

WARNING CONCERNING COPYRIGHT RESTRICTIONS: The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproduction of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If electronic transmission of reserve material is used for purposes in excess of what constitutes "fair use", that user may be liable for copyright infringement.

INTRODUCTION
TO
TYPOLOGY

THE
UNITY
AND
DIVERSITY
OF
LANGUAGE

*For Ida Mae Heemstra
I tried not to split too many infinitives.*

LINDSAY J. WHALEY



SAGE Publications
International Educational and Professional Publisher
Thousand Oaks London New Delhi

THE WORLD'S LANGUAGES IN OVERVIEW

The bread and butter of typology is cross-linguistic comparison, so it should come as no surprise that this book contains information and data on a great many languages. No doubt, there will be some languages that are familiar, but probably far more that most readers have never encountered before. For this reason, I offer some basic facts about the languages of the world and the relationships among them. Readers who have had little or no exposure to language classification are encouraged to read this prefatory material. It may provide them with enough of a grounding in the topic to feel more comfortable with my continual reference to lesser-known languages.

Perhaps one of the most common questions that linguists are asked is, how many languages are there? The answer is that no one really knows. This is partly due to the fact that some areas of the globe have not yet been surveyed in a systematic manner to determine the various dialects and languages that are spoken in them. An even bigger obstacle to answering the question about how many languages exist is that there is no consensus on when two varieties of speech are best analyzed as dialects of one another and when they should be taken as separate languages. From a linguistic standpoint, the choice of

how to label two speech types has much to do with the degree of intelligibility between them. Thus, it seems patently obvious that the variety of English spoken in Manchester, New Hampshire, and that spoken in Manchester, England, are dialects of a single language. After all, a speaker of one of these dialects can understand nearly everything that the other one says. Likewise, everyone agrees that English and Japanese are not dialects of a single language, but rather two distinct languages, because the degree of mutual intelligibility between them is about zero. These cases are relatively straightforward, but what about those instances in which speakers grasp about 95% of the content of another speech variety? How about 80% or 70%? Where does one draw the line? Situations in which there is imperfect comprehension between members of two speech communities pose an intractable problem for any simple counting of the number of languages in the world.

With these difficulties in mind, it is possible only to provide an estimate of roughly 4,000 to 6,000 languages that are currently in use. There is, of course, no way to know how many additional languages may have been spoken previously but have disappeared without leaving any trace.

Because no individual, no matter how strong their expertise in linguistics, knows about each of these languages, it has become a common practice in Linguistics to provide a genetic identification of a language when it is being described for those who may not be familiar with it. The genetic identity of a language is the language family to which it belongs. A language family is a group of languages or dialects that have arisen from a common ancestor. For example, at some point in the distant past (prior to 1,000 BC), Danish, English, German, Gothic, and Swedish (as well as several others) were not distinct tongues, but rather formed a single language that is commonly referred to as Proto-Germanic. We do not possess any written material from Proto-Germanic. We know a great deal, however, about the sounds of the language and the rules of its grammar because historical linguists have meticulously developed a reconstruction of what the language would have been like. Over time, dialects of Proto-Germanic formed, just as they do with any language. These dialects became more and more differentiated until they were no longer mutually intelligible—that is, they became distinct languages.

The evolution of languages from a shared ancestor is commonly depicted by a family tree. Figure A, for instance, is a family tree for Germanic languages.

The family tree in Figure A captures the genetic affinity between all the languages that are listed by having them all ultimately branch from the node

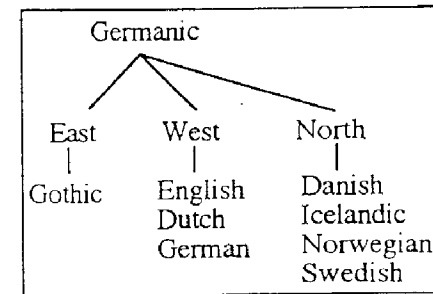


Figure A. The Germanic Languages

labeled “Germanic.” The tree also reflects that certain members of the Germanic family are more closely related than others through subgroupings, called branches, such as West Germanic, East Germanic, and North Germanic.

The entire Germanic family is itself nested in a larger family named Indo-European, which includes branches such as Italic (French, Spanish, Portuguese, etc.), Balto-Slavic (Russian, Polish, etc.), and Indic (Hindi, Bengali, etc.) as well as many others. Consequently, one can classify English as West Germanic, Germanic, or Indo-European. All are accurate labels; they simply reflect different degrees of association.

When languages are introduced in this book, I will use labels of genetic relatedness that are roughly equivalent to the level of Germanic. These groupings are largely, although not completely, uncontroversial and can be established quite easily using the conventional tools of historical linguistics. They reflect a time depth (the point at which languages start branching off from the common ancestor) of about 2,500 to 4,000 years. Where my sources for language data did not provide sufficient information to determine an appropriate label of family membership, I relied on Ruhlen (1987).

In addition to furnishing the genetic affiliation of a language, I also give the geographic area with which the language is most commonly associated—for example, French (Italic: France). For languages that are no longer spoken, the genetic affiliation is furnished, but there is no geographic data. The identificational information is only provided the first time a language is discussed in a chapter. I have also only included it in cases in which some linguistic feature of the language is exemplified or discussed.

I end this brief overview with a list of the languages that appear in the book. To provide a sense for how the various language families cluster together into larger groups, I have organized this list by language phyla (also called macrofamilies). Some of these phyla are generally accepted (e.g., Indo-European), whereas others are highly controversial (e.g., Altaic and Amerind). The phyla names are in all capital letters, the family names are in italics, and individual language names are in regular type.

INDO-EUROPEAN

Albanian:

Albanian

Armenian:

Armenian

Balto-Slavic:

Bulgarian, Lithuanian, Polish, Rumanian, Russian, Serbian

Celtic:

Welsh

Germanic:

Danish, German, Swedish

Hellenic:

Greek

Indo-Iranian:

Bengali, Hindi, Persian, Punjabi

Italic:

French, Latin, Spanish

URALIC

Finno-Ugric:

Finnish, Hungarian, Komi

NIGER-KORDOFANIAN

Niger-Congo:

Akan, Awutu, Bambara, Bamileke, Beembe, Dewoin, Ewe, Kinyarwanda, Kirundi, KiVunjo-Chaga, Lobala, Mende, Sesotho, Swahili, Wolof, Yoruba

NILO-SAHARAN

Nilotic:

Maasai

Saharan:

Kanuri

KHOISAN

Nama

AFRO-ASIATIC

Chadic:

Hausa, Ngizim

Cushitic:

Somali

Semitic:

Akkadian, Arabic, Hebrew, Tigre, Tigrinya

CAUCASIAN

South:

Georgian

Northwest:

Abaza, Abkhaz

Northeast:

Avar, Tabassaran

ALTAIC

Japanese-Ryukyuan:

Japanese

Korean:

Korean

Manchu-Tungusic:

Even, Evenki, Oroqen

Mongolian:

Mongolian

Turkic:

Turkish

ESKIMO-ALEUT

Aleut:

Aleut

Eskimo:

Greenlandic Eskimo, Iñupiaq

ELAMO-DRAVIDIAN

Dravidian:

Malayalam, Tamil, Telegu

SINO-TIBETAN

Sinitic:

Mandarin Chinese

Tibeto-Burman:

Burmese, Gurung, Lisu, Manipuri, Tamang, Tangut

AUSTRIC

Miao-Yao:

Yao

Mon-Khmer:

Khmer

Munda:

Mundari

Daic:

Thai, Yay

Austronesian:

Achenese, Agutaynen, Chamorro, Enga, Fijian, Futunu-Aniwa, Hawaiian, Indonesian, Malagasy, Paamese, Palauan, Tagalog

INDO-PACIFIC

New Guinea:

Barai, Daga, Dani, Kobon, Taoripi

AUSTRALIAN

Burarran:

Burera

Kalkatungic:

Kalkatungu

Karnic:

Diyari

Pama-Nyungen:

Dyirbal, Mparntwe Arrente, Wangkumara, Warlpiri, Yidiny

NA-DENE

Athabaskan:

Navaho

Tlingit:

Tlingit

AMERIND

Almosan-Keresiouan:

Blackfoot, Cayuga, Halkomalem, Kutenai, Lakota, Nootka, Oneida

Carib:

Carib, Hixkaryana, Makusi

Chibchan:

Guaymi

Equatorial-Tucanoan:

Bare, Guarani, Inga, Kaiowa-Guarani, Quechua, Tuyuca, Urubú, Yagua

Ge-Pano:

Cashibo

Hokan:

Atsugewi, Eastern Pomo, Mohave, Seri, Yuma

Oto-Manguean:

Isthmus Zapotec, Otomi, Mixtec

Penutian:

Choctaw, K'ekchi, Miwok, Mixe, Sierra Popoluca, Tepehua, Yokuts

Tanoan:

Southern Tiwa

Uto-Aztecan:

Comanche, Hopi, Kawaiisu, Michoacan Nahuatl, 'O'odham, Shoshone

ISOLATES/UNKNOWN

Burushaski

Illyrian (unclassified)

CREOLES/PIDGINS

Melanesian Pidgin, Papiamentu

ABBREVIATIONS

<i>Abbreviation</i>	<i>Explanation</i>	<i>Abbreviation</i>	<i>Explanation</i>
1	First person	AUX	Auxiliary
1D	First-person dual	BEN	Benefactive
1P	First-person plural	C1, C2	Noun Class 1, Class 2, and so on
1S	First-person singular	CAUS	Causative
2	Second person	CLASS	Classifier
2D	Second-person dual	CM	Case marker
2P	Second-person plural	CMPLT	Completive
2S	Second-person singular	COM	Comitative
3	Third person	COMP	Complementizer
3P	Third-person plural	CONTR	Contrastive
3S	Third-person singular	D	Dual
ABL	Ablative	DAT	Dative
ABS	Absolutive	DECL	Declarative
ACC	Accusative	DEF	Definite
ACT	Active	DEM	Demonstrative
ADJ	Adjective	DET	Determiner
AFFIRM	Affirmative	DIR	Directional
AGR	Agreement	DO	Direct object
AGT	Agent	DPST	Distant past
ANIM	Animate	DS	Different subject
ANT	Anterior	ERG	Ergative
AOR	Aorist	EXCL	Exclusive
APPL	Applicative	FACT	Factive
ART	Article	FEM	Feminine
ASP	Aspect	FOC	Focus
ATTN	Attention		

<i>Abbreviation</i>	<i>Explanation</i>	<i>Abbreviation</i>	<i>Explanation</i>
FUT	Future	PART	Participle
GEN	Genitive	PASS	Passive
HAB	Habitual	PAT	Patient
HUM	Human	PERF	Perfect(ive)
HYP	Hypothetical	PFV	Perfective
IMPF	Imperfect	PNoun	Proper noun
INCH	Inchoative	POSS	Possessive
INCL	Inclusive	POT	Potential
IND	Indicative	PRES	Present
INF	Infinitive	PROG	Progressive
INST	Instrument(al)	PROHIB	Prohibitive
INTER	Interrogative	PROP	Propriative
INV	Inverse	PST	Past
IRR	Irrealis	PTL	Particle
LOC	Locative	QUES	Question word
MOD	Modalis case	RCP	Reciprocal
MSC	Masculine	REAL	Realis
NEG	Negative	REL	Relativizer/relative pronoun
NEUT	Neuter	S-O REV	Subject-object reversal
NOM	Nominative	SBJV	Subjunctive
NOML	Nominalizer	SEQ	Sequential
nonFUT	Nonfuture	SS	Same subject
nonPST	Nonpast	STAT	Stative
NPOSSD	Nonpossessed	SUB	Subject
OBJ	Object	TM	Tense marker
OBL	Oblique	TNS	Tense
OBV	Obviative	TOP	Topic
OH	Object higher	TRANS	Transitive
OPT	Optative	VOL	Volitional
P	Plural		

P A R T

I

Basics of Language Typology

Introduction to Typology and Universals

What is language? On the face of it, the question seems simple. After all, language is so much a part of our everyday experience, so effortlessly employed to meet our impulses to communicate with one another, that it cannot be too intricate a task to figure out how it works. Hidden below the surface of the “what is language” question, however, is a web of mysteries that have taxed great minds from the beginning of recorded history. Plato, Lucretius, Descartes, Rousseau, Darwin, Wittgenstein, and Skinner, to name just a few, have all probed into some aspect of the human capacity for speech, yet none of them were able to explain the origin of language, why languages

differ, how they are learned, how they relay meaning, or why they are the way they are and not some other way. Language largely remains an enigma that awaits further exploration.

This is not to say that we have learned nothing or know nothing about our ability to utter meaningful sequences of sound. Centuries of careful observation and experimentation on language have revealed some extraordinary insights into its fundamental properties, some of them quite surprising. Perhaps most significant, language has no analogs in the animal kingdom. Nothing remotely similar to language has been discovered in the vast array of communication systems utilized by the fauna of our planet. Language, it seems, is uniquely human, a fact summarized well by Bertrand Russell (1948) when he exclaimed, “A dog cannot relate his autobiography; however eloquently he may bark, he cannot tell you that his parents were honest though poor.”

In addition to the species-specific quality of language, a second basic notion about language might be highlighted that has become foundational for modern linguistics: There is a basic unity that underlies the awesome diversity of the world’s languages. Whether it be Apache or Zulu or Hindi or Hebrew, there are certain core properties that languages have in common. These properties, often referred to as **language universals**, allow us to say that all languages are, in some sense, the same.

This is, in many ways, an astonishing claim, especially when confronted by the immense variety in the structures of the world’s languages. Consider the following sentences, the first from Lobala (Niger-Congo: Zaire) and the second from Hixkaryana (Carib: Brazil).¹

- (1) a. moto me t-a-iká mo-phé ná baphalnágá ná ntóma
 man DEM NEG-3S-PST CI-give and money and food
 The man didn’t give him either money or food.

(Data from Morgan 1994, 133)

- b. apaytara y-ari-hira nex-e-ye wekoko
 chicken 3S/3S-take-NEG be-DPST hawk
 The hawk didn’t take the chicken.

(Data from Derbyshire 1985, 138)

Despite the fact that the two sentences in (1) are both simple negative clauses, they appear to have little in common with each other or with the equivalent

English sentences. For instance, the order of words varies: The subject of the Lobala example, *moto me* (“the man”) occurs at the beginning of the sentence, whereas the Hixkaryana subject *wekoko* (“hawk”) is located at the end. The negation in Lobala is indicated by an auxiliary verb *t-*, but in Hixkaryana there is a negative suffix (*-hira*) on the main verb, *-ari-* (“take”). Both languages exhibit verb agreement, but in Lobala the agreement suffix (*-a*) is found on the negative auxiliary, and it only reveals information about the subject (namely, that the subject is third-person singular). In Hixkaryana, the agreement marker is a prefix (*y-*) rather than a suffix. Furthermore, it is located on the main verb, and it reveals information about both the subject and the object (namely, that both are third-person singular). With regard to these, and many other, differences, the concept of “language universals” may seem hard to accept. Nevertheless, most linguists would claim that there is an underlying homogeneity to language that is far more striking than differences like those just described. Discovering instances of this homogeneity and determining why it exists constitutes one of the major research goals of modern linguistics in general and typology specifically. It also represents the primary concern of this book.

The consensus in linguistics about the underlying unity of language is not paralleled by agreement over how the unity is to be explained or, even more fundamentally, what even constitutes an explanation for the unity. On this point, there are profound philosophical and methodological differences. For example, Noam Chomsky, perhaps the most significant figure in modern linguistics,² has argued that the unity is due to human biology. In his view, all humans are genetically endowed with a “language faculty,” which is distinct from other cognitive capacities. As children are exposed to the particular language (or languages) of their speech communities, this language faculty directs them in the rapid acquisition of a complex and mature grammatical system. To accomplish this, the language faculty must contain enough information (called **Universal Grammar**) to ensure that the child can learn a language accurately and learn it in the space of just 4 or 5 years. On the other hand, the innate language faculty must be flexible enough to give rise to the diverse array of structures that we actually find in the world’s languages.

Chomsky has appealed to an extraterrestrial authority on several occasions (e.g., Chomsky 1988, 1991) to drive this point home. Chomsky (1991) suggests that a Martian scientist who visited Earth would reach the following conclusion about a human’s inborn capacity for language:

Surely if some Martian creature, endowed with our capacities for scientific inquiry and theory construction but knowing nothing of humans, were to observe what happens to a child in a particular language community, its initial hypothesis would be that the language of that community is built-in, as a genetically-determined property, in essentials And it seems that this initial hypothesis may be very close to true. (26)

In contrast to Chomsky, other linguists have argued that the unity that underlies languages is better explained in terms of how languages are actually put to use. To be sure, languages are all employed for like purposes: asking questions, scolding bad behavior, amusing friends, making comparisons, uttering facts and falsehoods, and so on. Because languages exist to fulfill these types of functions, it stands to reason that speakers will develop grammars that are highly effective in carrying them out. Consequently, under the pressure of the same communicative tasks, languages evolve such that they exhibit grammatical similarities. Language universals, under this “functional” perspective, result from commonalities in the way language is put to work. Closely allied with this view is the proposal that common experiences among humans can account for certain universals in language structure. Lee (1988) articulates this view well:

Despite the fact that I come into contact with quite a different set of objects than a Kalahari bushman, the possible divergence between our experiences in the world is circumscribed by a number of factors independent of us both, and even of our speech communities as a whole. For example, we can both feel the effects of gravity and enjoy the benefits of stereoscopic vision. These shared experiences exert a force on the languages of all cultures, giving rise to linguistic universals. (211-12)

Which explanation for the similarities between languages is right? In all likelihood, the unity of language, and consequent language universals, arises from a slate of interacting factors, some innate, others functional, and still others cognitive, experiential, social, or historical. What this means in practical terms is that there are a number of legitimate ways to approach the question, “What is language?” This book examines the nature of language from a “typological” approach. In the following section, a better sense for what this means is developed.

1.0. Defining “Typology”

What exactly is meant by typology in the context of linguistics? In its most general sense, **typology** is

- (2) The classification of languages or components of languages based on shared formal characteristics.

As a point of departure, it is important to note that typology is not a theory of grammar.³ Unlike Government and Binding Theory, Functional Grammar, Cognitive Grammar, Relational Grammar, or the many other frameworks that are designed to model how language works, typology has the goal of identifying cross-linguistic patterns and correlations between these patterns. For this reason, the methodology and results of typological research are in principle compatible with any grammatical theory. The relationship between typology and theories of grammar is further explored later in this chapter and in Chapter 3.

Having described something that typology is not, we now must come to understand what it is. There are three significant propositions packed into the dense definition in (2): (a) Typology utilizes cross-linguistic comparison, (b) typology classifies languages or aspects of languages, and (c) typology examines formal features of languages. These parts of the definition will be examined one at a time with an eye to better understanding what is involved in performing language typology.

Proposition 1: Typology involves cross-linguistic comparison.

Ultimately, all typological research is based on comparisons between languages. Consider the following data:

- (3) a. I met *the man who taught you French*.
 b. *The dog which licked Cora* has become her friend.
 c. I sent the story to *the newspaper that your mother owns*.

From these sentences, we could form the generalization that English relative clauses (in bold type) follow the nouns that they modify (in italics). This description is of import to someone investigating English, but it is incomplete

as a typological claim because it is not grounded in a cross-linguistic perspective. Instead, in a typological approach we expect to find a description such as “English is typical in placing relative clauses after the nouns which they modify.” Note that to employ a term such as “typical” properly, one must first have gathered data on relative clauses from a representative sample of the world’s languages. Compiling an adequate sample remains one of the central methodological issues in typological research, an issue to which we return in Chapter 3.

Proposition 2: A typological approach involves classification of either (a) components of languages or (b) languages.

In the first case—classification of components of language—attention is directed toward a particular construction that arises in language—for example, reflexive verbs, oral stops, or discourse particles. Then, using cross-linguistic data, all the types of these specific phenomena are determined. The goal is to better comprehend how this facet of language operates by identifying the degrees of similarity and the degrees of variance that one finds among languages. There is also keen interest in determining whether there are correlations between the various patterns that one finds in a language.

For instance, we might do a typological investigation on oral plosive sounds.⁴ These are sounds, also called “stops,” that are produced when the airstream is completely impeded in the vocal tract, as in English [p] and [g]. If we were to examine the distribution of oral stops in the world’s languages, we would immediately be struck by the fact that all languages have at least one plosive sound. Thus, we would have discovered a universal about sound systems in human language. It is important to realize that this fact is not a logical requisite for language because we can easily conceive of a language that does not have any oral plosives. Therefore, our empirical discovery that all languages have at least one stop leads to an ontological question: Why should language be structured in this way? I return to the problem of explanation presently, but first let us determine what other sorts of facts about stops we would learn from our typological investigation.

There are over 50 distinct oral stops that occur in language, but individual languages utilize only a small proportion of this universal set, with languages such as Punjabi (Indo-Iranian: India, Pakistan), which has 24 plosives, being highly exceptional in how many plosives it contains (Gill and Gleason 1963). As we continued our investigation, we would further discover that plosive

sounds are not equally distributed in the languages of the world. Some are extremely common, such as [p], [t], and [k]. In fact, nearly all languages have at least one of them. In contrast, some plosives are relatively rare, such as the voiced uvular stop [G], which is found, for example, in Somali (Cushitic: Somalia). We now might notice certain intriguing facts about our emerging typology of plosives such as “unexpected” gaps. For instance, plosives created by bringing the lower teeth into contact with the upper lip are nonexistent, even though they are physically possible to articulate. Finally, we might identify certain stops as being “dominant”—for example, the voiceless alveolar stop [t] appears to be an especially dominant plosive sound because, if a language has only two voiceless stops, one of them is bound to be [t].

From this simple typological study we have learned a host of important facts about sound systems. Not all these facts are of the same sort. For example, some were absolute universals (e.g., all languages have at least one stop); some were universal tendencies (e.g., almost all languages have [p], [t], or [k]); and some were implicational universals (e.g., if a language has two voiceless stops, then one is a [t]). Implicational universals have played a particularly prominent role in typology because they commonly suggest connections between two or more aspects of language. We discuss the differences between these types of universals in Chapter 3.

Having ascertained some universals about stops, the next task is to provide an explanation for them. After all, the ultimate goal is to understand why language is the way it is. For present purposes, attention is restricted to one of the facts provided previously—that [p], [t], and [k] are extremely common in the languages of the world. In dealing with sounds, simple facts about human speech anatomy are generally a good place to begin when seeking an account for universals. Keating, Linker, and Huffman (1983) have proposed that [p], [t], and [k] are so widespread because they are aerodynamically efficient and they require less effort to produce than other stop sounds. Then, assuming that there is a tendency for “efficient” sounds to be incorporated into languages, we now have a reasonable account for the commonness of [p], [t], and [k].

The simplicity of our account thus far is a bit misleading. There are many further complexities that enter into a comprehensive explanation for the frequency of [p], [t], and [k] in language. First, note that the “efficiency” of sounds cannot be the sole factor involved in the evolution of sound systems. If it were, then the existence of sounds such as [G] in Somali, which are not particularly efficient aerodynamically, would be completely inexplicable.

Therefore, the assumption that efficient sounds tend to arise in language must be tempered with a proposal about how inefficient sounds get into phonological systems in the first place.

The explanation for common plosive sounds falls short in another way: We have not indicated how it is that efficient sounds are “incorporated” into languages. After all, a community of speakers does not consciously determine which sounds they will use in their language. Rather, new sounds gradually develop from existing sounds over time, the whole process being largely imperceptible to individual speakers. When we make a statement such as “efficient sounds tend to be incorporated into languages,” we are using a sort of convenient shorthand. In a full exposition, we also want to indicate the set of mechanisms through which efficient sounds enter a language.

For simplicity’s sake, I will not attempt to remedy the deficiencies with the original account of common plosive sounds. To do so would require a detailed discussion that does not directly relate to the purposes of this chapter. Nonetheless, it is important to keep in mind that simple statements that purport to explain language universals are usually intended as highly generalized accounts of the phenomena being investigated. As such, they are just the beginning of a satisfactory explanation.

The modest typological study just described focused on a single feature of language. The aim was not so much to classify languages as to understand some basic facts about phonology. A second kind of typological classification has the goal of classifying entire languages into categories based on shared properties. For instance, in the investigation of oral plosives we would instantly have found that different languages have different numbers of stops. We could take a sample of languages—for example, two languages from each of the major language families⁵—and divide them into types based on the number of oral stops that exist in their phonological inventories (Table 1.1).

The left-hand column of Table 1.1 indicates the number of oral stop sounds found in the language. The middle column furnishes the name(s) of the language(s) that possesses this number of stops. The right-hand column reflects the total number of languages that have a given stop inventory.

From Table 1.1, one can get a sense for the most common types of languages in terms of the number of oral stops (38% of our sample has between 6 and 8 oral stops) as well as the rarest (e.g., languages with over 14 stops). Furthermore, we begin to see the overall range of oral stop inventories in language (between 3 and 17).

TABLE 1.1 Number of Oral Stops in Languages^a

Number of Stops	Language	Total
3	Taoripi (IP)	1
4	Burera (Australian), Nama (Khoisan)	2
5	Beembe (NK), Dani (IP)	2
6	Evenki (UA), Greek (IE), Hopi (NAM), Mandarin (ST), Tagalog (AT)	5
7	Carib (SAM), Hebrew (AfA), Kanuri (NS), Maasai (NS)	4
8	Diyari (Australian)	1
9	Finnish (UA), Khmer (AuA), Thai (AT)	3
10	Hausa (AfA), Wolof (NK)	2
11		
12	Quechua (SAM)	1
13	Tlingit (NAM), Yao (ST)	2
14		
15	Telegu (Dravidian)	1
16	Bengali (IE)	1
17	Mundari (AuA)	1

a. AfA, Afro-Asiatic; AT, Austro-Tai; AuA, Austro-Asiatic; IE, Indo-European; IP, Indo-Pacific; NAM, Northern Amerind; NK, Niger-Kordofanian; NS, Nilo-Saharan; SAM, Southern Amerind; ST, Sino-Tibetan; UA, Ural-Altai.

To keep this example simple, it is necessary to ignore a host of potential problems surrounding the database of 26 languages. (The difficulties in developing an adequate database are described in Chapter 3). Therefore, the conclusions drawn from the study are purely impressionistic. We need to implement a much more rigorous study to confirm them.

Proposition 3: Typology is concerned with classification based on formal features of language.

There are many conceivable ways that one can talk about relationships between languages. For instance, languages can be placed into classes on the basis of their genetic relationships. Were this our concern, we would group together all languages that demonstrably have a common origin. In doing so, we would produce a set of “language families”: Indo-European, Afro-Asiatic, Manchu-Tungus (as in Figure 1.1), and so on.

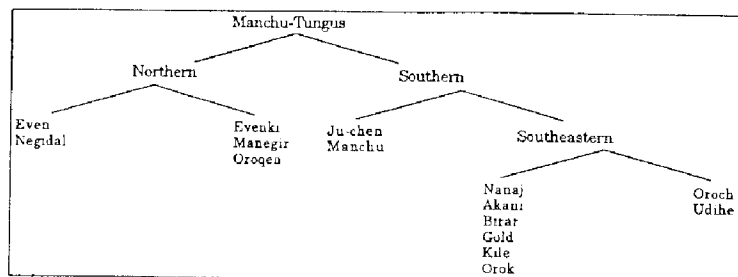


Figure 1.1. Manchu-Tungus Languages

In other circumstances we might choose to classify languages by their geographic location. We might then talk about Australian languages or the languages spoken in Northwestern Nigeria (Table 1.2).

In still other cases, we might classify languages in terms of demographic features—for example, languages with over 100 million speakers (Table 1.3).

Of course, all these methods of classification are useful devices for a particular goal. Their potential significance should be quite clear. They are not typology, however. Typologists, in contrast, classify languages in terms of the forms out of which a language is composed—its sounds, morphemes, syntax, or discourse structure.

This is not to say that these other kinds of classification are entirely unrelated to typology. The strong association between typological and genetic classification is most obvious. It is no surprise that Spanish (Italic: Spain and Latin America) and French (Italic: France) both have articles that reveal gender or that they both have subject agreement marked on verbs because we know that both languages have inherited these traits from Latin (Italic). The typological similarity of the two languages is a function of their genetic association.

The relationship between typology and areal classification is less well understood. To what extent the structure of one language can be affected by the languages around it is an area of intense current research. There is plenty of evidence, however, to demonstrate that grammars are flexible enough to adopt some features of other languages that are close to them spatially. One well-known instance of a group of languages that share grammatical features

TABLE 1.2 Languages of Northwestern Nigeria

Asu	Gurmana	Lela	Shama
Baangi	Gwamhi	Lopa	Shanga
Baatonun	Hausa	Madaka	Sorko
Basa	Hugworo	Nupe	Tiyal
Busa	Hun-Saare	Pongu	Wayam Rubu
Cinda	Kag	Regi	Zarma
Dendi	Kambari	Reshe	
Fungwa	Koromba	Sagamuk	
Gbagyi	Laru	Sambuga	

TABLE 1.3 Languages With Over 100 Million Speakers

Language	No. of Speakers (in Millions)
Mandarin	907
English	456
Hindi	383
Spanish	362
Russian	293
Arabic	208
Bengali	189
Portuguese	177
Indonesian	148
Japanese	126
French	123
German	119

because of their geographic proximity is found in the Balkans. In this region (often referred to as a **Sprachbund**), one encounters among other languages Albanian (Albanian: Albania), Bulgarian (Balto-Slavic: Bulgaria), and Rumanian (Balto-Slavic: Romania), all of which come from different subfamilies of Indo-European. Certain linguistic patterns permeate the languages in this area despite their different genetic affiliations. For example, many of them have definiteness marked by a suffix on the noun:

- | | | | |
|-----|-----------|---------|--------------|
| (4) | Albanian | mik-u | "friend-the" |
| | Bulgarian | trup-at | "body-the" |
| | Rumanian | om-ul | "man-the" |

(Data from Bynon 1977)

What is amazing about this particular example is that the use of suffixes to mark definiteness is not a trait of any of the language branches from which the languages come. Indeed, the origin of this formal trait is still somewhat of a mystery. What is crucial for present purposes, however, is that, from a genetic standpoint, none of these languages are expected to employ this morphological strategy. The fact that they share it can only be due to their geographical connection.

This type of linguistic similarity between languages of different genetic stock is particularly common in speech communities where two or more languages coexist and there is a high degree of multilingualism; in such cases, it is well attested that parts of the grammar of one language can be adopted by another (see Myers-Scotton 1993).

Therefore, although typological classification is a different sort of procedure than are genetic, geographic, and demographic classifications, it must be recognized that the typological characteristics of languages can be greatly influenced by these other factors.

One final point about typology's focus on the formal features of language requires comment. "Formal features" are the chunks of information that one finds in language, its phrases, sentences, and so on. These features are, of course, used to convey meaning. Consequently, typologists have always been concerned with semantic categories, such as "tense," "agent," or "gender," and how these categories are manifested by the formal units of language. Therefore, the emphasis on formal features in the definition of typology given previously should not be taken to exclude semantic considerations.

2.0. Summary

Breaking down the definition of typology into three parts has helped to clarify what is involved in exploring language from a typological perspective. Before we move on to the next chapters, one reminder is needed: Typology is more than mere taxonomy (listing)—what we find in a typological analysis should have implications for our question, "What is language?" To get to the heart of this question requires one to seek explanations for typological patterns. It is not enough to say "labiodental stops do not exist" or "relative clauses typically follow the noun which they modify." We must seek a

plausible explanation for why these things are so. This type of "why" question, however, often forces us to go beyond grammar itself for an answer. We commonly must enter other "extragrammatical" domains: discourse, pragmatics, physiology, cognition, speech processing, language contact, social influences, and so on.

A reliance on extragrammatical explanation closely aligns typology with functional approaches to language, which are also based on the premise that language structure cannot be properly understood without reference to its communicative functions. Even so, formal linguistic theories which have traditionally eschewed explanations based on extragrammatical factors, are increasingly integrating typological research as well (Kayne 1994 and Fukui 1995 are representative examples). I return to this topic in Chapter 3.

Typology has also enjoyed some prominence in the area of historical linguistics. When linguists attempt to reconstruct languages that no longer exist and for which there is no written record, they are essentially involved in making a series of "educated guesses" about what these languages looked like. The principles and methods of historical linguistics provide the tools necessary to do reliable reconstructive work. Typological findings are one of these tools. Recalling the previous discussion on oral stops, any reconstruction of a phonological system that posited an inventory with only the two voiceless plosives [p] and [k] would immediately be suspect because it is a violation of the implicational universal, "if a language has two voiceless stops, then one of them is [t]."

Perhaps most important, typology is extremely useful as a tool in linguistic fieldwork. Most linguists never have the opportunity to organize large samples of languages to test for universals, nor do they ever propose completely new grammatical theories. All linguists, however, do research on individual languages. How is typology useful in this endeavor? First, it aides in learning about languages that previously have not been documented by making the linguist aware of what is "out there" and what is "typical." Many phenomena that might seem quite exotic in comparison to one's native language may actually be typologically common. Such phenomena are not likely to cause confusion or frustration for the field-worker familiar with typology. This is not only true for typological claims about language in general but also with analyses of typological characteristics of individual language groups. Today, the vast majority of fieldwork is carried out with languages in language families about which a fair amount is known. Becoming an expert on the typical characteristics of a given family before arriving on the field is vital.

In the same vein, if linguists know what is common in language, they will be quick to recognize unexpected patterns in the language(s) being examined. The recognition of these unusual data leads to the search for explanations, which in turn can reveal important details of the language history, contact with other people groups, or other features of the grammatical system. In an extreme case, the aberrant construction may ultimately prove to be something that typologists have thought not to exist. The linguist who knows the claims that have been made about language universals can grasp the significance of such a discovery and pass it on to the field as a whole. This is precisely what happened with Des Derbyshire's work with Amazonian languages. Languages that ordered direct objects before both verbs and subjects (see (1b)) were not thought to exist until Hixkaryana data proved otherwise (Derbyshire 1977; Derbyshire and Pullum 1981).

Finally, a knowledge of typology provides access to a major corpus of linguistic literature. To be able to use this research correctly, one must have a good background in the terminology and methodology of the discipline.

3.0. Key Terms

Language universals
Sprachbund

Typology
Universal Grammar

Notes

1. A key to the abbreviations used in the morpheme-by-morpheme glosses may be found at the beginning of the book. Because the data are drawn from a variety of traditions, all with their own terminology, I have taken the liberty of unifying the glosses in many instances and simplifying them in others. Throughout the book, language names are followed by an indication of genetic affiliation (i.e., the language family to which the language belongs) and the location of the largest concentration of native speakers or the area with which the language is commonly associated. The classification and geographic information appears only with the first mention of the language in a chapter. For details on the genetic affiliations I have ascribed to the languages see "The World's Languages in Overview".

2. This terse description does not do Chomsky's name full justice. In a very real sense, Chomsky has directed the course of linguistics, especially in North America, for the past 30 years.

In addition, he has had a major influence on political thinking, psychology, and the philosophy of language. See Lyons (1991) and Salkie (1990) for overviews of Chomsky's impact on the intellectual community.

3. Croft (1990, 2-3) points out that the term *typology* is often used to denote a perspective on language that is unique from that which undergirds grammatical frameworks such as Government and Binding Theory, and in this way typology might be considered a theory of grammar.

4. Data to support the generalizations about oral stops can be found in Maddieson (1984).

5. The seemingly innocuous decision to draw two languages from each of the major language families is, in fact, highly problematic because there is little consensus on how many major language families there are or on the internal constitution of these families. For convenience, I have employed the genetic classification used by Maddieson (1984) because the data are also taken from this source.

2

A (Brief) History of Typology

Broadly speaking, typology has a twofold purpose: to identify universals and to establish the potential range of variation among languages. These research aims first arose in linguistics back in the 1800s, particularly in the work of two German linguists, Friedrich von Schlegel and Wilhelm von Humboldt (Ramat 1995). Unlike contemporary typologists, they were interested almost exclusively in morphology. By examining the processes of word formation that occur in language, they developed a scheme to categorize languages on the basis of how much morphology was used in the construction of a word and how this morphology was used. Although contemporary linguistics still uses some of the terminology they developed in their investigations, modern typology has little in common with the research of these pioneers. The assumptions, methods, and focus of current typological research have all changed dramatically.

Understanding the transitions from early typology to the present day conveniently highlights certain central assumptions that underlie the analyses presented in this book and deepens our awareness of what typology is all about. Therefore, in this chapter, I investigate the changes that have occurred in typology during the past 100 years and have given rise to modern typological analysis.¹

1.0. Early Typologists

No scholarly work occurs outside of a certain worldview. Today, for example, more and more research in linguistics begins with the belief that language (and all other mental activity) is explicable in purely physical terms. Under this view, which would have been incomprehensible for a good share of human history and laughable for most of the rest, the production and comprehension of sentences is ultimately nothing more than the firing of neurons. The neurons themselves are subject to the same physical laws that account for planetary motion, the properties of light, and reproduction.

The work of Schlegel, Humboldt, and their contemporaries was carried out under different presuppositions. Most significant for typology, they believed language to have an abstract organic unity. That is, the formal aspects of language (its sounds, morphemes, grammar, and the like) and the changes that happened to these forms over time were not random or arbitrary because they all were reflections of an inner character. Like any organism, a language could develop over time, but it would always have the same essence. They believed Chinese, English, Kiowa, and Yoruba, as well as any other language, differed because the inner character of the people, the fundamental spirit of their culture, which gave rise to the languages, was not the same.

Such a notion is difficult to grasp because it is far removed from the current conception of language, but it is not nearly as “strange” as one might think. If we move to a different sphere of human activity—for example, economics—we find that the dominant metaphors of today reflect a conception of an organic entity. Hence, certain behaviors such as fluctuations in the stock market indicate something about the “health” of our economy. It might be described as “shrinking” or “growing” (ideally the latter!). Not only do our metaphors depict the economy as a living being but we also treat it as a being

that has power. It causes people to lose their jobs, to save or spend money, to invest, to go broke, etc. We, of course, must be careful not to push the analogy between the Humboldtian view of language and our view of the economy too far. The point is only to demonstrate that it is, in fact, not so “weird” to attribute coherence and life to abstract notions.

Perhaps because morphological differences between languages are so striking, it was this realm of grammar that was presumed to best mirror the organic essence behind languages. The basic distinction, due to Schlegel and his brother, was among **affixal** (1a), **inflectional** (1b), and **no structure** (1c) languages.

- (1) a. *Affixal*: Kirundi (Niger-Congo: Burundi)
 Y-a-bi-gur-i-ye abâna
 C1-PST-C8.them-buy-APPL-ASP C2.children
 He bought them for the children. (Adapted from Sabimana 1986)
- b. *Inflectional*: Attic Greek (Hellenic: Greece)
 hoi stratiôtai ēgoradz-on ta epitēdeia
 the soldiers buy-3P:IMPF:ACT:IND the provisions
 The soldiers were buying the provisions. (Xenophon, *Anabasis* 1.5.10)
- c. *No structure*: Mandarin Chinese (Sinitic: China)
 wǒ mǎi le shuǐguǒ le
 I buy ASP fruit PTL
 I have bought the fruit. (Adapted from Li and Thompson 1981)

Kirundi (1a) is representative of an affixal language in that it permits a series of morphemes to be affixed to a lexical head (i.e., a verb, noun, or adjective). Consider the verb, *yabiguriye*. There are three prefixes: *y-*, which indicates that the subject of the verb belongs to noun class 1; the past tense marker *a-*; and *bi-*, a morpheme that denotes a direct object belonging to noun class 8. In addition, there are two suffixes. The applicative morpheme, *-i*, being used here to identify the noun *abâna* (children) as the beneficiary of the act of buying, and *-ye*, which is an aspect marker (the concept of **aspect** is discussed in Chapter 8).

Inflectional languages, like Greek (1b), also evince affixation, but the affixes that are employed typically contain a great deal of semantic information. For example, the suffix *-on* reveals that the subject is third person (i.e., refers to someone other than the speaker or listener), that the subject is plural, that the verb is past tense and has a durative aspect, and that the sentence is a

statement of fact rather than a command or a condition. In inflectional languages, all this meaning is fused into a single affix, unlike affixal languages which tend to employ affixes that provide one piece of information each.

In no structure languages, as the name suggests, little affixation is used at all. Note that Mandarin Chinese (1c), which is commonly used as the quintessential example of a no structure language, has no verb agreement with the subject and that aspect marking, when it occurs at all, arises as a separate particle rather than a verbal affix.

Because language was thought to be unified, morphological classification such as that discussed here was thought to serve as a handy means of categorizing languages in their entirety. An examination of the syntax, for instance, would ultimately reveal the same inner character of the language as the morphology and, consequently, there was little reason to study it. This assumption permitted a benign neglect of syntax in typology that was not corrected for roughly a century.²

Like linguists today, Humboldt assumed that language had an unseverable association with the human mind. In fact, he believed that universals of language were manifestations of universals present in human thought (Brown 1967). Unlike modern linguists, however, Humboldt (1971) also thought that differences among languages reflected basic differences in the mental life of various speech communities:

Languages must have evolved along with the flourishing tribes from the respective intellectual peculiarities of the latter, which imposed numerous restrictions on them. . . . Languages are depicted as bound to, and dependent on, the nations to which they pertain. (2)

Even more removed from current perspectives on language, there was also a clear evaluative component to Humboldt's brand of typology. The quality of languages, he thought, could be determined by how closely they resembled an idealized linguistic system. Quite to the contrary, it is the consensus in modern linguistics that there is no qualitative difference between languages—no “better” or “worse.”

Because Humboldt also held that language structure was revelatory of intellectual capacity, his linguistic philosophy could easily be manipulated into claims of cultural superiority using the following logic:³ Because German more closely matches the structure of the perfect language than Chinese, it is superior to Chinese. Also, because language structure derives from intellec-

tual prowess, it follows that German thought is superior to Chinese thought. Having rejected both the assumption that languages can be judged against an ideal and the claim that variations in language structure relate to differences in intellectual capacity, linguists in the present age find it absurd to make any judgments about the quality of a culture on the basis of how words are formed and sentences composed. Several other major shifts in thought have occurred that separate modern-day typology from its roots. This is the topic of the following section.

2.0. Revolutions in Typology

Even in Humboldt's era, linguistics was becoming dominated by an historical-comparative approach to language study. That is, the major goals of linguistics were seen as understanding the processes that gave rise to language change and determining the historical relationship among languages. For this reason, typology was marginal to linguistics in the first half of the 1900s.

Nevertheless, in the early twentieth century several important changes transpired in linguistics that radically altered the assumptions of the Humboldtian typology discussed in Section 1. Grounded in the pioneering work of Ferdinand de Saussure, linguists began to argue that, although language may be organic and therefore changing, at any given point in time it is a self-contained system. Thus, Leonard Bloomfield (1933, 19) wrote, "In order to describe a language one needs no historical knowledge whatever." This is a shift from a **diachronic** (historical) perspective to a **synchronic** perspective (looking at a language at a single stage in its development).

Although linguists like Bloomfield—cumulatively referred to as the American Structuralists—continued to emphasize morphology in their research on languages, they wholeheartedly rejected any belief that differences in morphological form revealed differences in the "inner form" of the language (or anything about the intellect of the people who spoke it). Discarding this assumption meant going beyond morphology in linguistic analysis. It was no longer proper to ignore the other aspects of language such as syntax. Moreover, the possibility of languages of mixed types arose. Although two languages might be similar with respect to their morphology, they might be

radically diverse in terms of sentence structure. As a result, work in typology, when it arose, switched from a focus on languages as wholes (**holistic typology**) to features of languages (**partial typology**).

Across the Atlantic, another nucleus of linguistic thought, the Prague School, argued that certain characteristics of language are inherently linked.⁴ Roman Jakobson (1929, 1963) pointed out that the vowel inventory and consonant inventory in languages are connected in predictable ways. For example, if a language has nasal vowels, it will also have nasal consonants. Statements like this capture facts about language that are always true. Later work by the Prague School, particularly by Skalička (1935, 1979), recognized that many language properties are associated in probabilistic rather than absolute fashion. In describing them, then, one can only propose a universal tendency—for example, if a language has only one fricative (i.e., a sound made by the airstream passing through a small aperture in the vocal tract), it is probably [s].⁵ Although we can expect this "universal" to be true most of the time, there are languages that constitute counter-examples, such as Hawaiian (Austronesian: United States), which has the single fricative [h].

Although the American Structuralists and the Prague School furnished ideas that transformed typological thought, it was Joseph Greenberg who infused the field of linguistics with an optimism about typology's potential to deliver major discoveries about the nature of language. Although his impact on typology has been immense and varied, for present purposes I mention just a few of his more significant contributions. First, Greenberg (1954) sought to establish a quantificational basis for typological study. Until the time of Greenberg, typology was highly subjective—based almost entirely on the observations and intuitions of individual linguists.⁶ The usefulness of such research to the burgeoning field of linguistics was limited, in part, because it did not meet the "scientific" standards that American linguists were trying so desperately to achieve in the 1940s and 1950s. Greenberg developed a strategy to measure numerically both the degree and the types of morphology present in a language. Although Greenberg's method is laudable in its own right, the lasting significance of his quantitative approach has been that it showed that languages did not fall into discrete morphological types (Croft 1990). That is, a language such as English cannot be said absolutely to be an inflecting or no structure language (to borrow terms from the nineteenth century). Rather, it is closer to being a no structure language than Greenlandic Eskimo (Eskimo-Aleut: Greenland) but more inflecting than Khmer (Mon-Khmer: Cambodia). In this early work of Greenberg, another equally important assumption was

implicit—namely, that the proper task of typology is not comparing languages per se but instead comparing constructions. The point of typology is not to answer “What kinds of languages are there?” but to answer “What kinds of structures are in languages?” This assumption has become explicit in the work of many current typologists and also in several theories of grammar (e.g., Relational Grammar). There is ample opportunity to observe this fact in the chapters that follow.

Greenberg made full use of the Prague School notion that certain aspects of structure in language correlate and that implicational universals can be stated in terms of the correlation. These implicational universals have the form, “given X in a language, Y is also found.” His seminal paper, “Some Universals of Grammar With Particular Reference to the Order of Meaningful Elements” (Greenberg 1966), laid out 45 implicational universals. Universal 2 provides an example:

- (2) Universal 2 (Greenberg 1966): In languages with prepositions, the genitive almost always follows the governing noun, while in languages with postpositions, it almost always precedes.

Kinyarwanda (Niger-Congo: Rwanda) exemplifies the first type of language described by this universal:

- (3) a. Umugore y-oohere-je umubooyi **kw'** iisoko
 CI.woman CI-send-ASP cook to market
 The woman sent the cook to market.
 b. umwaana *w' umugore*
 child of woman
 the woman's child

In accordance with Universal 2, Kinyarwanda is a language that uses prepositions (in bold type) and places the genitive (i.e., the possessor—in italics) after the noun that governs it (the possessee), as can be seen in (3b). Japanese (Japanese-Ryukyuan: Japan) demonstrates the opposite pattern. It employs postpositions, and genitives are placed before the nouns that they modify (4b).

- (4) a. Yuuko wa Mitiko **ni** Koobe de dekuwasita
 TOP to in ran.into
 Yuko ran into Mitiko in Kobe.

- b. Tanaka no hisyo
 POSS secretary
 Tanaka's secretary

(Data from Dubinsky 1990)

Like the Prague School linguists, Greenberg (1966) made much use of probabilistic statements. In Universal 2, for example, he claims that “in languages with prepositions, the genitive almost always follows the governing noun.” Languages such as Swedish (Germanic: Sweden) in (5) are atypical in terms of the universal statement, but they are not altogether unexpected.

- (5) a. Han ramlade i sjön
 he fell into water
 He fell into the water.
 b. *Eriks* mor
 Eric's mother

(Data from Björkhagen 1962)

Greenberg's Universal 2 is also indicative of his intent to incorporate syntax into typological study. In fact, over half of the 45 universal statements Greenberg provided in his 1966 paper dealt exclusively with word order, and many of the statements that made reference to morphology were concerned with the relationship between affix ordering and syntax. The move toward syntactic analysis, however, was not unique to Greenberg. As mentioned previously, it was also true of Prague School typology and the American Structuralists. In point of fact, however, the present-day dominance of syntactic phenomena in typology is probably not so much due to Greenberg and those before him as to another American linguist, Noam Chomsky, who is discussed later in this chapter.

Another characteristic trait of modern typology that is represented well in Greenberg's work is a focus on the ways that language changes through time (see in particular Greenberg 1978). Greenberg's interest in diachrony was in many ways a throwback to the earlier days of typology in which historical-comparative linguistics predominated. The uniqueness of Greenberg's work, however, was in his use of language change as an explanation for language universals. The basic insight is the following: Because the form that a language takes at any given point in time results from alterations that have occurred to a previous stage of the language, one should expect to find some explanations for (or exceptions to) universals by examining the processes of language change. In other words, many currently existing properties of a

language can be accounted for in terms of past properties of the language. Examples of typological explanations based on language change arise in many places throughout this book. I specifically return to this important topic in Section 4 of Chapter 3.

Finally, Greenberg helped to draw attention to the importance of a proper database in the search for language universals. He made at least some attempt to remove the genetic biases from his claims about universals by using what at the time was considered a large sample of languages (30 languages altogether) and including languages from many language families. Although his sampling techniques have ultimately been shown to be inadequate in that they clearly fail to avoid a genetic bias (Dryer 1989b, 1992; Hawkins 1983), they had the effect of drawing attention to the importance of solid methodology when making cross-linguistic claims. We examine many of the methodological issues that are still being debated today in Chapter 3.

The last development in linguistics that has had a shaping influence on typology is Noam Chomsky's model of linguistic competence (its evolution can be traced through Chomsky 1957, 1965, 1970, 1981, 1988, 1992). For those who are familiar with the field of linguistics, the inclusion of Chomsky as one of the major molders of typology may appear awkward or even objectionable. After all, Chomsky himself has never engaged in typological research and has seemed generally skeptical about typology's capacity to inform him in his own work on syntax. The fact remains, however, that the cornerstone concept of Chomsky's model, Universal Grammar, has greatly affected typology.

Chomsky's understanding of Universal Grammar (UG), which he originally outlined in 1965, has changed slightly over time. In general, however, UG is taken to be the linguistic structures that are shared by all languages and a limited set of **parameters** over which languages are permitted to vary. To get a clearer idea of how UG operates, consider a simple example. In the English question in (6), the question word *whom* is at the beginning of the sentence.

(6) Whom did you see ____?

We recognize the placement of question words sentence initially to be a special feature of questions. Normally, the direct object in English is placed just after the verb, where the blank is found in (6). Thus, we might say that *whom* has moved from its typical position to the first position in the sentence. Compare this with Mandarin Chinese:

(7) ni kanjian-le *shei*?
 you see-ASP who
 Who do you see?

(Data from Huang 1982)

In (7), the question word *shei* remains in the usual spot for direct objects. Unlike English, the question word does not need to move to a sentence-initial position. This raises an intriguing question: Why should languages differ with regard to where question words appear?

Under Chomsky's view of Universal Grammar, the question is answered in the following manner. At an abstract level, the English and Chinese sentences have precisely the same word order. Indeed, this order follows principles that hold true of all languages. English and Chinese, however, are different in terms of the parameter of verb agreement (Agr). English is +Agr as can be seen by the appearance of the agreement marker *-s* in the present tense (I run vs. He runs). Chinese, however, is -Agr. The language never displays agreement on the verb. It has been claimed that the variation along this parameter is what triggers the required fronting of question words. +Agr languages such as English require it, whereas -Agr languages do not (Huang 1982).

It should be fairly obvious how the notion of Universal Grammar relates to typology. Research on Universal Grammar is aimed at discovering what all languages hold in common and the boundaries of their differences, which is precisely the same goal as typology. By bringing the search for universals to the center of formal syntactic theory, Chomsky established a point of contact with typological study. Ideally, typology should inform the model of Universal Grammar, and Universal Grammar should inform typology on motivations for the patterns in language it uncovers. Unfortunately, this ideal has not often been achieved (see Pullum 1979 and Newmeyer 1983, 67-72, for discussion) because of major disagreements about how universals are to be explained. Among other topics, we return to the issue of explanation in the next chapter.

3.0. Summary

In this chapter, some of the guiding assumptions of modern typology have been introduced by tracing developments in the field during the past century.

The starting point for all typology is the presupposition that there are recurrent structural patterns across languages that are not random or accidental. These patterns can be described in statements called *language universals*.

Once one grants this simple assumption, myriad questions arise. The first type of question is “What kinds of universals are there?” This is the topic of Section 1 in Chapter 3. The beginnings of the answer, however, were already hinted at previously; typologists explore both absolute properties of language and probabilistic properties. In addition, they are concerned with the connections between two or more properties.

A second key question about universals is “How are they determined?” I return to this question in Section 2 of Chapter 3. For now, it is sufficient to say that this question has become central to typology in the past few decades, and its answer has profound implications, particularly for universals that are based on statistical probability.

The final basic question that concerns modern typology is “How are universals explained?” A protracted debate over issues of explanation has been occurring since the 1950s. The most acrimonious elements of the debate have concerned the relationship between diachrony and synchrony (i.e., to what degree does an explanation require reference to past stages of a language?) and the need to go outside the language system itself in forming satisfying explanations (see Croft 1995 for a discussion on this latter topic). In Sections 3 and 4 of Chapter 3, I present some of the fundamental concepts involved in the debate.

4.0. Key Terms

Affixal languages	Parameters
Diachronic	Partial typology
Holistic typology	Synchronic
Inflectional languages	Universal Grammar
No structure languages	

Notes

1. See Greenberg (1974) for a more thorough treatment of the history of typology.
2. Martin Haspelmath (personal communication, 1995) rightly reminds me that the relative neglect of syntax at the time in favor of other aspects of language was also a function of the intense interest in Indo-European languages that were highly affixal.
3. Humboldt himself, however, explicitly denied that it was his intent to make claims about the ascendancy of one culture over another (see chapter 19 of Humboldt 1971).
4. Sgall (1995) surveys current and past contributions of Prague School linguistics to typology.
5. According to Maddieson (1984), this tendency holds true of 83.8% of the languages in his database that contain a single fricative.
6. Sapir (1921) sensed the dangers inherent in the overly subjective nature of the typology of his day: “It is dangerous to generalize from a small number of selected languages. To take as the sum total of our material Latin, Arabic, Turkish, Chinese, and perhaps Eskimo or Sioux as an afterthought, is to court disaster” (122).