

# L95: Introduction to Natural Language Syntax and Parsing

## Lecture 3

Simone Teufel

Department of Computer Science and Technology  
University of Cambridge

Michaelmas 2019/20

# Organisational

- Read Section 5.1 to 5.6
- Do Exercises 1-3 in section 5.9
- Familiarise yourselves with PTB Guidelines
- Submit Assignment 1 (POS-tagging) on Monday 12noon
- Work through logic worksheet
- Today: Assignment 2 (Phrase structure analysis of NPs)

# About Assignment 1/Recap

- Tokenization
- Ambiguity
- Idioms
- Multi Word Units
- Finite vs. non-finite forms of the verb
- Use vs Mention: quoted material, titles, ...
- **Discuss** cases of uncertainty

# Particles vs. Prepositions

- The man up the ladder fell.
- Kim ran up the stairs.
- Kim ran up a large bill.
- Kim slipped up.
- Kim washed up the dishes.
- Kim washes the dishes up.

# Lexical Features

Feature	Values	Examples
Num(ber)	Sg / Pl	boy(+s)
N-type	Mass, Count, Name	boy, research *research+s, Fred
Per(son)	1,2,3	I (1sg), you (2sg) (s)he (3sg)
Case	Nom, Acc	he (nom), him (acc)
Valence	Intrans, Trans, Ditrans, Scomp,...	smile, kiss, give, believe...
A-type	base / comparative / superlative	- old, older, oldest

Tag	Description	Example	Tag	Description	Example	Tag	Description	Example
CC	coordinating conjunction	<i>and, but, or</i>	PDT	predeterminer	<i>all, both</i>	VBP	verb non-3sg present	<i>eat</i>
CD	cardinal number	<i>one, two</i>	POS	possessive ending	<i>'s</i>	VBZ	verb 3sg pres	<i>eats</i>
DT	determiner	<i>a, the</i>	PRP	personal pronoun	<i>I, you, he</i>	WDT	wh-determ.	<i>which, that</i>
EX	existential 'there'	<i>there</i>	PRP\$	possess. pronoun	<i>your, one's</i>	WP	wh-pronoun	<i>what, who</i>
FW	foreign word	<i>mea culpa</i>	RB	adverb	<i>quickly</i>	WP\$	wh-possess.	<i>whose</i>
IN	preposition/ subordin-conj	<i>of, in, by</i>	RBR	comparative adverb	<i>faster</i>	WRB	wh-adverb	<i>how, where</i>
JJ	adjective	<i>yellow</i>	RBS	superlatv. adverb	<i>fastest</i>	\$	dollar sign	\$
JJR	comparative adj	<i>bigger</i>	RP	particle	<i>up, off</i>	#	pound sign	#
JJS	superlative adj	<i>wildest</i>	SYM	symbol	<i>+, %, &amp;</i>	“	left quote	‘ or “
LS	list item marker	<i>1, 2, One</i>	TO	“to”	<i>to</i>	”	right quote	’ or ”
MD	modal	<i>can, should</i>	UH	interjection	<i>ah, oops</i>	(	left paren	[, (, {, <
NN	sing or mass noun	<i>llama</i>	VB	verb base form	<i>eat</i>	)	right paren	], ), }, >
NNS	noun, plural	<i>llamas</i>	VBD	verb past tense	<i>ate</i>	,	comma	,
NNP	proper noun, sing.	<i>IBM</i>	VBG	verb gerund	<i>eating</i>	.	sent-end punc	. ! ?
NNPS	proper noun, plu.	<i>Carolinas</i>	VBN	verb past part.	<i>eaten</i>	:	sent-mid punc	: ; ... --

**Figure 8.1** Penn Treebank part-of-speech tags (including punctuation).

Eh?

From the film “The Martian”.



(An instance of productive derivational morphology (zero-derivation).)

## Moving on...

- Constituency
- Phrase structure Grammar
- Phrase structure Trees
- Tests for Constituency



# Linguistic Methodology

- Descriptive, not prescriptive
- Example: should we (or should we not) split infinitives?
  - Captain Kirk has gone beyond our galaxy
  - Captain Kirk's mission is to go beyond our galaxy
  - Captain Kirk has been travelling the universe for 30 years
-

# Linguistic Methodology

- Descriptive, not prescriptive
- Example: should we (or should we not) split infinitives?
  - Captain Kirk (**boldly**) has (**boldly**) gone beyond our galaxy
  - Captain Kirk's mission is (**boldly**) to (**boldly**) go beyond our galaxy
  - Captain Kirk (**boldly**) has (**boldly**) been (**boldly**) travelling (**\*boldly**) the universe for 30 years
- Which rule can we derive from this?

# Distributional analysis

- Algorithm:
  - Create a template
  - Perform substitutions
  - Test for grammaticality
  - Ungrammaticality, semantic oddness/implausibility
- The \_\_\_\_\_ can run
- \_\_\_\_\_ can run
- collect an equivalence class of strings that can go in the slot.
- They are called Constituents
- This is an entirely bottom-up approach championed in the 1930s
- Today, two different dominant approaches:
  - Generative grammar (phrase structure)
  - Headedness (relational)

# Generative Methodology

- Noam Chomsky (1957): Syntactic Structures
- Finite sets of rules predict all and only grammatical sentences
- Generative grammars: mappings between sentences and meaning

# A context-free grammar

## Rules:

S → NP VP  
VP → VP PP  
VP → V  
VP → V NP  
VP → V VP  
NP → NP PP  
PP → P NP

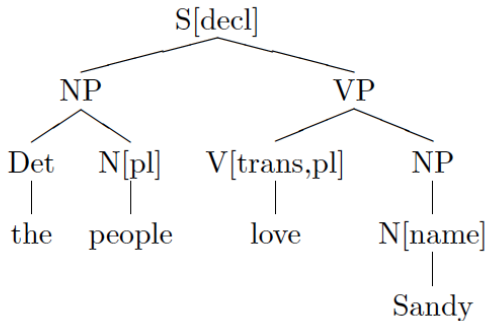
## Lexicon:

V → can  
V → fish  
NP → fish  
NP → they  
NP → rivers  
NP → pools  
NP → December  
NP → Scotland  
NP → it  
P → in

# Headedness

- The heads of NPs are nouns.
- Heads are the only constituent of a constituent that cannot be dropped:
  - The castle is old
  - \*The is old
  - The big castle is old
  - \*The big is old
  - The castle by the hill is old
  - \*The by the hill is old
  - Castles are interesting

# Phrase Marker Trees



# Constituency Tests

## **Substitution test**

- use “proform” (eg “do so” stands in for a VP; eg. “that” stands in for an NP)
- If substitution is felicitous, then phrase is a constituent (of same category as the proform).
- What are other NPs (like “the people”)?
- What are other transitive verbs (like “love”)?



# Constituency Tests

## **Movement test**

- Constituents can be moved around in the sentence.
  - The old man has come to dinner.
  - Has the old man come to dinner?
  - \*The has old man come to dinner

# Constituency Tests

## Insertion test

- Appositions are parentheticals.
- They cannot be inserted into constituents, only at the end of constituents.
  - The President of America, Ronald Reagan, is over 70.
  - \*The President, Ronald Reagan, of America is over 70.
  - \*The President of America is, Ronald Reagan, over 70.

# Constituency Tests

## **Omissibility test** (only suitable for some constituent types)

- Some constituents can be omitted
- Non-constituents cannot be omitted
  - Some friends of the old man came to dinner.
  - Some friends came to dinner.
  - \*Some friends man came to dinner.

# Constituency Tests

## **Coordination test** (well-known exceptions)

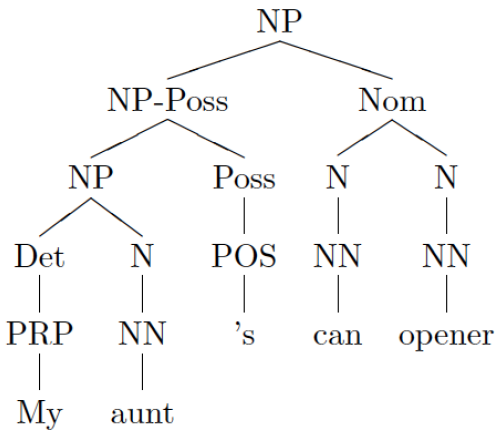
- Constituents of the same type can be coordinated
  - Kim and Sandy kissed each other
  - The old men and women came to dinner
  - The old man and his young nephew came to dinner
  - Kim and Sandy divorced and remarried each other
  - Kim kissed Sandy and remarried her
  - That rather old and very unreliable car belongs to Kim
  - Kim washed up and Sandy watched the TV

## Problems for Coordination Test

- Kim is a conservative and proud of it
- Kim became a conservative and arrogant
- Kim enjoys chess and watching football
- Kim gave Sandy a pen and Fido a bone
- “To hell with them and be dammed” , he said.

## Assignment 2

- Perform a phrase structure analysis of all noun phrases in your chosen sentences (the same ones from assignment 1)
- First bracket all NPs
- Recursively embedded
- Draw a Phrase Structure tree for each NP
- Reuse your Tokenisation and POS analysis from assignment 1
- Submit by Monday 28 October



# Noun compounds





## Reading for next time

- Read 5.7 and 5.8
- We will discuss exercises in 5.9 (1-3), so if you haven't done them, another chance.
- Keep working on logic worksheet