

University of Hue
College of Foreign Languages

An Introduction
to
English Syntax

University of Hue
College of Foreign languages

**An Introduction
to
English Syntax**

Compiled by
Nguyen Van Huy
Than Trong Lien Nhan

HCFL

CHAPTER 1

INTRODUCTION TO SYNTAX

I. Introduction

SYNTAX is the central component of human language. Language has often been characterized as the systematic correlation between certain types of oral/graphic forms for spoken/written language; and, for signed language, they are manual.

It is not the case that every possible meaning that can be expressed is correlated with a unique, un-analyzable form. Rather, each language has a stock of meaning-bearing elements and different ways of combining them to express different meanings, and these ways of combining them are themselves meaningful. The two English sentences *Chris gave the notebook to Dana* and *Dana gave the notebook to Chris* contain exactly the same meaning-bearing elements, i.e. words, but they have different meanings because the words are combined differently in them. These different combinations fall into the realm of syntax; the two sentences differ not in terms of the words in them but rather in terms of their syntax. Syntax can thus be given the following characterization, taken from Matthews (1982:1):

The term ‘syntax’ is from the Ancient Greek *syntaxis*, a verbal noun which literally means ‘arrangement’ or ‘setting out together’. Traditionally, it refers to the branch of grammar dealing with the ways in which words, with or without appropriate inflections, are arranged to show connections of meaning within the sentence.

First and foremost, syntax deals with how **sentences** are constructed, and users of human languages employ a striking variety of possible arrangements of the elements in sentences. One of the most obvious yet important ways in which languages differ is the order of the main elements in a sentence. In English, for example, the **subject** comes before the **verb** and the **direct object** follows the verb. In Lakhota (a Siouan language of North America), on the other hand, the subject and direct object both precede the verb, while in Toba Batak (an Austronesian language of Indonesia; (Schachter 1984b), they both follow the verb.

In Lakhota, the subject comes first followed by the direct object, whereas in Toba Batak the subject comes last in the sentence, with the direct object following the verb and preceding the subject. The **basic word order** in Toba Batak is thus the opposite of that in Lakhota. There are also languages in which the order of words is normally irrelevant to the interpretation of which element is subject and which is object. This is the case, for example, in Russian sentences.

In Russian the order of the words is not the key to their interpretation, as it is in the sentences from the other languages. Rather, it is the form of the words that is crucial.

The changes in the form of the words to indicate their function in the sentence are what Matthews referred to as ‘inflections’, and the study of the formation of words and how they may change their form is called **morphology**. The relationship between syntax and morphology is important: something which may be expressed syntactically in some languages may be expressed morphologically in others. Which element is subject and which is object is signaled syntactically in these languages, while it is expressed morphologically in the others.

Syntax and morphology make up what is traditionally referred to as ‘**grammar**’; an alternative term for it is **morphosyntax**, which explicitly recognizes the important relationship between syntax and morphology.

1. *Definition*

SYNTAX is the study of how words are combined to form sentences in a language. Thus, syntax concerns the system of rules and categories that underlies sentence formation.

1.1. Grammaticality and Ungrammaticality

A central part of the description of what speakers do is characterizing the **grammatical** (or **well-formed**) sentences of a language and distinguishing them from **ungrammatical** or (**ill-formed**) sentences. Grammatical sentences are those that are in accord with the rules and principles of the syntax of a particular language, while ungrammatical sentences violate one or more syntactic rules or principles. For example, *The teacher is reading a book* is a grammatical sentence of English, while *Teacher the book a reading is* would not be. Ungrammatical sentences are marked with an asterisk, hence **Teacher the book a reading is*. This sentence is ungrammatical because it violates some of the word order rules for English, that is (i) basic word order in English clauses is subject-verb-object, (ii) **articles** like *the* and *a* precede the **noun** they modify, and (iii) **auxiliary verbs** like *is* precede the **main verb**, in this case *reading*. It is important to note that these are English-specific syntactic rules.

Well-formed sentences are those that are in accord with the syntactic rules of the language; this does not entail that they always make sense semantically. For example, the sentence *The book is reading the teacher* is nonsensical in terms of its meaning, but it violates no syntactic rules or principles of English; indeed, it has exactly the same syntactic structure as *The teacher is reading a book*. Hence it is grammatical (well-formed), despite being semantically odd.

1.2. Grammaticality

A sentence is grammatical if native speakers judge it to be a possible or acceptable sentence of their language.

The dog bit the man.
The man barks.
* The dog the man bit.

- Grammaticality is not based on what is taught in school but on the rules acquired or constructed unconsciously as children. Much grammatical knowledge is ‘in place’ before we learn to read.

The ability to make grammaticality judgments does not depend on having heard the sentence before. You may never have heard or read *Enormous crickets in pink socks were dancing at the ball* but your syntactic knowledge will tell you the sentence is grammatical.

- Grammaticality judgments do not depend on whether the sentence is meaningful or not, as shown by the following sentences:

*Colorless green ideas sleep furiously.
A verb crumpled the milk.*

Although the sentences do not make much sense, they are syntactically well formed. They sound 'funny' but they differ in their 'funniness' from the following strings:

*Furiously sleep ideas green colorless,
*Milk the crumpled verb a.

The grammaticality of this case is based on the ordering of words and morphemes of a sentence.

- Grammatical sentences may be uninterpretable if they include nonsense strings, that is, words with no agreed-on meaning, as shown by the first two lines of 'Jabberwocky' by Lewis Carroll:

*'Twas brillig, and the slithy toves
Did gyre and gimble in the wabe.*

Such nonsense poetry is amusing because the sentences 'obey' syntactic rules and sound like good English. Ungrammatical strings of nonsense words are not entertaining:

**Toves slithy the brilltg 'twas
wabe the in gimble and gyre did.*

- Grammaticality does not depend on the truth of sentences either - if it did, lying would be impossible - nor on whether real objects are being discussed, nor on whether something is possible or not.

*You all have had 10 marks for the midterm examination.
Those fathers have been pregnant for 3 months.*

Unconscious knowledge of the syntactic rules of grammars permits speakers to make grammaticality judgments.

Thus syntactic rules in a grammar must at least account for:

- i. the grammaticality of sentences;
- ii. the ordering of words and morphemes;
- iii. structural ambiguity;

*synthetic buffalo hides (synthetic buffalo hides ≠ synthetic buffalo hides)
Visiting professors can be interesting.*

- iv. the fact that sentences with different structures can have the same meaning;

Learning syntax is interesting. = It's interesting to learn syntax.

- v. the grammatical and logical relations within a sentence;

The student solved the problem.
The problem was solved by the students.

- vi. speaker's creative ability to produce and understand any of an infinite set of possible sentences.

2. Syntactic Categories & Word Classes

2.1. Aspects of Syntactic Structure

In the syntactic structure of sentences, two distinct yet interrelated aspects must be distinguished. The first one has already been mentioned: the function of elements as subject and direct object in a sentence. 'Subject' and 'direct object' have traditionally been referred to as **grammatical relations**. Hence this kind of syntax will be referred to as **relational structure**. It includes more than just grammatical relations like subject and direct object; it also encompasses relationships like **modifier-modified**, e.g. *tall building* or *walk slowly* (*tall, slowly*=modifier, *building, walk*=modified) and **possessor-possessed**, e.g. *Pat's car* (*Pat's* = possessor, *car* = possessed). The second aspect concerns the organization of the units which constitute sentences. A sentence does not consist simply of a **string** of words; that is, in a sentence like *The teacher read a book in the library*, it is not the case that each word is equally related to the words adjacent to it in the string. There is no direct relationship between *read* and *a* or between *in* and *the*; *a* is related to *book*, which it modifies, just as *the* is related to *library*, which it modifies. *A* is related to *read* only through *a book* being the direct object of *read*, and similarly, *the* is related to *in* only through *the library* being the object of the **preposition** *in*. The words are organized into units which are then organized into larger units. These units are called **constituents**, and the hierarchical organization of the units in a sentence is called its **constituent structure**. This term will be used to refer to this second aspect of syntactic structure. Consider the eight words in the sentence *The teacher read a book in the library*. What units are these words organized into? Intuitively, it seems clear that the article *the* or *a* goes with, or forms a unit with, the noun following it. Is there any kind of evidence beyond a native speaker's intuitions that this is the case?

If the article forms a unit with the noun that follows it, we would expect that in an alternative form of the same sentence the two would have to be found together and could not be split up. Thus in the passive version of this sentence, *A book was read by the teacher in the library*, the unit *a book* serves as subject, and the unit *the teacher* is the object of the preposition *by*. The constituent composed of a noun and an article is called a **noun phrase** [NP]; as will be shown later, NPs can be very complex. The preposition *in* and the NP following it also form a constituent in this sentence (*in the library*); it is called a **prepositional phrase** [PP]. The fact that the PP is a constituent can be seen by looking at another alternative form. *In the library the teacher read a book*. Finally, the verb plus the NP following it form a unit as well, as shown by a sentence like *I expected to find someone reading the book, and reading the book was a teacher*. The constituent composed of a verb plus following NP is called a **verb phrase** [VP]. As with NPs, VPs can be quite complex. In each of these alternative forms, a combination of words from the original sentence which one might intuitively put together in a single unit also occurs together as a unit, and this can be taken as evidence that they are in fact constituents. Using square brackets to group the words in constituents together, the constituent structure of *The teacher read a book in the library* may be represented as follows ('S' stands for 'sentence'):

[S [NP [N The [N teacher]]] [VP [V read] [NP a [N book]]] [PP [P in] [NP the [N library]]] PP] VP] S]

Note the nesting of constituents within constituents in this sentence, e.g. the NP *the library* is a constituent of the PP *in the library* which is a constituent of the VP *read a book in the library*.

At the beginning of this section it was noted that the two aspects of syntactic structure, relational structure and constituent structure, are ‘distinct yet interrelated’, and it is possible now to see how this is the case. For example, a VP was described as being composed of a verb and the following NP, but it could alternatively be characterized as involving the verb and its direct object. Similarly, a PP is composed of a preposition and its object. NPs, on the other hand, involve modifiers, and accordingly the relation between *the* and *teacher* could be described as one of modifier-modified. Thus, these two aspects of syntactic structure are always present in a sentence, and when one or the other is emphasized, the sentence is being described from one of the two perspectives. It will be seen later that different grammatical phenomena seem to be more easily analyzed from one perspective rather than the other.

2.2 Lexical Categories

In the discussion of the constituents of sentences, reference has been made to nouns and noun phrases, verbs and verb phrases, and prepositions and prepositional phrases. Nouns, verbs and prepositions are traditionally referred to as ‘parts of speech’ or ‘word classes’; in contemporary linguistics they are termed **lexical categories**. The most important lexical categories are **noun**, **verb**, **adjective**, **adverb** and **prepositions** and **postpositions** (being subsumed **adposition**). In **traditional grammar**, lexical categories are given **notional definitions**, i.e. they are characterized in terms of their semantic content. For example, *noun* is defined as ‘the name of a person, place or thing’, *verb* is defined as an action word’, and *adjective* is defined as ‘a word expressing a property or attribute’. In modern linguistics, however, they are defined morpho-syntactically in terms of their grammatical properties.

Nouns may be classified in a number of ways. There is a fundamental contrast between nouns that refer uniquely to particular entities or individuals and those that do not; the best example of the first kind of noun is a proper name, e.g. *Sam*, *Elizabeth*, *Paris* or *London*, and nouns of this type are referred to as **proper nouns**. Nouns which do not refer to unique individuals or entities are called **common nouns**, e.g. *dog*, *table*, *fish*, *car*, *pencil*, *water*. One of the important differences between proper and common nouns in a language like English is that common nouns normally take an article, while proper nouns do not, e.g. *The boy left* versus **The Sam left* (cf. **Boy left* versus *Sam left*). Common nouns may be divided into **mass nouns** (or non-count nouns) and **count nouns**. Count nouns, as the name implies, denote countable entities, e.g. *seven chairs*, *six pencils*, *three dogs*, *many cars*. Mass nouns, on the other hand, are not readily countable in their primary senses, e.g. **two waters*, **four butters*, **six snows*. In order to make them countable, it is necessary to add what is sometimes called a ‘measure word’, which delimits a specific amount of the substance, e.g. *two glasses/bottles/drops of water*, *four pats / sticks of butter*, *six shovelfuls of snow*. Measure words can be used with count nouns only when they are plural, e.g. **six boxes of pencil* versus *six boxes of pencils*, **two cups of peanut* versus *three jars of peanuts*. **Pronouns** are closely related to nouns, as they both function as NPs. Pronouns are traditionally characterized as ‘substitutes’ for nouns or as ‘standing for’ nouns, e.g. *John went to the store*, and *he bought some milk*, in which *he* substitutes or stands for *John* in the second clause. This, however, is true only of **third-person** pronouns like *he*, *she*, *it*, or *they*; it is not true of

first-person pronouns like *I* or **second-person** pronouns like *you*. First- and second-person pronouns refer to or index the speaker and addressee in a speech event and do not replace or stand for a noun.

Verbs can likewise be categorized along a number of dimensions. One very important dimension is whether a verb takes just a subject (an **intransitive** verb), or a subject and a direct object (a **transitive** verb), or a subject, direct object and **indirect object** (a **ditransitive** verb). This will be referred to as the ‘valence’ of the verb. Another dimension concerns the kind of situation it represents. Some verbs represent static situations which do not involve anyone actually doing anything, e.g. *know* as in *Chris knows the answer*, or *see* as in *Pat sees Dana over by the bookcase*. Some symbolize actions, e.g. *run* as in *Kim ran around the track*, or *sing* as in *Leslie sang a beautiful aria*. Others refer to a change of state, e.g. *freeze* as in *The water froze* (the change in the state of the water is from liquid to solid), or *dry* as in *The clothes dried quickly* (the change in the state of the clothes is from wet to dry). Some represent complex situations involving an action plus a change of state, e.g. *break* as in *Larry broke the window with a rock* (Larry does something with a rock [action] which causes the window to break [change of state]). This classification of verbs is quite complex and is more appropriately in the domain of semantics rather than syntax.

Some examples of adjectives in English include *red*, *happy*, *tall*, *sick*, *interesting*, *beautiful*, and many others. Adjectives typically express properties of entities, e.g. *a red apple*, *a tall woman*, *a beautiful sunset*. Some properties are inherent attributes of an entity; for example, some apples are red because they are naturally so, whereas some barns are red because they have been painted red, not because they are inherently red. Hence color is an inherent property of apples but not of barns. Some languages signal this distinction overtly. In Spanish, for example, the adjective *feliz* means ‘happy’, and whether it is an inherent or permanent property of the person referred to is signaled by the verb it is used with, i.e. *Maria es feliz* ‘Maria is happy (a happy person)’ versus *Maria esta feliz* ‘Maria is happy (now, at this moment but not necessarily always)’. Spanish has two verbs meaning ‘be’, *ser* and *estar*, and one of the differences between them is that *ser* plus adjective (*es* in this example) is used to signify inherent or permanent attributes, while *estar* plus adjective (*esta* in this example) serves to indicate non-permanent, transitory attributes.

English adverbs typically, but not always, end in *-ly*, e.g. *quickly*, *happily*, *beautifully*, *rapidly* and *carefully*. *Fast* and *friendly* are exceptions; *fast* is an adverb without *-ly* (it can also be an adjective), and *friendly*, despite the admonitions of road signs in Texas to ‘drive friendly’, is an adjective, e.g. *a friendly waiter*. Adverbs modify verbs, adjectives and even other adverbs, and they can be classified in terms of the nature of this modification; manner adverbs, for example, indicate the manner in which something is done, e.g. *The detective examined the crime scene carefully*, or *The ballerina danced beautifully*, while temporal adverbs, as the name implies, express when something happened, e.g. *Kim talked to Chris yesterday*, or *Dana will see Pat tomorrow*. *Yesterday* and *tomorrow* do not end in *-ly* and have the same form when functioning as an adverb that they have when functioning as a noun, e.g. *Yesterday was a nice day*, *Tomorrow will be very special*. The most common adverbial modifiers of adjectives and adverbs are words like *very*, *extremely*, *rather*, e.g. *a very tall tree*, *the extremely clever student*, *rather quickly*. This class of adverbs is referred to as **degree modifiers**.

Prepositions are adpositions that occur before their object, while postpositions occur after their object. English (and Spanish) have only prepositions, e.g. English *in*, *on*, *under*, *to*,

(Spanish *en, a, con,*) whereas Japanese and Korean have only postpositions. German has both: *in dem Haus* ‘in the house’ (preposition *in*) versus *dem Haus gegenüber* ‘over across from the house’ (postposition *gegenüber*).

There are a number of minor categories. The category of **determiners** includes articles like *a* and *the*, and **demonstratives** like *this* and *that*. Determiners modify nouns in relation to their referential properties. Articles indicate roughly whether the speaker believes her interlocutor(s) can identify the referent of the NP or not; an **indefinite article** like *a(n)* signals that the speaker does not assume the interlocutor(s) can identify the referent of the NP, while a **definite article** like *the* indicates that the speaker does assume that the interlocutor(s) can identify it. Demonstratives, on the other hand, refer to entities in terms of their spatial proximity to the speaker; English *this* refers to an entity close to the speaker, while *that* refers to one farther away. (*Which book do you mean? This one here or that one over there?* versus **This one over there or that one here?*) Many languages make a three-way distinction: close to the speaker (English *this*, Spanish *esta* [FEM]), away from the speaker but not far (English *that*, Spanish *esa* [FEM]), and farther away from the speaker (archaic English *yon*, Spanish *aquella* [FEM]). These distinctions are also expressed by locative demonstratives, e.g. English *here*, German *hier*, Spanish *aquí* versus English *there*, German *da*, Spanish *ahí* versus English *yonder*, German *dort*, Spanish *allí*. **Quantifiers**, as the label implies, express quantity-related concepts. English quantifiers include *every, each, all, many, and few*, as well as the numerals *one, two, three, etc.*, e.g. *every boy, many books, the seven sisters*. **Classifiers** serve to classify the nouns they modify in terms of shape, material, function, social status and other properties. They are found in many East and Southeast Asian and Mayan languages, among others. They are similar in many respect to the measure words that occur with English mass nouns, but they occur with all nouns regardless of the count-mass distinctions. **Conjunctions**, like *and, but* and *or*, serve to link the elements in a conjoined expression. There are conjoined NPs, e.g. *a boy and his dog*, conjoined verbs, e.g. *Leslie danced and sang*, and conjoined adjectives, e.g. *Lisa is tall and slender*. All major lexical categories can be linked by conjunctions to form conjoined expressions. **Complementizers** mark the dependent clause is a complex sentence, e.g. English *that* as in *Sally knows that Bill ate the last piece of pizza*. The final category is **particles**, which is a classification often given to elements which do not fall into any of the other categories. Many particles have primarily **discourse** functions, e.g. English *indeed*, German *doch*, Spanish *entonces*.

There is an important opposition that divides lexical categories into two general classes, based on whether the membership of the class can readily be increased or not. Languages can usually increase their stock of nouns, for example, by borrowing nouns from other languages or creating new ones through **compounding** (e.g. *black + board* yields *blackboard*) or other morphological means (e.g. *rapid + -ly = rapidly*), but they do not normally create or borrow new adpositions, conjunctions or determiners. Lexical categories such as noun and verb whose membership can be enlarged are termed **open class** categories, whereas categories such as adposition, determiner or conjunction, which have small, fixed membership, are called **closed class** categories.

The definitions of lexical categories given so far are primarily the notional ones from traditional grammar. These definitions seem intuitively quite reasonable to speakers of Indo-European languages, and they seem to correlate nicely with the syntactic functions of the different parts of speech. Let us define three very general syntactic functions: **argument**, modifier and **predicate**. In a sentence like *the teacher read an interesting book, the teacher* and *an interesting book* are the arguments, *read* is the predicate, and *the, an* and *interesting*

are modifiers. Similarly, in *Kim is tall*, *Kim* is the argument and (is) *tall* is the predicate. The term ‘argument’ here includes NPs and PPs functioning as subject, direct object or indirect object. The notions of predicate and argument will be discussed in more detail in the following chapters, but for now one can say simply that in a sentence the predicate expresses the state of affairs that the referents of the arguments are involved in. (The terms ‘predicate’ and ‘argument’ are also used in semantics with a different meaning; they are being used here and elsewhere to refer to syntactic notions, unless otherwise noted.) It is usual to distinguish 1 -place, 2-place and 3-place predicates, depending on how many participants there are in the state of affairs depicted by the predicate. *Being sick* is a state of affairs involving only one participant, hence *be sick* is a 1-place predicate which takes one argument, e.g. *Kim is sick*. In *the teacher destroyed the note*, there is an action of destroying involving a *teacher* and a *note*. Destroying involves a destroyer and something destroyed; hence *destroy* is a 2-place predicate and takes two arguments. Finally, giving involves a giver, something given and a recipient, and therefore *give* is a 3-place predicate and takes three arguments, e.g. *The teacher gave an interesting book to Kim*. Given these distinctions, it seems intuitively clear that nouns would be arguments, verbs would be predicates and adjectives would be modifiers, and this is in fact the case very often. But not always. Nouns and adjectives can function as part of a predicate, as in *Dana is a phonologist* and *Chris was sick*. Even though they are part of the predicate, they are still formally distinct from verbs; they do not take **tense suffixes** like verbs do, i.e. **Dana phonologists* or **Chris sicked*. The **copula** *be*, a kind of verb, carries these verbal inflections. [...]

Every language has noun and verb as lexical categories. This reflects the fundamental role of reference and predication in communication. One of the most important functions of language is to allow speakers to depict states of affairs in the world, and in order for them to do this, there must be linguistic devices which refer to the participant(s) in a state of affairs and other devices which denote the action, event or situation in a state of affairs. Lexical items specialized for the first task are nouns, those specialized for the second are verbs.

What about the other major lexical categories? There are languages which lack adpositions altogether; they express the semantic content of prepositions and postpositions by means of the kind of **suffixes on nouns** such as in the Russian language. The concepts expressed by these endings are called ‘**case**’, and the endings are called ‘case markers’. Russian has both case suffixes and prepositions, but Dyirbal, an Australian Aboriginal language (Dixon 1972), has only case suffixes and no adpositions at all. Hence the lexical category ‘adposition’ is not universal. It also appears that adjective is not universal. In Lakhota, for example, the words expressing properties like ‘red’, ‘tall’, ‘big’, etc., are formally verbs and have basically the same morphosyntactic properties as verbs. [...]. Finally, there has been much less research done on adverbs cross-linguistically than the other major categories, and therefore it is difficult to draw any conclusions about their universality.

Thus, it appears that noun and verb are universal lexical categories, but adposition and adjective are not. It is crucial to keep in mind that when it is claimed that adjective is not a universally valid lexical category, it does not mean that there are languages which lack words expressing properties like ‘red’, ‘big’, ‘happy’, etc. Rather, it means that the words expressing these notions behave morphosyntactically like members of one of the other classes (verb in Lakhota, noun in Dyirbal and Quechua).