The Scientific Study of Language

The scientific study of language is not a familiar undertaking. Ordinarily when you study a language, you are trying to learn how to speak and understand it. So an important question arises right at the start. Why would anyone want to study language scientifically? Why not simply *use* language and get on with the business of living?

Before trying to answer that question, a distinction must be drawn: namely, the distinction between specific, individual languages and language in general. Language in general is a generic and highly abstract idea; specific languages are instances of language in general, just as specific breeds of dogs are instances of dog in general. A specific language is a particular set of social conventions governing the formation of grammatical utterances and their use in achieving personal goals. People learn particular languages—English, Urdu, Swahili, Japanese—and use them to interact with others in their society. The scientific study of language is the study of language in general, the study of what is common to all languages.

So why would anyone care about language in general? Nobody speaks language in general. It is useless for communication; you cannot order a pizza in language in general. Yet a moment's reflection establishes the central role that language in general plays in all human affairs. People want to know about language in general, not to use it in social interaction, but in the hope of understanding something that is uniquely human about human beings.

There are good reasons to study language scientifically. Language in general is important not only because it distinguishes human beings from all other animals on the earth, but because, directly or indirectly, it makes possible the elaborate organization of civilized society. Because the human capacity for acquiring language is innate—every human group has one—when you investigate language, you are investigating something universal, with a solid base in the biological nature of *Homo sapiens sapiens*. And language in general is interesting because, although everyone knows and uses a specific language, few people understand what they know. Becoming self-consciously aware of what is known unself-consciously carries a special brand of excitement.

A skeptic might resist such claims, of course. "It is not speech," the skeptic might say, "but human intelligence—the amazing human ability to learn and to adapt—that is so extraordinary and unique." And to support that objection the skeptic could point to people who are born deaf and never acquire spoken language, yet communicate splendidly by means of hand gestures. They achieve the same results, and do so almost as expeditiously as their noisier cousins. But what this argument shows is merely that speech and language are different. Systems of hand gestures that are called sign languages—American Sign Language, Danish Sign Language, British Sign Language—are in every sense as much specific languages as are any systems of spoken words. It is true that congenitally deaf people can do quite well without speech, but they cannot get along without language.

One need not be a communications expert to recognize that different signals can carry the same message, so it is tempting to conclude that it is the message, not the signal, that is important. But that overlooks the fact that without a signal there could be no message. The important point is that different kinds of signals—spoken, gestured, written, whatever—can make possible the externalization of thought and perception. A gorilla might be enormously intelligent, but without some system of signals to communicate that intelligence—without language—its remarkable gift would remain unshared. Intelligence alone does not explain the human ability to communicate.

A system of externalizing thought that uses the voice and ear has many advantages over a system that uses the hand and eye. Speech is the biologically given signal system for human communication, and much will be made of spoken words in the pages that follow. But when the voice and ear are not available, the human need to communicate finds other signals to replace them. A theory of language in general should have room for this diversity of signals.

The evolution of language enabled many individuals to think together. The externalization achieved by language is not perfect, of course, but it is

good enough that many people can share, enjoy, and profit from one person's experience. Thus, social units could form and work together in novel ways, cooperating as if they were a single superordinate individual. The survival value of human language was very great.

If language is important, then it should follow that words, the building blocks of language, are also important. Indeed, the very fact that something is a word means that it must have been important to somebody. Or, to put it differently, a word is an idea that some group of people thought was impor-

The Eskimo Vocabulary for Snow

When an idea is important, people are likely to have a word for it. Mountain people will have a word for mountain; people who live on the plains and have never seen a mountain will not have such a word. The more important something is, moreover, the more words there are likely to be. If, for example, a language has many words for different kinds and states of bamboo, an anthropologist knows that bamboo plays a central role in the lives of that people. But it is not necessary to study other cultures to find lexical specialization. Painters have many words for colors, chemists have many words for chemical compounds, horsemen recognize many different kinds of horses. Every professional group develops its own technical jargon for talking about matters of critical concern.

It is remarkable, therefore, that one particular example of lexical specialization has captured the popular imagination and been referred to so often in the press—namely, the Eskimo vocabulary for snow. It is plausible to suppose that snow plays an important role in the lives of the Eskimo, so their language should have several words for it. The interesting question is: Precisely how many different snow words do they have? According to the myth that has grown up, the Eskimo language has hundreds of words for different kinds and grades of snow, an extravagant specialization that is sometimes cited to illustrate how primitive minds categorize reality differently.

Anthropologist Franz Boas (1858–1942).

The anthropologist Laura Martin traced this myth back to a passage in Franz Boas's *Handbook of North American Indians* (1911), where he comments that Eskimo has apparently distinct words for snow: *aput* for snow on the ground, *qana* for falling snow, *piqsirpoq* for drifting snow, and *qimuqsuq* for a snow drift. The number began to grow in 1940 when Benjamin Lee Whorf published a popular article claiming that Eskimo has distinct words for falling snow, snow on the ground, packed snow, slushy snow, wind-driven snow, and other kinds of snow. As interest in the matter spread, the published claims grew vaguer: "Eskimo languages have many words for snow." Thereafter "many" was translated into nine, forty-eight, one hundred, two hundred.

The linguist Geoffrey Pullum advises his readers to fight this Eskimological falsehood. When you hear the claim, he advises, you should stand up and announce that the best dictionary of the Eskimo language gives just two roots: *qanik* for snow in the air and *aput* for snow on the ground. It will not make you the most popular person in the room, but it will strike a blow for truth and the standards of evidence.

Snow in the air, snow on the ground.

tant enough to enter in the lexicon. (This notion of what a word is does not hold up under careful scrutiny, as Chapter 2 takes pains to point out, but it is more vague than wrong.) People who know a word can share that idea with other members of their group, and a shared vocabulary is part of the glue that holds people together and allows them to create a shared culture.

An inescapable fact that will surely impress anyone who carefully considers the operations of the human mind is how much people know. Some people know more than others, of course, but everybody knows a lot. And one thing that everybody knows is a language, which is itself a very large chunk of knowledge. The major part of that large chunk of knowledge consists of knowing the words of the language. It is not the speech sounds or the rules for generating grammatical sentences that require the most extensive learning. It is the vocabulary: thousands of words, each with its own sound, its own spelling, its own meaning, its own role, its own use, its own history.

An intriguing thing about this knowledge is that people know so much that they do not realize they know. In English, for example, Subject-Verb, or Subject-Verb-Object, is a general pattern for grammatical sentences. Subject and Object phrases contain nouns, Verb phrases contain verbs. To conform to this pattern, therefore, people who speak English must know the difference between nouns and verbs. Why, then, is it so difficult to teach this distinction to schoolchildren, who conform to it consistently in their everyday use of language? Apparently there are different ways of knowing. Knowing how to speak grammatically and knowing how to characterize grammatical speech are not the same.

Just as none of the reasons for studying language scientifically depend on any particular language, so none of the reasons for studying words depend on any particular words. What is at issue in a scientific discussion of words is not so much specific words as wordiness: Why are all languages wordy? Why are words a universal design feature of languages? It is words in general, not particular words, that are scientifically important. Humanistic scholars may be interested in individual words in particular documents written in specific languages, but a scientist must search for generalizations and invariants.

A Little History

Curiosity about the origins of language, and about why there are so many different languages, has a long history. The whole idea of studying language in general scientifically grew out of studying specific languages historically.

The desire to learn and understand specific languages is as old as recorded history. In classical Greece every educated man studied grammar and rhetoric, and the scholarly pursuit of those subjects has persisted down to the present day. But correct usage, persuasive oratory, and the accurate preservation of important texts are not the stuff from which the science of linguistics was

built. Foundations for linguistic science had to await comparative studies of different languages.

The scientific study of language began just two hundred years ago with systematic attempts to trace historical relations among languages. Medieval scholars studied Latin, of course, and with the Renaissance came knowledge of ancient Greek. But it was not until the eighteenth century that the rediscovery of the ancient Sanskrit language of India made the comparison of languages an active field of study. Sir William Jones, who learned Sanskrit as Chief Justice of Bengal, wrote that no one who was familiar with Sanskrit, Greek, and Latin "could examine all three without believing them to have sprung from some common source, which, perhaps, no longer exists." Sir William's observation, made in 1786, stirred many imaginations and identified an ideal starting point for comparing languages and tracing their historical changes.

Two questions became salient. First, what was the lost "common source" from which Sanskrit, Greek, and Latin had sprung? Second, what other languages had sprung from the same source?

The initial demonstration that Sanskrit, Greek, and Latin are related was easy. But reconstructing their common source was not easy. Many scholars had to follow Jones's lead before the details became clear. The most important approach is to compare the attested languages, looking for shared features. This comparative method is based on the assumption that if two languages share a special feature, they probably inherited it from a common ancestor. Comparative reconstruction always starts with vocabulary, where resemblances are particularly striking. Lists of similarities were drawn up, and from them the phonology and vocabulary of the ancestral language were inferred. For example, the words for numbers from one to ten show how similar the three languages are, and how different they are from Japanese, which is not related.

At first it was thought that Greek and Latin were descended from Sanskrit, but after extensive discussion and debate over the accumulating evidence it was finally agreed that all three are descended from some mother language, called Proto-Indo-European (PIE), a language spoken before the invention of writing and now lost forever.

How do linguists decide such things? It is generally accepted that languages change, of course: Old English changed into Middle English and then into Modern English; Latin changed into Italian, French, Spanish. Those changes are well documented in the written record. But what do linguists do when the evidence is less compelling?

Suppose, for example, that someone wanted to say that Italian is not really the sister of Spanish, but its mother. The challenge would be to prove that all the changes that were introduced in going from Latin to Italian are also present in Spanish, and also to find other innovations in Spanish that have not

Numbers from One to Ten in Five Languages

English	Latin	Greek	Sanskrit	Japanese
one	unus	heis	ekas	hitotsu
two	duo	duo	dva	futatsu
three	tres	treis	tryas	mittsu
four	quattuor	tettares	catvaras	yottsu
five	quinque	pente	panca	itsutsu
six	sex	heks	sat	muttsu
seven	septem	hepta	sapta	nanatsu
eight	octo	okto	asta	yattsu
nine	novem	ennea	nava	kokonotsu
ten	decem	deka	dasa	to

occurred in Italian. That case cannot be made, of course, but if it could, linguists would declare that an early form of Italian was an older stage of modern Spanish. The point of this counterfactual example, however, is that all such judgments are comparative, not absolute. There are no absolute linguistic properties that are primitive, primordial, or imperfect and that will therefore mark one language as older than another. A line of linguistic descent is not some simple "given." It must be inferred indirectly as a special kind of relatedness.

By comparative arguments, therefore, scholars decided that Greek, Latin, and Sanskrit were sister languages descended from a common ancestral language. The more languages there were to be compared, of course, the more reliable such inferences were judged to be. So the first question led naturally to the second: What other languages were descended from PIE? The same kinds of comparisons that were used to reconstruct PIE from languages now dead were also used to establish common ancestors of living languages. The large family of modern languages that descended from this lost PIE—including Kurdish, Persian, Urdu, Hindi, modern Greek, French, Spanish, Italian, Portuguese, German, Dutch, Norwegian, Swedish, Russian, Ukrainian, Bulgarian, and many others in addition to English—are called Indo-European languages. Tracing out relations among these different and widely distributed languages became an exciting intellectual adventure, and the implications for European prehistory are still debated by archeologists.

One of the first fruits of this budding science of language were the *sound laws*, which resulted from attempts to establish that modern Germanic languages (including English) belong to the Indo-European family. Proto-Germanic, an early form of German that might have split off from PIE sometime prior to 1000 B.C., can be reconstructed from early records of Gothic, Old English, Old High German, and Old Norse, written between A.D. 200 and 1200. Scholars such as Rasmus Rask noted that many words in these early Germanic languages bear a systematic relation to words in Latin if you assume that the voiceless stops in Latin [p, t, k] became voiceless fricatives in Germanic [f, θ, h], as in the initial sounds of these:

Latin *pater* → Old Norse *fathir* (with voiced *th*, ð)
Latin *tres* → Old Norse *thrir* (with voiceless *th*, θ)
Latin *cornu* → Old High German *horn*

Rask's observations were summarized by Jakob Grimm in 1822 in what is usually called, somewhat unfairly, Grimm's Law. It summarized these systematic sound shifts and established that the Germanic languages do belong in the Indo-European family. However, the changes were so drastic and the number of important German words that have no known Indo-European source was so large that something more than ordinary linguistic evolution must have been going on. Some scholars have speculated that Proto-Germanic

developed when some people who spoke a non-Indo-European language rich in voiceless fricatives came in contact with speakers of PIE.

It was immediately recognized that there were exceptions to Grimm's Law. For example, according to the Law, the PIE voiceless stop *t* in *pater* should have become the voiceless fricative *th*-sound (*θ*, as in English *thin*), but instead became the voiced *th*-sound (ð, as in Modern English *father*). The *t* in Latin *frater*, on the other hand, was replaced by the voiceless *θ* in the Gothic *brothar*, as Grimm's Law specified. In view of such contradictions, Grimm's Law was assumed to describe merely a general tendency.

In 1875, however, many of the exceptions were explained by Verner's Law, which said that if the PIE voiceless stop is not initial, or is not immediately followed by a stressed vowel, then it becomes a voiced fricative in the Germanic languages. In other words, the Danish linguist Karl Verner (1846–1896) discovered that a whole series of noninitial consonant shifts in Proto-Germanic depended on which syllable was accented. When it was realized that German has *t* in *vater*, but *d* in *bruder*, because the two words had a different accentuation three or four thousand years ago, it served not only to heighten respect for the linguistic science that was able to demonstrate such truths, but also to increase the feeling that the world of spoken sounds is subject to laws as strict as those of natural science.

This strategy—to account for sound changes as dependent on the contexts in which the sounds had originally occurred—was so successful that one enthusiastic group, the Neogrammarians (nicknamed the "young grammarians"), claimed that linguistics had become a precise science and that there

Modern English evolved from early Germanic languages. Here for comparison are three versions of Luke 2 : 8–14. The Gothic was translated from Greek, the Old English from Latin. The Middle English version is from the Wyclif Bible, which was the first complete translation of the Latin Vulgate into English.

Gothic, c. 350

8 Jah hairdjos wesun in pamma samin landa, pairhwakandans jah witandans wahtwom nahts ufaro hairdai seinai.
9 Ip aggilus fraujins anaqam ins, jah wulpus fraujins biskain ins, jah ohtedun agisa mikilamma.
10 Jah qap du im sa aggilus: Ni ogeip; unte sai, spillo izwis faheid mikila, sei wairpip allai managein,
11 Patei gabaurans ist izwis himma daga nasjands, saei ist Xristus frauja, in baurg Daweidis.
12 Jah pata izwis taikns; bigitid barn biwundan, jah galagid in uzetin.
13 Jah anaks warp mip pamma aggilau managei harjis himinakundis, hazjandane gup, jah qipandane:
14 Wulpus in hauhistjam gupa, jah ana airpai gawairpi in mannam godis wiljins.

For *q* read "*qu*".

Old English (West Saxon Dialect), End of Tenth Century

8 And heirdas wǣron on pām ilcan rīce, waciende and nihtwæccan healdende ofer heora heorda.
9 Pā stōd dryhtnes engel wip hī, and godes beorhtnes him ymbescān; and hī him micelum ege ādrēdon.
10 And sē engel him tōcwæp: Nelle gē ēow ādrǣdan, sōplice nū, ic ēow bodie micelne gefēan, sē bip eallum folce,
11 forpām tōdæg ēow is hælend ācenned, sē is dryhten Christ, on Davides ceastre.
12 And pis tācen ēow bip: gē gemētap ān cild hrœglum bewunden, and on binne ālēd.
13 And pā wæs færinga geworden mid pām engle micelnes heofonlices werodes, god heriendra and pus cwependra:
14 Gode sī wuldor on hēahnesse, and on eorpan sibb mannum gōdes willan.

Spelling regularized and marks of length added.

Middle English, Wyclif, 1389

8 And schepherdis weren in the same cuntre, wakinge and kepinge the watchis of the nyzt on her flok.
9 And loo, the aungel of the Lord stood by sydis hem, and the clerenesse of God schynede aboute hem; and thei dredden with greet drede.
10 And the aungel seide to hem: Nyle ze drede; lo, sothli I euangelise to zou a grete ioye, that schal be to al peple.
11 For a sauyour is borun to day to vs, that is Crist the Lord, in the cite of Dauith.
12 And this a tokene to zou; ze schulen fynde a zong child wlappid in clothis, and put in a cracche.
13 And sudenly ther is maad with the aungel a multitude of heuenly knyzthod, heriynge God, and seyinge,
14 Glorie be in the hizeste thingis to God, and in erthe pees be to men of good wille.

could be NO exceptions to the sound laws. They did not think of sound changes as more or less haphazard events affecting some words but not others. Instead, they believed that a sound change was simply a change in the way speakers produced the speech sound (or a series of speech sounds), and so would affect that sound (or series of sounds) anywhere it happened to occur. When some apparent exception was noted, it was assumed that the relevant contexts had not been properly analyzed; a correct analysis would show it to be regular after all. This attitude led to acrimonious debate over endless details of the pronunciation of different languages, most of which were no longer spoken by any living people.

In 1878 the Swiss linguist Ferdinand de Saussure, then only twenty-two years old, published a *Memoir* on the PIE vowel system in which he tried to explain the irregular behavior of certain sounds in the daughter languages. After a highly technical analysis, Saussure proposed that PIE must have had another speech sound, whose pronunciation he could not determine by formal analysis. This hypothetical speech sound had been lost in the daughter languages, but not before leaving traces on the sounds that had preceded or followed it. Saussure's hypothesis, which came to be known as the "laryngeal theory" because it was thought that the lost sound might have had a laryngeal pronunciation, solved a number of problems in the development of various Indo-European languages. It remained purely hypothetical, however, for almost fifty years, until cuneiform Hittite had been discovered and deciphered. In 1927 it was demonstrated that ancient Hittite still had laryngeal consonants, written *h* or *hh*, in just those places where Saussure had claimed that the lost sound must have been in PIE. By purely formal analysis, Saussure had discovered the laryngeal consonants of PIE—and the validity of Neogrammarian principles had been upheld.

By necessity, these linguists dealt with written texts. As the new science emerged, however, the sound laws focused attention on pronunciation and on the need for more precise descriptions of speech sounds. Toward the end of the nineteenth century, the foundations were laid for phonology, the description of systems of speech sounds used in different languages. It was probably inevitable that attempts would be made to apply this tool for describing spoken languages to the many languages of the world that had no accumulated literatures because they had never been written down. Thus a bond was forged between the young science of linguistics and the young science of anthropology. An anthropologist in the field must know how to deal with the language of whatever exotic people he or she wishes to study. Training in linguistics became part of the technical education of every professional anthropologist.

As a consequence of this extension of linguistic methods to the study of languages that had no literary heritage, the original definition of linguistic science as the study of historical change in language had to be amended. The exotic languages that interested anthropologists obviously had histories, but no one would ever know what they were. Even without knowing its history, however, there is much of interest that can be said about a language. So two ways of studying language gradually sorted themselves out. The traditional study of linguistic change came to be called historical linguistics. The newer approach—which took a cross-sectional slice of a language, treating each language as a complex symbolic system existing within a limited period of time—came to be known simply as linguistics.

It was Ferdinand de Saussure who first insisted on the importance of this distinction between two different approaches to the study of language. To Saussure, historical or diachronic facts about language must be derived from nonhistorical or synchronic facts, or from a long succession of synchronic facts. As he defined the terms, synchronic linguistics is concerned with the

logical and psychological relations that organize coexisting language elements as a system in the minds of language users, whereas diachronic linguistics is concerned with relations that organize successive language elements, but are not present in the minds of any language users and so do not form a system. By giving priority to the synchronic approach, Saussure defined the central task of linguistic science.

By the 1930s synchronic linguistics had established itself as a respectable scientific discipline, related to but independent of diachronic linguistics and the humanistic study of written languages and their literatures. Synchronic descriptions of a language treated three major domains: pronunciation, grammar, and vocabulary. The new science of phonology handled pronunciation; it provided systems of phonetic writing in which to transcribe and analyze spoken utterances. Long utterances transcribed in this way were analyzed into constituent parts—words, phrases, clauses—and a theory of syntax was used to write rules for forming grammatical phrases and sentences. Finally, the accumulation of an alphabetical list of words and their meanings provided information needed to reveal morphology, the rules for forming words.

In this way, linguistic anthropologists were able to record and preserve many languages that had never been written down before. They could learn the language, could reduce it to writing, and could return from the field with a large collection of recorded utterances—a corpus—to analyze at their leisure. (Today, of course, tape recorders relieve them of the tedium of phonetic transcription in the field.) By limiting their generalizations to the actual corpus of utterances that they had recorded, these linguistic anthropologists always had solid empirical evidence to fall back on.

As part of their basic training, therefore, linguistic anthropologists mastered a synchronic theory of language in general, but primarily as an aid in learning specific languages. Linguistic theory served largely as a guide for drawing generalizations about some particular language on the basis of finite samples of recorded speech in that language.

Not until the 1950s did a linguist present a persuasive alternative to this anthropological approach. Then Noam Chomsky put forward the argument that language, properly conceived, is not a collection of texts that someone has written or a corpus of utterances that someone has transcribed. A language is something that people know, something that children learn and adults use. Any particular corpus can contain but a small sample of the infinite variety of sentences that a speaker of the language could produce and understand. In short, Chomsky redefined the subject matter of linguistics. No longer would synchronic linguistics be limited to the study of recorded instances. The subject matter of linguistics for Chomsky was the competence of language users, not their performance. Performances are merely the evidence from which their shared competence can be inferred.

Chomsky placed grammar at the center of his new formulation and named his new approach to it "generative grammar." A generative grammar consists of explicit rules that assign structural descriptions to sentences. An

ideal generative grammar would describe all and only the grammatical sentences of a particular language, and so could be considered a (highly abstract) description of what a person must know in order to speak and understand that language.

Describing abilities is a responsibility of psychology, so Chomsky's redefinition effectively made linguistics a branch of cognitive psychology. In recent years, linguistics and psychology have forged a bond as strong and as valuable to both as was the earlier bond between linguistics and anthropology. This book is a product of that new conception of linguistic science.

Psycholinguistics

To psychologists, these new ideas became known as psycholinguistics. Even before experimental psychology emerged from philosophy late in the nineteenth century, the central role of language in the minds of human beings was generally acknowledged. But few practitioners of the young science of psychology knew enough linguistics to design and execute studies that could reveal the role of language in controlling people's attention, memory, imagery, thought, or behavior.

The earliest attempts by psychologists to incorporate language into their experiments concentrated heavily on words. Human learning was studied by asking people to memorize lists of words, or to remember pairs of words; vocabulary size was found to be a good indicator of mental age; human intelligence was measured with vocabulary tests, and every high school graduate was expected to know the most frequently used words; the acuity of hearing and the effects of acoustic interference were assessed by asking people to write down the words they could hear; aberrant human emotions were diagnosed with word association tests; reading ability was calibrated in terms of words per minute. In these and other ways psychologists demonstrated their appreciation of the importance of language, even though the only unit of language they found comfortable to work with was the word.

Psycholinguistics, dedicated to understanding both the psychological foundations of language and the linguistic foundations of psychology, was initially populated by psychologists and anthropologists who had discovered their shared interest in synchronic descriptions of language. Psychologists brought to this new science their experimental techniques and their extensive studies of words; anthropologists brought to it their comparative methods and their broader perspective on what a language can be and do. Together they quickly discovered the great psychological complexity of human languages, and many psychologists who had initially assumed that speech must be some kind of conditioned reflex were forced to revise their opinions. But what to substitute for existing theories of conditioning and learning was far from clear. Psycholinguists were ready for the kind of theory that Chomsky soon provided.

Chomsky developed the conception of generative grammar to describe people's linguistic competence, and it held obvious implications for psychology. How is it possible, he asked, that people can know as much as they do when their contacts with the world are so personal and limited? In particular, how can children learn a language so readily when their exposure to its subtleties is so brief and impoverished? Chomsky's answer was that the capacity for language is part of the genetic endowment of all human beings, and that this innate competence can be characterized in terms of explicit principles governing the kinds of linguistic constructions that are possible—that is to say, in terms of generative grammar. His generative theory not only reformulated what psycholinguistic experiments should be investigating, but it suggested

mechanisms of language comprehension and language acquisition that were novel and exciting.

One effect of Chomsky's ideas was to redirect psycholinguistic research toward grammar and away from vocabulary. The claim was accepted that the indefinite variety of grammatical sentences a language user is competent to utter and understand can only be described (and, presumably, can only be learned) in terms of generative rules. What those rules are, how children acquire them, how those rules organize adult language, what happens when brain injuries limit them—these were new and challenging questions that suddenly seemed open to investigation. Words, by comparison, were uninteresting. It was assumed that the number of words must be limited, that words are subject to too many exceptions to support any interesting system of rules, and that there is nothing a child can do but memorize them. Psycholinguistic research on words almost disappeared.

Excessive interest in words was followed by excessive neglect, so it was probably inevitable that these basic building blocks of language would again come into favor. Psycholinguists began to recognize that words are linguistic universals just as surely as sentences are, that their number is not limited in most languages, and that learning them is anything but a tedious exercise in rote memorization. Even generative grammarians revived their interest in words as they came to realize that many of the syntactic rules they studied could just as well be stated as features of words and so regarded as lexical knowledge.

For example, a generative grammar for English must have a rule stating:

R1. Sentences have a noun phrase as the subject followed by a verb phrase as the predicate.

That rule is needed to account for such sentences as *The woman wept*. Note that *The gun wept* also follows this rule, but it is not an acceptable sentence. In the lexical entry for *weep*, therefore, it is necessary to specify that this particular verb requires an animate, perhaps even a human, noun phrase in the subject position. But if the lexical entry for *weep* already specifies that it must take a noun phrase as its subject, R1 is simply redundant and can be discarded. The grammatical rule has been caught in the lexical entry.

The need to treat morphology also drew attention back to words. Initial accounts of generative grammar drew heavily on English examples—and English, of course, relies largely on word order to signal the grammatical roles of various noun phrases. No doubt as a consequence of this emphasis on word order, the inflectional morphology of English has grown progressively simpler over the centuries; plural, possessive, and tense inflections are about all that is left. It was probably an advantage that the early versions of generative grammar did not need to struggle too much with the intricacies of inflectional morphology, but eventually such matters must be faced; in many languages,

the morphology is more complex than the syntax. To extend the theory to account for rules of word formation as well as rules of sentence formation, a closer examination of words was unavoidable.

In recent years, therefore, there has been a revival of scientific interest in words, both in linguistics and in psycholinguistics. This book, a product of that revival, attempts to bring together existing scientific knowledge of words. The approach is synchronic, not diachronic. That is to say, the concern here is not with where words come from or how they change, but rather with the much less familiar science of words as living components of the reader's own mental life.

Overview

Clearly, the study of words is part of the study of language, and so falls directly into the province of linguistic science. Any serious discussion of words must rely heavily on linguistic generalizations and hypotheses. But words are too important to leave to linguists. Words concern everyone. And because everyone is interested in them, words have been studied from many perspectives. A variety of approaches are assembled in this book in the hope that they will combine to yield a more comprehensive appreciation of this ubiquitous and essential unit of language. But, since selection is unavoidable, in these pages a synchronic view of words is favored over diachronic accounts of their histories.

The discussion should begin with a definition of "word." Of course, everyone knows what words are—and it is fortunate that they do because, as Chapter 2 will show, a good definition is hard to find. Part of the difficulty arises from the fact that words lead a double life. On the one hand, words are simply physical things or events—noises, gestures, marks; on the other hand, they express meanings. To know a word is to know (at least) two different kinds of things: First, it is to be able to produce and recognize physical tokens representing the word; second, it is to understand the meanings that those tokens can be used to communicate. The basic structure of lexical knowledge, therefore, is a mapping between two sets: the set of word forms and the set of word meanings.

That cognitive structure provides the organization for this book. Chapters 3 through 7 delve into the word forms themselves: written and spoken. These linguistic units are characterized by formal properties and relations—formal in the sense that they pertain to the forms of words, not their meanings. Since these forms are the most tangible manifestations of linguistic competence, they have been described and analyzed with great precision, yet there is much we still do not understand about them. But we do know that they control important aspects of human language.

Given that background, Chapter 8 turns to the system of word meanings—the lexical concepts that word forms can be used to express. These are intangi-

ble linguistic units whose physical instantiations are still unknown, but whose properties and relations can be inferred from the use that is made of them. The story is complicated, however, by the fact that a word's meaning interacts in complex ways with the grammatical role that it plays. In English and other Indo-European languages, words in different syntactic categories express different kinds of meanings; Chapters 9 through 11 discuss the semantic structures of nouns, verbs, and modifiers.

A final chapter pulls together some of these ideas in terms of what is known about the way words are learned—about the growth of vocabulary in children and adults.