Structural Linguistics 2: the *function*
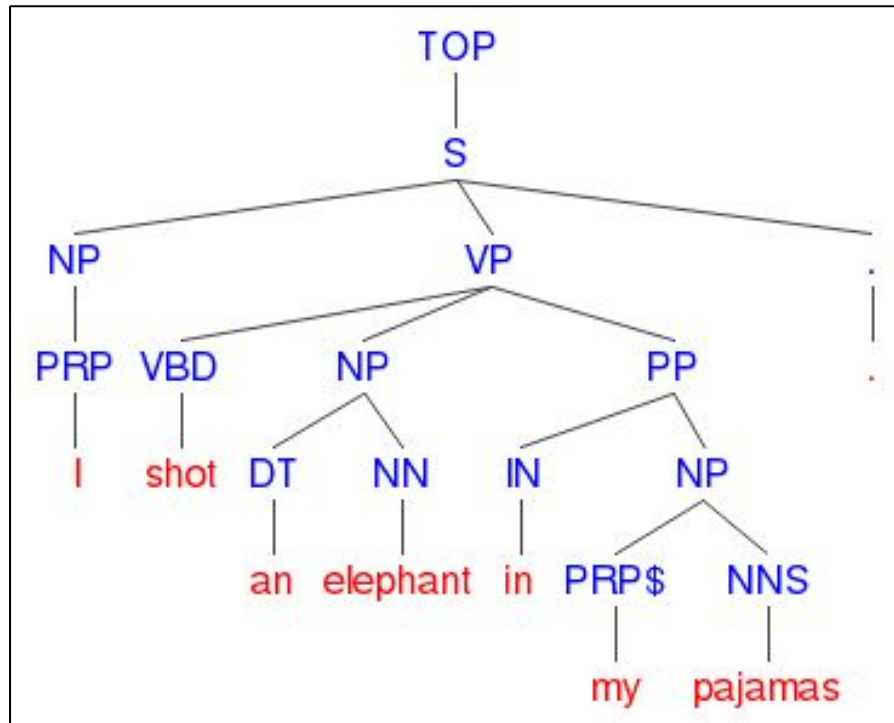
Mariana Romanyshyn,
Computational Linguist at Grammarly

Contents

A word is its…

1. form
2. *function*
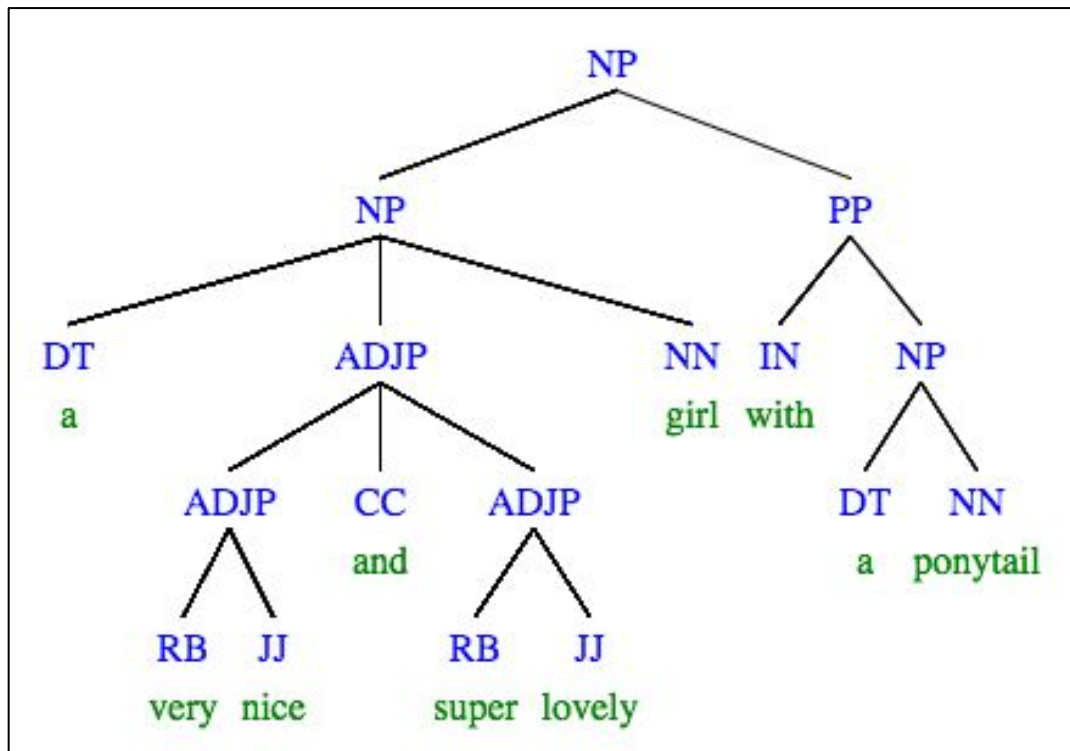3. meaning

Intro

Function

A word's *function* is defined by:

- lexical and grammatical properties
  - part of speech (POS)
  - gender, number, animacy, etc.
- role in the sentence
  - what the word modifies
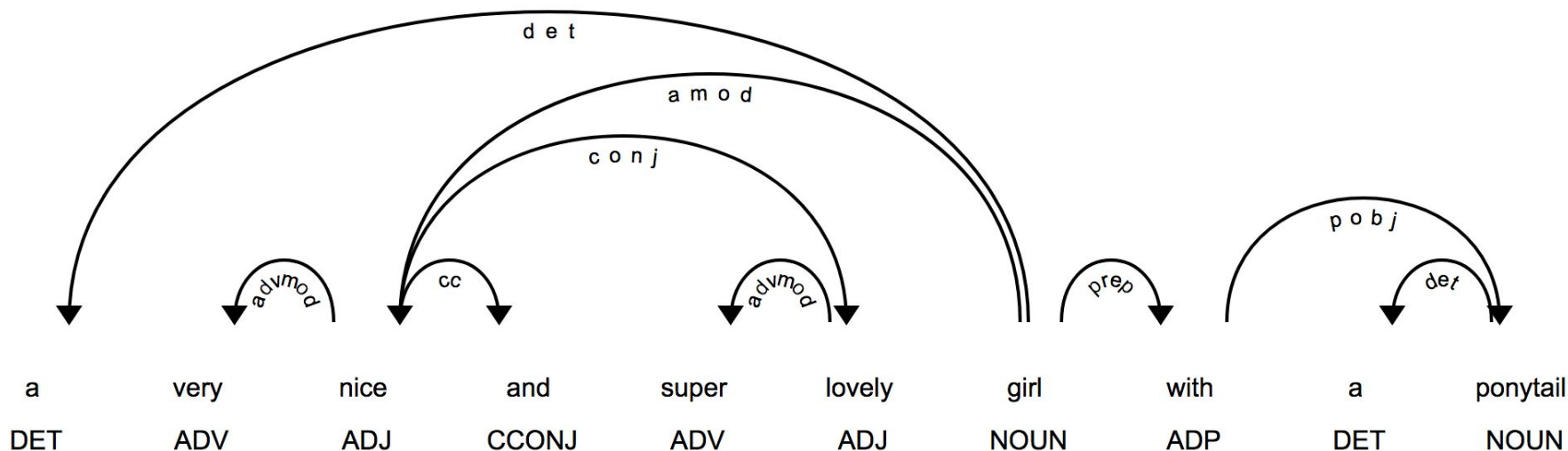 - what modifies the word

Methods of syntactic analysis

- constituency tree
  - phrase-based
  - 1 or more children in each node
  - type of phrase depends on the main element

Methods of syntactic analysis

- dependency tree
  - arc-based; arc label depends on the child and parent
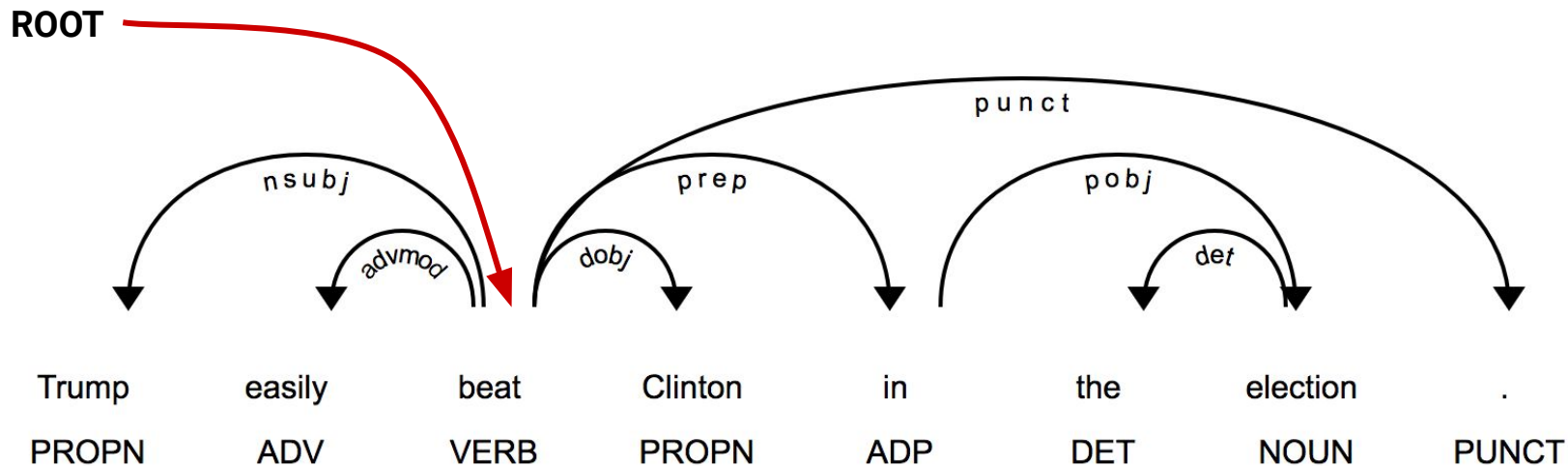 - every child has exactly one parent

How do I know the correct parse?

Use linguistic analysis!

Linguistic Analysis

- the tree ends with a ***root*** (or a ***TOP***, etc.)
- the ***verb*** is always the main element (when present)

Linguistic Analysis

- use substitution
  - *Her advice seems strange, **yet** I believe she's right.*

Linguistic Analysis

- use substitution
  - *Her advice seems strange, **{yet=>but}** I believe she's right.*
- ask questions
  - *The **guy** that I **met** yesterday was very funny.*

Linguistic Analysis

- use substitution
  - *Her advice seems strange, {yet=>but} I believe she's right.*
- ask questions
  - *The **guy (which guy?)** that I **met** yesterday was very funny.*
- remove elements
  - *Mary was hiding in the room **behind the shelves**.*
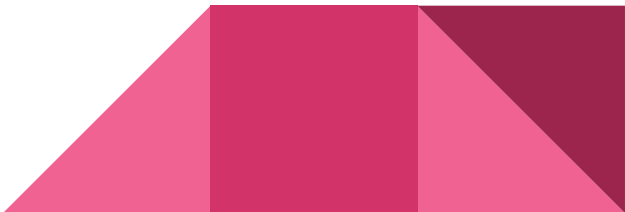 - *Kids were running with water **in their hats**.*

Linguistic Analysis

- use substitution
  - *Her advice seems strange, {yet=>but} I believe she's right.*
- ask questions
  - *The **guy (which guy?)** that I **met** yesterday was very funny.*
- remove elements
  - *Mary was hiding ~~in the room~~ **behind the shelves**.*
  - *Kids were running with water **in their hats**.*
- change the word order
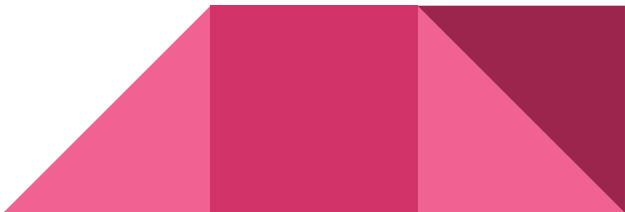 - *She left the room **singing happily**.*

Linguistic Analysis

- apply transformations
  - *Іван іде з другом => Друг іде з Іваном*
  - *Іван іде з палкою => * Палка йде з Іваном*

  - *учитель школи => шкільний учитель*
  - *прибуття потяга => * потяжне прибуття*

  - *He told me about the meeting tomorrow. => tomorrow's meeting*
  - *He told me about my mother tomorrow. => * tomorrow's mother*

Notation

- Language-specific:
  - Penn Treebank POS tags and phrase labels
  - Original Stanford dependencies
  - CLEAR NLP dependencies
  - Languagetool POS vs. pymorphy2 POS

- Language-independent:
  - Universal POS tags
  - Universal Dependencies
 - Universal Stanford dependencies

Notation in spaCy

- Language-specific:
  - [Penn Treebank POS tags](#) and [phrase labels](#) **<<==**
  - [Original Stanford dependencies](#)
  - [CLEAR NLP dependencies](#) **<<==**
  - [Languagetool POS](#) vs. [pymorphy2 POS](#)

- Language-independent:
  - [Universal POS tags](#) **<<==**
  - [Universal Dependencies](#)
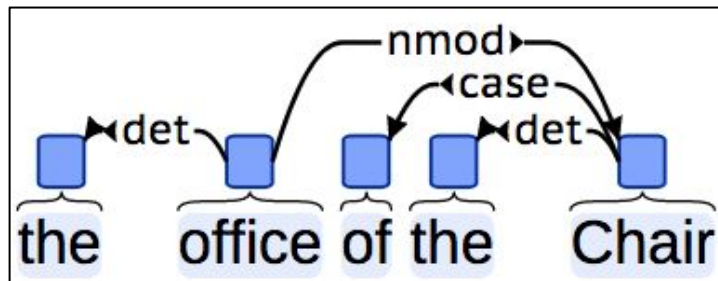 - [Universal Stanford dependencies](#)

Notation

Compare:

- Universal POS
  - *"cats"*: NOUN: {Animacy: Anim, Number: Plur...}
  - *"котики"*: NOUN: {Gender: Masc, Animacy: Anim...}
- Penn
  - *"cats"*: NNS
- Languagetool
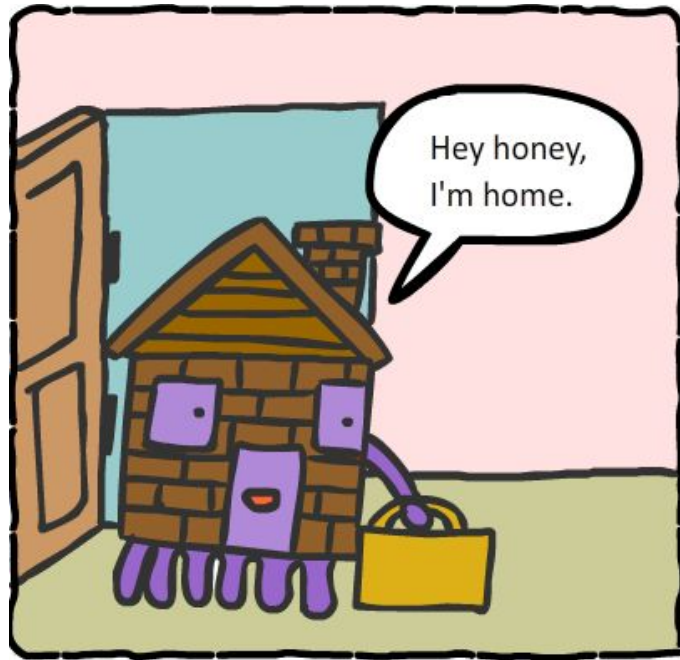 - *"котики"*: noun:anim:p:v_naz

Notation

Compare:

● Universal dependencies

● Original Stanford / CLEAR
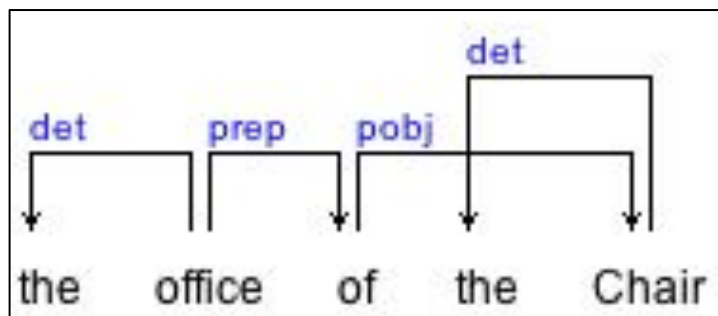
Let's talk parts of speech!

1. Adverb

Adverb Tags & Labels

- Tags
  - Penn: RB, RBR, RBS
  - Universal: ADV

- Phrase-level constituent
  - ADVP - adverbial phrase label

Lexical and Grammatical Properties

- degree
  - *мало, менше, найменше, щонайменше*
  - *slowly, more slowly, the most slowly*

- category
  - time, place, manner, degree, frequency, focusing, evaluative, linking
  - *today, upstairs, loudly, quite, rarely, mainly, surprisingly, however*

Adverb Modifiers

- adverb
  - ADVMOD, NEG *(quite often, not really)*
- determiner
  - DET *(no sooner)*
- noun phrase
  - NPADVMOD *(three years later)*
- prepositional phrase
  - PREP *(separately from others)*

2. Adjective

Adjective Tags & Labels

- Tags
  - Penn: JJ, JJR, JJS
  - Universal: ADJ

- Phrase-level constituent
  - ADJP - adjectival phrase label

Lexical and Grammatical Properties

- degree
  - *малий, менший, найменший, щонайменший, якнайменший*
  - *fast, faster, the fastest*
  - gradable vs. non-gradable

- qualitative/relative
  - *tall, dark* vs. *Swedish, wooden*

Lexical and Grammatical Properties

- gender, number, declension
  - *малий, мала, малому, малими*

- order
  - *beautiful big old green Italian couch*
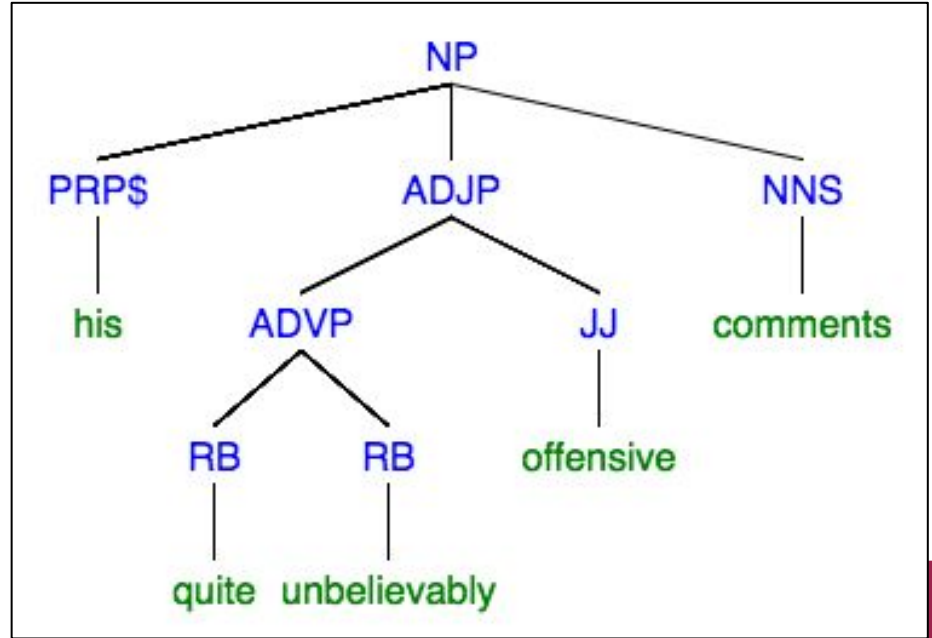 - *yellow wild roses != wild yellow roses*

Adjective Modifiers

- adverb
  - ADVMOD *(very nice)*
- noun phrase
  - NPADVMOD *(3 years old)*
- prepositional phrase
  - PREP *(angry with you)*
- clausal complement
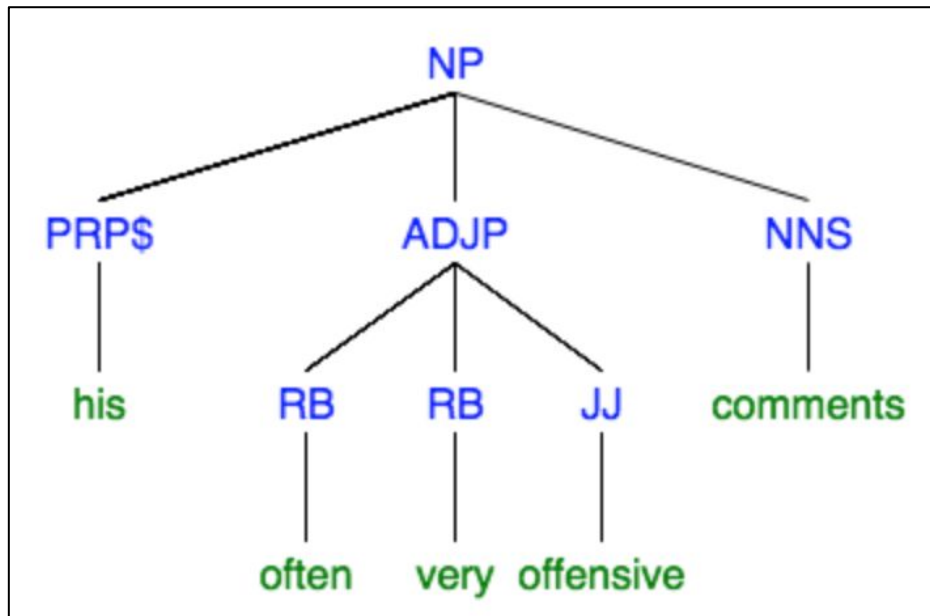 - CCOMP, XCOMP *(happy to help, sure that we'll succeed)*

Types of Modification

- submodification

- stacked modification

Types of Modification

- submodification

- stacked modification

3. Noun

Noun Tags & Labels

- Tags
  - Penn: NN, NNS, NNP, NNPS
  - Universal: NOUN, PROPN

- Phrase-level constituent
  - NP - noun phrase label

Lexical and Grammatical Properties

- number
  - *foxes* vs. *fox*, *квітка* vs. *квіти*
  - Singularia Tantum vs. Pluralia Tantum
  - count vs. mass ambiguity

- case/declension
  - *светрик* vs. *светрику, светриком*
  - *woman* vs. *woman's*

Lexical and Grammatical Properties

- gender
  - *дуб* vs. *die Eiche*
  - *каліка, вітрище, агакало*
  - *This is my car. She is a beauty!*

Lexical and Grammatical Properties

- gender
  - *дуб* vs. *die Eiche*
  - *каліка, вітрище, агакало*
  - *This is my car. She is a beauty!*

Beating gender bias
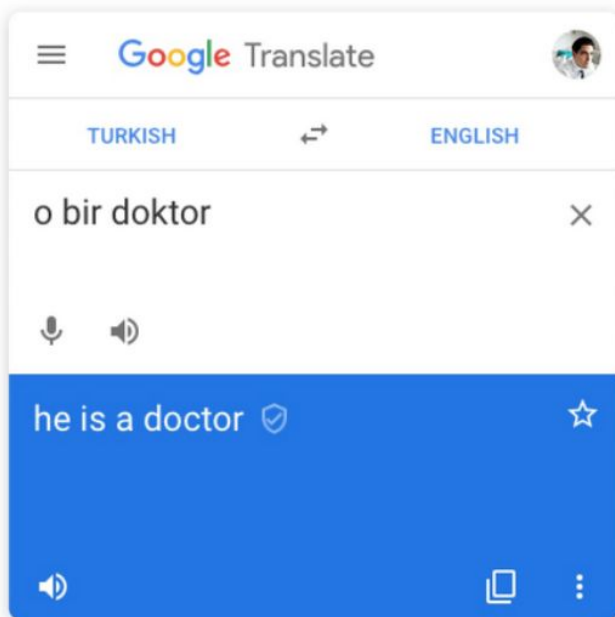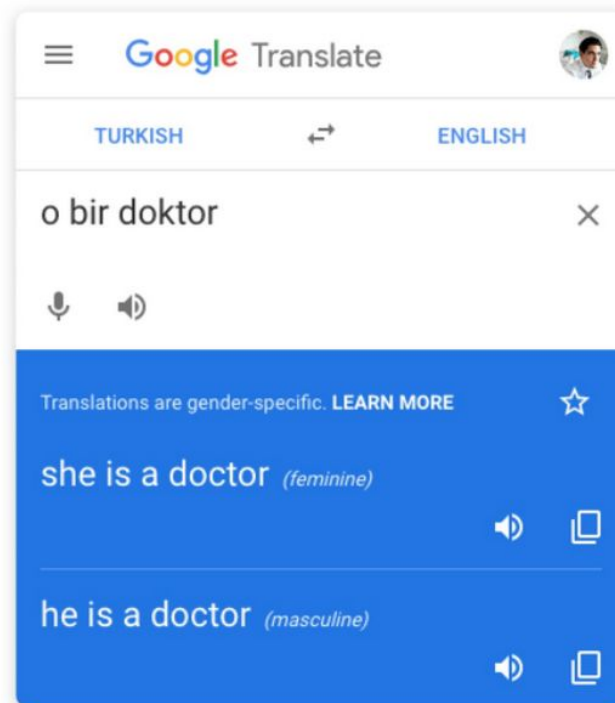
Beating gender bias

Before

After

Lexical and Grammatical Properties

- common/proper
  - *Apple* vs. *apple, Роман* vs. *роман*

- animacy
  - *car* vs. *cat*

- concrete/abstract
  - *car* vs. *love*

Noun Modifiers

- predeterminer/determiner/possessive modifier
  - PREDET, DET, NEG, POSS *(all those/my/people's dreams)*
- numeral
  - NUMMOD *(million samples)*
- case marker
  - CASE *(Alexa 's)*
- adjective
  - AMOD *(artificial intelligence)*

Noun Modifiers

- noun adjunct
  - COMPOUND   *(data science, language model)*
- appositive
  - APPOS   *(Seva, my co-teacher,)*
- prepositional phrase
  - PREP   *(part of speech, question about homework)*
- subordinate clause
  - ACL, RELCL   *(tips to follow, students who come on time)*

Practice

Analyze the following phrases:

- *Natural Language Processing*

- *Peter Norvig, Google's director of research*

- *Use a less greasy lotion sunscreen that blocks sun rays.*

4. Verb

Verb Tags & Labels

- Tags
  - Penn: VB, VBP, VBZ, VBG, VBD, VBN
  - Universal: VERB

- Phrase-level constituents
  - VP - verb phrase label
  - S, SQ, SINV - clause
  - SBAR, SBARQ - subordinate clause

Lexical and Grammatical Properties

- Person, number, and gender
  - *читаю, читаєш, читає, читаємо, читають*
  - *become, becomes*

- Mood
  - indicative, imperative, subjunctive, interrogative, conditional

Lexical and Grammatical Properties

- Aspect: perfect vs. imperfect
- Voice: active vs. passive
- Tense: past vs. present vs. future
- Notional/auxiliary
- Transitivity
  - transitive vs. intransitive vs. reflexive
  - monotransitive, ditransitive

Verb Modifiers

- subject
  - NSUBJ, NSUBJPASS, CSUBJ, CSUBJPASS
- objects
  - DOBJ, DATIVE *(make me a sandwich)*
- clausal complement
  - CCOMP, XCOMP *(decided that I'll come, decided to come)*
- adverbial clause
  - ADVCL *(I'll come if you come too.)*

Verb Modifiers

Catenative structures with XCOMP:

От було взяло заманулося піти спробувати навчитися готувати їсти. 8 дієслів підряд. #ГраничнаМова

Verb Modifiers

- prepositional phrase
  - PREP, AGENT *(agree with you, agreed by you)*
- adverb, negation
  - ADVMOD, NEG *(move fast, not move)*
- object predicate
  - OPRD *(make him king)*
- noun phrase as an adverbial
  - NPADVMOD *(this week I'll learn about linguistics)*

Verb Modifiers

- auxiliary verb
  - AUX, AUXPASS   *(don't move, to move, is moved)*
- subordinating conjunction
  - MARK   *(I'll come if you come too.)*
- predicative complement
  - ACOMP   *(is interesting)*
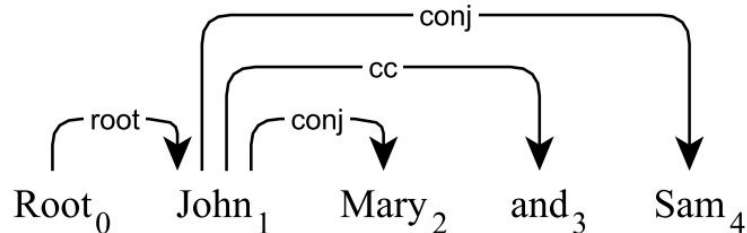- particle
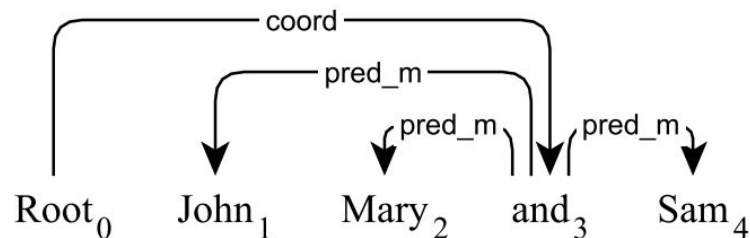 - PRT   *(come on)*

5. Coordination

Coordination

- constituents
  - the label stays the same: (VP VP CC VP)
  - … or not: (UCP PP CC SBAR)

- dependencies
  - PRECONJ *(both X and Y)*
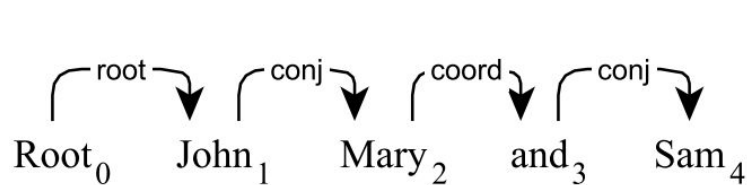 - CC *(both X and Y)*
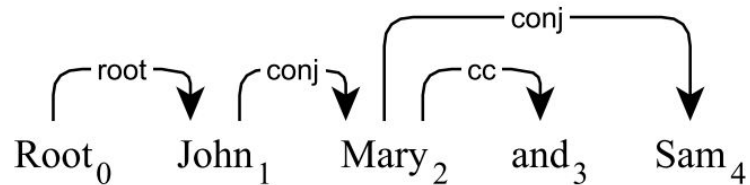 - CONJ *(both X and Y)*

Coordination

Practice

Analyze the following sentences:

- *Now, if you want to receive e-mails about my upcoming shows, then please give me money so I can buy a computer.*

- *Так її мій приятель у розмові й називав, а коли я запитав про її ім'я та прізвище, то Микола зирнув на мене нетямкувато — здається, він забув тим поцікавитися.*

- *All the food tasted excellent, and with the new renovation of chairs and the bathroom, it is awesome.*

6. Functional parts of speech

Preposition

- Tags
  - Penn: IN
  - Universal: ADP
- Phrase-level constituent
  - PP - prepositional phrase
- Modifiers
  - adverb: ADVMOD   *(back to, long before)*
  - object: POBJ, PCOMP   *(with you, about what to do next)*

Conjunction

- Tags
  - Penn: IN, CC
  - Universal: SCONJ, CCONJ *(sadly, is rarely used)*
- Labels
  - CONJP - coordinate conjunction phrase
- Modifiers
  - adverb: ADVMOD   *(just because I said so)*
  - noun phrase: NPADVMOD   *(one day before we arrived)*

Determiner

- Tags
  - Penn: DT
  - Universal: DET
- Modifiers
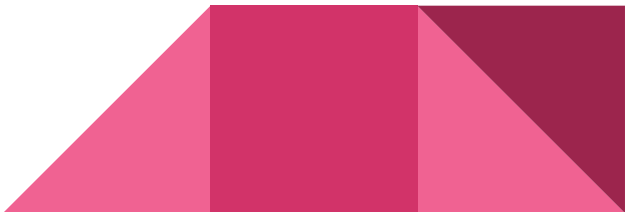 - adverb: ADVMOD *(almost all)*

Numeral

- Tags
  - Penn: CD
  - Universal: NUM
- Labels
  - QP - quantifier phrase
- Modifiers
  - adverb: QUANTMOD *(over 5,000)*

Additional notation (Penn)

- PRP, PRP$ - pronoun, possessive pronoun
- WP, WP$, WDT, WRB - "who", "whose", "which", "when"
- UH - interjection
- RP, TO - particle
- EX - existential "there"
- FW - foreign word
- SYM - non-standard symbol
- LS - list marker
- . , : " `` - punctuation
- $ - currency

7. How to use

Constituency Trees

```
(TOP (S (SBAR (IN "If")
              (S (NP (PRP "you"))
                 (VP (VBP "want")
                     (S (VP (TO "to")
                            (VP (VB "receive")
                                (NP (NP (NNS "e-mails"))
                                    (PP (IN "about")
                                        (NP (PRP$ "my") (JJ "upcoming") (NNS "shows")))))))))))
        (, ",")
        (ADVP (RB "then"))
        (INTJ (UH "please"))
        (VP (VB "give")
            (NP (PRP "me"))
            (NP (NN "money"))
            (SBAR (IN "so")
                  (S (NP (PRP "I"))
                     (VP (MD "can")
                         (VP (VB "buy")
                             (NP (DT "a") (NN "computer")))))))
        (. ".")))
```

Dependency Trees

1	If	if	IN	3	mark
2	you	you	PRP	3	nsubj
3	want	want	VBP	14	advcl
4	to	to	TO	5	aux
5	receive	receive	VB	3	xcomp
6	e-mails	e-mail	NNS	5	dobj
7	about	about	IN	6	prep
8	my	my	PRP$	10	poss
9	upcoming	upcoming	JJ	10	amod
10	shows	show	NNS	7	pobj
11	,	,	,	14	punct
12	then	then	RB	14	advmod
13	please	please	UH	14	intj
14	give	give	VB	0	root
15	me	me	PRP	14	dative

...

Dependency Trees
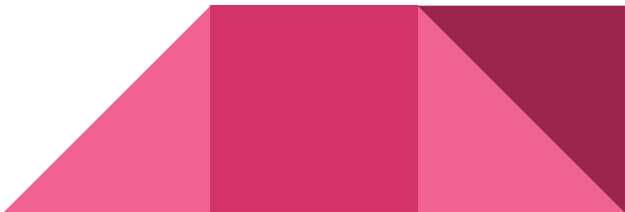
```
mark(want-3, If-1)
nsubj(want-3, you-2)
advcl(give-14, want-3)
mark(receive-5, to-4)
xcomp(want-3, receive-5)
dobj(receive-5, e-mails-6)
case(shows-10, about-7)
nmod:poss(shows-10, my-8)
amod(shows-10, upcoming-9)
nmod(e-mails-6, shows-10)
```

```
nsubj(give-14, then-12)
discourse(give-14, please-13)
root(ROOT-0, give-14)
iobj(give-14, me-15)
dobj(give-14, money-16)
dep(give-14, so-17)
nsubj(buy-20, I-18)
aux(buy-20, can-19)
parataxis(give-14, buy-20)
det(computer-22, a-21)
dobj(buy-20, computer-22)
```

POS taggers and Parsers

- [Spacy](#) (7 languages; Python)
- [Stanford CoreNLP](#) (6 languages; Java, available for [other languages](#) too)
- [StanfordNLP](#) (53 languages; Python)
- [OpenNLP](#) (7 languages; Java)
- [Emory NLP](#) (English; Java)
- only POS tagging: [nltk](#) (English; Python) or [TextBlob](#) (English; Python)

Only parts of speech and no disambiguation:

- [pymorphy2](#) (Russian, Ukrainian; Python)
- [nlp_uk](#) (Ukrainian; Groovy)

POS tagging in spaCy

```
In [28]:  import en_core_web_md

          nlp = en_core_web_md.load()
```

```
In [31]:  sentence = nlp("I like turtles because they are cute.")
          " ".join(token.text + "_" + token.tag_ for token in sentence)
```

Out[31]: 'I_PRP like_VBP turtles_NNS because_IN they_PRP are_VBP cute_JJ ._.'

```
In [32]:  " ".join(token.text + "_" + token.pos_ for token in sentence)
```

Out[32]: 'I_PRON like_VERB turtles_NOUN because_ADP they_PRON are_VERB cute_ADJ ._PUNCT'
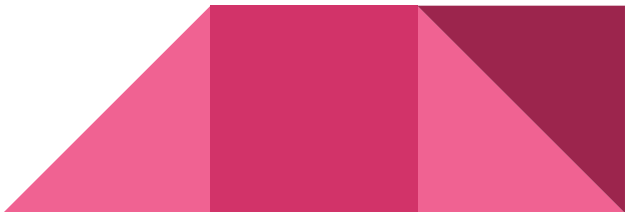
Dependency parsing in spaCy

```
In [34]: for token in sentence:
             print("{:10}{:10}{:10}".format(
                 token.text, token.dep_, token.head.text))
```

```
I          nsubj      like
like       ROOT       like
turtles    dobj       like
because    mark       are
they       nsubj      are
are        advcl      like
cute       acomp      are
.          punct      like
```

POS tagging in pymorphy

```
In [1]:   import tokenize_uk
          import pymorphy2
```

```
In [13]:  text = "Обговорімо основні етапи купівлі та розмитнення \"євробляхи\"."
          morph = pymorphy2.MorphAnalyzer(lang='uk')
```

```
In [14]:  tokens = tokenize_uk.tokenize_words(text)
          print(tokens)
```
```
          ['Обговорімо', 'основні', 'етапи', 'купівлі', 'та', 'розмитнення', '"', 'євроблях
          и', '"', '.']
```

```
In [15]:  print([(word, morph.parse(word)[0].tag.POS) for word in tokens])
```
```
          [('Обговорімо', 'VERB'), ('основні', 'ADJF'), ('етапи', 'NOUN'), ('купівлі',
          'NOUN'), ('та', 'CONJ'), ('розмитнення', 'NOUN'), ('"', None), ('євроблях
          и', 'NOUN'), ('"', None), ('.', None)]
```

POS tagging in pymorphy

```
In [25]:   morph.parse("того")[0].tag.POS
```

```
Out[25]:   'NOUN'
```

```
In [26]:   morph.parse("того")
```

```
Out[26]: [Parse(word='того', tag=OpencorporaTag('NOUN,inan femn,voct'), normal_form='тога', score=1.0, m
         ethods_stack=((DictionaryAnalyzer(), 'того', 175, 6),)),
          Parse(word='того', tag=OpencorporaTag('NOUN,Fixd,masc,Sgtm,inan nomn'), normal_form='того', sc
         ore=1.0, methods_stack=((DictionaryAnalyzer(), 'того', 165, 0),)),
          Parse(word='того', tag=OpencorporaTag('NOUN,Fixd,masc,Sgtm,inan gent'), normal_form='того', sc
         ore=1.0, methods_stack=((DictionaryAnalyzer(), 'того', 165, 1),)),
          Parse(word='того', tag=OpencorporaTag('NOUN,Fixd,masc,Sgtm,inan datv'), normal_form='того', sc
         ore=1.0, methods_stack=((Dictionar
          Parse(word='того', tag=Opencorpora
         ore=1.0, methods_stack=((Dictionar
          Parse(word='того', tag=Opencorpora
         ore=1.0, methods_stack=((Dictionar
          Parse(word='того', tag=Opencorpora
         ore=1.0, methods_stack=((Dictionar
          Parse(word='того', tag=Opencorpora
          Parse(word='того'. tag=Opencorpora
```
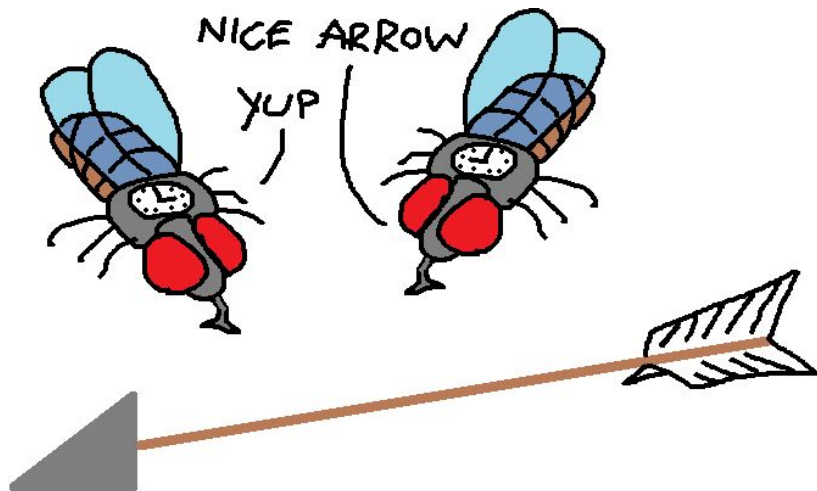
¯_(ツ)_/¯

Conclusion

Ambiguities: English

- 400K unique word forms
- 30K words can have
  >1 possible POS
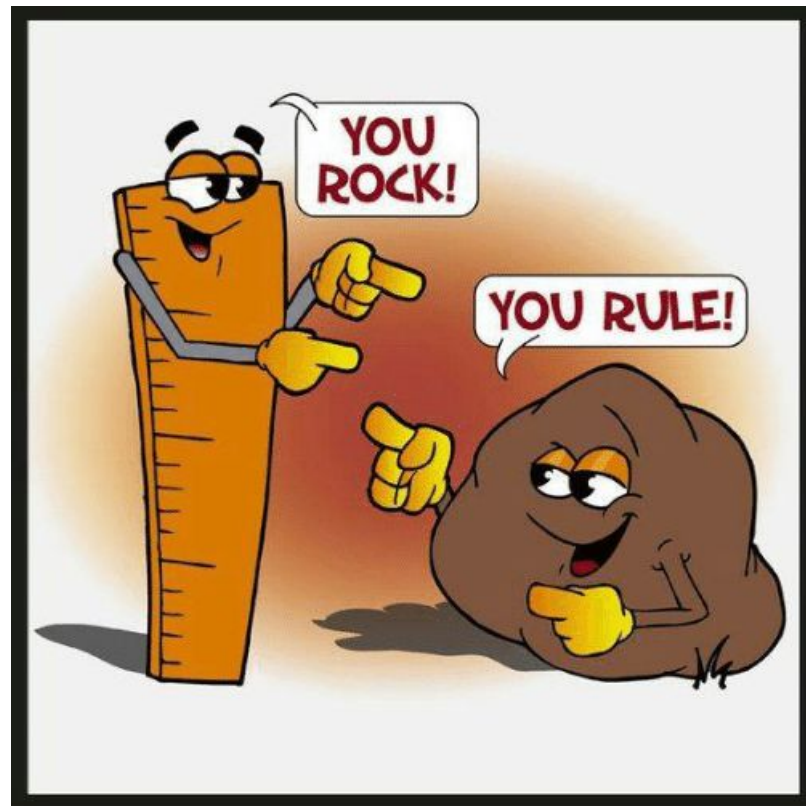
Ambiguities: English

- *Time flies like an arrow.*
- *I saw her duck with a telescope.*
- *She is calculating.*
- *We watched an Indian dance.*
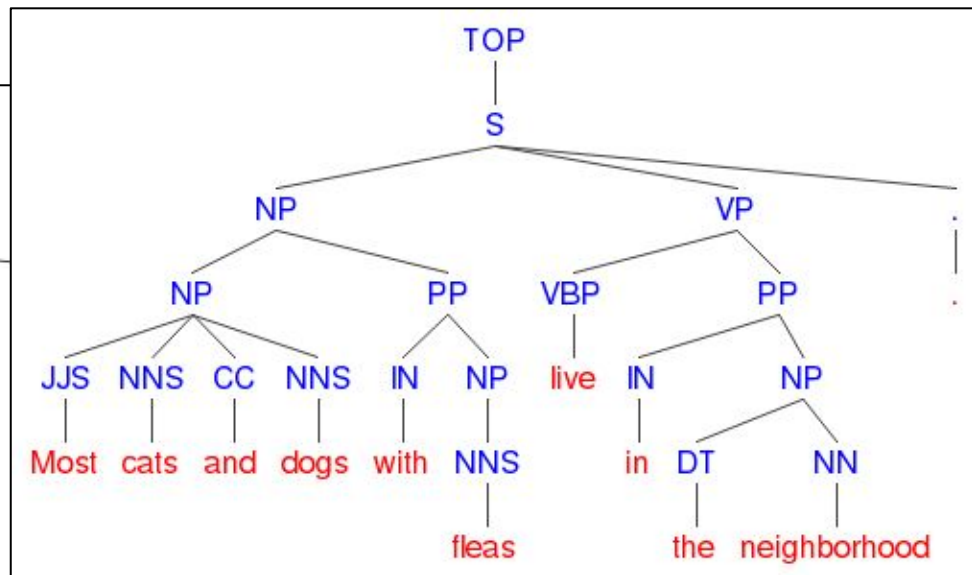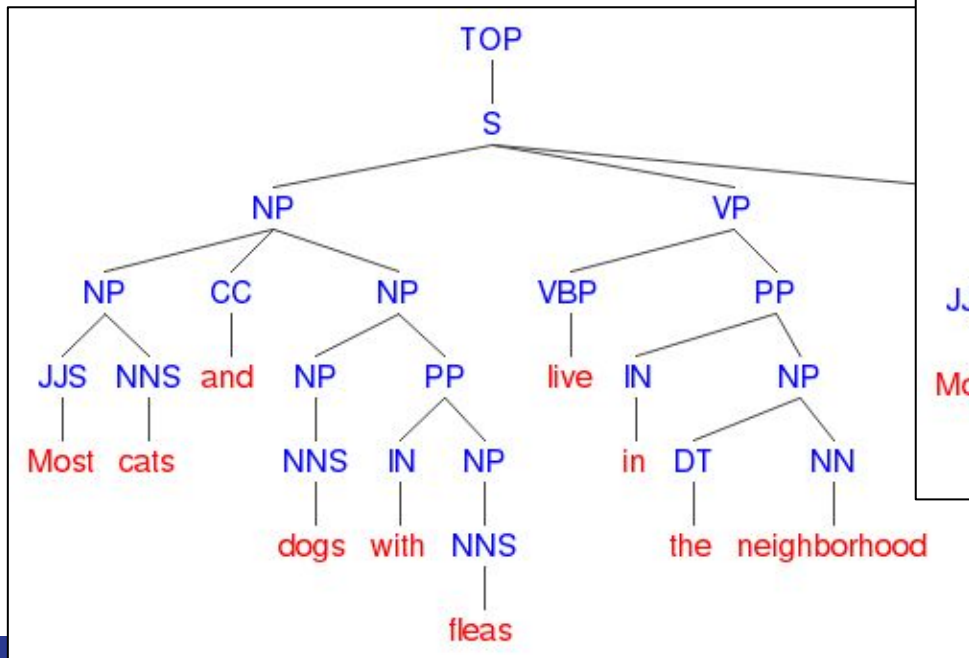- *They can fish.*
- *More lies ahead...*

Ambiguities: English

Most cats and dogs with fleas live in the neighborhood.

Ambiguities: English
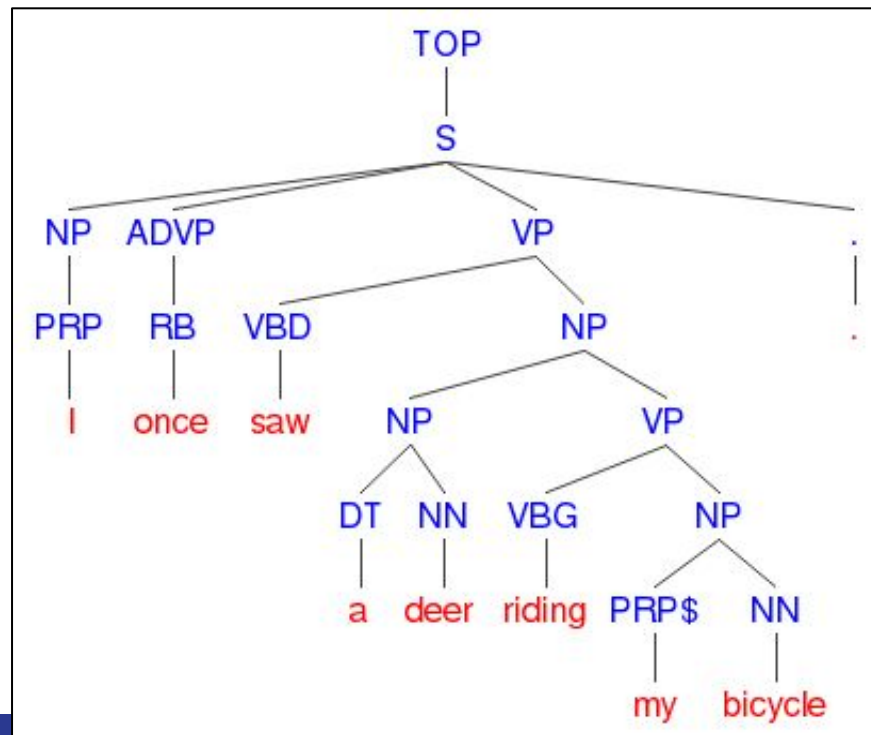
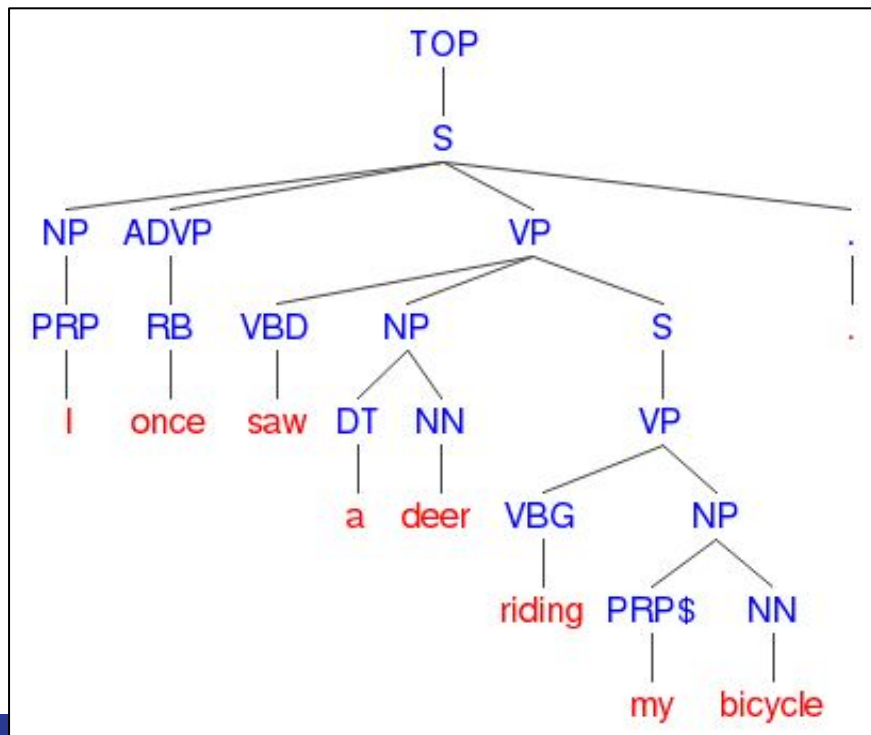*Most cats and dogs with fleas live in the neighborhood.*

Ambiguities: English

I once saw a deer riding my bicycle.

Ambiguities: English
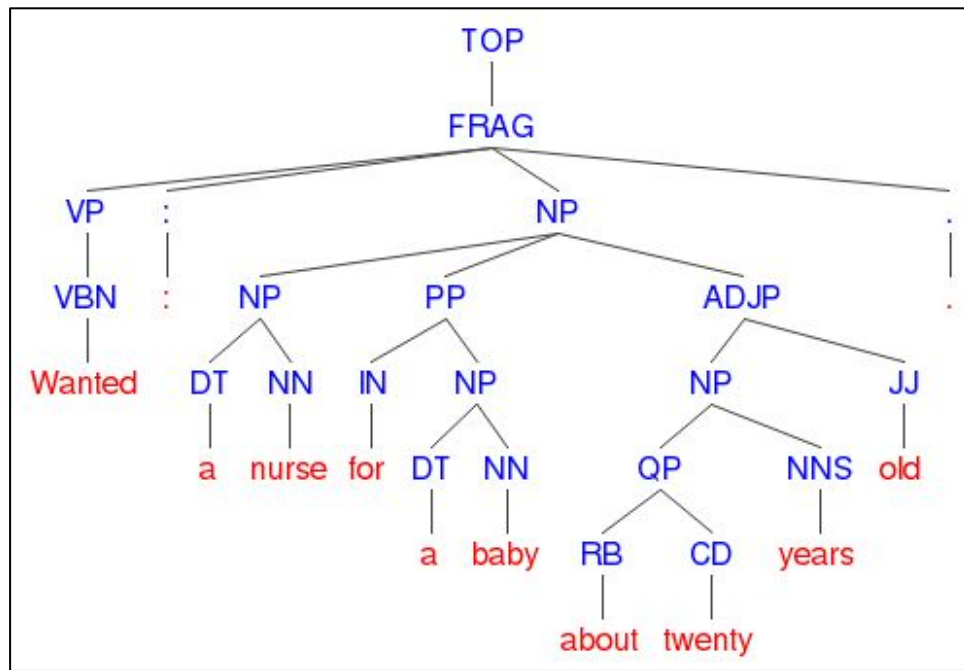
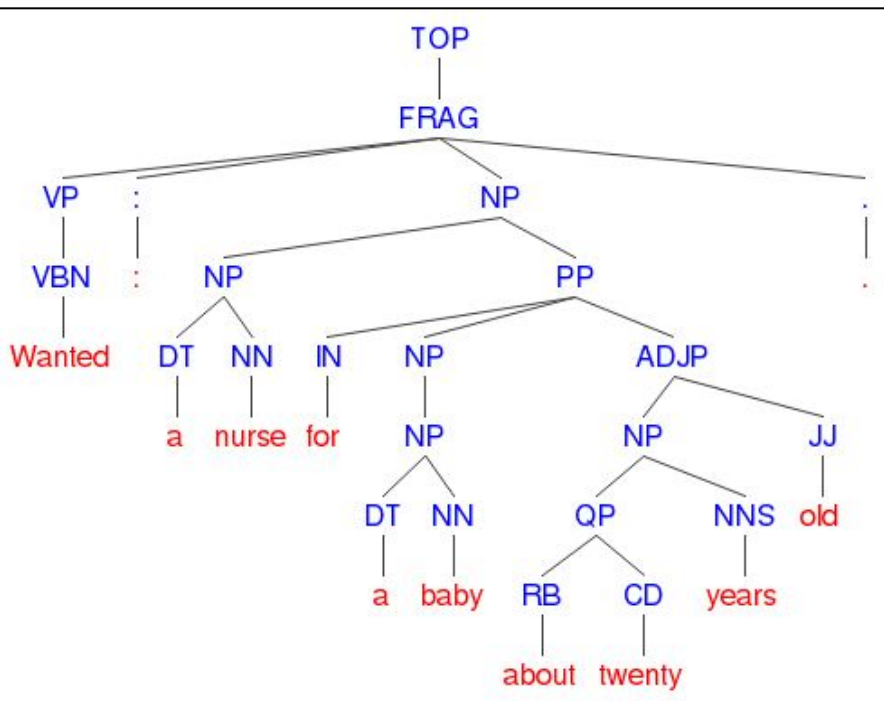*I once saw a deer riding my bicycle.*

Ambiguities: English

Wanted: a nurse for a baby about twenty years old.

Ambiguities: English
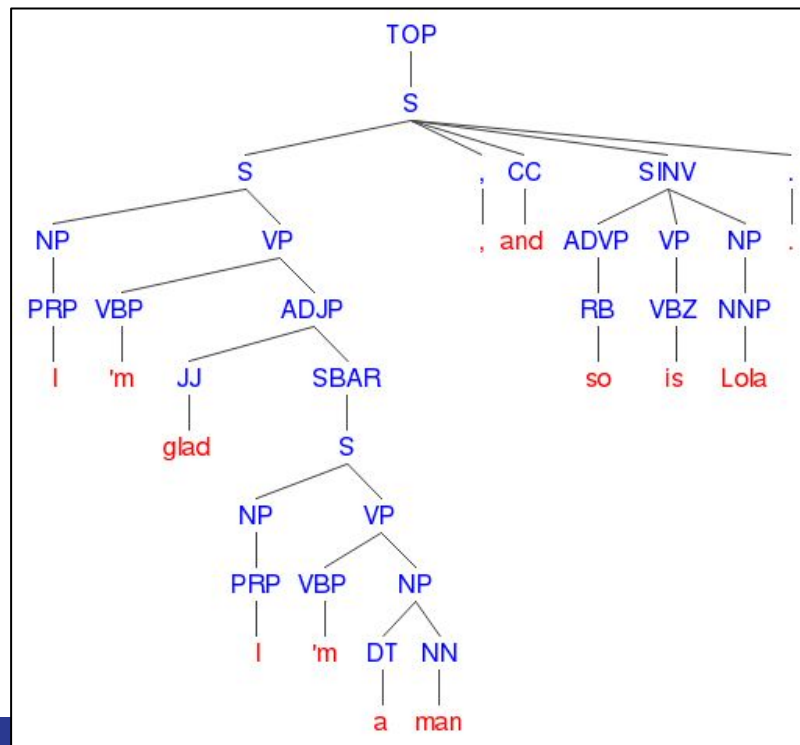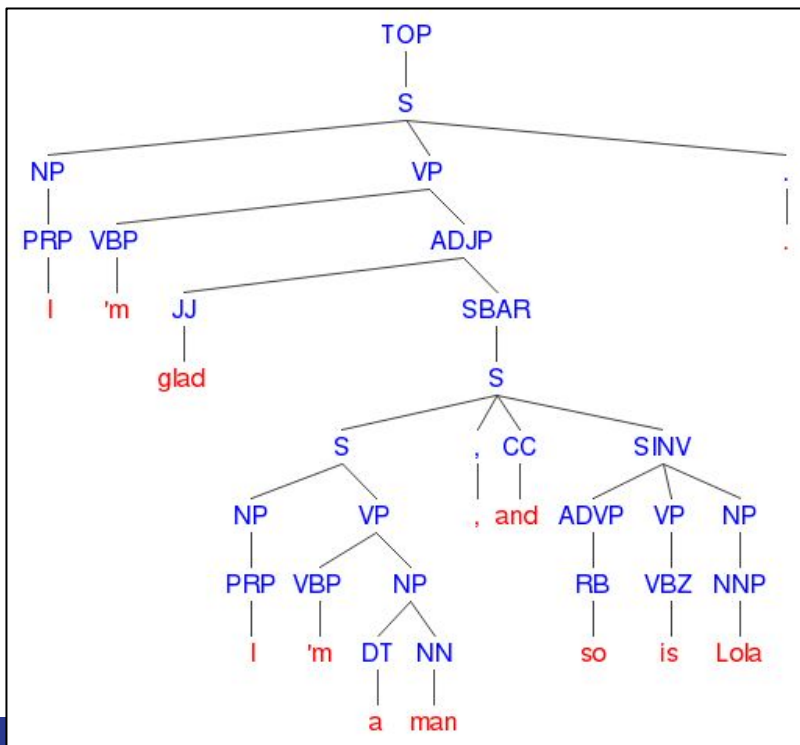
Wanted: a nurse for a baby about twenty years old.

Ambiguities: English

I'm glad I'm a man, and so is Lola.

Ambiguities: English

I'm glad I'm a man, and so is Lola.

Ambiguities: English

- *We decided **immediately** to buy this house.*
- *You can **only** access the web at this workstation.*
- *In Kyiv **alone** there are 3 mln people.*

Ambiguities: Ukrainian

- 2.1 mln unique word forms
- 10K word forms have > 1 possible POS

What is the most ambiguous word?

Ambiguities: Ukrainian

- 2.1 mln unique word forms
- 10K word forms have > 1 possible POS

What is the most ambiguous word?

```
коли ['adv', 'conj', 'part', 'noun', 'verb']
прецінь ['adv', 'insert', 'conj', 'part']
тому ['adv', 'noun', 'conj', 'adj']
усе ['adv', 'conj', 'part', 'adj']
як ['adv', 'conj', 'part', 'noun']
ага ['excl', 'part', 'noun']
але ['conj', 'part', 'excl']
багатій ['noun', 'verb', 'adj']
вагітній ['adj', 'verb', 'noun']
варт ['adj', 'noun', 'predic']
власне ['insert', 'part', 'adj']
властиво ['insert', 'part', 'predic']
відколи ['adv', 'conj', 'verb']
гай ['excl', 'verb', 'noun']
гайну ['noun', 'verb', 'adj']
десь ['adv', 'insert', 'part']
доки ['adv', 'conj', 'noun']
доросла ['noun', 'verb', 'adj']
жила ['adj', 'verb', 'noun']
знайомим ['adj', 'verb', 'noun']
лютим ['noun', 'verb', 'adj']
милим ['noun', 'verb', 'adj']
мов ['conj', 'part', 'noun']
```
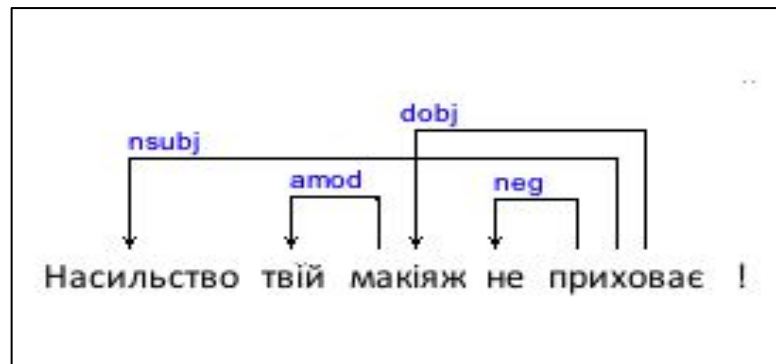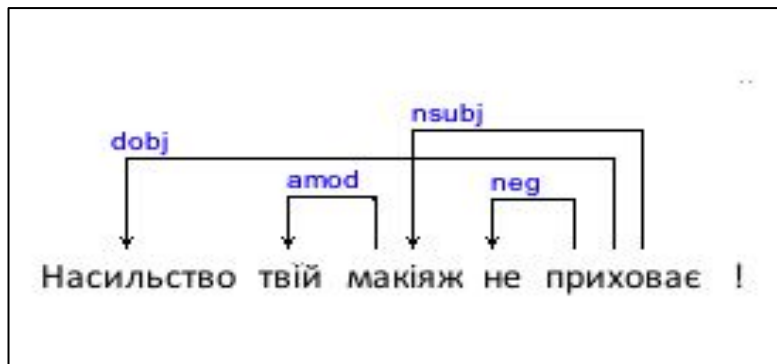
Ambiguities: Ukrainian

- *Це мало мало значення.*
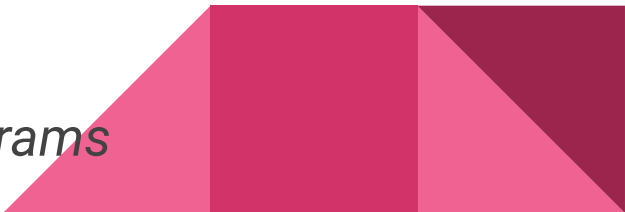- *Коло друзів та незнайомців.*

Ambiguities: Ukrainian

- *Це мало мало значення.*
- *Коло друзів та незнайомців.*
- *Насильство твій макіяж не приховає.*

Features

- Part of speech, part-of-speech tag
- Morphological properties:
  - *gender, animacy, number, person, case*
  - *aspect, voice, tense, degree of comparison*
- Constituents
  - *parents, children*
- Direct and indirect dependencies
  - *parents, children, type of relation*
- Depth of the syntactic tree
- Statistics: *POS+word, POS ngrams, syntactic ngrams*

Questions?