7 Language typology

A **TYPOLOGY** is simply a categorization of some range of phenomena into various types. To "typologize" something is to group its parts into types. For example, we often hear jokes like the following: "There are three kinds of people – those who can count, and those who can't." Typological linguists are people who like to group languages into well-defined and useful types.

But what makes a typology useful? A typology is useful when it makes "predictions" about multiple characteristics of the items being typologized. For example, suppose we were to typologize motorized vehicles. Which would be the most meaningful typology, A or B?:

- Typology A: bus, van, automobile, tractor
- Typology B: red ones, green ones, blue ones, white ones

If you know that a motor vehicle is a bus, what else do you know about it? Quite a lot actually – it is probably going to be a large vehicle, with lots of seats, designed primarily to carry people, etc. If, on the other hand, you know some random motor vehicle is blue in color, there is not much else you can guess about its characteristics. Therefore, typology A is more useful, because it reflects "clusters" of structural and functional characteristics that go together, rather than simply indicating isolated properties.

Turning to a linguistic example, we could say that there are two kinds of languages in the world – those that have the sound [r] in their phonetic inventory and those that don't. However, knowing whether a language has an [r] is not likely to have many repercussions in other parts of the language, therefore this is not a particularly interesting or useful typology. However, there are several other linguistic typologies that have been very helpful to people interested in exploring the characteristics of the human mind. These are typologies that identify *clusters* of characteristics that languages are likely to possess.

The value of typologizing languages is that it helps linguists understand the range and limits of possible variation among human languages. If logically possible types are found to be very rare or nonexistent, that may provide some insight into how the human mind works. Thus language typology can give us a "window" on the mind and communication. To extend our non-linguistic example, if we typologized all the motorized vehicles in the world according to number of wheels, we might find that there are no, or extremely few, vehicles with five wheels. This fact would invite us to investigate *why* motorized vehicles are restricted in exactly

this respect. What is it about the origin, history, or function of motor vehicles that seems to rule out the existence of five-wheeled vehicles?

Several typologies of language have been proposed in the history of linguistic science. In this chapter, we will discuss morphological and syntactic typology. In later chapters we will discuss a typology of grammatical relations (chapter 8), voice and valence (chapter 9), and clause combining (chapter 10). Syntactic typology has proven particularly fruitful in stimulating the subfields of **TYPOLOGICAL LINGUISTICS**, and **FUNCTIONAL LINGUISTICS**.

Morphological typology

There are two parameters by which the morphological typology of a language may be measured. These are described by Comrie (1989) as the **INDEX OF SYNTHESIS** and the **INDEX OF FUSION**. The index of synthesis refers to how many morphemes tend to occur per word in a language, while the index of fusion refers to how many meanings tend to be associated with each morpheme.

The index of synthesis defines a continuum from **ISOLATING** languages at one extreme to highly **POLYSYNTHETIC** languages at the other. Figure 7.1 illustrates this continuum.

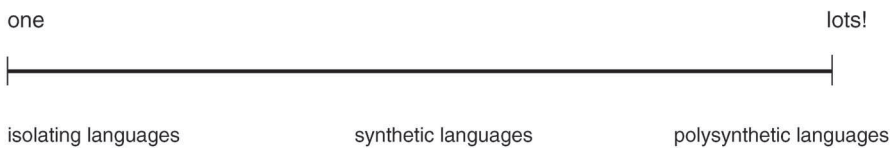


Figure 7.1 *The index of synthesis (the number of morphemes per word)*

A strictly isolating language is one in which every word consists of only one morpheme. The Chinese languages come close to this extreme. A highly polysynthetic language is one in which words tend to consist of several morphemes. The Quechuan and Eskimo-Aleut languages are good examples of highly polysynthetic languages. The following is an example of a polysynthetic structure in Central Yup'ik (thanks to Eliza Orr):

(1) Tuntussuqatarniksaitengqiggtuq
tuntu-ssur-qatar-ni-ksaite-ngqiggte-uq.
reindeer-hunt-FUT-say-NEG-again-3SG.IND
'He had not yet said again that he was going to hunt reindeer.'

The index of fusion (figure 7.2) describes a continuum between highly **AGGLUTINATIVE** languages to highly **FUSIONAL** languages. A highly agglutinative language is one in which most morphemes express one and only one meaning. A highly fusional language (sometimes called "inflectional," but since this has other connotations, we will use the term fusional) is one in which morphemes often express several meanings. For example, in Spanish the suffix *-ó* in a word like *habló* expresses at least five conceptual categories: indicative mood, third person, singular, past tense, and perfective aspect. If any one of these conceptual

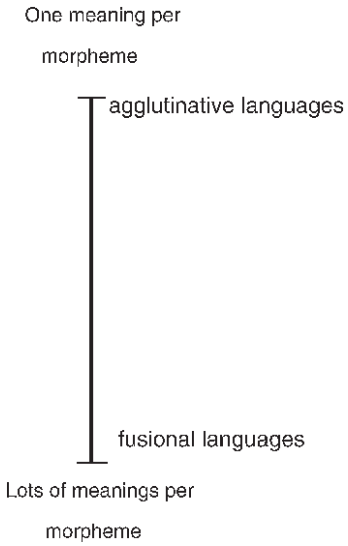


Figure 7.2 *The index of fusion*

categories changes, the form of the suffix must change. Turkish is a language for which each lexical meaning and conceptual category is, in general, expressed by its own morpheme. Therefore, Turkish is a highly agglutinative language. For highly isolating languages, the index of fusion just doesn't apply. If anything, English is agglutinative rather than fusional, e.g., in *anti-dis-establish-ment-ari-an-ism* each morpheme has a specific and fairly clear meaning. But then, such words in English are mostly of Latin origin. Fusion is apparent in English in the present tense, third person, singular suffix *-s*, as in *he walks the line*, and in the paradigm for the verb *be*, but not much else.

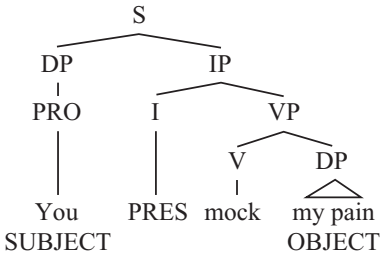
There is no generally accepted quantitative method for precisely establishing the indices of synthesis and fusion for a given language. A rule of thumb for the index of synthesis is that if the language can express a whole sentence with just a verb, it is polysynthetic. If it can't, then it is isolating. Adjectives such as "somewhat" or "highly" can then be added in order to give a sense of where a language falls on each continuum, e.g., English is "somewhat isolating," Mandarin is "highly isolating." Turkish is "somewhat polysynthetic and highly agglutinative" while Yup'ik is "highly polysynthetic and somewhat fusional." Knowing something about the morphological typology of a language helps linguists make better hypotheses about the likely meanings of various structures and helps tremendously in understanding the historical roots and development of a language.

Syntactic typology

Linguists have long noticed that some languages tend to place the verb at the end of a clause, others at the beginning, still others place it somewhere in the middle. Finally, many languages seem to place the verb just about anywhere.

Among the nominal ("noun-like") elements in a clause, an important distinction has traditionally been made between subject and object (abbreviated S and O in early typological research).[1] In terms of tree diagrams, you can think of the subject as the DP that is directly under the S node, and the object as the DP that is directly under the VP node:

(2)



It turns out that there is a very major **TYPOLOGICAL** distinction between languages in which the object follows the verb (VO languages), and those in which the object precedes the verb (OV languages; Greenberg 1963, inter alia). In terms of phrase structure rules, this can be thought of as a distinction between languages, like English, in which the VP rule has an optional DP following the head verb, and others, like Japanese, in which the optional DP precedes the head verb:

(3)

VP	→	(DP) V	OV Languages (Japanese, Finnish, Hindi . . .)
VP	→	V (DP)	VO Languages (Mandarin, Indonesian, English . . .)

What is interesting about this typology is that the order of object and verb in the verb phrase tends to correlate with other aspects of the syntax of the language. For example, if a language has OV order, it will almost certainly have postpositions, rather than prepositions. Conversely, if a language has VO order, it will almost certainly have prepositions. Also, in OV languages, inflected auxiliaries almost always come after the verb, whereas in VO languages, auxiliaries usually precede the verb. In short, there are, generally speaking, two major types of languages in the world: those in which syntactic heads normally precede their complements, and those in which syntactic heads follow their complements:

(4)

		VO languages:		OV languages:	
		Head	Complement	Complement	Head
VP	→	V	DP	DP	V
IP	→	AUX	VP	VP	AUX
PP	→	P	DP	DP	P
DP	→	D	NP	NP	D

Because every language is always in a state of change, and the order of head and complement in a particular phrasal category is one variable that may change over time, these correlations are not absolute. However, they are highly significant from a statistical point of view. It is certainly not mere coincidence that languages correlate in this way. The problem for linguistic theory is *why* this should be the case. Many linguists have approached this problem from different directions, and

Table 7.1 *Summary of Greenberg's Universals (from appendix 2 of Greenberg 1963)*

Greenberg's Universal	Parameter	correlation	
#1	Main clauses	V-O	O-V
#3,4	Adpositions	Prepositions	Postpositions
#2	Genitive (possessor) and head noun	N-G	G-N
#17	Head noun & adjective	N-Adj	Adj-N
#24	Relative clauses and head noun	N-RelCL	RelCL-N
#22	Comparatives	Adj-Mkr-Std	Std-Mkr-Adj
#16	Inflected auxiliaries	Aux-V V-Aux	
#9	Question particles	Sentence-initial	Sentence-final
#12	Question words	Sentence-initial or elsewhere	Sentence-initial
#27	Affixes	Prefixes	Suffixes

we will not try to summarize these here. Rather, we will simply present the findings of some very important foundational research, and then give some examples of languages that represent each of the major types.

The foundational work in syntactic typology was done by Joseph Greenberg in the early 1960s. Greenberg compared the syntactic characteristics of thirty languages and found several interesting correlations. In particular, he noticed that the languages in his sample tend to have a basic, or unmarked, syntactic structure, and that the order of certain elements in this basic structure correlate with the orders of other elements. Table 7.1 summarizes the correlations that Greenberg (1963) observed for VO and OV languages. These have come to be known as "Greenberg's Universals," since they were assumed to represent correlations that hold true universally, i.e., for all languages.

It is important to recognize that Greenberg simply observed certain correlations. He did not attempt to provide a reason for (i.e., to "motivate") those correlations, or even to test them for statistical significance. In this sense, Greenberg did not attempt to *predict* constituent orders in as yet unstudied languages. Since 1963, much research has revealed problems with Greenberg's original typology. Significant revisions, criticisms, and extensions of Greenberg's work are found in Hawkins (1983), D. Payne (1985), Mithun (1987), Dryer (1988, 1992), and Hawkins (1994). In an important correction, Dryer (1988) shows that Greenberg's Universal number 17 (the order of adjective and head noun) does not hold when a larger sample of languages is considered. Nevertheless, Greenberg's work stimulated the field of typological linguistics and has continued to be very influential.

In the following sections we will explain and illustrate some of the correlations described in table 7.1, using examples from two typologically distinct languages – Japanese and Malagasy.

Examples of an OV and a VO language

Japanese and Malagasy are two languages that conform to Greenberg's observations very closely. Most languages are not this ideal, but these will suffice to exemplify a rather remarkable recurring pattern in the syntactic structures of the world's languages.

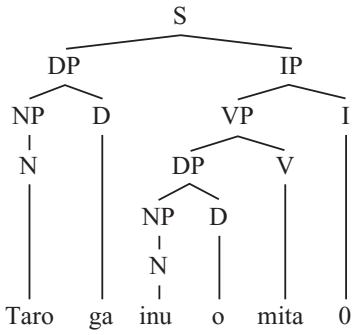
The following example illustrates OV constituent order in Japanese. In this clause, the inflected VP is in brackets. Notice that the object, *inu*, precedes the Verb, *mita*:

(5)

O V:	Taro	ga	[	inu	o	mita	]	'Taro saw a dog.'
	Taro	NOM		dog	ACC	saw		

Here is a possible tree diagram of this clause:

(6)



Notice that under the IP node, the tree "branches" to the left. The syntactic head of each phrasal category is on the right, and its complement branches off to the left. In tree diagrams of English sentences, you will notice that the branching tends to extend to the right (see, e.g., example 72, in chapter 6). For this reason, languages like Japanese are sometimes called **LEFT-BRANCHING** languages, in contrast to English and other VO languages, which can be termed **RIGHT-BRANCHING** languages. Other terms sometimes used for these two language types are head-final and head-initial languages, respectively, or complement+head and head+complement languages, respectively.

In the next Japanese example, we see that the possessor, *hito*, precedes the possessed item, *inu*, in a noun phrase. The grammatical relation that corresponds most closely to the functional notion of "possessor" is GENitive. Therefore, in Greenberg's terminology, Japanese employs GENitive+Noun order in the noun phrase:

(7)

GEN N:	Taro	ga	[hito	no	inu	o	]	mita
		NOM	man	GEN	dog	ACC		saw

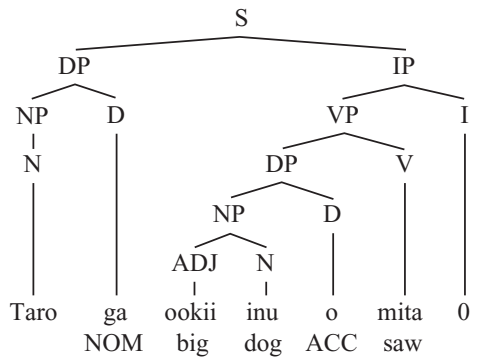
'Taro saw the man's dog.'

In example 8 the word *ookii*, meaning 'big,' comes before the noun that it modifies. Therefore, Japanese exhibits ADJective + Noun order in noun phrases:

(8) ADJ N: Taro ga [ookii inu o] mita. 'Taro saw a big dog.'
NOM big dog ACC saw

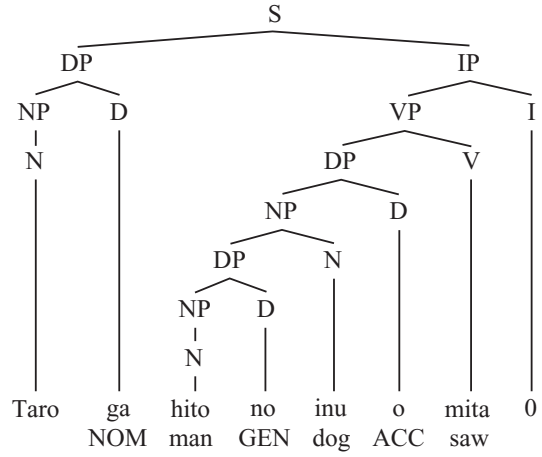
In both of the above examples we see again that branching extends to the left in Japanese. Here are some corresponding tree diagrams:

(9)



'Taro saw (or looked at) a/the big dog.'

(10)



'Taro saw (or looked at) a/the man's dog.'

Notice that under the IP node of example 10, there is a VP, then a DP, then an NP, then another DP. Among many other facts, this diagram captures the important fact that genitive nouns (those for which the determiner in Japanese is *no*) are nested within other NPs. In other words, genitive nouns are a kind of modifier, or optional complement, of other nouns.

Example 11 illustrates a **RELATIVE CLAUSE** in Japanese. Relative clauses are clauses that modify nouns and are embedded within noun phrases. We will have much more to say about relative clauses in chapter 10. For now, just notice that the Relative Clause (bracketed by {curly braces}) comes before the Noun it modifies, *inu*:

(11) RC N: Taro ga [{niku o tabeta} inu o] mita
NOM meat ACC ate dog ACC saw
'Taro saw the dog that ate the meat.'

Japanese has *post*positional phrases rather than *pre*positional phrases. This is another common characteristic of verb-final languages. In 12 we see the POST-position *kara* following the Noun it is related to:

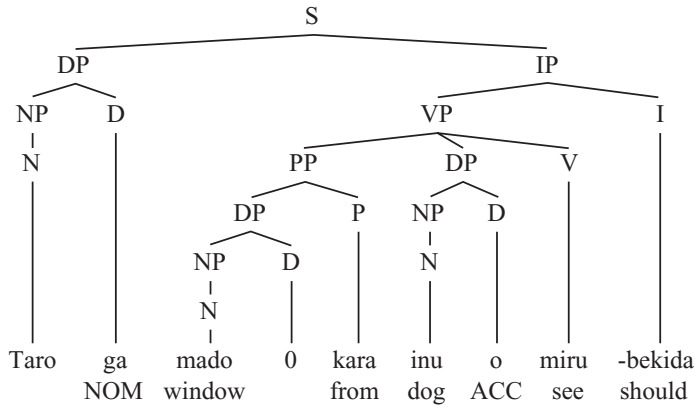
(12) N Postposition: Taro ga [mado kara] inu o mita
NOM window from dog ACC saw
'Taro saw a dog from the window.'

AUXiliaries in Japanese also follow the main Verb:

(13) V AUX: Taro-ga inu o [miru bekida]
-NOM dog ACC see should
'Taro should see a dog.'

Here is a plausible tree diagram of a Japanese clause with a postpositional phrase and an inflectional element (something like an auxiliary) following the verb:

(14)



'Taro should see a dog from the window.'

Now we will look at some examples from Malagasy, a language that in many respects exhibits syntactic structure that is the "mirror image" of Japanese.

In Malagasy the verb normally comes first in the clause. Then comes the O, and finally the subject (examples courtesy of Andoveloniaina Rasolofo):

(15) V O: [Nahita alika] Rashu. 'Rasoa saw a dog.'
saw dog Rasoa

In the noun phrase, the GENitive (possessor) follows the possessed Noun:

(16) N GEN: Nahita [ni alika n'ilai rangahi] Rashu
saw the dog the man Rasoa
'Rasoa saw the man's dog.'

Also, ADJectives follow their head Nouns:

(17) N ADJ: Nahita [alika be] Rashu 'Rasoa saw a big dog.'
saw dog big Rasoa

Auxiliaries in Malagasy come before the Verb:

(18) AUX V: [afaka maita] alika be Rashu 'Rasoa can see a big dog.'
can see

Again, Malagasy is exactly the opposite of Japanese in placing relative clauses after their head Nouns. In this example, the head noun is *alika*, 'dog,' and the Relative Clause which modifies it follows:

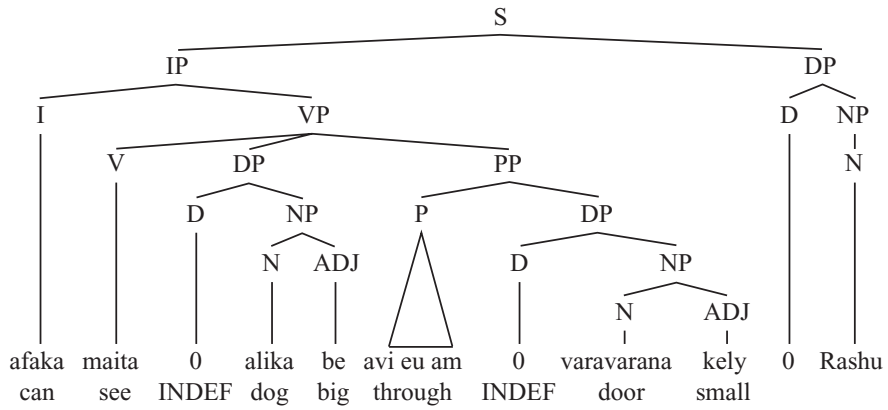
(19) N RC: Nahita ilai [alika { nihinana ilai hena }] Rashu
saw the dog ate the meat Rasoa
'Rasoa saw the dog that ate the meat.'

Finally, Malagasy exhibits PREpositions rather than postpositions:

(20) PREP N: Nahita alika [avi varavarana kely] Rashu
saw dog through door small Rasoa
'Rasoa saw a dog through the window.'

Thus we see that Malagasy exhibits exactly the mirror image of Japanese in terms of basic syntactic structure. Here is a possible tree diagram of a Malagasy clause that illustrates all of the phrasal structures we have mentioned:

(21)



'Rasoa can see a big dog through a window (lit. small door).'

Notice that in this diagram the branches under the IP node extend to the right, rather than the left. In this respect, Malagasy can be considered a "right-branching" language.

It is important to emphasize two facts about the pioneering work of Greenberg. First, most languages are not as consistent with Greenberg's general findings as Japanese and Malagasy are. Second, Greenberg's sample was quite inadequate in a number of respects. More recent work (e.g., Dryer 1988, 1992) has shown that some of Greenberg's observations simply cannot be sustained when a larger, more statistically significant, sample of languages is taken into account.

Pragmatic constituent order languages

While the work of Greenberg (1963) and other early typologists was instrumental in establishing the fields of typological and functional linguistics, there were several conceptual problems with this early work. In particular, one of the assumptions was that all languages employ constituent order to express grammatical relations. In other words, the very use of the terms SVO, SOV, and others to describe language types takes it for granted that "Subject" and "Object" are the relevant terms that determine the order of words in languages. This does hold true for English and many other languages, but is it necessarily true for all? Could it be the case that a language may use the order of words in a clause to express some other communicationally important information? What would such a language be like?

As a thought experiment, imagine a language that uses word order to express relative size rather than grammatical relations. In this language, when a clause describes two participants interacting, the larger participant is mentioned first, and the smaller one is mentioned later, regardless of which one is subject and which is object. Therefore, a clause like the following would be ambiguous:

(22) HYPOTHETICAL DATA: Bear kill man.

By putting *bear* early in the clause, the speaker is asserting only that the bear is larger than the man, not whether the bear is the subject or the object of the clause. This sentence could conceivably mean 'The bear kills the man' or 'The man kills the bear.'

This kind of ambiguity is not often tolerated in languages, because it is dysfunctional – expressing who acts and who or what is acted upon is such an important communicative task that it is not likely to be ignored by the morphosyntax of any natural language. Therefore, if the language is determined to use word order to distinguish size, it would probably come up with some other solution to the problem of expressing which participant is the actor, and which one is the affected participant.

How about a morphological solution? This is the kind of solution illustrated in chapter 1 with the Naga language:

HYPOTHETICAL DATA:

(23)	a.	Bear-a kill man.	'Bear kills man.'	(A suffix -a marks the AGENT.)
	b.	Bear kill man-a.	'Man kills bear.'	
(24)	a.	Bear kill man-p.	'Bear kills man.'	(A suffix -p marks the PATIENT.)
	b.	Bear-p kill man.	'Man kills bear.'	
(25)	a.	Béár kill màn.	'Bear kills man.'	(High tone marks AGENT; low tone marks PATIENT.)
	b.	Bèàr kill mán.	'Man kills bear.'	

The examples in 23, 24, and 25 illustrate three possible solutions to the problem of how to express the AGENT and PATIENT if word order is used for some other purpose. Of course, any of the morphological, lexical, or syntactic expression types discussed in chapters 1 and 2 may be used. These are just random possibilities, out of an infinite number. We will have a lot more to say about how languages actually do express grammatical relations and semantic roles in chapters 8 and 9.

As you may suspect by now, there are in fact languages that use word order for purposes other than to express grammatical relations. There are probably none that use word order to express relative size, though such a language is conceivable. What is much more common are languages that use the linear order of words in clauses to express **PRAGMATIC** information, such as **IDENTIFIABILITY**, **TOPICALITY**, **REFERENTIALITY** and others. In this section we will give a few examples from such languages, and then present some suggestions for how to analyze pragmatically based constituent order languages.

In Ngandi (Heath 1978:206 as cited in Mithun 1987), constituent order is governed by the following principle:

(26) New, indefinite, or otherwise "newsworthy" information is placed early in the clause. Given, definite, or otherwise already introduced information is placed later.

In example 27 from Ngandi, we see subject-verb order when the subject refers to a non-specific, newly mentioned participant (27a), and verb-subject order when the subject refers to a specific identified item (27b):

(27) a.

	Subject	Verb
Načuweleñ-uŋ	gu-jark-yuŋ	gu-ja-walk, . . .
then-ABS	GU-water-ABS	GU-now-go:through

'Then water passes through,' (first mention of water)

b.

	Verb	Subject
Načuweleñ-uŋ	gu-ja-geyk-da-ni	gu-jark-yuŋ
then-ABS	GU-now-throw-AUG-PR	GU-water-ABS

'Then the water rushes through,' (subsequent mention of water)

Coos (Frachtenberg 1913:7) also follows this "indefinite early" principle. In example 28a, the matting is the affected participant, and it comes before the verb when it is mentioned for the first time. In 28b, the matting is again the affected participant, but this time it comes after the verb, because in this clause the matting has already been mentioned in the discourse, and therefore can be taken as **GIVEN INFORMATION**:

(28) a.

Object	Subject	Verb
TE tc!i'cil	yüL is	yö'qat . . .
that matting	we two	split:it

'Let's split this mat.' (first mention of mat)

(they did so, and went down to examine the earth. The earth was still not solid, even . . .)

b.

			Verb	Object	
i	lau	tci	uxhi'touts	hE	tc!icil.
when	that	there	they:two:put:it:down	the	matting

'after they had put down the mat,' (subsequent mention of mat)

In both of these languages, the positions of all nominal clause constituents (i.e., subject, object, and other elements) are determined to a large extent by pragmatic factors.

For some languages, one nominal element exhibits a fairly fixed position (variable only under extreme pragmatic pressures), while another is more variable. Some languages that operate in this way are:

- Spanish. Fairly fixed verb-object, pragmatically variable subject (Bentivoglio 1983).
- Guaymí (Chibchan, Costa Rica, and Panama). Fixed object-verb, pragmatically variable subject.
- Panare (Cariban, Venezuela). Fixed verb-subject, pragmatically variable object.
- Apuriná (Arawakan, Brazil). Fixed subject-verb, pragmatically variable object (Aberdour 1985).

The areas of the world in which languages seem particularly sensitive to pragmatic ordering principles are the Americas, Australia, and to a lesser extent Austronesia and South Asia. Not enough studies of constituent order in discourse have been conducted in Africa to allow generalizations regarding the sensitivity of African languages to pragmatic principles in constituent ordering (though some interesting work has been done; see, e.g., Watters 1979 and Dooley and Levinsohn 2001). The Slavic languages tend to be the most pragmatically sensitive in the Indo-European family.

It should be emphasized, however, that pragmatic factors influence constituent order in all languages to one degree or another. It is just that in some languages pragmatic factors are so dominant that it is difficult or impossible to describe the

"basic" constituent order in terms of subject and object. On the other hand, even languages in which pragmatics dominates may show sensitivity to grammatical relations to some extent.

Finally, some languages have pragmatically determined constituent order that is overridden by syntactic considerations only when ambiguity would result. To understand how this might work, let's extend our thought experiment a little. Imagine the hypothetical language represented in example 22, without any of the morphosyntactic expressions of semantic roles proposed in examples 23 through 25. In this language, would a sentence like the following be ambiguous?:

(29) Car drove Lucretia.

Remember that physical size determines the order of constituents, therefore, *car* comes first because it refers to a participant that is larger than Lucretia. Nevertheless, quite apart from word order there is good reason to guess that *Lucretia* refers to the actor, and *car* refers to the affected participant. What is that? Hello?! People drive cars, but cars don't drive people! This is a pragmatic fact about the world that everyone (at least those who understand what cars and people are) implicitly know. When you think about it, the vast majority of two-participant clauses that we use are of this sort – only one of the participants is the pragmatically plausible AGENT. For many common activities, such as reading books, eating apples, preparing meals, cleaning house, sweeping floors, carrying suitcases, etc., morphosyntactic signals are not even needed to distinguish which participant acts and which is acted upon. It is only in those relatively rare situations in which either participant could fulfill either role that ambiguity may result. It is in those cases only that morphosyntax may be needed to express the distinction. For example people both control and are affected by activities such as chasing, insulting, hugging, slapping, etc., therefore clauses that describe these kinds of activities are more likely to be ambiguous in our hypothetical language:

(30) Hypothetical language in which largest participant comes first:
a. Apollo chased Daphne.
b. Cyclops insulted Sinbad.
c. Goliath killed David.

These examples would all be ambiguous in our hypothetical language in which size determines constituent order, whereas an example like the following would not be:

(31) Camel rode Ali.

Why not? Because the examples in 30 describe activities in which either participant could conceivably be the AGENT. In 31, on the other hand, Ali (a man's name) can plausibly ride a camel, but a camel can't plausibly ride Ali. Therefore

there may be (and actually are) some languages that invoke a morphosyntactic solution to the problem of distinguishing participant roles only in situations like 30, but don't bother in situations like 31, because in these cases context and common sense are sufficient to make the intended meaning clear.

How to analyze the syntactic typology of a language

Most linguists would consider the "basic" constituent orders of a language to be exhibited at least in **PRAGMATICALLY NEUTRAL** clauses. Pragmatically neutral clauses are those that do not present any part of the clause as being unusually highlighted, emphasized, or contrasted. For example, a clause like 32 is not pragmatically neutral in English:

(32)

O	S	V
Beans	I	like.

A clause like this places special contrast on the O argument, as illustrated in a context such as *Beans I like*; *rice I hate*. We would not want to analyze the syntactic typology of English based on this kind of clause. If we did, we would probably say that English is an OV language, which is clearly incorrect. Similarly, we wouldn't want to use sentences like the following:

(33)

	V S	
a.	Once there was a Hobbitt.	**EXISTENTIAL CLAUSE**
	O S V	
b.	Whom did Frodo see?	Question
	V S O	
c.	Have you a match?	Question

All of these structures are **PRAGMATICALLY MARKED**. That is, they are used only in special circumstances in a conversation, such as when participants are being brought onto the discourse stage for the first time, or when some specific piece of information is being questioned. As you can see, these English clauses express unusual constituent orders (VS, OSV, and VSO). For this reason, we would not want to use examples such as these to determine the "basic" syntactic typology of a language. Instead we want to use pragmatically neutral clauses.

However, identifying one clause type as pragmatically neutral is often difficult. A general way to approach this problem is to start with a large corpus of texts (stories, transcribed conversations, or other discourse types) and eliminate from consideration clause types that are known to exhibit marked constituent orders in some languages. These would include:

- Dependent clauses (see chapter 10)
- Clauses that introduce participants onto the discourse stage (33a)

- Questions (33b, c)
- Negative clauses
- Clearly contrastive clauses (e.g., 32)
- Clauses in which a pronoun is used to express O and/or S (i.e., basic constituent order concerns order of **FULL NOUNS** and verbs).

Here is an example from Spanish to show that clauses with pronouns can exhibit unusual orders. In Spanish, the basic order when S and O are full nouns is SVO (34a). However, as is the case in many Romance languages, when O is expressed as a clitic pronoun, it comes right before the inflected verb (34b):

(34)	a.	Frodo	vió	a	Gandolfo.	'Frodo saw Gandolph.'
		Frodo	see.PAST.3SG	DAT	Gandolph	
	b.	Frodo	*lo* vió.	(*Frodo vió lo).		'Frodo saw him.'
			3SG			

Once all the clause types listed above have been eliminated from consideration, it is probable that the clauses that remain are largely pragmatically neutral. If in these remaining clauses there are examples of transitive verbs with full NPs expressing S and O, *and* if those NPs exhibit a consistent order relative to the verb, then that order can be considered basic. Unfortunately, in reality this is a rare situation. Once you have eliminated all of the clause types listed above, you are often left with very few clauses indeed. Such clauses as are left often lack overt expression of one or more **CORE ARGUMENTS** (S or O in Greenberg's terms). D. Payne (1986) has observed that pragmatically neutral clauses tend to consist of a verb and one or fewer noun phrases. In many languages use of any full nouns in discourse is pragmatically marked.

If you can't decide on a basic constituent order using the above criteria, the language is probably one of the many languages of the world that employs constituent order to express pragmatic status. In this case you will need to conduct a statistical study of the use of the various constituent order possibilities in discourse. Analytical methods for conducting such studies lie outside the scope of the present book. However, if you go on in linguistics you will undoubtedly have opportunities to take courses in discourse analysis. In those courses you will learn, among other things, how to conduct a study to determine the pragmatic principles that underlie the ordering of constituents in any language.

There is one last point that needs to be made concerning how to determine the basic constituent order of a language. The orders of elements within verb phrases, noun phrases, or adpositional phrases is not evidence for a particular order in main clauses. For example, Greenberg observed that languages with postpositions are always (in his sample) of the OV type. However, if we know the language has postpositions rather than prepositions, we cannot use Greenberg's observations to claim that the basic order in main clauses must be OV. Greenberg did not make predictions – only observations based on a very small sample. Languages are too

often inconsistent for us to take non-main-clause orders as evidence for main-clause constituent order.

Conceptual outline of chapter 7

I. Languages can be typologized (classified into types) according to a number of parameters. The most interesting typological parameters are those that describe "clusters" of grammatical properties. The two typologies discussed in this chapter are:
- Morphological typology
- Syntactic typology (also known as "constituent order typology")

II. Morphological typology consists of two parameters:
- The index of synthesis (the number of morphemes per word)
- The index of fusion (the number of meanings per morpheme)

III. There are three broad types of languages according to their syntactic typology:
- Languages in which syntactic heads follow their complements (called OV languages, left-branching languages, head-final languages, or complement+head languages).
- Languages in which syntactic heads precede their complements (variously termed VO languages, right-branching languages, head-initial languages, or head+complement languages).
- Languages for which constituent order is determined by some principle other than grammatical relations.

IV. In the last type of language, pragmatic statuses such as referentiality, identifiability, and contrastiveness are the functional variables most likely to be expressed by constituent order. In these languages, the semantic roles of participants must be expressed in some other morphosyntactic way, at least in those situations where more than one participant could plausibly be an AGENT.

V. In order to determine the "basic constituent order" (if any) of a language, it is important to isolate pragmatically neutral clauses.

Exercise 7.1: Yagua

Tom Payne and Matthew Dryer

Yagua is a **LANGUAGE ISOLATE** spoken by about 4,000 people in northeastern Peru.

A. On the basis of the examples on the following page, identify whether Yagua is basically an O-V or a V-O language, and whether it is S-V or V-S. Give the evidence for your claims.

B. List the other constituent order characteristics of Yagua illustrated in these data. For each characteristic, indicate whether it is expected, unexpected, or neither, given the basic order of V and O identified above. Cite examples that illustrate each of your claims.

1. a. Sa-munaa-dee Alchíco. 'Alchico's placenta.'
   3SG-placenta-DIM Alchico
   b. Alchíco munaadee. 'Alchico's placenta.'
   *Munaadee Alchíco, *Alchíco samunaadee.
   c. Samunaadee. 'His placenta.'
2. Jirya munaadee. 'This placenta.'
   *Munaadee jirya.
3. Tɨnkɨɨ munaadee. 'One placenta.'
   *Munaadee tɨnkɨɨ.
4. Samunaadee kúútya. 'His placenta whispers.'
   *Kúútya samunaadee.
5. Sakúútya Alchíco munaadee 'Alchico's placenta whispers.'
6. Jíryoonú sʊ́ʊ́y-anú sa-roori-myú Alchíco-níí
   bushmaster bite-PAST 3SG-house-LOC Alchico-3SG
   'A bushmaster (snake) bit him in Alchico's house.'
   *Jíryoonú sasʊ́ʊ́yanuníí. ('A bushmaster he bit him.')
7. Sa-sʊ́ʊ́y-anú jíryoonu Alchico roori-myú-níí
   3SG-bite-PAST bushmaster-3SG Alchico house-LOC-3SG
   'A bushmaster (snake) bit him in Alchico's house.'
8. są-ą rą́ą́-kyu. 'He will jump!'
   3SG-FUT jump-POT
   *rą́ą́kyu sąą, *sarą́ą́kyu sąą, *sarą́ą́kyu ą.
9. Sa-niy suvʊ́-tyąa jiñu munátya sʊ-ʊmutẹẹsá
   3SG-MALF fear-INTS this ancestor 3SG-behind
   munaa játiy sa-rẹẹ-ñíí.
   placenta REL 3SG-jump-3SG
   'This ancestor is really afraid behind the placenta that makes him jump.'
10. Rą-a ją́ą́-charatá jiyu-dáy koodí-vyiimú.
    1SG-FUT fall-might here-DAY snake-inside
    'I might fall here inside a snake.'

Exercise 7.2: Tshangla

Eric Andvik

1. Ja-ga ata yigi ringmu thur dri-ba.
   1SG-DAT eld.brother letter long one write-PAST
   'My elder brother wrote one long letter.'

2. Ro-ka gari otha phai yanglu jap-kai tsuk-pa cha.
   3SG-DAT car that house green behind-ABL put-PAST is
   'His car is parked behind that green house.'

3. Ja-ga usin-ga chharo nan-gi ye-khan
   1SG-DAT young.sis-DAT friend 2SG-ERG speak-REL
   echha ngo-le re-be.
   book buy-NPAST can-NPAST
   'My younger sister's friend can buy the book that you talked about.'

A. Where is Tshangla spoken? What is its genetic affiliation? How many speakers are there?

B. Describe the head and complement orders in the various syntactic constituents illustrated. In what respects do these data conform to Greenberg's (1963/1966) observations, and in what respects do they *not* conform? Note any ambiguous or problematic data.

Exercise 7.3: Shugnan

M. E. Alexeev, adapted by Tom Payne

Here are some noun phrases in Shugnan and their translations into English:

1. kuzaa hats 'jar of water'
2. chalak zimaadj 'bucket of dirt'
3. tambal byuyun 'beard of a lazybones'
4. biig dyuyunaa 'pot of corn'
5. kuzaa gjev 'lid of a jar'
6. beechoraa zimaadj 'dirt of a beggar'

A. What language family does Shugnan belong to, and where is it spoken?

B. Translate into Shugnan:

7. 'bucket of water'
8. 'corn of a beggar'
9. 'jar of a lazybones'

C. What determines the order of genitive and head noun in Shugnan?

Exercise 7.4: translating trees into bracketing

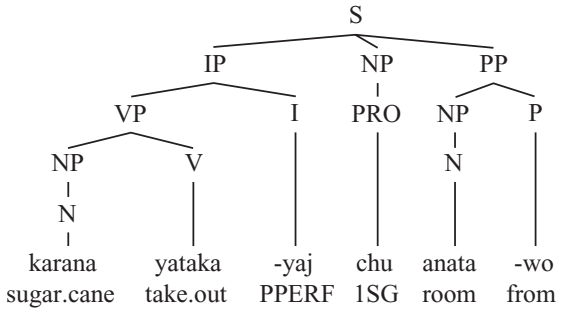
Tom Payne

The following are constituent structure trees from three typologically distinct languages.

A. "Translate" each tree into a labeled and bracketed string.

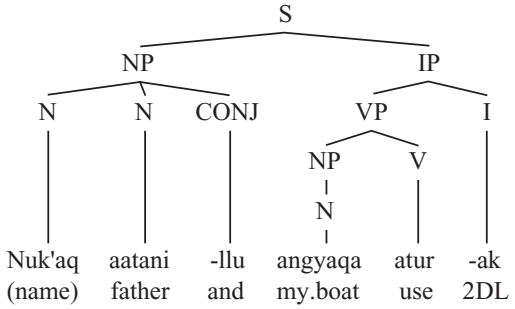
B. For each language, indicate its syntactic type (head+complement or complement+head).

a. Panare: ‘I took the sugar cane out of the room.’



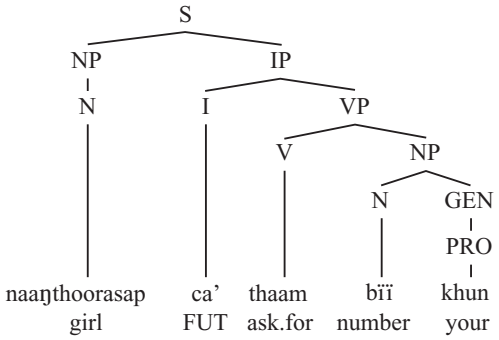
Labeled, bracketed string: Karana yataka -yaj chu anata -wo
Language type: _____

b. Yupʹik: ‘Nukʹaq and father are using my boat.’



Labeled, bracketed string: Nukʹaq aatani -llu angyaqa atur -ak
Language type: _____

c. Thai: ‘The girl will ask for your number.’



Labeled, bracketed string: naaŋthoorasap ca’ thaam bïï khun
Language type:

Exercise 7.5: translating bracketing into trees

Tom Payne

The following are labeled and bracketed sentences from three typologically distinct imaginary languages.

A. "Translate" these strings into constituent structure trees. Be sure to label each node correctly.

B. Give a plausible gloss and free translation for each example.

a. $[_{s}\ [_{ip}\ [_{aux}$ vro$]_{aux}\ [_{vp}\ [_{v}$ olin$]_{v}\ [_{dp}\ [_{np}\ [_{n}$ fim$]_{n}]_{np}\ [_{d}$ gron$]_{d}]_{dp}]_{vp}]_{ip}\ [_{dp}\ [_{np}\ [_{n}$ smu$]_{n}\ [_{adj}$ nid$]_{adj}]_{np}\ [_{d}$ rad$]_{d}]_{dp}]_{s}$

b. $[_{s}\ [_{np}[_{n}$ olnik$]_{n}]_{np}\ [_{ip}\ [_{np}\ [_{adj}$iops$]_{adj}\ [_{n}$ poms$]_{n}]_{np}\ [_{vp}\ [_{v}$ anterrettim$]_{v}]_{vp}\ [_{aux}$ bim$]_{aux}]_{ip}]_{s}$

c. $[_{s}\ [_{ip}\ [_{i}$ e$]_{i}\ [_{vp}\ [_{v}$ apa'a$]_{v}\ [_{dp}\ [_{d}$ ek$]_{d}\ [_{np}\ [_{n}$ eliam$]_{n}\ [_{adj}$ gib$]_{adj}]_{np}]_{dp}\ [_{pp}\ [_{p}$ iva$]_{p}\ [_{dp}\ [_{d}$ ila$]_{d}\ [_{np}\ [_{n}$i' ot$]_{n}\ ]_{np}]_{dp}]_{pp}]_{vp}]_{ip}\ [_{dp}\ [_{d}$ ila$]_{d}[_{np}\ [_{n}$anar$]_{n}\ [_{adj}$ ipso$]_{adj}]_{np}]_{dp}]_{s}$

Exercise 7.6: a small grammar of Japanese – an OV language

Mitsuyo Hamaya, Naoaki Tai, and Tom Payne

A. Fill in the missing translations of clauses 12 (Japanese) and 13 (English).

B. Write phrase structure rules and a lexicon that will sanction all of the following clauses of Japanese.

C. Give three more clauses that your rules and lexicon allow, and three strings that would be ungrammatical.

1. Taro ga hashitta.	'Taro ran.'
2. Taro ga ookina inu o mitsuketa.	'Taro found the big dog.'
3. Mitsuyo ga neta.	'Mitsuyo slept.'
4. Taro ga ookina koen de hashitta.	'Taro ran in the big park.'
5. Mitsuyo ga boru o nageta.	'Mitsuyo threw the ball.'
6. Taro ga boru o koen de nageta.	'Taro threw the ball in the park.'
7. Taro ga nageta.	'Taro threw (something).'
8. Taro ga inu to hashitta.	'Taro ran with the dog.'
9. Taro ga Mitsuyo ni ookina boru o nageta.	'Taro threw the big ball to Mitsuyo.'
10. Mitsuyo ga inu ni gohan o koen de ageta.	'Mitsuyo gave rice to the dog in the park.'
11. Mitsuyo ga inu ni hone o ageta.	———–
12. ———–	'Taro gave the ball to the dog.'
13. Taro ga Mitsuyo ni inu o ageta.	

Exercise 7.7: Hungarian

A. N. Zhurinsky, adapted by Tom Payne

Here are six Hungarian sentences (in the official alphabet) and their translations into English:

1. Az asztalon a térkép van.	'The map is on the desk.'
2. Az asztalokon a térképek vannak.	'The maps are on the desks.'
3. A füzetnél az újság van.	'The newspaper is near the notebook.'
4. Az újságokon a füzetek vannak.	'The notebooks are on the newspapers.'
5. Az ablakoknál a pad van.	'The bench is near the windows.'
6. A székeken a kasok vannak.	'The baskets are on the chairs.'

A. Translate the following English sentences into Hungarian (Hints: Remember that Hungarian has vowel harmony. There is no irregularity in this problem):

7. The notebook is on the desk.
8. The newspapers are on the notebook.
9. The chairs are near the desk.
10. The benches are near the chairs.
11. The basket is on the window.

B. Describe the difference in usage between *a* and *az* in Hungarian.

C. What do you know about the basic constituent orders of Hungarian, just based on these data? Is this consistent or inconsistent with Greenberg's (1963) observations?

Note

1. It is perhaps an unfortunate fact of the history of linguistics that multiple terminologies have been used in the domain of grammatical relations. The focus in the present chapter is Greenberg's (1963) pioneering work on syntactic typology. In that work, and in much subsequent research, the terms S, O, and V were used to refer to the notions of Subject, Object, and Verb respectively. It is important for all students of linguistics to be aware of these terms, and be able to use them confidently. However, later work revised this terminology significantly, and we will be discussing the newer terminologies starting in chapter 8. It is also important not to confuse the term S as a syntactic category (the "highest level" category label) with S as an abbreviation for "subject."