CHAPTER

1 What is language typology?

CHAPTER OUTLINE

KEY TERMS

distribution
crosslinguistic
    generalizations
genetic relations
language contact
linguistic typology
language types
existential
    statements
unrestricted
    universals
restricted
    (implicational)
    universals
absolute
    universals
statistical
    (probabilistic)
    universals
implicans
implicatum
language sample
data base

In this chapter, the goals of language-typological research will be defined as studying similarities and differences among languages that do not stem from shared genetic relationship, language contact, or shared environmental conditions. Some basic research tools will be introduced: language-typological generalizations of various kinds, ways of constructing language samples, and sources for obtaining language data.

1.1 Goals

You are riding in a crowded elevator; next to you stand two people conversing in a foreign language. You don't understand a word of what they are saying and couldn't even repeat any of it: their speech strikes you as just plain noise. Yet, the two people obviously communicate. One person says something whereupon the other breaks into a peal of laughter; he then responds and the first person comes back with another round of what sounds like complete gibberish. How can these odd noises make any sense to anybody?

What you have just experienced is a true fact: languages are different. The following examples further illustrate how different they can be. (2), (3), and (4) are Polish, Hungarian, and Turkish translations of the English sentence in (1).

(1)   *Give us today our daily bread.*   English

(2)   *Chleba   naszego   powszedniego   daj   nam   dzisiaj.*   Polish
      bread   our   daily   give   to.us   today

(3)   *Mindennapi   kenyerünket   add   meg   nekünk   ma.*   Hungarian
      daily   our.bread   give   PREF   to.us   today

(4)   *Gündelik   ekmeğimizi   bize   bogün   ver.*   Turkish
      daily   our.bread   to.us   today   give

While these examples and the "elevator-experience" suggest that languages are very different, languages also show surprising similarities. Look at the translations of sentence (1) in two additional languages.

(5)   *Unser   tägliches   Brot   gib   uns   heute.*   German
      our   daily   bread   give   us   today

(6)   *Vårt   dagliga   brōd   giv   oss   idag.*   Swedish
      our   daily   bread   give   us   today

Several of the words are similar in English, German, and Swedish:

(7)   ENGLISH   GERMAN   SWEDISH
      our   *unser*   *vårt*
      bread   *Brot*   *brōd*
      give   *gib*   *giv*
      us   *uns*   *oss*

The fact that languages are both different and similar is a puzzle. Two questions arise:

(8)   (a)   How are languages different from each other and how are they similar?
      (b)   What is the reason for their differences and for their similarities?

The first question addresses the **distribution** of structural properties among languages: what occurs and where? The second question in turn asks for an

explanation of the distributional facts: why does a structural property occur where it does? There is hardly any issue more central to the science of linguistics than these two; and they are also the focus of this book.

The data cited in (1)–(7) have begun to answer the question in (8a) by showing crosslinguistic similarities and differences in vocabulary, word structure, and word order. Let us now turn to (8b), which asks for the reasons for crosslinguistic similarities. The vocabulary resemblances among English, German, and Swedish illustrated above have a straightforward explanation. About twenty-five-hundred years ago, these languages did not exist separately; instead, there was a single ancestral language – linguists call it Proto-Germanic – from which all three subsequently derived. The vocabulary similarities are due to inheritance: ancestral words have survived in the daughter languages albeit their forms have been somewhat altered over the centuries. The gradualness with which the three languages have changed away from each other is shown by the Old English version of the same sentence.

(9)　*urne*　*daghwamlican*　*hlaf*　*syle*　*us*　*to*　*dæg*　Old English
　　　our　daily　　　　　bread　give　us　to　day

The table in (10) shows that some of the Old English words are more similar to their German and Swedish equivalents than the Modern English ones because Old English, spoken about a thousand years ago, was closer in time to Proto-Germanic – their shared mother language.

(10)　GERMAN　　SWEDISH　　　OLD ENGLISH　　MODERN ENGLISH
　　　unser　　　*vårt*　　　　*urne*　　　　　*our*
　　　tägliches　*dagliga*　　　*daghwamlican*　*daily*

Polish, Hungarian, and Turkish (illustrated in (2), (3), and (4) above) are not Germanic languages; hence the differences.

The above materials point at one reason for similarities among languages: **shared historical origin**. However, languages may resemble each other even if they are not known to have evolved from the same ancestor. Take the word *sugar*. As we might expect, in Germanic languages (English, German, Swedish, Dutch, and others) it has roughly the same form.

(11)　ENGLISH　　*sugar*
　　　GERMAN　　*Zucker*
　　　SWEDISH　　*socker*
　　　DUTCH　　　*suiker*

But the word has similar forms even in languages outside the Germanic family.

(12)　SPANISH:　　*azúcar*
　　　FRENCH:　　*sucre*
　　　ITALIAN:　　*zucchero*
　　　POLISH:　　*cukier*

A possible explanation may still be shared genetic origin: along with the Germanic family, the Romance languages (Spanish, French, and Italian)

and Polish, a Slavic language, are all members of the larger group of Indo-European. Thus, the word for 'sugar' could be a legacy of Proto-Indo-European, their shared mother language (spoken about 5000–4000 BCE).

If this were the case, we would expect languages outside Indo-European to have completely different words for 'sugar.' However, this is not so.

(13) HUNGARIAN: *cukor*
      TURKISH: *şeker*
      HEBREW: *sukkar*
      ARABIC: *soukar*
      JAPANESE: *satoo*
      SWAHILI: *sukari*
      INDONESIAN: *sakar*

These languages are not Indo-European: Hungarian is Finno-Ugric, Turkish belongs to the Turkic family, Hebrew and Arabic are Semitic, Japanese is an isolate, Swahili is Bantu, and Indonesian is Malayo-Polynesian. The extent to which some of these languages differ from English was illustrated in (3) on the example of Hungarian and in (4) for Turkish. Yet, the words for 'sugar' in these languages are still similar to the English word.

The explanation is again historical but of a different kind: not shared origin but **language contact**. The words for 'sugar' all come from Sanskrit *śarkar*ā. Sugarcane was first cultivated in India, the home of Sanskrit. In the eighth century CE, Indian merchants began to export sugarcane; crusaders then brought it to Europe and traders spread it around the globe along with the word itself, with spelling and pronunciation somewhat modified according to the conventions of each language.

So far, we have seen two sources of similarities among languages: **shared historical origin** and **contact**. Might there be other reasons? In pondering this question, let us consider resemblances among things outside language. Take people. If you compare your friends, you will find that some are more like each other than others. In some cases, this may be due to the fact that they are related. If they are siblings, they may have inherited certain features – such as black hair or musical ability – from their parents or from their more remote ancestors. Thus, just as in the case of languages, genetic relatedness is a possible explanation for resemblances among people.

Now suppose you know two unrelated individuals who share an interest in butterflies. A different kind of explanation is needed here: they may have been long-time friends and one of them came to be interested in butterflies when prompted by the other. Contact and the attendant spread of characteristics from one individual to another can explain similarities among people as it does between languages shown on the example for the word for 'sugar.'

But let us consider a third scenario. Suppose you have two friends who are not related nor have they ever met; yet, they are both devoted to ice-fishing. Chances are that both came from parts of the world where there are severe winters that cover lakes and rivers with thick ice. Their shared interest is likely to be related to the environment that they both come from. Could the **same environment** – natural or cultural – result in similarities also among languages as it does among people?

Features of the natural setting of a speech community are often reflected in the vocabulary of the language. Nicholas Evans reports that in Kayardild (an Australian aboriginal language) there are five different verbs to describe hopping, one for each subtype of macropods – an animal species specific to Australia that includes kangaroos, wallaroos, and wallabies (1998: 164). Socio-cultural setting can also have an effect on language: if these conditions are similar, so may be some aspects of the languages. An example is word forms differentiated by degrees of respect. Two of the many languages that have a broad range of vocabulary items whose use is determined by social considerations are Guugu Yimidhirr, an Australian aboriginal language of Queensland, and Japanese. In Japanese, several kinship terms have alternative forms depending on whether you speak to members of your own family or to people outside it. For example, 'grandfather' is *ojiisan* when talking to family and *sofu* when talking to outsiders; 'father' is *otoosan* when talking to family but *chichi* when talking to outsiders (Inoue 1979: 282). In Guugu Yimidhirr, some words have special forms for talking to one's brother-in-law or father-in-law as opposed to talking to others. For example, the word for 'to go' is *balil* in the respectful brother-in-law style; the everyday form is *dhadaa* (Haviland 1979: 217–218). Note that Japanese and Guugu Yimidhirr are neither genetically related nor have they been in direct contact. Instead, their socially-conditioned vocabulary distinctions correlate with the stratified societies where these word distinctions developed.

These culturally conditioned vocabulary distinctions are somewhat comparable to the difference between the French second-person pronouns *tu* and *vous*, where the former is used to address a close family member or friend while the latter is reserved for formal relations. Similar distinctions hold in Spanish (*tu* and *usted*) and German (*du* and *Sie*). That such usages respond to societal conditions can be seen most clearly when social structure changes. In some European countries, such as Austria and Germany, where the second person pronoun of the language has an intimate and a polite form, the former is gaining over the polite version, very likely in response to societal leveling.

The three factors of **genetic relatedness**, **language contact**, and **shared cultural environment** go a long way addressing the question in (8b) about why languages are similar. However, they do not work in all cases: two additional reasons need to be invoked: **types** and **universals**.

Consider the following sentences from Hindi, Japanese, and Turkish, all translations of 'They bring water for the girl's mother.' Square brackets set off phrases.

(14) [Ve]   [larkiki   make   liye]   [pani]   [late   hai].   Hindi
      they   girl's   mother   for   water   bring   are

(15) [karera   wa]   [ano   onnanoko   no   haha   ni]   [mizu   o]   [motte   kuru]   Japanese
      they   SUBJ   the   girl   GEN   mother   for   water   ACC   bring   give

(16) [*Kız* *annesi* *için]* *[su]* *[getiriler.]*   Turkish
   girl her.mother for water they.bring

Although the words in these sentences are very different from each other, note that they are placed in the same order. All three are so-called "SOV languages," which means that first comes the subject of the sentence (if there is one; it is not present in the Turkish sentence), then the indirect and direct object, and then the verb.

In addition, the three languages share two other order patterns given in (17b) and (17c). (The symbol & indicates linear order.)

(17) (a) Subject & Object & Verb
     ("they water bring")
   (b) Possessor & Possessum
     ("girl's mother")
   (c) Noun Phrase & Adposition
     ("mother for")

The identical orderings of sentence constituents in these three languages is not due either to shared origin, or contact, or shared environment. These languages are not genetically related: Hindi is Indo-European and, as noted above, Turkish is Turkic and Japanese is an isolate. They have not been in close contact, nor are their natural and cultural conditions particularly similar.

Strikingly, these languages contrast with others that have near-mirror-image orders for all three sets of constituents, as in (18).

(18) (a) Verb & Subject & Object
     ("bring they water")
   (b) Possessum & Possessor
     ("mother girl's")
   (c) Adposition & Noun Phrase
     ("for mother")

Here are examples from Arabic and Rapa-Nui (the language of Easter Island; data from Chapin 1978). Arabic is Semitic, Rapa-Nui is Malayo-Polynesian; they are both genetically and geographically separate and share little by way of socio-cultural conditions.

(19) (a) Verb & Subject & Object   Arabic
     axaδa *aşşinijjuna* *almala*
     took the.Chinese the.money
     'The Chinese took the money.'
   (b) Possessum & Possessor
     bajtu *arraʒuli*
     house man.GEN
     'the house of the man'
   (c) Adposition & Noun Phrase
     ila bosʈon
     to Boston
     'to Boston'

(20) (a) Verb & Subject & Object   Rapa Nui
          He   to'o   te   tenitō   i   te   moni.
          PAST take the Chinese ACC the money
          'The Chinese took the money.'
     (b) Possessum & Possessor
          te   hoi   o   te   tagata
          the horse GEN the man
          'the horse of the man'
     (c) Adposition & Noun Phrase
          ki Boston
          'to Boston'

Although the correlations between the alternative positions of the verb and the other two pairs of constituents are only a tendency across languages, most SOV languages do place the possessor before the possessum and use postpositions, and most verb-initial languages put the possessor after the possessum and have prepositions.

What might be the reason for genetically unrelated SOV languages in various parts of the world sharing these order patterns? And similarly, why do unrelated VSO languages scattered around the globe tend to have the opposite orders? An obvious idea is that the constituents whose ordering patterns form consistent clusters have something in common. If so, the otherwise puzzling clustering of THREE different orders would be reduced to just ONE pattern: the particular orders would fall out of a single regularity.

Several explanations along these lines have been proposed in the literature; they will be discussed in Chapter 7 (Section 7.3). One hypothesis is that the regularity has to do with the uniform ordering of heads and dependents (Vennemann 1973). What is meant by the "head" of a construction is the indispensable part and the one that determines the category of the entire construction. The "dependent" in turn is of a different category than the entire construction and it is often optional. The sorting of verbs, objects, possessums, possessors, adpositions, and noun phrase complements into the two categories of head and dependent is given in (21).

(21) HEAD   DEPENDENT
      Verb   Object
      Possessum   Possessor
      Adposition   Noun Phrase

In other words, the common denominator of verbs, possessums, and adpositions is that they are all heads, with object, possessor, and noun phrase complement being their respective dependents.

According to this theory, languages tend to adopt a single order rule for heads and dependents from which the ordering of verb and object, possessum and possessor, and adposition and noun phrase automatically follows. Languages thus belong to two different **types**: SOV languages (e.g. Hindi, Turkish, and Japanese), which adopt dependent & head order and therefore have "water bring," "John's book," and "Boston in"; and

verb-initial languages (e.g. Arabic, Rapa-Nui), which opt for head & dependent order ("bring water," "book John's," "in Boston").

By way of a re-cap, the four types of explanations of crosslinguistic similarities discussed above are schematized in (22).

(22) Explaining crosslinguistic similarities...
    (a)   ... by shared inheritance
        QUESTION: Why do English and German have similar words for 'bread'?
        ANSWER: Because both English and German are Germanic languages and they inherited this word from Proto-Germanic, their shared ancestral language.
    (b)   ... by language contact
        QUESTION: Why do English and Swahili have similar words for 'sugar'?
        ANSWER: Because both languages adopted the Sanskrit word through contact.
    (c)   ... by shared environmental conditions
        QUESTION: Why do Japanese and Guugu Yimidhirr have alternative words where the choice between them depends on the social relationship between speaker and addressee?
        ANSWER: Because in both languages, these distinctions evolved in response to the demands of stratified societies.
    (d)   ... by reference to language types
        QUESTION: Why do both Hindi and Japanese place the possessor before the possessum?
        ANSWER: Because possessors are dependents and possessums are heads and both languages are of the dependent & head type.

These four types of explanations may be invoked in case we want to explain that **some** languages are similar to each other as opposed to others. But what about similarities that hold for **all languages**? As an example, consider the fact that all known languages have personal pronouns, such as *I*, *you*, and so forth. Let's try to apply the answer types discussed above to this fact.

(23) QUESTION: Why do all known languages have personal pronouns?
        ANSWER:
    (a)   Because all languages are genetically related and the ancestral language from which they derived had personal pronouns.
    (b)   Because all languages have been in direct or indirect contact with each other and the presence of personal pronouns has spread from one to the other.
    (c)   Because all languages are spoken in the same cultural conditions that call for personal pronouns.
    (d)   Because all languages belong to the same language type.

Let us evaluate these possible explanations. The first one posits a single source for all human languages: if this ancient language had personal

pronouns, all of its daughter languages could have retained it. This is not an impossible hypothesis but it still leaves two questions open: why the source language had personal pronouns and why personal pronouns have consistently survived in the course of millennia when so many other structural properties have changed. The second hypothesis may also be correct: perhaps personal pronouns first evolved in one language and the idea then spread to all other languages; but the questions of ultimate origin and universal survival still remain open. The third hypothesis posits similar cultural conditions for all languages. In a sense, it is true that all languages share some of their environment: for example, they are all spoken in a human community. But why would this fact require personal pronouns?

The fourth explanation is correct by definition: to the extent that all languages have personal pronouns, we must say that they form a single type. But here we are not talking about a (sub)-type of languages but about all languages being a (sub)-type of communication systems. This yields a fifth kind of answer to why languages are similar.

(24)   QUESTION: Why do all known languages have personal pronouns?
        ANSWER: Because all languages belong to a type of communication
        systems where the presence of personal pronouns is required.

However, (24) simply states the fact rather than explaining it. The ultimate explanation must have to do with some or perhaps all of the three factors considered above: the origins of human languages, their contact with each other and their shared natural and social conditions, as well as with the universally manifested function of personal pronouns in thought and expression. The presence of personal pronouns in a language does not appear necessary: names or demonstrative pronouns (such as 'this' and 'that') could do just as well to identify 'me,' 'you,' and others. However, a name may be shared by several individuals and demonstratives like 'this' and 'that' also do not point at speaker and hearer as clearly as 'I' and 'you' do. Thus, preference for clarity in identifying speaker and hearer provides only a probabilistic explanation for the universal genesis and survival of personal pronouns across languages. More will be said about personal pronouns in Section 2.3.2. of Chapter 2.

Let us summarize the above discussion. First, two central questions of linguistics were raised: the crosslinguistic distribution of structural properties across languages and the reasons for their distribution. In contemplating possible reasons for similarities among languages, we explored five kinds of explanations: shared inheritance, contact, shared environment, language types, and **language universals**.

The field of linguistic research called **language typology** is the study of the latter two phenomena: typologically and universally shared features of languages. It focuses on the concept of a language type. The term "**type**" in everyday usage is synonymous with "kind": it refers to a subclass of a class of entities. In this broad sense, two languages belong to the same type if they have at least one characteristic in common regardless of whether this shared characteristic is due to shared inheritance or borrowing or similar

environmental conditions. In actual linguistic usage, however, two languages are generally said to belong to the same type if their similarities hold across various genetic, areal, and cultural groups.

Before we begin to study language-typological implications and language universals in detail, we need to identify the conceptual tools needed for this study.

1.2 Tools

1.2.1 Statement types

In the previous section, we laid out the task of language typology: it is to find similarities among languages that are independent of genetic origin, areal influence, and shared environmental conditions. How do we capture the results of this investigation?

Let us begin by looking at the speech sound inventories of languages. We find the following:

(25) (a) Some languages have oral stops (e.g. /t/).
      (b) Some languages have alveolar nasals (/n/).

These statements simply declare the existence of languages that have such sounds. However, such existential statements do not provide us with distributional information: they do not tell us which languages have oral stops and which languages have /n/. They say that such sounds are possible in human languages since if at least one language has them, they must of course be possible. Thus, if we encounter a new language, what we know is that it may or may not have oral stops and alveolar nasals but we do not know if that particular language does or does not have them.

How could we turn these existential statements into distributional ones? Here is one attempt:

(26) (a) All languages have oral stops.
      (b) All languages have alveolar nasals.

These statements would be very useful: they define the set of human languages that have a particular property – namely, all languages. If we encounter a new language, these statements make predictions regarding its consonant inventory: that it will include oral stops and alveolar nasals.

But are these predictions true? As it turns out, (26a) is true: all languages known to us have oral stops. However, (26b) is untrue: indigenous languages spoken in the North-West area of the North-American continent – for example, Tlingit, a language of Alaska – have no alveolar nasals. Here is our problem: (25b), which states that some languages have alveolar nasals, is true but not predictive; (26b) is predictive but not true. Could we somehow combine the valuable universal scope of (26b) with the truth of (25b)?

The problem and its solution can be easily illustrated from everyday life. Suppose you are in a foreign city trying to learn the opening hours of food stores. Here is what you find: