CHAPTER

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

(3) Mindennapi kenyerünket add meg nekünk ma. Hungarian daily our.bread give PREF to.us today (4) Gündelik ekmeğimizi bize bogün Turkish ver. 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 our	tägliches daily	Brot bread	gib give	uns us	heute. today	German
(6)	Vårt our	dagliga daily	<i>brōd</i> bread	giv give	oss us	idag. today	Swedish

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.

ENGLISH	sugar
GERMAN	Zucker
SWEDISH	socker
DUTCH	suiker
	ENGLISH GERMAN SWEDISH DUTCH

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 *sarkarā*. 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] they	[larkiki girl's	mak mot	ke ther	liye] for	par wat	ij [er 1	llate bring	ç	hai]. are	Hin	di		
(15)	[karera	wa]	[ano	onna	noko	no	haha	n	ıi]	[mizu	o]	[motte	kuru]	Japanese
	they	SUBJ	the	girl		GEN	moth	ner f	or	water	ACC	bring	give	

(16)	$[K\iota z$	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-mirrorimage 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 axada aşşinijjuna almala took the.Chinese the.money 'The Chinese took the money.'
 (b) Possessum & Possessor bajtu arrazuli house man.GEN 'the house of the man'
 - (c) Adposition & Noun Phrase
 ila boston to Boston
 'to Boston'

Arabic

- (20) (a) Verb & Subject & Object
 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

Rapa Nui

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 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: (27) Some food stores are open 7 days a week.

Well, but which are those stores? Your first guess may be this:

(28) All food stores are open 7 days a week.

It then turns out that this is not so: some stores close on Sunday. In order to find out which are the seven-day stores, you will naturally look for a common characteristic of these businesses that distinguishes them from the others. Here is what you may discover:

(29) All food stores that carry fresh produce are open 7 days a week.

By logical structure, (29) is like (28): it is a universal statement (since it includes the quantifier *all*) but it has an advantage over (28): it is correct. Both (28) and (29) are universal generalizations for sets of stores but they differ in how the sets are defined. (28) says something about the universe of all food stores in town; (29) says something about a sub-universe of the town's food stores: those that share the common denominator of carrying fresh produce. (28) is an unconditional, or unrestricted, universal; (29) is a conditional, or restricted, universal.

In this example, once the unrestricted universal (28) turned out to be incorrect for the entire domain of food stores, the solution was to find a characteristic that defined a subdomain of food stores for which the universal generalization held. Let us apply the same idea to solving the problem of the distribution of alveolar nasals. Given that we want to determine the universe of languages within which all languages have alveolar nasals but we know that this universe does not include all languages, we need to identify a characteristic that carves out the proper subdomain of languages within which the universal holds.

As it turns out, this characteristic is the presence of labial nasals. Thus, the following restricted universal holds true:

(30) All languages **that have labial nasals** (/**m**/) also have alveolar nasals (/**n**/).

What the above examples show is that crosslinguistic generalizations may be existential or universal; and if they are universal, they may be unrestricted or restricted. Restricted universals are also called conditional or implicational; unrestricted ones are unconditional. These alternative terms will be used interchangeably throughout the book. Here are the schemata for these statement types:

(31) (a) Existential statements:

Some languages have X.

- (b) Universal statements:
 - (aa) unrestricted universals: All languages have X.
 - (bb) implicational universals: All languages that have Y also have X.

The typological clusters and universals mentioned in Section 1.1 can now be re-cast into these schemata.

- (32) Unrestricted universal: All languages have personal pronouns.
- (33) Implicational universals:
 - (a) Most languages that have SOV order also have Possessor& Possessum and Noun Phrase & Adposition orders.
 - (b) Most languages that have verb-initial order also have Possessum & Possessor and Adposition & Noun Phrase order.

(33a) and (33b) may be collapsed into the unrestricted universal in (34).

(34) Unrestricted universal: Most languages order heads and dependents uniformly.

What is the predictive force of these statement types? Unrestricted universals are about two logically possible languages types: languages that have X and those that do not have X; and the claim is that only one of the two exists. This is shown in (35). The star in front of a type means it is claimed not to occur; X is a structural feature; + and – stand for the presence and absence of a feature.

(35) Unrestricted universals

X Type I + *Type II –

Implicational universals in turn make a claim about four logically possible language types:

(36) Implicational universals

Y X Type I – – Type II + + Type III – + *Type IV + –

Tables like (36) that compare the logically possible co-occurrence patterns of two properties with their actual occurrences are referred to in the literature as **tetrachoric** tables (from Greek *tettares* That is, languages that have both characteristics (Type I) or that have neither (Type II) are both predicted to occur. However, of languages that have only one of the two features, only one type is said to occur (Type III); the other (Type IV) does not. In other words, the presence of Y is said to imply the presence of X: Y cannot occur without X.

Both restricted and unrestricted universals map out the logically possible distribution patterns of a structural characteristic: two for unrestricted universals – languages having the characteristic and those not having it – and four for implicational ones: languages having both characteristics, or neither, or one or the other. The actually observed patterns are then compared with the logical possibilities. As Frans Plank has remarked, "typology confronts possibility with reality" (Plank 1999: 285). If there is a gap – something that is logically possible does not actually occur – this is a highly valued finding because it calls for an explanation. If it seems something COULD occur, why does it NOT occur? More will be said about this at the end of the last chapter of this book.

What we have hit upon here is a powerful conceptual tool: distributional statements of a universal kind, either unrestricted or implicational. The applicability of such statements is not specific to studying the distribution of structural characteristics across languages: unrestricted and implicational universals can be usefully stated for any other domain of the world's phenomena as well. (28) and (29) showed how we use these tools in everyday life, such as in discovering opening hours of stores. These statement types are also basic staples in science. Here are some zoological universals.

(37) (a) An unrestricted universal

All animals have reproductive systems.

(b) An implicational universal All animals that have feathers are bipedal.

The respective charts are given in (38) and (39).

(38) All animals have reproductive systems.

	reproductive systems	EXAMPLES
Type I	+	all animals
*Type II	-	0

(39) All animals that have feathers are bipedal.

	having feathers	being bipedal	EXAMPLES
Type I	-	—	dogs
Type II	+	+	birds
Type III	-	+	humans
*Type IV	+	—	0

So far we have seen two basic divisions among crosslinguistic statements: whether they are **existential or universal**; and if the latter, whether they are **unrestricted or implicational**. There are three more variables that differentiate universal statements. One is their **modality**: whether they are absolute or statistical. Another has to do with **the relationship between X and Y**, called the implicans and the implicatum. And, thirdly, statements differ in whether X and Y are **simple or complex**. Let us look at examples.

An example of the varying **modality** of universals comes from the crosslinguistic distribution of nasal consonants as opposed to oral ones. Based on familiarity with English, German, Spanish, and so forth, we may state (40).

(40) Some languages have nasal consonants.

(40) is true; but it is not a universal statement and thus makes no firm prediction. Let's try an unrestricted universal.

(41) All languages have nasal consonants.

(41) makes a prediction but it is not true: languages in the NW areas of the North-American continent lack not only alveolar nasals (as discussed above) but they lack nasals in general. So let's try an implicational universal.

(42) All languages that have Y have nasal consonants.

Existential statements tell us about possibilities: the next language that you look at MAY or MAY NOT have a particular feature. Statistical statements are about probabilities: it is not only possible that the next language that you encounter has a particular feature it is said to be PROBABLE. Absolute universals in turn are about certainties: they say that the next language you look at WILL have a particular feature. This is a promising approach; but the problem is that no Y-feature has yet been found: that is, no property has been spotted that would differentiate languages with nasals and those without them. Do we therefore have to be content with the existential statement in (40)? No; there is a better option: we can state (43).

(43) Most languages have nasal consonants.

By logical structure, (43) falls between an existential statement and a universal one: its scope does not include all languages but it says more than just the existential claim that some languages have them. It is a **statistical** statement. It does not make a sure-fire prediction about whether the next language that you look at does or does not have nasal consonants, but it makes it more likely that it will have them. The statements formulated about word order regularities in (33) and (34) are similarly of the statistical rather than absolute sort.

Next, let's turn to another factor by which crosslinguistic generalizations may differ: **the relationship between implicans and implicatum** of implicational statements. Here are three statements from outside language that differ in this respect.

- (44) (a) All parts of the world that have bees also have flowering plants.
 - (b) All animals that have feathers have two legs.
 - (c) All animal limbs that belong to primates are articulated.

(44a) relates two different objects – bees and flowering plants – that occur **next to** each other: the presence of one calls for the presence of the other. (44b) also relates two objects but they are **parts** of the **same** entity: an animal body. And (44c) relates **two properties** of the **same animal body part**: primate limbs are articulated. We will label the three types **paradigmatic**, **syntagmatic**, and **reflexive** universals (the last ones also called provisions).

Here are corresponding examples from language.

- (45) (a) A paradigmatic universal In all languages in which the inflected verb precedes the subject in yes/no-questions, it does so in *wh*-questions as well. (Greenberg 1966a: #11a)
 - (b) A syntagmatic universal In all languages in which the inflected verb precedes the subject in wh-questions, the wh-word is normally initial. (Greenberg 1966a: #11b)
 - (c) A reflexive universal In all languages in which yes-no questions are differentiated from declaratives by an intonation pattern, the position of this pattern is reckoned from the end of the sentence rather than from the beginning. (Greenberg 1966a: 110, #8)

These generalizations, just as those in (44), differ in how their implicans and implicatum are related to each other. In (45a), the claim is about the

construction repertoire of languages: verb-before-subject order in yes/noquestions predicts the presence of the same pattern in another construction: wh-questions. Thus, implicans and implicatum are properties of **different constructions**: wh-questions and yes -no-questions. In (45b), this is not so: implicans and implicatum are both properties of a **single construction**: wh-questions. (45c) also applies within a sentence but implicans and implicatum are not distinct constituents of a construction; instead, they are properties of a **single constituent**. The statement refers to one feature of a constituent – an intonation pattern – and adds a detail about that feature: its position.

The diagrams in (46) show these three kinds of implicational universals. The rectangles are constructions of a language; Y and X are implicans and implicatum; arrows highlight the direction of prediction.

(46) (a) paradigmatic implication:

If Y, then X (where Y and X are features of different constructions of the same language).



(b) syntagmatic implication:

If Y, then X (where Y and X are features of different constituents co-occurring in the same construction).

(c) reflexive implication: If Y, then X (where Y and X are features of the same constituent).

$$\begin{bmatrix} 1 & \dots & [Y] & \dots & | & | & | & | \\ 1 & \dots & [X] & | & | & \dots & | & | \\ \end{bmatrix}$$

We now turn to the last kind of division among implicational universals. Consider (47).

(47) In all languages that have prepositions and where the demonstrative follows the noun, the adjective also follows the noun. (Hawkins 1983: 71)

At first blush, the structure of this statement seems to deviate from a normal implication since it mentions not two but three structural characteristics: prepositions, noun-before-demonstrative order, and noun-before-adjective order. But notice that it still has the two basic terms: implicans and implicatum; it is just that the implicans consists of two conditions rather than one. (47) has a **complex** implicans: more than one condition needs to be met before a prediction results.

Similarly, implicata may also be complex, as in (48).

(48) In most languages where the adjective precedes the noun, both the demonstrative and the numeral also precede the noun. (Greenberg 1966a: #18)

(48) is like a "two for the price of one" deal: from a single condition – Adjective & Noun – two predictions fall out: Demonstrative & Noun and Numeral & Noun. Needless to say, this pattern is preferable over (47), which yields "one for the price of two." The statistical universals about constituent order stated in (33) were also of the more useful kind: the position of the verb in the sentence predicts both the order of Possessor and Possessum and the order of Adposition and Noun Phrase.

Let us recapitulate the above survey of the typology of crosslinguistic statements.

- (49) (A) Crosslinguistic statements may be
 - existential statements:
 - In some languages, there is X.
 - universal statements: In all languages, there is X.
 - (B) Universal statements differ
 - (a) in the universe they pertain to
 - unrestricted universals:
 - In all languages, there is X.
 - implicational universals:
 In all languages where there is Y, there is also X.
 - (b) in their modality
 - absolute universals:
 - In all languages, there is X.
 - statistical universals:
 - In most (or in 60% etc. of) languages, there is X.
 - (C) Implicational universals differ
 - (a) in the relationship between their terms
 - paradigmatic implications: In all languages, if there is Y, there is also X, where Y and X are different constructions.
 - syntagmatic implications:
 In all languages, if there is Y, there is also X,
 where Y and X are parts of the same construction.
 - reflexive implications:
 In all languages, if there is Y, there is also X,
 where Y and X are features of the same constituent
 within a construction.
 - (b) in the complexity of their terms
 - single implicans and/or implicatum:
 In all languages where there is Y, there is also X.
 - complex implicans and/or implicatum
 In all languages where there is Y (and/or W), there is also X (and/or Z).

There is one more task to attend to before we are ready to embark on the study of actual crosslinguistic generalizations that have been proposed in the literature: we should take a closer look at the terms mentioned in our statements. Consider the shared schema of unrestricted and implicational universals:

(50) In all (or most) languages (where there is Y), there is also X.

First, what exactly do we mean by "all languages"? And, second, how do we obtain information about the grammatical properties X and Y? The next two subsections will take up these questions in turn.

1.2.2 Language samples

In formulating language-universal statements, our goal is to find generalizations that hold for all human languages (or for most of them; or for (most of) a well-defined subset of them). Thus, our domain of inquiry includes not only languages that exist at the present but also those that existed in the past but have died out or changed into a different language, and even those that will evolve in the future. But clearly, we cannot possibly inspect all these languages. First and most obviously, we cannot know what future languages will be like. Second, we cannot know about all the languages that have ever existed in human history. By Daniel Nettle's estimate, human language first evolved at least 50,000 years ago and possibly before that. Since that time, about 233,000 languages have simply disappeared. Adding to this figure the 7,000 languages that are now in the world, the total number is about 240,000, of which today's languages form only about 3%.

So what about this 3%? Here is the third reason why we cannot inspect all languages: of the roughly 7,000 languages spoken today, we have descriptions for only about a third of them (Bakker 2011: 101–102). Large areas of Australia, South America, and other parts of the world are still *terra incognita* from a linguistic point of view. However, there are many doctoral dissertations and other monographs aiming at closing the gap between known and unknown. Three of the major institutions that produce new knowledge about hitherto unknown or insufficiently described languages are the Research Centre for Language Typology at La Trobe University in Melbourne, Australia, the Language and Culture Research Centre at the Cairns Institute of James Cook University also in Australia, and the Summer Institute of Linguistics International (SIL). At the two Australian institutions, descriptive work focuses on – but is by no means restricted to – the Aboriginal languages of Australia.

So far we have seen that claims about "all languages" cannot actually be assessed due to lack of information about future languages, about all past languages, and about all languages that exist today. This means that our largest available data base consists of some past languages and those present-day languages that have been described. In addition, there are also more practical considerations in the way of testing universal claims: even with today's extensive data bases, no linguist can consult all descriptions that are available. If we tried, we would have to deal with thousands of languages and the amount of time and effort involved would be enormous. The first extant written documentation of a language is Sumerian tablets dating from around 3,000 BCE; we have no direct information about earlier languages. Nor do we have detailed information about some languages that were spoken in the more recent past, such as Etruscan (inscriptions dated from about 700 BCE).

SIL was founded in 1934. The goals of this faith-based but nondenominational organization include both language development in the speaker communities and linguistic research. Its workers currently over 5,500 from about 60 countries have produced linguistic accounts of 2,700 or more languages of the world.

To make the task more manageable, language typologists work with selected **samples** of languages. The question is: what languages should be chosen as part of a language sample?

The principles that guide sampling directly follow from the goals of language typology: we are interested in identifying similarities and differences among languages that are independent of genetic origin, language contact, and environment. Thus, a proper sample must be representative of all language families, all geographic areas, and all cultures.

A thoughtfully designed and widely influential sampling technique is one proposed by Matthew Dryer (1989). Dryer's focus is on establishing universal tendencies – that is, statistical universals – rather than absolute universals. His question is: under what conditions can the crosslinguistic distribution of a structural pattern be declared a universally valid tendency?

The first step is to insure genetic balance.

(a) GENETIC GROUPS

All languages are assigned to one of 322 groups called genera. Each genus contains related languages that can be traced back to an ancestor about 2,500 years ago. For example, Romance languages (Latin, Spanish, French, Italian, and others) form a genus and so do Germanic languages (English, German, Dutch, Swedish, and so forth).

The second step is to make sure the emerging sample is areally balanced.

(b) GEOGRAPHIC GROUPS

Each genus is assigned to one of five continent-size areas of the world: Africa, Eurasia, Australia & New Guinea, North America, and South America. For example, Romance and Germanic languages are part of the Eurasian area and so are some non-Indo-European genera such as Finno-Ugric (Finnish, Estonian, Hungarian, and so forth).

How is it then decided whether a particular structural feature's crosslinguistic distribution indicates a statistically significant tendency?

(c) MEASURE OF UNIVERSAL TENDENCIES

Given a particular structural feature – say front-rounded vowels, e.g. French / \ddot{u} / in *tu* 'you' – languages in the various genera are checked for the presence versus absence of this property. The pattern is said to represent a significant universal tendency if in all five geographic areas, the majority of the genera exhibit that feature.

If a genus has both languages that have that feature and languages that do not, the genus is divided into two sub-genera each counted separately as a genus.

Here is one of Dryer's examples. The question is whether SOV order (as in Turkish, Hindi, and so forth) is a crosslinguistically significant tendency over SVO order (as in English). The table in (51) presents his results (269–270). The five areas are listed on the top; the numbers are of the genera in each area that have SOV and SVO order. The boxes highlight the majority figures.

(51)		Afr	Eura	Austr-NG	NorthAm	SouthAm	Total
	SOV	22	26	19	26	18	111
	SVO	21	19	6	6	5	57

Since SOV order is exhibited in more genera in every area than SVO order, SOV emerges as a significant crosslinguistic tendency.

This is in contrast with the distribution of SVO (English) and VSO (Rapa Nui, Arabic) orders. Here are the results (270–271).

(52)		Afr	Eura	Austr-NG	NorthAm	SouthAm	Total
	SVO	21	19	6	6	5	57
	VSO	5	3	0	12	2	22

In this case, SVO may be viewed as a trend but, because in North America, VSO genera are more numerous than SVO genera, SVO fails to reach the level of a significant crosslinguistic tendency.

Given that our knowledge of the entire set of human languages, past, present, and future, is only partial and unavoidably so, our universal statements are mere hypotheses whose validity can never be proven. This holds regardless of whether the statement is absolute or statistical: the next language may be a counterexample to an absolute statement or may change the probabilities of a statistical one. They must be viewed as bestpossible guesses. They involve extrapolations from what is KNOWN about SOME languages onto what ALL languages MIGHT be like.

1.2.3 Data sources

Here is again the general schema of language-typological statements:

(53) (repeated from (50))In all (or most) languages (where there is Y), there is also X.

In the preceding section, we probed into the meaning of the phrase "all (or most) languages." The other fundamental terms of this statement type are X and Y – the structural properties of languages whose distribution is at issue. The question is: given a language, what are sources of information for X and Y?

The primary data for general linguistic research come from oral language. Ideally, typological studies, too, should be based on live data gathered orally from speakers. In actuality, this is barely feasible given the extensive crosslinguistic samples that typological work requires. Another, more doable way of obtaining primary data is by relying on written information, such as questionnaires that request translations of relevant materials into various languages.

Most commonly, however, data for typological work are derived from secondary sources: published grammars, dictionaries, journal articles, and the like. In addition, information can also be obtained from the data bases available on the Internet. The website of the Association for Linguistic Typology (www.linguistic.typology.org) lists a number of relevant data bases. One of these is the World Atlas of Language Structures (www.wals.info), which provides articles and maps for the areal distribution of 142 grammatical features, such as relative clauses or consonant inventories. Another valuable source is the website of the Surrey Morphology Group (www.surrey.ac.uk), which offers crosslinguistic data and analyses on a number of morphosyntactic patterns, such as syncretism and agreement. The Typological Database System, whose home is the University of Amsterdam, is a collection of independently developed typological data bases (www.hum.uva.nl/TDS).

These sources are relevant for **discovering** new language-typological generalizations. Another endeavor crucial to the advancement of our knowledge about the distribution of grammatical properties is **testing** generalizations that have already been proposed in the literature. A prime source for such generalizations is The Universals Archive based in Konstanz, Germany (http://typo.uni-konstanz.de/archive). At the time of this writing (Summer 2012), it lists 2029 crosslinguistic generalizations gleaned from the typological literature. A separate branch of this website is an inventory of rare grammatical characteristics (to date, 147 are listed).

A printout of a crosslinguistic generalization in the Language Universals Archive is given in (54) (Figure 1.1). It has to do with the crosslinguistic distribution of certain body-part terms. The first line gives the serial number of the statement in the archive. Next comes the verbatim quote from the source followed by a restatement that makes the implicational structure explicit. The rest of the entries are self-explanatory; "achronic" means the generalization pertains to synchronic stages of present, past, and future languages. References are to the bibliographic list given on the website.

(54)	Number	1180 (used to be 1184 in the old version)				
	Original	If in a given language there is a separate term for				
		'leg' (as opposed to 'foot') then there is also a term				
		for 'arm' (as opposed to 'hand').				
	Standardized	If there is a separate term for 'leg' (as opposed to 'foot'),				
		THEN there is a term for 'arm' (as opposed to 'hand')				
	Formula	'leg' \Rightarrow 'arm'				
	Keywords	body parts				
	Domain	Lexicon				
	Туре	Implication				
	Status	Achronic				
	Quality	Absolute				
	Basis	41 languages in Brown 1976 (12 American Indian				
		languages, 10 European, 5 sub-Saharan African,				
		Mideastern and Western Asian, 5 Southeast Asian,				
		2 Chinese, 2 Micronesian)				
	Source	Brown 1976, also mentioned in Anderson 1978: 352				
	Counterexamples –					
	Comments	By Frans Plank 03.08.2006, 09:49				
		A term for 'leg' is present in most but not all language.				
		According to Andersen's data, at least three languages,				
		Hopi (Uto-Aztecan), Inupik (Eskimo-Aleut), and				
		Tarascan (Chibchan), do not have 'leg' though they do				
		have terms for subparts (e.g. 'thigh' \Rightarrow 'calf').				

Figure 1.1 The distribution of terms for 'leg' and 'arm'.

Summary

This chapter presented the goals and tools of language typological research. The goals involve establishing the distribution of grammatical properties across genetically, areally, and culturally independent languages. The means whereby the results of this endeavor are captured are crosslinguistic statements of various kinds.

Existential statements tell us what can occur in languages; universals tell us what occurs under what conditions. Universals are unrestricted if they have all languages in their scope; they are implicational if they have a well-defined subset of languages in their scope. Implicational statements may vary in terms of the relationship between implicans and implicatum and in whether their terms are simple or complex. All of these statements may also differ in their modality: statistical statements hypothesize what is probable while absolute statements are hypotheses about what may be certain. For formulating crosslinguistic hypotheses, we work with genetically, areally and, as much as possible, culturally balanced language samples. Language data come mostly from published sources and from data bases.

In everyday discourse, we often think we know a lot of things – or at least we speak as if we think we do. When we stereotype people or institutions, we presume to know all from a few. We tend to think that we know the future based on past experience, that we know the causes of things including people's intentions behind their acts, and how things would have been if they had been different. In actuality, such inferences are mere possibilities or probabilities rather than certainties. In scientific discourse, researchers are more careful about distinguishing knowledge states: they form hypotheses rather than declare things as certain and, instead of sweeping generalizations, they describe what is possible (since it has occurred) and quantify probabilities as much as they can. These issues will be taken up again in the closing section of Chapter 7.

Activities

- 1. Look up the word for 'salt' in dictionaries of different languages. Are there any similarities? If so, what might be the reason?
- 2. Universals both unrestricted and restricted can also be stated for the distribution of structural characteristics within languages. An unrestricted universal for English words is that they all contain at least one vowel. But now consider the following: "All consonant-initial words of English start with /s/." This is clearly untrue: there are thousands of words like *table* or *paper* that do not start with /s/.

Try to formulate a restricted universal of the following type: "All English words that have characteristic X start with /s/." For identifying X, consider words like *string*, *sprain*, *splint*, and others; also non-existing words like **ptring*, **tprain*, **kplint*, and others.

- Consider the crosslinguistic generalizations in (a), (b), (c), and (d).
 A. Determine for each whether it is an unrestricted or an
 - implicational universal and whether it is absolute or statistical.B. Determine each statement's predictive force for English by
 - choosing one of the following answers:
 - (i) This statement makes a correct prediction about English.
 - (ii) This statement makes an incorrect prediction about English.
 - (iii) This statement makes no prediction about English.
 - Here are the statements:
 - (a) In most languages where the adjective precedes the noun, both the demonstrative and the numeral also precede the noun. (Greenberg 1966a, #18; cited above in (48))
 - (b) In all languages in which the inflected verb precedes the subject in wh-questions, the wh-word is normally initial. (Greenberg 1966a: #11b; cited above in (45b))
 - (c) Whenever the verb agrees with the subject or the object in gender, it also agrees in number. (Greenberg 1966a, #32)
 - (d) In most languages, interdentals are fricatives.
- 4. In Section 1.1, it was noted that SOV and VSO languages tend to have mirror-image orders. Consider the order of Subject, Object, and Verb, Possessor and Possessum, and Noun Phrase and Adposition in English. Which of the two types does English belong to or stand closer to?
- 5. Here is a paradigmatic implicational universal (Greenberg 1966a: #24). If the relative clause precedes the noun either as the only construction or as an alternative construction, either the language is postpositional or the adjective precedes the noun or both.

What does this statement say about the language types "mimicked" by the following sentences? For each type, circle your answer. Relative clauses are bracketed; * indicates the structure is ungrammatical.

- TYPE A: (a) The [yesterday I bought] apples are sweet.
 - (b) sweet apples
 - (c) the store in
- Answer: i. predicts this type
 - ii. excludes this type
 - iii. there is not enough data to decide
 - iv. does not say anything about this type
- TYPE B: (a) The [yesterday I bought] apples are sweet.
 - (b) apples sweet
 - (c) in the store
- Answer: i. predicts this type
 - ii. excludes this type
 - iii. there is not enough data to decide
 - iv. does not say anything about this type
- TYPE C: (a) The apples [I bought yesterday] are sweet.
 - (b) *The [yesterday I bought] apples are sweet.
 - (c) sweet apples

Answer:	i. ii. iii. iv.	predicts this type excludes this type there is not enough data to decide does not say anything about this type
TYPE D:	(a) (b)	The [yesterday I bought] apples are sweet. apples sweet
Answer:	i. ii. iii. iv.	predicts this type excludes this type there is not enough data to decide does not say anything about this type
TYPE E:	(a) (b) (c) (d)	The apples [I bought yesterday] are sweet. The [yesterday I bought] apples are sweet. sweet apples in the store
Answer:	i. ii. iii. iv.	predicts this type excludes this type there is not enough data to decide does not say anything about this type

Further reading

- Two comprehensive handbooks of language typology are Haspelmath *et al.* 2001 and Song 2011. See also issue 11/1, 2007, of the journal *Linguistic Typology* devoted in its entirety to what language typology is and how it relates to other endeavors within the science of linguistics. A careful assessment of the extent to which language universals exist at all is Evans and Levinson 2009 along with the detailed responses following the article.
- Textbooks on language typology include Ramat 1987, Comrie 1989, Whaley 1997, Song 2001, and Croft 2003.
- Nutshell grammars written in popular style are provided by Shopen (ed.) 1979a and 1979b. The Atlas of Languages (Comrie et al. 2003) surveys the languages of the world by continents; it is written in an easy style with lots of pictures. Comrie 1990 provides brief grammars of the world's major languages. The Book of a Thousand Tongues (Nida 1972) contains passages from the Bible in 1431 languages.
- The Internet addresses of the three institutions mentioned in the text are as follows:
 - Research Centre for Language Typology (La Trobe University): www.latrobe.edu.au/rclt
 - Language and Culture Research Centre (Cairns Institute, James Cook University): https://eresearch.jcu.edu.au/spaces/TLA
 - Summer Institute of Linguistics International: http://www.sil.org
- For detailed discussions of sampling issues including the optimal size of a sample, and of data sources, see Song 2001: 17–41, Croft 2003: 19–30, and Bakker 2011.
- For a comprehensive survey of linguistic diversity across genetic and areal groups, see Nichols 1992.
- For a survey of the languages of the world, see Pereltsvaig 2012.