

FNLP Tutorial 3

Question 1. Constituency Tests

In the following sentences, are the underlined sequences constituents? Support your answer by referring to the constituency tests of substitution, (wh-)movement, coordination and clefting. Apply multiple tests where possible.

1. The lecture that the professor prepared was absolutely brilliant.
2. The lecture that the professor prepared was absolutely brilliant.
3. Students are reading the paper by Alan Turing.
4. I joined the meeting in my office on Tuesday.
5. I joined the meeting in my office on Tuesday.
6. I joined the meeting in my office on George Street.
7. The young man the GPUs.
8. The young man the GPUs.

Solution

1. Substitution: *It was absolutely brilliant.*
Clefting: *It was the lecture that the professor prepared that was absolutely brilliant.*
(Wh-)movement: *What was absolutely brilliant? The lecture that the professor prepared.*
Coordination: *The lecture that the professor prepared and the course manual were absolutely brilliant.*
2. Substitution: *The lecture that they prepared was absolutely brilliant.*
Coordination: *The lecture that the professor and the teaching assistant prepared was absolutely brilliant.*
Movement and clefting will yield ungrammatical results (e.g. **The professor, the lecture that prepared was absolutely brilliant.*, **It was the professor that the lecture that prepared was absolutely brilliant.*). If you are curious why that happens: it is because “the professor” is part of a relative clause island.
3. Substitution: *Students are reading it.*
Clefting: *It is the paper by Alan Turing that students are reading.*
(Wh-)Movement: *What are the students reading? The paper by Alan Turing.*
Coordination: *Students are reading the paper by Alan Turing and the book.*
4. Two readings of this sentence that impact constituenthood are 1) one joined the meeting while they were in the office (e.g. the meeting could have been online), 2) the meeting was in their office and they joined it. In the latter reading, this is a constituent.
Substitution: *I joined that on Tuesday.*
Clefting: *It was the meeting in my office that I joined on Tuesday.*
(Wh-)movement: *What did I join on Tuesday? The meeting in my office.*
Coordination: *I joined the meeting in my office and the lecture in Appleton Tower on Tuesday.*

5. This is not a constituent. “In my office” modifies the VP or “the meeting” (dependent on the interpretation), while “on Tuesday” modifies the VP. If this were a constituent, “on Tuesday” would modify “my office”, which is inaccurate.

At first glance, it may seem like substitution gives a desirable result: *I joined the meeting there.* However, “there” cannot substitute both the location and the timing of an event, as is the case in this sentence. Note that substituting with “there then” isn’t a legitimate application of the substitution test.

Similarly, clefting and (wh-)movement yield odd results: **It was in my office on Tuesday that I joined the meeting.*, **Where was it that I joined the meeting? In my office on Tuesday.*

What about coordination? Although we could form *I joined the meeting in my office on Tuesday and in Appleton Tower on Friday.*, sounding somewhat plausible, this is a corner case. As you may notice, whether the tests succeed or not is not always clear cut, and to some, the tests’ results may sound more acceptable than to others. When in doubt about whether something is acceptable or not, it can help to ask a native speaker of the language in which one performs constituency tests (although even native speakers may disagree).

6. In contrast to (5), (6) passes all three tests, because “on George Street” modifies the NP instead of the VP.

Substitution: *I joined the meeting there.*

Clefting: *It was in my office on George Street that I joined the meeting.* Movement: *In my office on George Street, I joined the meeting.*

Coordination: *I joined the meeting in my office on George Street and in the lecture hall on Princess Street.*

7. This is not a constituent. At first glance, it may seem like it, when reading “young” as the adjective and “man” as the noun but one will quickly discover that this is a garden path sentence.

Substitution: **He the GPUs.*

Clefting: **It was the young man that the GPUs.*

(Wh-)movement: **Who the GPUs? The young man.*

Coordination: **The young man and the old lady the GPUs.*

8. This, on the other hand, *is* a constituent. Notice that when clefting, the result is often very awkward when testing verb phrases: the remaining tests are better suited for this.

Substitution: *The young do so.*

Clefting: *?It is man the GPUs that the young do.*

(Wh-)movement: *What do the young do? Man the GPUS.*

Coordination: *The young man the GPUs and use the machines.*

Question 2. Attachment Ambiguity in Trees

A common source of attachment ambiguity in English comes from prepositional phrases, as you may have already noticed in Question 1. The relevant grammar rules include:

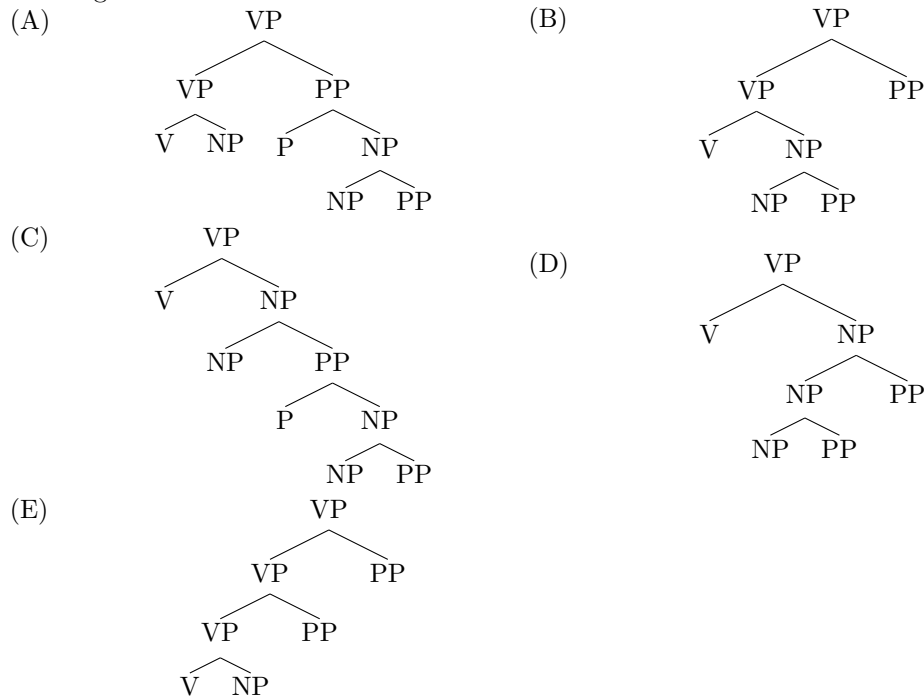
$$VP \rightarrow V NP$$
$$VP \rightarrow VP PP$$
$$NP \rightarrow NP PP$$
$$PP \rightarrow P NP$$

Here are five verb phrases:

1. read the paper about the Turing test on my laptop
2. read the paper about the success of Transformer

3. read the paper by Alan Turing about the Turing test
4. joined the meeting in the office with the blue door
5. joined the online meeting from my office via Zoom

Pictured below are five partial trees. Match the phrases to the trees which best capture their meanings.



Solution 1B; 2C; 3D; 4A; 5E

Question 3. Writing a context free grammar

Consider this small corpus of English noun phrases.

1. this book
2. the great green dragon
3. these green leaves
4. the leaves in the hallway

- Write a context-free grammar that generates at least the above noun phrases. Use part-of-speech tags as preterminal symbols, as we did in the lecture.
- Show a derivation tree of a grammatical noun phrase that is not included in the corpus. Does your grammar overgenerate? That is, does it generate sequences that are not grammatical English noun phrases? If so, refine your grammar to fix the issue that you mentioned.

Solution We will start with the following set of POS tags:

DT determiner, e.g. this and these

A adjective, e.g. great and green

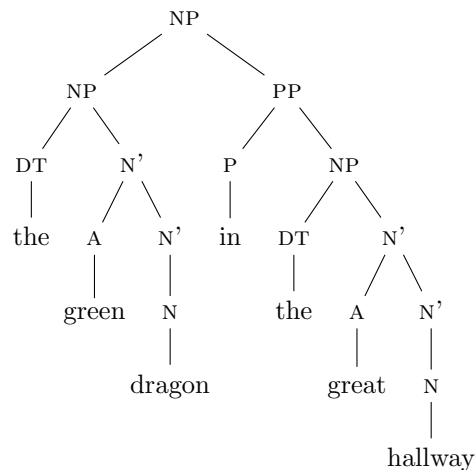
P preposition, e.g. in

N noun, e.g. book

Our first attempt might look something like this (NP is the start symbol). We write $X \rightarrow Y|Z$ as a shorthand for the two productions $X \rightarrow Y$ and $X \rightarrow Z$.

NP \rightarrow DT N'
NP \rightarrow NP PP
N' \rightarrow A N'
N' \rightarrow N
PP \rightarrow P NP
DT \rightarrow the|this|these
P \rightarrow in
A \rightarrow great|green
N \rightarrow book|dragon|leaves|hallway

We can derive the noun phrase "the green dragon in the great hallway" from this grammar:



Unfortunately, this grammar also generates ungrammatical noun phrases like:

- *this leaves
- *these great book

To prevent this from happening, our grammar must encode number information (if something is singular or plural). In particular, we need to encode this information for determiners and nouns and make sure that only singular determiners and singular nouns can combine (and analogously for plural).

We first refine our set of POS tags slightly to encode number information, i.e. we introduce DT_{SG} , DT_{PL} and N_{SG} and N_{PL} . We also modify production rules to track this information:

$$\begin{aligned} NP &\rightarrow DT_{SG} N'_{SG} \\ NP &\rightarrow DT_{PL} N'_{PL} \\ NP &\rightarrow NP PP \\ N'_{SG} &\rightarrow A N'_{SG} \\ N'_{PL} &\rightarrow A N'_{PL} \\ N'_{SG} &\rightarrow N_{SG} \\ N'_{PL} &\rightarrow N_{PL} \\ PP &\rightarrow P NP \\ DT_{SG} &\rightarrow \text{the|this} \\ DT_{PL} &\rightarrow \text{these} \\ P &\rightarrow \text{in} \\ A &\rightarrow \text{great|green} \\ N_{SG} &\rightarrow \text{book|dragon|hallway} \\ N_{PL} &\rightarrow \text{leaves} \end{aligned}$$

This grammar still generates noun phrases that are disfavoured if not ungrammatical, like:

- ??the green great dragon