1 Some basic ideas in syntax

UNIT 1 Defining syntax
UNIT 2 Syntactic data
UNIT 3 Hypotheses

UNIT 1 DEFINING SYNTAX

Objectives:

- Understand the definition of "syntax."
- Understand the scientific method as applied to sentence structure.
- Understand the role of negative evidence.
- Understand the importance of structure in sentence construction.

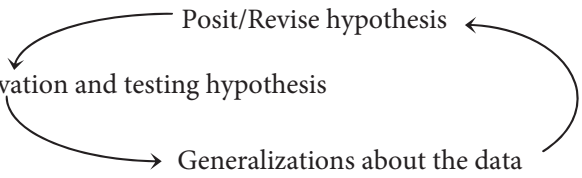
1.1 Introduction

Definition Syntax is the scientific study of sentence structure.

Comment There are two important parts to this definition: science and sentence structure. Let's look at each of these parts in some detail.

1.2 Science

Discussion When hearing the word *science*, what leaps to mind for most people are the hard sciences like biology, chemistry, physics. But the word *science* actually refers to a methodology for study. The study of syntax uses this methodology, so it is properly considered a science. The scientific method is expressed in the following diagram:



The scientific method involves taking a hypothesis about the subject matter, testing it by observing and gathering data, making generalizations about the patterns in that data and then revising the hypothesis to account for the new generalizations.

We are going to apply this definition to an example from syntax, but first we have to start with some definitions.

Definition A **declarative sentence** asserts that an event or state of affairs has occurred (or hasn't occurred): e.g. *Susan ate an apple. Susan didn't eat an apple.*

Definition A *yes/no* **question** is a question that can be answered by *yes*, *no* or *maybe*: e.g. *Did Susan eat an apple?*

Exercise **Q1** ✏[1] Identify which of the following sentences are *yes/no* questions, which are declarative sentences and which are neither.

(1) John hasn't eaten anything.	Y/N	Decl.	Neither
(2) Does Bill really prefer meatballs?	Y/N	Decl.	Neither
(3) Has Peter eaten his smoked salmon yet?	Y/N	Decl.	Neither
(4) What has Peter done now?	Y/N	Decl.	Neither
(5) Heather smokes too much.	Y/N	Decl.	Neither
(6) John did WHAT?	Y/N	Decl.	Neither

With this background about *yes/no* and declarative questions in mind, consider the following hypothesis:

Hypothesis 1: Yes/no *questions are formed by moving the second word in the equivalent declarative sentence to the front.*

Now look at the following sentences:

(7) Frodo will eat the magic beans. (declarative)

(8) Will Frodo eat the magic beans? (*yes/no* question)

Q2 ✏ Are sentences (7-8) consistent with hypothesis 1?
(Pay careful attention to the wording of the hypothesis!) Y N

Now consider the next two sentences

(9) The little hobbit will eat the magic beans. (declarative)

(10) Will the little hobbit eat the magic beans? (*yes/no* question)

Q3 ✏ Are these two sentences consistent with hypothesis 1?
(Pay careful attention to the wording of the hypothesis!) Y N

Q4 ✏ Instead of (10), what sentence does hypothesis 1 actually predict to be the grammatical *yes/no* question equivalent to (9):

(11) .....

In order to explain why the sentence you wrote above on line (11) is ungrammatical, but the one in (10) is OK, we will need to revise the hypothesis.

[1] This symbol means that this question can be answered directly in this book.

Q5 🕮[2] Try to come up with a hypothesis that accounts for the grammaticality of (10). (Hint 1: words such as *will* are called **auxiliaries.** Hint 2: use as much of the language in hypothesis 1 as you can, making only minimal changes.)

Hypothesis 2: Yes/no *questions are formed by moving . . . (complete this sentence)*

Comment Once a hypothesis has been revised, we re-evaluate it and see if it needs further revisions. This involves considering the **predictions** of the hypothesis. There are two kinds of predictions for syntactic theory. A good hypothesis will predict that some sentences are grammatical (more on this notion in unit 2), and others will be ungrammatical. For example, hypothesis 1 (incorrectly) predicted that sentence (11) would be grammatical, and sentence (10) would be ungrammatical. Hypothesis 2, by contrast, correctly predicts that sentence (10) is grammatical and sentence (11) is ungrammatical.

If we consider the case above again, we observe that often it is the *un*grammatical sentences that inform us as to how to revise our hypotheses.

Notation In syntax, ungrammatical sentences are marked with an asterisk (*).

Exercise Consider the following set of sentences:

(12) The hobbit who will dance at the party has eaten the magic beans.

(13) *Will the hobbit who dance at the party has eaten the magic beans?

(14) Has the hobbit who will dance at the party eaten the magic beans?

Q6 ✏ Does hypothesis 2 predict that sentence (13) will be grammatical? Y N

Q7 ✏ Does hypothesis 2 predict that sentence (14) will be grammatical? Y N

Comment In order to revise our hypothesis we're going to have to make reference to the **structure** of the sentence. The subject in sentence (12) contains a *relative clause*. We'll define relative clauses precisely in a later chapter. In this case, the relative clause is *[who will dance at the party]*. This relative clause is embedded (contained) in the main clause (the entire sentence).

Exercise **Q8** 🕮 Try to come up with a hypothesis that accounts for the grammaticality of (12) and (14). (Hint: you should refer to whether the auxiliary is embedded inside of a relative clause or appears in the main clause.)

[2] This symbol means that you should answer this question in your notebook or on a separate piece of paper.

Hypothesis 3: Yes/no *questions are formed by moving . . . (complete this sentence)*

Challenge Can you think of a sentence that hypothesis 3 *fails* to predict correctly? The sentence could be grammatical or ungrammatical. How would you revise the hypothesis to account for this sentence? (There is no feedback for this question.)

1.3 Structure

Comment This leads us to the second part of our definition of syntax: the structure of sentences. The formation of *yes/no* questions makes reference to whether the auxiliary is embedded or not. Sentences are hierarchically structured.

Comment If you add up the values of a series of numbers, it doesn't matter what order they are added in:

7 + 8 + 14 + 2 = 2+ 14 + 8 +7 = 8 + 7 +2 +14 etc.

This is a property of simple addition.

Exercise Take the following words:

yellow, singing, a, the, elephant, mouse, sniffed

Q9 ✏ Using each word only once, and using every word, try to come up with as many grammatical sentences as possible (there are at least eight; more may be possible). One is done for you as an example

(15) A singing elephant sniffed the yellow mouse.

(16) .....

(17) .....

(18) .....

(19) .....

(20) .....

(21) .....

(22) .....

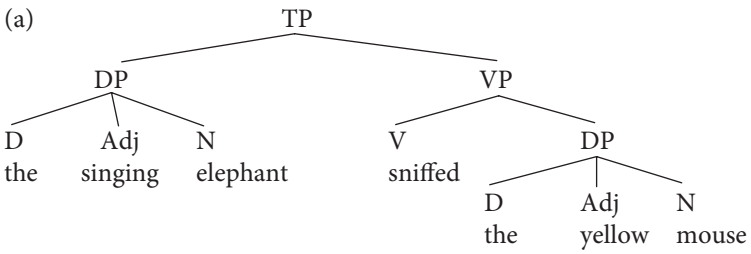
Q10 ✏ Do these sentences mean the same thing? Y N

Q11 ✏ Is syntax like addition, in the sense that order is irrelevant? Y N

Comment The fact that these sentences do not mean the same thing indicates that the sentence isn't just a linear string of symbols like addition. The position of the words relative to one another makes a difference in meaning. This, taken together with the importance of structure for hypothesis 3 above, is evidence

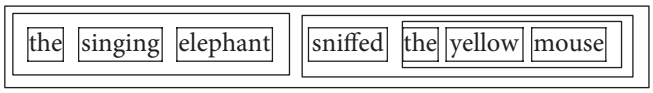
that sentences are structured entities. In particular we're going to claim that certain words are more closely grouped together than others. We can represent this with a tree structure (a), bracketing (b) or boxes (c) (we will return to the notations and the abbreviations you see below in detail in later parts of the book):

(a)



(b) $[_{TP}[_{DP}[_{D}$ the$][_{Adj}$ singing$][_{N}$ elephant$]][_{VP}[_{V}$ sniffed$][_{DP}$ $[_{D}$ the$][_{Adj}$ yellow$]$ $[_{N}$ mouse$]]]]$

(c)



Summary

In this unit, we've defined the study of syntax as the scientific approach to sentence structure. We looked at the scientific method, and how it involves looking for the predictions that a hypothesis makes in terms of the grammaticality and ungrammaticality. In probing this using *yes/no* questions as an example, we also found that sentences aren't merely strings of words, but are objects with structure. This was confirmed by noting that the meaning of a sentence is more than the sum of the meanings of its parts.

In the next unit, we'll look at where syntacticians get their data.

Suggested further reading

(full references are given at the end of the book)

- Carnie (2006), chapter 1
- Larson, (2010), units 1-2
- Sapir (1929)
- Tallerman (2005), chapter 1
- Wikipedia[3] article on syntax: http://en.wikipedia.org/wiki/Syntax

[3] Wikipedia is a mixed bag as far as academic sources go. Sometimes Wikipedia articles are well written and knowledgeable, and sometimes they definitely are not. Sometimes Wikipedia contains some helpful basic information, but you should always use it with caution.

- Wikipedia article on Scientific Method: http://en.wikipedia.org/wiki/Syntax
- Wikipedia article on Hypotheses: http://en.wikipedia.org/wiki/Hypothesis

Answers to questions

Q1 (2) and (3) are Y/N questions, (1) and (5) are declaratives, (4) and (6) are neither (they are questions but they cannot be answered with *yes*, *no* or *maybe*).

Q2 and Q3 Sentences (7) and (8) are predicted by the hypothesis: the first word in the declarative form is the second word in the Y/N question, and vice versa. Sentences (9) and (10), however, are not predicted. In sentence (10), it is the fourth word of sentence (9) that appears first.

Q4 and Q5 Hypothesis 1 predicts that the *yes/no* question form of sentence (9) would be **Little the hobbit will eat the magic beans.* The second word (*little*) is inverted with the first (*the*). Hypothesis 2 should be something like "*Yes/no* questions are formed by moving the auxiliary of the equivalent declarative sentence to the front."

Q6 and Q7 Hypothesis 2 isn't specific about which auxiliary in a sentence will move to the front. So both sentences are predicted to be grammatical. We need to revise our hypothesis to explain why (13) is ungrammatical.

Q8 Hypothesis 3 needs to make reference to the difference between the embedded auxiliary and the main clause auxiliary. There are several ways to phrase this, but one way would be: Yes/no *questions are formed by moving the main clause auxiliary to the beginning of the sentence.*

Q9, Q10 and Q11 You should have eight sentences, manipulating whether you use *a* or *the* to modify *elephant* or *mouse*, and the same for *yellow* and *singing*. (*A yellow elephant sniffed the singing mouse; A singing elephant sniffed the yellow mouse; A yellow mouse sniffed the singing elephant; A singing mouse sniffed the yellow elephant; The yellow elephant sniffed a singing mouse; The singing elephant sniffed a yellow mouse; The yellow mouse sniffed a singing elephant; The singing mouse sniffed a yellow elephant.)* Either noun phrase can be used in either subject or object position. The answers to the next two questions are both No.

UNIT 2 SYNTACTIC DATA

Objectives:

- Understand the role of corpora vs. judgment tasks.
- Learn to read and analyze foreign language examples.
- Use and apply judgment tasks.
- Distinguish syntactic from semantic judgment tasks.

Comment The scientific method requires data, so it's reasonable to ask how we gather that data and what kind of data we use. One obvious source of data is what we hear spoken around us or find written in books and newspapers.

2.1 Corpora

Definition A collection of written or spoken material representing real-world usage of a language is called a **corpus** (plural: **corpora**).

Discussion Corpora have a wide variety of uses, but also have a wide variety of limitations. In this unit, we will look at the role of corpora and the role of another data-gathering technique called the "judgment task" in the analysis of sentence structure.

Definition A corpus with an **interlinear gloss** or word-aligned gloss has three lines: (a) The example in the original language; (b) a word-by-word gloss, where the English for each word (or morpheme[1]) is aligned with the original language; (c) an idiomatic translation into English.

(1) (a) Níor bhuail mé Seán. *Actual language data*
(b) NEG strike I John *Word-by-word gloss*
(c) "I didn't strike John." *Idiomatic translation*

Discussion For most syntacticians the most important part of this is the *second line*: the word-by-word gloss. The glosses are lined up word for word (and sometimes morpheme for morpheme) with the foreign language on the line above. This line tells you (i) what each word in the foreign language example means, and more importantly, (ii) the order of the words in the foreign language.

[1] A morpheme is a word part such as a suffix or prefix or the root of a word.

When trying to determine the phrase structure of a foreign language or the behavior of a word or phrase, this is the line to look at! Remember: do not do an analysis of the idiomatic translation of the sentence, because then you are only doing an analysis of English!

Comment Sometimes you will also run into example sentences where there is no word-by-word gloss, only an idiomatic gloss. This often happens with "side-by-side" corpora. Often you can deduce the meaning of words that aren't glossed by comparing and contrasting these forms with other sentences that you already know the meaning of. Consider the Irish sentence in (2) and compare it to (1):

(2) Níor rith mé.
"I didn't run."

I haven't provided a word-by-word gloss here. However, from sentence (1) we see that *mé* is glossed as "I" and *Níor* is glossed as "NEG" (i.e. "not"). What's left is *rith*. Since the verb *bhuail* "strike" appears between the negation and the word meaning "I," we might conclude that *rith* is also a verb, and means "run" which we take from the idiomatic translation.

Exercise Consider the following sentences from Japanese.

Q1 ✏ Fill in the blanks for the meanings of the words without word-by-word interlinear glosses. For the moment ignore the *-ga*, *-o* and *-ni* particles.

(3) Taroo-ga Mieko-o mita.
Taro Mieko saw
"Taro saw Mieko."

(4) Taroo-ga Mieko-o sensei-ni shookaisata.
Taro Mieko teacher introduced
"Taro introduced Mieko to the teacher".

(5) Mieko-ga sensei-o mita.
Mieko ..... saw
"Mieko saw the teacher."

(6) Taroo-ga tuita.
Taro .....
"Taro arrived."

(7) Taroo-ga isu-ni. suwatta
Taroo ....
"Taro sat on the chair."

Comment Now let's figure out the meanings of the suffixes *-ga, -o* and *-ni*. First some definitions to help you; for now we'll assume an intuitive understanding of what a "subject" and a "direct object" and "indirect object" are. In the English sentence *Calvin gave the peanuts to Scott*, *Calvin* is the subject, *the peanuts* is the direct object, and *Scott* is the indirect object.

Definition **Nominative case** is the marking associated with subjects. For example, in English the pronouns *I, he, she, we* and *they* are in the nominative case; they only ever appear in subject position (in other words, before the verb).

Definition **Accusative case** is the marking associated with direct objects. For example, in English the pronouns *me, him, her, us* and *them* are in the accusative case. They appear in object position (in other words, after the verb).

Definition **Dative case** is the marking (often) associated with indirect objects and nouns marking the location of the event. In English we mark the dative case by adding prepositions such as *to, in, at* or *on* to the noun.

Exercise **Q2** ✏ Using the above terms and looking at sentences (3–7), define the Japanese suffixes:

-o .....

-ga .....

-ni .....

Exercise In English, the order of words is

Subject + Verb + Direct Object + Dative.

Q3 🕮 What is the order of words in Japanese?

Comment It is possible to extract a lot of information from corpora. For example, they often can give us a rough idea how frequent a particular construction is and how that construction is used relative to other constructions. Corpora are often a good place to look for counterexamples to negative predictions, for example, if the theory predicts that a sentence is impossible. Checking a corpus allows an efficient method of seeing if there are frequent counterexamples.

However, there are limits to the kinds of generalizations that can be found in corpora. The next exercise probes the advantages and limitations of corpus data.