Exercise Sheet 2 – Linguistic Concepts – Solutions

1. Learning Objectives

In this sheet we are going to:

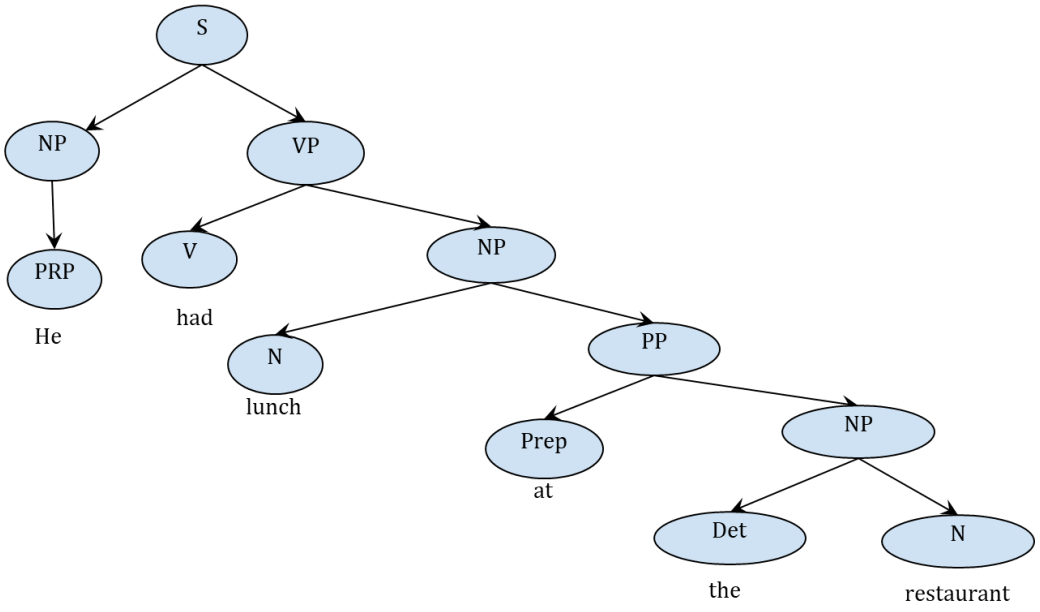
- know more about linguistic structure, analysis and data
- do some exercises on basic linguistic notions
- practice some tasks on linguistic analysis

2. Formal Grammar

Define a formal grammar (grammar rules + lexicon) that can be used to analyse/generate the following sentence:

He had lunch at the restaurant.

Refer to slides 33 - 35



Grammar Rules

S -> NP VP
NP -> PRP
VP-> V NP
NP-> N PP
PP-> P NP
NP-> DT N

Lexicon

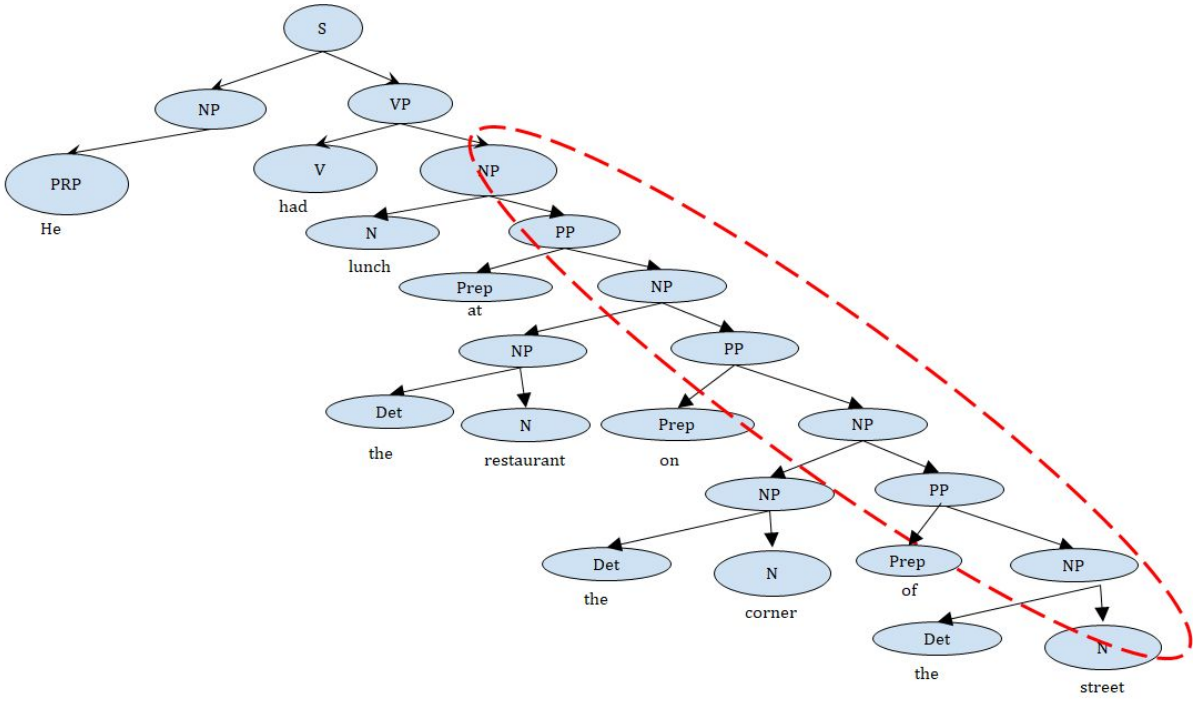
Noun -> lunch, restaurant
Verb -> had [“to have “, 3rd, pres]
Preposition -> at
Determiner -> the
Pronoun -> he

3. Formal Grammar (recursion)

Extend the grammar so that it can be used to analyse/generate the following sentence:

He had lunch at the restaurant on the corner of the street.

Refer to slides 39 & 40



Additional Rules:

NP-> NP PP

Recursive grammar rules:

PP-> P NP
NP-> NP PP

4. Part of Speech Tagging

Use the PoS tag set from the slides to annotate each word in the following sentence with the correct part of speech:

He had an expensive, but very good lunch at the Thai restaurant with the big windows that is opposite the church.

Refer to slides 33

He(PRP) had(VBD) an(DT) expensive(JJ), but(CC) very(RB) good(JJ) lunch(NN) at(IN) the(DT) Thai(**JJ**/NNP) restaurant(NN) with(IN) the(DET) big(JJ) windows(NNS) that(Wh-Det) is(VBP) opposite(IN/RB) the(DET) church(NN).

note the ambiguity of "Thai" and "opposite".

6. Morphology

Refer to slide 17, 18 & 19

6.1. Can you find the stem the following words: amusing, amusement, and amused?

the stem would be **amus**.

6.2. Can you get the lemma for the following words: amusement, amusing, and amused?

the lemma for each and all would be **amuse.**

Some languages have complex morphology such as the Arabic language. Consider the word

وسيكتبونها

/wasayaktubuwnahA/

6.3. Can you guess the PoS tag for this word?

although it is a single word, but it translates to a phrase which is "and they will write it". Segmenting it will lead to the following PoS tags.

و+ س + يكتب + ون + ها

wa+sa+yaktub+oun+ha

and+will+write+(they)+it

CC+RP+VBP+PRP

6.5. Name some applications that might use these morphological processes?

Information Retrieval (query expansion in search engines), Feature Extraction, Information Extraction, Language modelling.

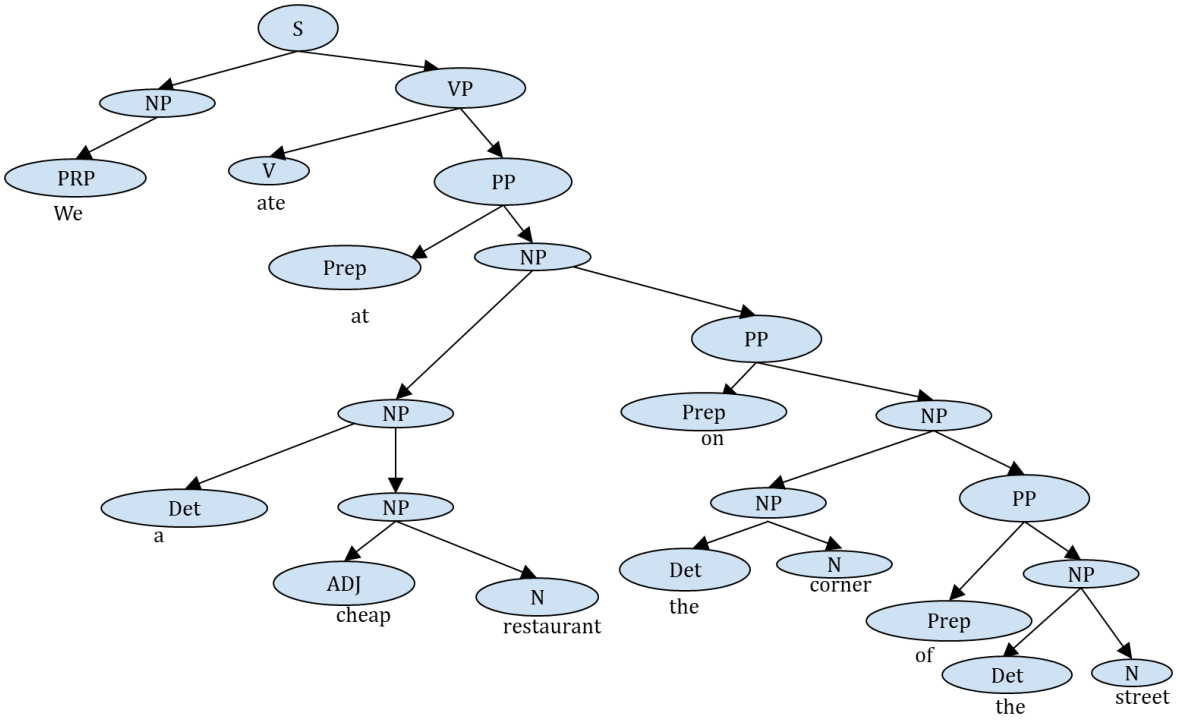
7. Phrase/Constituency Structure vs. Dependency Structure

Provide a phrase/constituency structure (tree) and dependency structure (tree) for the following sentence.

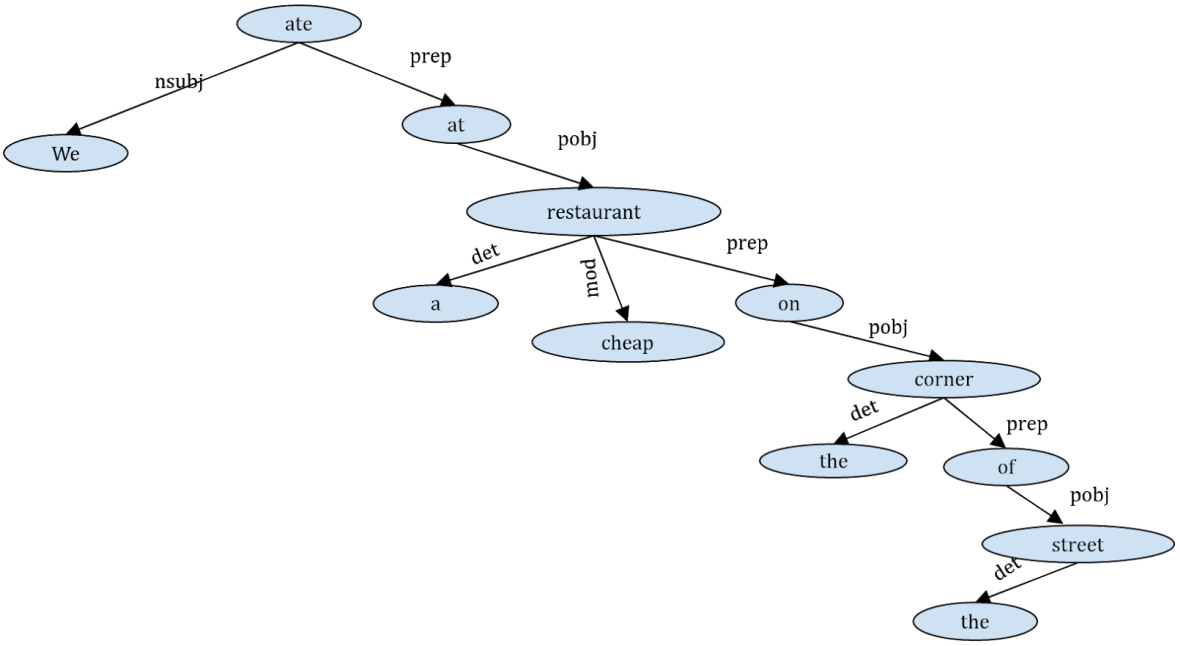
We ate at a cheap restaurant on the corner of the street.

Refer to slides 34 & 41

Phrase structure:



Dependency Structure:


ate
nsubj
prep
We
at
pobj
restaurant
det
mod
prep
a
cheap
on
pobj
corner
det
prep
the
of
pobj
street
det
the