The hierarchical structure of sentences

Intro to Generative Syntax
Instructor: Suzana Fong

7–July–2025

1 Overview

- Recall that we were trying to model how English speakers form interrogative sentences and formulated the tentative rule in (1).

  (1) Place the *leftmost* verb *V* of the sentence in the *beginning* of the sentence. Subsequently, place an interrogative phrase *Q* to the *immediate left* of *V*.

- While (1) accounted for some interrogative sentences, it incorrectly yielded the sentence (2).

  (2) *Wrong prediction that (1) makes:*

  * **What** has [that Solfrid __ bought __] will surprise Seb?

- The problem with rule (1) is that it is stated in terms of **linear order**.
  - ▷ Evidence indicates that sentences actually have an underlying complex **structure**.

2 Linear order vs. structure

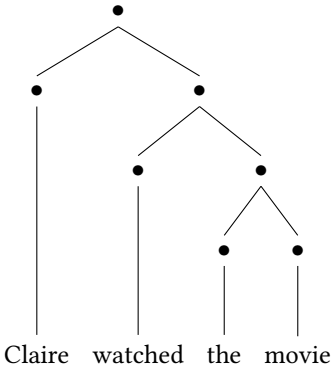
2.1 Introduction

- When we utter a sentence, one word follows another, in a linear string.
- Nonetheless, there is evidence that the syntax of natural languages operates in terms of **hierarchical structure**.
- Words are grouped together into **constituents** that exclude other parts of the sentence.

(3) a. Sentences are realized as a *linear string of words*...

Claire → watched → the → movie

b. ... but underlying them is a *hierarchical organization*.



- Is there evidence of underlying syntactic structure?
- Yes, from the interpretive effects of arranging the same words in a different way, so that the same string of words corresponds to different underlying structures, each with its own interpretation.

> **Exercise 1**
>
> The sentence *The enraged cow injured the farmer with an ax* is ambiguous. Can you paraphrase the two possible readings of this sentence?

(4) a.

'The cow injured the farmer who was holding an ax.'

b.

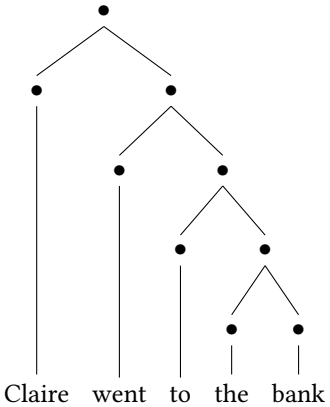
'The cow used an ax to injure the farmer.'

2.2 Accidental homophony

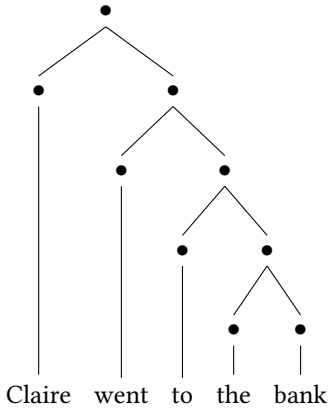
> **Exercise 2**
>
> A sentence such as *Claire went to the bank* is also ambiguous. Can you paraphrase the two possible readings of this sentence?

- A sentence such as *Claire went to the bank* is also ambiguous, though in a *different* way.
- The reason is that there are two words with the same form *bank*, but the syntactic structure is the same, in both readings paraphrased in (5).

(5) a. 'Claire went to a financial institution (e.g. to deposit a check).'



b. 'Claire went to the bank alongside the river.'



- ▷ *The enraged cow injured the farmer with an ax*: an example of structural ambiguity.
- ▷ *Claire went to the bank*: an example of accidental homophony.

Definition 1

Two linguistic units α and β are **accidentally homophonous** when α and β are independent from each other, but happen to have the same form or phonological realization.

Accidental homophony ≈ doppelgänger (i.e. different, in fact, independent individuals who happen to have similar external appearances).

(6)

(7) *Other examples of accidental homophony*

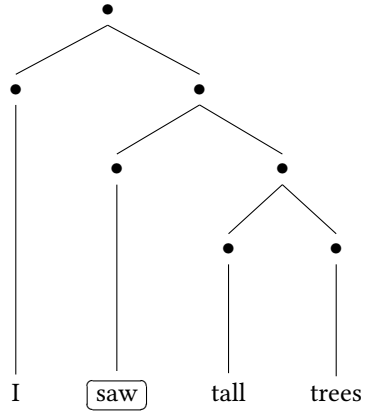
a. I saw tall trees.

b. Eyglo is looking for mixed nuts.

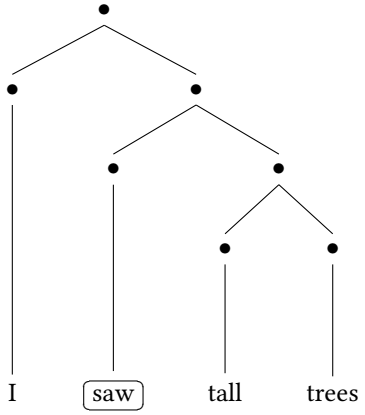
c. Eyglo is looking for a match.

(8) I saw tall trees.

a.



b.



Exercise 3

Explain the humorous effect of the following joke:

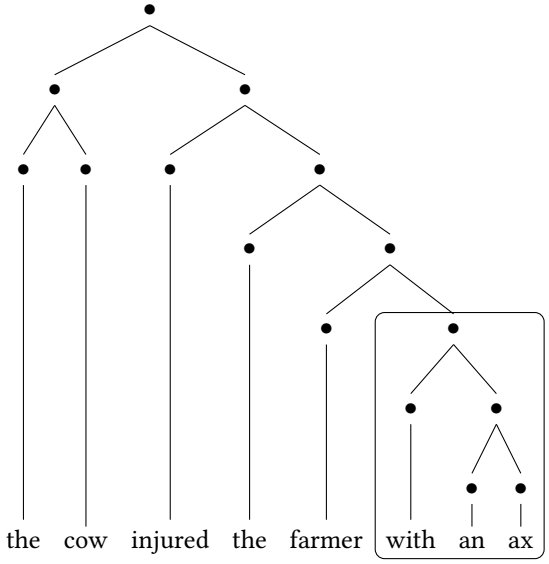
(9) A. How do you make a turtle fast?
B. Take away his food.

2.3 Structural ambiguity

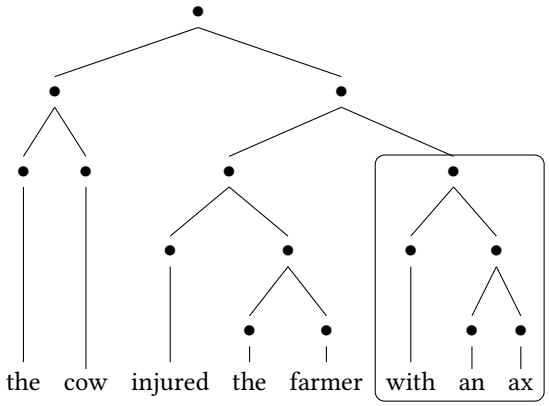
- Conversely, in *The enraged cow injured the farmer with an ax*, all words mean virtually the same.
- It is how *with an ax* combines with the rest of the sentence that causes the different readings.
- This is called *syntactic* or *structural ambiguity*.
- Intuitively, in the reading where the cow injured the farmer who was holding an ax (depicted in (4a)), *with an ax* specifies which farmer the cow injured (e.g. in the scenario there are two farmers, one is holding an ax and the other isn't holding anything).
  - ▷ There is no characterization about how the cow inflicted the injury.
- In the reading where the cow used an ax to attack the farmer (depicted in (4b)), *with an ax* characterizes the event of injuring the farmer.
  - ▷ There is no characterization of the injured farmer—we don't know what they were holding or if they were holding anything at all.

(10) The enraged cow injured the farmer with an ax.

a. 'The cow injured the farmer who was holding an ax.'



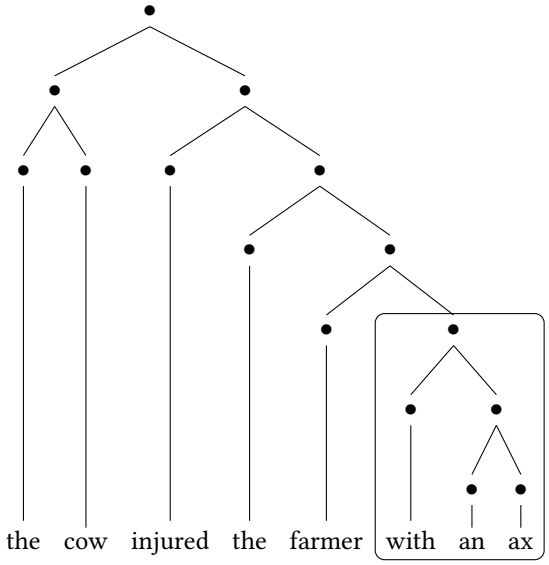
b. 'The cow used an ax to injure the farmer.'



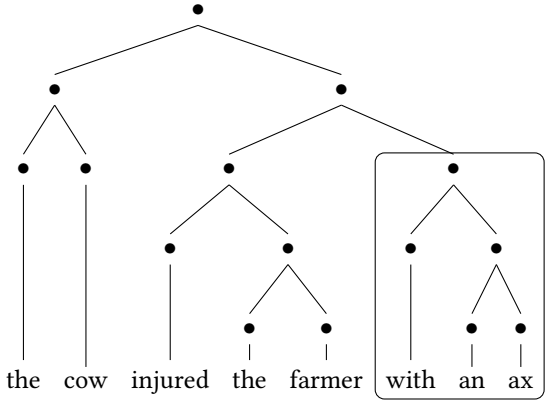
- ▷ One of the main goals of this course is to learn how to draw diagrams like (10a–10b).
- ▷ For now, it suffices to appreciate the fact that they represent different ways in which the components of the string *The enraged cow injured the farmer with an ax* are arranged, yielding a different structure and a corresponding reading.

- Importantly, despite the difference in arrangement and in internal structure, there is no difference in how the sentence is pronounced.

(11) a. The → cow → injured → the → farmer → with → an → ax



b. The → cow → injured → the → farmer → with → an → ax



Definition 2

A sentence is **structurally ambiguous** when there are different ways for its constituents to be arranged, yielding different configurations/structures, each with its own meaning. The difference in structure is *not* reflected in how the sentence is pronounced/realized.

Structural ambiguity ≈ *Lego Creator* (i.e. same set of Lego blocks, which you can build three different figures with)

(12)

a. 

b. 

c. 

Exercise 4

The sentences below are structurally ambiguous. Provide paraphrases for their readings.

(13) a. Flying planes can be dangerous.
b. The professor said she would give an exam on Monday.
c. Solfrid baked the cake in the freezer.
d. Visiting relatives can be boring.
e. Come meet our new French pastry chef.

2.4 Takeaway

- Sentences are realized as linear strings, where one word follows the other.
- However, if sentences were reduced to a linear string, we would not be able to account for the ambiguity that certain sentences such as *The enraged cow injured the farmer with an ax* exhibit.
- Sentences like this have more than one interpretation and, yet, they are realized by the *same* linear string.
- Conversely, if we assume that this sentence is ambiguous due to the way in which its subcomponents are arranged, giving rise to *structural ambiguity*, then we are able to account for the fact that there is a one-to-many relationship between the linear realization of a sentence like *The enraged cow injured the farmer with an ax* and its possible interpretations.

3 Constituency tests

- The paraphrases in (4a–4b) indicate that the string *The enraged cow injured the farmer with an ax* has more than one reading.
- But how can we probe into the internal structure that corresponds to each reading?
- Using **constituency tests**.

Exercise 5

Consider the following slightly modified versions of the ambiguous sentence we have been investigating:

(14) a. [The farmer [with an ax]], the cow injured.
b. [The farmer], the cow injured [with an ax].

These sentences (14a–14b) are no longer ambiguous. What is the only reading that each sentence has?

- (14a) only has the reading 'The cow injured the farmer who was holding an ax,' while (14b) only has the reading 'The cow used an ax to injure the farmer.'
- (14a–14b) are formed through a syntactic operation called **topicalization**, whereby some string is displaced and pronounced at the beginning of the sentence.

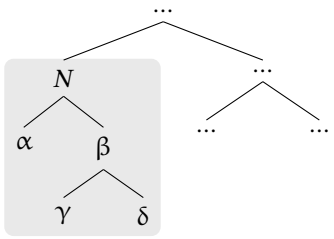
Topicalization *can only target constituents.* This means that:

- ▷ [The farmer [with an ax]] is topicalized in (14a), so it is a constituent.
- ▷ In (14b), only [the farmer] is topicalized. As such, it is a constituent separate from [with an ax].

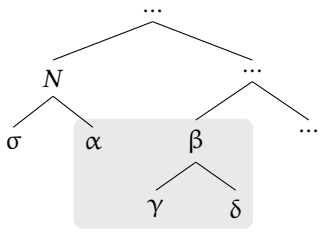
Definition 3

(15) a. Given a node N and a set of nodes $\{\alpha, \beta, \gamma, \dots\}$, then N is a **constituent** *iff* all members of $\{\alpha, \beta, \gamma, \dots\}$ are dominated by N and there is no node e such that e is dominated by N and e is not a member of $\{\alpha, \beta, \gamma, \dots\}$.
b. A node x *dominates* a node y *iff* there is one single descending line from x to y.

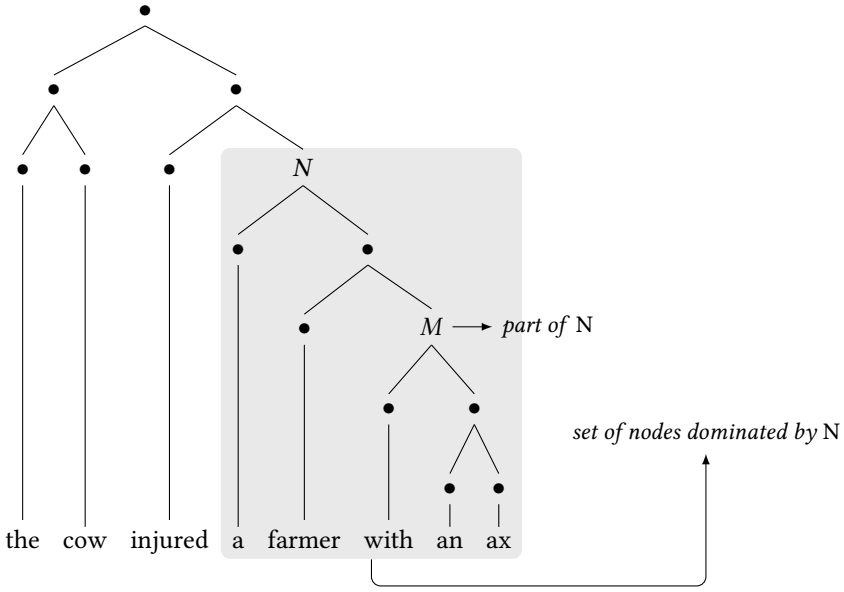
(16) a. $\{\alpha, \beta, \gamma, \delta\}$ *form a constituent, i.e.* N



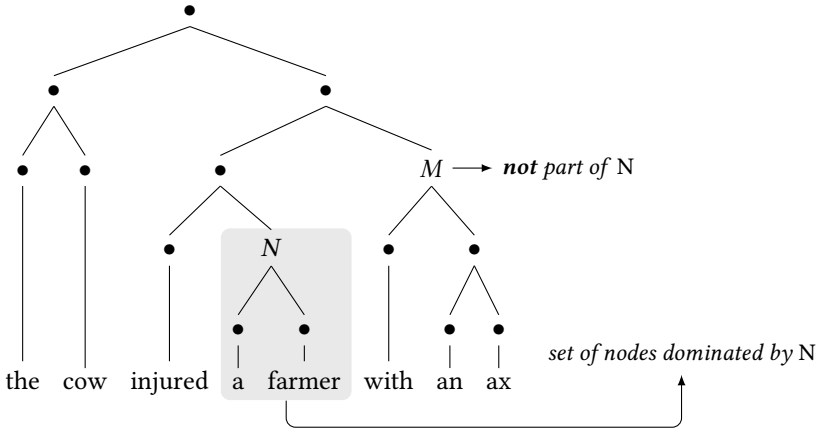
b. $\{\alpha, \beta, \gamma, \delta\}$ *do* ***not*** *form a constituent*



(17) 'The cow injured the farmer who was holding an ax.'



(18) 'The cow used an ax to attack the farmer.'



- Constituency tests such as topicalization allows us to identify the constituents of a sentence and, by extension, how they are arranged in the internal organization of each sentence.

3.1 Topicalization

- As we just saw, *topicalization* consists in fronting (i.e. moving to the beginning or the front of a sentence) a *constituent* to the beginning of the sentence.
- Here is another example:

(19) Tasneem sold an old collection of vinyls to the antique shop.

a. [An old collection of vinyls], Tasneem sold __ to the antique shop.

b. An old collection of vinyls to the antique, Tasneem sold __ shop.

c. [The antique shop], Tasneem sold an old collection of vinyls to __.

- Something else (19b) demonstrates is that linear order does not determine constituency, since the words in the string *an old collection of vinyls to the antique* follow one another in the baseline and, yet, the result of topicalization in (19b) is ungrammatical.

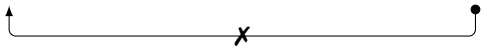
3.2 Clefting

- **Clefting** is a similar constituency test, in that, like topicalization, it involves the fronting of a constituent to the front of the sentence.
- However, in addition, clefting involves the expletive *it*, a form of the verb *be* (e.g. *is*, *was*, etc), and an appropriate relative word (e.g. *who*, *which*, *that*, etc).

(20) Tasneem sold an old collection of vinyls to the antique shop.

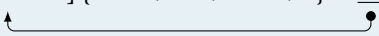
a. It was [an old collection of vinyls] which Tasneem sold __ to the antique shop.

b. It was an old collection of vinyls to the antique which Tasneem sold __ shop.



- More generally, clefting has the following template:

(21) It_{EXPL} be [*constituent*] {which/who/where/...} ... __ ...



  - ▷ The form of *be* matches the tense (and the agreement) of the verb in the baseline.
  - ▷ The choice of interrogative word is determined by the features of the constituent that was clefted (HUMAN = *who*, LOCATION = *where*, etc)

3.3 Proform substitution

- Yet another constituency test is **proform substitution**.
- A **proform** is a word like a pronoun (e.g. *he*, *she*, *it*, *them*, etc), as well as words like *then* or *there*, which stands for a constituent of a given type.

(22) Faatu will buy a new book tomorrow.

a. **She** will buy a new book tomorrow. [3SG FEM. HUMAN]

b. Faatu will buy **it** tomorrow. [3SG INANIMATE]

c. Faatu will buy a new book **then**. [TEMPORAL]

- Like other tests, proform substitution can only target constituents.

(23) Faatu will buy a new book tomorrow.

a. * Faatu will buy a new **it**.

b. * Faatu will buy a new **then**.

➟ *Book tomorrow* is not a constituent.

3.4 Fragment answer

- Only a constituent can be the answer to a question, while retaining the meaning of the original sentence.

(24) Faatu will buy a new book tomorrow.

a. Q: **Who** will __ buy a new book tomorrow? A: [Faatu]
b. Q: **What** will Faatu buy __ tomorrow? A: [A new book]
c. Q: **When** will Faatu buy a new book __? A: [Tomorrow]

- A string that is not a constituent cannot be the answer to a question.

(25) Faatu will buy a new book tomorrow.

Q: **What** will Faatu buy __? A: *A new book tomorrow

3.5 Interim summary

- While sentences are realized as strings of words and morphemes that are realized one after the other, linear order is not what determines how they are merged together in the underlying structure of a given sentence.
- Rather, words are grouped together in constituents.
- **Constituency** is evidenced by a few diagnostics. The ones we investigated were the following:
  1. Topicalization
  2. Clefting
  3. Proform substitution
  4. Fragment answer

Exercise 6

Apply constituency diagnostics in the sentence below, targeting the string in boldface in each case.

(26) a. **Claire** read a book about Stoicism in January.
b. Claire read **a book about Stoicism** in January.
c. Claire read a book about Stoicism **in January**.

Exercise 7

Consider the same sentence (26), repeated below.

(26) Claire **read a book about Stoicism** in January.

Is the string *read a book about Stoicism* a constituent? Apply the four constituency tests we just examined.

NB: you will need to use an appropriate form of the auxiliary *do* in the topicalization, clefting, and fragment answer tests.

3.6 Structural ambiguity and constituency tests

Exercise 8

The sentence (27) is structurally ambiguous.

(27) I called the student with a microphone.

What are its possible readings? Use constituency diagnostics to distinguish between them.

Takeaway

- Sentences can be structurally ambiguous, i.e. the words that compose it can be organized into different constituents. Each structure corresponds to a particular reading.
- Each structure and corresponding reading can be teased apart with constituency tests.