Pinker, Steven. 1995. The Language Instinct: The New Science of Language and Mind. Penguin UK.

Chapters 5-6, pages 126-191

5

Words, Words, Words

The word *glamour* comes from the word *grammar,* and since the Chomskyan revolution the etymology has been fitting. Who could not be dazzled by the creative power of the mental grammar, by its ability to convey an infinite number of thoughts with a finite set of rules? There has been a book on mind and matter called *Grammatical Man,* and a Nobel Prize lecture comparing the machinery of life to a generative grammar. Chomsky has been interviewed in *Rolling Stone* and alluded to on *Saturday Night Live*. In Woody Allen's story "The Whore of Mensa," the patron asks, "Suppose I wanted Noam Chomsky explained to me by two girls?" "It'd cost you," she replies.

Unlike the mental grammar, the mental dictionary has had no cachet. It seems like nothing more than a humdrum list of words, each transcribed into the head by dull-witted rote memorization. In the preface to his *Dictionary,* Samuel Johnson wrote:

> It is the fate of those who dwell at the lower employments of life, to be rather driven by the fear of evil, than attracted by the prospect of good; to be exposed to censure, without hope of praise; to be disgraced by miscarriage, or punished for neglect, where success would have been without applause, and diligence without reward.
>
> Among these unhappy mortals is the writer of dictionaries.

Johnson's own dictionary defines *lexicographer* as "a harmless drudge, that busies himself in tracing the original, and detailing the signification of words."

In this chapter we will see that the stereotype is unfair. The world of words is just as wondrous as the world of syntax, or even more so. For not only are people as infinitely creative with words as they are with phrases and sentences, but memorizing individual words demands its own special virtuosity.

Recall the wug-test, passed by any preschooler: "Here is a wug. Now there are two of them. There are two___." Before being so challenged, the child has neither heard anyone say, nor been rewarded for saying, the word *wugs*. Therefore words are not simply retrieved from a mental archive. People must have a mental rule for generating new words from old ones, something like "To form the plural of a noun, add the suffix *-s*." The engineering trick behind human language—its being a discrete combinatorial system—is used in at least two different places: sentences and phrases are built out of words by the rules of syntax, and the words themselves are built out of smaller bits by another set of rules, the rules of "morphology."

The creative powers of English morphology are pathetic compared to what we find in other languages. The English noun comes in exactly two forms *(duck* and *ducks),* the verb in four *(quack, quacks, quacked, quacking)*. In modern Italian and Spanish every verb has about fifty forms; in classical Greek, three hundred and fifty; in Turkish, two million! Many of the languages I have brought up, such as Eskimo, Apache, Hopi, Kivunjo, and American Sign Language, are known for this prodigious ability. How do they do it? Here is an example from Kivunjo, the Bantu language that was said to make English look like checkers compared to chess. The verb "Näïkìmlyìïà," meaning "He is eating it for her," is composed of eight parts:

- N-: A marker indicating that the word is the "focus" of that point in the conversation.
- -ä-: A subject agreement marker. It identifies the eater as falling into Class 1 of the sixteen gender classes, "human singular." (Remember that to a linguist "gender" means kind, not sex.) Other genders embrace nouns that pertain to several humans, thin or extended objects, objects that come in pairs or clusters, the pairs or clusters themselves, instruments, animals, body parts, diminutives (small or cute versions of things), abstract qualities, precise locations, and general locales.

- -ï-: Present tense. Other tenses in Bantu can refer to today, earlier today, yesterday, no earlier than yesterday, yesterday or earlier, in the remote past, habitually, ongoing, consecutively, hypothetically, in the future, at an indeterminate time, not yet, and sometimes.
- -kì-: An object agreement marker, in this case indicating that the thing eaten falls into gender Class 7.
- -m-: A benefactive marker, indicating for whose benefit the action is taking place, in this case a member of gender Class 1.
- -lyì-: The verb, "to eat."
- -ï-: An "applicative" marker, indicating that the verb's cast of players has been augmented by one additional role, in this case the benefactive. (As an analogy, imagine that in English we had to add a suffix to the verb *bake* when it is used in *1 baked her a cake* as opposed to the usual *I baked a cake.)*
- -à : A final vowel, which can indicate indicative versus subjunctive mood.

If you multiply out the number of possible combinations of the seven prefixes and suffixes, the product is about half a million, and that is the number of possible forms per verb in the language. In effect, Kivunjo and languages like it are building an entire sentence inside a single complex word, the verb.

But I have been a bit unfair to English. English is genuinely crude in its "inflectional" morphology, where one modifies a word to fit the sentence, like marking a noun for the plural with *-s* or a verb for past tense with *-ed*. But English holds its own in "derivational" morphology, where one creates a new word out of an old one. For example, the suffix *-able,* as in *learnable, teachable,* and *huggable,* converts a verb meaning "to do X" into an adjective meaning "capable of having X done to it." Most people are surprised to learn how many derivational suffixes there are in English. Here are the more common ones:

-able	-ate	-ify	-ize
-age	-ed	-ion	-ly
-al	-en	-ish	-ment

-an	-er	-ism	-ness
-ant	-ful	-ist	-ory
-ance	-hood	-ity	-ous
-ary	-ic	-ive	-y

In addition, English is free and easy with "compounding," which glues two words together to form a new one, like *toothbrush* and *mouse-eater*. Thanks to these processes, the number of possible words, even in morphologically impoverished English, is immense. The computational linguist Richard Sproat compiled all the distinct words used in the forty-four million words of text from Associated Press news stories beginning in mid-February 1988. Up through December 30, the list contained three hundred thousand distinct word forms, about as many as in a good unabridged dictionary. You might guess that this would exhaust the English words that would ever appear in such stories. But when Sproat looked at what came over the wire on December 31, he found no fewer than thirty-five new forms, including *instrumenting, counterprograms, armhole, part-Vulcan, fuzzier, groveled, boulderlike, mega-lizard, traumatological,* and *ex-critters*.

Even more impressive, the output of one morphological rule can be the input to another, or to itself: one can talk about the *unmicrowaveability* of some French fries or a *toothbrush-holder fastener box* in which to keep one's toothbrush-holder fasteners. This makes the number of possible words in a language bigger than immense; like the number of sentences, it is infinite. Putting aside fanciful coinages concocted for immortality in *Guinness,* a candidate for the longest word to date in English might be *floccinaucinihilipilification,* defined in the *Oxford English Dictionary* as "the categorizing of something as worthless or trivial." But that is a record meant to be broken:

floccinaucinihilipilificational: pertaining to the categorizing of something as worthless or trivial

floccinaucinihilipilificationalize: to cause something to pertain to the categorizing of something as worthless or trivial

floccinaucinihilipilificationalization: the act of causing something to pertain to the categorizing of something as worthless or trivial *floccinaucinihilipilificationalizationai.* pertaining to the act of causing something to pertain to the categorizing of something as worthless or trivial

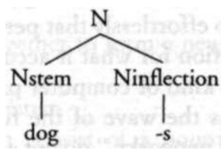
floccinaucinihilipilificationalizationalize: to cause something to pertain to the act of causing something to pertain . . .

Or, if you suffer from sesquipedaliaphobia, you can think of your *great-grandmother,* your *great-great-grandmother,* your *great-great-great-grandmother,* and so on, limited only in practice by the number of generations since Eve.

What's more, words, like sentences, are too delicately layered to be generated by a chaining device (a system that selects an item from one list, then moves on to some other list, then to another). When Ronald Reagan proposed the Strategic Defense Initiative, popularly known as Star Wars, he imagined a future in which an incoming Soviet missile would be shot down by an *anti-missile missile.* But critics pointed out that the Soviet Union could counterattack with an *anti-anti-missile-missile missile.* No problem, said his MIT-educated engineers; we'll just build an *anti-anti-anti-missile-missile-missile missile.* These high-tech weapons need a high-tech grammar—something that can keep track of all the *anti's* at the beginning of the word so that it can complete the word with an equal number of *missile's,* plus one, at the end. A word structure grammar (a phrase structure grammar for words) that can embed a word in between an *anti-* and its *missile* can achieve these objectives; a chaining device cannot, because it has forgotten the pieces that it laid down at the beginning of the long word by the time it gets to the end.

Like syntax, morphology is a cleverly designed system, and many of the seeming oddities of words are predictable products of its internal logic. Words have a delicate anatomy consisting of pieces, called morphemes, that fit together in certain ways. The word structure system is an extension of the X-bar phase structure system, in which big nounish things are built out of smaller nounish things, smaller nounish things are built out of still smaller nounish things, and so on. The biggest phrase involving nouns is the noun phrase; a noun phrase contains an N-bar; an N-bar contains a noun—the word. Jumping from syntax to morphology, we simply continue the dissection, analyzing the word into smaller and smaller nounish pieces.

Here is a picture of the structure of the word *dogs:*



The top of this mini-tree is "N" for "noun"; this allows the clocking maneuver in which the whole word can be plugged into the noun slot inside any noun phrase. Down inside the word, we have two parts: the bare word form *dog,* usually called the stem, and the plural inflection *-s*. The rule responsible for inflected words (the rule of wug-test fame) is simply

> N —> Nstem Ninflection
> "A noun can consist of a noun stem followed by a noun inflection."

The rule nicely interfaces with the mental dictionary: *dog* would be listed as a noun stem meaning "dog," and *-s* would be listed as a noun inflection meaning "plural of."

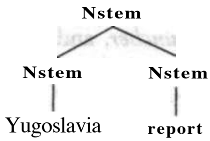
This rule is the simplest, most stripped-down example of anything we would want to call a rule of grammar. In my laboratory we use it as an easily studied instance of mental grammar, allowing us to document in great detail the psychology of linguistic rules from infancy to old age in both normal and neurologically impaired people, in much the same way that biologists focus on the fruit fly *Drosophila* to study the machinery of genes. Though simple, the rule that glues an inflection to a stem is a surprisingly powerful computational operation. That is because it recognizes an abstract mental symbol, like "noun stem," instead of being associated with a particular list of words or a particular list of sounds or a particular list of meanings. We can use the rule to inflect any item in the mental dictionary that lists "noun stem" in its entry, without caring what the word means;

we can convert not only *dog* to *dogs* but also *hour* to *hours* and *justification* to *justifications*. Likewise, the rule allows us to form plurals without caring what the word sounds like; we pluralize unusual-sounding words as in *the Gorbachevs, the Bachs,* and *the Mao Zedongs*. For the same reason, the rule is perfectly happy applying to brand-new nouns, like *faxes, dweebs, wugs,* and *zots*.

We apply the rule so effortlessly that perhaps the only way I can drum up some admiration for what it accomplishes is to compare humans with a certain kind of computer program that many computer scientists tout as the wave of the future. These programs, called "artificial neural networks," do not apply a rule like the one I have just shown you. An artificial neural network works by analogy, converting *wug* to *wugged* because it is vaguely similar to *hug-hugged, walk-walked,* and thousands of other verbs the network has been trained to recognize. But when the network is faced with a new verb that is unlike anything it has previously been trained on, it often mangles it, because the network does not have an abstract, all-embracing category "verb stem" to fall back on and add an affix to. Here are some comparisons between what people typically do and what artificial neural networks typically do when given a wag-test:

VERB	TYPICAL PAST-TENSE FORM GIVEN BY PEOPLE	TYPICAL PAST-TENSE FORM GIVEN BY NEURAL NETWORKS
mail	mailed	membled
conflict	conflicted	conflafted
wink	winked	wok
quiver	quivered	quess
satisfy	satisfied	sedderded
smairf	smairfed	sprurice
trilb	trilbed	treelilt
smeej	smeejed	leefloag
frilg	frilged	freezled

Stems can be built out of parts, too, in a second, deeper level of word assembly. In compounds like *Yugoslavia report, sushi-lover, broccoli-green,* and *toothbrush,*



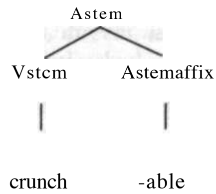
two stems are joined together to form a new stem, by the rule

> Nstem —> Nstem Nstem
>
> "A noun stem can consist of a noun stem followed by another noun stem."

In English, a compound is often spelled with a hyphen or by running its two words together, but it can also be spelled with a space between the two components as if they were still separate words. This confused your grammar teacher into telling you that in *Yugoslavia report,* "Yugoslavia" is an adjective. To see that this can't be right, just try comparing it with a real adjective like *interesting*. You can say *This report seems interesting* but not *This report seems Yugoslavia!* There is a simple way to tell whether something is a compound word or a phrase: compounds generally have stress on the first word, phrases on the second. A *dark room* (phrase) is any room that is dark, but a *dark room* (compound word) is where photographers work, and a darkroom can be lit when the photographer is done. A *black board* (phrase) is necessarily a board that is black, but some *blackboards* (compound word) are green or even white. Without pronunciation or punctuation as a guide, some word strings can be read either as a phrase or as a compound, like the following headlines:

> Squad Helps Dog Bite Victim
> Man Eating Piranha Mistakenly Sold as Pet Fish
> Juvenile Court to Try Shooting Defendant

New stems can also be formed out of old ones by adding affixes (prefixes and suffixes), like the *-al, -ize,* and *-ation* I used recursively to get longer and longer words ad infinitum (as in *sensationalizationalization).* For example, *-able* combines with any verb to create an adjective, as in *crunch-crunchable*. The suffix *-er* converts any verb to a noun, as in *crunch-cruncher,* and the suffix *-ness* converts any adjective into a noun, as in *crunchy-crunchiness*.



The rule forming them is

> Astern → Stem Astemaffix
> "An adjective stem can consist of a stem joined to a suffix."

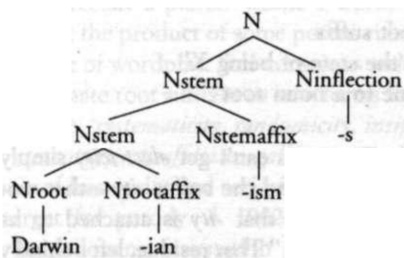
and a suffix like *-able* would have a mental dictionary entry like the following:

> *-able:*
> adjective stem affix
> means "capable of being X'd"
> attach me to a verb stem

Like inflections, stem affixes are promiscuous, mating with any stem that has the right category label, and so we have *crunchable, scrunchable, shmooshable, wuggable,* and so on. Their meanings are predictable: capable of being crunched, capable of being scrunched, capable of being shmooshed, even capable of being "wugged," whatever *wug* means. (Though I can think of an exception: in the sentence *I asked him what he thought of my review of his book, and his response was unprintable,* the word *unprintable* means something much more specific than "incapable of being printed.")

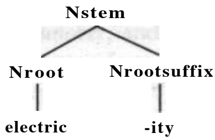
The scheme for computing the meaning of a stem out of the meaning of its parts is similar to the one used in syntax: one special element is the "head," and it determines what the conglomeration refers to. Just as the phrase *the cat in the hat* is a kind of cat, showing that *cat* is its head, a *Yugoslavia report* is a kind of report, and *shmooshability* is a kind of ability, so *report* and *-ability* must be the heads of those words. The head of an English word is simply its rightmost morpheme.

Continuing the dissection, we can tease stems into even smaller parts. The smallest part of a word, the part that cannot be cut up into any smaller parts, is called its root. Roots can combine with special suffixes to form stems. For example, the root *Darwin* can be found inside the stem *Darwinian*. The stem *Darwinian* in turn can be fed into the suffixing rule to yield the new stem *Darwinianism*. From there, the inflectional rule could even give us the word *Darwinianisms,* embodying all three levels of word structure:



Interestingly, the pieces fit together in only certain ways. Thus *Darwinism,* a stem formed by the stem suffix *-ism,* cannot be a host for *-ian,* because *-ian* attaches only to roots; hence *Darwinismian* (which would mean "pertaining to Darwinism") sounds ridiculous. Similarly, *Darwinsian* ("pertaining to the two famous Darwins, Charles and Erasmus"), *Darwinsianism,* and *Darwinsism* are quite impossible, because whole inflected words cannot have any root or stem suffixes joined to them.

Down at the bottommost level of roots and root affixes, we have entered a strange world. Take *electricity*. It seems to contain two parts, *electric* and *-ity:*



But are these words really assembled by a rule, gluing a dictionary entry for *-ity* onto the root *electric,* like this?

> Nstem —> Nroot Nrootsuffix
> "A noun stem can be composed of a noun root and a suffix."
>
> *-ity:*
> noun root suffix
> means "the state of being X"
> attach me to a noun root

Not this time. First, you can't get *electricity* simply by gluing together the word *electric* and the suffix *-ity*—that would sound like "electrick itty." The root that *-ity* is attached to has changed its pronunciation to "electríss." That residue, left behind when the suffix has been removed, is a root that cannot be pronounced in isolation.

Second, root-affix combinations have unpredictable meanings; the neat scheme for interpreting the meaning of the whole from the meaning of the parts breaks down. *Complexity* is the state of being complex, but *electricity* is not the state of being electric (you would never say that the electricity of this new can opener makes it convenient); it is the force powering something electric. Similarly, *instrumental* has nothing to do with instruments, *intoxicate* is not about toxic substances, one does not recite at a *recital,* and a five-speed *transmission* is not an act of transmitting.

Third, the supposed rule and affix do not apply to words freely, unlike the other rules and affixes we have looked at. For example, something can be *academic* or *acrobatic* or *aerodynamic* or *alcoholic,* but *academicity, acrobaticity, aerodynamicity,* and *alcoholicity* sound horrible (to pick just the first four words ending in *-ic* in my electronic dictionary).

So at the third and most microscopic level of word structure, roots and their affixes, we do not find bona fide rules that build words according to predictable formulas, *wug-style*. The stems seem to be stored in the mental dictionary with their own idiosyncratic meanings attached. Many of these complex stems originally were formed after the Renaissance, when scholars imported many words and suffixes into English from Latin and French, using some of the rules appropriate to those languages of learning. We have inherited the words, but not the rules. The reason to think that modern English speakers mentally analyze these words as trees at all, rather than as homogeneous strings of sound, is that we all sense that there is a natural break point between the *electric* and the *-ity*. We also recognize that there is an affinity between the word *electric* and the word *electricity,* and we recognize that any other word containing *-ity* must be a noun.

Our ability to appreciate a pattern inside a word, while knowing that the pattern is not the product of some potent rule, is the inspiration for a whole genre of wordplay. Self-conscious writers and speakers often extend Latinate root suffixes to new forms by analogy, such as *religiosity, criticality, systematicity, randomicity, insipidify, calumniate, conciliate, stereotypy, disaffiliate, gallonage,* and *Shavian*. The words have an air of heaviosity and seriosity about them, making the style an easy target for parody. A 1982 editorial cartoon by Jeff MacNelly put the following resignation speech into the mouth of Alexander Haig, the malaprop-prone Secretary of State:

> I decisioned the necessifaction of the resignatory action/option due to the dangerosity of the trendflowing of foreign policy away from our originatious careful coursing towards consistensivity, purposity, steadfastnitude, and above all, clarity.

Another cartoon, by Tom Toles, showed a bearded academician explaining the reason verbal Scholastic Aptitude Test scores were at an all-time low:

> Incomplete implementation of strategized programmatics designated to maximize acquisition of awareness and utilization of communications skills pursuant to standardized review and assessment of languaginal development.

In the culture of computer programmers and managers, this analogy-making is used for playful precision, not pomposity. *The New*

Hacker's Dictionary, a compilation of hackish jargon, is a near-exhaustive catalogue of the not-quite-freely-extendible root affixes in English:

> *ambimoustrous* adj. Capable of operating a mouse with either hand.
> *barfulous* adj. Something that would make anyone barf.
> *bogosity* n. The degree to which something is bogus.
> *bogotify* v. To render something bogus.
> *bozotic* adj. Having the quality of Bozo the Clown.
> *cuspy* adj. Functionally elegant.
> *depeditate* v. To cut the feet off of (e.g., while printing the bottom of a page).
> *dimwittery* n. Example of a dim-witted statement.
> *geekdom* n. State of being a techno-nerd.
> *marketroid* n. Member of a company's marketing department.
> *mumblage* n. The topic of one's mumbling.
> *pessimal* adj. Opposite of "optimal."
> *wedgitude* n. The state of being wedged (stuck; incapable of proceeding without help).
> *wizardly* adj. Pertaining to expert programmers.

Down at the level of word roots, we also find messy patterns in irregular plurals like *mouse-mice* and *man-men* and in irregular past-tense forms like *drink-drank* and *seek-sought*. Irregular forms tend to come in families, like *drink-drank, sink-sank, shrink-shrank, stink-stank, sing-sang, ring-rang, spring-sprang, swim-swam,* and *sit-sat,* or *blow-blew, know-knew, grow-grew, throw-threw, fly-flew,* and *slay-slew*. This is because thousands of years ago Proto-Indo-European, the language ancestral to English and most other European languages, had rules that replaced one vowel with another to form the past tense, just as we now have a rule that adds *-ed*. The irregular or "strong" verbs in modern English are mere fossils of these rules; the rules themselves are dead and gone. Most verbs that would seem eligible to belong to the irregular families are arbitrarily excluded, as we see in the following doggerel:

Sally Salter, she was a young teacher who taught,
And her friend, Charley Church, was a preacher who praught;
Though his enemies called him a screecher, who scraught.

His heart, when he saw her, kept sinking, and sunk;
And his eye, meeting hers, began winking, and wunk;
While she in her turn, fell to thinking, and thunk.

In secret he wanted to speak, and he spoke,
To seek with his lips what his heart long had soke,
So he managed to let the truth leak, and it loke.

The kiss he was dying to steal, then he stole;
At the feet where he wanted to kneel, then he knole;
And he said, "I feel better than ever I fole."

People must simply be memorizing each past-tense form separately. But as this poem shows, they can be sensitive to the patterns among them and can even extend the patterns to new words for humorous effect, as in Haigspeak and hackspeak. Many of us have been tempted by the cuteness of *sneeze-snoze, squeeze-squoze, take-took-tooken,* and *shit-shat,* which are based on analogies with *freeze-froze, break-broke-broken,* and *sit-sat*. In *Crazy English* Richard Lederer wrote an essay called "Foxen in the Henhice," featuring irregular plurals gone mad: *booth-beeth, harmonica-harmonicae, mother-methren, drum-dra, Kleenex-Kleenices,* and *bathtub-bathtubim*. Hackers speak of *faxen, VAXen, boxen, meece,* and *Macinteesh*. *Newsweek* magazine once referred to the white-caped, rhinestone-studded Las Vegas entertainers as *Elvii*. In the *Peanuts* comic strip, Linus's teacher Miss Othmar once had the class glue eggshells into model *igli*. Maggie Sullivan wrote an article in the *New York Times* calling for "strengthening" the English language by conjugating more verbs as if they were strong:

Subdue, subdid, subdone: Nothing could have subdone him the way her violet eyes subdid him.

Seesaw, sawsaw, seensaw: While the children sawsaw, the old man thought of long ago when he had seensaw.

Pay, pew, pain: He had pain for not choosing a wife more carefully.

Ensnare, ensnore, ensnorn: In the 60's and 70's, Sominex ads ensnore many who had never been ensnorn by ads before.

Commemoreat, commemorate, commemoreaten: At the banquet to commemoreat Herbert Hoover, spirits were high, and by the end of the evening many other Republicans had been commemoreaten.

In Boston there is an old joke about a woman who landed at Logan Airport and asked the taxi driver, "Can you take me someplace where I can get scrod?" He replied, "Gee, that's the first time I've heard it in the pluperfect subjunctive."

Occasionally a playful or cool-sounding form will catch on and spread through the language community, as *catch-caught* did several hundred years ago on the analogy of *teach-taught* and as *sneak-snuck* is doing today on the analogy of *stick-stuck*. (I am told that *has tooken* is the preferred form among today's mall rats.) This process can be seen clearly when we compare dialects, which retain the products of their own earlier fads. The curmudgeonly columnist H. L. Mencken was also a respectable amateur linguist, and he documented many past-tense forms found in American regional dialects, like *heat-het* (similar to *bleed-bled), drag-drug (dig-dug),* and *help-holp (tell-told)*. Dizzy Dean, the St. Louis Cardinals pitcher and CBS announcer, was notorious for saying "He slood into second base," common in his native Arkansas. For four decades English teachers across the nation engaged in a letter-writing campaign to CBS demanding that he be removed, much to his delight. One of his replies, during the Great Depression, was "A lot of folks that ain't sayin' 'ain't' ain't eatin'." Once he baited them with the following play-by-play:

> The pitcher wound up and flang the ball at the batter. The batter swang and missed. The pitcher flang the ball again and this time the batter connected. He hit a high fly right to the center fielder The center fielder was all set to catch the ball, but at the last minute his eyes were blound by the sun and he dropped it!

But successful adoptions of such creative extensions are rare; irregulars remain mostly as isolated oddballs.

Irregularity in grammar seems like the epitome of human eccentricity and quirkiness. Irregular forms are explicitly abolished in "rationally designed" languages like Esperanto, Orwell's Newspeak, and Planetary League Auxiliary Speech in Robert Heinlein's science fiction novel *Time for the Stars*. Perhaps in defiance of such regimentation, a woman in search of a nonconformist soulmate recently wrote this personal ad in the *New York Review of Books:*

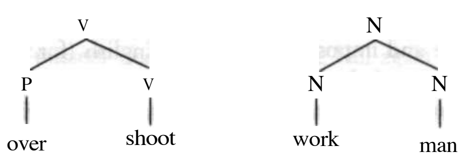
> Are you an irregular verb who believes nouns have more power than adjectives? Unpretentious, professional DWF, 5 yr. European resident, sometime violinist, slim, attractive, with married children... . Seeking sensitive, sanguine, youthful man, mid 50's-60's, health-conscious, intellectually adventurous, who values truth, loyalty, and openness.

A general statement of irregularity and the human condition comes from the novelist Marguerite Yourcenar: "Grammar, with its mixture of logical rule and arbitrary usage, proposes to a young mind a foretaste of what will be offered to him later on by law and ethics, those sciences of human conduct, and by all the systems wherein man has codified his instinctive experience."

For all its symbolism about the freewheeling human spirit, though, irregularity is tightly encapsulated in the word-building system; the system as a whole is quite cuspy. Irregular forms are roots, which are found inside stems, which are found inside words, some of which can be formed by regular inflection. This layering not only predicts many of the possible and impossible words of English (for example, why *Darwinianism* sounds better than *Darwinismian);* it provides a neat explanation for many trivia questions about seemingly illogical usage, such as: Why in baseball is a batter said to have *flied out*—why has no mere mortal ever *flown out* to center field? Why is the hockey team in Toronto called the *Maple Leafs* and not the *Maple Leaves?* Why do many people say *Walkmans,* rather than *Walkmen,* as the plural of *Walkman?* Why would it sound odd for someone to say that all of his daughter's friends are *low-lives?*

Consult any style manual or how-to book on grammar, and it will give one of two explanations as to why the irregular is tossed aside—both wrong. One is that the books are closed on irregular words in English; any new form added to the language must be regular. Not true: if I coin new words like *to re-sing* or *to out-sing,* their pasts are *re-sang* and *out-sang,* not *re-singed* and *out-singed.* Similarly, I recently read that there are peasants who run around with small tanks in China's oil fields, scavenging oil from unguarded wells; the article calls them *oil-mice,* not *oil-mouses.* The second explanation is that when a word acquires a new, nonliteral sense, like baseball's *fly out,* that sense requires a regular form. The oil-mice clearly falsify that explanation, as do the many other metaphors based on irregular nouns, which steadfastly keep their irregularity: *sawteeth* (not *saw-tooths), Freud's intellectual children* (not *childs), snowmen* (not *snow-mans),* and so on. Likewise, when the verb *to blow* developed slang meanings like *to blow him away* (assassinate) and *to blow it off* (dismiss casually), the past-tense forms remained irregular: *blew him away* and *blew off the exam,* not *blowed him away* and *blowed off the exam.*

The real rationale for *flied out* and *Walkmans* comes from the algorithm for interpreting the meanings of complex words from the meanings of the simple words they are built out of. Recall that when a big word is built out of smaller words, the big word gets all its properties from one special word sitting inside it at the extreme right: the head. The head of the verb *to overshoot* is the verb *to shoot,* so *overshooting* is a kind of *shooting,* and it is a verb, because *shoot* is a verb. Similarly, a *workman* is a singular noun, because *man,* its head, is a singular noun, and it refers to a kind of man, not a kind of work. Here is what the word structures look like:



Crucially, the percolation conduit from the head to the top node applies to *all* the information stored with the head word: not just its nounhood or verbhood, and not just its meaning, but any irregular form that is stored with it, too. For example, part of the mental dictionary entry for *shoot* would say "I have my own irregular past-tense form, *shot*." This bit of information percolates up and applies to the complex word, just like any other piece of information. The past tense of *overshoot* is thus *overshot* (not *overshooted).* Likewise, the word *man* bears the tag "My plural is *men*." Since *man* is the head of *workman,* the tag percolates up to the N symbol standing for *workman,* and so the plural of *workman* is *workmen*. This is also why we get *out-sang, oil-mice, sawteeth,* and *blew him away*.

Now we can answer the trivia questions. The source of quirkiness in words like *fly out* and *Walkmans* is their *headlessness*. A headless word is an exceptional item that, for one reason or another, differs in some property from its rightmost element, the one it would be based on if it were like ordinary words. A simple example of a headless word is a *low-life*—not a kind of life at all but a kind of person, namely one who leads a low life. In the word *low-life,* then, the normal percolation pipeline must be blocked. Now, a pipeline inside a word cannot be blocked for just one kind of information; if it is blocked for one thing, nothing passes through. If *low-life* does not get its meaning from *life,* it cannot get its plural from *life* either. The irregular form associated with *life,* namely *lives,* is trapped in the dictionary, with no way to bubble up to the whole word *low-life*. The all-purpose regular rule, "Add the *-s* suffix," steps in by default, and we get *low-lifes*. By similar unconscious reasoning, speakers arrive at *saber-tooths* (a kind of tiger, not a kind of tooth), *tenderfoots* (novice cub scouts, who are not a kind of foot but a kind of youngster that has tender *ieet),flatfoots* (also not a kind of foot but a slang term for policemen), and *still lifes* (not a kind of life but a kind of painting).

Since the Sony Walkman was introduced, no one has been sure whether two of them should be *Walkmen* or *Walkmans*. (The nonsexist alternative *Walkperson* would leave us on the hook, because we would be faced with a choice between *Walkpersons* and *Walkpeople.)* The temptation to say *Walkmans* comes from the word's being headless: a Walkman is not a kind of man, so it must not be getting its meaning from the word *man* inside it, and by the logic of headlessness it shouldn't receive a plural form from *man,* either. But it is hard to be comfortable with any kind of plural, because the relation between *Walkman* and *man* feels utterly obscure. It feels obscure because the word was not put together by any recognizable scheme. It is an example of the pseudo-English that is popular in Japan in signs and product names. (For example, one popular soft drink is called Sweat, and T-shirts have enigmatic inscriptions like CIRCUIT BEAVER, NURSE MENTALITY, and BONERACTIVE WEAR.) The Sony Corporation has an official answer to the question of how to refer to more than one Walkman. Fearing that their trademark, if converted to a noun, may become as generic as *aspirin* or *kleenex,* they sidestep the grammatical issues by insisting upon *Walkman Personal Stereos*.

What about flying out? To the baseball cognoscenti, it is not directly based on the familiar verb *to fly* ("to proceed through the air") but on the noun *a fly* ("a ball hit on a conspicuously parabolic trajectory"). To *fly out* means "to make an out by hitting a fly that gets caught." The noun *a fly,* of course, itself came from the verb *to fly*. The word-within-a-word-within-a-word structure can be seen in this bamboo-like tree:

Since the whole word, represented by its topmost label, is a verb, but the element it is made out of one level down is a noun, *to fly out,* like *low-life,* must be headless—if the noun *fly* were its head, *fly out* would have to be a noun, too, which it is not. Lacking a head and its associated data pipeline, the irregular forms of the original verb *to fly,* namely *flew* and *flown,* are trapped at the bottommost level and cannot bubble up to attach to the whole word. The regular *-ed* rule rushes in in its usual role as the last resort, and thus we say that Wade Boggs *flied out*. What kills the irregularity of *to fly out,* then, is not its specialized meaning, but its being a verb based on a word that is not a verb. By the same logic, we say *They ringed the city with artillery* ("formed a ring around it"), not *They rang the city with artillery,* and *He grandstanded to the crowd* ("played to the grandstand"), not *He grandstood to the crowd.*

This principle works every time. Remember Sally Ride, the astronaut? She received a lot of publicity because she was America's first woman in space. But recently Mae Jemison did her one better. Not only is Jemison America's first *black* woman in space, but she appeared in *People* magazine in 1993 in their list of the fifty most beautiful people in the world. Publicity-wise, she has out-Sally-Rided Sally Ride (not *has out-Sally-Ridden Sally Ride)*. For many years New York State's most infamous prison was Sing Sing. But since the riot at the Attica Correctional Facility in 1971, Attica has become even more infamous: it has out-Sing-Singed Sing Sing (not *has out-Sing-Sung Sing Sing)*.

As for the Maple Leafs, the noun being pluralized is not *leaf,* the unit of foliage, but a noun based on the *name* Maple Leaf, Canada's national symbol. A name is not the same thing as a noun. (For example, whereas a noun may be preceded by an article like *the,* a name may not be: you cannot refer to someone as *the Donald,* unless you are Ivana Trump, whose first language is Czech.) Therefore, the noun *a Maple Leaf* (referring to, say, the goalie) must be headless, because it is a noun based on a word that is not a noun. And a noun that does not get its nounhood from one of its components cannot get an irregular plural from that component either; hence it defaults to the regular form *Maple Leafs*. This explanation also answers a question that kept bothering David Letterman throughout one of his recent *Late Night* shows: why is the new major league baseball team in Miami called the Florida Marlins rather than the Florida Marlin, given that those fish are referred to in the plural as *marlin?* Indeed, the explanation applies to all nouns based on names:

I'm sick of dealing with all the *Mickey Mouses* in this administration, [not *Mickey Mice]*

Hollywood has been relying on movies based on comic book heroes and their sequels, like the three *Supermans* and the two *Batmans*. [not *Supermen* and *Batmen]*

Why has the second half of the twentieth century produced no *Thomas Manns?* [not *Thomas Menn]*

We're having Julia Child and her husband over for dinner tonight. You know, *the Childs* are great cooks. [not *the Children]*

Irregular forms, then, live at the bottom of word structure trees, where roots and stems from the mental dictionary are inserted. The developmental psycholinguist Peter Gordon has capitalized on this effect in an ingenious experiment that shows how children's minds seem to be designed with the logic of word structure built in.

Gordon focused on a seeming oddity first noticed by the linguist Paul Kiparsky: compounds can be formed out of irregular plurals but not out of regular plurals. For example, a house infested with mice can be described as *mice-infested,* but it sounds awkward to describe a house infested with rats as *rats-infested*. We say that it is *rat-infested,* even though by definition one rat does not make an infestation. Similarly, there has been much talk about *men-bashing* but no talk about *gays-bashing* (only *gay-bashing),* and there are *teethmarks,* but no *clawsmarks*. Once there was a song about a *purple-people-eater,* but it would be ungrammatical to sing about a *purple-babies-eater*. Since the licit irregular plurals and the illicit regular plurals have similar meanings, it must be the grammar of irregularity that makes the difference.

The theory of word structure explains the effect easily. Irregular plurals, because they are quirky, have to be stored in the mental dictionary as roots or stems; they cannot be generated by a rule. Because of this storage, they can be fed into the compounding rule that joins an existing stem to another existing stem to yield a new stem. But regular plurals are not stems stored in the mental dictionary; they are complex words that are assembled on the fly by inflectional rules whenever they are needed. They are put together too late in the root-to-stem-to-word assembly process to be available to the compounding rule, whose inputs can only come out of the dictionary.

Gordon found that three- to five-year-old children obey this restriction fastidiously. Showing the children a puppet, he first asked them, "Here is a monster who likes to eat mud. What do you call him?" He then gave them the answer, *a mud-eater,* to get them started. Children like to play along, and the more gruesome the meal, the more eagerly they fill in the blank, often to the dismay of their onlooking parents. The crucial parts came next. A "monster who likes to eat mice," the children said, was a *mice-eater*. But a "monster who likes to eat rats" was never called a *rats-eater,* only a *rat-eater*.

(Even the children who made the error *mouses* in their spontaneous speech never called the puppet a *mouses-eater.)* The children, in other words, respected the subtle restrictions on combining plurals and compounds inherent in the word structure rules. This suggests that the rules take the same form in the unconscious mind of the child as they do in the unconscious mind of the adult.

But the most interesting discovery came when Gordon examined how children might have acquired this constraint. Perhaps, he reasoned, they learned it from their parents by listening for whether the plurals that occur inside the parents' compounds are irregular, regular, or both, and then duplicate whatever kinds of compounds they hear. This would be impossible, he discovered. Motherese just doesn't have any compounds containing plurals. Most compounds are like *toothbrush,* with singular nouns inside them; compounds like *mice-infested,* though grammatically possible, are seldom used. The children produced *mice-eater* but never *rats-eater,* even though they had no evidence from adult speech that this is how languages work. We have another demonstration of knowledge despite "poverty of the input," and it suggests that another basic aspect of grammar may be innate. Just as Crain and Nakayama's Jabba experiment showed that in syntax children automatically distinguish between word strings and phrase structures, Gordon's mice-eater experiment shows that in morphology children automatically distinguish between roots stored in the mental dictionary and inflected words created by a rule.

A word, in a word, is complicated. But then what in the world *is* a word? We have just seen that "words" can be built out of parts by morphological rules. But then what makes them different from phrases or sentences? Shouldn't we reserve the word "word" for a thing that has to be rote-memorized, the arbitrary Saussurean sign that exemplifies the first of the two principles of how language works (the other being the discrete combinatorial system)? The puzzlement comes from the fact that the everyday word "word" is not scientifically precise. It can refer to two things.

The concept of a word that I have used so far in this chapter is a linguistic object that, even if built out of parts by the rules of morphology, behaves as the indivisible, smallest unit with respect to the rules of syntax—a "syntactic atom," in *atom's* original sense of something that cannot be split. The rules of syntax can look inside a sentence or phrase and cut and paste the smaller phrases inside it. For example, the rule for producing questions can look inside the sentence *This monster eats mice* and move the phrase corresponding to *mice* to the front, yielding "*What did this monster eat?* But the rules of syntax halt at the boundary between a phrase and a word; even if the word is built out of parts, the rules cannot look "inside" the word and fiddle with those parts. For example, the question rule cannot look inside the word *mice-eater* in the sentence *This monster is a mice-eater* and move the morpheme corresponding to *mice* to the front; the resulting question is virtually unintelligible: *What is this monster an -eater?* (Answer: mice.) Similarly, the rules of syntax can stick an adverb inside a phrase, as in *This monster eats mice quickly*. But they cannot stick an adverb inside a word, as in *This monster is a mice-quickly-eater*. For these reasons, we say that words, even if they are generated out of parts by one set of rules, are not the same thing as phrases, which are generated out of parts by a different set of rules. Thus one precise sense of our everyday term "word" refers to the units of language that are the products of morphological rules, and which are unsplittable by syntactic rules.

The second, very different sense of "word" refers to a rote-memorized chunk: a string of linguistic stuff that is arbitrarily associated with a particular meaning, one item from the long list we call the mental dictionary. The grammarians Anna Maria Di Sciullo and Edwin Williams coined the term "listeme," the unit of a memorized list, to refer to this sense of "word" (their term is a play on "morpheme," the unit of morphology, and "phoneme," the unit of sound). Note that a listeme need not coincide with the first precise sense of "word," a syntactic atom. A listeme can be a tree branch of any size, as long as it cannot be produced mechanically by rules and therefore has to be memorized. Take idioms. There is no way to predict the meaning of *kick the bucket, buy the farm, spill the beans, bite the bullet, screw the pooch, give up the ghost, hit the fan,* or *go bananas* from the meanings of their components using the usual rules of heads and role-players. *Kicking the bucket* is not a kind of kicking, and buckets have nothing to do with it. The meanings of these phrase-sized units have to be memorized as listemes, just as if they were simple word-sized units, and so they are really "words" in this second sense. Di Sciullo and Williams, speaking as grammatical chauvinists, describe the mental dictionary (lexicon) as follows: "If conceived of as the set of listemes, the lexicon is incredibly boring by its very nature. . . . The lexicon is like a prison—it contains only the lawless, and the only thing that its inmates have in common is their lawlessness."

In the rest of this chapter I turn to the second sense of "word," the listeme. It will be a kind of prison reform: I want to show that the lexicon, though a repository of lawless listemes, is deserving of respect and appreciation. What seems to a grammarian like an act of brute force incarceration—a child hears a parent use a word and thenceforth retains that word in memory—is actually an inspiring feat.

One extraordinary feature of the lexicon is the sheer capacity for memorization that goes into building it. How many words do you think an average person knows? If you are like most writers who have offered an opinion based on the number of words they hear or read, you might guess a few hundred for the uneducated, a few thousand for the literate, and as many as 15,000 for gifted wordsmiths like Shakespeare (that is how many distinct words are found in his collected plays and sonnets).

The real answer is very different. People can recognize vastly more words than they have occasion to use in some fixed period of time or space. To estimate the size of a person's vocabulary—in the sense of memorized listemes, not morphological products, of course, because the latter are infinite—psychologists use the following method. Start with the largest unabridged dictionary available; the smaller the dictionary, the more words a person might know but not get credit for. Funk & Wagnall's *New Standard Unabridged Dictionary,* to take an example, has 450,000 entries, a healthy number, but too many to test exhaustively. (At thirty seconds a word, eight hours a day, it would take more than a year to test a single person.) Instead, draw a sample—say, the third entry from the top of the first column on every eighth left-hand page. Entries often have many meanings, such as "*hard:* (1) firm; (2) difficult; (3) harsh; (4) toilsome . . ." and so on, but counting them would require making arbitrary decisions about how to lump or split the meanings. Thus it is practical only to estimate how many words a person has learned at least one meaning for, not how many meanings a person has learned altogether. The testee is presented with each word in the sample, and asked to choose the closest synonym from a set of alternatives. After a correction for guessing, the proportion correct is multiplied by the size of the dictionary, and that is an estimate of the person's vocabulary size.

Actually, another correction must be applied first. Dictionaries are consumer products, not scientific instruments, and for advertising purposes their editors often inflate the number of entries. ("Authoritative. Comprehensive. Over 1.7 million words of text and 160,000 definitions. Includes a 16-page full-color atlas.") They do it by including compounds and affixed forms whose meanings are predictable from the meanings of their roots and the rules of morphology, and thus are not true listemes. For example, my desk dictionary includes, together with *sail,* the derivatives *sailplane, sailer, sailless, sailing-boat,* and *sailcloth,* whose meanings I could deduce even if I had never heard them before.

The most sophisticated estimate comes from the psychologists William Nagy and Richard Anderson. They began with a list of 227,553 different words. Of these, 45,453 were simple roots and stems. Of the remaining 182,100 derivatives and compounds, they estimated that all but 42,080 could be understood in context by someone who knew their components. Thus there were a total of 44,453 + 42,080 = 88,533 listeme words. By sampling from this list and testing the sample, Nagy and Anderson estimated that an average American high school graduate knows 45,000 words—three times as many as Shakespeare managed to use! Actually, this is an underestimate, because proper names, numbers, foreign words, acronyms, and many common undecomposable compounds were excluded. There is no need to follow the rules of Scrabble in estimating vocabulary size; these forms are all listemes, and a person should be given credit for them. If they had been included, the average high school graduate would probably be credited with something like 60,000 words (a tetrabard?), and superior students, because they read more, would probably merit a figure twice as high, an octobard.

Is 60,000 words a lot or a little? It helps to think of how quickly they must have been learned. Word learning generally begins around the age of twelve months. Therefore, high school graduates, who have been at it for about seventeen years, must have been learning an average of ten new words a day continuously since their first birthdays, or about a new word every ninety waking minutes. Using similar techniques, we can estimate that an average six-year-old commands about 13,000 words (notwithstanding those dull, dull *Dick and Jane* reading primers, which were based on ridiculously lowball estimates). A bit of arithmetic shows that preliterate children, who are limited to ambient speech, must be lexical vacuum cleaners, inhaling a new word every two waking hours, day in, day out. Remember that we are talking about listemes, each involving an arbitrary pairing. Think about having to memorize a new batting average or treaty date or phone number every ninety minutes of your waking life since you took your first steps. The brain seems to be reserving an especially capacious storage space and an especially rapid transcribing mechanism for the mental dictionary. Indeed, naturalistic studies by the psychologist Susan Carey have shown that if you casually slip a new color word like *olive* into a conversation with a three-year-old, the child will probably remember something about it five weeks later.

Now think of what goes into each act of memorization. A word is the quintessential symbol. Its power comes from the fact that every member of a linguistic community uses it interchangeably in speaking and understanding. If you use a word, then as long as it is not too obscure I can take it for granted that if I later utter it to a third party, he will understand my use of it the same way I understood yours. I do not have to try the word back on you to see how you react, or test it out on every third party and see how they react, or wait for you to use it with third parties. This sounds more obvious than it is. After all, if I observe that a bear snarls before it attacks, I cannot expect to scare a mosquito by snarling at it; if I bang a pot and the bear flees, I cannot expect the bear to bang a pot to scare hunters. Even within our species, learning a word from another person is not just a case of imitating that person's behavior. Actions are tied to particular kinds of actors and targets of the action in ways that words are not. If a girl learns to flirt by watching her older sister, she does not flirt with the sister or with their parents but only with the kind of person that she observes to be directly affected by the sister's behavior. Words, in contrast, are a universal currency within a community. In order to learn to use a word upon merely hearing it used by others, babies must tacitly assume that a word is not merely a person's characteristic behavior in affecting the behavior of others, but a shared bidirectional symbol, available to convert meaning to sound by any person when the person speaks, and sound to meaning by any person when the person listens, according to the same code.

Since a word is a pure symbol, the relation between its sound and its meaning is utterly arbitrary. As Shakespeare (using a mere tenth of a percent of his written lexicon and a far tinier fraction of his mental one) put it,

> What's in a name? that which we call a rose
> By any other name would smell as sweet.

Because of that arbitrariness, there is no hope that mnemonic tricks might lighten the memorization burden, at least for words that are not built out of other words. Babies should not, and apparently do not, expect *cattle* to mean something similar to *battle,* or *singing* to be like *stinging,* or *coats* to resemble *goats*. Onomatopoeia, where it is found, is of no help, because it is almost as conventional as any other word sound. In English, pigs go "oink"; in Japanese, they go "boo-boo." Even in sign languages the mimetic abilities of the hands are put aside and their configurations are treated as arbitrary symbols. Residues of resemblance between a sign and its referent can occasionally be discerned, but like onomatopoeia they are so much in the eye of ear of the beholder that they are of little use in learning. In American Sign Language the sign for "tree" is a motion of a hand as if it was a branch waving in the wind; in Chinese Sign Language "tree" is indicated by the motion of sketching a tree trunk.

The psychologist Laura Ann Petitto has a startling demonstration that the arbitrariness of the relation between a symbol and its meaning is deeply entrenched in the child's mind. Shortly before they turn two, English-speaking children learn the pronouns *you* and *me*. Often they reverse them, using *you* to refer to themselves. The error is forgivable. *You* and *me* are "deictic" pronouns, whose referent shifts with the speaker: *you* refers to you when I use it but to me when you use it. So children may need some time to get that down. After all, Jessica hears her mother refer to her, Jessica, using *you;* why should she not think that *you* means "Jessica"?

Now, in ASL the sign for "me" is a point to one's chest; the sign for "you" is a point to one's partner. What could be more transparent? One would expect that using "you" and "me" in ASL would be as foolproof as knowing how to point, which all babies, deaf and hearing, do before their first birthday. But for the deaf children Petitto studied, pointing is not pointing. The children used the sign of pointing to their conversational partners to mean "me" at exactly the age at which hearing children use the spoken sound *you* to mean "me." The children were treating the gesture as a pure linguistic symbol; the fact that it pointed somewhere did not register as being relevant. This attitude is appropriate in learning sign languages; in ASL, the pointing hand-shape is like a meaningless consonant or vowel, found as a component of many other signs, like "candy" and "ugly."

There is one more reason we should stand in awe of the simple act of learning a word. The logician W. V. O. Quine asks us to imagine a linguist studying a newly discovered tribe. A rabbit scurries by, and a native shouts, "Gavagai!" What does *gavagai* mean? Logically speaking, it needn't be "rabbit." It could refer to that particular rabbit (Flopsy, for example). It could mean any furry thing, any mammal, or any member of that species of rabbit (say, *Oryctolagus cuniculus),* or any member of that variety of that species (say, chinchilla rabbit). It could mean scurrying rabbit, scurrying thing, rabbit plus the ground it scurries upon, or scurrying in general. It could mean footprint-maker, or habitat for rabbit-fleas. It could mean the top half of a rabbit, or rabbit-meat-on-the-hoof, or possessor of at least one rabbit's foot. It could mean anything that is either a rabbit or a Buick. It could mean collection of undetached rabbit parts, or "Lo! Rabbithood again!," or "It rabbiteth," analogous to "It raineth."

The problem is the same when the child is the linguist and the parents are the natives. Somehow a baby must intuit the correct meaning of a word and avoid the mind-boggling number of logically impeccable alternatives. It is an example of a more general problem that Quine calls "the scandal of induction," which applies to scientists and children alike: how can they be so successful at observing a finite set of events and making some correct generalization about all future events of that sort, rejecting an infinite number of false generalizations that are also consistent with the original observations?

We all get away with induction because we are not open-minded logicians but happily blinkered humans, innately constrained to make only certain kinds of guesses—the probably correct kinds—about how the world and its occupants work. Let's say the word-learning baby has a brain that carves the world into discrete, bounded, cohesive objects and into the actions they undergo, and that the baby forms mental categories that lump together objects that are of the same kind. Let's also say that babies are designed to expect a language to contain words for kinds of objects and words for kinds of actions—nouns and verbs, more or less. Then the undetached rabbit parts, rabbit-trod ground, intermittent rabbiting, and other accurate descriptions of the scene will, fortunately, not occur to them as possible meanings of *gavagai*.

But could there really be a preordained harmony between the child's mind and the parent's? Many thinkers, from the woolliest mystics to the sharpest logicians, united only in their assault on common sense, have claimed that the distinction between an object and an action is not in the world or even in our minds, initially, but is imposed on us by our language's distinction between nouns and verbs. And if it is the word that delineates the thing and the act, it cannot be the concepts of thing and act that allow for the learning of the word.

I think common sense wins this one. In an important sense, there really are things and kinds of things and actions out there in the world, and our mind is designed to find them and to label them with words. That important sense is Darwin's. It's a jungle out there, and the organism designed to make successful predictions about what is going to happen next will leave behind more babies designed just like it. Slicing space-time into objects and actions is an eminently sensible way to make predictions given the way the world is put together. Conceiving of an extent of solid matter as a thing—that is, giving a single mentalese name to all of its parts—invites the prediction that those parts will continue to occupy some region of space and will move as a unit. And for many portions of the world, that prediction is correct. Look away, and the rabbit still exists; lift the rabbit by the scruff of the neck, and the rabbit's foot and the rabbit ears come along for the ride.

What about kinds of things, or categories? Isn't it true that no two individuals are exactly alike? Yes, but they are not arbitrary collections of properties, either. Things that have long furry ears and tails like pom-poms also tend to eat carrots, scurry into burrows, and breed like, well, rabbits. Lumping objects into categories—giving them a category label in mentalese—allows one, when viewing an entity, to infer some of the properties one cannot directly observe, using the properties one *can* observe. If Flopsy has long furry ears, he is a "rabbit"; if he is a rabbit, he might scurry into a burrow and quickly make more rabbits.

Moreover, it pays to give objects several labels in mentalese, designating different-sized categories like "cottontail rabbit," "rabbit," "mammal," "animal," and "living thing." There is a tradeoff involved in choosing one category over another. It takes less effort to determine that Peter Cottontail is an animal than that he is a cottontail (for example, an animallike motion will suffice for us to recognize that he is an animal, leaving it open whether or not he is a cottontail). But we can predict more new things about Peter if we know he is a cottontail than if we merely know he is an animal. If he is a cottontail, he likes carrots and inhabits open country or woodland clearings; if he is merely an animal, he could eat anything and live anywhere, for all one knows. The middle-sized or "basic-level" category "rabbit" represents a compromise between how easy it is to label something and how much good the label does you.

Finally, why separate the rabbit from the scurry? Presumably because there are predictable consequences of rabbithood that cut across whether it is scurrying, eating, or sleeping: make a loud sound, and in all cases it will be down a hole lickety-split. The consequences of making a loud noise in the presence of lionhood, whether eating or sleeping, are predictably different, and that is a difference that makes a difference. Likewise, scurrying has certain consequences regardless of who is doing it; whether it be rabbit or lion, a scurrier does not remain in the same place for long. With sleeping, a silent approach will generally work to keep a sleeper—rabbit or lion—motionless. Therefore a powerful prognosticator should have separate sets of mental labels for kinds of objects and kinds of actions. That way, it does not have to learn separately what happens when a rabbit scurries, what happens when a lion scurries, what happens when a rabbit sleeps, what happens when a lion sleeps, what happens when a gazelle scurries, what happens when a gazelle sleeps, and on and on; knowing about rabbits and lions and gazelles in general, and scurrying and sleeping in general, will suffice. With *m* objects and *n* actions, a knower needn't go through $m \times n$ learning experiences; it can get away with $m + n$ of them.

So even a wordless thinker does well to chop continuously flowing experience into things, kinds of things, and actions (not to mention places, paths, events, states, kinds of stuff, properties, and other types of concepts). Indeed, experimental studies of baby cognition have shown that infants have the concept of an object before they learn any words for objects, just as we would expect. Well before their first birthday, when first words appear, babies seem to keep track of the bits of stuff that we would call objects: they show surprise if the parts of an object suddenly go their own ways, of if the object magically appears or disappears, passes through another solid object, or hovers in the air without visible means of support.

Attaching words to these concepts, of course, allows one to share one's hard-won discoveries and insights about the world with the less experienced or the less observant. Figuring out which word to attach to which concept is the *gavagai* problem, and if infants start out with concepts corresponding to the kinds of meanings that languages use, the problem is partly solved. Laboratory studies confirm that young children assume that certain kinds of concepts get certain types of words, and other kinds of concepts cannot be the meaning of a word at all. The developmental psychologists Ellen Markman and Jeanne Hutchinson gave two- and three-year-old children a set of pictures, and for each picture asked them to "find another one that is the same as this." Children are intrigued by objects that interact, and when faced with these instructions they tend to select pictures that make groups of role-players like a blue jay and a nest or a dog and a bone. But when Markman and Hutchinson told them to "find another *dax* that is the same as this *dax*" the children's criterion shifted. A word must label a *kind* of thing, they seemed to be reasoning, so they put together a bird with another type of bird, a dog with another type of dog. For a child, a *dax* simply cannot mean "a dog or its bone," interesting though the combination may be.

Of course, more than one word can be applied to a thing: Peter Cottontail is not only a *rabbit* but an *animal* and a *cottontail*. Children have a bias to interpret nouns as middle-level kinds of objects like "rabbit," but they also must overcome that bias, to learn other types of words like *animal*. Children seem to manage this by being in sync with a striking feature of language. Though most common words have many meanings, few meanings have more than one word. That is, homonyms are plentiful, synonyms rare. (Virtually all supposed synonyms have some difference in meaning, however small. For example, *skinny* and *slim* differ in their connotation of desirability; *policeman* and *cop* differ in formality.) No one really knows why languages are so stingy with words and profligate with meanings, but children seem to expect it (or perhaps it is this expectation that causes it!), and that helps them further with the *gavagai* problem. If a child already knows a word for a kind of thing, then when another word is used for it, he or she does not take the easy but wrong way and treat it as a synonym. Instead, the child tries out some other possible concept. For example, Markman found that if you show a child a pair of pewter tongs and call it *biff,* the child interprets *biff* as meaning tongs in general, showing the usual bias for middle-level objects, so when asked for "more biffs," the child picks out a pair of plastic tongs. But if you show the child a pewter cup and call it *biff,* the child does not interpret *biff* as meaning "cup," because most children already know a word that means "cup," namely, *cup*. Loathing synonyms, the children guess that *biff* must mean something else, and the stuff the cup is made of is the next most readily available concept. When asked for more *biffs,* the child chooses a pewter spoon or pewter tongs.

Many other ingenious studies have shown how children home in on the correct meanings for different kinds of words. Once children know some syntax, they can use it to sort out different kinds of meaning. For example, the psychologist Roger Brown showed children a picture of hands kneading a mass of little squares in a bowl. If he asked them, "Can you see any sibbing?," the children pointed to the hands. If instead he asked them, "Can you see a sib?," they point to the bowl. And if he asked, "Can you see any sib?," they point to the stuff inside the bowl. Other experiments have uncovered great sophistication in children's understanding of how classes of words fit into sentence structures and how they relate to concepts and kinds.

So what's in a name? The answer, we have seen, is, a great deal. In the sense of a morphological product, a name is an intricate structure, elegantly assembled by layers of rules and lawful even at its quirkiest. And in the sense of a listeme, a name is a pure symbol, part of a cast of thousands, rapidly acquired because of a harmony between the mind of the child, the mind of the adult, and the texture of reality.

6

The Sounds of Silence

When I was a student I worked in a laboratory at McGill University that studied auditory perception. Using a computer, I would synthesize trains of overlapping tones and determine whether they sounded like one rich sound or two pure ones. One Monday morning I had an odd experience: the tones suddenly turned into a chorus of screaming munchkins. Like this: (beep boop-boop) (beep boop-boop) (beep boop-boop) HUMPTY-DUMPTY-HUMPTY-DUMPTY-HUMPTY-DUMPTY (beep boop-boop) (beep boop-boop) HUMPTY-DUMPTY-HUMPTY-DUMPTY-HUMPTY-HUMPTY-DUMPTY-DUMPTY (beep boop-boop) (beep boop-boop) (beep boop-boop) HUMPTY-DUMPTY (beep boop-boop) HUMPTY-HUMPTY-HUMPTY-DUMPTY (beep boop-boop). I checked the oscilloscope: two streams of tones, as programmed. The effect had to be perceptual. With a bit of effort I could go back and forth, hearing the sound as either beeps or munchkins. When a fellow student entered, I recounted my discovery, mentioning that I couldn't wait to tell Professor Bregman, who directed the laboratory. She offered some advice: don't tell anyone, except perhaps Professor Poser (who directed the psychopathology program).

Years later I discovered what I had discovered. The psychologists Robert Remez, David Pisoni, and their colleagues, braver men than I am, published an article in *Science* on "sine-wave speech." They synthesized three simultaneous wavering tones. Physically, the sound was nothing at all like speech, but the tones followed the same contours as the bands of energy in the sentence "Where were you a year ago?" Volunteers described what they heard as "science fiction sounds" or "computer bleeps." A second group of volunteers was told that the sounds had been generated by a bad speech synthesizer. They were able to make out many of the words, and a quarter of them could write down the sentence perfectly. The brain can hear speech content in sounds that have only the remotest resemblance to speech. Indeed, sine-wave speech is how mynah birds fool us. They have a valve on each bronchial tube and can control them independently, producing two wavering tones which we hear as speech.

Our brains can flip between hearing something as a bleep and hearing it as a word because phonetic perception is like a sixth sense. When we listen to speech the actual sounds go in one ear and out the other; what we perceive is *language*. Our experience of words and syllables, of the "b"-ness of *b* and the "ee"-ness of *ee,* is as separable from our experience of pitch and loudness as lyrics are from a score. Sometimes, as in sine-wave speech, the senses of hearing and phonetics compete over which gets to interpret a sound, and our perception jumps back and forth. Sometimes the two senses simultaneously interpret a single sound. If one takes a tape recording of *da,* electronically removes the initial chirplike portion that distinguishes the *da* from *ga* and *ka,* and plays the chirp to one ear and the residue to the other, what people hear is a chirp in one ear and *da* in the other—a single clip of sound is perceived simultaneously as *d*-ness and a chirp. And sometimes phonetic perception can transcend the auditory channel. If you watch an English-subtitled movie in a language you know poorly, after a few minutes you may feel as if you are actually understanding the speech. In the laboratory, researchers can dub a speech sound like *ga* onto a close-up video of a mouth articulating *va, ba, tha,* or *da*. Viewers literally *hear* a consonant like the one they see the mouth making—an astonishing illusion with the pleasing name "McGurk effect," after one of its discoverers.

Actually, one does not need electronic wizardry to create a speech illusion. All speech is an illusion. We hear speech as a string of separate words, but unlike the tree falling in the forest with no one to hear it, a word boundary with no one to hear it has no sound. In the speech sound wave, one word runs into the next seamlessly; there are no little silences between spoken words the way there are white spaces between written words. We simply hallucinate word boundaries when we reach the edge of a stretch of sound that matches some entry in our mental dictionary. This becomes apparent when we listen to speech in a foreign language: it is impossible to tell where one word ends and the next begins. The seamlessness of speech is also apparent in "oronyms," strings of sound that can be carved into words in two different ways:

> The good can decay many ways.
> The good candy came anyways.
>
> The stuffy nose can lead to problems.
> The stuff he knows can lead to problems.
>
> Some others I've seen.
> Some mothers I've seen.

Oronyms are often used in songs and nursery rhymes:

> I scream,
> You scream,
> We all scream
> For ice cream.
>
> Mairzey doats and dozey doats
> And little lamsey divey,
> A kiddley-divey do,
> Wouldn't you?
>
> Fuzzy Wuzzy was a bear,
> Fuzzy Wuzzy had no hair.
> Fuzzy Wuzzy wasn't fuzzy,
> Was he?
>
> In fir tar is,
> In oak none is.
> In mud eel is,
> In clay none is.
> Goats eat ivy.
> Mares eat oats.

And some are discovered inadvertently by teachers reading their students' term papers and homework assignments:

Jose can you see by the donzerly light? [Oh say can you see by the dawn's early light?]
It's a doggy-dog world. [dog-eat-dog]
Eugene O'Neill won a Pullet Surprise. [Pulitzer Prize]
My mother comes from Pencil Vanea. [Pennsylvania]
He was a notor republic. [notary public]
They played the Bohemian Rap City. [Bohemian Rhapsody]

Even the sequence of sounds we think we hear within a word are an illusion. If you were to cut up a tape of someone saying *cat,* you would not get pieces that sounded like *k, a,* and *t* (the units called "phonemes" that correspond roughly to the letters of the alphabet). And if you spliced the pieces together in the reverse order, they would be unintelligible, not *tack.* As we shall see, information about each component of a word is smeared over the entire word.

Speech perception is another one of the biological miracles making up the language instinct. There are obvious advantages to using the mouth and ear as a channel of communication, and we do not find any hearing community opting for sign language, though it is just as expressive. Speech does not require good lighting, face-to-face contact, or monopolizing the hands and eyes, and it can be shouted over long distances or whispered to conceal the message. But to take advantage of the medium of sound, speech has to overcome the problem that the ear is a narrow informational bottleneck. When engineers first tried to develop reading machines for the blind in the 1940s, they devised a set of noises that corresponded to the letters of the alphabet. Even with heroic training, people could not recognize the sounds at a rate faster than good Morse code operators, about three units a second. Real speech, somehow, is perceived an order of magnitude faster: ten to fifteen phonemes per second for casual speech, twenty to thirty per second for the man in the late-night Veg-O-Matic ads, and as many as forty to fifty per second for artificially sped-up speech. Given how the human auditory system works, this is almost unbelievable. When a sound like a click is repeated at a rate of twenty times a second or faster, we no longer hear it as a sequence of separate sounds but as a low buzz. If we can hear forty-five phonemes per second, the phonemes cannot possibly be consecutive bits of sound; each moment of sound must have several phonemes packed into it that our brains somehow unpack. As a result, speech is by far the fastest way of getting information into the head through the ear.

No human-made system can match a human in decoding speech. It is not for lack of need or trying. A speech recognizer would be a boon to quadriplegics and other disabled people, to professionals who have to get information into a computer while their eyes or hands are busy, to people who never learned to type, to users of telephone services, and to the growing number of typists who are victims of repetitive-motion syndromes. So it is not surprising that engineers have been working for more than forty years to get computers to recognize the spoken word. The engineers have been frustrated by a tradeoff. If a system has to be able to listen to many different people, it can recognize only a tiny number of words. For example, telephone companies are beginning to install directory assistance systems that can recognize anyone saying the word *yes,* or, in the more advanced systems, the ten English digits (which, fortunately for the engineers, have very different sounds). But if a system has to recognize a large number of words, it has to be trained to the voice of a single speaker. No system today can duplicate a person's ability to recognize both many words and many speakers. Perhaps the state of the art is a system called DragonDictate, which runs on a personal computer and can recognize 30,000 words. But it has severe limitations. It has to be trained extensively on the voice of the user. You . . . have . . . to . . . talk . . . to . . . it . . . like . . . this, with quarter-second pauses between the words (so it operates at about one-fifth the rate of ordinary speech). If you have to use a word that is not in its dictionary, like a name, you have to spell it out using the "Alpha, Bravo, Charlie" alphabet. And the program still garbles words about fifteen percent of the time, more than once per sentence. It is an impressive product but no match for even a mediocre stenographer.

The physical and neural machinery of speech is a solution to two problems in the design of the human communication system. A person might know 60,000 words, but a person's mouth cannot make 60,000 different noises (at least, not ones that the ear can easily discriminate). So language has exploited the principle of the discrete combinatorial system again. Sentences and phrases are built out of words, words are built out of morphemes, and morphemes, in turn, are built out of phonemes. Unlike words and morphemes, though, phonemes do not contribute bits of meaning to the whole. The meaning of *dog* is not predictable from the meaning of *d,* the meaning of o, the meaning of *g,* and their order. Phonemes are a different kind of linguistic object. They connect outward to speech, not inward to mentalese: a phoneme corresponds to an act of making a sound. A division into independent discrete combinatorial systems, one combining meaningless sounds into meaningful morphemes, the others combining meaningful morphemes into meaningful words, phrases, and sentences, is a fundamental design feature of human language, which the linguist Charles Hockett has called "duality of patterning."

But the phonological module of the language instinct has to do more than spell out the morphemes. The rules of language are discrete combinatorial systems: phonemes snap cleanly into morphemes, morphemes into words, words into phrases. They do not blend or melt or coalesce: *Dog bites man* differs from *Man bites dog,* and believing in God is different from believing in Dog. But to get these structures out of one head and into another, they must be converted to audible signals. The audible signals people can produce are not a series of crisp beeps like on a touch-tone phone. Speech is a river of breath, bent into hisses and hums by the soft flesh of the mouth and throat. The problems Mother Nature faced are digital-to-analog conversion when the talker encodes strings of discrete symbols into a continuous stream of sound, and analog-to-digital conversion when the listener decodes continuous speech back into discrete symbols.

The sounds of language, then, are put together in several steps. A finite inventory of phonemes is sampled and permuted to define words, and the resulting strings of phonemes are then massaged to make them easier to pronounce and understand before they are actually articulated. I will trace out these steps for you and show you how they shape some of our everyday encounters with speech: poetry and song, slips of the ear, accents, speech recognition machines, and crazy English spelling.

One easy way to understand speech sounds is to track a glob of air through the vocal tract into the world, starting in the lungs.

When we talk, we depart from our usual rhythmic breathing and take in quick breaths of air, then release them steadily, using the muscles of the ribs to counteract the elastic recoil force of the lungs. (If we did not, our speech would sound like the pathetic whine of a released balloon.) Syntax overrides carbon dioxide: we suppress the delicately tuned feedback loop that controls our breathing rate to regulate oxygen intake, and instead we time our exhalations to the length of the phrase or sentence we intend to utter. This can lead to mild hyperventilation or hypoxia, which is why public speaking is so exhausting and why it is difficult to carry on a conversation with a jogging partner.

The air leaves the lungs through the trachea (windpipe), which opens into the larynx (the voice-box, visible on the outside as the Adam's apple). The larynx is a valve consisting of an opening (the glottis) covered by two flaps of retractable muscular tissue called the vocal folds (they are also called "vocal cords" because of an early anatomist's error; they are not cords at all). The vocal folds can close off the glottis tightly, sealing the lungs. This is useful when we want to stiffen our upper body, which is a floppy bag of air. Get up from your chair without using your arms; you will feel your larynx tighten. The larynx is also closed off in physiological functions like coughing and defecation. The grunt of the weightlifter or tennis player is a reminder that we use the same organ to seal the lungs and to produce sound.

The vocal folds can also be partly stretched over the glottis to produce a buzz as the air rushes past. This happens because the high-pressure air pushes the vocal folds open, at which point they spring back and get sucked together, closing the glottis until air pressure builds up and pushes them open again, starting a new cycle. Breath is thus broken into a series of puffs of air, which we perceive as a buzz, called "voicing." You can hear and feel the buzz by making the sounds *ssssssss,* which lacks voicing, and *zzzzzzzz,* which has it.

The frequency of the vocal folds' opening and closing determines the pitch of the voice. By changing the tension and position of the vocal folds, we can control the frequency and hence the pitch. This is most obvious in humming or singing, but we also change pitch continuously over the course of a sentence, a process called intonation. Normal intonation is what makes natural speech sound different from the speech of robots in old science fiction movies and of the Coneheads on *Saturday Night Live*. Intonation is also controlled in sarcasm, emphasis, and an emotional tone of voice such as anger or cheeriness. In "tone languages" like Chinese, rising or falling tones distinguish certain vowels from others.

Though voicing creates a sound wave with a dominant frequency of vibration, it is not like a tuning fork or a test of the Emergency Broadcasting System, a pure tone with that frequency alone. Voicing is a rich, buzzy sound with many "harmonics." A male voice is a wave with vibrations not only at 100 cycles per second but also at 200 cps, 300 cps, 400 cps, 500 cps, 600 cps, 700 cps, and so on, all the way up to 4000 cps and beyond. A female voice has vibrations at 200 cps, 400 cps, 600 cps, and so on. The richness of the sound source is crucial—it is the raw material that the rest of the vocal tract sculpts into vowels and consonants.

If for some reason we cannot produce a hum from the larynx, any rich source of sound will do. When we whisper, we spread the vocal folds, causing the air stream to break apart chaotically at the edges of the folds and creating a turbulence or noise that sounds like hissing or radio static. A hissing noise is not a neatly repeating wave consisting of a sequence of harmonics, as we find in the periodic sound of a speaking voice, but a jagged, spiky wave consisting of a hodgepodge of constantly changing frequencies. This mixture, though, is all that the rest of the vocal tract needs for intelligible whispering. Some laryngectomy patients are taught "esophageal speech," or controlled burping, which provides the necessary noise. Others place a vibrator against their necks. In the 1970s the guitarist Peter Frampton funneled the amplified sound of his electric guitar through a tube into his mouth, allowing him to articulate his twangings. The effect was good for a couple of hit records before he sank into rock-and-roll oblivion.

The richly vibrating air then runs through a gantlet of chambers before leaving the head: the throat or "pharynx" behind the tongue, the mouth region between the tongue and palate, the opening between the lips, and an alternative route to the external world through the nose. Each chamber has a particular length and shape, which affects the sound passing through by the phenomenon called "resonance." Sounds of different frequencies have different wavelengths (the distance between the crests of the sound wave); higher pitches have shorter wavelengths. A sound wave moving down the length of a tube bounces back when it reaches the opening at the other end. If the length of the tube is a certain fraction of the wavelength of the sound, each reflected wave will reinforce the next incoming one; if it is of a different length, they will interfere with one another. (This is similar to how you get the best effect pushing a child on a swing if you synchronize each push with the top of the arc.) Thus a tube of a particular length amplifies some sound frequencies and filters out others. You can hear the effect when you fill a bottle. The noise of the sloshing water gets filtered by the chamber of air between the surface and the opening: the more water, the smaller the chamber, the higher the resonant frequency of the chamber, and the tinnier the gurgle.

What we hear as different vowels are the different combinations of amplification and filtering of the sound coming up from the larynx. These combinations are produced by moving five speech organs around in the mouth to change the shapes and lengths of the resonant cavities that the sound passes through. For example, *ee* is defined by two resonances, one from 200 to 350 cps produced mainly by the throat cavity, and the other from 2100 to 3000 cps produced mainly by the mouth cavity. The range of frequencies that a chamber filters is independent of the particular mixture of frequencies that enters it, so we can hear an *ee* as an *ee* whether it is spoken, whispered, sung high, sung low, burped, or twanged.

The tongue is the most important of the speech organs, making language truly the "gift of tongues." Actually, the tongue is three organs in one: the hump or body, the tip, and the root (the muscles that anchor it to the jaw). Pronounce the vowels in *bet* and *butt* repeatedly, *e-uh, e-uh, e-ub*. You should feel the body of your tongue moving forwards and backwards (if you put a finger between your teeth, you can feel it with the finger). When your tongue is in the front of your mouth, it lengthens the air chamber behind it in your throat and shortens the one in front of it in your mouth, altering one of the resonances: for the *bet* vowel, the mouth amplifies sounds near 600 and 1800 cps; for the *butt* vowel, it amplifies sounds near 600 and 1200. Now pronounce the vowels in *beet* and *bat* alternately. The body of your tongue will jump up and down, at right angles to the *bet-butt* motion; you can even feel your jaw move to help it. This, too, alters the shapes of the throat and mouth chambers, and hence their resonances. The brain interprets the different patterns of amplification and filtering as different vowels.

The link between the postures of the tongue and the vowels it sculpts gives rise to a quaint curiosity of English and many other languages called phonetic symbolism. When the tongue is high and at the front of the mouth, it makes a small resonant cavity there that amplifies some higher frequencies, and the resulting vowels like *ee* and *i* (as in *bit)* remind people of little things. When the tongue is low and to the back, it makes a large resonant cavity that amplifies some lower frequencies, and the resulting vowels like *a* in *father* and o in *core* and in *cot* remind people of large things. Thus mice are t*ee*ny and squeak, but elephants are hum*o*ngous and ro*a*r. Audio speakers have small tw*ee*ters for the high sounds and large w*oo*fers for the low ones. English speakers correctly guess that in Chinese *ch'ing* means light and *ch'ung* means heavy. (In controlled studies with large numbers of foreign words, the hit rate is statistically above chance, though just barely.) When I questioned our local computer wizard about what she meant when she said she was going to *frob* my workstation, she gave me this tutorial on hackerese. When you get a brand-new graphic equalizer for your stereo and aimlessly slide the knobs up and down to hear the effects, that is *frobbing*. When you move the knobs by medium-sized amounts to get the sound to your general liking, that is *twiddling*. When you make the final small adjustments to get it perfect, that is *tweaking*. The *ob, id,* and *eak* sounds perfectly follow the large-to-small continuum of phonetic symbolism.

And at the risk of sounding like Andy Rooney on *Sixty Minutes,* have you ever wondered why we say *fiddle-faddle* and not *faddle-fiddle?* Why is it *ping-pong* and *pitter-patter* rather than *pong-ping* and *patter-pitter?* Why *dribs and drabs,* rather than vice versa? Why can't a kitchen be *span and spic?* Whence *riff-raff, mish-mash, flim-flam, chit-chat, tit for tat, knick-knack, zig-zag, sing-song, ding-dong, King Kong, criss-cross, shilly-shally, see-saw, hee-haw, flip-flop, hippity-hop, tick-tock, tic-tac-toe, eeny-meeny-miney-moe, bric-a-brac, clickety-clack, hickory-dickory-dock, kit and kaboodle,* and *bibbity-bobbity-boo?* The answer is that the vowels for which the tongue is high and in the front always come before the vowels for which the tongue is low and in the back. No one knows why they are aligned in this order, but it seems to be a kind of syllogism from two other oddities. The first is that words that connote me-here-now tend to have higher and fronter vowels than verbs that connote distance from "me": *me* versus *you, here* versus *there, this* versus *that*. The second is that words that connote me-here-now tend to come before words that connote literal or metaphorical distance from "me" (or a prototypical generic speaker): *here and there* (not *there and here), this and that, now and then, father and son, man and machine, friend or foe, the Harvard-Yale game* (among Harvard students), *the Yale-Harvard game* (among Yalies), *Serbo-Croatian* (among Serbs), *Croat-Serbian* (among Croats). The syllogism seems to be: "me" = high front vowel; me first; therefore, high front vowel first. It is as if the mind just cannot bring itself to flip a coin in ordering words; if meaning does not determine the order, sound is brought to bear, and the rationale is based on how the tongue produces the vowels.

Let's look at the other speech organs. Pay attention to your lips when you alternate between the vowels in *boot* and *book*. For *boot,* you round the lips and protrude them. This adds an air chamber, with its own resonances, to the front of the vocal tract, amplifying and filtering other sets of frequencies and thus defining other vowel contrasts. Because of the acoustic effects of the lips, when we talk to a happy person over the phone, we can literally hear the smile.

Remember your grade-school teacher telling you that the vowel sounds in *bat, bet, bit, bottle,* and *butt* were "short," and the vowel sounds in *bait, beet, bite, boat,* and *boot* were "long"? And you didn't know what she was talking about? Well, forget it; her information is five hundreds years out of date. Older stages of English differentiated words by whether their vowels were pronounced quickly or were drawn out, a bit like the modern distinction between *bad* meaning "bad" and *baaaad* meaning "good." But in the fifteenth century English pronunciation underwent a convulsion called the Great Vowel Shift. The vowels that had simply been pronounced longer now became "tense": by advancing the tongue root (the muscles attaching the tongue to the jaw), the tongue becomes tense and humped rather than lax and flat, and the hump narrows the air chamber in the mouth above it, changing the resonances. Also, some tense vowels in modern English, like in *bite* and *brow,* are "diphthongs," two vowels pronounced in quick succession as if they were one: ba-eet, bra-oh.

You can hear the effects of the fifth speech organ by drawing out the vowel in *Sam* and *sat,* postponing the final consonant indefinitely. In most dialects of English, the vowels will be different: the vowel in *Sam* will have a twangy, nasal sound. That is because the soft palate or velum (the fleshy flap at the back of the hard palate) is opened, allowing air to flow out through the nose as well as through the mouth. The nose is another resonant chamber, and when vibrating air flows through it, yet another set of frequencies gets amplified and filtered. English does not differentiate words by whether their vowels are nasal or not, but many languages, like French, Polish, and Portuguese, do. English speakers who open their soft palate even when pronouncing *sat* are said to have a "nasal" voice. When you have a cold and your nose is blocked, opening the soft palate makes no difference, and your voice is the opposite of nasal.

So far we have just discussed the vowels—sounds where the air has clear passage from the larynx to the world. When some barrier is put in the way, one gets a consonant. Pronounce *ssssss*. The tip of your tongue—the sixth speech organ—is brought up almost against the gum ridge, leaving a small opening. When you force a stream of air through the opening, the air breaks apart turbulently, creating noise. Depending on the size of the opening and the length of the resonant cavities in front of it, the noise will have some of its frequencies louder than others, and the peak and range of frequencies define the sound we hear as *s*. This noise-making comes from the friction of moving air, so this kind of sound is called a fricative. When rushing air is squeezed between the tongue and palate, we get *sh;* between the tongue and teeth, *th;* and between the lower lip and teeth, *f*. The body of the tongue, or the vocal folds of the larynx, can also be positioned to create turbulence, defining the various "ch" sounds in languages like German, Hebrew, and Arabic *(Bach, Chanukah,* and so on).

Now pronounce a *t*. The tip of the tongue gets in the way of the airstream, but this time it does not merely impede the flow; it stops it entirely. When the pressure builds up, you release the tip of the tongue, allowing the air to pop out (flutists use this motion to demarcate musical notes). Other "stop" consonants can be formed by the lips (*p*), by the body of the tongue pressed against the palate *(k),* and by the larynx (in the "glottal" consonants in *uh-oh)*. What a listener hears when you produce a stop consonant is the following. First, nothing, as the air is dammed up behind the stoppage: stop consonants are the sounds of silence. Then, a brief burst of noise as the air is released; its frequency depends on the size of the opening and the resonant cavities in front of it. Finally, a smoothly changing resonance, as voicing fades in while the tongue is gliding into the position of whatever vowel comes next. As we shall see, this hop-skip-and-jump makes life miserable for speech engineers.

Finally, pronounce *m*. Your lips are sealed, just like for *p*. But this time the air does not back up silently; you can say *mmmmm* until you are out of breath. That is because you have also opened your soft palate, allowing all of the air to escape through your nose. The voicing sound is now amplified at the resonant frequencies of the nose and of the part of the mouth behind the blockage. Releasing the lips causes a sliding resonance similar in shape to what we heard for the release in *p*, except without the silence, noise burst, and fade-in. The sound *n* works similarly to *m*, except that the blockage is created by the tip of the tongue, the same organ used for *d* and *s*. So does the *ng* in *sing*, except that the body of the tongue does the job.

Why do we say *razzle-dazzle* instead of *dazzle-razzle?* Why *super-duper, helter-skelter, harum-scarum, hocus-pocus, willy-nilly, hully-gully, roly-poly, holy moly, herky-jerky, walkie-talkie, namby-pamby, mumbo-jumbo, loosey-goosey, wing-ding, wham-bam, hobnob, razzamatazz,* and *rub-a-dub-dub?* I thought you'd never ask. Consonants differ in "obstruency"—the degree to which they impede the flow of air, ranging from merely making it resonate, to forcing it noisily past an obstruction, to stopping it up altogether. The word beginning with the less obstruent consonant always comes before the word beginning with the more obstruent consonant. Why ask why?

Now that you have completed a guided tour up the vocal tract, you can understand how the vast majority of sounds in the world's languages are created and heard. The trick is that a speech sound is not a single gesture by a single organ. Every speech sound is a *combination* of gestures, each exerting its own pattern of sculpting of the sound wave, all executed more or less simultaneously—that is one of the reasons speech can be so rapid. As you may have noticed, a sound can be nasal or not, and produced by the tongue body, the tongue tip, or the lips, in all six possible combinations:

	Nasal (Soft Palate Open)	Not Nasal (Soft Palate Closed)
Lips	*m*	*p*
Tongue tip	*n*	*t*
Tongue body	*ng*	*k*

Similarly, voicing combines in all possible ways with the choice of speech organ:

	Voicing (Larynx Hums)	No Voicing (Larynx Doesn't Hum)
Lips	*b*	*p*
Tongue tip	*d*	*t*
Tongue body	*g*	*k*

Speech sounds thus nicely fill the rows and columns and layers of a multidimensional matrix. First, one of the six speech organs is chosen as the major articulator: the larynx, soft palate, tongue body, tongue tip, tongue root, or lips. Second, a manner of moving that articulator is selected: fricative, stop, or vowel. Third, configurations of the other speech organs can be specified: for the soft palate, nasal or not; for the larynx, voiced or not; for the tongue root, tense or lax; for the lips, rounded or unrounded. Each manner or configuration is a symbol for a set of commands to the speech muscles, and such symbols are called features. To articulate a phoneme, the commands must be executed with precise timing, the most complicated gymnastics we are called upon to perform.

English multiplies out enough of these combinations to define 40 phonemes, a bit above the average for the world's languages. Other languages range from 11 (Polynesian) to 141 (Khoisan or "Bushman"). The total inventory of phonemes across the world numbers in the thousands, but they are all defined as combinations of the six speech organs and their shapes and motions. Other mouth sounds are not used in any language: scraping teeth, clucking the tongue against the floor of the mouth, making raspberries, and squawking like Donald Duck, for instance. Even the unusual Khoisan and Bantu clicks (similar to the sound of *tsk-tsk* and made famous by the Xhosa pop singer Miriam Makeba) are not miscellaneous phonemes added to those languages. Clicking is a manner-of-articulation feature, like stop or fricative, and it combines with all the other features to define a new layer of rows and columns in the language's table of phonemes. There are clicks produced by the lips, tongue tip, and tongue body, any of which can be nasalized or not, voiced or not, and so on, as many as 48 click sounds in all!

An inventory of phonemes is one of the things that gives a language its characteristic sound pattern. For example, Japanese is famous for not distinguishing *r* from *l*. When I arrived in Japan on November 4, 1992, the linguist Masaaki Yamanashi greeted me with a twinkle and said, "In Japan, we have been very interested in Clinton's erection."

We can often recognize a language's sound pattern even in a speech stream that contains no real words, as with the Swedish chef on *The Muppets* or John Belushi's samurai dry cleaner. The linguist Sarah G. Thomason has found that people who claim to be channeling back to past lives or speaking in tongues are really producing gibberish that conforms to a sound pattern vaguely reminiscent of the claimed language. For example, one hypnotized channeler, who claimed to be a nineteenth-century Bulgarian talking to her mother about soldiers laying waste to the countryside, produced generic pseudo-Slavic gobbledygook like this:

> Ovishta reshta rovishta. Vishna beretishti? Ushna barishta dashto. Na darishnoshto. Korapshnoshashit darishtoy. Aobashni bedetpa.

And of course, when the words in one language are pronounced with the sound pattern of another, we call it a foreign accent, as in the following excerpt from a fractured fairy tale by Bob Belviso:

GIACCHE ENNE BINNESTAUCCHE

> Uans appona taim uase disse boi. Neimmese Giacche. Naise boi. Live uite ise mamma. Mainde da cao.
>
> Uane dei, di spaghetti ise olle ronne aute. Dei goine feinte fromme

no fudde. Mamma soi orais, "Oreie Giacche, teicche da cao enne traide erra forre bocchese spaghetti enne somme uaine."

Bai enne bai commese omme Giacche. I garra no fudde, i garra no uaine. Meichese misteicche, enne traidese da cao forre bonce binnese.

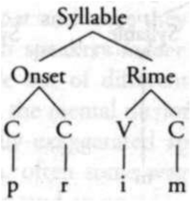
Giacchasse!

What defines the sound pattern of a language? It must be more than just an inventory of phonemes. Consider the following words:

ptak	thale	hlad
plaft	sram	mgla
vlas	flutch	dnom
rtut	toasp	nyip

All of the phonemes are found in English, but any native speaker recognizes that *thale, plaft,* and *flutch* are not English words but could be, whereas the remaining ones are not English words and could not be. Speakers must have tacit knowledge about how phonemes are strung together in their language.

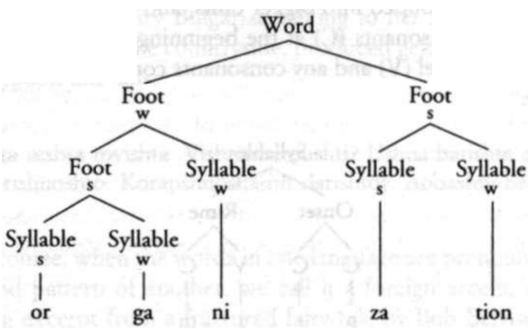
Phonemes are not assembled into words as one-dimensional left-to-right strings. Like words and phrases, they are grouped into units, which are then grouped into bigger units, and so on, defining a tree. The group of consonants (C) at the beginning of a syllable is called an onset; the vowel (V) and any consonants coming after it are called the rime:



The rules generating syllables define legal and illegal kinds of words in a language. In English an onset can consist of a cluster of consonants, like *flit, thrive,* and *spring,* as long as they follow certain restrictions. (For example, *vlit* and *sring* are impossible.) A rime can consist of a vowel followed by a consonant or certain clusters of consonants, as in *toast, lift,* and *sixths*. In Japanese, in contrast, an onset can have only a single consonant and a rime must be a bare vowel; hence *strawberry ice cream* is translated as *sutoroberi aisukurimo, girlfriend* as *garufurendo*. Italian allows some clusters of consonants in an onset but no consonants at the end of a rime. Belviso used this constraint to simulate the sound pattern of Italian in the Giacche story; *and* becomes *enne, from* becomes *fromme, beans* becomes *binnese*.

Onsets and rimes not only define the possible sounds of a language; they are the pieces of word-sound that are most salient to people, and thus are the units that get manipulated in poetry and word games. Words that rhyme share a rime; words that alliterate share an onset (or just an initial consonant). Pig Latin, eggy-peggy, aygo-paygo, and other secret languages of children tend to splice words at onset-rime boundaries, as does the Yinglish construction in *fancy-shmancy* and *Oedipus-Shmoedipus*. In the 1964 hit song "The Name Game" ("Noam Noam Bo-Boam, Bonana Fana Fo-Foam, Fee Fi Mo Moam, Noam"), Shirley Ellis could have saved several lines in the stanza explaining the rules if she had simply referred to onsets and rimes.

Syllables, in turn, are collected into rhythmic groups called feet:



Syllables and feet are classified as strong (s) and weak (w) by other rules, and the pattern of weak and strong branches determines how much stress each syllable will be given when it is pronounced. Feet, like onsets and rhymes, are salient chunks of word that we tend to manipulate in poetry and wordplay. Meter is defined by the kind of feet that go into a line. A succession of feet with a strong-weak pattern is a trochaic meter, as in *Mary had a little lamb;* a succession with a weak-strong pattern is iambic, as in *The rain in Spain falls mainly in the plain*. An argot popular among young ruffians contains forms like *fan-fuckin-tastic, abso-bloody-lutely, Phila-fuckin-delphia,* and *Kalama-fuckin-zoo*. Ordinarily, expletives appear in front of an emphatically stressed word; Dorothy Parker once replied to a question about why she had not been at the symphony lately by saying "I've been too fucking busy and vice versa." But in this lingo they are placed inside a single word, always in front of a stressed foot. The rule is followed religiously: *Philadel-fuckin-phia* would get you laughed out of the pool hall.

The assemblies of phonemes in the morphemes and words stored in memory undergo a series of adjustments before they are actually articulated as sounds, and these adjustments give further definition to the sound pattern of a language. Say the words *pat* and *pad*. Now add the inflection *-ing* and pronounce them again: *patting, padding*. In many dialects of English they are now pronounced identically; the original difference between the *t* and the *d* has been obliterated. What obliterated them is a phonological rule called flapping: if a stop consonant produced with the tip of the tongue appears between two vowels, the consonant is pronounced by flicking the tongue against the gum ridge, rather than keeping it there long enough for air pressure to build up. Rules like flapping apply not only when two morphemes are joined, like *pat* and *-ing;* they also apply to one-piece words. For many English speakers *ladder* and *latter,* though they "feel" like they are made out of different sounds and indeed are represented differently in the mental dictionary, are pronounced the same (except in artificially exaggerated speech). Thus when cows come up in conversation, often some wag will speak of an udder mystery, an udder success, and so on.

Interestingly, phonological rules apply in an ordered sequence, as if words were manufactured on an assembly line. Pronounce *write* and *ride*. In most dialects of English, the vowels differ in some way. At the very least, the *i* in *ride* is longer than the *i* in *write*. In some dialects, like the Canadian English of newscaster Peter Jennings, hockey star Wayne Gretzky, and yours truly (an accent satirized a few years back, eh, in the television characters Bob and Doug McKenzie), the vowels are completely different: *ride* contains a diphthong gliding from the vowel in *hot* to the vowel *ee; write* contains a diphthong gliding from the higher vowel in *hut* to *ee*. But regardless of exactly how the vowel is altered, it is altered in a consistent pattern: there are no words with long/low *i* followed by *t,* nor with short/high *i* followed by *d*. Using the same logic that allowed Lois Lane in her rare lucid moments to deduce that Clark Kent and Superman were the same, namely that they are never in the same place at the same time, we can infer that there is a single *i* in the mental dictionary, which is altered by a rule before being pronounced, depending on whether it appears in the company of *t* or *d*. We can even guess that the initial form stored in memory is like the one in *ride,* and that *write* is the product of the rule, rather than vice versa. The evidence is that when there is no *t* or *d* after the *i,* as in *rye,* and thus no rule disguising the underlying form, it is the vowel in *ride* that we hear.

Now pronounce *writing* and *riding*. The *t* and *d* have been made identical by the flapping rule. But the two ***i's*** are still different. How can that be? It is only the difference between *t* and *d* that causes a difference between the two ***i's,*** and that difference has been erased by the flapping rule. This shows that the rule that alters *i* must have applied *before* the flapping rule, while *t* and *d* were still distinct. In other words, the two rules apply in a fixed order, vowel-change before flapping. Presumably the ordering comes about because the flapping rule is in some sense there to make articulation easier and thus is farther downstream in the chain of processing from brain to tongue.

Notice another important feature of the vowel-altering rule. The vowel *i* is altered in front of many different consonants, not just *t*. Compare:

prize	price
five	fife
jibe	hype
geiger	biker

Does this mean there are five different rules that alter *i*—one for *z* versus *s,* one for *v* versus *f,* and so on? Surely not. The change-triggering consonants *t*, *s, f, p,* and *k* all differ in the same way from their counterparts *d, z, v, b,* and *g:* they are unvoiced, whereas the counterparts are voiced. We need only one rule, then: change *i* whenever it appears before an *unvoiced* consonant. The proof that this is the real rule in people's heads (and not just a way to save ink by replacing five rules with one) is that if an English speaker succeeds in pronouncing the German *ch* in *the Third Reich,* that speaker will pronounce the *ei* as in *write,* not as in *ride.* The consonant *ch* is not in the English inventory, so English speakers could not have learned any rule specifically applying to it. But it is an unvoiced consonant, and if the rule applies to any unvoiced consonant, an English speaker knows exactly what to do.

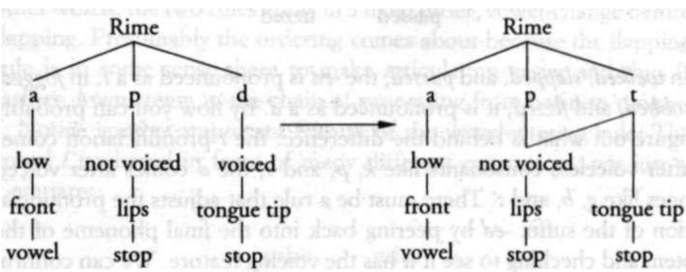
This selectivity works not only in English but in all languages. Phonological rules are rarely triggered by a single phoneme; they are triggered by an entire class of phonemes that share one or more features (like voicing, stop versus fricative manner, or which organ is doing the articulating). This suggests that rules do not "see" the phonemes in a string but instead look right through them to the features they are made from.

And it is features, not phonemes, that are manipulated by the rules. Pronounce the following past-tense forms:

walked	jogged
slapped	sobbed
passed	fizzed

In *walked, slapped,* and *passed,* the *-ed* is pronounced as a *t;* in *jogged, sobbed,* and *fizzed,* it is pronounced as a *d*. By now you can probably figure out what is behind the difference: the *t* pronunciation comes after voiceless consonants like *k, p,* and *s;* the *d* comes after voiced ones like *g, b,* and *z*. There must be a rule that adjusts the pronunciation of the suffix *-ed* by peering back into the final phoneme of the stem and checking to see if it has the voicing feature. We can confirm the hunch by asking people to pronounce *Mozart out-Bached Bach*. The verb *to out-Bach* contains the sound *ch,* which does not exist in English. Nonetheless everyone pronounces the *-ed* as a *t*, because the *ch* is unvoiced, and the rule puts a *t* next to any unvoiced consonant. We can even determine whether people store the *-ed* suffix as a t in memory and use the rule to convert it to a *d* for some words, or the other way around. Words like *play* and *row* have no consonant at the end, and everyone pronounces their past tenses like *plade* and *rode,* not *plate* and *rote*. With no stem consonant triggering a rule, we must be hearing the suffix in its pure, unaltered form in the mental dictionary, that is, *d*. It is a nice demonstration of one of the main discoveries of modern linguistics: a morpheme may be stored in the mental dictionary in a different form from the one that is ultimately pronounced.

Readers with a taste for theoretical elegance may want to bear with me for one more paragraph. Note that there is an uncanny pattern in what the *d-to-t* rule is doing. First, *d* itself is voiced, and it ends up next to voiced consonants, whereas *t* is unvoiced, and it ends up next to unvoiced consonants. Second, except for voicing, t and *d* are the same; they use the same speech organ, the tongue tip, and that organ moves in the same way, namely sealing up the mouth at the gum ridge and then releasing. So the rule is not just tossing phonemes around arbitrarily, like changing a *p* to an *l* following a high vowel or any other substitution one might pick at random. It is doing delicate surgery on the *-ed* suffix, adjusting it to be the same in voicing as its neighbor, but leaving the rest of its features alone. That is, in converting *slap* + *ed* to *slapt,* the rule is "spreading" the voicing instruction, packaged with the *p* at the end of *slap,* onto the *-ed* suffix, like this:



The voicelessness of the *t* in *slapped* matches the voicelessness of the *p* in *slapped* because they are the *same* voicelessness; they are mentally represented as a single feature linked to two segments. This happens very often in the world's languages. Features like voicing, vowel quality, and tones can spread sideways or sprout connections to several phonemes in a word, as if each feature lived on its own horizontal "tier," rather than being tethered to one and only one phoneme.

So phonological rules "see" features, not phonemes, and they adjust features, not phonemes. Recall, too, that languages tend to arrive at an inventory of phonemes by multiplying out the various combinations of some set of features. These facts show that features, not phonemes, are the atoms of linguistic sound stored and manipulated in the brain. A phoneme is merely a bundle of features. Thus even in dealing with its smallest units, the features, language works by using a combinatorial system.

Every language has phonological rules, but what are they for? You may have noticed that they often make articulation easier. Flapping a *t* or a *d* between two vowels is faster than keeping the tongue in place long enough for air pressure to build up. Spreading voicelessness from the end of a word to its suffix spares the talker from having to turn the larynx off while pronouncing the end of the stem and then turn it back on again for the suffix. At first glance, phonological rules seem to be a mere summary of articulatory laziness. And from here it is a small step to notice phonological adjustments in some dialect other than one's own and conclude that they typify the slovenliness of the speakers. Neither side of the Atlantic is safe. George Bernard Shaw wrote:

> The English have no respect for their language and will not teach their children to speak it. They cannot spell it because they have nothing to spell it with but an old foreign alphabet of which only the consonants—and not all of them—have any agreed speech value. Consequently it is impossible for an Englishman to open his mouth without making some other Englishman despise him.

In his article "Howta Reckanize American Slurvian," Richard Lederer writes:

> Language lovers have long bewailed the sad state of pronunciation and articulation in the United States. Both in sorrow and in anger, speakers afflicted with sensitive ears wince at such mumblings as

guvmint for *government* and *assessories* for *accessories*. Indeed, everywhere we turn we are assaulted by a slew of slurrings.

But if their ears were even more sensitive, these sorrowful speakers might notice that in fact there is no dialect in which sloppiness prevails. Phonological rules give with one hand and take away with the other. The same bumpkins who are derided for dropping *g's* in *Nothin' doin'* are likely to enunciate the vowels in *pó-lice* and *accidént* that pointy-headed intellectuals reduce to a neutral "uh" sound. When the Brooklyn Dodgers pitcher Waite Hoyt was hit by a ball, a fan in the bleachers shouted, "Hurt's hoit!" Bostonians who pahk their cah in Hahvahd Yahd name their daughters Sheiler and Linder. In 1992 an ordinance was proposed that would have banned the hiring of any immigrant teacher who "speaks with an accent" in—I am not making this up—Westfield, Massachusetts. An incredulous woman wrote to the *Boston Globe* recalling how her native New England teacher defined "homonym" using the example *orphan* and *often*. Another amused reader remembered incurring the teacher's wrath when he spelled "cuh-rée-uh" *k-o-r-e-a* and "cuh-rée-ur" *c-a-r-e-e-r,* rather than vice versa. The proposal was quickly withdrawn.

There is a good reason why so-called laziness in pronunciation is in fact tightly regulated by phonological rules, and why, as a consequence, no dialect allows its speakers to cut corners at will. Every act of sloppiness on the part of a speaker demands a compensating measure of mental effort on the part of the conversational partner. A society of lazy talkers would be a society of hard-working listeners. If speakers were to have their way, all rules of phonology would spread and reduce and delete. But if listeners were to have their way, phonology would do the opposite: it would enhance the acoustic differences between confusable phonemes by forcing speakers to exaggerate or embroider them. And indeed, many rules of phonology do that. (For example, there is a rule that forces English speakers to round their lips while saying *sh* but not while saying *s*. The benefit of forcing everyone to make this extra gesture is that the long resonant chamber formed by the pursed lips enhances the lower-frequency noise that distinguishes *sh* from *s,* allowing for easier identification of the *sh* by the listener.) Although every speaker soon becomes a listener, human hypocrisy would make it unwise to depend on the speaker's foresight and consideration. Instead, a single, partly arbitrary set of phonological rules, some reducing, some enhancing, is adopted by every member of a linguistic community when he or she acquires the local dialect as a child.

Phonological rules help listeners even when they do not exaggerate some acoustic difference. By making speech patterns predictable, they add redundancy to a language; English text has been estimated as being between two and four times as long as it has to be for its information content. For example, this book takes up about 900,000 characters on my computer disk, but my file compression program can exploit the redundancy in the letter sequences and squeeze it into about 400,000 characters; computer files that do not contain English text cannot be squished nearly that much. The logician Quine explains why many systems have redundancy built in:

> It is the judicious excess over minimum requisite support. It is why a good bridge does not crumble when subjected to stress beyond what reasonably could have been foreseen. It is fallback and failsafe. It is why we address our mail to city and state in so many words, despite the zip code. One indistinct digit in the zip code would spoil everything. ... A kingdom, legend tells us, was lost for want of a horseshoe nail. Redundancy is our safeguard against such instability.

Thanks to the redundancy of language, yxx cxn xndxrstxnd whxt x xm wrxtxng xvxn xf x rxplxcx xll thx vxwxls wxth xn "x" (t gts lttl hrdr f y dn't vn kn whr th vwls r). In the comprehension of speech, the redundancy conferred by phonological rules can compensate for some of the ambiguity in the sound wave. For example, a listener can know that "thisrip" must be *this rip* and not *the srip* because the English consonant cluster *sr* is illegal.

So why is it that a nation that can put a man on the moon cannot build a computer that can take dictation? According to what I have explained so far, each phoneme should have a telltale acoustic signature: a set of resonances for vowels, a noise band for fricatives, a silence-burst-transition sequence for stops. The sequences of phonemes are massaged in predictable ways by ordered phonological rules, whose effects could presumably be undone by applying them in reverse.

The reason that speech recognition is so hard is that there's many a slip 'twixt brain and lip. No two people's voices are alike, either in the shape of the vocal tract that sculpts the sounds, or in the person's precise habits of articulation. Phonemes also sound very different depending on how much they are stressed and how quickly they are spoken; in rapid speech, many are swallowed outright.

But the main reason an electric stenographer is not just around the corner has to do with a general phenomenon in muscle control called coarticulation. Put a saucer in front of you and a coffee cup a foot or so away from it on one side. Now quickly touch the saucer and pick up the cup. You probably touched the saucer at the edge nearest the cup, not dead center. Your fingers probably assumed the handle-grasping posture while your hand was making its way to the cup, well before it arrived. This graceful smoothing and overlapping of gestures is ubiquitous in motor control. It reduces the forces necessary to move body parts around and lessens the wear and tear on the joints. The tongue and throat are no different. When we want to articulate a phoneme, our tongue cannot assume the target posture instantaneously; it is a heavy slab of meat that takes time to heft into place. So while we are moving it, our brains are anticipating the next posture in planning the trajectory, just like the cup-and-saucer maneuver. Among the range of positions in the mouth that can define a phoneme, we place the tongue in the one that offers the shortest path to the target for the next phoneme. If the current phoneme does not specify where a speech organ should be, we anticipate where the next phoneme wants it to be and put it there in advance. Most of us are completely unaware of these adjustments until they are called to our attention. Say *Cape Cod*. Until now you probably never noticed that your tongue body is in different positions for the two *k* sounds. In *horseshoe,* the first s becomes a *sh;* in *NPR,* the n becomes an *m;* in *month* and *width,* the *n* and *d* are articulated at the teeth, not the usual gum ridge.

Because sound waves are minutely sensitive to the shapes of the cavities they pass through, this coarticulation wreaks havoc with the speech sound. Each phoneme's sound signature is colored by the phonemes that come before and after, sometimes to the point of having nothing in common with its sound signature in the company of a different set of phonemes. That is why you cannot cut up a tape of the sound *cat* and hope to find a beginning piece that contains the *k* alone. As you make earlier and earlier cuts', the piece may go from sounding like *ka* to sounding like a chirp or whistle. This shingling of phonemes in the speech stream could, in principle, be a boon to an optimally designed speech recognizer. Consonant and vowels are being signaled simultaneously, greatly increasing the rate of phonemes per second, as I noted at the beginning of this chapter, and there are many redundant sound cues to a given phoneme. But this advantage can be enjoyed only by a high-tech speech recognizer, one that has some kind of knowledge of how vocal tracts blend sounds.

The human brain, of course, is a high-tech speech recognizer, but no one knows how it succeeds. For this reason psychologists who study speech perception and engineers who build speech recognition machines keep a close eye on each other's work. Speech recognition may be so hard that there are only a few ways it could be solved in principle. If so, the way the brain does it may offer hints as to the best way to build a machine to do it, and how a successful machine does it may suggest hypotheses about how the brain does it.

Early in the history of speech research, it became clear that human listeners might somehow take advantage of their expectations of the kinds of things a speaker is likely to say. This could narrow down the alternatives left open by the acoustic analysis of the speech signal. We have already noted that the rules of phonology provide one sort of redundancy that can be exploited, but people might go even farther. The psychologist George Miller played tapes of sentences in background noise and asked people to repeat back exactly what they heard. Some of the sentences followed the rules of English syntax and made sense:

> Furry wildcats fight furious battles.
> Respectable jewelers give accurate appraisals.
> Lighted cigarettes create smoky fumes.
> Gallant gentlemen save distressed damsels.
> Soapy detergents dissolve greasy stains.

Others were created by scrambling the words within phrases to create colorless-green-ideas sentences, grammatical but nonsensical:

> Furry jewelers create distressed stains.
> Respectable cigarettes save greasy battles.
> Lighted gentlemen dissolve furious appraisals.
> Gallant detergents fight accurate fumes.
> Soapy wildcats give smoky damsels.

A third kind was created by scrambling the phrase structure but keeping related words together, as in

> Furry fight furious wildcat battles.
> Jewelers respectable appraisals accurate give.

Finally, some sentences were utter word salad, like

> Furry create distressed jewelers stains.
> Cigarettes respectable battles greasy save.

People did best with the grammatical sensible sentences, worse with the grammatical nonsense and the ungrammatical sense, and worst of all with the ungrammatical nonsense. A few years later the psychologist Richard Warren taped sentences like *The state governors met with their respective legislatures convening in the capital city,* excised the first *s* from *legislatures,* and spliced in a cough. Listeners could not tell that any sound was missing.

If one thinks of the sound wave as sitting at the bottom of a hierarchy from sounds to phonemes to words to phrases to the meanings of sentences to general knowledge, these demonstrations seem to imply that human speech perception works from the top down rather than just from the bottom up. Maybe we are constantly guessing what a speaker will say next, using every scrap of conscious and unconscious knowledge at our disposal, from how coarticulation distorts sounds, to the rules of English phonology, to the rules of English syntax, to stereotypes about who tends to do what to whom in the world, to hunches about what our conversational partner has in mind at that very moment. If the expectations are accurate enough, the acoustic analysis can be fairly crude; what the sound wave lacks, the context can fill in. For example, if you are listening to a discussion about the destruction of ecological habitats, you might be on the lookout for words pertaining to threatened animals and plants, and then when you hear speech sounds whose phonemes you cannot pick out like "eesees," you would perceive it correctly as *species*—unless you are Emily Litella, the hearing-impaired editorialist on *Saturday Night Live* who argued passionately against the campaign to protect endangered feces. (Indeed, the humor in the Gilda Radner character, who also fulminated against saving Soviet jewelry, stopping violins in the streets, and preserving natural racehorses, comes not from her impairment at the bottom of the speech-processing system but from her ditziness at the top, the level that should have prevented her from arriving at her interpretations.)

The top-down theory of speech perception exerts a powerful emotional tug on some people. It confirms the relativist philosophy that we hear what we expect to hear, that our knowledge determines our perception, and ultimately that we are not in direct contact with any objective reality. In a sense, perception that is strongly driven from the top down would be a barely controlled hallucination, and that is the problem. A perceiver forced to rely on its expectations is at a severe disadvantage in a world that is unpredictable even under the best of circumstances. There is reason to believe that human speech perception is, in fact, driven quite strongly by acoustics. If you have an indulgent friend, you can try the following experiment. Pick ten words at random out of a dictionary, phone up the friend, and say the words clearly. Chances are the friend will reproduce them perfectly, relying only on the information in the sound wave and knowledge of English vocabulary and phonology. The friend could not have been using any higher-level expectations about phrase structure, context, or story line because a list of words blurted out of the blue has none. Though we may call upon high-level conceptual knowledge in noisy or degraded circumstances (and even here it is not clear whether the knowledge alters perception or just allows us to guess intelligently after the fact), our brains seem designed to squeeze every last drop of phonetic information out of the sound wave itself. Our sixth sense may perceive speech as language, not as sound, but it *is* a sense, something that connects us to the world, and not just a form of suggestibility.

Another demonstration that speech perception is not the same thing as fleshing out expectations comes from an illusion that the columnist Jon Carroll has called the mondegreen, after his mis-hearing of the folk ballad "The Bonnie Earl O'Moray":

> Oh, ye hielands and ye lowlands,
> Oh, where hae ye been?
> They have slain the Earl of Moray,
> And laid him on the green.

He had always thought that the lines were "They have slain the Earl of Moray, And Lady Mondegreen." Mondegreens are fairly common (they are an extreme version of the Pullet Surprises and Pencil Vaneas mentioned earlier); here are some examples:

> A girl with colitis goes by. [A girl with kaleidoscope eyes. From the Beatles song "Lucy in the Sky with Diamonds."]
>
> Our father wishart in heaven; Harold be they name . . . Lead us not into Penn Station. [Our father which art in Heaven; hallowed be thy name . . . Lead us not into temptation. From the Lord's Prayer.]
>
> He is trampling out the vintage where the grapes are wrapped and stored. [. . . grapes of wrath are stored. From "The Battle Hymn of the Republic."]
>
> Gladly the cross-eyed bear. [Gladly the cross I'd bear.]
>
> I'll never be your pizza burnin'. [. . . your beast of burden. From the Rolling Stones song.]
>
> It's a happy enchilada, and you think you're gonna drown. [It's a half an inch of water . . . From the John Prine song "That's the Way the World Goes 'Round."]

The interesting thing about mondegreens is that the mis-hearings are generally *less* plausible than the intended lyrics. In no way do they bear out any sane listener's general expectations of what a speaker is likely to say or mean. (In one case a student stubbornly mis-heard the Shocking Blue hit song "I'm your Venus" as "I'm Your Penis" and wondered how it was allowed on the radio.) The mondegreens do conform to English phonology, English syntax (sometimes), and English vocabulary (though not always, as in the word *mondegreen* itself). Apparently, listeners lock in to some set of words that fit the sound and that hang together more or less as English words and phrases, but plausibility and general expectations are not running the show.

The history of artificial speech recognizers offers a similar moral. In the 1970s a team of artificial intelligence researchers at Carnegie-Mellon University headed by Raj Reddy designed a computer program called HEARSAY that interpreted spoken commands to move chess pieces. Influenced by the top-down theory of speech perception, they designed the program as a "community" of "expert" subprograms cooperating to give the most likely interpretation of the signal. There were subprograms that specialized in acoustic analysis, in phonology, in the dictionary, in syntax, in rules for the legal moves of chess, even in chess strategy as applied to the game in progress. According to one story, a general from the defense agency that was funding the research came up for a demonstration. As the scientists sweated he was seated in front of a chessboard and a microphone hooked up to the computer. The general cleared his throat. The program printed "Pawn to King 4."

The recent program DragonDictate, mentioned earlier in the chapter, places the burden more on good acoustic, phonological, and lexical analyses, and that seems to be responsible for its greater success. The program has a dictionary of words and their sequences of phonemes. To help anticipate the effects of phonological rules and coarticulation, the program is told what every English phoneme sounds like in the context of every possible preceding phoneme and every possible following phoneme. For each word, these phonemes-in-context are arranged into a little chain, with a probability attached to each transition from one sound unit to the next. This chain serves as a crude model of the speaker, and when a real speaker uses the system, the probabilities in the chain are adjusted to capture that person's manner of speaking. The entire word, too, has a probability attached to it, which depends on its frequency in the language and on the speaker's habits. In some versions of the program, the probability value for a word is adjusted depending on which word precedes it; this is the only top-down information that the program uses. All this knowledge allows the program to calculate which word is most likely to have come out of the mouth of the speaker given the input sound. Even then, DragonDictate relies more on expectancies than an able-eared human does. In the demonstration I saw, the program had to be coaxed into recognizing *word* and *worm,* even when they were pronounced as clear as a bell, because it kept playing the odds and guessing higher-frequency *were* instead.

Now that you know how individual speech units are produced, how they are represented in the mental dictionary, and how they are rearranged and smeared before they emerge from the mouth, you have reached the prize at the bottom of this chapter: why English spelling is not as deranged as it first appears.

The complaint about English spelling, of course, is that it pretends to capture the sounds of words but does not. There is a long tradition of doggerel making this point, of which this stanza is a typical example:

> Beware of heard, a dreadful word
> That looks like beard and sounds like bird,
> And dead: it's said like bed, not bead—
> For goodness' sake don't call it "deed"!
> Watch out for meat and great and threat
> (They rhyme with suite and straight and debt).

George Bernard Shaw led a vigorous campaign to reform the English alphabet, a system so illogical, he said, that it could spell *fish* as "ghoti"—*gh* as in *tough, o* as in *women, ti* as in *nation.* ("Mnomnoupte" for *minute* and "mnopspteiche" for *mistake* are other examples.) In his will Shaw bequeathed a cash prize to be awarded to the designer of a replacement alphabet for English, in which each sound in the spoken language would be recognizable by a single symbol. He wrote:

> To realize the annual difference in favour of a forty-two letter phonetic alphabet . . . you must multiply the number of minutes in the year, the number of people in the world who are continuously writing English words, casting types, manufacturing printing and writing machines, by which time the total figure will have become so astronomical that you will realize that the cost of spelling even one sound with two letters has cost us centuries of unnecessary labour. A new British 42 letter alphabet would pay for itself a million times over not only in hours but in moments. When this is

> grasped, all the useless twaddle about enough and cough and laugh and simplified spelling will be dropped, and the economists and statisticians will be set to work to gather in the orthographic Golconda.

My defense of English spelling will be halfhearted. For although language is an instinct, written language is not. Writing was invented a small number of times in history, and alphabetic writing, where one character corresponds to one sound, seems to have been invented only once. Most societies have lacked written language, and those that have it inherited it or borrowed it from one of the inventors. Children must be taught to read and write in laborious lessons, and knowledge of spelling involves no daring leaps from the training examples like the leaps we saw in Simon, Mayela, and the Jabba and *mice-eater* experiments in Chapters 3 and 5. And people do not uniformly succeed. Illiteracy, the result of insufficient teaching, is the rule in much of the world, and dyslexia, a presumed congenital difficulty in learning to read even with sufficient teaching, is a severe problem even in industrial societies, found in five to ten percent of the population.

But though writing is an artificial contraption connecting vision and language, it must tap into the language system at well-demarcated points, and that gives it a modicum of logic. In all known writing systems, the symbols designate only three kinds of linguistic structure: the morpheme, the syllable, and the phoneme. Mesopotamian cuneiform, Egyptian hieroglyphs, Chinese logograms, and Japanese kanji encode morphemes. Cherokee, Ancient Cypriot, and Japanese kana are syllable-based. All modern phonemic alphabets appear to be descended from a system invented by the Canaanites around 1700 B.C. No writing system has symbols for actual sound units that can be identified on an oscilloscope or spectrogram, such as a phoneme as it is pronounced in a particular context or a syllable chopped in half.

Why has no writing system ever met Shaw's ideal of one symbol per sound? As Shaw himself said elsewhere, "There are two tragedies in life. One is not to get your heart's desire. The other is to get it." Just think back to the workings of phonology and coarticulation. A true Shavian alphabet would mandate different vowels in *write* and *ride,* different consonants in *write* and *writing,* and different spellings for the past-tense suffix in *slapped, sobbed,* and *sorted*. *Cape Cod* would lose its visual alliteration. A *horse* would be spelled differently from its *horseshoe,* and National Public Radio would have the enigmatic abbreviation *MPR*. We would need brand-new letters for the *n* in *month* and the *d* in *width*. I would spell *often* differently from *orphan,* but my neighbors here in the Hub would not, and their spelling of *career* would be my spelling of *Korea* and vice versa.

Obviously, alphabets do not and should not correspond to sounds; at best they correspond to the phonemes specified in the mental dictionary. The actual sounds are different in different contexts, so true phonetic spelling would only obscure their underlying identity. The surface sounds are predictable by phonological rules, though, so there is no need to clutter up the page with symbols for the actual sounds; the reader needs only the abstract blueprint for a word and can flesh out the sound if needed. Indeed, for about eighty-four percent of English words, spelling is completely predictable from regular rules. Moreover, since dialects separated by time and space often differ most in the phonological rules that convert mental dictionary entries into pronunciations, a spelling corresponding to the underlying entries, not the sounds, can be widely shared. The words with truly weird spellings (like *of, people, women, have, said, do, done,* and *give)* generally are the commonest ones in the language, so there is ample opportunity for everyone to memorize them.

Even the less predictable aspects of spelling bespeak hidden linguistic regularities. Consider the following pairs of words where the same letters get different pronunciations:

electric-electricity
photograph-photography
grade-gradual
history-historical
revise-revision
adore-adoration
bomb-bombard
nation-national
critical-criticize
mode-modular
resident-residential
declare-declaration
muscle-muscular
condemn-condemnation
courage-courageous
romantic-romanticize
industry-industrial
fact-factual
inspire-inspiration
sign-signature
malign-malignant

Once again the similar spellings, despite differences in pronunciation, are there for a reason: they are identifying two words as being based on the same root morpheme. This shows that English spelling is not completely phonemic; sometimes letters encode phonemes, but sometimes a sequence of letters is specific to a morpheme. And a morphemic writing system is more useful than you might think. The goal of reading, after all, is to understand the text, not to pronounce it. A morphemic spelling can help a reader distinguishing homophones, like *meet* and *mete*. It can also tip off a reader that one word contains another (and not just a phonologically identical impostor). For example, spelling tells us that *overcome* contains *come,* so we know that its past tense must be *overcame,* whereas *succumb* just contains the sound "kum," not the morpheme *come,* so its past tense is not *succame* but *succumbed.* Similarly, when something *recedes,* one has a *recession,* but when someone *re-seeds* a lawn, we have a *re-seeding.*

In some ways, a morphemic writing system has served the Chinese well, despite the inherent disadvantage that readers are at a loss when they face a new or rare word. Mutually unintelligible dialects can share texts (even if their speakers pronounce the words very differently), and many documents that are thousands of years old are readable by modern speakers. Mark Twain alluded to such inertia in our own Roman writing system when he wrote, "They spell it Vinci and pronounce it Vinchy; foreigners always spell better than they pronounce."

Of course English spelling could be better than it is. But it is already much better than people think it is. That is because writing systems do not aim to represent the actual sounds of talking, which we do not hear, but the abstract units of language underlying them, which we do hear.