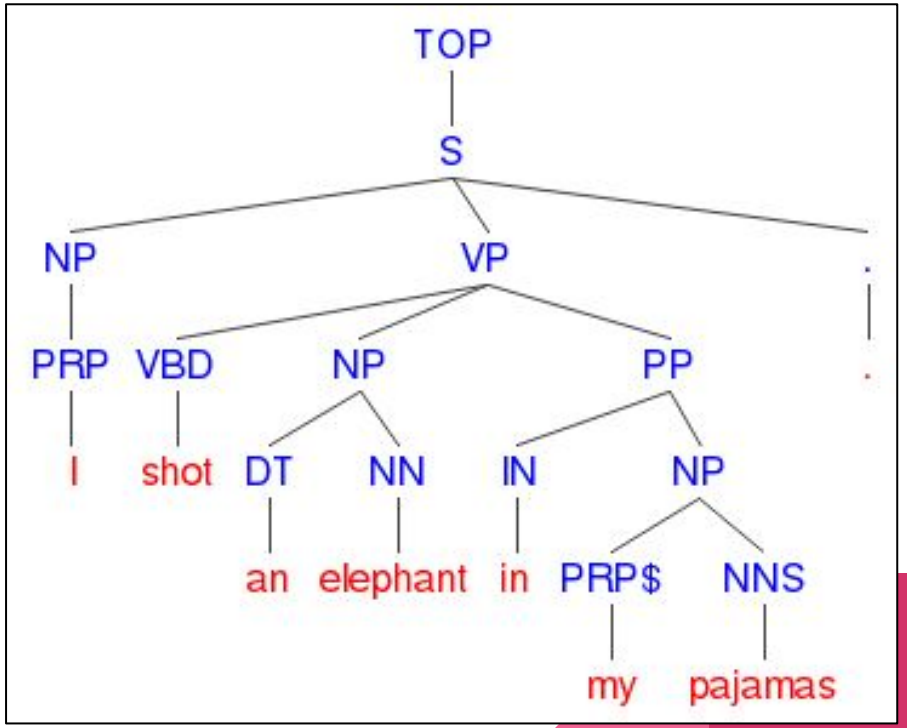
Structural Linguistics 2: the *function*

Mariana Romanyshyn, *Grammarly, Inc.*

Contents

A word is its...

1. form
2. ***function***
3. meaning

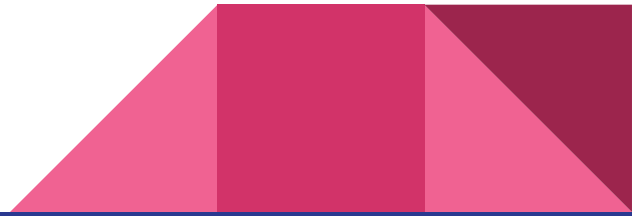


Intro

Function

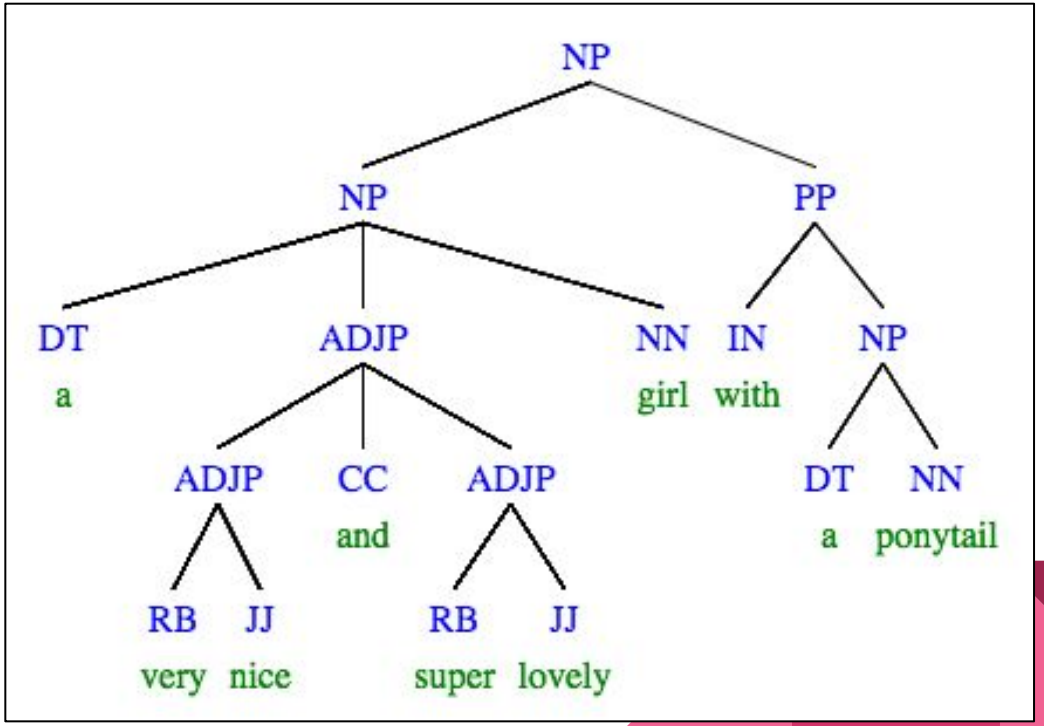
A word's ***function*** is defined by:

- 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

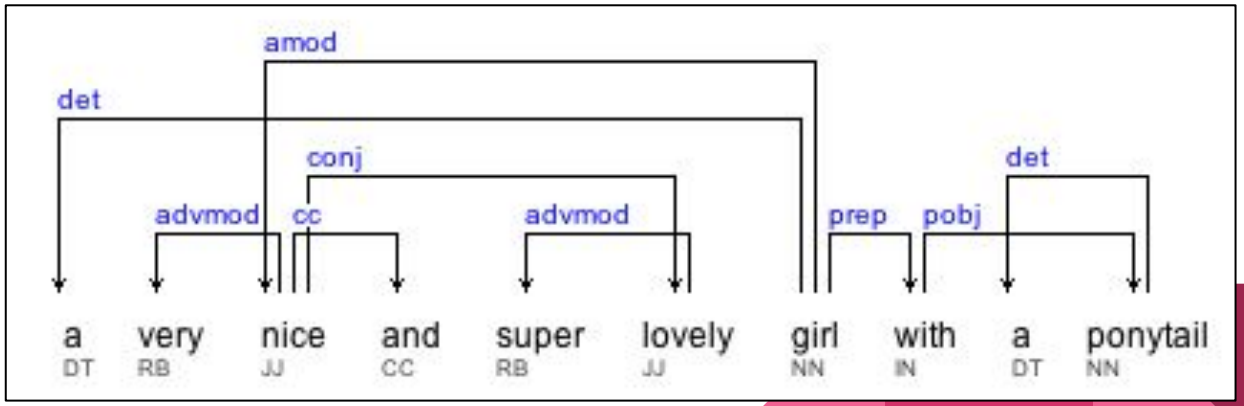
- constituents
- dependencies



Methods of syntactic analysis

- constituents

- dependencies

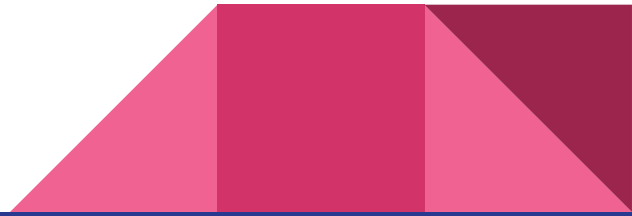


Linguistic Tips

- use substitution
  - *Her advice seems strange, **{yet=>but}** I believe she's right.*

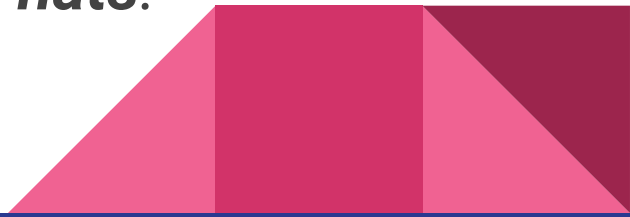
Linguistic Tips

- 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.*



Linguistic Tips

- 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 Tips

- 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 Tips

- 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 and Universal Stanford dependencies
  - Languagetool POS vs. pymorphy2 POS

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

Notation

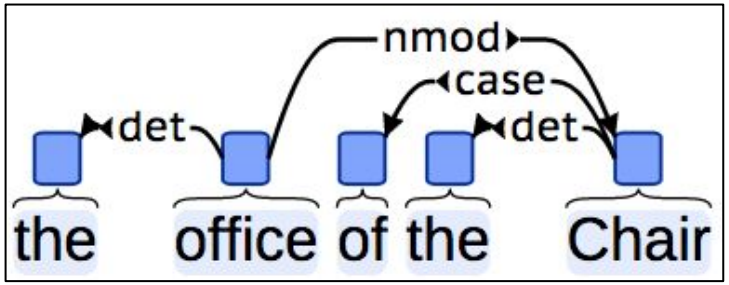
Compare:

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

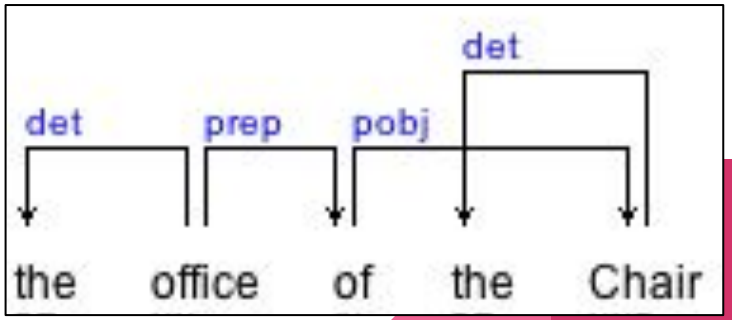
Notation

Compare:

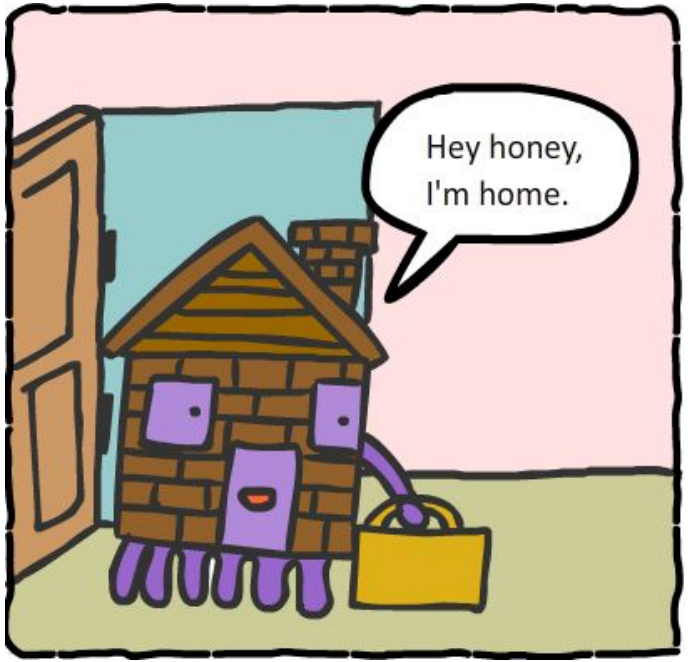
- Universal dependencies



- Original Stanford



Let’s talk parts of speech!



1. Adverb

Adverb Tags & Labels

- Tags
  - Penn: RB, RBR, RBS
  - Universal: ADV (Degree)

- Labels
  - ADVP - adverbial phrase label

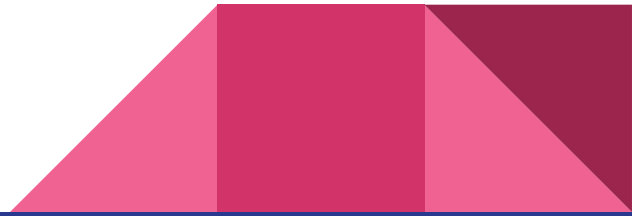
Lexical and Grammatical Properties

- 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
- determinative
  - DET
- noun phrase
  - NPADVMOD *(UD: npmod, tmod)*
- prepositional phrase
  - PREP *(UD: nmod)*

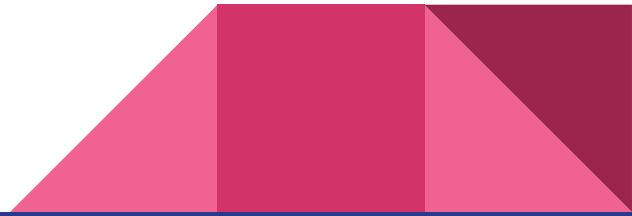


2. Adjective

Adjective Tags & Labels

- Tags
  - Penn: JJ, JJR, JJS
  - Universal: ADJ (Degree, Gender, Number, Case)

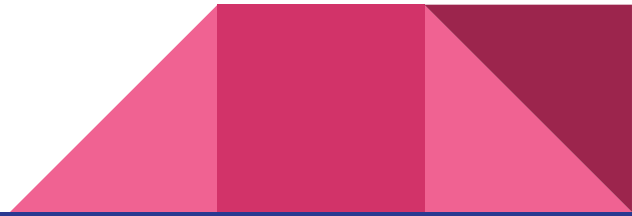
- Labels
  - ADJP - adjectival phrase label



Lexical and Grammatical Properties

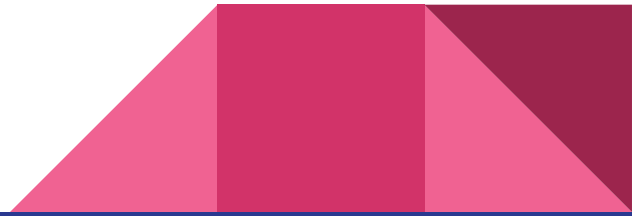
- degree
  - *малий, менший, найменший, щонайменший, якнайменший*
  - *fast, faster, the fastest*
  - gradable vs. non-gradable

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



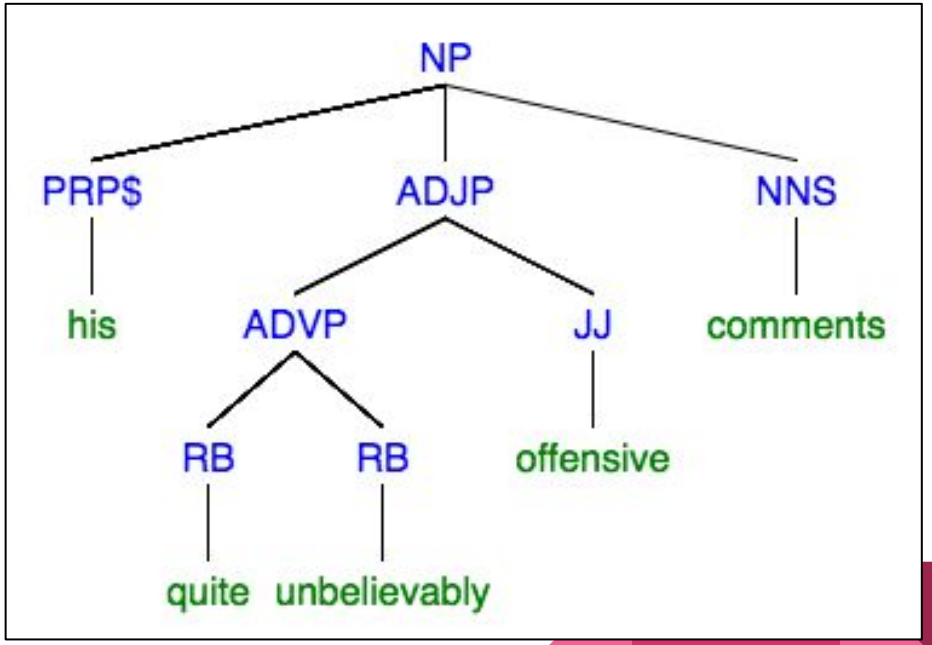
Adjective Modifiers

- adverb
  - ADVMOD
- noun phrase
  - NPADVMOD *(UD: npmod, tmod)*
- prepositional phrase
  - PREP *(UD: nmod)*
- clausal complement
  - CCOMP, XCOMP



Types of Modification

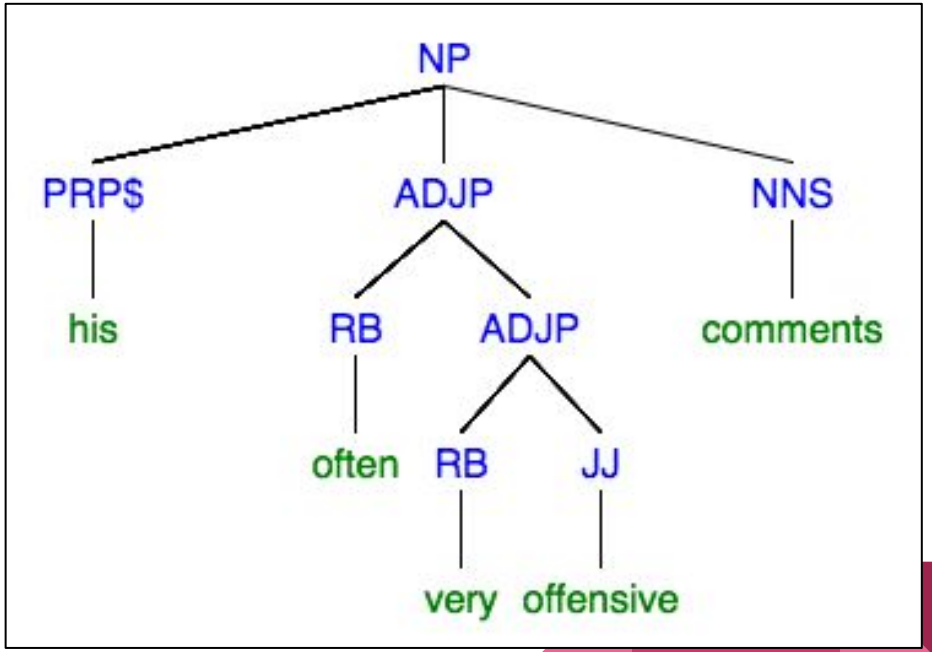
- submodification
- stacked modification



Types of Modification

- submodification

- stacked modification

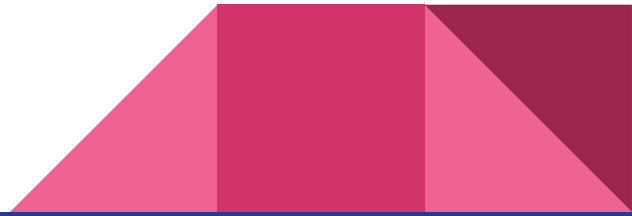


3. Noun

Noun Tags & Labels

- Tags
  - Penn: NN, NNS, NNP, NNPS
  - Universal: NOUN, PROPN (Animacy, Gender, Number, Case)

- Labels
  - 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!*

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

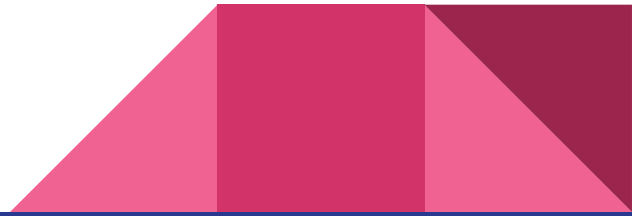
Lexical and Grammatical Properties

- animacy
  - *car* vs. *cat*

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

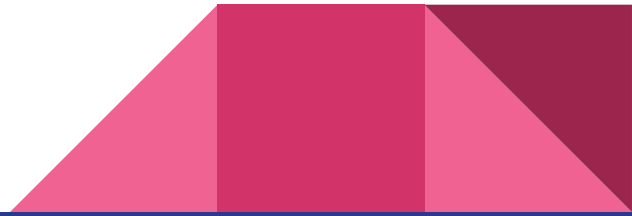
Noun Modifiers

- predeterminer/determiner/case
  - PREDET, DET, NEG, CASE
- numeral
  - NUMMOD
- possessive noun/adjective
  - POSS
- adjective
  - AMOD



Noun Modifiers

- noun adjunct
  - COMPOUND
- appositive
  - APPOS
- prepositional phrase
  - PREP *(UD: nmod)*
- subordinate clause
  - ACL, RELCL



Practice

Analyze the following phrases:

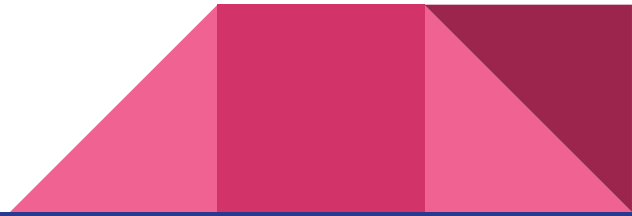
- *his at times very offensive behaviour*
- *all those grossly overrated tax advisers in the city that attacked me*

4. Verb

Verb Tags & Labels

- Tags
  - Penn: VB, VBP, VBZ, VBG, VBD, VBN
  - Universal: VERB (Mood, Tense, Aspect, Voice, etc.)

- Labels
  - VP - verb phrase label
  - S, SQ, SINV - clause
  - SBAR, SBARQ - subordinate clause



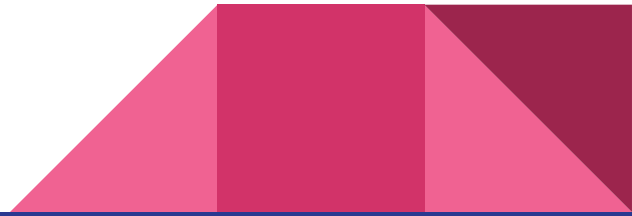
Lexical and Grammatical Properties

- 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



Lexical and Grammatical Properties

- Finite vs. non-finite
- Notional/auxiliary
- Transitivity
  - transitive vs. intransitive vs. reflexive
  - monotransitive, ditransitive

Verb Modifiers

- subject
  - NSUBJ, NSUBJPASS, CSUBJ, CSUBJPASS
- object
  - DOBJ, DATIVE *(UD: iobj)*
- clausal complement
  - CCOMP, XCOMP
- adverbial clause
  - ADVCL

Verb Modifiers

- prepositional phrase
  - PREP, AGENT *(UD: nmod)*
- negation
  - NEG
- adverb
  - ADVMOD
- noun phrase as an adverbial
  - NPADVMOD *(UD: npmod, tmod)*

Verb Modifiers

- auxiliary verb
  - AUX, AUXPASS
- subordinating conjunction
  - MARK
- predicative complement
  - ACOMP *(UD: cop)*
- particle
  - PRT

5. Coordination

Coordination

- constituents
  - the label stays the same: (VP VP CC VP)
  - (UCP PP CC SBAR)

- dependencies
  - CC, PRECONJ, CONJ

Practice

Analyze the following phrases:

- *both books*
- *both the books*
- *both books and newspapers*

Practice

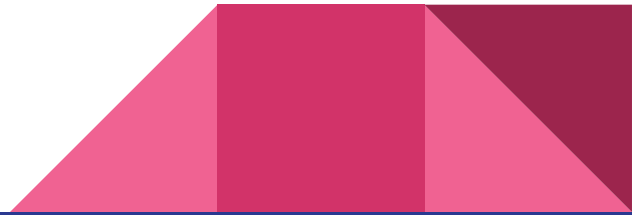
Analyze the following sentences:

- *Now, if you want to receive e-mails about my upcoming shows, then please give me money so I can buy a computer.*
- *The biggest room in the house, the living room, looks out on to a beautiful garden.*
- *All the food tasted excellent, and with the new renovation of chairs and the bathroom, it is awesome.*

6. Functional parts of speech

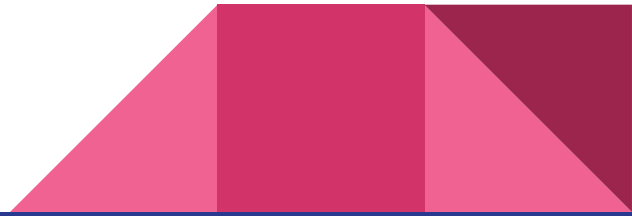
Preposition

- Tags
  - Penn: IN
  - Universal: ADP
- Labels
  - PP - prepositional phrase
- Modifiers
  - adverb: ADVMOD
  - object: POBJ, PCOMP *(UD: absent)*



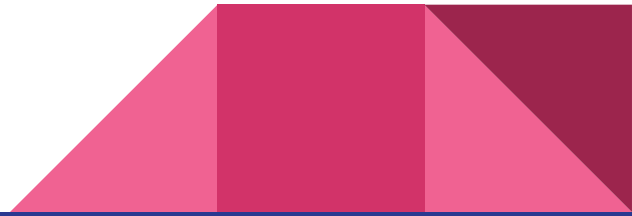
Conjunction

- Tags
  - Penn: IN, CC
  - Universal: SCONJ, CCONJ
- Labels
  - CONJP - coordinate conjunction phrase
- Modifiers
  - adverb: ADVMOD
  - noun phrase: NPADVMOD



Determiner

- Tags
  - Penn: DT
  - Universal: DET
- Modifiers
  - adverb: ADVMOD

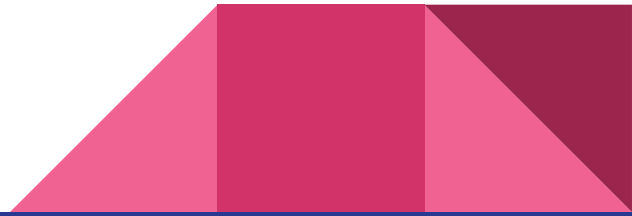


Numeral

- Tags
  - Penn: CD
  - Universal: NUM
- Labels
  - QP - quantifier phrase
- Modifiers
  - adverb: QUANTMOD *(UD: absent)*

Additional notation (Penn)

- INTJ - interjection
- RP, TO - particle
- FW - foreign word
- SYM - non-standard symbol
- LS - list marker



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)
```

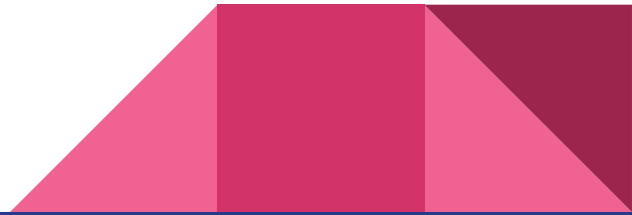
```
nsubj(give-14, then-12)
discourse(give-14, please-13)
root(ROOT-0, give-14)
iobj(give-14, me-15)
dobj(give-14, money-16)
dep(give-14, so-17)
nsubj(buy-20, I-18)
aux(buy-20, can-19)
parataxis(give-14, buy-20)
det(computer-22, a-21)
dobj(buy-20, computer-22)
```

POS taggers and Parsers

- Stanford CoreNLP (6 languages; Java)
- Spacy (7 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:

- languagetool (30 languages; Java), or nlp_uk
- pymorphy2 (Russian, Ukrainian)



POS tagging and Parsing: spaCy

```
>>> import en_core_web_sm
>>> nlp = en_core_web_sm.load()
>>> sentence = nlp("I like turtles because they are cute.")
>>> print(" ".join(token.text + "_" + token.pos_ for token in sentence))
I_PRON like_VERB turtles_NOUN because_ADP they_PRON are_VERB cute_ADJ ._PUNCT
>>> print(" ".join(token.text + "_" + token.tag_ for token in sentence))
I_PRP like_VBP turtles_NNS because_IN they_PRP are_VBP cute_JJ ._.
```

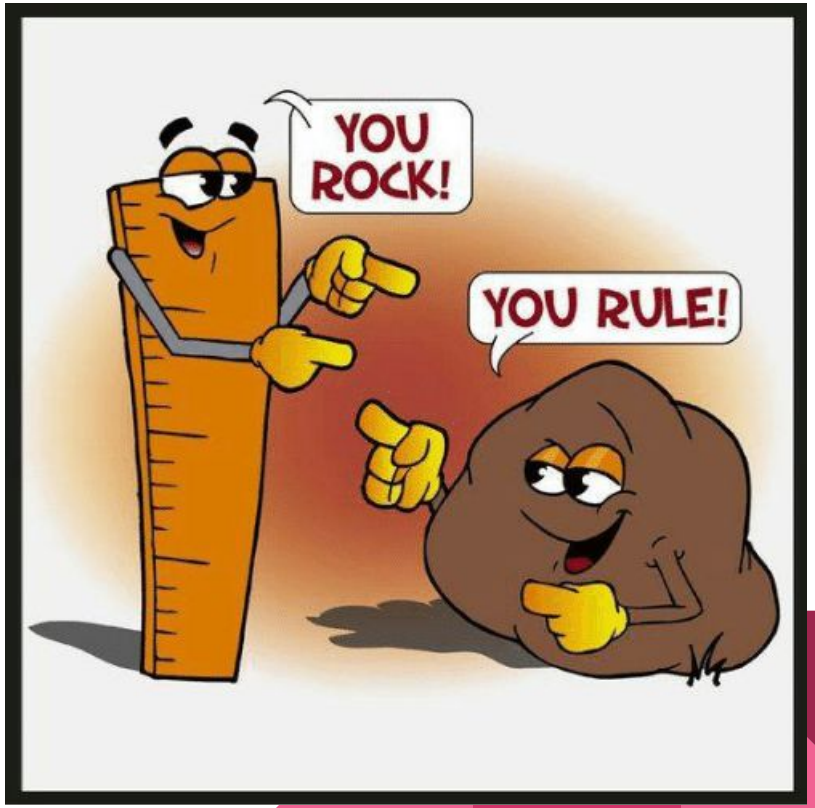
POS tagging and Parsing: spaCy

```
>>> for node in sentence:
...     if node.pos_ == "VERB":
...         for child in node.children:
...             print("{}({}, {})".format(child.dep_, node.text, child.text))
...
nsubj(like, I)
dobj(like, turtles)
advcl(like, are)
punct(like, .)
mark(are, because)
nsubj(are, they)
acomp(are, cute)
```

Conclusion

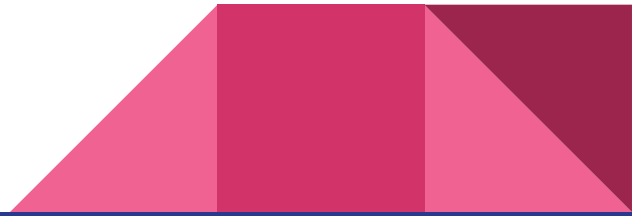
Ambiguities: English

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



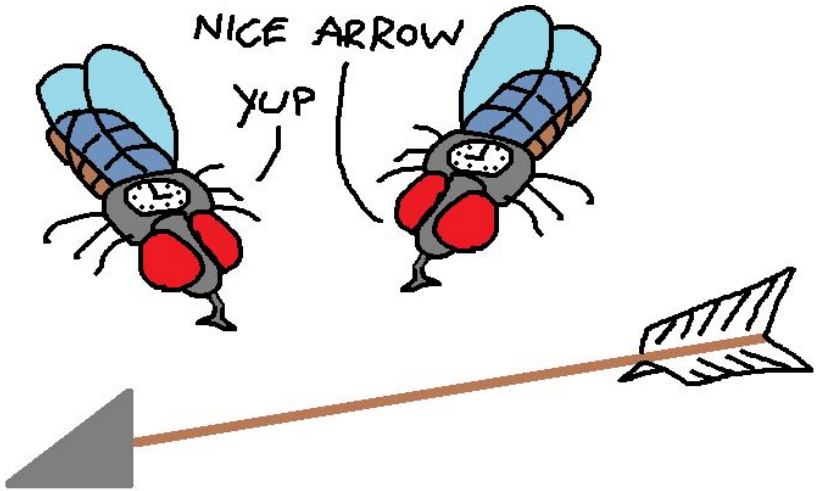
Ambiguities: English

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



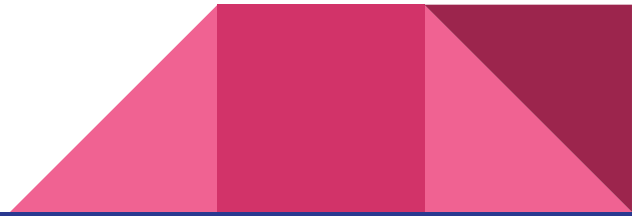
Ambiguities: English

- *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...*



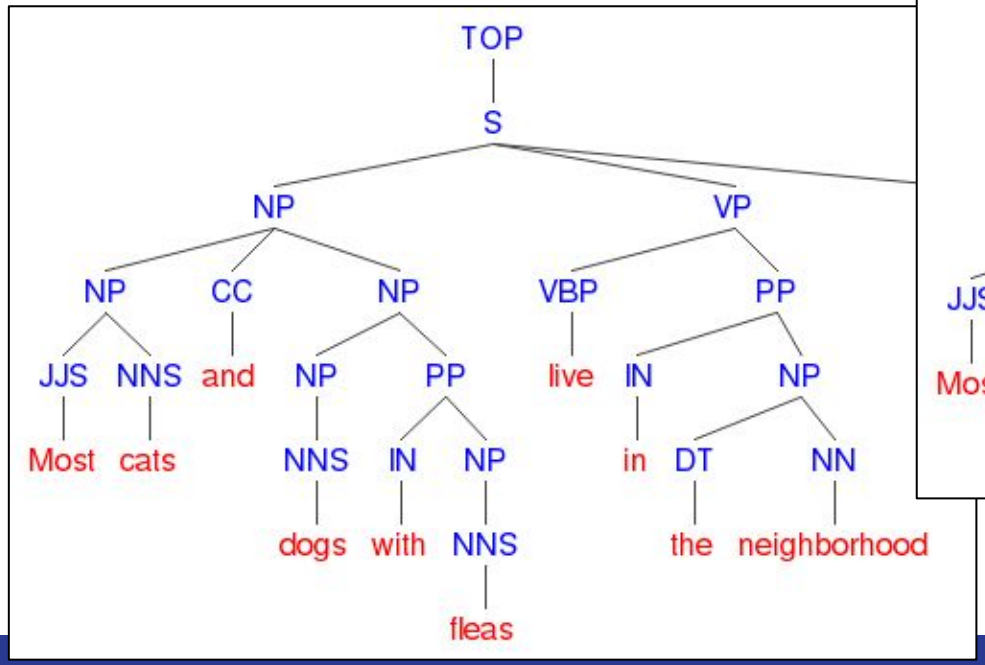
Ambiguities: English

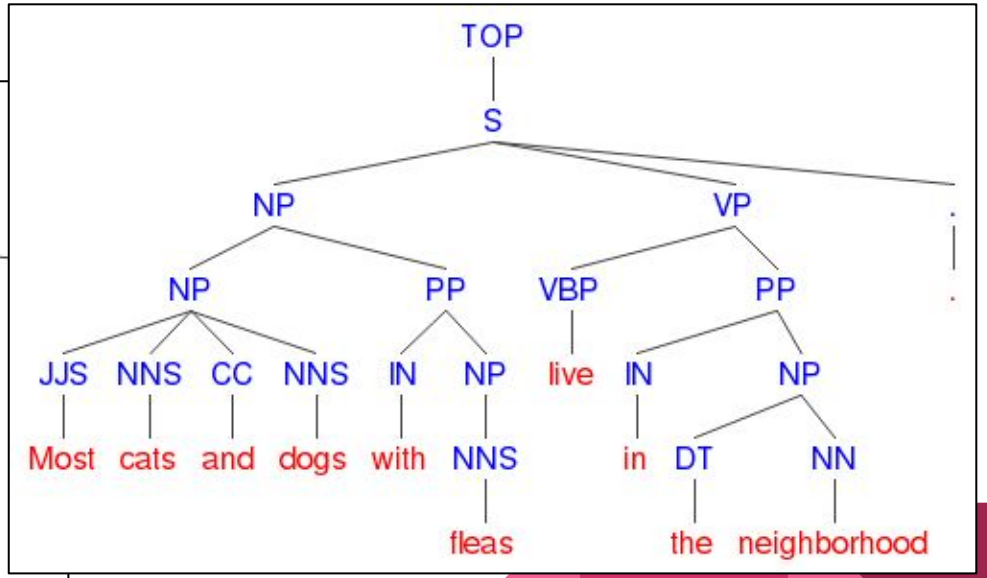
Most cats and dogs with fleas live in the neighborhood.



Ambiguities: English

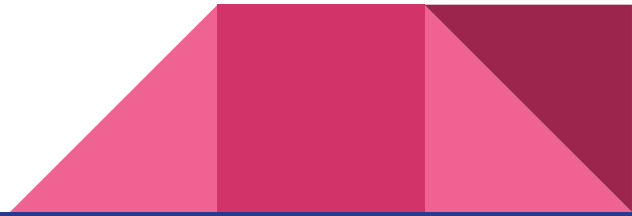
Most cats and dogs with fleas live in the neighborhood.





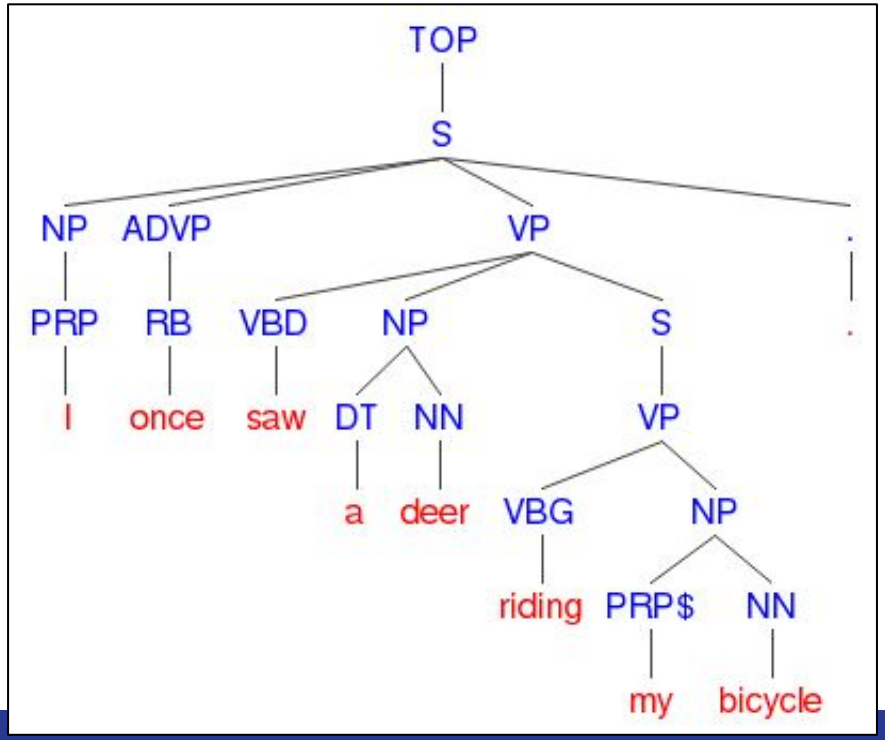
Ambiguities: English

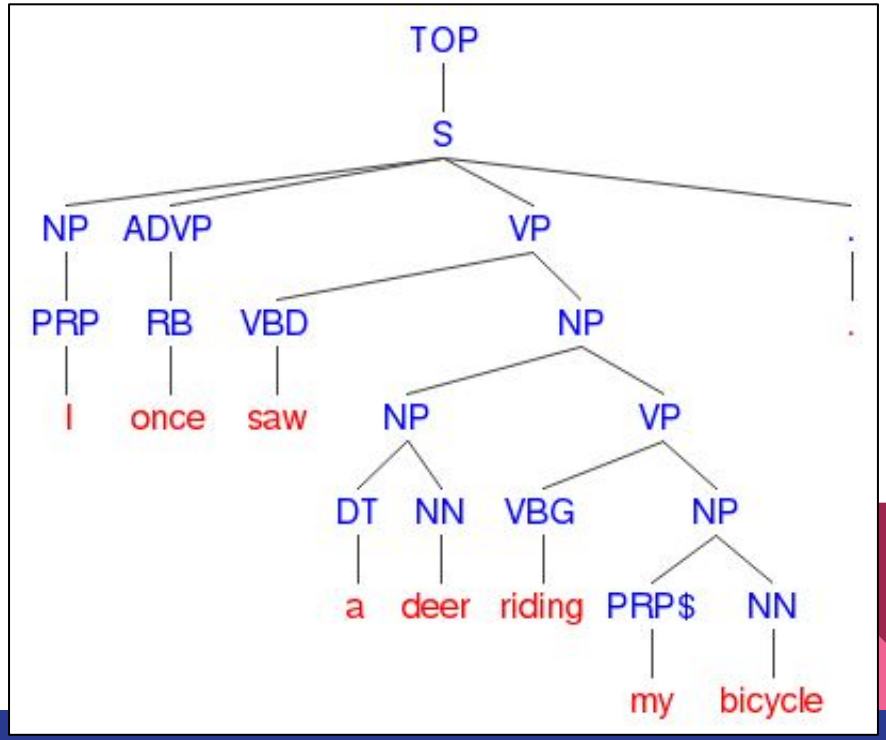
I once saw a deer riding my bicycle.



Ambiguities: English

I once saw a deer riding my bicycle.



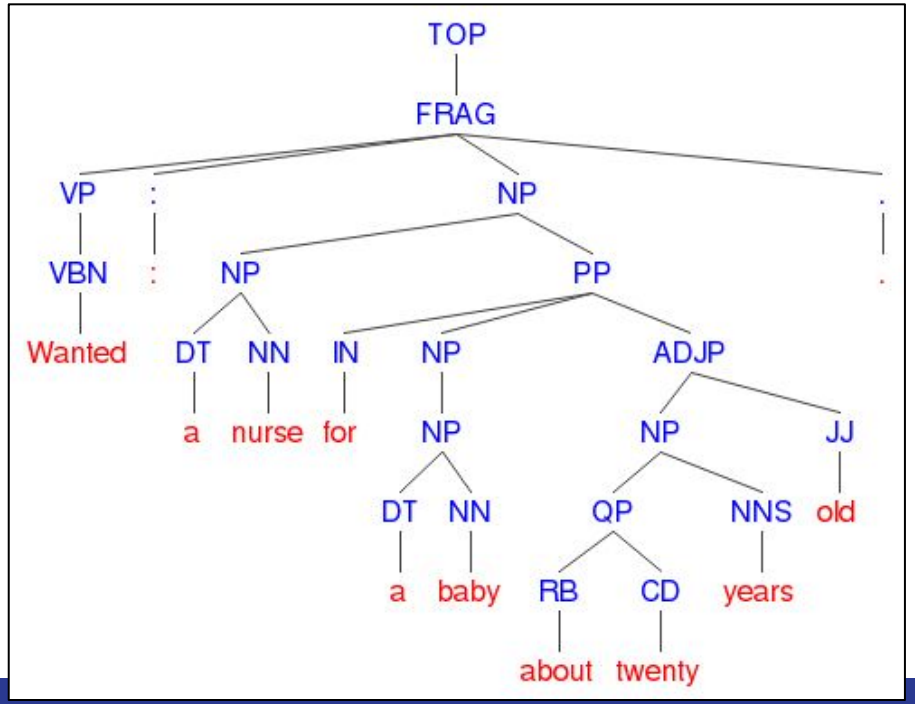


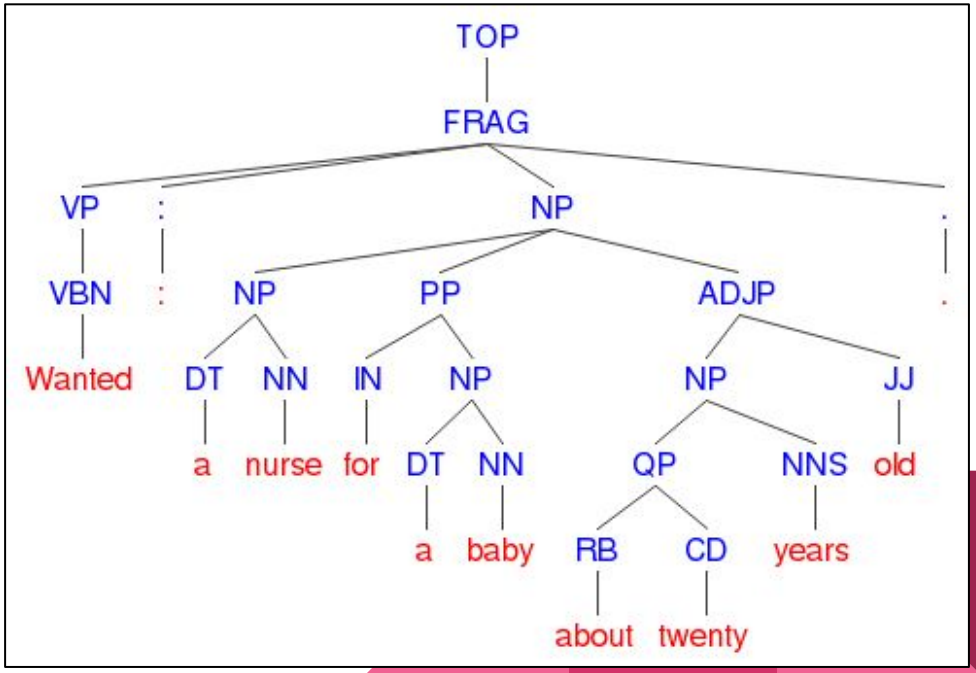
Ambiguities: English

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

Ambiguities: English

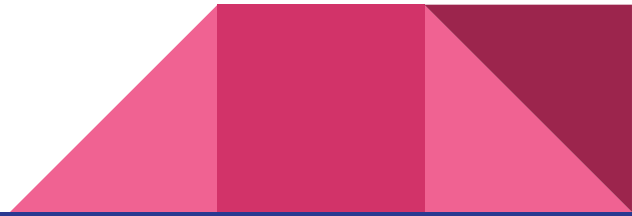
Wanted: a nurse for a baby about twenty years old.





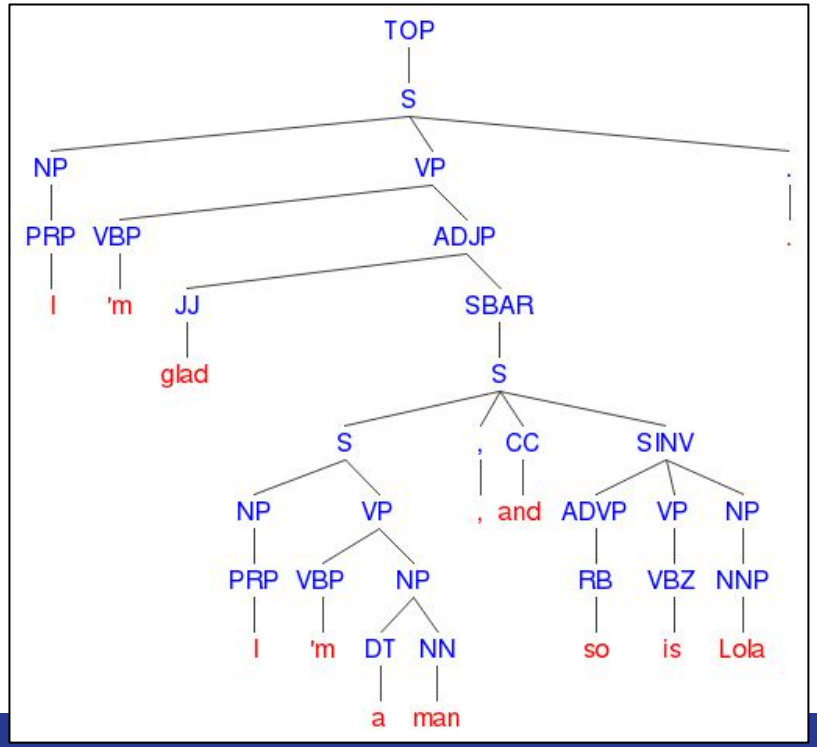
Ambiguities: English

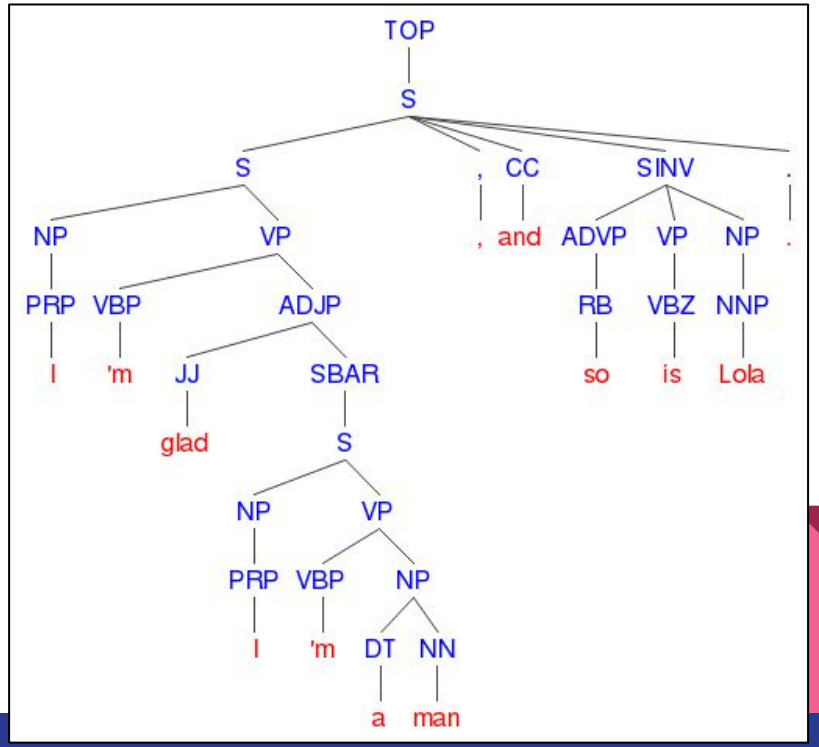
I’m glad I’m a man, and so is Lola.



Ambiguities: English

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

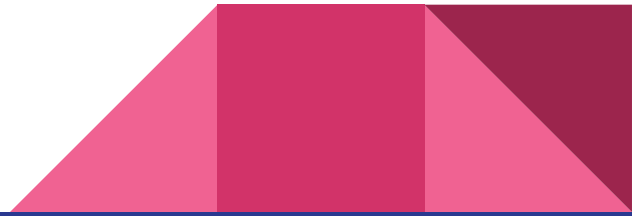




Ambiguities: Ukrainian

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

What is the most ambiguous word?



Ambiguities: Ukrainian

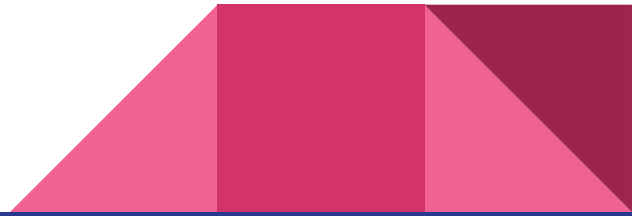
- 2.1 mln unique word forms
- 10K word forms have > 1 possible POS

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']
```

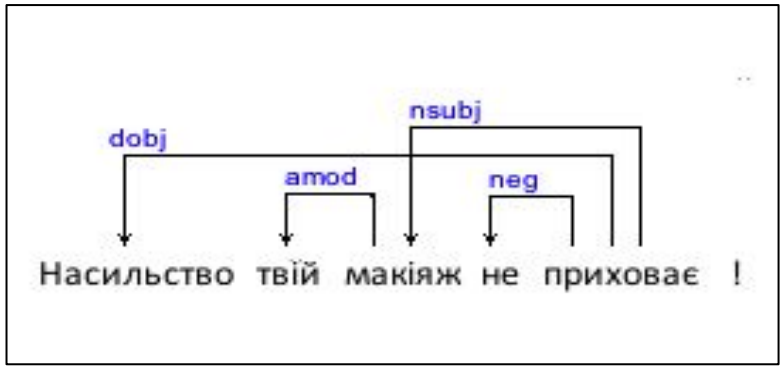
Ambiguities: Ukrainian

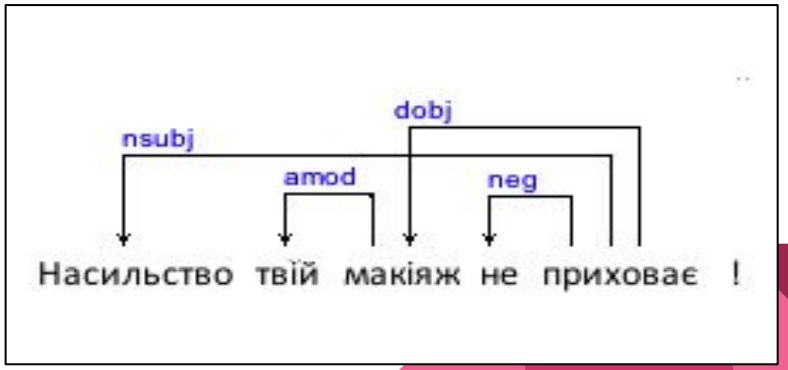
- *Це мало мало значення.*
- *Коло друзів та незнайомців.*



Ambiguities: Ukrainian

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





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?