

Introduction

Introduction to Syntax, EGG 2011, Lecture 1

July 25th, 2011

1 An exercise to start

Let's remind ourselves of what syntax is all about.

- Break up into groups of 3 or 4 – just the people sitting near you.
- Give a description of the structure of the sentences that follow in 1 below:
 - How do the pieces fit together?
 - Do some pieces belong more closely together than others?
 - Which pieces are more central or important to the sentence and how it is put together?
 - How do the various pieces function?
- Do not use **any** technical terms. Try to explain the structures in a way that a non-linguist would understand.

Exercise:

- (1)
 - a. Ellie ate all the bacon.
 - b. Anna thinks that Teddy is a friendly dog.
 - c. Cole seems to like everything.

2 What is Syntax?

Subfields of linguistics are normally divided up according to the levels of structure they deal with:¹

Phonetics: the physical properties of the sounds used in language: how they are produced, perceived and transmitted

¹Of course there are also subfields of linguistics which deal with the relationship between language and external factors. Historical linguistics deals with how languages change over time, sociolinguistics with how languages vary in relation to social factors, computational linguistics with how human language can be dealt with using computers and computer-related techniques and ideas, neurolinguistics with how language is manifested in the brain, etc.

Phonology: how these sounds are organized and utilized in language systems

Morphology: how words are put together out of smaller elements

Syntax: how sentences are put together out of their constituent elements

Semantics: how the sounds and structures of words and sentences are associated with their meanings

So syntax is about how sentences are put together. Before we get into details, let's clarify some terminology:

Utterance: a concrete instance of linguistic output, whether spoken, written, signed. . .

Sentence: an abstraction over a series of (potential) utterances whose forms are identical as far as is linguistically significant

Proposition: a message or chunk of meaning which is complete in the sense that we can in principle determine whether it is true or false in a particular context

- (2) John saw Stephan.
- (3) Jean a vu Stephan.
- (4) Stephan was seen by John.

- 2, 3 and 4 are three distinct utterances on the page. They are also three distinct sentences, but if spoken at the same time could express the same proposition.
- 2 (or 3 or 4) individually spoken at different times would represent distinct utterances, and could represent distinct propositions, but would always be the same sentence.

The basic data syntacticians work with are the sentences and the non-sentences of a given language.

- (5) John saw Stephan.
- (6) * John Stephan saw.
- (7) # The car saw Stephan.

- If an utterance does not constitute a sentence of the language in question, we say it is **ungrammatical** and mark it with a *.
- If it does constitute a sentence syntactically, but corresponds to an odd or nonsensical proposition, we say it is **deviant** and mark it with a #.

Modern theoretical syntax has as its goal a complete and explicit account of sentence structures in all languages.

- So we have to deal with big issues like word order:

(8) Her **finner** *du* publikasjoner. (Norwegian)

(9) Here *you* **find** publications. (English)

- But we also need to worry about things that may seem too obvious to mention, e.g. the fact that you can't say 10 if John is male:

(10) * John likes herself.

- And also stuff that is so weird and obscure that you may have never noticed it:

(11) Who do you think Ciaran will ask first?

(12) Who do you think that Ciaran will ask first?

(13) Who do you think will ask Seamus first?

(14) * Who do you think that will ask Seamus first?

3 The scientific method

Modern theoretical syntax approaches its goals from a scientific perspective. Being 'scientific' just means following the scientific method:

1. Gather some empirical observations X about a phenomenon.
 2. Propose hypotheses about why we observed X (and not some other imaginable things Y).
 3. Test these hypotheses on the basis of new observations Z.
 4. Revise the hypotheses to account for Z.
 5. Repeat steps 3 and 4 until all available data can be accounted for.
- ☞ This process generally does not terminate. Our understanding of the phenomenon gets better and better, but we can never be sure that our knowledge is complete.

To see the scientific method at work in syntax, let's consider yes/no question formation in English. We'll start with the following data:

- (15) Galahad can speak French.
(16) Can Galahad speak French?

These two sentences are clearly related:

- 16 questions the statement made in 15.
- In terms of form, the two sentences have exactly the same words, but they are in a slightly different order.

This is probably not an accident, so let's formulate a hypothesis to account for how the sentences are related to each other:

Hypothesis 1: In English, yes/no questions are like statements, but have the verb at the beginning of the sentence.

That handles 15 and 16, so let's test it on some new data.

- (17) * Speak Galahad can English.

- ☞ Our hypothesis predicts that this should be a valid way to form a question from 15 – we've moved a verb to the beginning of the sentence.
- ☞ The problem here is clearly that we've moved the wrong verb. So we need to modify our hypothesis to capture this restriction:²

Hypothesis 2: In English, yes/no questions are like statements, but have the **first** verb of the statement at beginning of the sentence.

Again, we're ready for some new data:

- (18) Annabel knows Galahad.
(19) * Knows Annabel Galahad?
(20) Does Annabel know Galahad?

- ☞ Hypothesis 2 predicts that the yes/no question formed from 18 should be 19, but 19 is not a sentence of English.
- ☞ The real yes/no question related to 18 is in 20, where the form *does* has appeared out of nowhere and gotten itself to the front of the sentence.

The issue here is the identity of the verb, *knows* as opposed to *can*. Consider questions with some other verbs:

²As we refine our hypotheses, I'll use **boldface** for things that change.

- (21) a. Ferris has eaten Pizza.
b. Has Ferris eaten Pizza?
c. * Does Ferris have eaten Pizza?
- (22) a. Ferris will eat Pizza.
b. Will Ferris eat Pizza?
c. * Does Ferris will eat Pizza?
- (23) a. Ferris ate Pizza.
b. * Ate Ferris Pizza?
c. Did Ferris eat Pizza?
- (24) a. Ferris likes Pizza.
b. * Likes Ferris Pizza?
c. Does Ferris like Pizza?

⇨ So-called **auxiliary verbs** like *can*, *have* and *will* can come at the front of questions, but normal verbs like *know* and *eat* can't.

So again, we have to adjust our hypothesis:

Hypothesis 3: In English, yes/no questions are like statements, but have the first **auxiliary** verb of the statement at the beginning of the sentence. **If there is no auxiliary verb in the statement, a form of *do* appears as auxiliary at the beginning of the question.**

We're far from finished here, but what we've got already works pretty well and gets the basic idea of English yes/no question formation down.

4 The problem of linguistic knowledge

Something that sets linguistics apart from other sciences like physics and biology is the nature of the object of study

- ☞ Unlike stars and algae, languages aren't independent physical entities out in the world.
 - ☞ A language is really a body of knowledge in the minds of its speakers.
 - ☞ Specific utterances can have a concrete existence in the world – in the form of sound waves or marks on paper.
 - ☞ But we're really more interested in the sentences that underly utterances, and these are abstract mental objects.
- ⇨ So what we really want to understand is not just the languages themselves, but the knowledge that people have of their languages.

But what kind of knowledge do speakers have?

- ? When I say that I can speak English, what does that imply about what is in my mind?
- ? And how did that knowledge get there in the first place?

5 Towards explanation

Linguistic knowledge is crucial to the explanations we're after in theoretical syntax. An influential proposal from Noam Chomsky evaluates theories based on this insight:

Observational adequacy: being able to simply identify which utterances correspond to grammatical sentences of a language

- I.e. accurately cataloguing what's out there in a concise and systematic fashion
- This takes a lot of work and forms the basis for everything else, but it doesn't necessarily reflect much deep insight.

Descriptive adequacy: additionally determining the structures for those sentences and showing how they are systematically related to each other

- I.e. a solid analysis of the systems at work in a given language, understood completely in its own terms.
- This still doesn't get at language in general or the role of knowledge.

Explanatory adequacy: additionally providing a framework which distinguishes possible languages from impossible ones and provides an explanation for how languages can be acquired

- I.e. setting each language in a universal, comparative context, showing what all languages have in common and how they can differ.
- An understanding of the knowledge underlying language and how it is acquired underpins the theory. All and only those languages which a child can acquire are the possible human languages.

Now, the toy investigation of yes/no questions we started above is going to have difficulty achieving explanatory adequacy.

- ☞ All of the hypotheses specifically mention English, because all of the data so far have come from English.

So it is appropriate to reconsider our hypotheses in terms of data from other languages.

How about some German sentences...

- (25) Reinhold kann Französisch.
Reinhold can French
'Reinhold can speak French.'
- (26) Kann Reinhold Französisch?
can Reinhold French
'Can Reinhold speak French?'

- ⇒ So (at least some) other languages form yes/no questions in a similar way, by moving a verb to the beginning of the sentence (with important differences of detail).

- ☞ In fact, similar patterns of question formation are found in lots of languages. There are languages that do things differently, but the variation seems to be limited.

- ☞ The following are commonly attested ways of forming yes/no questions:

- Move a verb to the beginning of the sentence (English, German...)
- Leave the word order the same as in statements, but add a special question-marker (Latin, Tamil).
- Change the morphology on the verb (Greenlandic).
- Use the same syntax as in a statement, but with a special question intonation (Modern Greek, colloquial French).

- ☞ But lots of other patterns we can imagine do not seem to be used in any language to form a yes/no question:

- Move one or more words to the **end** of the sentence
- Reverse the order of all the words in the sentence
- Change the morphology on the subject

So our theory should explain why so many languages show the same 3 or 4 patterns, and why other patterns are impossible.

6 Generative Grammar and Universal Grammar

The most influential approach to these issues is due to Noam Chomsky and his colleagues. Chomsky's idea is that the basic capacity for language is a human instinct, encoded in our genes.

- That is, we are born knowing something like a template for human languages, what Chomsky calls Universal Grammar (UG).
- When we learn a language as children, we only have to learn the details of how it fills in the UG template.

So for Chomsky, the goal of linguistic theory is to identify the properties of UG, and to determine the specific settings that yield languages like English, German and Chinese.

- ☞ The theory of a language is a model of the knowledge of its speakers, a set of rules that can account for – or **generate** – all the sentences of that language and no others.
- ⇒ Hence the Chomskyan enterprise as a whole is often called **Generative Grammar**.

Current versions of Chomskyan Generative Grammar instantiate the approach known as the **Minimalist Program** or simply **Minimalism**, which we will also pursue in this course.

7 Formalism and formal properties of language

Generative Grammar has an additional very important property: it is a **formal** theory.

- This means that rules and principles of the theory will be formulated in concise and precise terms, using a fair amount of mathematical notation.³
- E.g. rather than a prose description like 27, we might have a formal phrase-structure rule like 28:

³The word 'formal' has (at least) two distinct senses when used to describe theories, in particular linguistic theories, which often leads to confusion. In addition to the sense described in the main text, 'formal' can refer to a theory which manipulates symbols purely in terms of their form as defined within the theory, with no sensitivity to the meanings those symbols might have (and which by extension pursues the thesis that this is a productive way to approach its subject matter). Generative Grammar is also formal in this sense, but it is not always the case that a theory which is formal one sense is also formal in the other.

- (27) A sentence consists of a noun phrase followed by a verb phrase.
- (28) $S \rightarrow NP VP$

There are a number of good reasons for this:

- ☞ If written properly, formal rules are clear, concise and unambiguous.
- ☞ This makes them easier to test and potentially disconfirm, and it makes it easier to see what sort of unexpected consequences they might have.

Consider again English yes/no questions and our attempts thus far to explain them:

- Our hypotheses so far have been anything but formal, involving simple prose descriptions without any explicit definitions of the terms used.
- E.g., we haven't been clear about what an "auxiliary verb" is. Given this, it's not actually clear what our hypothesis has to say about sentences like 30:

- (29) Galahad wants to speak French.
- (30) *Wants Galahad to speak French?

- We also haven't clarified what the "first" auxiliary verb in a sentence means. A proper formal understanding of the issue is necessary to account for data like these:

- (31) The pie which is made with molasses is the most delicious.
- (32) *Is the pie which _ made with molasses is the most delicious?
- (33) Is the pie which is made with molasses _ the most delicious?

But it's not just clarity that makes a formal theory of language advantageous.

- ☞ Language has certain properties which are very naturally treated in formal terms, and would be rather cumbersome to deal with otherwise.

E.g., language exhibits clear evidence for hierarchical structure:

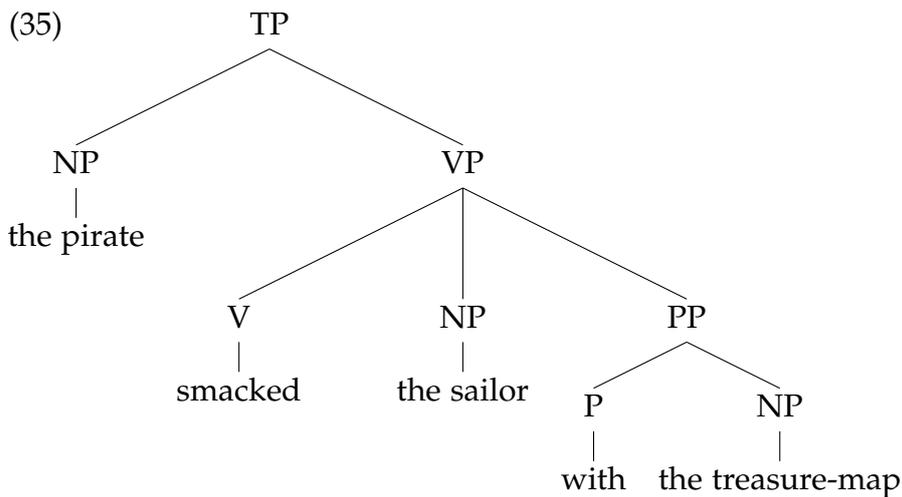
- (34) The pirate smacked the sailor with the treasure map.

Note that this sentence can mean two different things:

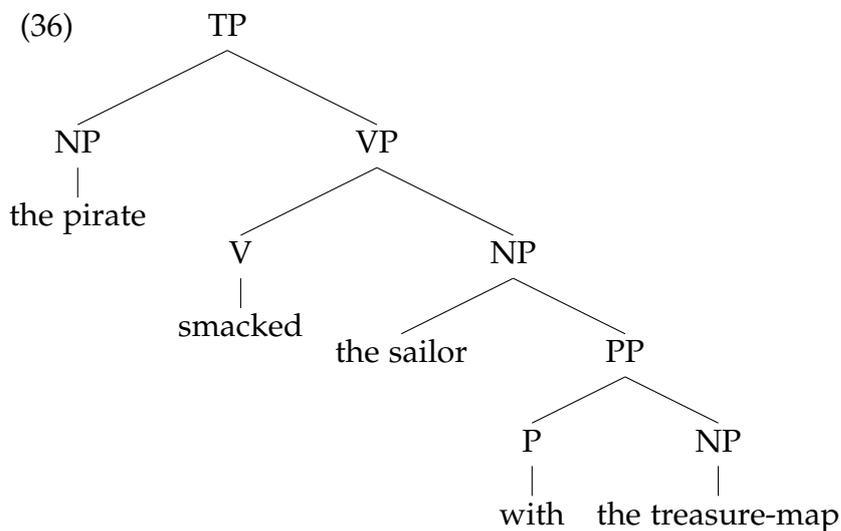
1. The pirate took the treasure map, rolled it up and used it to smack the sailor.
2. There was a sailor holding the treasure map, and the pirate smacked him (using the back of his hand or a bottle of rum or a dead parrot).

The best explanation for this ambiguity is that the two different meanings have structures which differ in terms of how *with the treasure map* fits in with everything else:

1. The PP is part of the VP. The treasure map was the instrument of the smacking.



2. The PP is part of the NP. The sailor who got smacked had the treasure map.



Language also exhibits rampant **recursion**, which is when something contains a copy of itself. Here's an example of visual recursion:

What's going on here?

- English sentences can be formed from a NOUN PHRASE followed by a VERB followed by another NOUN PHRASE, e.g. 37 or, with different words, *The cat saw the dog* or *The Flyers win the Stanley Cup*.
- But NOUN PHRASES can be more complex, e.g. including a PREPOSITIONAL PHRASE like *from the store*.
- Note, then, that a PREPOSITIONAL PHRASE actually consists of a PREPOSITION plus a NOUN PHRASE. As in 39, the NOUN PHRASE could itself contain another PREPOSITIONAL PHRASE. . .

Crucially, we can capture this complicated pattern with two extremely simple formal rules:

(40) $NP \rightarrow D N PP$

(41) $PP \rightarrow P NP$

- ☞ Describing this in terms of prose is actually much more difficult, takes up more space on the page, and is more likely to lead to misunderstanding.

8 A caveat on developing theories

In this course we'll be primarily dealing with theories, and syntactic theory is very much a work in progress.

- ☞ Even the most advanced theories out there – in linguistics and in other sciences as well – are incomplete, and on some points certainly wrong.
- ☞ It is always possible to find data from some language or another that don't fit with the a given theory and force us to revise it.
- ☞ But this isn't bad. It's totally normal, and in fact it's the way that we make progress.

The approach that we take here will basically follow what we just did with questions.

- We'll start with a very simple, basic set of hypotheses about how sentences are put together.
- Every day we'll consider new kinds of data that don't fit, and revise our hypotheses to account for them.

- In this way, we'll work our way up towards something like the theories that specialists in the field are working on.

So be prepared:

- ☞ At every step, you'll probably be able to come up with problems for the theory we've developed to that point.
- ☞ When you do, think about how we could modify the theory to handle it.
And bring it up in class.
- ☞ Of course, our analysis of particular data will evolve over the next two weeks.

This will take some time to get used to, but it's the only way to really understand how syntax and science in general are done.