspired Prof. Bernard Comrie’s lecture entitled “Endangered Numeral Systems,” presented in January 2004 at the annual meeting of the Linguistic Society of America, in Oakland, CA.
61. Supyire: data from Carlson 1994.
62. Endangered non-decimal systems: Comrie 2005.
63. Iqwaye losing their numbers: Mimica 1988: 11.
64. Mimica 1988: 11–12.

#### **Case Study: The Leaf-Cup People, India’s Modern ‘Primitives’**

1. Of India’s 17 official languages, only the smallest, Kashmiri, has less than 10 million speakers, and even then, it is many times the size of Ho, the largest of the Munda languages. *The Ethnologue* has population estimates for many of the languages of India: Gordon 2005.
2. Small languages, and even not so small ones like Ho, can find it difficult to break into the computer age if they use a non-latin alphabet or writing system that differs from those used by economically important world languages. For the sake of language revitalization and access to computers by speakers of endangered languages, we hope to see greater progress in ushering the writing systems of small and endangered languages into the worldwide Unicode standard.
3. The Ho origin myth by Mr. K. C. Naik Biruli (born 1957), resident Mayurbanj district, was told in Bhubaneshwar, India, on September 13, 2005. It was recorded in audio and video by Gregory D. S. Anderson and me. This is an abridged version of a yet unpublished translation by Biruli and Anderson. The ‘ten months’ of pregnancy are lunar months (see chapter 3). To the best of my knowledge, this Ho origin story has not been previously published.

#### **7. Worlds within Words**

1. Valuing small languages: Linguist and endangered language expert Nancy Dorian (2002) points out that scientists who valorize the diversity of languages for the sake of advancing a scientific research agenda are indulging

a highly culture-specific (e.g., Western) set of values. Linguists are typically “outside experts” on language loss rather than experiencers of it, so there is a danger that we may be setting our own values ahead of those of the speech communities themselves. For example, Dorian notes linguists tend to “dote upon structural rarity and overemphasize it to the exclusion of other significant matters” (Dorian 2002: 136). I agree with Dorian’s assessment and have thus placed this chapter about language structures, traditionally an object of great concern to linguists, at the very end of this book. I also heartily endorse Dorian’s challenge that “the rhetoric of advocacy needs to broaden for all audiences so as to acknowledge the vastness of the research challenge, while the investigations need to widen so as to encompass more of the social and cultural as well as the structural range that each language represents” (Dorian 2002: 139). The choice and ordering of the chapter topics in this book, as well as the inclusion of native speakers’ points of view are intended as part of such a broadening effort.

2. Chomskyan preference for innate structures over culturally shaped content: As Eve Danziger notes, “In the view of Chomsky (1975) and his followers, all significant linguistic categories exist independently of the situation of their users—including the particular language learned. The categories are present from birth, encoded in human DNA in a form that is autonomous of any subsequent experience. . . . But linguistic and cultural categories are collective conventions that themselves can inspire and construct, as well as reflect, aspects of the individual’s experience” (Danziger 2005:66–67).
3. Holistic approach to language: Haas 1976: 43.
4. As linguist Doug Whalen (2004) predicts, the study of endangered languages will also revolutionize the field of linguistics.
5. Homer Simpson’s speech: For a linguistic analysis, see Yu 2004.
6. Queen’s English: Harrington, Palethorpe, and Watson 2000.
7. Biatch: The most popular spelling by far for this recently coined, two-syllable version of the word ‘bitch’ is ‘biatch’ with 640,000 Google hits, while ‘biotch’ has 74,300; ‘beyotch’ 17,400; ‘bioootch’ 4,490; and ‘biooootch’ 1,270. Even exaggerated spellings like ‘beeeeeeatch’ get 1,130 hits; ‘beeeeeeeatch’ 184; ‘beeeeeeeatch’ 213; and ‘beeeeeeeatch’ 80 (as of August 2005). No matter how it is spelled, this word appears to be solidly part of the English spoken and written lexicon, though not yet acknowledged in mainstream dictionaries.
8. Puhleeze: An alternative two-syllable form of ‘please’ indicating exasperation gets 24,500 Google hits. However, this spelling is not winning by such a landslide as ‘biatch’. Google also yielded 18,200 hits for alternate spelling ‘puhlease’, 12,700 for ‘puleeze’ and solid numbers for longer spellings including 3,130 for ‘puhleeeeeeze’ (as of August 2005).
9. Equal complexity: Fromkin, Rodman, and Hyams 1998.
10. Equal complexity: Fromkin, Rodman, and Hyams 1998. For an interesting counter-argument, see Dixon 1997.

11. Nichols 1992 proposes one method for assessing complexity based on available inflection sites within a typical sentence. Many more such models are needed before linguistic complexity becomes truly quantifiable.
12. Fromkin, Rodman, and Hyams (1998) set forth equal complexity as a fundamental principle: "There are no 'primitive' languages—all languages are equally complex and equally capable of expressing any idea." These ideas are widely repeated, for example, in this statement from a Gallaudet College website (Gallaudet College 1977, 1978): "All languages are equally complex and capable of expressing any idea. A language which appears simple in some respects is likely to be more complex in others."
13. Child language acquisition: Stromswold 2000: 910, notes that "Children who are acquiring languages like Turkish, which have rich, regular, and perceptually salient morphological systems, generally begin to use functional category morphemes at a younger age than children acquiring morphologically poor languages . . . For example, in striking contrast to . . . English-speaking children, Turkish-speaking children often begin to produce morphologically complex words before they begin to use multiword utterances (Aksu-Koc and Slobin 1985)."
14. Rotokas: The reported phonetic consonant inventory is [p], [t], [k], [ʔ], [~], and [g], with allophonic variation; for example, /r/ may be pronounced as [~], [d], [n], or [l]. In the proposed orthography for Rotokas, six consonant symbols are used: {*p, t, k, v, r, g*} (Firchow and Firchow 1969). Robinson 2006 and Robinson, in communication with me in 2005, states that his research shows the Rotokas phoneme inventory now includes [m], [n], and [s], and was either originally misanalyzed or may have recently acquired phonemes from contact with Tok Pisin or English. Ingush consonants: University of California, Berkeley Ingush Project under the direction of Prof. Johanna Nichols (website in bibliography) and Johanna Nichols, "A brief overview of Ingush phonology," at <http://ingush.berkeley.edu:7012/orthography.html#Phonology> (accessed August 2006).
15. Ejectives: Ladefoged 2001: 131–33. Notes that the trade-off between ease of articulation and acoustic distinctness disfavors bilabial ejectives, which differ only slightly in their acoustics from regular bilabial plosives but require more effort. Ladefoged further notes that if a language does adopt ejectives, it tends to have them at places of articulation where it already has other plosives.
16. Ubykh: Dumézil 1959, Vogt 1963.
17. Rotokas: Stuart Robinson 2006, and Robinson personal communication. The full glossed form is:  

ora-rugorugo-pie-pa-a-veira

REF/REC-think.REDUP-CAUS-PROG-3.PL-HABITUAL

"They were always thinking back."
18. Ingush words: Nichols 2004.
19. Equal complexity: For a brief, general discussion of equal complexity, see Dixon 1997: ch. 3. Attempts to assess complexity across languages include

- Nichols 1992, Juola 1998, Shosted 2005. A discussion of language-specific loci of complexity (and clustering of speakers' errors therein) is found in Wells-Jensen 1999.
20. Biological analogy: In fact, the study of biology offers much the same dichotomy. Some critics decry the modern emphasis on genetic studies, arguing that an understanding of an organism's DNA is useless if we ignore the actual organism: its morphology, behavior, ecological adaptivity, etc. For more on the hooded seal see Lavigne and Kovacs 1988.
  21. For a cogent statement on why linguists need a diversity of languages in order to recognize the "strange phenomena," see Corbett 2001.
  22. "Language is a mirror of mind": Chomsky 1975: 4.
  23. Tuvan acoustic sensibility and throat-singing: Levin 2006 and personal communication; van Tongeren 2002 and personal communication.
  24. Tuvan reduplication and sound symbolism: Harrison 2000, 2004.
  25. White Hmong: Ratliff 1992: 136–63 and personal communication.
  26. Reduplication patterns from many languages: Raimy 2000 and personal communication; Rubino 2005.
  27. Rotokas reduplication: Firchow 1987: 13, 54.
  28. Eleme reduplicative negation: Data from Gregory Anderson (personal communication); data collected in collaboration with Oliver Bond (Bond 2006). The orthography has been simplified above; the full phonetic forms are: [mT-rT], [mT-mT-rT]; [(bQi) r[-kQ-d'u], [(bQi) r[-kQ-kQ-d'u].
  29. Carrier: Poser 2005.
  30. Cantonese classifiers: The Internet discussion site may be viewed at Sheik 2005. Disagreement among speakers about which classifiers to use for novel objects is not surprising, since unless they know the classifier in advance, they have to make a decision based on how they perceive the object. For example, the choice of classifier for glue will depend on the glue's current form. If the 'glue' has congealed and solidified, then even classifiers for solid things might be applicable, e.g., *yati juem gaau* ('one brick glue; a brick of gelatin stuff'); a bottle of glue = *yati joem gaau seuiz* ('one bottle glue water'); a stick of glue = *yati jiu gaau seuiz* ('one stick glue water'); a drop of glue = *yati dik6 gaau seuiz* ('one drop glue water'). Thanks to linguist Alan C. L. Yu (personal communication).
  31. Yupno 'hot', "cold" and "cool": Wassmann and Dasen 1994b.
  32. Salish reduplication: Anderson 1999, Kuipers 1967.
  33. Nivkh and Squamish: These languages also each provide a generic class for novel or hard-to-classify objects.
  34. Rotuman kinship terms and inheritance: Churchward 1940; Rensel 1991.
  35. Tabasaran (also spelled Tabassaran) and Tsez: Comrie and Polinsky 1998. The authors conclude (pp. 105–106): "in both Tabasaran and Tsez, we have a moderately rich number of cases: 14 or 15 in Tabasaran, depending on dialect, and 18 in Tsez. The richness that gives rise to claims such as Tabasaran having 48, 47, or 53 cases, or Tsez having 126 cases, derives from the possi-

- bilities of combining these cases with one another” (*Guinness Book of Records* by Young 1997).
36. Arrernte: Henderson 2002: 108. Thanks to Alice C. Harris for bringing this example to my attention.
  37. Sora incorporation: Ramamurti 1931. Some data simplified in the text for legibility. Sora also makes incorporated words like *e-jir-ten-e-mandra* ‘the man that is going’ and *jeruu-lunger-kid-en* ‘a tiger that dwells in a deep cave’. Actual phonemic transcriptions given as: [jəru-ˈluŋər-ˈkid-ən] ‘deep+cave+dwells+tiger’ (p. 48 ex. 6); [kuŋ-kuŋ-ˈded-u-ˈbo:b-ˈmər] ‘shave+shave+remove hairs+head+man’ (p. 49 para. 169 and footnote 2); [ə-ˈjir-t-e-ŋ-ə-ˈman(d)ra:], lit. ‘that+goes+that+man’ (p. 49 para. 170 ex. 1). The ‘belly stab’ example is [ˈpo:-ˈpoŋ-ˈkun -t-am] ‘stab-belly-knife-[will]-thee’, i.e. ‘I will stab you with a knife in your belly’ (p. 44, para. 139). Thanks to Gregory Anderson for pointing out this phenomenon.
  38. Gta<sup>1/4</sup>: externally modified incorporated arguments are briefly mentioned in Sadock 1991 and discussed in greater detail in Anderson (2007). Gregory Anderson (personal communication) notes that the syntactic interpretation of these structures remains controversial, and that theories of incorporation other than Sadock 1991, e.g., M. Baker 1988, disallow (and fail to account for) such ‘syntactic transparency’ of an incorporated noun. Clearly, further study needs to be done of Gta<sup>1/4</sup> and other little-documented Munda languages. A linguistic field expedition to India in September 2005 by Gregory Anderson and me yielded some promising new data on Munda incorporation patterns.
  39. “Dude”: Kiesling 2004.
  40. Arapaho gendered words: Conathan 2006 and personal communication.
  41. Arapesh kin terms: Fortune 1942: 24. See also Dobrin 2001: 35.
  42. Gros Ventre: Driver 1961: 353–54.
  43. Gros Ventre words: Spellings adapted here for the general reader. Flannery (1946) spells Gros Ventre ‘bread’ phonetically as [dja’tsa] and [kya’tsa]; for ‘answer to a hail’ she writes [wei’] (male) and [ao’] (female).
  44. Gros Ventre: Taylor 1982. Sound samples of Gros Ventre, including male and female speech (though these do not always sound distinctly different where differences are expected), may be downloaded from a Gros Ventre language website (Fort Belknap College 2005). Similar cases of striking differences between men’s and women’s speech are reported for Chuckchi (10,000 speakers in Siberia) by Dunn 2000 and citations therein, and for 19<sup>th</sup> century “Esquimaux” (Inuit) of Alaska by Parry 1824:553.
  45. Yanyuwa: Kirton 1988.
  46. Yanyuwa: Data from Bradley 1998. The young man’s quote is attributed to Yanyuwa consultant “J.T.”
  47. Javanese: For an excellent discussion of formality levels and the linguistic differences they entail, see Errington 1998.

48. Sasak: According to linguist Peter Austin (personal communication), you cannot just say 'sit' in Sasak (2.1 million speakers). Instead, you must use special forms to specify who is doing the sitting, in what part of the house, and with what particular body posture. The Sasak verb system thus encodes physical, topographic, and social information about the sitter. See also Syahdan 2000: 89.
49. Sign languages: For general discussions of sign languages vis-à-vis spoken ones, see Jackendoff 1994: ch. 7; Napoli 2003: ch. 4; and S. Anderson 2004: ch. 9. For a recent typological study of sign languages see Zeshan 2006.
50. Number of sign languages worldwide is unknown: Linguist and sign language expert Ulrike Zeshan (personal communication) writes: "Currently, nobody knows how many sign languages exist in the world. My personal estimate is that there are probably several hundred sign languages, most of which are undocumented. This also includes small village-based sign languages in village communities with a high percentage and long history of hereditary deafness. . . . Leaving a good margin to cover the many white areas on our world map of sign languages, I doubt we would go beyond, say, about 700 sign languages in the final count, if indeed we ever reach that stage."
51. Spontaneous sign: A recent example is Al-Sayyid Bedouin Sign Language, which arose spontaneously over the past 70 years with no apparent outside influence and has evolved the full range of complex structures expected in any human language (Sandler et al. 2005).
52. Italian Sign language simultaneous signs: Russo 2004. For a discussion of sign simultaneity see Zeshan 2002; in Irish sign, Leeson and Saeed 2002; in Quebec sign, Miller 1994.
53. Italian Sign Language signs for CAR and TRAFFIC LIGHT: Radutzky 2001.
54. Thai sign languages: Nonaka 2004: 741 and personal communication.
55. Polynesian canoe words found in Southern California languages Chumashan and Gabrielino: Klar and Jones 2005.
56. Genetic similarities between Native Americans and Native Central Siberians: Schurr 2004; Zegura et al. 2004.
57. Ket (Yeniseic) in relation to Tlingit and Eyak: Vajda 1999, 2005 and personal communication.
58. Mlabri reversion: Oota et al. 2005.
59. Mlabri words: Rischel 1995.