Part One

Introduction

1. Grammatical Constructions, Semantic Classes and Information Packaging

1.1. What is morphosyntax?

The term **morphosyntax** refers to the combination of morphology and syntax. **Syntax** is the analysis of the internal structure of utterances/sentences, more specifically, how words are put together. **Morphology** is the analysis of the internal structure of words, including prefixes, suffixes and other internal changes to words that generally have a meaning (elusive as that meaning sometimes is). Therefore, **morphosyntax** is the analysis of the internal structure of utterances, both above the word level and below it.

Why combine morphology and syntax? Because grammatical **constructions** involve both. Consider the examples of the English Numeral Modification Construction in (1):

(1) *English Numeral Modification*:

one tree
two tree-s
three tree-s
etc.

The English Numeral Modification Construction involves both syntax—the order of numeral and noun—and morphology—the form of the noun, singular or plural. A description or analysis of the English Numeral Modification Construction must include reference to both: the relative position of numeral and noun, and the inflection of the noun for number. A construction is often represented schematically, in this case as [NUM NOUN-NMB]: the abbreviations NUM, NOUN and -NMB represent categories of words (*one, two, three* etc. for NUM; *tree, bush* etc. for NOUN) or bound morphemes (-NMB for the number suffix). NUM, NOUN and -NMB are also described as **roles** in the construction (Croft 2001:11, 24, 175-76, drawing on unpublished work by Paul Kay; also called a 'function' or 'slot').

Of course, some constructions in languages seem to involve "only syntax": order and grouping of words. Other constructions seem to involve "only morphology": the inflectional forms of words, for example.

Another reason to combine syntax with morphology is that bound morphemes almost always originate in free words that originally combined with other words into constructions. Those constructions were reduced by the process of **grammaticalization** (see §2.3). An example of grammaticalization in progress in English can be seen in the contracted forms of auxiliaries and negation: *will not > won't, I am > I'm,* etc. As a result, we will see the same sorts of meanings and semantic combinations in stem+inflection combinations that we also find in multiword constructions. In fact, it is sometimes difficult to draw the line between syntactic constructions and morphological constructions: language change, including grammaticalization, is gradual.

The English Numeral Modification Construction does not consist solely of a morphosyntactic form. The construction also conveys a **meaning**, that is, **semantic content** or **information content**. The noun denotes a set of individuals of the noun

category (a set of trees), and the numeral specifies the cardinality of that set (one, two, three, etc.). In addition, the information is **packaged** so that the construction as a whole denotes the set of the trees, and the specific number of trees is secondary information added about the set. The English Numeral Modification Construction contrasts with a sentence such as *The trees number fifteen*, in which the number of trees functions as the primary information, predicated of the trees. An important hypothesis of this book is that (morpho)syntactic constructions not only communicate meaning (information); they also **package** that **information** in different ways for the purposes of communication. The modification function illustrated in (1) is an example of information packaging; see §1.3 and §1.5.

The focus of this textbook is primarily on **syntax**, that is, **how words are put together into utterances, and what those combinations of words mean**. We will not discuss those aspects of morphology that have to do with the phonological form of morphemes, such as bound vs. free morphemes, morphophonological processes, conjugational or declensional classes and word formation (for morphology, see Haspelmath and Sims 2010; for word formation, see Štekauer et al. 2012). Our focus will be on morphology that serves syntax.

Another reason for the focus on syntax in this textbook is that much of the description of morphological meaning is typically covered in courses on semantics. While a case can be made that linguistics curricula should be organized in terms of a full-year (or longer) sequence that provides a survey of morphology, syntax and semantics combined, I will proceed on the assumption that most linguistics curricula divide (morpho)syntax from semantics. Nevertheless, there will be a significant amount of discussion of semantic content in this textbook, since semantics plays a major role in shaping morphosyntax.

This textbook proceeds from three basic assumptions about the analysis of morphosyntax. The first is that the proper unit for grammatical analysis is a (morphosyntactic) **construction**, such as the numeral modification construction described above. The second assumption is that one must always investigate a construction with respect to how its morphosyntactic **form** expresses its **function**, which in our analysis includes both meaning and information packaging.

These two assumptions are shared by **construction grammar** (Fillmore, Kay and O'Connor 1988; Goldberg 1995, 2006; Croft 2001), and the second assumption is characteristic of **functionalist** theories of grammatical structure (e.g. Givón 2001a,b). The basic structure of a construction is presented in Figure 1.1. In the contemporary construction grammar approach, constructions include single-word constructions, such as the construction consisting of English noun stems plus their inflection for number in (1). In fact, a construction is any conventionalized pairing of form and function. Reasons for using a constructional approach will be discussed in §1.2. The basic analysis of the functions of constructions is outlined in §1.3.

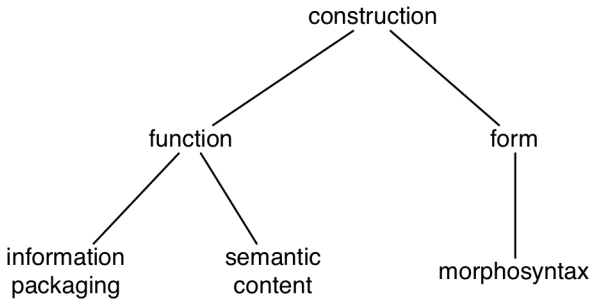

Figure 1.1. The basic structure of a construction.

The third assumption is that one must always examine how the morphosyntactic expression of a function varies across languages. The third assumption, combined with the first two, is the hallmark of **linguistic typology**. Linguistic typology is the study of the diversity of languages across the world, and universals that structure that diversity. Linguistic typology has primarily though not exclusively examined the diversity of grammatical structures in the world's languages. In fact, much of the content of this book describes the results of more than half a century of research on crosslinguistic variation and universals of grammatical structure.

There are several reasons for taking a typological approach. The study of syntax shoud and must be inclusive, that is, accommodate the diversity of grammatical structures in all languages from all cultures. Perhaps most important, we can understand why grammatical structures are the way they are, even in a single language, only by placing them in the context of patterns of global syntactic diversity. Also, even a single language contains variation, and all languages are changing all of the time. The way a language varies and has changed, and will come to vary and change, is an instance of the cross-linguistic variation and change that is described in this textbook. The basis for doing crosslinguistic comparison is discussed in §1.4, and some basic concepts for crosslinguistic comparison, also known as comparative concepts, are introduced in §1.5.

The relationship between grammatical, i.e. morphosyntactic, form and the function expressed by that form is a very complex one. There is no simple one-to-one mapping between the function to be expressed and the morphosyntactic structure of an utterance. This poses major challenges in organizing a textbook such as this one. The primary reasons for this complexity are given in §§2.3-2.4. The motivation for the organization of this textbook will be presented in this chapter and chapter 2.

1.2. Why constructions?

In this section we will discuss some of the reasons for adopting the constructional approach. To do, however, we must first discuss word classes and their relationship to semantic classes on the one hand and constructions on the other (a more detailed discussion can be found in Croft 2001, chapters 1-2).

1.2.1. Word classes and semantic classes

Most descriptions of both familiar European languages and less familiar languages from other parts of the world use **word class** terms, also known as parts of speech or syntactic categories, as a means to capture morphosyntactic patterns: 'Parts of speech tell us how a word is going to function in the sentence' (Carnie 2013:44). The major word classes of English and other long-studied languages have been established in the Western grammatical tradition for a long time: noun, verb, adjective, adverb. Many other word classes have also been proposed by linguists, recently and not so recently. However, problems with the use of word classes to capture grammatical patterns arise when facing the grammatical diversity found across the languages of the world.

When one reads grammatical descriptions of lesser-known and previously undocumented languages (and even of better-known languages), one often finds that word class terms are used in confusing and seemingly conflicting ways, as seen in the quotations from reference grammars found in (2a-d):

(2) a. 'Sidaama numerals are adjectives' (Kawachi 2007:135)
b. 'Numerals [in Iñupiaq] are a subclass of nouns…numerals behave like nouns…Iñupiaq numerals are nouns' (Lanz 2010:106, 107, 108)
c. 'adjectives [in Mamainde] are encoded as verbs' (Eberhard 2009:324)
d. 'Acehnese has no class of adjectives' (Durie 1985:101)

These quotations challenge students and scholars who are using reference grammars in order to understand syntax across languages or to analyze particular constructions. For the most part, these problems are not due to unsatisfactory or inconsistent application of syntactic analysis to these languages by the authors. They are basically due to problems with using word class terms to describe how words are used in grammatical constructions.

Statements of the form 'Word class X is word class Y', such as the statement about Sidaama in (2a) and the last statement about Iñupiaq in (2b), appear puzzling at first. What is meant by 'Word class X is word class Y'? These statements are meant to be interpreted as follows, for example for (2a): Sidaama numerals and Sidaama adjectives function the same way in a sentence. But if they function the same way in a sentence, then what do the terms 'numeral' and 'adjective' mean in this context? They mean that the Sidaama translation equivalents of English words in the English Numeral class and the Sidaama translation equivalents of English words in the English Adjective class are members of single word class in Sidaama.

That is, the term 'numeral' in (2a-b) are being used to refer to **semantic classes**—translation equivalents—in these quotations: numeral concepts and property concepts respectively. The same is true of 'noun' in the second statement about Iñupiaq in (2b), where 'noun' refers to object concepts (persons and things). It is also true of the statement about 'adjectives' in (2c): (2c) means that the Mamainde translation equivalent of English Adjectives do not form a distinct word class—specifically, they are not distinct from the word class to which the Mamainde translation equivalents of English Verbs belong.

On the other hand, the term 'adjective' in (2a), 'noun' in the first and third statements in (2b), and 'verb' in (2c) are being used to describe a syntactic category (a word class). What is intended can be determined by looking at more careful statements like the second statement in (2b), 'numerals behave like nouns', or the statement in (2c), 'adjectives are encoded like verbs'. In these statements, 'noun' and 'verb' still are being used to express semantic classes of words (object concept words and action concept words respectively). But these assertions more clearly state that words of two different semantic classes belong to a single word class.

The statement in (2d) is also confusing. The term 'adjective' in (2d) is not referring to the semantic class of property concepts. The paraphrase 'Acehnese has no property concepts', would be nonsensical; every language has a way to express property concepts. Even so, (2d) is not entirely straightforward to interpret. What (2d) means is that property concept words in Acehnese do not form a distinct word class—not unlike the assertions in (2a-c).

Because 'noun', 'verb' and 'adjective' are used to mean different things, there are two terminological problems give rise to confusion. The first is that the same term is being used for a syntactic category, that is, a word class defined by their syntactic patterning, and for a semantic category, or more precisely, a class of words determined by their meaning. In the case of 'adjective', there exists a distinct term for the semantic class, namely 'property (concept)', although this term was not used in any of the descriptions in (2). The use of 'adjective' in (2a), (2c) and (2d) is at best confusing—and a more suitable term is available, namely 'property concept words'. For example, (2c) could be rephrased as 'property concept words are encoded as verbs'.

In the case of 'numeral', we have a different kind of problem. Linguists use the term 'numeral' for both semantic class (any translation equivalent for '1', '2', '3', etc.) and a word class (e.g., the English word class which has *one, two, three*... as members). There is no widely accepted distinct pair of terms for the syntactic category and the semantic category when it comes to 'numerals'.

Unfortunately both of these terminological problems are pervasive in linguistic discussions of syntax. The solution to the first problem is simply to be careful and consistent in using semantic terms—'property'—for semantic classes and grammatical terms—'adjective'—for syntactic categories. We will consistently distinguish terms for semantic categories and terms for grammatical (morphosyntactic) categories in this textbook.

A solution to the second problem commonly found in typological writings is to use the lower case form of the term for the semantic class—'numeral'—and to use the capitalized form of the same term for a (language-specific) word class—'Numeral'. This convention has been proposed by a number of typologists, including Lazard (1975), Comrie (1976), Bybee (1985) and Croft (2001). We will follow this rule of thumb for naming (language-specific) word classes in this textbook, even when the terms for word class and semantic class are different (for additional rules of thumb for naming language-specific word classes, see Croft 2016a).

1.2.2. Word classes and constructions

In the preceding section, we were deliberately vague about what a 'word class' or 'syntactic category' actually is: it was described as 'a means to capture morphosyntactic patterns', 'how words are used in grammatical constructions', and as 'defined by their syntactic patterning'. In this section, we will be more precise about what a word class, or more generally a syntactic category is. We will also discuss two important consequences of the definition.[1]

In linguistic analysis, word classes are defined not by their semantics but by their **occurrence in constructions**—more precisely, a word's occurrence in a particular role in a construction. For example, English Adjectives such as *tall* are defined, not as words denoting property concepts like height, but instead in terms of their occurrence in certain English constructions:

(3) a. as modifiers of nouns: *a* ***tall*** *tree*
b. as the complement of a copula *be* in predication: *That tree* ***is tall***
c. they inflect in a certain way (a morphological construction): ***tall-er, tall-est***
d. they can in turn be modified by certain degree expressions: ***very tall, a little tall***).

The term 'word class' presupposes that the linguistic units defined by occurrence in constructions are always words. This is not the case, however. Any type of syntactic unit can be defined in terms of occurrence in constructions. For example, an English Subject phrarse is defined by occurrence in certain English constructions, for example those in (4a-b):

(4) a. Occurs in immediate preverbal position in an Active sentence: ***John*** *congratulated Mary.*
b. Controls the form of the Verb or Auxiliary: ***You are*** *tall* vs. ***She is*** *tall.*

Also, a morphological form smaller than a word, such as a root or a stem, is defined by its occurrence in morphological constructions (sometimes called 'morphological categories'). For example, (regular) English Verbs are defined by their occurrence in morphological constructions such as Third Person Singular Present [__-*s*], Past [__-*ed*], Participle [__-*ing*] and so on. (More abstract descriptions of the constructions would be necessary to include morphologically irregular English Verbs such as *is/are/was/were/being.*) Thus, the comments in this section about 'word classes' are generalizable to syntactic categories that may be larger units than a word or smaller units than a word.

The constructional basis of word classes or syntactic or morphological categories is often obscured by the terms used in syntactic analyses. The construction(s) used to define word classes are called many different things: 'criteria' (Givón 2001a:49; Dixon 2010b:38), 'tests' (McCawley 1998; Carnie 2013:47, 98-100), 'evidence', 'phenomena', 'operation', and 'process' (Mulder 1994:114). The pattern of occurrence of words in

[1]How word classes are defined, and the consequences of how they are defined, is discussed in detail in Croft (2001, chapters 1-4).

certain constructions and not others are said to be the words' 'distribution' (Harris 1951:5; Carnie 2013:47), 'behavior' (McCawley 1998:186), 'properties' (McCawley 1998:18; Evans and Osada 2005:452; Schachter and Shopen 2007:2), 'features' (Amha 2001:89), 'use' (Jagersma 2010:268), or 'function' (Palmer 2009:94).

There is a major shortcoming in using word classes and other syntactic categories for a cross-linguistic approach to morphosyntax—this is the first consequence of the definition of word classes alluded to at the beginning of this section. The constructions defining a word class of a language are also constructions of that language. English Adjectives are constructions of English, not of Sidaama, Iñupiaq, Mamainde or Acehnese. So English Adjective is an English word class; the other languages have their own word classes, defined by their own constructions. If so, then how can we compare English syntax to the syntax of Sidaama, Iñupiaq, Mamainde or Acehnese?

All hope is not lost. Just as one can compare words that are translation equivalents across languages, one can compare functionally equivalent constructions across languages. For example, for *tall tree*, we could compare the functional equivalents in other languages of the construction illustrated for English in (3a), namely the function of modifying a referent. And we can compare word class membership across languages by identifying the semantically equivalent words that occur (or do not occur) in the functionally equivalent constructions in question—for *a tall tree*, a property word meaning 'tall' and an object word meaning 'tree'. In this way one can, for example, compare the constructions used for modification across languages, and observe how property concept words and other semantic classes of words are used in those constructions. This is a fundamental characteristic of a constructional and typological approach to syntactic analysis: using function to identify equivalent constructions and equivalent classes of words across languages. The quotations in (2) show that the mapping between words (translation equivalents) and the constructions (functional equivalents) they occur in varies in complex ways across languages.

There is another wrinkle that has to be addressed before going on, however— this is the second consequence of the definition of word classes (and other syntactic categories). We assumed above that English Adjectives are defined by their occurrence in four constructions—(3a) through (3d) above—not just one. But not all English Adjectives that occur as modifiers—the construction in (3a)—also occur in the other three constructions (and vice versa). Compare *tall* to *alive* in (5), *entire* in (6), *intelligent* in (7), and *even* in (8):

(5) *Modification of a referent:*
a. This insect is **alive**.
b. *an **alive** insect

(6) *Predication with a copula:*
a. An **entire** chapter is devoted to this problem.
b. *This chapter is **entire**.

(7) *Degree inflections:*
a. tall-er, tall-est
b. ***intelligent**-er, ***intelligent**-est

(8) *Degree modifiers:*
a. a very tall tree
b. *a very **even** number

In other words, the different constructions in (3a-d) do not define a single word class of English Adjectives, as shown in (5)-(8). Instead, each construction defines its own distinct word class. Of course, in this case, the word classes overlap; but they cannot be equated without losing important information about the syntax of English. In other words, the mapping between words and constructions in even a single language varies in complex ways.

Again, all hope is not lost for finding patterns and even cross-linguistic universals in this variation. If we recognize that the basic fact is the (complex) mapping between words and their occurrence in constructions—that is, the role a word fills in a construction—then we can look for patterns and explanations for those patterns in **the mapping between words and the relevant roles in the constructions**. For example, a straightforward explanation for the unacceptability of *even* in the degree modifier construction in (8b) is the semantic incompatibility between the degree of a property that is expressed by the degree modifier construction and the meaning of *even (number):* a number is either even or odd. Other differences in word-construction mappings may have historical explanations—explanations that may be idiosyncratic to a single language, or represent recurrent processes of change across languages.

Reference grammars do not always give a full description of the complex mapping between words and constructions. Reference grammars often substitute an enumeration of word classes for a full description of the constructions used to define those word classes, the functions of those constructions, and the range of semantic classes of words that occur in those constructions. Even careful typological analyses often use the same term to describe a semantic class of words and a construction used to describe a word class including that semantic class but not identical to it. An example of this problem is illustrated by the first paragraph of a questionnaire used in a crosslinguistic survey of 'ditransitive constructions' (see §7.5), such as *I gave her a book*:

> *Ditransitive construction* is defined semantically as 'a construction with a recipient (R) and a theme (T) argument', where these semantic role labels are understood broadly. Typical ditransitive verbs are 'give', 'sell', 'show', 'promise', 'teach', but languages may treat other verbs in the same way, so that **these verbs would also count as ditransitive for current purposes** (e.g. in English *deny, envy* [as in *She denied me a kiss; I envy you your success*]; in German *entziehen* 'withdraw from'). (Comrie, Haspelmath and Malchukov 2010:65, emphasis added)

The first part of the paragraph defines ditransitive verbs as a semantic class, namely verbs describing events with a recipient participant and a theme participant (the theme is what the recipient comes to possess), with an illustrative list. But the second part of the paragraph, emphasized here, switches to a definition of ditransitive verbs as any verb that occurs in the ditransitive construction in a language (i.e. the construction that includes

semantically ditransitive verbs), even if they are of a different semantic type—denying, envying and withdrawing events do not have a recipient.

The critical comments about language description in this and the preceding sections are only about consistency in description, not completeness of description. It is impossible to make a complete description of all of the variation in the mapping between words and constructions in a language, especially if one also wants takes into account variation between speakers and even across different utterances of the same speaker, and grammatical changes in progress. A grammatical description is an overview of the major patterns in a language. Likewise, this textbook is an even higher-level overview of the major grammatical patterns found across languages. The reader should not assume that what is presented about the constructions described in this textbook automatically extends to other constructions, even closely related constructions; or that the words that occur in the roles of the constructions described here can be assumed to occur in analogous roles even in related constructions.

We take a thoroughly construction-based approach to syntactic analysis and language description in this textbook. Words occur in certain constructions and don't occur in other constructions. Words have meanings; grammatical constructions have meanings, or functions, as well. We can compare languages by comparing what happens with the translational equivalents of English words, and comparing the functional equivalents of English constructions. Semantic and functional translation equivalence allows us to identify general patterns of syntax across languages, and offer explanations for those general patterns. These explanations are primarily functional and diachronic.

1.2.3. Constructions and the organization of this textbook

The chapters in Parts Two through Four of this textbook present syntactic constructions of the world's languages: what their function is, and the range of morphosyntactic structures used to express those functions. The basis for dividing the presentation of constructions into separate chapters, and the order of chapters in this textbook, is primarily in terms of the function of the constructions. Of course, language is a general purpose communication system: in principle we can express anything we want in language. We have to use an analysis of functions that allows us to divide and organize all the functions that we can express in language into categories that are useful for syntactic analysis. That is, the categories of functions should not only be conceptually related, but also divide constructions into groups that are related to each other by their morphosyntatic form. In the next section, we argue that the analysis of functions most useful for syntax is in terms of information packaging.

1.3. Why information packaging?

A central hypothesis of this textbook is that all constructions convey both semantic or information **content** and a **packaging** of that content (the latter is also called **'**discourse function'). For example, the division of words into those denoting object concepts like 'tree', property concepts like 'tall' and action concepts like 'fall' represents a categorization of words by their semantic content. In contrast, in a sentence like *The tall tree fell,* one can categorize the same words as referring (*tree*), modifying (*tall*) and

predicating (*fell*). This categorization represents the packaging of that semantic content. In the most general syntactic constructions, the packaging of the semantic content is globally organized around the following skeletal structure:

reference - what the speaker is talking about
predication - what the speaker is asserting about the referents in a particular utterance
modification - additional information provided about the referent

These fundamental information packaging functions are called the **(major) propositional act** functions (following Searle 1969 and Croft 1991). These functions are discussed in more detail in §2.1.

The separation of semantic or information content and information packaging is not generally recognized in the analysis of sentence function. The primary reason to separate meaning and information packaging is that, in principle at least, **any semantic content can be packaged in any information packaging function** (see §2.3). In *a tall tree,* the property of tallness is being used to modify the referent of *tree.* But in *That tree is tall*, the property of tallness is predicated of the tree.

The semantic classes of objects, properties and actions can all refer, modify, or predicate (Croft 1991), as can be seen in Table 1.1 for the words in boldface.

	reference	**modification**	**predication**
object	*the sharp* ***thorns***	*the* ***thorn's*** *color*	*It* ***is a thorn****.*
property	***sharpness***	*the* ***sharp*** *thorns*	*Those thorns* ***are sharp.***
action	*I said* [***that*** *the thorns* ***scratched*** *me*]. *the* [***scratching*** *of the thorns*]	*the thorns* [***that*** ***scratched*** *me*] *the thorns* [***scratching*** *me*]	*The sharp thorns* ***scratched*** *me.*

Table 1.1. Packaging of semantic classes in different propositional act functions.

The facts presented in Table 1.1 explain why the traditional grammar definitions of 'nouns' as object words, 'adjectives' as property words, and 'verbs' as action words fails. For example, properties and actions can be referred to (*sharpness, scratching*). The facts in Table 1.1 also suggest which constructions are relevant for a crosslinguistically useful analysis of parts of speech, namely constructions used for reference, modification and predication. Recognizing these two dimensions of linguistic function allows for a crosslinguistically valid approach to the vexing problem of parts of speech.

Nevertheless, it is true that referring expressions are most frequently objects, predicates are most frequently actions, and modifiers are most frequently properties. They are the constructions that occur on the diagonal from upper left to lower right in Table 1.1: *the sharp* ***thorns*** (noun)*, the* ***sharp*** *thorns* (adjective)*, The sharp thorns* ***scratched*** *me* (verb). As the parenthetical terms indicate, these frequent combinations

form the basis of a crosslinguistically valid analysis of nouns, verbs and adjectives. We will develop this analysis in more detail in §2.2 and §2.3.

The separation of semantic content and information packaging also allows us to reformulate the problem in the statements in (2a), (2b) and (2d) in §1.2.1, such as 'Sidaama numerals are adjectives'. Many word class terms, especially those for major parts of speech, are used for either the semantic class or the information packaging function of the construction defining the word class. Hence 'Sidaama numerals are adjectives' really means 'Sidaama numerals occur in the same modification construction used for prototypical modifiers'. 'Acehnese has no adjectives' doesn't mean 'Acehnese has no property words' or 'Acehnese has no modification constructions'. It means 'Acehnese does not have a modification construction used specifically for property words'.[2]

Many traditional uses of terms for word classes and parts of speech are ambiguous between denoting a semantic class and denoting an information packaging function. For example, 'demonstrative adjective' means 'deictic word ["demonstrative"] used as a modifier ["adjective"]', as in *that book*. In a parallel fashion, 'demonstrative pronoun' means 'deictic word used as a referring expression', as in *That is a book*. On the other hand, 'predicate adjective' means: 'property word ["adjective"] used as a predication ["predicate"], as in *That tree is tall.* In a parallel fashion, 'attributive adjective' means: 'property word used as a modifier', as in *a tall tree.* So the term 'adjective' is used to denote the modification function in 'demonstrative adjective' but the property concept class in 'predicate adjective'.

Another example of use of grammatical terms for either information packaging function or semantic class is in a description of the words *ano* and *wola/wolata* in Sabanê (Nambikwaran; Antunes de Araujo 2004:96; VS = verbal suffix):

> The morphemes **ano** and **wola/wolata** are used to express plurality, meaning 'much/many/a lot'…In the sentences (26-31), **ano** is a quantifier, while **wola(ta)** is an adverb…:
>
> (27) naysunum-ka ano-it-al-i
> land-OBJ much-VS-PRS.N-ASSR
> 'There is plenty of land.'
> …
> (30) wolata amayl-i-al-i
> a_lot_more to_rain-VS-PRS.N-ASSR
> 'It is raining very hard.'

The word *ano* is defined by its semantic class ('quantifier'), but the near-synonym *wola(ta)* is defined by the information packaging function of the construction it occurs in ('adverb', i.e. modifier of a predicate). Thus, a major part of the difficulty in interpreting statements about word classes in reference grammars involves figuring out whether the author is using the term 'adjective', say, to describe a language-specific word class

[2] Actually, for Durie it also means 'Acehnese does not have a predication construction used specifically for property words'. One doesn't always know which constructions are implicitly referred to by statements like 'Acehnese has no adjectives'.

defined by occurrence in one or more constructions of the language; a semantic class (property concepts); or an information-packaging function (modification).

The separation of semantic content and information packaging in describing the function of constructions has a number of theoretical and practical consequences. The information packaging functions are generally much more isomorphic to syntactic structures than lexical semantic classes: object concepts can be predicated, and action concepts can be packaged as referents, patients can be packaged as subjects, etc. The information packaging functions are also less variable across languages—i.e. more universal—than lexical semantics. Although the match is not perfect, **information packaging is the function of the morphosyntactic form of sentences**. For this reason, the organization of the constructions in this textbook is according to their information packaging functions.

The information packaging functions in Table 1.1 provide the basic skeleton for most constructions (see §1.5). This basic skeleton is described in more detail in Chapter 2. The information packaging functions in Table 1.1 also provide the basic skeleton for this textbook. Part Two describes constructions used for reference (chapter 3) and for modification (chapters 4-5). These are **phrasal constructions** (**phrases** for short). For example, *the furry kitten* is an instance of a phrase: it refers to a kitten, and modifies that concept with the property of being furry. Part Three describes constructions used for predication. These are **clausal constructions** (**clauses** for short). An example of a clausal construction is *The birds were singing*: singing is predicated of the birds. Part Four describes constructions used for multiple predications. These are **complex sentence constructions**, made up of multiple clauses. An example of a complex sentence construction is [*The birds were singing*] [*when I went out to get the newspaper*]. It consists of two clauses, indicated by the square brackets; the second clause is dependent on the first one and bears a semantic relation of overlapping occurrence in time.

The functions in Table 1.1 are not the only information packaging functions found in grammatical constructions. **All constructions encode both semantic content—information—and the way that information is packaged**. The information packaging functions associated with different types of constructions will be described in the chapters making up Parts Two, Three and Four of this textbook.

The organization of this textbook is summarized in Table 1.2. There are construction types and information packaging functions in Table 1.2 that are probably not familiar to you; they will be explained as they are introduced.

Part One: Introduction	(*overview, framework of analysis*)	Chapters 1-2
Part Two: Phrases	reference	Chapter 3
	modification	Chapters 4-5
Part Three: Clauses	predication (topic-comment)	Chapters 6-9
	predication, thetic, identificational	Chapters 10-11
	speech act functions, modality	Chapter 12
	complex predicates	Chapters 13-14
Part Four: Complex Sentences	coordination, subordination	Chapters 15-17
	events as arguments (complements)	Chapter 18
	events as modifiers (relative clauses)	Chapter 19

Table 1.2. Information packaging functions and the organization of this textbook.

There still remains the question of how to organize the chapters themselves. Constructions are a pairing of form and function, and function represents both semantic content and information packaging. But morphosyntactic form varies considerably across languages. What is the best way to compare and hence organize that crosslinguistic diversity of constructional form? In the next section, we argue that the comparison involves distinguishing grammatical constructions from grammatical strategies.

1.4. How do we compare constructions within and across languages?

Semantic content and information packaging allows us to compare constructions within and across languages. For example, we can compare the predication of object concepts in English (9), Spanish (10) and Classical Nahuatl (11a; cf. Nahuatl action predication in 11b; Andrews 1975:148, 25; cf. Stassen 1997:46, glossed by Stassen):

(9) I am not **a doctor**.

(10) Yo no soy médico.
I NEG am **doctor**
'I am not a doctor.'

(11) a. ah-ni-tīcitl
NEG-1SG-**doctor**
'I am not a doctor.'

b. ah-ni-chōco
NEG-1SG-**cry**
'I am not crying.'

But how do we talk about the constructions of the world's languages? That is, how do we talk about not just the function (semantics and information packaging) of the constructions in (9)-(11), but also the morphosyntactic structures in (9)-(11)? We need **comparative concepts** (Haspelmath 2010b) for object predication constructions, and also

action predication constructions, object reference constructions, and any grammatical construction that we want to compare across languages. Moreover, we will also want to compare different constructions with similar functions in a single language, such as the English constructions in (12) used for object modification, that is, the modification of one object concept (regulations) by another object concept (university):

(12) a. the regulations of the university
b. the university's regulations
c. university regulations

Haspelmath describes comparative concepts as concepts that are not language-specific constructions, nor the language-specific word classes that those constructions define. Instead, they are theoretical concepts used for cross-linguistic comparison. The most important property of comparative concepts is that they are cross-linguistically consistent. That is, their definition is not dependent on language-specific constructions and categories.

The functions that speakers want to express, that is, semantics or information packaging, can serve as comparative concepts (see §1.2.2; Croft 2001, chapter 3). But in order to talk about the morphosyntactic form of constructions, we need to define grammatical comparative concepts: comparative concepts that refer to the pairing of form and function, not just function by itself.

Typologists generally use two general types of comparative concepts (Croft 2014, 2016). The first general type is a **construction**:

> **construction:** any pairing of form and function in a language (or any language) used to express a particular combination of semantic content and information packaging

The term 'construction' when unmodified refers to any pairing of morphosyntactic form and function (the latter is a combination of semantic content and information packaging). When 'construction' is modified, as in 'relative clause construction', it refers to the category of form-function pairings that express the function described by the modifying phrase, but any form expressing that function in any language.

That is, a construction is a set of morphosyntactic forms that all have in common a particular combination of semantic content, such as object concepts, and information packaging, such as predication. However, a construction does not define any particular form. That is, a construction does not define **how** the function is expressed; a construction just defines **what** function is expressed. For example, if one finds an article titled "A typology of relative clause constructions", one will expect to find a survey of all the different morphosyntactic ways that languages express a particular function (in this case, action modification; see chapter 19).

Since every construction expresses a combination of a particular semantic content and a particular information packaging function, an obvious solution to the naming of this comparative concept is to use a compound of the semantic content and the information packaging function, i.e. 'object predication construction', 'object reference construction', 'action predication construction' and so on. In many cases, this will suffice. But in fact the traditional grammatical terms for many constructions are used in basically this

fashion: a 'predicate nominal construction' is a construction used for object predication, and so on. In particular, reference grammars use grammatical construction terms such as 'predicate nominal', 'relative clause', and so on to describe how the relevant combination of semantic structure and information packaging function is expressed in the language being described. We will respect this usage as much as possible when giving names for constructions as grammatical comparative concepts.

For example, we will use the term **predicate nominal construction** to describe a grammatical construction in a language used to express object predication. A predicate nominal construction is a pairing of grammatical form with a function. The function taken by itself is called 'object predication'. But for a language-specific construction to be categorized as a predicate nominal construction, all that matters is the function it is used to express. Since 'predicate nominal construction' is a crosslinguistic comparative concept, not a language-specific concept, it is not capitalized.

Constructions as comparative concepts come in different degrees of generality. A predicate nominal construction is defined in terms of the function of predicating an object concept. One can also define a predication construction, which is defined in terms of the function of predicating any type of concept: object concepts, property concepts, action concepts, etc.

Much discussion of constructions outside of typology as well as in typology pertains to language-specific form-function pairings, such as the English Numeral Modification Construction. In order to distinguish these language-specific form-function pairings from the functionally defined constructions that serve as comparative concepts, we will capitalize the term 'Construction' when it is part of the name of a language-specific construction, e.g. the English Predicate Nominal Construction. We will return to the relationship between language-specific and cross-linguistic form-meaning pairings in §1.5.

The range of constructions found in the world's languages represents the range of meanings that are communicated in linguistic utterances. Human languages are general-purpose communication systems and thus are used to express everything from fundamental experiences common to all human beings to highly specialized knowledge found only in a single culture or, even more narrowly, to a subcommunity in a culture with a particular expertise. No single grammar textbook or reference grammar of a language can possibly capture this full range of human experience even for one speech community. Nevertheless, focusing on the morphosyntax (rather than the lexicon) does delimit a more manageable subset of the grammatical structure of a language, namely how the meanings found in individual words and morphemes are combined and packaged. Traditions of grammatical analysis in major literary languages and in linguistics, and of grammatical description of languages around the world, allow us to devise a framework encompassing the broad range of information packaging functions that grammatical constructions perform.

The introduction of grammatical constructions as comparative concepts is not enough, however. We also want to be able to compare constructions in terms of their form as well as their function. For example, English *I am not a doctor* and the Spanish translation equivalent *Yo no soy médico* are structurally similar in that they both contain an inflecting form (English *am*, Spanish *soy*) distinct from the object concept word (English *doctor*, Spanish *médico*). English and Spanish differ from the Classical Nahuatl

translation equivalent *ah-ni-tīcitl*, in which the object concept word (*tīcitl*) itself is the inflecting form. In other words, calling *I am not a doctor* and *Yo no soy médico* instances of the predicate nominal construction tells us **what** the English and Spanish speaker are trying to say—that is, the semantic content and how they package that content. But this means that the definition of a construction does not tell us **how** the English and Spanish speakers express the function, let alone the similarities in how they express it.

Thus, we also need grammatical comparative concepts that describe not just **what** a speaker intends to convey—that is, the function—but also **how** they convey it—that is, also the grammatical form. To describe the 'how' as well as the 'what', we introduce the notion of a **strategy**:[3]

> **strategy**: a construction in a language (or any language), used to express a particular combination of semantic content and information packaging (the 'what'), that is further distinguished by certain characteristics of grammatical form that can be defined in a crosslinguistically consistent fashion (the 'how').

For example, we will say that both English and Spanish employ a **verbal copula strategy** for their predicate nominal constructions (see §10.1.2 for details).

The verbal copula strategy has certain characteristics of grammatical form that can be defined independently of language-specific word classes or constructions. Those characteristics are the ones described above: (i) the predication function is expressed in part by the presence of a morpheme different from the object concept word, and (ii) this additional word is inflected for at least some of the categories that the construction for the predication of actions (i.e., verbs) also inflect for.

A single language may have multiple constructions using different strategies for a single construction, as we saw in (12a-c) for object modification (broadly construed; see §4.1.4) in English. Thus these grammatical comparative concepts are for comparing different morphosyntactic types of constructions, whether in different languages or in the same language. A single language may even have multiple constructions using the same strategy. For example, Spanish has two verbal copula strategies, one using *ser* and one using *estar*. (They differ semantically, but also overlap in some contexts.) These two Spanish verbal copula strategies will have to be differentiated by language-specific construction names, such as the *Ser* Copula Construction and the *Estar* Copula Construction.

Strategies come in different kinds. One type of strategy is defined like the verbal copula strategy, using properties of grammatical structure that are **cross-linguistically valid**, that is, defined independently of language-specific word classes or constructions. For example, a copula can be defined crosslinguistically as a separate morpheme that encodes the predication function. These types of strategies we will call **encoding strategies**: they represent different morphosyntactic means for expressing the same function in different languages, or as alternative means for expressing the same function in a single language.

Another type of strategy is defined in terms of two or possibly more constructions with closely related functions. In this case, the strategy is defined in terms of similarities

[3] The term 'strategy' is one long used with this meaning in typology (at least as far back as Keenan and Comrie 1977 and Givón 1979).

and differences in morphosyntactic form (again, defined in a cross-linguistically valid fashion) between the two constructions. The morphosyntactic similarities and differences between the strategies found in two (or more) functionally closely related constructions forms a **system** of strategies for the contrasting functions of the constructions.

For example, what are called alignment strategies (see §6.3.1) are defined in terms of the similarities and differences in the forms of intransitive and transitive constructions. For example, consider the transitive and intransitive constructions of Yuwaalaraay, illustrated in (13)-(14) (Williams 1980:36):

(13) ḍuyu-gu ṇama ḍayn-Ø yi:-y
snake-**ERG** that man-**ABS** bite-NFUT
'The snake [ergative] bit the man [absolutive].'

(14) wa:l ṇama yinar-Ø banaga-ṇi
NEG that woman-**ABS** run-NFUT
'The woman [absolutive] didn't run.'

In the Yuwaalaraay intransitive, the form of the lone argument phrase of an intransitive predicate—'woman' in (14)—is the same as the form of the argument phrase expressing the patient-like participant of the event denoted by a transitive predicate—'man' in (13). This co-expression (see Hartmann et al. 2014:476, fn. 5) of arguments is called the absolutive. In contrast, the argument phrase expressing the agent-like participant of the same event—'snake' in (13)—is expressed with a different form from that of the lone argument phrase of the intransitive predicate, namely the suffix *-gu*. This distinct expression, or lack of co-expression, is called the ergative.

This second type of strategy describes similarities and differences in morphosyntactic form that apply across two (or more) constructions—in this case, the intransitive and transitive constructions. Whenever a strategy is defined in synchronic comparison to another strategy, then a system of strategies is being described.

A third type of strategy is defined in terms of using another construction in the language to express the function in question. For example, the presentation of a possession relation in Russian is expressed using a locative construction, that is, a construction also used for the presentation of an object at a location (see §10.4):

(15) u menja mašina
at **1SG.GEN** car:NOM
'I have a car.' [lit. '**At me (is)** a car']

We will call these **recruitment strategies**: the construction used for one function—presentative locative—is recruited for use in a different function—presentative possession. Another way of describing a recruitment strategy is that the construction being recruited for the new function is being 'extended' to the new function.

Recruitment strategies are very common, since recruitment of an existing construction reduces the number of different strategies needed for the vast range of meanings one wants to express. Recruitment strategies represent a fundamentally

diachronic relationship between two constructions: the recruitment event actually occurred in the past.

A strategy that is recruited from one function will adapt to its new function. For example, the predication of physiological sensation is recruited from the possessive construction in some languages. In French, *J'ai faim*, literally 'I have hunger', means 'I am hungry.' One can modify *faim* with the degree intensifier *très* 'very': *J'ai très faim* 'I am very hungry'. One cannot use a degree intensifier with ordinary possession: 'I have very a car' makes no sense (Croft 2001:115). This is the first step in divergence of a recruited strategy from it source construction. As the recruited construction's form comes to be altered, it diverges so much from its source construction that it evolves into an encoding strategy unique to that construction.

There is a sense in which all strategies are probably ultimately recruitment strategies diachronically. Much of the diversity of grammatical strategies, particularly the less common or more unusual strategies, exists because grammatical change is gradual, and "hybrid" strategies appear as one strategy gradually evolves into another. Systems of strategies have a more complex history, involving divergence or convergence of constructions that come to be used for closely related functions. For these reasons, it is sometimes difficult to differentiate strategies that are historically related. We will discuss diachronic processes that lead to linguistic diversity in constructions in many places in this textbook.

Finally, strategies come in varying degrees of generality. The inflected copula strategy for the predicate nominal constructions in English and Spanish was described above specifically for the function of object predication. However, the verbal copula strategy is also used in both languages for predicate adjective constructions (to express property predication): *Ella **es** inteligente* 'She **is** intelligent'. Should we define strategies independent of constructions, that is, independent of the function that the construction expresses?

We do not do so here: **strategies are always strategies for a particular construction**, the 'how' for a particular 'what'. Nevertheless, we can say that a strategy is used for a more general construction. For example, we can define the verbal copula strategy as a strategy for the more general predication construction: a construction used to predicate any kind of semantic content, whether it is object predication or property predication or predication of some other concept. We can also define strategies for even more general constructions (see in particular §2.4-§2.5, §4.2-§4.5, §6.2.2 and §15.2.3). The recognition of strategies for highly general constructions represents some of the insights of grammatical theory as to how human beings verbalize their experiences in linguistic form.

Studying the strategies used for constructions is the heart of grammatical analysis. The range of strategies found in the world's languages represents the variation in grammatical structure that may occur. It also implies constraints or at least dispreferences in how function is encoded in grammatical form that constitute generalizations or universals about human language, and it reveals the rich network of relationships among constructions in a language (and in language in general). This textbook will survey the major strategies for a wide range of constructions that have been found in crosslinguistic research. It is not intended to be exhaustive: languages are often surprisingly diverse, and unusual strategies are sometimes employed for certain constructions.

In sum, we have two categories of grammatical comparative concepts: constructions and strategies. The grammatical comparative concepts allow us to talk about grammatical constructions (form as well as function) across languages, even though particular grammatical constructions are language-specific. Since they are comparative concepts, they are in lower case, unlike language-specific construction names, which are capitalized just like language-specific word classes. Constructions are defined solely by the combination of meaning and information packaging they express, while strategies are defined in addition by certain crosslinguistically definable formal characteristics they have in common.

1.5. How do we analyze the structure of sentences in a particular language?

In §1.1, we characterized our method of analyzing morphosyntax as constructional and functional-typological, and in the following sections, we described in more detail the structure of constructions, and how to compare constructions cross-linguistically. We now return to what this means for the analysis of the structure of sentences in particular languages.

The first step in any linguistic analysis is classification of the phenomena in question, in our case sentences, by certain parameters, in our case the function and form of the sentence.[4] Ideally, analysis proceeds by inductive generalization over a corpus of utterances, although further sentences may be obtained by elicitation to clarify points of analysis. In constructional analysis, the first step is to classify sentences as instances of a construction, that is, a pairing of form and function. In other words, we begin holistically, by categorizing complex structures as a whole.

Sentences can be classified by function and by form. We use the example of the object predication function that we introduced in §1.4. We can classify the English sentences in (16a-c) as instances of the object predication function:

(16) a. She is a doctor.
b. Bill is a teacher.
c. That flower is a penstemon.
etc.

The sentences in (16a-c) are also similar in morphosyntactic form. That is, each of (16a-c) can be analyzed into parts that correspond both in form and function: *She, Bill* and *That flower* correspond to the referent of which the object category is being predicated; *a doctor, a teacher* and *a penstemon* correspond to the object category that is predicated; and *is,* or more generally a form of *be*, recurs across all the sentences.

The classification process involves choices about how narrowly to define the combination of function and form that describes the construction category. Consider the functions expressed by the sentences in (17):

[4] For a more detailed discussion of grammatical analysis and explanation as presented in this section, see Greenberg (1968, chapter 10) and Croft (2003:284-86; the 1990 edition of this book has a more detailed discussion on pp. 246-59).

(17) a. She rode her bicycle across campus.
b. That slope is very steep.
c. She became a doctor.
d. The mayor of San Agustín is my aunt.
e. You are more of a dancer than she is.

The function of action predication in (17a) could be included to form a more general construction of predication of any type. If so, the number of shared functional and formal properties of this more general construction will be reduced in comparison to the narrower type illustrated in (16a-c). The function of property predication in (17b) could be grouped with object predication, especially as property predication shares a formal element, namely a form of *be.* Example (17c) has a different meaning, of process rather than state, as well as replacing *be* by *become*. Example (17d) is not strictly predicational but equational (see §10.1.2 for the distinction), although it shares the formal element *be*. Example (17e) adds a different semantic dimension, of comparison, as well as a more complex form.

Constructions can be defined at different levels of generality, as long as there are shared similarities of function and/or similarity of form. One could broaden the definition of the construction one wants to analyze to include these additional examples, on the assumption that there is some degree of similarity of function and form between all the members of the broader construction category. There will be also cross-cutting patterns of similarity, which can be captured by a hierarchy or a lattice of constructions, or a constructional paradigm. Here we stick with the narrow definition of the English Predicate Nominal Construction that excludes sentences like (17a-e).

The next step of analysis is what makes the approach here typological, and represents the focus of this textbook, which is intended for an audience that is already familiar with the sort of single-language analysis described in the preceding paragraphs. We now compare the English construction illustrated in (16a-c) to parallel constructions in other languages. At this point, the process of classifying constructions is based on properties of function and form that are shared across languages, and not specific to English. The process of analysis is still one of classification. In (16)-(17), we grouped certain English sentences together as instances of an English Predicate Nominal Construction, and exclude others. Now we group the English sentences in (15a-c) with sentences in other languages, such as (18) from French and (19) from Kiowa (Watkins 1984:227; cf. Stassen 1997:94, which corrects the gloss):

(18) Elle est médécin
she **is** doctor
'She is a doctor.'

(19) té: kɔ́ygú bà-dɔ́:
all Kiowa.INV 2PL-**be**
'You are all Kiowas.'

The sentences in (18) and (19) correspond to the English sentences in (16a-c) in function—object predication—and in form. All the sentences contain a form denoting a

referent, a form denoting the predicated object category, and a form inflected with the same categories that action predications ("verbs"; see §2.5) inflect for in the language. This cross-linguistic category corresponds to a construction strategy, namely the verbal copula strategy, introduced in §1.4 and described in more detail in chapter 10.

A further step is to compare other sentences in still other languages that express the same function but with different formal structures. This is a still broader classification process. Now we classify sentences such as the English sentences in (15a-c) and the French and Kiowa sentences in (18)-(19) with other sentences in other languages that use different formal structures, such as the zero strategy in (20) from Burmese (Okell 1969:177; cf. Stassen 1997:70) and the verbal strategy in (21) from Classical Nahuatl (Andrews 1975:146; gloss added):

(20) thu si?thà
he soldier
'He is a soldier.'

(21) ni-cihuātl
1SG-woman
'I am a woman.'

This more general category is also a cross-linguistic category, the predicate nominal construction (all lower case, because it is a cross-linguistic category).

This way of analyzing the English Predicate Adjective Construction typologically is illustrated in Figure 1.2.

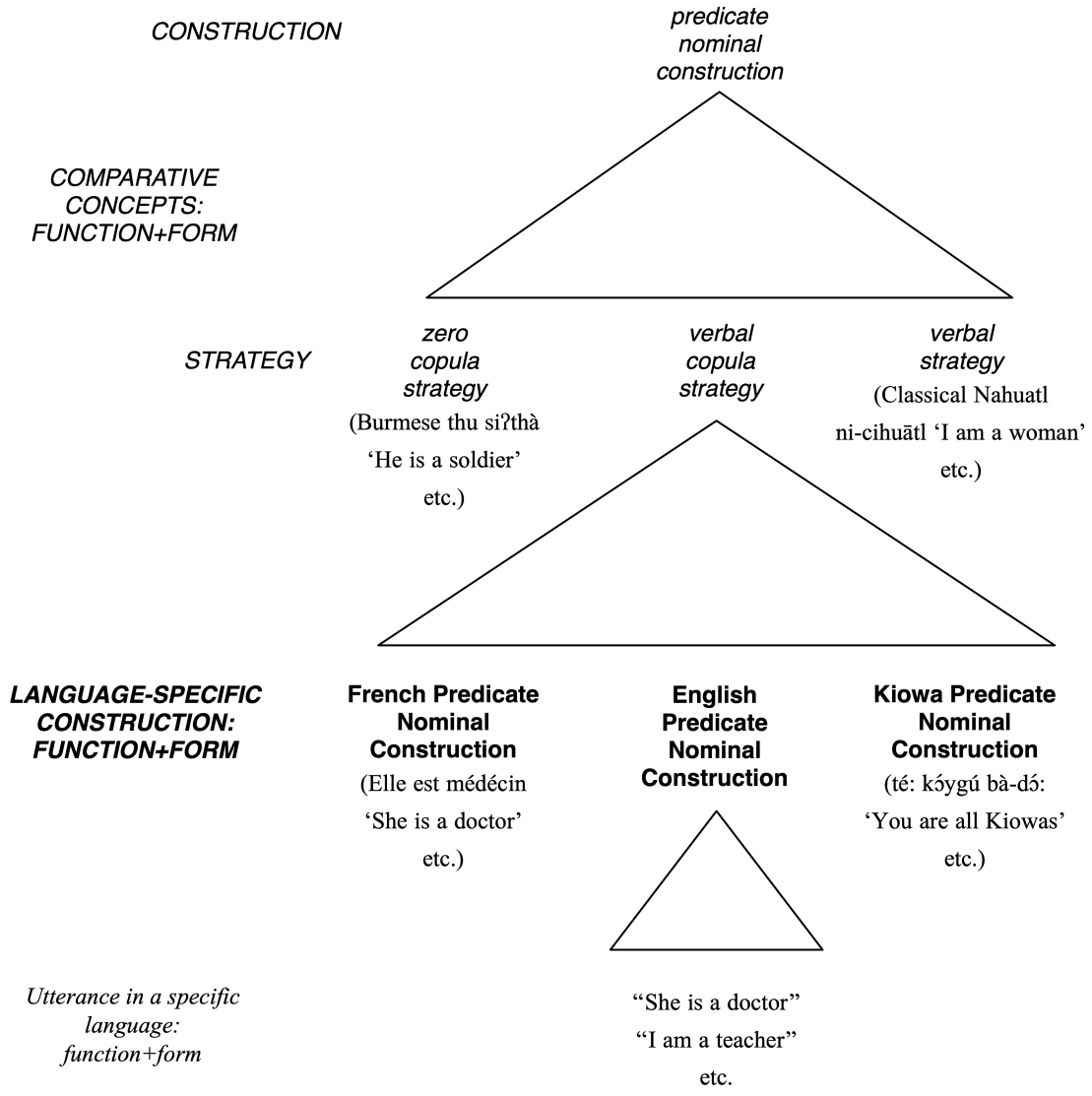

Figure 1.2. Analysis of English She is a doctor *via constructional classification.*

All of the categories in Figure 1.2 are instances of constructions, that is, form-function pairings. Only constructions (lower case and specified by function) and strategies are cross-linguistically defined form-meaning pairings. The English, French and Kiowa Predicate Nominal Constructions are form-meaning pairings defined language-specifically, i.e. by their language-specific function and form. Language-specific form-meaning pairings such as the English Predicate Nominal Construction (specified by language name and capitalized) and (cross-linguistic) strategies are defined by form as well as function. Only (cross-linguistic) constructions are defined by function alone.

The analysis of the English Predicate Adjective Construction in Figure 1.2 describes the construction as a whole and its function as a whole. However, we already saw that in

order to reach this analysis, the parts of the sentences that belong to the construction were classified in terms of both function and form.

Any predication construction predicates something of a referent. In *She is a doctor*, the referent is expressed by *She* and the predicated concept is expressed by *a doctor*. In addition, the English Predicate Nominal Construction contains another word, namely *is*. This additional word identifies the English Predicate Nominal Construction as an instance of one particular strategy of the predicate adjective construction, the verbal copula strategy. This strategy entails that the construction contains a verbal copula, as defined in §1.4: a word distinct from the predicated concept word, but bearing inflections also found in action predication.

A visual representation of the analysis of the sentence *She is a doctor,* an instance of the English Predicate Nominal Construction, and its parts is given in Table 1.3:

Sentence	*She*	*is*	*a doctor*
Construction	English Predicate Nominal Construction		
Roles	PrNomSbj	*be*	PrNomPred
semantics	object		object
information packaging	reference	predication	
construction	predicate nominal (object predication)		
strategy	verbal copula		

Table 1.3. Analysis of She is a doctor *as an instance of the English Predicate Nominal Construction.*

The first row is the specific sentence being analyzed, *She is a doctor,* with its breakdown into parts. The second and third rows, the rows in white, classify the sentence as an instance of a language-specific construction, the English Predicate Nominal Construction, and give the roles of that construction which each part of the sentence *She is a doctor* fills. The labels "PrNomSbj" (for "Predicate Nominal Subject") and "PrNomPred" ("Predicate Nominal Predicate") indicate that these English syntactic categories are defined specifically by the roles in the English Predicate Nominal Construction (see §1.2). The form *be* is intended to subsume all forms of this English Verb that occur in that role in the English Predicate Nominal Construction.

The rows in light gray represent the function of the construction, that is, the combination of semantics (fourth row) and information packaging (fifth row) that the construction expresses. These two dimensions of function are both language-specific and cross-linguistic. They are language-specific in that this is the function expressed by English speakers when they use the English Predicate Nominal Construction. They are cross-linguistic in that the function of the English Predicate Nominal Construction also serves as a functional comparative concept (see §1.4).

Finally, the rows in dark gray represent the analysis of *She is a doctor* in cross-linguistic perspective, as hybrid (form + function) comparative concepts. That is, they represent the English Predicate Nominal Construction as an instance of the cross-linguistic predicate nominal (= object predication) construction (sixth row), specifically

the predicate nominal construction using the verbal copula strategy (seventh row). The cross-linguistic analysis is what is displayed in Figure 1.2.

Parts of a construction are themselves constructions. For example, *she* is an instance of the English Personal Pronoun construction, a particular type of referring expression construction (see chapter 3). And *is a doctor* is an instance of the English Support Verb Construction, where an inflecting form (*be*) combines with or "supports" a word or phrase (in this case, the phrase *a doctor*) expressing a predicated concept (see §13.5). In other words, a sentence may be the instance of different constructions, capturing different dimensions of the morphosyntactic structure of the sentence and of parts of the sentence.

What is the role, if any, of the comparative concepts in dark gray in Table 1.3 in a single-language morphosyntactic analysis? This is what this book is basically all about. We will give a simple example here. The syntax of a single language is complex and varied. For example, English uses two different predication constructions, the verbal copula for object and property predication, and inflection without a copula for action predication:

(22) a. She **is** a doctor.
b. She **is** intelligent.
c. She sing-**s**.

The single-language description of English can answer the **what** question, namely what function is being expressed here; and the **how** question, namely what morphosyntax is used to express the function. But we also want to know the answer to the **why** question. Why is there a copula form for object and property predication, but not for action predication? Languages are also not fixed: speakers vary in how they express functions, and they also change over time how they express functions. For example, Cantonese uses a copula *háih* for object predication only; but it is now coming to use the word *hóu* as a copula for property predication (Matthews and Yip 1994:158; see §10.3, footnote 2). Why is Cantonese developing a copula for property predication, but not (yet) for action predication?

The reason is that the patterns of variation and change found in the grammar of a particular language are in many cases simply instances of patterns of variation and change found across languages (Croft 2001:105-7; Croft 2016b). In order to explain the structure, variation and change occurring in a single language, it is easiest (and empirically safest) to look at variation and change across languages, and see what universal patterns and explanations for those patterns there are. This is a major purpose of this textbook: not just to present the major functions expressed in languages, and the commonest strategies to express those functions, but to answer (where possible) **why** languages have come to be the way they are.

The reader will no doubt have realized that a thorough analysis of a language-specific construction, both the whole and its parts, in terms of both form and function will lead to a plethora of technical terms. Specifically, there will need to be terms for the semantic content; the information packaging; the construction; and the strategy used for the construction. These four analytical categories correspond to the four rows in light and dark gray in Table 1.3.

As much as possible in this textbook, theoretical terms already in use will be employed, and as closely to traditional use as possible. However, traditional use of such terms may sometimes be ambiguous, for example between semantic vs. constructional definitions, information packaging vs. constructional definitions, and construction vs. strategy definitions. For this reason, the reader must become familiar with the precise use of the terms in this textbook; major differences with traditional use will be indicated when the term is introduced. In order to focus on the accurate use of terms, when each term is introduced, it will be emphasized in boldface. The term will be identified as semantic **(*sem*)**, information packaging **(*inf*)**, construction **(*cxn*)** or strategy **(*str*)**, or simply as a general theoretical concept. At the end of each chapter there will be a summary of the terms and concepts introduced in the chapter, classified by type and listed by the sections in which they were introduced. Finally, there is an online alphabetic glossary of all of the terms in the book, classified by type, with examples in English or other languages.

1.6. Appendix—Interpreting language examples: interlinear morpheme translations

Since this textbook describes crosslinguistic variation in how semantic content and information packaging are encoded in morphosyntactic form, one must be able to interpret example sentences from other languages, including languages that one is otherwise unfamiliar with. The now widely used method to make the interpretation process easier is the **interlinear morpheme translation**, or **IMT**. An IMT can be illustrated by example (23) from Amharic (see §2.4):

(23)	yä-pitär	yäfəqr-u-n	zäfän	azzäfafän	← **object language**
	GEN-Peter	love-DEF-ACC	song	sing:NML	← **IMT**
	‘Peter’s singing [i.e. his way of singing] the love song’				← **(free) translation**

The first line gives the sentence or phrase in the original language, usually called the **object language**, in a morphological analysis. If the morpheme boundaries are obscured by morphophonological processes, the author sometimes gives two lines for the object language: the first line is the sentence as spoken, and the second line is a morphological analysis of the sentence before the application of the morphophonological rules.

There is substantial variation in the notation of different kinds of morpheme boundaries in language descriptions. However, an emerging standard is given by the Leipzig Glossing Rules (http://www.eva.mpg.de/lingua/resources/glossing-rules.php). The Leipzig Glossing Rules should be studied carefully; only the briefest summary of the rules is given here (relevant rules are cited in parentheses). The following list gives a summary of the morpheme boundary notation found in the object language line:

=	links a clitic (enclitic or proclitic) to the word to which it is cliticized (Rule 2)
-	links an affix (suffix or prefix) to the stem to which it is affixed (Rule 2)
< >	encases an infix inside a word or stem (Rule 9)
~	links the reduplicated part of a word or stem to the word/stem (Rule 10)
< >~	links an infixed reduplication to the word/stem (not in the Leipzig Glossing Rules)

The second line is the IMT. The IMT gives a schematic description of the morphosyntactic structure of the example in the object language. Each part gives the translation of the morphemes in the object language line. The commonest convention is to translate word bases in lower case, and to translate grammatical morphemes in small capitals or all capitals. Abbreviations are generally used for grammatical morphemes (Rule 3). Abbreviations used in IMTs in this book are given in the List of Abbreviations.

The same morpheme boundary notation is used in the IMT, except that the translation of an infix is moved to just before the base. However, in many cases there is not a one-to-one mapping between the object language morphemes and the IMT elements. In some cases, the grammatical meaning is zero-coded (see §2.4). Zero coding is annotated either by Ø in the object language line, or by putting the category in square brackets [ ] in the IMT line (Rule 6). In other cases, the grammatical meaning is expressed by two discontinuous morphemes in the object language (bipartite elements and circumfixes). This may represented by repeating the gloss for the meaning in the IMT line (Rule 8).

In still other cases, grammatical categories are expressed by nonconcatenative morphology in the object language, such as transfixes, suppletion, subtraction and other types of base modification (for an introduction to morphology, see Haspelmath and Sims 2010). The IMT also notates **cumulation** (Haspelmath and Sims 2010:98), that is, the expression of more than one grammatical category by a single morpheme in the object language (e.g. English *-s* in *She sing-s* cumulates third person, singular number and present tense in a single morpheme); and **false cumulation**, that is, the translation of an object language morpheme by more than one English word because English lacks a one-word translation (e.g. Spanish *buscar* and its English translation 'look for').

The best practice for notating these mismatches in form and meaning in the Leipzig Glossing Rules are given below:

\	morphophonological modification of the base (Rule 4D)
.	cumulation (Rule 4); conventionally, a period is not used between person and number (e.g. 1PL) (Rule 5)
>	portmanteau morpheme expressing one person acting on another (e.g. 1PL>3SG); this is a special case of cumulation (Rule 4E)
_	false cumulation (e.g. look_for) (Rule 4A)

In actual practice, the period is often used for all of these form-meaning mismatches (Rule 4). In addition, when there are two morphemes in the object language but the author does not want to show the formal segmentation, a colon can be used, as in *azzäzafän* [sing:NML] in (23) (Rule 4C). Nevertheless, it is valuable to distinguish the different types of form-meaning mismatches, so best practice is encouraged here.

The last line gives the "free" translation of the object language example. This translation essentially gives you the meaning of the example, ideally both the semantic content and the information packaging. In effect, the language of the translation—in this book, English—acts as the **metalanguage** to represent the meaning of the object language examples.

English (or other languages used in language descriptions written in French, Spanish, Russian, Japanese etc.) is not always the best metalanguage for describing the object

language function. English, and the other languages used in translations, does not always make lexical or grammatical semantic distinctions that are found in the object language, and sometimes the metalanguage's way of representing the information packaging in the object language is not always clear, or not carefully replicated by the author. A language description should provide careful translations of object language examples, based on an understanding of semantics and information packaging in the metalanguage used as well as in the object language. This is not always the case. Nevertheless translations are the only thing we can go on unless the author provides a discussion of the meaning of the example, or we have access to native speaker consultants. For this reason, translations must be used with reasonable caution when analyzing an unfamiliar language.

In sum, the three lines in an example like (23) gives the grammatical structure of the object language sentence (the object language line), the semantic analysis of the parts of the object language sentence (the interlinear morpheme translation/IMT), and the semantic structure including information packaging of the whole sentence (the free translation).

Terms introduced in this chapter:

1.1. What is morphosyntax?
morphology, syntax, morphosyntax, construction (*see also §1.4)*, role, grammaticalization, meaning (semantic/information content), information packaging, form (*aka* expression), function, construction grammar, functionalism, linguistic typology

1.2. Why constructions?

1.2.1. Word classes and semantic classes
word class, semantic class

1.2.3. Word classes and constructions

1.2.4. Constructions and the organization of this textbook

1.3. Why information packaging?
reference/referent (*inf*), predication (*inf*), modification (*inf*), propositional act (*inf*), phrase (*cxn*) , clause (*cxn*) (*see also §2.1, §6.1.1)*, complex sentence (*cxn*)

1.4. How do we compare constructions within and across languages?
comparative concept, construction, strategy, verbal copula (*str*), crosslinguistically valid, encoding strategy, system, recruitment strategy

1.5. How do we analyze the structure of sentences in a particular language?

1.6. Appendix—Interpreting language examples: interlinear morpheme translations
interlinear morpheme translation (IMT), object language, (free) translation, cumulation, false cumulation, metalanguage