Intro to Syntax, PART ONE

Omer Preminger, *MIT*

EGG 2009 / COST-A33, Poznań

Introduction to [Introduction to Syntax] 2 / 39

General stuff

- class in "parts"
- handouts
- English as the *object language*
- English as the teaching language
- ~~interruptions~~ questions/comments/contributions!
  - there's no part of syntax (or even linguistics in general) that you're supposed to "already know" before this course
  - ➢ if you don't understand something — ask!
- the ***why***

Generative Linguistics: An Absurdly Brief Overview 4 / 39

What is it all about? Some examples

- We possess ***tacit knowledge*** about language
  ≡ knowledge that we might not even know we have

(1) a. **Peter** forgets to lock the door every time **he** leaves the house.
b. **He** forgets to lock the door every time **Peter** leaves the house.

- ➢ *he* and *Peter* can ***co-refer*** in (1a), but not in (1b)
  - ***co-refer*** := refer to the same individual
- Is this about *precedence* ("what comes first")?

(2) a. Every time **Peter** leaves the house **he** forgets to lock the door.
b. Every time **he** leaves the house **Peter** forgets to lock the door.

- ➢ *he* and *Peter* can co-refer in both (2a) and (2b)
  - *precedence* cannot explain why (1b) is bad

What is it all about? Some examples

- Another example:

(3) a. The judges chose a picture of Stewie.
b. A picture of Stewie won first prize.

(4) a. Who did the judges choose a picture of?
b. * Who did a picture of win first prize?

[the asterisk (*) means that the sentence is judged as **unacceptable** by native speakers of the language]

- ➢ Is this about the linear position of *Stewie*/*who*?
  - ○ in other words, could the reason (4b) is bad be something like *"you can only ask about Stewie/who if it is at the end of the sentence"*?

(5) a. The judges chose a picture of Stewie to represent the school.
b. Who did the judges choose a picture of ____ to represent the school?

⇒ it is **not** the linear position of *Stewie*/*who* that makes (4b) bad

What is it all about? Some examples

- Yet another example:

(6) Peter wants to keep the car in the garage.

- This sentence has (at least) two possible meanings:
  (i) *Peter wants the car to continue to be in the garage.*
  (ii) *There is a car that Peter wants to retain ownership of; and that car happens to be in the garage.*
- Now consider (7):

(7) Which car does Peter want to keep in the garage?

  - ○ the meaning of (7) can be based on (i), but not on (ii)
    - – i.e., (7) cannot mean *Which car is the car that Peter wants to retain ownership of, and happens to be in the garage?*
      - · even though that is a completely coherent meaning, logically speaking

The learnability problem

- We've seen several examples that demonstrate:
  - we have certain kinds of knowledge about language that we're not necessarily conscious of
    - this knowledge is, at least in some cases, fairly intricate
      - resisting what might be considered "naïve" analyses

➢ **HERE'S THE KICKER:**

- Nobody's parents/teachers/TV shows them a sentence like (4b) . . .

(4) b. * Who did a picture of win first prize?

. . . and told them *"oh by the way, don't ever do this; it's not allowed"*

but note that:

1. there's no phonological problem in uttering (4b)
2. we can perfectly well understand what (4b) **would** mean, if it were grammatical

The learnability problem

- Similarly, nobody shows children sentences like (1b) and (7) . . .

(1) b. **He** forgets to lock the door every time **Peter** leaves the house.

(7) Which car does Peter want to keep in the garage?

. . . and says *"here's what a sentence like this **cannot** mean"*

⇒ ***SO HOW DO SPEAKERS COME TO HAVE THIS KNOWLEDGE?***

- POSSIBLE EXPLANATION:
  speakers know that these sentences are bad because they never encounter them in the course of their natural exposure to language

  (it's reasonable to claim that a sentence like (4b) is never encountered; as for (1b)/(7), we could say speakers never encounter them with the intended meaning being the one that is, in fact, impossible; though one might wonder about the relationship between *intended meaning* and *understood meaning*. . .)

➢ **BUT. . .**

The learnability problem

(8) A three-headed giraffe would make a formidable prime minister.

- I bet none of you have ever encountered (8), either
  - but you all know that (8) is a well-formed sentence in English

- speakers know, even when confronted with a sentence for the first time, whether it is well-formed or not
  - and what meanings it can and cannot have

⇒ ***SO AGAIN, HOW DO SPEAKERS COME TO HAVE THIS KNOWLEDGE?***

- This is an especially vexing question, given certain properties of the language acquisition process
  - This isn't a language acquisition course
    - Bart Hollebrandse is teaching one right here @ EGG!
  - ➢ but it's worthwhile to consider language acquisition, at this juncture...

An interlude on language acquisition

DISCLAIMER: Anything and everything I say here is superseded by what Bart Hollebrandse tells you!

Some properties of the language acquisition process:

I. Children acquire language blindingly fast, and at a very early age

A MENTAL EXERCISE:

- consider the degree of difficulty involved in learning a new language (right now, as adults)
- and imagine what might be a comparably difficult topic to learn, for example, in mathematics (*multi-variate integration*...?)

➢ how outlandish is it to imagine a 5-year-old who has mastered a mathematical task of this sort?

- yet almost all 5-year-olds have mastered at least one language
  - at a level that would make any foreign language learner proud

An interlude on language acquisition

II. There is a critical period, age-wise, for language acquisition (more or less until puberty)

- if the child receives no significant linguistic input by a certain age, she or he will be severely impaired in their linguistic capabilities
- this mirrors the state of affairs with other systems that are *biological*
  - e.g., the visual system and exposure to light

III. The so-called *"poverty of the stimuli"* properties

III.a. The child receives little-to-no *negative evidence*

- as mentioned earlier: situations where a child is presented with a sentence that is unacceptable (e.g., (4b)) — and told *"this is impossible"/"don't do this"* — are, at best, exceedingly rare

(4) b. * Who did a picture of win first prize?

An interlude on language acquisition

III.b. Moreover, even when systematic teaching is provided, children often ignore it

- especially if they are not at the proper stage of acquisition to absorb this knowledge

(9) **CHILD:** Want other one spoon, Daddy. *[Braine 1971:160–161]*
FATHER: You mean, you want the other spoon.
CHILD: Yes, I want other one spoon, please Daddy.
FATHER: Can you say "the other spoon"?
CHILD: Other… one… spoon.
FATHER: Say "other".
CHILD: Other.
FATHER: "Spoon".
CHILD: Spoon.
FATHER: "Other spoon".
CHILD: Other… spoon. Now give the other one spoon?

An interlude on language acquisition

III.c. The data to which the child is exposed contains production errors ("slips of the tongue"), aborted utterances, etc.

- Chomsky (1980) compares this to trying to learn the rules of chess by only observing a game, in which both players occasionally (albeit rarely) make moves that violate the rules
  - without necessarily indicating that a violation has occurred
- ➢ and of course, language is a much more complex system than the set of rules of chess
  - the rules of chess would probably fit on a single sheet of paper
    - if that was true for language, linguistics would be pretty boring!

An interlude on language acquisition

IV. Children acquiring different languages have been shown to go through certain uniform stages

- regardless of which language it is they are acquiring

V. Despite what seem to be pretty vast differences between different languages, children "figure out" the properties of the particular language they are acquiring

- at the same time, infants of course have the capacity to acquire **any** human language
  - if somebody switched 2 infants, one living in an English-speaking environment, and the other living in a Russian-speaking one —
    - each infant would, of course, perfectly master the "other" language, to which they were actually exposed

What we will and won't cover: Some important distinctions

- Again, this isn't a course about language acquisition
  - ⇒ we won't be focusing, for the most part, on the question of how speakers come to have this (tacit) knowledge about language
    - but rather, simply investigating what that knowledge is
- **BUT:** as you can see, the acquisition question is a very interesting one
  - Moreover, it's a very important criterion for evaluating theories of linguistic knowledge (so-called *"explanatory adequacy"*):
    - our theory of the adult speaker's linguistic knowledge should make it possible for a child to arrive at that knowledge
      - based only on the kind of data available during acquisition
- This is one example of a larger issue: **syntax does not exist in a vacuum**
  - it interacts with other modules of language
  - and language, in turn, interacts with other modules of cognition

What we will and won't cover: *syntax* vs. *semantics*

⇒ Some important distinctions to keep in mind:

I. *syntax* vs. *semantics*

(10) Colorless green ideas sleep furiously.

(11) Furiously sleep ideas green colorless. *[Chomsky 1957]*

- Neither (10) nor (11) is a felicitous English sentence; but:
  - (10) is just meaningless
  - (11), in addition to being meaningless, has other problems…
- ➢ We will be concerned with the kind of problems that plague (11), but not (10)
  - **what goes wrong with (10) — as well as with (12a–b) — falls outside the domain of *syntax***

(12) a. The square root of Milly's desk drinks humanity.

b. Being a theorem frightens consternation.

What we will and won't cover: *descriptive* vs. *prescriptive*

II. *descriptive* vs. *prescriptive*

- The term "grammar" is sometimes used to refer to something completely different
  - namely, how some people think you "ought" to speak, or what's acceptable in the eyes of your 5th grade English teacher
- So, for example, English speakers are told not to say — or at least, not to write — sentences like (13), which *end in a preposition*:

(13) What are you talking about?

- ➢ But every English speaker would instantly recognize (13) as part of their native language
- we are interested in the *descriptive* study of the knowledge that speakers actually have, not what others might try to impose upon them

⇒ these *prescriptive* rules are **not** part of what we'll be studying

What we will and won't cover: *competence* vs. *performance*

III. *competence* vs. *performance*

- When we are speaking, all kinds of things (linguistic and non-linguistic) might happen, that affect what we say:

(14) This is the

. . . speaker is then interrupted by loud sound in other room, or
. . . speaker then forgets what they wanted to say, or
. . . speaker then accidentally inhales a fly, etc. —
and the sentence is never completed.

  - at this point, we could (in principle) declare that *This is the* must be part of the English language

    (maybe with a footnote stating that it's only grammatical if you inhale a fly after saying it)

    - and that our grammatical theory must account for it
      - since, after all, a native speaker has uttered it

What we will and won't cover: *competence* vs. *performance*

- But it seems at least **as reasonable** to say that our grammar should only account for complete sentences
  - and that the grammar interacts with other facts about the world
    - in a way that sometimes gives rise to incomplete sentences, like *This is the*

➢ in fact, we probably **don't want** our theory of grammar to have to account for these outside factors (e.g., the flight patterns of flies)

⇒ we will make a distinction between —

***competence*:** what the grammar would produce in a perfect world

***performance*:** actual linguistic behavior — the result of grammar interacting with interruptions, limited memory, fatigue, inhaled flies, etc.

What we will and won't cover: *competence* vs. *performance*

- Another such example has to do with the length of sentences and noun-phrases, and/or the depth of embedding:

(15) a. John likes apples.
b. Mary hopes that John likes apples.
c. Bill knows that Mary hopes that John likes apples.
d. Sue doubts that Bill knows that Mary hopes that John likes apples.
⋮ ⋮

(16) a. John's friend
b. John's mother's friend
c. John's mother's sister's friend
d. John's mother's sister's fiancé's friend
e. John's mother's sister's fiancé's aunt's friend
⋮ ⋮

What we will and won't cover: *competence* vs. *performance*

- It's pretty clear that there's no fixed bound on the length of a sentence or noun-phrase
  - if you disagree, take what you think is *the longest possible sentence*, and add *She said that* in front of it, and… *voilà*: a longer sentence
- ➢ But of course, in practice, we cannot actually utter infinite sentences
  - you need to eat, people will stop paying attention to you, you will eventually die, etc.

⇒ **but it's not at all clear that we want our theory of grammar to account for things like the life-expectancy of the speakers**

- ➢ Again, it seems that in terms of *competence* (i.e., "in a perfect world"), there is no limit on the length of a sentence/noun-phrase
  - the fact that, in reality, there probably is an upper bound — after which you lose concentration, etc. — is a fact about *performance*

Other (possibly immaterial) distinctions

- Since we are trying to investigate the *competence* of speakers, but have access only to *performance*, this investigation is imperfect
  - this obviously includes asking people for grammaticality judgments

➢ But is this a particular property of some kinds of scientific investigation, and not others?

- Imagine an astronomer/astrophysicist looking through a telescope
  - Obviously, this person has indirect access to what is really the object of study (stars, planets, black holes, etc.)
  - Moreover, this access can be further affected by, e.g., the person's contact-lens prescription being outdated, or their eyes being tired

➢ Would you say that observations made through a telescope do not constitute ***external evidence*** for astronomy/astrophysics?

Other (possibly immaterial) distinctions

- In what sense, then, do psycholinguistic/neurolinguistic/acquisition data constitute ***external evidence***?
  - what are they *external* to?
  - and what, exactly, is this supposed to contrast with?
- All of these areas of research, just like theoretical linguistics, have their own kinds of confounds
  - and these confounds are not any more or less severe, a priori, than *performance* effects are to *competence*

⇒ so the only sense in which these are *external* is that they are (perhaps) *external* to theoretical linguistics (though they probably shouldn't be. . .)

- in which case, theoretical linguistics is also *external evidence*
  - *external* to neurolinguistics/psycholinguistics, that is

How are words organized into sentences?

HYPOTHESIS #1
a sentence is simply a collection of words

. . . and we're smart enough to figure out, from the meanings of those individual words, what their combination means.

(17) Stewie mocked Brian about his unfinished novel today. ≅
{*Stewie*, *mocked*, *Brian*, *about*, *his*, *unfinished*, *novel*, *today*}

- **PROBLEM:** order seems to matter for grammaticality
  (reminder: the asterisk (*) means that the sentence is judged as unacceptable)

(18) * Stewie novel about today his mocked Brian unfinished. ≅
{*Stewie*, *novel*, *about*, *today*, *his*, *mocked*, *Brian*, *unfinished*} = **(17)** ✗

⇒ if our goal is to account for the speakers' linguistic knowledge, **HYPOTHESIS #1** is insufficient
  – since speakers' knowledge includes the fact that (18) doesn't have the same status as (17)

How are words organized into sentences?

- **ANOTHER PROBLEM:** order seems to matter for meaning

(19) Stewie mocked Brian about his unfinished novel today. ≠
Brian mocked Stewie about his unfinished novel today.

but: {*Stewie*, *mocked*, *Brian*, *about*, *his*, *unfinished*, *novel*, *today*} =
{*Brian*, *mocked*, *Stewie*, *about*, *his*, *unfinished*, *novel*, *today*} ✗

- **HYPOTHESIS #1** predicts that the two utterances in (19) would be linguistically equivalent
  - but they are not

How are words organized into sentences?

HYPOTHESIS #2
a sentence is simply an *ordered* collection of words

(20) "Stewie mocked Brian." $\cong \langle Stewie_1, mocked_2, Brian_3 \rangle$

$\neq$

(21) "Brian mocked Stewie." $\cong \langle Brian_1, mocked_2, Stewie_3 \rangle$

$\Rightarrow$ this representation captures the fact that when we change (20) to (21), we change something rather profound about the utterance itself

- it is also able to distinguish (17) from (18):

(17) "Stewie mocked Brian about his unfinished novel today."
$\cong \langle Stewie_1, mocked_2, Brian_3, about_4, his_5, unfinished_6, novel_7, today_8 \rangle$

$\neq$

(18) * "Stewie novel about today his mocked Brian unfinished."
$\cong \langle Stewie_1, novel_2, about_3, today_4, his_5, mocked_6, Brian_7, unfinished_8 \rangle$

Beyond linear order

- Is there any reason to go beyond **HYPOTHESIS #2**?
  - **HYPOTHESIS #2** tells us that (17) is *different* from (18)... (which it is)

(17) Stewie mocked Brian about his unfinished novel today.

(18) * Stewie novel about today his mocked Brian unfinished.

  - but it doesn't tell us why (17) is *good* and (18) is *bad*
    - as opposed to, say, (18) being *good* and (17) being *bad*
  - ➢ we will come back to this issue a bit later

- ***but:*** there are other reasons why **HYPOTHESIS #2** is insufficient —

(17) Stewie mocked Brian about his unfinished novel today. $\stackrel{?}{=}$
Stewie mocked Brian about his unfinished novel today. $\stackrel{?}{=}$
Stewie mocked Brian about his unfinished novel today. $\stackrel{?}{=}$
$\langle Stewie_1, mocked_2, Brian_3, about_4, his_5, unfinished_6, novel_7, today_8 \rangle$

  - ***an intuition:*** not every substring of (17) is equal
    - A string like *his unfinished novel* forms a unit, in a way that *Brian about his* doesn't

Constituency: from intuition to evidence

- This *intuition* can be fleshed-out empirically:

(17) Stewie mocked Brian about his unfinished novel today.
Stewie mocked Brian about [his unfinished novel] today.
Stewie mocked [Brian about his] unfinished novel today.

I. SUBSTITUTION

II. MOVEMENT

III. FRAGMENT-ANSWER

(and there are more...)

(22) Stewie mocked Brian about it today.

(23) * Stewie mocked him/it unfinished novel today.

Constituency: from intuition to evidence

- This *intuition* can be fleshed-out empirically:

(17) Stewie mocked Brian about his unfinished novel today.
Stewie mocked Brian about [his unfinished novel] today.
Stewie mocked [Brian about his] unfinished novel today.

I. SUBSTITUTION

II. MOVEMENT

III. FRAGMENT-ANSWER

(and there are more...)

(24) ? His unfinished novel, Stewie mocked Brian about $\hat{}$ today
(...and his unfinished self-portrait, yesterday).

(25) * Brian about his, Stewie mocked $\hat{}$ unfinished novel today
(...and Meg about her/hers, yesterday).

- crucially, almost all speakers judge (24) to be better than (25)

[*the question-mark (?) means that the sentence is judged as* ***marginally acceptable*** *by native speakers of the language*]

Constituency: from intuition to evidence

➢ This *intuition* can be fleshed-out empirically:

(17) Stewie mocked Brian about his unfinished novel today.
Stewie mocked Brian about [his unfinished novel] today.
Stewie mocked [Brian about his] unfinished novel today.

I. SUBSTITUTION

II. MOVEMENT

III. FRAGMENT-ANSWER

(and there are more…)

(26) **A:** What did Stewie mock Brian about today?
B: His unfinished novel.

(27) **A:** …?
* **B:** Brian about his.

Constituency: from intuition to evidence

➢ This *intuition* can be fleshed-out empirically:

(17) Stewie mocked Brian about his unfinished novel today.

I. SUBSTITUTION

II. MOVEMENT

III. FRAGMENT-ANSWER

(and there are more…)

- These *diagnostics* distinguish some strings of words from others…
  … and each one seems, for the most part, to pick out **the same strings**
  ➢ the same strings about which we had the *intuition* that they constitute **cohesive units**

Constituency: from intuition to evidence

- We now have *evidence* that the grammar treats some sequences of words differently than others
  - and that some sequences of words form **cohesive units**

➢ we will call these units *constituents*

- How can a representation of sentences as a flat, ordered set of words (**HYPOTHESIS #2**) capture the existence of constituents?

(17) Stewie mocked Brian about his unfinished novel today. $\stackrel{?}{=}$
$\langle Stewie_1, mocked_2, Brian_3, about_4, his_5, unfinished_6, novel_7, today_8 \rangle$

  - after all, $Brian_3$ $about_4$ his_5 are three consecutive words in the ordered set, just like his_5 $unfinished_6$ $novel_7$

⇒ we need a more articulated representation of sentence structure

How constituents are organized

- What representation should we need to replace **HYPOTHESIS #2** with?
  - we need to explore what strings of words do and don't form constituents in various sentences
- ➢ It turns out that when we apply the diagnostics we've seen (and others), we find patterns like this:

(28) John read two very interesting books about linguistics last week.
John read two very interesting books about linguistics last week.

(each underline represents a constituent)

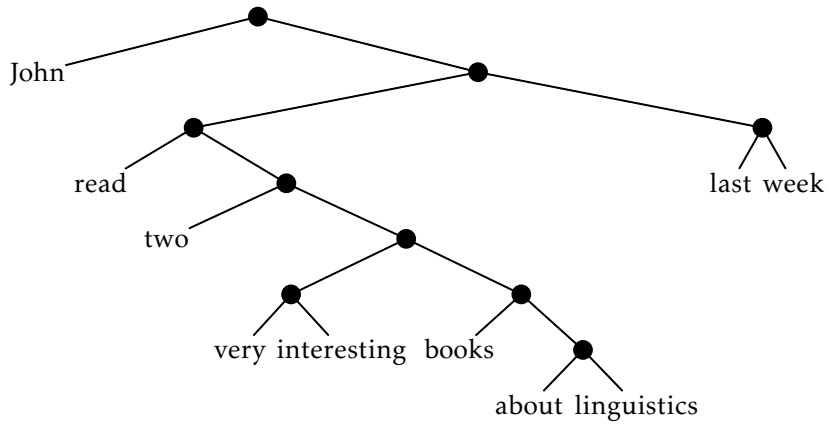
- ***observation:*** if a constituent α in (28) *overlaps* with a constituent β, then α and β stand in a *containment* relation
  - i.e., we don't find patterns of this sort: * word1 word2 word3
- ➢ in other words, constituents are always *nested* (in their entirety) within other constituents

How constituents are organized

- ➢ A good way to represent such a nested structure is with a *tree diagram*

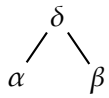
(28) John read two very interesting books about linguistics last week.

(29)

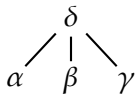


How constituents are organized

- the tree in (29) has another interesting property
  - every node has exactly two *daughters*
    - each sub-tree looks like, e.g., (30a)
    - and not, e.g., like (30b)

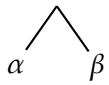
(30) a. 

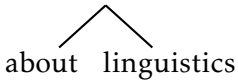
(α and β are *daughters* of δ)

b. 

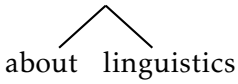
(α, β and γ are *daughters* of δ)

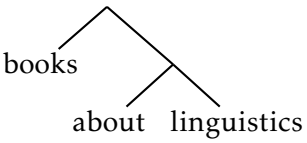
⇒ the entire tree can be constructed by successive application of a basic, building-block operation:

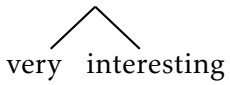
(31) *Merge*(α, β) = 

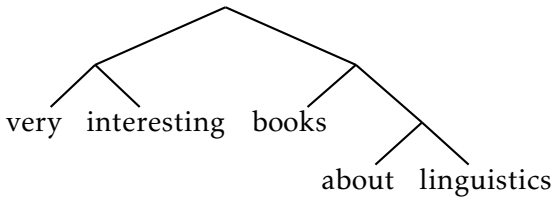
(32) *Merge*(about, linguistics) = 

How constituents are organized

(32) *Merge*(about, linguistics) = 

(33) *Merge*(books, (32)) = 

(34) *Merge*(very, interesting) = 

(35) *Merge*((34), (33)) = 

And so forth…

References

Abels, Klaus. 2008. *Introduction to Syntax*. Unpublished class materials, Debrecen: EGG.
Braine, Martin D. S. 1971. The acquisition of language in infant and child. In *The learning of language*, 7–95. New York, NY: Appleton-Century-Crofts.
Chomsky, Noam. 1957. *Syntactic structures*. The Hague: Mouton.
Chomsky, Noam. 1980. *Rules and representations*. New York, NY: Columbia University Press.
Richards, Norvin. 2008. *Introduction to Syntax*. Unpublished class materials, Cambridge, MA: MIT.
Siloni, Tal. 2003. *Introduction to Syntax*. Unpublished class materials, Tel-Aviv: Tel-Aviv University.

This is svn-revision 1082.