Chapter 1   Classification of Languages

W̲hy bother classifying languages? There are several reasons why it is advantageous to do so. First, without a meaningful way of classifying languages we would not have an efficient framework within which to compare and contrast the numerous languages of the world. Second, a meaningful classificatory system allows us not only to arrange languages very neatly in their "pigeonholes" or "sample cases" like butterflies, but perhaps also to discover something new that we did not know before. Of course, not all classificatory frameworks are "meaningful" in this way; they may not always lead us to discover something new about the languages we are classifying or even display the truly important similarities and differences of these languages in a revealing and efficient way.

In the case of languages, it seems, it is necessary to have several different frameworks by means of which different important aspects of the world's languages are displayed for our inspection and important and perhaps we may be led to discover important correlations among various features.

GENETIC CLASSIFICATION

The oldest scientific way of classifying languages is into "language families." This method is called "genetic classification." Languages are considered to be "genetically related" if they can be shown (by using methods which will be discussed in some detail shortly) to be descended from the same parent

language, or "proto-language." Thus, it can be shown that Russian and English are both descended from the same source language, dubbed Proto-Indo-European by linguists, and are therefore "genetically related" or "belong to the same language family."

How do linguists go about determining whether a given pair of languages belongs to the same language family? Basically, the process employed seeks to establish that there exist between such languages "systematic correspondences" that cannot be explained by any means save common origin of the languages in question.

A very common misconception concerning genetic relationship is that the related languages must somehow be "similar," especially in a superficial sort of way, that is, words in one language must show some phonetic resemblance to words with the same meaning in the other language. However, although very often this *may* be the case, especially if the languages in question are very closely related, it is *not* an absolute requirement for establishing genetic relationship. Furthermore, superficial resemblances can often be explained on the basis of such phenomena as borrowing, and therefore not only are they not necessarily present in the case of related languages, but they do not by themselves constitute valid proof that the two languages are genetically related. Linguistics departments often receive letters from well-meaning amateurs who, struck by some chance superficial similarities between languages spoken in widely separated corners of the globe, propose a new genetic alignment for the languages in question.

To establish that a pair of languages are genetically related one needs to demonstrate that *there are recurring sound correspondences between the words of the two languages which have roughly the same meaning and belong to the basic vocabulary.* The more such sound correspondences recur, the stronger the proof of genetic relationship.

Why *sound correspondences?* For the most part it seems that the connection between form (sound) and meaning of words is quite arbitrary. For example, there is no good reason why the English sequence of sounds *tree* should mean what it does, or for that matter that the same meaning is conveyed in Chinese by the sound sequence *shù*. It is true that there are some words in each language in which the connection between sound and meaning is not so arbitrary. For example, each language has some words like English *boom* and *buzz* (onomatopoeic expressions) which imitate the sounds they represent. Also, an exception to the general arbitrary nature of the connection between form and meaning are the so-called nursery words, such as *mama* and *papa*, which are found in most of the languages encountered so far.

These words are not entirely arbitrary because they have their origin in infants' babbling stage, which is governed by universal rules and tendencies built into each human baby's linguistic repertoire and are therefore independent of particular language. (Even so, there is still a large dose of arbitrariness in both onomatopoeia and nursery words. Take, for example, the onomatopoeia for rooster's crowing in English and Japanese: *cock-a-doodle-doo*

and *kokekokko*, respectively. Or consider the following fact: in the Georgian language, spoken in the Caucasus region between Europe and Asia, *mama* means 'father' and *deda* means 'mother'.)

Another reason that sound correspondences are used to establish genetic relationship and not other formal characteristics of language, such as word order or structure of relative clauses, is that number of possible differences among languages in regard to these formal characteristics is surprisingly small: not all logically possible word orders actually occur among natural languages, and there seem to be only a few possible types of relative clause constructions. Thus, since the number of possibilities is so small, the likelihood of chance similarities and chance "correspondences" in these aspects of language structure is rather great. As we shall see when we discuss typological classification of languages, there are very compelling reasons for *not* taking typological parallels as proof of genetic relationship except in very special situations.

The sound correspondences have to 'recur'—some linguists prefer to say that the sound correspondences have to be "regular"—to ensure that they are not due to chance. It would be highly unlikely that by sheer chance there would arise a recurring correspondence between two sounds in two different languages in words meaning the same thing. For example, one can establish the following correspondence between English and Russian: English *s*:Russian *s*. This correspondence appears in such words as *son* and *sister* (Russian *syn* and *sestra*, respectively) as well as some other words. Of course, linguists do not look only for a couple of recurring correspondences to establish genetic relationship; they look for as many as they can find in order to strengthen their case in support of the genetic relationship claim. Also—and this cannot be emphasized enough—linguists do not look at individual correspondences in isolation: the comparativist who notes the existence of the *s:s* correspondence between English and Russian will also have to demonstrate that the rest of the sounds in the words which exhibit this correspondence also exhibit a recurring relationship. Thus, in our example it is possible to demonstrate that there is another recurring sound correspondence, *r:r*, in the word for *sister* (e.g., compare English *three* and Russian *tri*).

Why do we insist on "basic vocabulary," and what does this term mean? We have already eliminated two kinds of words from consideration—onomatopoeia and nursery words—because these words exhibit crosslinguistic similarities due to universal tendencies and a certain degree of nonarbitrariness in the connection between form and meaning. The reason we need to eliminate all but the basic, everyday type of vocabulary (usually said to consist of items such as names of body parts, kinship terms, natural phenomena not limited to a particular climate or place on earth, bodily functions, etc.) is that this type of vocabulary is not as readily borrowed from language to language as is other type of vocabulary.

Unfortunately for linguists, in certain circumstances everything may end up being borrowed, and therefore insisting on the basic vocabulary will not absolutely guarantee that one will eliminate all possible borrowings. There

are two types of borrowing. In the usual situation only nonbasic vocabulary items tend to be borrowed: those items for which the borrowing language may not have handy terms, as in the case of new inventions, foreign philosophical concepts, imported fads, fashions, and so on. In the less usual case (but by no means very rare) there may be a prolonged contact, such that there is a high degree of bilingualism, or one language has such a marked prestige over the other language that the lower-prestige language borrows even basic vocabulary items. Not only is basic vocabulary borrowed in such situations but there is usually massive borrowing of vocabulary, which in turn may affect the rest of the grammar of the borrowing language. Such massive borrowings have occurred in the history of English (from French) and Japanese (from Chinese).

Massive borrowings usually leave some clues that they have occurred. For example, very often languages that have borrowed extensively from other languages, including basic vocabulary items, will have many doublets: two ways of referring to the same thing, one borrowed and one native. Thus, in Japanese there are two sets of numerals, one set borrowed from Chinese and the other the native Japanese set. Although the use of the two sets in Japanese is not random but usually depends on what is being counted, either set can be used when one is not counting anything in particular. The existence of such doublets immediately raises the suspicion that one member of such doublet has been borrowed, for languages usually avoid the luxury of having two different words for everything unless there is some external factor, such as prestige, that compels the borrowing language to tolerate such lexical redundancies.

If an intimate contact between various languages (whether genetically related or not) continues for a long time in a certain geographic area, that area may develop into a "linguistic area," that is, an area in which languages share a number of linguistic traits in common *not because these traits have been inherited from the same parent language but because these traits have diffused from one language into another.* On the Balkan peninsula, in southeastern Europe, we have an example of a linguistic area which involves related languages belonging to various branches of the Indo-European language family. One of the several features that is shared among the languages of this region is the presence of a postposed definite article, which is not an inherited feature from their ultimate parent language Proto-Indo-European but must have been a local innovation in one of the languages that somehow spread to other languages in the area.

An example of a linguistic area which involves languages that are not recognized as being genetically related is India. In India, there are languages belonging to three different language families: the Indo-Aryan group (belonging to the Indo-European language family), Dravidian (Dravidian language family), and Munda languages (Austroasiatic language family). In spite of being genetically unrelated, these language groups all share some linguistic features in common which must have diffused throughout the area from one language to another. One such feature here is the presence of series of

retroflex consonants in all the languages of the subcontinent. Since neither Indo-European languages nor Austroasiatic languages outside of the area are noted as having this feature, and since it is clear that we have to reconstruct a retroflex series of consonants for Proto-Dravidian, it is reasonable to assume that retroflex consonants have diffused into Indo-Aryan and Munda languages from Dravidian. This is yet another reason why it is not very convincing to argue for genetic relationship solely on the basis of typological similarities.

Finally, we should ask ourselves whether it is possible to prove that some languages are *not* genetically related. Upon a moment's reflection one should realize that such a proof is strictly speaking not possible. After all, the fact that we have never seen a Martian does not necessarily *prove* that Martians do not exist. (In general, it is much harder to prove a negative hypothesis than a positive one: the discovery of even a single Martian would prove that Martians exist, but one would have to show that he has searched high and low in all the likely places for Martians and not found any before most people would finally accept the proposition that there are no Martians.) In the same way, just because one cannot at a given moment find the necessary evidence that a given pair of languages are genetically related does not necessarily mean that the languages in question do not have their ultimate origin in the same protolanguage. It may simply mean that the necessary evidence is no longer available. After several millennia of separation the two languages in question as well as their basic vocabularies may have changed so much that the sound correspondences linguists can set up recur so few times that it is hard to convince anyone that any genetic relationship exists between these languages. It may well be that *all* the languages presently spoken in the world are genetically related; however, at present we do not have the means either to prove or disprove this hypothesis. It is estimated by some linguists that our present methods of establishing genetic relationship among languages work only for languages that have been separated from each other for less than five thousand years; at the same time it seems highly probable that human beings have been speaking for tens of millennia.

GENETIC SUBGROUPING

After establishing which languages are genetically related and which appear not to be, there are several other things that comparativists can do. First, they can attempt to reconstruct as much of the parent language of the related languages as possible. At present fairly reliable, though by no means perfect, methods of reconstructing the sound system, lexicon, and morphology are available. Much work is being done on establishing comparable methods for the reconstruction of the syntax of the parent language, but there is not yet any widely accepted comparative methodology for the reconstruction of *syntax*. This may be so because we seem to know more about the *synchronic* syntax

working of both phonology and morphology and therefore understand more how these components of grammar may change through time, whereas in syntax we haven't yet been able to hit upon the correct model for the description of the synchronic workings of syntax. (One of the requirements for a syntactic theory, then, would be that it must, among other things, offer a basis for explaining how syntactic change takes place.)

Second, comparativists may attempt to ascertain which among those languages that they have already determined to be genetically related are "more closely related" among themselves and which are more "distantly related." Again, what linguists mean by "closely related" is commonly misunderstood to mean 'superficially more similar to each other', and therefore it must be stressed that this expression, "more closely related," has a special, technical meaning in this context.

Later we shall look at some concrete examples from real languages; for the sake of clarity, let us look at a simple hypothetical case first. Suppose that you have just examined three different languages X, Y, and Z and determined that all three of them are genetically related. As you begin to reconstruct the sound system of their protolanguage you come to the conclusion that several sound changes that you have posited must have taken place in the history of both X and Y. Both X and Y appear to have changed a word-final -*m* to -*n* and a word-final -*k* to a glottal stop. In addition, both X and Y have devoiced all obstruents (stops, fricatives, and affricates).

How can we explain why X and Y share these three changes? There are several possible answers. First, it could be a coincidence that both X and Y underwent the same changes. However, the more such shared changes there are, the less likely it is that this phenomenon results entirely from coincidence. Coincidence as an explanation is likely only if all the changes that X and Y have in common are very common sound changes that are well-motivated on phonetic grounds. Even so, it would be highly improbable that two separate languages would undergo a whole series of *identical* sound changes through sheer coincidence. If the sound changes shared are not so usual or the number of such shared changes argues against chance, another possible explanation offers itself: the shared changes must have taken place not when X and Y were already separate languages but before they became separate languages. That is, there was an *intermediate protolanguage*, Proto-XY, during the existence of which these changes took place, and then X and Y split into separate entities each of which had its own separate development from that point on.

The relationship among X, Y, and Z can be represented by means of a "family tree diagram" (see Figure 1.1). This diagram allows us to explain the shared changes between X and Y by proposing that the split between X and Y was a later phenomenon that the split of the ancestor of X and Y, Proto-XY, and Z. Thus languages X and Y are considered to be 'more closely related' to each other than either is to Z. (It should also be noted that although for the sake of illustration we have used sound changes, the shared changes or innovations may consist of other types of changes, say, lexical or morphological, or even a combination of all types of changes.) Note that even though
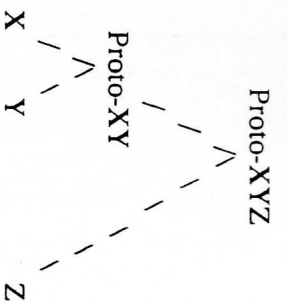
X and Y have been declared to be more closely related to each other than either is to Z, X and Z may share a number of similarities that are not shared by X and Y, the 'closely related' languages. Thus, superficially at least, X and Z may look as if *they* were the more closely related pair, and usually beginners jump to the conclusion that these indeed are the most closely related languages. What tends to be forgotten is that the basis for determining which languages are more closely related is *not* superficial resemblances, which may be due largely or even entirely to *shared retentions* of the features of the ancestor language of the entire family, but *shared innovations or changes*. Accordingly, unless one has reconstructed the protolanguage (at least in outline form) of the entire family one cannot do genetic subgrouping because the latter task requires a determination of which features of the protolanguage have changed in each language and which have been retained in each language.

The method of genetic subgrouping just described and the family tree diagram that is usually used to display the results of such subgrouping both suffer from serious deficiencies because they assume certain things that are not always actually true. First, it is quite possible that the reason why X and Y have some innovations in common is that various changes diffused from X to Y and vice versa *after* X and Y had become differentiated languages. In such a case it would, of course, be erroneous to ascribe the shared innovations to a period when X and Y were still a part of the same language. The tree diagram, too, implies that once split, languages cannot influence each other, which is quite incorrect. The existence of the Balkan linguistic area and similar phenomena elsewhere disproves this. In addition, the family tree diagram seems to imply that splits between languages are neat, clean breaks which happen at a particular moment in time. Such "breaks" may occasionally occur when an entire section of population moves suddenly to a far off location, or when for some other reason communication is cut off between various groups of people speaking the same language, but usually we encounter situations in which dialectal differences gradually build up over the course of time; different dialects continue to be spoken in the vicinity of each other and continue to influence each other linguistically. Finally, there is no specific point at which we can say that two dialects of the same language have become separate languages in their own right. One may argue that one such

Figure 1.1

point is reached when the two variants are no longer mutually intelligible. However, mutual intelligibility itself is a continuum without sharp, neat breaks.

Another method of subgrouping, called "lexicostatistics," is used quite often in those cases where the protolanguage of the family has not been reconstructed yet or the data on individual languages—especially on the historical changes that have taken place in them—are severely limited. (Very often in such cases the only thing a linguist has to work with are word lists of the languages involved.) Basically, this method involves calculating the percentages of "shared cognates," that is, words traceable to the same historical source, in the basic vocabularies of the languages being subgrouped. In the example that we discussed, the same family tree would be obtained if languages X and Y were to be found to share, say, 80 percent of basic vocabulary in common whereas the percentage of shared basic vocabulary that either X or Y shares with Z is significantly lower than 80 percent (say around 50 percent).

Most linguists do not consider lexicostatistics to be a very reliable tool for subgrouping for several reasons. Lexicostatistics simply counts the number of shared cognates without ascertaining whether this sharing of cognates is a result of lexical innovations, lexical retentions, or borrowing from each other or from outside sources. If we consider that in the first place genetic subgrouping was set up to explain the phenomenon of shared innovations in related languages, it becomes obvious why a method that ignores the distinction between innovations and retentions is regarded as being a very crude tool for subgrouping. Second, normally one looks at all kinds of innovations, sound changes, lexical changes, morphological changes, and the like, all of which strengthen the case for a particular subgrouping, whereas lexicostatistics looks only at the lexicon. Finally, lexicostatistics sometimes poses a very peculiar problem for anyone who wants to draw a family tree on the basis of lexicostatistical data—what to do when, say, X and Y have 75 percent of basic vocabulary in common and Y and Z share 75 percent, but X and Z share only 50 percent in common. (On the basis of the above, X and Y should belong to the same branch, and Y and Z should belong to the same branch, but X and Z should not, which is a paradoxical situation.) Difficulties such as these have made most linguists wary of lexicostatistics.

I now review the method of establishing genetic relationship and doing genetic subgrouping by examining some data from real languages. To keep things simple, only a small portion of the relevant data will be presented (see Table 1.1).

There are no consonant clusters or word-final consonants in Samoan and Hawaiian. Maranao is a language spoken in the Philippines.

ESTABLISHING GENETIC RELATIONSHIP

First of all, even though this may seem obvious from Table 1.1, we must *formally* show that these languages are all genetically related by demonstrating that there are at least some recurring sound correspondences among all of

---

Table 1.1

GLOSS		MALAY	SAMOAN	MARANAO	HAWAIIAN
1.	'two'	dua	lua	dua	lua
2.	'five'	lima	lima	lima	lima
3.	'sky'	laŋit	laŋi	laŋit	lani
4.	'to cry'	taŋis	taŋi	ʼəʼul	kani
5.	'louse'	kutu	ʼutu	kutu	ʼuku
6.	'lobster'	udaŋ	ula	udaŋ	ula
7.	'I'	aku	aʼu	aku	aʼu
8.	'rafters'	kasaw	ʼaso	kasaw	ʼaho
9.	'mistake'	salah	sala	salaʼ	hala
10.	'eye'	mata	mata	mata	maka
11.	'pandanus'	pandan	fala	fala	hala
12.	'hibiscus'	baru	fau	raguruy	hau
13.	'house'	balay	fale	walay	hale
14.	'coconut'	ñiur	niu	niug	niu
15.	'hardwood'	teras	toa	tegas	koa
16.	'die'	mati	mate	matay	make
17.	'way, path'	jalan	ala	lalan	ala
18.	'drink'	minum	inu	inum	inu

ŋ = velar nasal consonant
ñ = palatal nasal
ʼ = glottal stop

them. For example, one can establish the recurring correspondences for the data given (see Table 1.2).

Note that many other sets of recurring sound correspondences can be found in the data, but they do not necessarily involve basic vocabulary items. (For example, *b:f:w:h* correspondence which occurs in the items for 'hibiscus' and 'house'.) On the other hand, many of the nonbasic vocabulary items listed exhibit the same correspondences as the basic vocabulary items. (For example, the item for 'hardwood' exhibits the same initial consonant correspondence as correspondence 3 in Table 1.2.) This latter fact suggests that at least some of the nonbasic vocabulary items are not borrowed forms

Table 1.2

MALAY	SAMOAN	MARANAO	HAWAIIAN	ITEM
1. l	l	l	l	2, 3, 9, etc.
2. m	m	m	m	2, 10, 16
3. t	t	t	k	10, 16
4. u	u	u	u	1, 7, 18
5. a	a	a	a	1, 2, 9, etc.
6. i	i	i	i	2, 3, 18

but inherited from the common parent or protolanguage of these four languages.

It should also be pointed out that although it looks as if all four languages in question are genetically related, not all of their vocabulary items, whether basic or nonbasic, are necessarily cognate (i.e., traceable to the same source). Since even the basic vocabulary may be replaced through semantic change, we do not necessarily expect that all the basic vocabulary items be cognate in the related languages. (For example, compare the Maranao word for 'to cry' in Table 1.1).

GENETIC SUBGROUPING

Now that it has been formally established that all the languages in our sample seem to be genetically related, I will attempt to establish which of the languages are more closely related to each other, specifically, which of the languages share a period of common development after the split from the parent language of all four of the languages we have cited. In other words, we have to look for innovations which are shared by two or more related languages.

Once we have established that certain languages are genetically related, we no longer need to exclude the nonbasic vocabulary items from our consideration. Of course, we still have to be on the lookout for possible loanwords, but once we have established what the "regular" sound correspondences are in the basic vocabulary of these languages we can consider those nonbasic items that exhibit the same sound correspondences to be also cognate. If these languages are related, it stands to reason that at least some of their nonbasic vocabulary items are also cognate. (One of the ways we identify possible borrowings is by observing which items exhibit unexpected sound correspondences.)

The correspondences cited in Table 1.2 to establish genetic relationship are not useful for genetic subgrouping, with the possible exception of correspondence 3. All the other sound correspondences listed are *identities*, that is, each language has exactly the same reflex of the protophoneme, and therefore, if we are looking for shared innovations, or changes that took place only in *some* of the related languages, these correspondences are not very helpful.

Correspondence 3 turns out to be of little help also because it is not possible to determine without more evidence whether the Hawaiian language alone has innovated (by changing *t* to *k*) or whether Malay, Maranao, and Samoan have innovated by changing *k* to *t*. If we could show the latter case to be correct, we could then claim that the three languages just mentioned are more closely related to each other than any of them is to Hawaiian. (Given more evidence, which we won't cite here, it becomes clear that it is Hawaiian alone that has innovated in this case.)

By adding sound correspondences that are derived both from basic and nonbasic vocabulary items we can observe that the languages in question seem to fall into two groups (see Table 1.3).

Table 1.3

	MALAY	SAMOAN	MARANAO	HAWAIIAN	ITEM
7.	d	l	d	l	1,6
8.	s	∅	s	∅	15,14
9.	h	∅	'	∅	9
10.	t	∅	t	∅	3
11.	m	∅	m	∅	18
12.	ŋ	∅	ŋ	∅	6
13.	g	∅	r	∅	14

Let us first compare correspondence 1 in Table 1.2 with the correspondence 7 in Table 1.3. Correspondence 1 has *l* in all four languages, and it is therefore not unreasonable to assume that this correspondence most likely reflects *l of the protolanguage. In correspondence 7 on the other hand some languages have *l* (Samoan and Hawaiian) and some have *d* (Malay and Maranao). In this instance there are at least two possible explanations that we may consider: (1) Malay and Maranao have changed the original *l to *d* under some conditions but preserved it as *l* under some other conditions. (2) The original protophoneme *d is preserved as such in Malay and Maranao but is shifted everywhere to *l* in Samoan and Hawaiian.

Since both *l* and *d* appear in pretty much the same environments in Malay and Maranao—initially in items 1 and 2 or between vowels as in items 6 and 9 in Table 1.1—it is not possible to formulate a rule which would correctly predict when the hypothetical *l became *d* and when it stayed *l* in Malay and Maranao. Furthermore, since the comparative method does not allow us to posit an unconditioned phonemic split (positing that some *l's shifted to *d* irregularly) we must reject the first hypothesis and look at another alternative.

The second alternative, on the other hand, does not go counter to the usual assumption of the comparative method that sound changes are regular: there are no *d*'s in either Hawaiian or Samoan which have not shifted to *l*. Therefore, we accept the second hypothesis, which means that it is Hawaiian and Samoan that share an innovation in common: *d → l.

In addition, correspondences 8 through 13 offer additional evidence for a closer genetic relationship between Samoan and Hawaiian. Although most of these correspondences do not recur in our data, the words in which they occur are clearly cognate since they exhibit many of the regular correspondences established on the basis of basic vocabulary. What these cases all share is that Malay and Maranao exhibit various consonants whereas Samoan and Hawaiian have zero reflexes. Also, these correspondences all occur at the end of words. Here, too, there are two logical alternatives. (1) Malay and Maranao have added various consonants at the end of words; (2) Samoan and Hawaiian have deleted all word-final consonants.

Again, it is not possible to justify accepting the first hypothesis since it is obvious that this would go counter to the usual assumptions of the comparative method: Malay and Maranao did *not* add a consonant at the end of every word (compare 'five' and 'lobster'), and there is no way to account for different consonants being added after identical vowels (compare 'sky' and 'to cry' in Malay).

On the other hand, if we posit that Samoan and Hawaiian have dropped all word-final consonants, there are no counterexamples in our data to account for. Moreover, deletion of word-final consonants is a more common phenomenon than epenthesis, or insertion, of word-final consonants (although the latter does occasionally occur). The clincher is provided by some evidence from Samoan and Hawaiian that the word-final consonants in Malay and Maranao represent a shared retention from the protolanguage. For example, the passive form of the verb 'to drink' in Hawaiian is *inumia*, which retains the original ending of the verb stem (*m*) because of the following passive suffix -*ia*. (Similarly, Samoan also shows internal evidence suggesting that word-final consonants were deleted.)

To sum up, I have established that Samoan and Hawaiian share at least two innovations (sound changes) in common: (1) shift of *d to l in all environments, and (2) deletion of all word-final consonants. Therefore I conclude that Samoan and Hawaiian are more closely related to each other than either is to Malay or Maranao. But what about Malay and Maranao? Don't they look "more closely related" because their forms are so similar? Although it would be rather tedious to demonstrate, it is indeed not possible to establish on the basis of the data given that Maranao and Malay have any innovations in common. All that they share are the retentions. That is, they look very much alike because they are both comparatively conservative languages, not because they have changed in a similar manner. Thus in spite of their superficial similarity, they by definition do not qualify as 'more closely related'.

Proto-Austronesian

Proto-Polynesian

Samoan   Hawaiian   Malay   Maranao

Figure 1.2

Accordingly, the interrelations among these languages can be schematized by means of a tree diagram (see Figure 1.2). Other types of diagrams representing genetic subgrouping are possible, but while they are in many respects more accurate than the tree diagram, the tree diagram remains the most popular way of representing genetic subgrouping. In this book the tree diagram is the primary visual means of accounting for genetic subgrouping, but the reader is cautioned about the limitations of this kind of representation.

TYPOLOGICAL CLASSIFICATION OF LANGUAGES

Languages may also be classified according to the type of their structure. Such a classification is potentially much more revealing, since genetically related languages are not necessarily also similar in structure, whereas by claiming that two languages belong to the same *type* of language one does claim that they share at least some similarities. Thus, for example, saying that Russian and English are genetically related merely makes the claim that these two languages have a common origin but does not necessarily imply that the two languages have preserved many common features. Thus a person who knows English would not on the basis of such a claim expect that Russian would necessarily be similar to English in respect to word order. On the other hand, if one is told that both English and Russian are Subject + Verb + Object languages, then one would know that at least in regard to basic word order (and even in respect to the order of other types of syntactic constituents) the two languages are very similar.

Actually, genetic and typological classification supplement each other very well: the former tells us something about the origins of a particular language, whereas the latter tells us something about the structure of that language. Typological classification is not yet as fully developed a science as is genetic classification. The reason for that is that until relatively recently linguists had not identified the really important, fundamental features on which we should base the typological classification of languages. After all, languages may differ from each other in many ways, and since all languages differ from each other in at least *some* respects, a classification of languages that took into account *all* the features would be too cumbersome and lead to a taxonomy in which each individual language would be a separate type. In the past, there were many proposals for a meaningful typological taxonomy of languages focusing primarily on morphology; in recent years there have also been very interesting proposals for syntactic taxonomy, and even one for a holistic approach that combines salient features of phonology, morphology, and syntax.

PHONOLOGICAL TYPOLOGY

One may classify languages into tone languages, pitch-accent languages, and dynamic accent languages. Each of these categories can perhaps be further subdivided. For example, one can subdivide tone languages into those that have (underlyingly) only level tones and those that in addition to level tones,

have contour tones (i.e., tones which change pitch direction). One can also subdivide dynamic accent (stress) languages according to whether the placement of stress is predictable or unpredictable in them.

It is possible to classify languages also according to other features of their phonological systems. For example, one can classify languages according to the number of vowel phonemes in their phonological system, and then further subclassify the languages according to the particular vowels found in each system. For such type of language classification one may turn to the pioneering work of Trubetzkoy (1969), who, however, was not primarily interested in typological classification for its own sake. Also, Hockett (1955) presents a discussion of phonological typology based on the American structuralist model of phonology.

MORPHOLOGICAL TYPOLOGY

Morphological structure has been the traditional basis for language classification, and those readers who are interested in a fairly detailed overview of the history of language classification based on morphology may wish to consult Horne (1966). Here I provide only a summary of important aspects of morphological typology.

In spite of many terminological differences (and some terminological confusion) various scholars working in this field in the nineteenth and early twentieth centuries more or less agreed that there are three or four major types of language. Each of the four traditional morphological types listed below is followed by the name of a language usually cited as supposedly epitomizing that type. The usual terms referring to each type of language appear in parentheses. (Since different typologists define these terms somewhat differently, the terms in question are cited here only for reference.)

Type I: Classical Chinese (*analytic/isolating*)
Type II: Turkish (*agglutinative*)
Type III: Latin (*fusional/synthetic/inflected*)
Type IV: Eskimo (*polysynthetic/incorporating*)

Although one may discuss the main differences among the four language types under various rubrics, I focus here only on the most salient ones, emphasizing some aspects that were not necessarily emphasized by the scholars who set up the original typological schema.

The main division should be made between Type I and Type II languages, on the one hand, and the Type III and Type IV, on the other. The feature that is the main differentiator here is the transparency of word structure: In Type I and Type II languages the morphological makeup of words is crystal clear and the function of various morphemes is relatively easy to ascertain, whereas in Type III and Type IV languages word structure tends to be obscured by various factors. (Note that it *tends* to be obscured but is not necessarily obscured in all cases.)

Although Turkish and other Type II languages tend to have long, morphologically complex words consisting of series of affixes attached to the root or

stem morpheme, and Classical Chinese (Type I) has relatively short words and extremely few affixes, in both language types the morphological structure of words is completely transparent because the boundaries between adjacent morphemes are seldom obscured (i.e., morphological cuts are seldom in doubt), nonautomatic morpheme alternants (i.e., morphologically conditioned allomorphs) are few or even nonexistent, there are no morphemes whose status as independent morphemes is questionable (such as various morphemes which merely connect other morphemes to each other, thematic vowels, conjugation markers, etc.), and affixes usually have a well defined grammatical function.

In languages belonging to Type III and Type IV, the morphological structure of words may be obscured by the presence of "empty morphemes" (such as *cran-* in *cranberry*), suppletion, and other types of nonautomatic morpheme alternants. Also, such languages tend to have a fair number of portmanteau morphs and even what I call "portmanteau morphemes." An example of a portmanteau morph is French *au* which is the realization of a sequence of two morphemes: the preposition meaning 'to' and the masculine definite article, both of which have independent realizations in other environments. An example of a portmanteau morpheme is the suffix *-i* in the Latin word *fili-i* 'of the son', which signals at the same time masculine gender, singular number, and genitive case, none of which is ever signaled in Latin by an independent morpheme.

The difference between a Type I language and a Type II language is the same as the difference between a Type III and a Type IV language: Type I and Type III languages have relatively less affixation than Type II and Type IV; what in languages of Types I and III is expressed by independent words very often may, or even must, be expressed by bound affixes in languages belonging to Types II and Type IV. In general, Type III languages have more affixation than Type I languages but not as much as Type II or Type IV. Although languages belonging to Type I have very few affixes, it does not mean that their words are not morphologically complex—usually languages of this type allow quite free compounding of stem or root morphemes within a word.

Languages belonging to Type I and Type III place a relatively greater burden on syntax than do languages belonging to Type II and Type IV, which place a relatively greater burden on morphology. Type I languages especially tend to have a fairly rigid word order to signal various grammatical relations between words. That is not to say that languages belonging to other types may not have a fairly rigid word order as well; we are talking here of natural tendencies, not necessarily rigid laws. In this case "natural tendency" has a reasonable explanation: it is only natural that languages with little affixation should employ word order to mark at least some grammatical relations, and the fewer morphological devices there are in a language for marking such relations, the more likely it is that syntactic devices such as word order will be used instead.

In order to illustrate the "spirit" of the structure exemplifying each type of language, a few fairly typical sentences from Classical Chinese, Turkish, Latin,

and Yup'ik Eskimo are cited, together with a morpheme-by-morpheme translation and an idiomatic English translation, followed by a brief commentary on the salient aspects of the grammatical structure of the example sentences.

In all the examples to follow, the romanized version of the sentence or phrase uses spaces to mark word breaks and dashes to mark morpheme breaks. The morpheme-by-morpheme translation of the examples marks word breaks by slashes and morpheme boundaries by pluses. (This format is followed throughout the book.)

Type I: Classical Chinese

A. 下 馬 人 車 中

xià mǎ rù chē zhōng

descend/ horse/ enter/ chariot/ middle

'[He] got off the horse [and] got into the chariot.'

B. 松 下 問 童 子

sōng xià wèn tóng zǐ

pine/ under/ ask/ lad/ diminutive marker

'Under the pine trees [I] asked the boy.'

C. 楚 莊 王 賜 群 臣 酒

Chǔ Zhuāng Wáng cì qún chén jiǔ

name of a region/ proper name of a person/ king/ bestow/ flock/ official/ wine

'King Zhuang of Chu bestowed wine on his ministers.'

D. 豬 人 立 而 啼

zhū rén lì ér tí

pig/ man/ stand/ and/ cry

'The pig stood up [like a] man and cried.'

The morpheme *xia* can act as a verb (as in sentence A where it was translated as 'descend') or as a postposition (as in sentence B where it is translated as 'under'). Likewise *ren*, which usually functions as a noun ('man'), appears in an adverbial function in sentence D without any formal marking. This illustrates one of the secondary traits of Type I languages: One and the same morpheme may sometimes act as a different part of speech depending on the context *without any morphological marking to signal the different function.*

Sentence C illustrates the fact that word order alone marks three different grammatical relationships among the noun phrases: the subject (King Zhuang), the indirect object (his ministers), and the direct object (wine).

There are no true affixes in the preceding examples; all the words cited are stem morphemes. The only morpheme that seems to act as an affix is *zi*, the diminutive. However, this morpheme appears also as an independent stem meaning 'son' or 'offspring.'

No morphemes cited in the previous examples have any allomorphs (insofar as this can be determined from the logographic writing system); however, Early Classical Chinese does have some fused portmanteau morphemes.

Type II: Turkish

A. Köy-ün-den çık-ma-mış köy-lü bu mesele-ler-i anla-r-mı

village + third person possessive + from/ come + negative + past participle/ village + characterizing suffix/ demonstrative/ problem + plural + object marker/ understand + aorist + interrogative

'Does the villager who has not left his village understand these problems?'

B. Ev-ler-im-iz-den gel-mi-yor-d-um

house + plural + first person possessive + possessor pluralizer + from/ come + negative + progressive + past + first person

'I was not coming from our houses.'

It is fairly evident from the preceding examples that both nouns and verbs in Turkish may consist of relatively long strings of morphemes: a stem followed by a series of suffixes. What is not readily apparent from the examples is that allomorphy is overwhelmingly regular. (However, there are *some* irregularities present even in the two example sentences: The progressive aspect marker *yor*, for example, violates the rules of Turkish vowel harmony.) In addition there is a marked absence of portmanteau morphemes; concepts such as number, case, tense, aspect, and person are all marked by separate, independent morphemes.

Type III: Latin

A. Fili-us patr-em am-at

offspring + nominative singular masculine/ father + accusative singular masculine/ love + third person singular present indicative

'The son loves the father.'

which makes it look more like a Type III language. In addition, there are even a few examples of words containing long sequences of affixes, the allomorphs of which are perfectly regular, suggesting the type of structures common in Type II languages (e.g., the word *anti-dis-ESTABLISH-ment-ari-an-ism*). Thus, one may conclude that, typologically speaking, English is neither fish nor fowl: it is a typological anomaly. It turns out that most of the world's languages do not fit neatly within the categories of traditional morphological typology.

To remedy this problem, at least in part, Edward Sapir (1921) proposed a rather complicated refinement of the traditional morphological typology, but his schema provided up to 2,870 "pigeonholes" into which languages could be placed and was therefore too cumbersome. In addition, it did not completely do away with the problem of what to do with languages such as English which have features typical of different major typological classes of languages. Today it seems clear that any classificatory schema that attempts to pigeonhole an entire *language* within a typological schema cannot avoid at least some degree of arbitrariness.

Eventually, Joseph Greenberg (1960) hit upon the idea that one need not necessarily pigeonhole entire languages as belonging to one clearly delineated type or another. Instead, he proposed that one could calculate the *degree* to which a certain typologically important feature is present in a given language. Thus, for example, one could calculate the average ratio of prefixes per word by counting how many prefixes and how many words are found in a given representative passage in a given language. This is clearly a superior way of typologizing languages for the following reasons:

1. It avoids the arbitrariness of having to decide just how many prefixes (or whatever other feature) a language must have before it is labeled as "a prefixing language."

2. It reflects better the true nature of the language in that, for example, those languages that have a number of prefixes which are rarely used will have a very low index of prefixation, just like languages that have a very small inventory of prefixes to begin with. In the old schema, which was an "all or nothing" kind of classification, a language having only one or two prefixes would be classified as a prefixing language just like a language that had hundreds of different prefixes, and only languages that had no prefixes of any kind would be classified as nonprefixing.

3. It provides the possibility of comparing different *styles* in the same language in regard to certain typologically relevant features, thus recognizing that a language is not a perfectly homogeneous entity but has typologically quite distinct variants.

In his article, Greenberg proposes the following typological indexes:

1. *Degree of synthesis* involves the ratio of morphemes to words. Note that this index seems to combine the degree of affixation and the degree of compounding into one index.

2. *Index of agglutination* involves the ratio of agglutinative constructions (i.e., combinations of morphemes whose allomorphs are completely automatic in that environment) per morph junctures (i.e., the total number of morpheme junctures within a given word). Type II languages would naturally have a very high index of agglutination.

3. *Compositional index* is the ratio of roots per word. In other words, this index measures the degree of compounding in a given language.

4. *Derivational index* is the ratio of derivational morphemes per word. (We do not need to concern ourselves here with Greenberg's distinction between derivational and inflectional affixes.)

5. *Gross inflectional index* is the ratio of inflectional morphemes per word.

6. *Prefixial index* is the ratio of prefixes per word.

7. *Suffixial index* is the ratio of suffixes per word.

8. *Isolational index* is the ratio of instances of significant order per total nexus. Stated more simply, this index measures the degree of significant word order in a language. One might argue that this particular index is more syntactic than morphological in nature or that it at least straddles the line between the two.

9. *Pure inflectional index* is the ratio of instances of nonconcordial (see next index) inflectional morphemes per nexus.

10. *Concordial index* is the ratio of instances of concordial inflectional morphemes (in other words, morphemes which signal some kind of grammatical agreement) per nexus.

Curiously, despite some initial interest, nothing was done to improve and develop further Greenberg's indexes. That may be because the attention of linguists interested in language typology, including that of Greenberg himself, turned to syntactic typology, which blossomed because of a general shift of interest to syntactic phenomena stimulated by Noam Chomsky's generative syntax.

SYNTACTIC TYPOLOGY

The development of syntactic typology owes much both to Noam Chomsky, whose ideas greatly stimulated interest in matters syntactic in general, and to Joseph Greenberg, whose interest in language typology and language universals led him to take up the theory of *implicational universals* and to illustrate it by citing chiefly syntactic phenomena. (The theory of implicational universals was originally formulated by the Prague School linguist Roman Jakobson in 1958.) In an article entitled "Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements," published in 1963, Greenberg demonstrated that the presence of a certain syntactic feature often implies the presence of one or more other features. For example, he noted that languages with dominant VSO (Verb + Subject + Object) word order are almost always prepositional (i.e., they have prepositions rather than postpositions) and that there is a very strong tendency for SOV languages to be postpositional.

B. Pater fili-um vid-et

father + nominative singular masculine is here represented by a zero allomorph/ offspring + masculine accusative singular/ see + third person singular present indicative

'The father sees the son.'

The examples provided show only a moderate amount of affixation, which is normal for languages of this type. There are several examples of portmanteau morphemes (e.g., -us which at once shows case, gender, and number). Moreover, it is not always clear where the morphological cuts should be made. For example, the suffix representing third person indicative present has two allomorphs in the examples given: -at and -et. It may be argued that -t should be isolated as the marker of the third person singular, the vowels a and e would then be allomorphs of a present indicative morpheme. Alternatively, one could also recognize the vowels as independent morphemes signaling the verb conjugation. In either analysis there is a large element of arbitrariness involved, and the presence of such forms as am-ō ('I love'), in which the indicative present or conjugation marker morpheme is missing, complicates the morphological analysis since -ō is clearly a person marker regardless of what else it may be said to mark.

Finally, it is typical of Type III (as well as Type IV) languages that most root or stem morphemes are bound. That is, they cannot appear alone as independent words in the language. This is true of most of the roots given in the two preceding Latin examples, and perhaps one could make a very convincing argument that all Latin roots are bound. (For example, fili 'offspring' can appear as independent word in the vocative singular masculine, but only because the vocative singular masculine suffix is represented by a zero allomorph with this root.) Of course, affixes in all types of languages tend to be bound morphemes.

Type IV: Yup'ik Eskimo

A. Angyar-pa-li-yugnga-yugnar-quq-llu

boat + big + make + be able + probably + third person singular indicative intransitive + also

'Also, he can probably make big boats.'

B. Angut-em ner-aa neqa

man + relative singular case/ eat + third person singular transitive indicative/ fish

'The man eats the fish.'

Example A not only illustrates a long string of affixes but also the fact that

languages of this type often "incorporate" direct and indirect object nominals along with some of their modifiers by attaching them to a verb root. This is generally done only in case the object nominals are indefinite, as in example A. In those cases where the object is definite, the latter is expressed by an independent word (cf. example B).

It should be emphasized here that the presence of pronominal subject and object affixes attached to the verbs does not qualify a language as being incorporating.

Besides having portmanteau morphemes (e.g., third person singular indicative transitive), Yup'ik Eskimo has extremely complicated morphophonemics (e.g., suffixes which cause deletion of the preceding consonant and those that do not) which render many a morpheme boundary opaque.

In languages of this type, word order is not as important as the order of morphemes in a word. The same is true to a lesser extent in the case of Type II languages.

Perhaps it is not an exaggeration to say that languages of this type have a potentially infinite number of words. I could determine no theoretical limit on the number of affixes that can be appended to a single Yup'ik Eskimo word. In languages belonging to the other three types, a similar potential "infinity of words" effect may be achieved by means of compounding, which involves combining two or more stems or roots into single words. It should be noted, however, that languages differ significantly in the extent to which such processes are allowed to go. For example, English allows only a moderate amount of compounding, whereas German is famous for its very long compounds, which are often cited to amuse speakers of moderately compounding languages. The following example, cited from Uspenskij (1957:307), was seen in a German stamp catalogue. (Here plus signs mark the potential word breaks within the compound. The morpheme breaks are left unmarked.)

Kaiser + Wilhelm + Jerusalems + reise + gedächtnis + brief + karten + post + marke

'Kaiser Wilhelm's Jerusalem trip commemorative postcard stamp'

At this point the reader may legitimately wonder, To which of the above four types of languages does English belong? The answer to this question is not, unfortunately, a straightforward one. Although English shares many features with Type I languages such as Classical Chinese, it also shares some features with languages such as Latin. For example, the English words fish and man can both be used as either verbs or nouns without any special marking to indicate when they are used as verbs and when as nouns. In English, too, it is the word order alone that marks the case relations among the three noun phrases in the following sentence: "Peter gave John the book."

On the other hand, English has a lot of irregular morphology, including suppletion and portmanteau allomorphs (e.g., foot + PLURAL = feet),

The foregoing observations cited from Greenberg's article illustrate the principle of implicational universal very well: given that a language has VSO as the basic word order we can automatically assume that such a language has prepositions instead of postpositions, and given the fact that a language has SOV as the basic word order we can reasonably expect (though not assume with total certainty) that this language will have postpositions instead of prepositions. Moreover, in SOV languages attributive adjectives and relative clauses precede the nouns they modify, whereas in SVO languages they usually follow.

Of course, the implicational chaining need not stop with just one implication. It is quite possible, for example, to find a whole chain of implications where the presence of feature X implies the presence of feature Y, which in turn implies the presence of feature Z, and so on. It should also be noted that such implicational relationships are not limited solely to the domain of syntax but can be found in other components of grammar as well.

The discovery of implicational universals is a very important development for linguistic typology for two reasons. First, it allows for a more economical typological schema: if we choose the basic classificatory categories on the basis of those features that by universal simplication entail the presence of a large number of other important features, we need only state that a given language has this feature to imply, by universal convention, a whole series of other features as well. For example, by classifying a language as having SOV as its basic word order we imply that it is also a postpositional language.

Second, the existence of implicational universals alerts us to many phenomena that require explanation; that is, linguistic theory must explain why there are such interrelationships between various grammatical features. In turn, our search for such explanations usually leads us to discover even more hidden facts and relationships in language. Thus, what starts as a "merely" typological, classificatory endeavor eventually leads to ever expanding understanding of the workings of language.

It is not possible here to go into all the recent developments in syntactic typology. Here I shall only note that the basic word-order typology is the fundamental feature of syntactic typology. For a well written account of current developments in syntactic typology and typology in general, the reader is especially urged to consult the works of Bernard Comrie and William Croft listed in the bibliography at the end of this chapter.

LEXICOSEMANTIC TYPOLOGY

Among the various possible classifications of language on the basis of linguistic structure one may also mention classification on the basis of the features of the lexicon, though no one has yet proposed a fully developed classificatory schema based on the features of the lexicon. One often sees in the more popular works on linguistics claims that there are hundreds of monomorphemic words for camel in Arabic, similarly large number of monomorphemic words for different types of snow in Eskimo, large numbers of different words for coconut in Pacific Island languages, and so on. Such claims usually turn out to be grossly exaggerated since Eskimo and English are not

so different in the number of their monomorphemic words for 'snow', as Pullum (1991) points out, but to the extent that languages may actually differ somewhat in this respect, it may be worthwhile to consider the possibility of establishing some kind of classification on the basis of lexical differences among various languages.

At the moment it is not clear whether a comprehensive and meaningful classificatory schema could be constructed on the basis of such observations, but perhaps in the future a typology of languages based on the peculiar features of the lexicon will prove useful. Various works in ethnosemantics and cognitive anthropology certainly provide a good basis for such a schema. At present there are already many cross-language comparative studies of such phenomena—antonym systems (Wirth, 1982), kinship terminology and numeral systems (Greenberg, 1978), and basic color terms (Berlin and Kay, 1969)—that can form a basis for some kind of lexicosemantic typology.

HOLISTIC TYPOLOGY

The existence of implicational universals and the fairly lengthy implicational chains that are involved in some of them have stimulated many scholars to look for implicational universals that would tie together important features on all major levels of linguistic structure, from phonology to morphology and syntax. Of course, such a global, holistic approach, if feasible, would be the ultimate typological classification in terms of its economy and elegance. Unfortunately, since things usually turn out to be much more complicated and messy in "real life" than the elegant theories claim them to be, a careful scholar needs to exercise an even greater degree of skepticism when evaluating any new theory that appears to be too elegant and economical. Basically, it seems that various typological schemas that have been constructed on the basis of implicational universals suffer from the same malady, namely that many "universals" turn out to be merely universal tendencies rather than universals in the true sense of this term.

A good example of a holistic typological schema is found in Patricia Donegan and David Stampe's 1983 article. Simply stated, the authors' argument is that phrase and word accent are correlated with the basic word order, and determine also whether a language will be mostly prefixing or suffixing, whether it will have case-relation marking by means of affixation or not, whether it will tend to be tonal or not, and so on. The authors present quite interesting arguments in favor of their hypothesis that links all these seemingly disparate elements of language structure together into one coherent theory of language type. Much needs to be done, however, to validate the main claims of this theory as well as to make the necessary refinements.

SOCIOLINGUISTIC CLASSIFICATION

One may also classify languages according to their sociolinguistic roles, that is, what roles they play in the societies that use them. For example, some languages are now only used in written form or only in church services (liturgi-

cal languages such as Coptic or Latin, both of which were originally used for other purposes as well), some are used as second languages for prestige or for other special purposes. Languages have different degrees of prestige, official status, and different social uses to which they are put. What complicates matters in such a classification is the fact that one and the same language may have very different functions in different societies that use it. For example, Classical Arabic is mostly a liturgical language in non-Arab Muslim countries, whereas in Arab countries it (actually a modified version of it called Modern Literary Arabic) is used as a formal written language, in education, in formal discourse, and as an intermediary language to be used with Arabs from other countries whose colloquial language may be quite different from the local colloquial Arabic.

A preliminary attempt to formulate a sociolinguistic framework of language classification may be found in W. Stewart (1962). A brief summary of it is also given in Horne (1966:3–4). Kloss (1968) classifies entire nations in terms of the status of the languages spoken in them in a somewhat elaborate schema designed for "language planning" purposes.

Finally, we should mention the possibility of sociolinguistic/ethnolinguistic classification based on the ethnography of speaking or ethnography of communication studies. Simply stated, such studies concentrate on the complex interrelationship between language use in particular speech communities with various social, economic, ethnographic, and cultural factors in these speech communities. For example, the phenomenon of turn-taking in conversation may be studied crosslinguistically to determine what may be the universal and what are the speech-community particular phenomena involved. Another example is the series of studies concerning the dimension of volubility/taciturnity in different speech communities: Is volubility or taciturnity prized more in the community, and under what circumstances?

Ethnography of speaking is a field that was initiated by Dell Hymes in the 1960s because he felt that linguistics, anthropological linguistics, and sociolinguistics were all ignoring certain important language phenomena that were crucial for our understanding of how language is actually used in different speech communities. Since then there have been numerous studies in the field, many of which have focused on crosslinguistic cross-cultural aspects of the ethnography of speaking, thus laying the foundation for a sociolinguistic/ethnolinguistic language typology. (Only a small sample of the relevant literature will be cited in the bibliography at the end of this chapter. The works by Duranti and Philipsen and Carbaugh, however, may be consulted for references to other important work in this area.)

EXERCISES

1. Establishing Genetic Relationship

The data below contains lexical items from Finnish and Hungarian. Your task is to decide whether or not one can establish that these two languages are genetically related *based solely on the given data*. If you decide that it is not possible to do so, discuss the difficulties and your reasoning. If you decide

that it is possible to establish the genetic relationship in this case, cite the relevant evidence for your claim.

GLOSS	FINNISH	HUNGARIAN
1. child	lapsi	gyermek
2. church	kirkko	templom
3. six	kusi	hat
4. fire	tuli	tűz
5. head	pää	fő
6. three	kolme	három
7. cracker	keksi	keksz
8. water	vesi	víz
9. fish	kala	hal
10. sun	aurinko	nap
11. winter	talvi	tél
12. sugar	sokeri	czukor
13. tree	puu	fa
14. tongue	kieli	nyelv
15. eye	silmä	szem
16. blood	veri	vér
17. death	kuolema	halál
18. zero	nolla	nulla
19. son	poika	fiú
20. under	ala	allá
21. hear	kuulla	hall
22. heel	suoni	in
23. nose	nenä	orr
24. soap	saippua	szappan
25. soft	pehmeä	puha
26. ice	jää	jég
27. nest	pesä	fészek
28. bird	lintu	lud ('goose')
29. breast	süli	öl ('lap')
30. half	puoli	fél
31. egg	muna	mony
32. hand	käsi	kéz
33. heart	sydän	szív
34. two	kaksi	két
35. hundred	sata	szás
36. mouth	suu	száj
37. what	mitä	mi

Note: Forms are cited in the standard orthography for Finnish and Hungarian. In Finnish, [y] stands for the high front rounded vowel, and [ä] stands for a low front unrounded vowel. In Hungarian stress mark indicates vowel length; combination *sz* represents [s], whereas an *s* not followed by *z* is [š]; *cs* is [č], and *c* is [ts]. The letter *y* indicates that the preceding consonant is palatalized. In both languages, the letter *j* stands for the palatal semivowel.

II. Typology Exercise 1

Choose two different passages of English of about two hundred words each and perform a simple morphological analysis on them. One of the passages chosen should consist of a running conversation, whereas the other should be an example of a narrative in formal, literary style. Calculate the degree of synthesis and the compositional index for the two passages. If either of the indices is significantly different for the two passages, how do you account for the difference?

III. Typology Exercise 2

Examine the following data from four different languages and determine which morphological type is exemplified by each. Give a brief justification for your decision in each case.

Language A (based on Lillooet Salish)

tuxp-elíc'e-ʔen-č-eš

buy + clothing + transitive verb marker + first person singular object + third person singular transitive subject

'He bought me some clothes.'

Language B (Quechua)

'ika-y-ku-man-lya p"awa-sa-n-ku-ču

flower + first person + plural + to + also/ fly + progressive + present + plural + question

'Are they flying to my flowers also?'

Language C (Samoan)

ua alu le teine i le fale-ma'i [both *fale* and *ma'i* can be independent words]

perfective/ go/ the/ girl/ to/ the/ house + sick

'The girl has gone to the hospital.'

Language D (Nepali)

y-as khol-ā māthi sāgh-u thi-yo

that + oblique case/ river + oblique case/ across (postposition)/ bridge + non-oblique case, masculine ending/ past tense allomorph of *hunu* 'to be' + third person singular preterite indicative

'Across that river there was a bridge.'

SELECTED BIBLIOGRAPHY

Genetic Classification

Arlotto, Anthony. 1972. *Introduction to historical linguistics.* Boston: Houghton Mifflin. (A very readable elementary-level introduction to genetic classification, comparative method, and allied topics.)

Haas, Mary R. 1969. *The prehistory of languages.* Janua Linguarum, series minor no. 57. The Hague: Mouton. (A more technical work which emphasizes the comparative method as it is applied to unwritten languages, especially American Indian languages.)

Hock, Hans Heinrich. 1991. *Principles of historical linguistics.* 2d ed. Berlin: Mouton de Gruyter. (A very thorough and well-written advanced-level introduction to historical linguistics and comparative method.)

Lehmann, Winfred P. 1992. *Historical linguistics.* 3d ed. London: Routledge. (Elementary-level introduction to genetic classification, comparative method, and allied topics. One chapter is devoted to language typology.)

Typological Classification

Anderson, E. 1978. Lexical universals of body-part terminology. In J. H. Greenberg, ed., *Universals of human language.* vol. 3, 335–68. Stanford: Stanford University Press.

Berlin, Brent, and Paul Kay. 1969. *Basic color terms: Their universality and evolution.* Berkeley and Los Angeles: University of California Press.

Comrie, Bernard. 1989. *Language universals and linguistic typology: Syntax and morphology.* 2d ed. Chicago: University of Chicago Press. (An excellent overview of the more recent developments in language typology; very readable even though it is not an elementary introduction to the topic.)

Croft, William. 1990. *Typology and universals.* Cambridge: Cambridge University Press. (This work's emphasis and general orientation make it an excellent companion volume to Comrie's text.)

Donegan, Patricia Jane, and David Stampe. 1983. Rhythm and the holistic organization of language structure. In John F. Richardson et al., eds., *Papers from the parasession on the interplay of phonology, morphology and syntax,* 337–53. Chicago: Chicago Linguistic Society. (Somewhat difficult reading for those with little background in linguistics.)

Greenberg, Joseph. 1960. A quantitative approach to the morphological typology of language. *International Journal of American Linguistics* 26:178–94.

——. 1966. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph Greenberg, ed., *Universals of language,* 73–113. 2d ed. Cambridge: M.I.T. Press.

——. 1975. Research on language universals. *Annual Review of Anthropology* 4:75–94. (Explains the interconnection between language universals research and language typology.)

——. 1978. Generalizations about numeral systems. In J. H. Greenberg, ed., *Universals of human language,* vol. 3, 249–95. Stanford: Stanford University Press.

Hockett, Charles. 1955. *Manual of phonology.* Indiana University Publications in Anthropology and Linguistics, Memoir 2. Bloomington: Indiana University Press; rpt. Chicago: University of Chicago Press, 1979.

Horne, Kibbey M. 1966. *Language typology: 19th and 20th century views.* Washington, D.C.: Georgetown University Press. (A very condensed account of various language typology schemas, mostly based on morphology.)

Lehmann, Winfred P., ed. 1978. *Syntactic typology: Studies in the phenomenology of language*. Austin: University of Texas Press. (A collection of articles dealing primarily with word order typology and various implicational universals connected with the latter.)

Lyovin, Anatole. 1981. *Bibliography of linguistic typology*. *Working Papers in Linguistics* (Dept. of Linguistics, University of Hawaii at Manoa), 13(2):75–94. (Although not very accessible, this bibliography would be very useful for those who want to do further reading on the topic.)

Pullum, Geoffrey K. 1991. *The great Eskimo vocabulary hoax, and other irreverent essays in the study of language*. Chicago: University of Chicago Press.

Sapir, Edward. 1921. *Language*. New York: Harcourt, Brace and World. (Includes a lengthy section on language typology.)

Sociolinguistic Classification

Duranti, Alessandro. 1988. Ethnography of speaking: Towards a linguistics of the praxis. In Frederick J. Newmeyer, ed. *Linguistics: the Cambridge survey*; Vol. 4: *Language: The socio-cultural context*, 210–28. Cambridge: Cambridge University Press. (Contains a very good bibliography on ethnography of speaking as well as a good, short overview of the field.)

Gumperz, John J., and Dell Hymes, eds. 1972. *Directions in sociolinguistics: The ethnography of communication*. New York: Holt, Rinehart & Winston.

Hymes, Dell. 1974. *Foundations in sociolinguistics: An ethnographic approach*. Philadelphia: University of Pennsylvania Press. (The first two chapters are of special relevance. Chapter 2 illustrates how language communities differ from each other in their attitudes toward volubility and taciturnity.)

Kloss, Heinz. 1968. Notes concerning a language-nation typology. In Fishman, J. A., C. A. Ferguson, and J. Das Gupta, eds. *Language problems of developing nations*, 69–85 John Wiley.

Philipsen, G., and D. Carbaugh. 1986. A bibliography of fieldwork in the ethnography of speaking. *Language in Society* 15:387–97.

Saville-Troike, Muriel. 1989. *The ethnography of communication: An introduction*. 2d ed. Oxford: Basil Blackwell. (This is a standard introductory text in the field.)

Scollon, Ronald, and Suzanne B. K. Scollon. 1979. *Linguistic convergence: An ethnography of speaking at Fort Chipewyan, Alberta*. New York: Academic Press.

Stewart, William A. 1962. An outline of linguistic typology for describing multilingualism. In *Study of the role of second languages in Asia, Africa, and Latin America*, 15–25. Washington, D.C.: Center for Applied Linguistics of the Modern Language Association of America.

Chapter 2

Classification of Writing Systems

Writing is only a secondary aspect of language, that is, it is only a means of symbolizing spoken languages, often a very imperfect means at that. However, since we are more frequently confronted with unfamiliar and exotic looking foreign scripts than actual foreign languages as they are spoken, and since we shall encounter several exotic scripts in the language sketches, it will not do to omit discussing the various scripts of the world. Here the emphasis will be on the typology of the writing systems that are currently in use rather than on the historical development of writing and the ancient scripts.

Just like languages themselves, writing systems can be classified either genetically, according to their historical origin, or typologically, according to various criteria. In the case of writing systems, however, it is more revealing to discuss their typological classification before discussing their genetic classification.

TYPOLOGICAL CLASSIFICATION

Writing systems may differ from each other in many different ways, and therefore it is necessary to make a judicious choice of the aspects on which to base the typological criteria for classification. For example, one may choose to classify writing systems according to the inventory of shapes that they employ for their symbols. Although this turns out not to be a very enlightening classification, it does actually have some practical use—it may help us design