Until now:

- Introduction to a variety of NLP tasks and methods
- Hands on experience with tokenization
- Experimental setup

Recap: Tune-split

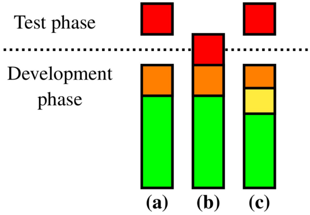


Figure 1: Overview of the use of data splits. red :test orange :dev green :train yellow :tune. a) standard splits for traditional machine learning models b) standard splits as used for neural network models c) our proposed splits for neural network models.

- Note that this is not universally accepted: however, it is important to keep a critical view on our setups!

Other solutions:

- Tune the number of epochs beforehand: might lead to suboptimal performance
- Train for the maximum number of epochs (dynamic learning rate)
- Use other methods to pick the final model: https://aclanthology.org/2021.emnlp-main.459.pdf

Other solutions:

- Tune the number of epochs beforehand: might lead to suboptimal performance
- Train for the maximum number of epochs (dynamic learning rate)
- Use other methods to pick the final model: https://aclanthology.org/2021.emnlp-main.459.pdf
- Note that in cross-domain/lingual settings, it is good practice to use the source data dev (which == tune!)

Takeaways (I hope):

- Tokenization and experimental setup are non-trivial
- Tokenization and experimental setup are important

Any other questions about last week?

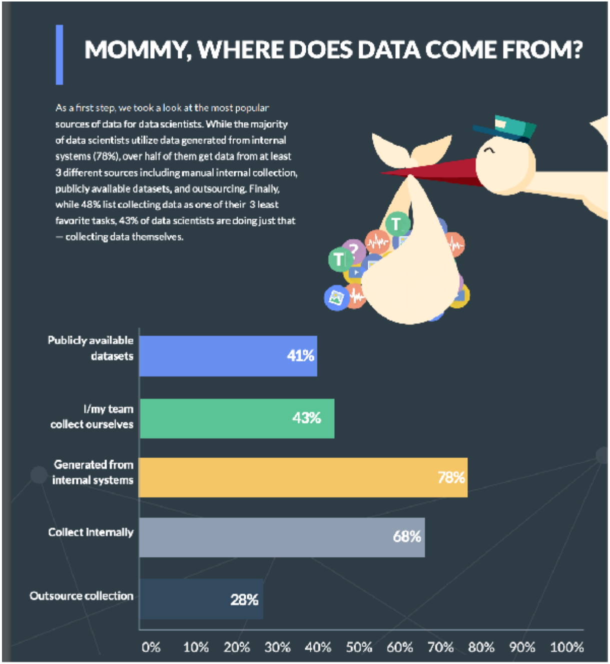
`www.menti.com/` 7727 3458

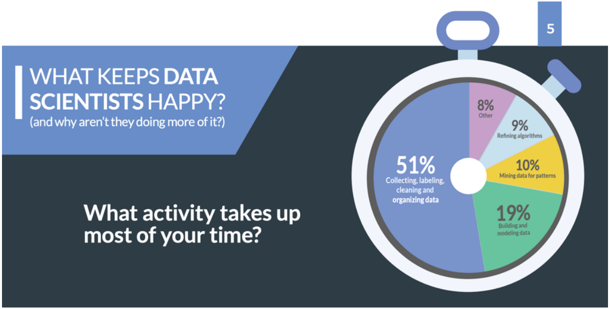
Ground truth

Today:

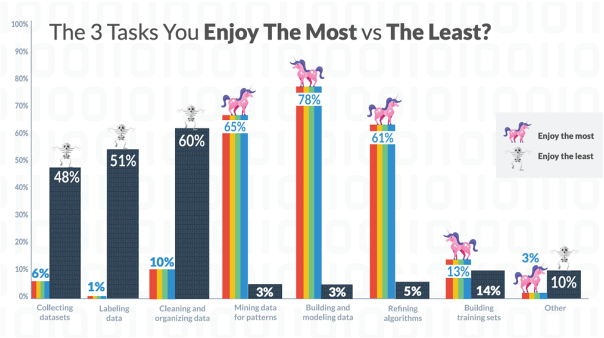
- Collecting data
- Annotating data
- Dataset statements
- POS tagging
- Words as features

Ground truth





From Figure-Eight Datascience report:
`https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport.pdf`



From Figure-Eight Datascience report:
`https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport.pdf`

Selection of data: Methodology

- **Top-down**: What do I want? What is the purpose?
- **Bottom-up**: What can I get? From where? Using what kind of method?

Scraping websites or using open APIs (! CAREFUL !)

- e.g. in your first year project
- Copyright issues: Published on the web != I can freely redistribute
- Moreover: data storage regulations, GDPR
- And: not being able to share the data leads to not reproducable results

What is the “population”? What is “representative”?

- A sample is representative if what’s true about the sample is also true in general
- But are our, say, collected tweets, representative of communication in general?
- Depending on the type of data it can be hard to determine whether a sample is representative in practice - be aware of BIASES
- Useful to document the composition of the dataset

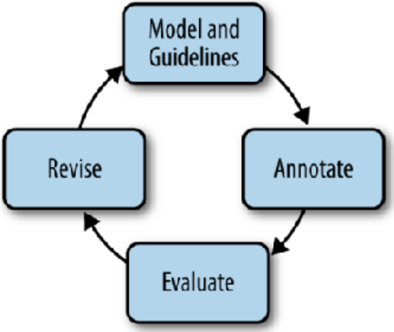
(dataset statements)

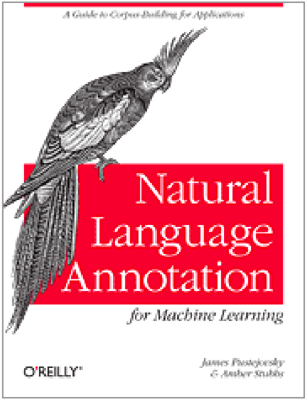
Examples of biases

Imagine doing a poll before an election. What are possible biases?

- Sampling bias (e.g. polled people over phone only)
- Non-response bias (who are the people who answer the survey?)

Annotating data





Types of annotation in NLP

- Create gold standard
- How well can humans do task X?

Types of annotation in NLP

Closely related to the types of tasks (non-exhaustive):

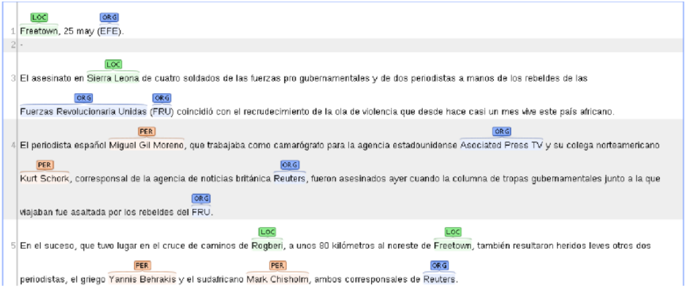
- **text classification**: sentiment, topic, intent, stance, language(variety)
- **relation between texts**: text similarity, textual entailment
- **structured prediction**: syntactic parsing, relation extraction, coreference resolution
- **sequence labeling**: Parts-Of-Speech, named entity recognition, semantic role labeling, language(variety)
- **text generation**: machine translation, question answering, dialogue, data to text, summarization
- **transformations**: grammatical error correction, tokenization, lemmatization

For some tasks, annotation is "freely" available!

- Sentiment analysis
- Emoji prediction
- Author identification
- (Machine translation)
- Language modeling (lecture 5,7,11,12)

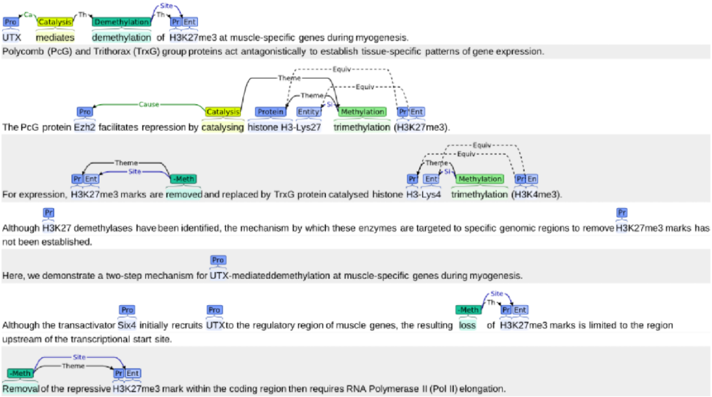
Does this mean that annotation is irrelevant?

Types of annotation



- `http://brat.nlplab.org/`
- WebAnno is a similar tool
  `https://webanno.github.io/webanno/`

Types of annotation



Lets give it a try: `www.menti.com/` 7727 3458

Informativity vs Correctness

Annotation involves a trade-off between:

- Informativity: useful for your task
- Correctness: annotation that is not too difficult for annotators to complete accurately

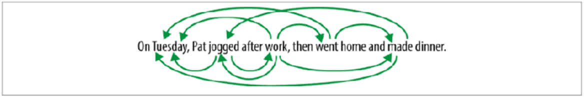
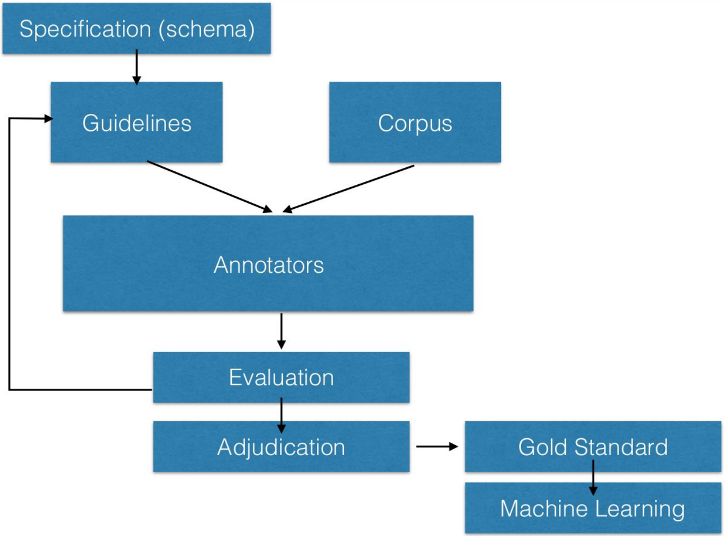


Figure 2-1. All temporal relations over events and times

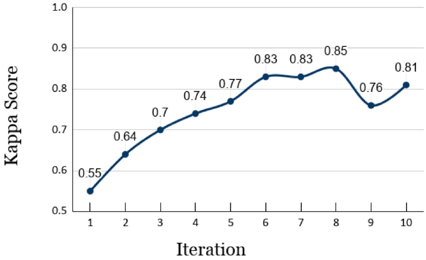
Annotation Process: Details



Make it cheaper:

- Split up data in chunks
- Annotate each chunk with multiple annotators
- Discuss and resolve disagreements (update guidelines)
- When Kappa converges: annotate rest of the data with 1 annotator

Example from Dialogue Act Classification of social media data:



picture from Vielsted and Wallenius (2021)

Guidelines

- Annotation guidelines are the instructions to your annotators that specify how to apply the model (the schema) to the data
- the annotation guidelines should answer the following questions
  - What is the goal (purpose)?
  - What is each label called and how is it used?
  - What parts of the text do you want annotated?
  - How are difficult cases handled?

Difficult cases

For example, Santorini et al. (1990) specified 'problematic cases' for the Penn Treebank parts-of-speech tags:

4 Problematic cases

This section discusses difficult tagging decisions. Section 4.1 discusses parts of speech that are easily confused and guidelines on how to tag such cases. Section 4.2 contains an alphabetical list of specific problematic words and collocations.

4.1 Confusing parts of speech

This section discusses parts of speech that are easily confused and gives guidelines on how to tag such cases.

CC or DT

When they are the first members of the double conjunctions *both ... and*, *either ... or* and *neither ... nor*, *both*, *either* and *neither* are tagged as coordinating conjunctions (CC), not as determiners (DT).

EXAMPLES: Either/DT child could sing.

But:

Either/CC a boy could sing or/CC a girl could dance.
Either/CC a boy or/CC a girl could sing.
Either/CC a boy or/CC girl could sing.

Be aware that *either* or *neither* can sometimes function as determiners (DT) even in the presence of *or* or *nor*.

EXAMPLE: Either/DT boy or/CC girl could sing.

CD or JJ

Number-number combinations should be tagged as adjectives (JJ) if they have the same distribution as adjectives.

EXAMPLES: a 50–3/JJ victory (cf. a handy/JJ victory)

Hyphenated fractions *one-half*, *three-fourths*, *seven-eighths*, *one-and-a-half*, *seven-and-three-eighths* should be tagged as adjectives (JJ) when they are prenominal modifiers, but as adverbs (RB) if they could be replaced by *double* or *twice*.

EXAMPLES: one-half/JJ cup; cf. a full/JJ cup
one-half/RB the amount; cf. twice/RB the amount; double/RB the amount

Inter-annotator agreement:

- How to compare annotations of multiple annotators
- Percentage of cases where they agree?

Does the difficulty of the task matter?

- Sentiment analysis: 2 classes
- Two annotators agree 70% of the cases, is this good? does this mean the task is difficult?

Cohen's Kappa

- Takes chance into account
- Usually between 0-1 (below 0 means worse than random)
- >2 annotators: Fleiss Kappa

$\kappa = \frac{\textit{How much agreement beyond chance was found}}{\textit{How much agreement beyond chance is possible}}$ $\kappa = \frac{p_o - p_e}{1 - p_e}$

- p_o = observed agreement (accuracy between 2 annotators)
- p_e = expected agreement

Expected agreement is sum of the expected score of all classes.
For 2-class sentiment analysis (+/-):

	A1+	A1-
A2+	40	10
A2-	20	30

Expected agreement +: $\frac{40+20}{100} * \frac{40+10}{100} = 0.3$
Expected agreement -: $\frac{10+30}{100} * \frac{20+30}{100} = 0.2$

Expected agreement is sum of the expected score of all classes.
For 2-class sentiment analysis (+/-):

	A1+	A1-
A2+	40	10
A2-	20	30

Expected agreement +: $\frac{40+20}{100} * \frac{40+10}{100} = 0.3$
Expected agreement -: $\frac{10+30}{100} * \frac{20+30}{100} = 0.2$
$p_e = 0.3 + 0.2 = 0.5$
$\kappa = \frac{0.7-0.5}{1-0.5} = .4$

Interpretation of κ, rule of thumb:

- $0 =$ random
- $< 0.4 =$ weak
- 0.4-0.6 $=$ moderate
- 0.6-0.8 $=$ strong
- 0.8-0.99 $=$ almost perfect
- $1 =$ perfect

Interpretation of κ, rule of thumb:

- $0 =$ random
- $< 0.4 =$ weak
- 0.4-0.6 $=$ moderate
- 0.6-0.8 $=$ strong
- 0.8-0.99 $=$ almost perfect
- $1 =$ perfect
- But don't forget to use common sense!

Note that:

- the same p_o does not always lead to the same κ
- when there are more classes, p_e is generally lower, thus a similar p_o generally leads to a higher κ

What can we learn (not all at the same time!):

- How good the annotators are at the task
- How difficult the task is
- Theoretical upperbound for our NLP systems
- Do we need to revise the guidelines/setup?

Dataset statements

Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science

Emily M. Bender
Department of Linguistics
University of Washington
ebender@uw.edu

Batya Friedman
The Information School
University of Washington
batya@uw.edu

https://www.aclweb.org/anthology/Q18-1041/

A. CURATION RATIONALE: which data and why?
B. LANGUAGE VARIETY/VARIETIES: BCP-47 language tags (639-3 ISO?)
C. SPEAKER DEMOGRAPHIC: age, gender, native tongue, socioeconomic status, race/ethnicity
D. ANNOTATOR DEMOGRAPHIC: age, gender, native tongue, socioeconomic status, race/ethnicity
E. SPEECH SITUATION: time, place, modality, spontaneity
F. TEXT CHARACTERISTICS: genre/topic
G. RECORDING QUALITY
H. OTHER
I. PROVENANCE APPENDIX

A. CURATION RATIONALE: which data and why?
B. LANGUAGE VARIETY/VARIETIES: BCP-47 language tags (639-3 ISO?)
C. SPEAKER DEMOGRAPHIC: age, gender, native tongue, socioeconomic status, race/ethnicity
D. ANNOTATOR DEMOGRAPHIC: age, gender, native tongue, socioeconomic status, race/ethnicity
E. SPEECH SITUATION: time, place, modality, spontaneity
F. TEXT CHARACTERISTICS: genre/topic
G. RECORDING QUALITY
H. OTHER
I. PROVENANCE APPENDIX

Can we "improve"? `www.menti.com/` 7727 3458

Is annotator disagreement "noise"?

Perhaps not!, (almost) all NLP tasks are subjective!: Disagreement can be used to improve:

- guidelines
- models
- evaluation

Source:	**To put it in the nutshell**, I believe that people should **have the obligation** to tell their relatives about **the** genetic **testing result** for the good of their health.
A1	To put it in **a** nutshell, I believe that people should **be obliged** to tell their relatives about **their** genetic **test results** for the good of their health.
A2	**In a nutshell**, I believe that people should have **an** obligation to tell their relatives about the genetic testing result for the good of their health.
A3	**In summary**, I believe that people should have the obligation to tell their relatives about the genetic testing result for the good of their health.
A4	**In a nutshell**, I believe that people should **be obligated** to tell their relatives about the genetic testing result for the good of their health.
A5	To put it in **a** nutshell, I believe that people should **be obligated** to tell their relatives about the genetic testing **results** for the good of their health.
A6	To put it in the nutshell, I believe that people should have **an** obligation to tell their relatives about **their** genetic **test results** for the good of their health.
A7	To put it in **a** nutshell, I believe that people should have the obligation to tell their relatives about the genetic testing result for the good of their health.
A8	To put it in **a** nutshell, I believe that people should **be obligated** to tell their relatives about the genetic testing result for the good of their health.
A9	To put it in **a** nutshell, I believe that people should have the obligation to tell their relatives about the genetic **test** result for the good of their health.
A10	To put it in **a** nutshell, I believe that people should have the obligation to tell their relatives about the genetic **test results** for the good of their health.

From Bryant and NG (2015):
`https://aclanthology.org/P15-1068.pdf`

Sequence Labeling

- Many real-world applications can be cast as sequence labelling problems that involve assigning labels to each element in a sequence.
- Sequence Labeling: assigns labels to each element in a sequence

Sequence Labeling

- Input Space X_s: sequences of items to label
- Output Space Y_s: sequences of output labels
- Model: $s_{params}(x, y)$
- Prediction: $argmax_y s_{params}(x, y)$
- Is a particular type of structured prediction problem

Parts-Of-Speech Tagging

In dictionaries every word has a syntactic function/class. In NLP we use(d) WordNet(s): every sense has a syntactic function:

WordNet Search - 3.1
- WordNet home page - Glossary - Help

Word to search for: duck [Search WordNet]

Display Options: (Select option to change) [Change]
Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations
Display options for sense: (gloss) "an example sentence"

Noun

- S: (n) **duck** (small wild or domesticated web-footed broad-billed swimming bird usually having a depressed body and short legs)
- S: (n) **duck**, duck's egg ((cricket) a score of nothing by a batsman)
- S: (n) **duck** (flesh of a duck (domestic or wild))
- S: (n) **duck** (a heavy cotton fabric of plain weave; used for clothing and tents)

Verb

- S: (v) **duck** (to move (the head or body) quickly downwards or away) *"Before he could duck, another stone struck him"*
- S: (v) **duck** (submerge or plunge suddenly)
- S: (v) dip, douse, **duck** (dip into a liquid) *"He dipped into the pool"*
- S: (v) hedge, fudge, evade, put off, circumvent, parry, elude, skirt, dodge, **duck**, sidestep (avoid or try to avoid fulfilling, answering, or performing (duties, questions, or issues)) *"He dodged the issue"; "she skirted the problem"; "They tend to evade their responsibilities"; "he evaded the questions skillfully"*

Example of Lexical Ambiguity

6 different parts-of-speech for the word back:

```
earnings growth took a back/JJ seat
a small building in the back/NN
a clear majority of senators back/VBP the bill
Dave began to back/VB toward the door
enable the country to buy back/RP about debt
I was twenty-one back/RB then
```

Syntactic analysis often considered as first step to **disambiguation** (and interpration)

Parts-Of-Speech Tagging

Assign each token in a sentence its **parts-of-speech tag**.

```
I      PRON
see    VERB
the    DET
light  NOUN
!      PUNCT
```

Parts-Of-Speech Tagging

Assign each token in a sentence its **parts-of-speech tag**.

```
I      PRON
see    VERB
the    DET
light  NOUN
!      PUNCT
```

General rule of thumb for classes:

- A word should be able to be replaced by any other word in its class, and the sentence should remain syntactically correct

How many word-classes exist?

How many word-classes exist? Most popular tag-sets:

- Penn Treebank tag-set (45-52 tags)
- Universal POS tags (12-17 tags)

How many word-classes exist? Most popular tag-sets:

- Penn Treebank tag-set (45-52 tags)
- Universal POS tags (12-17 tags)

Differences: 'VERB' in UPOS roughly corresponds to:

- VB: Verb, base form
- VBD: Verb, past tense
- VBG: Verb, gerund or present participle
- VBN: Verb, past participle
- VBP: Verb, non-3rd person singular present
- VBZ: Verb, 3rd person singular present

In recent datasets, UPOS is often combined with morphological information:

```
1  Hvor      hvor    ADV     _    _          2      advmod  _     _
2  kommer    komme   VERB    _    Mood=Ind|Tense=Pres  0     ro
3  julemanden        julemand     NOUN  _    Definite=Def|Nu
4  fra       fra     ADP     _    AdpType=Prep 1      case   _
5  ?         ?       PUNCT   _    _          2      punct   _     _
```

In recent datasets, UPOS is often combined with morphological information:

```
1  Hvor      hvor     ADV      _    _           2     advmod  _     _
2  kommer    komme    VERB     _    Mood=Ind|Tense=Pres  0     ro
3  julemanden         julemand     NOUN    _    Definite=Def|Nu
4  fra       fra      ADP      _    AdpType=Prep 1          case   _
5  ?         ?        PUNCT    _    _           2     punct   _     _
```

Interesting!

Two POS classes

Parts-of-speech can be divided into two broad supercategories: **closed class**and **open class** types. Closed classes are those with relatively fixed membership:

- prepositions: on, under, over, near, by, at, from, to, with
- particles: up, down, on, off, in, out, at, by
- determiners: a, an, the
- conjunctions: and, but, or, as, if, when
- pronouns: she, who, I, others
- auxiliary verbs: can, may, should, are
- numerals: one, two, three, first, second, third

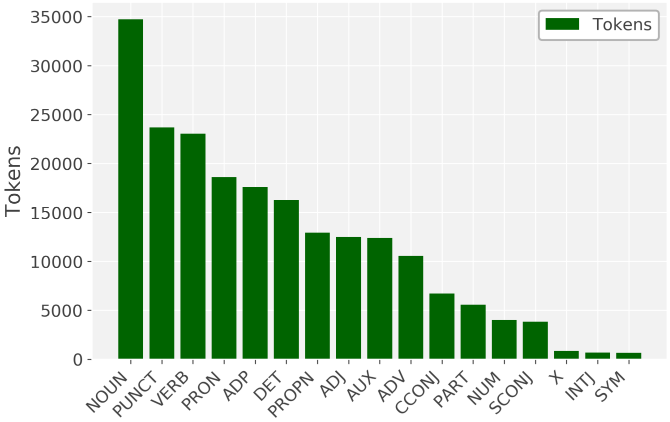
The Universal Dependencies tagset

We will use the Universal POS tag set:
`https://universaldependencies.org/u/pos/index.html` of the Universal Dependencies project (Nivre et al., 2016):

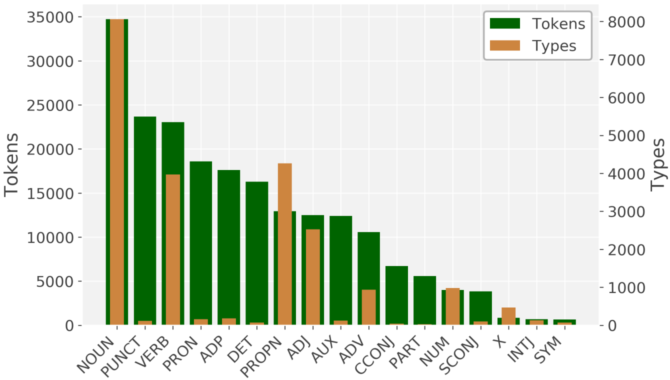
Open class	Closed class	Other
ADJ	ADP	PUNCT
ADV	AUX	SYM
INTJ	CCONJ	X
NOUN	DET	
PROPN	NUM	
VERB	PART	
PRON		
SCONJ		

Following slides are all based on the Univeral Dependencies (UD) version of English Web Treebank (EWT):

- weblogs
- newsgroups
- e-mails
- reviews
- yahoo! answers



Tokens
35000
30000
25000
20000
15000
10000
5000
0
NOUN
PUNCT
VERB
PRON
ADP
DET
PROPN
ADJ
AUX
ADV
CCONJ
PART
NUM
SCONJ
X
INTJ
SYM



Tokens
Types
35000
30000
25000
20000
15000
10000
5000
0
8000
7000
6000
5000
4000
3000
2000
1000
0
NOUN
PUNCT
VERB
PRON
ADP
DET
PROPN
ADJ
AUX
ADV
CCONJ
PART
NUM
SCONJ
X
INTJ
SYM

NOUN

Refers to a person, place, thing, animal or idea

NOUN

Refers to a person, place, thing, animal or idea Examples:

```
No    wine  glasses   .
DET   NOUN  NOUN      PUNCT
.......   the  rest  was  history   !
PUNCT     DET  NOUN  AUX  NOUN      PUNCT
```

VERB

Describe events or actions

VERB

Describe events or actions Examples:

```
The  service  stunk  .
DET  NOUN     VERB   PUNCT
Nobody  lived  there  .
PRON    VERB   ADV    PUNCT
```

PRON: Pronoun

Noun whose meaning is recoverable from the linguistic or extralinguistic context

PRON: Pronoun

Noun whose meaning is recoverable from the linguistic or extralinguistic context Examples:

```
We      heard   nothing   .
PRON    VERB    PRON        PUNCT
He      knows   his    bees   !
PRON    VERB    PRON   NOUN   PUNCT
```

PRON: Pronoun

Noun whose meaning is recoverable from the linguistic or extralinguistic context Examples:

```
We    heard  nothing  .
PRON  VERB   PRON     PUNCT
He    knows  his   bees  !
PRON  VERB   PRON  NOUN  PUNCT
```

Most frequent types: I, you, it, they, my, that, your, he, we, me

ADP: Adposition

Prepositions and postpositions: express relations of noun-phrase to other unit

ADP: Adposition

Prepositions and postpositions: express relations of noun-phrase to other unit Examples:

```
Call  dominos  in   your  town  .
VERB  PROPN    ADP  PRON  NOUN  PUNCT
Bad  for  business  .
ADJ  ADP  NOUN      PUNCT
```

ADP: Adposition

Prepositions and postpositions: express relations of noun-phrase to other unit Examples:

```
Call   dominos   in   your   town   .
VERB   PROPN     ADP  PRON   NOUN   PUNCT
Bad   for   business   .
ADJ   ADP   NOUN       PUNCT
```

Most frequent types: of, in, to, for, with, on, at, from, by, as

But first:

#	△	Team	Members	Score	Entries	Last	Solution
1	▲ 1	mie jonasson		0.48854	5	5d	
2	▼ 1	**robvanderg**		0.79770	5	4d	
3	—	Christian Weidemann		1.08778	1	4d	
4	—	Marcin Sroka		1.23282	1	16h	
5	—	Nicolai Kofod-Jensen		3.48473	1	11h	
6	—	viggo-gascou		8.24045	1	4d	

- Stroopwafels, Dumkes, Fristi, Rivella, roze koeken.

DET: Determiner

Articles (the,a,and) and determiners:

- demonstrative determiners: this, that
- interrogative/relative determiners: which, what
- quantifiers: many, some, all, no

DET: Determiner

Articles (the,a,and) and determiners:

- demonstrative determiners: this, that
- interrogative/relative determiners: which, what
- quantifiers: many, some, all, no

Examples:

```
These  guys  were  the  best  .
DET    NOUN  AUX   DET  ADJ   PUNCT
Any  plans  for  the  weekend  ?
DET  NOUN   ADP  DET  NOUN     PUNCT
```

PROPN: Proper Noun

Name of object, entity, person (Named Entity)

PROPN: Proper Noun

Name of object, entity, person (Named Entity) Examples:

```
Texas  Roadhouse  is   WAY  better  !!
PROPN  PROPN      AUX  ADV  ADJ     PUNCT
Just   Autos  Thank  you   !
PROPN  PROPN  VERB   PRON  PUNCT
```

ADJ: Adjective

Modifier of nouns (specifies properties or attributes). Can also be used as predicate.

ADJ: Adjective

Modifier of nouns (specifies properties or attributes). Can also be used as predicate. Examples:

`Great`	`work`	`and`	`honest`	`establishment`	`!`
`ADJ`	`NOUN`	`CCONJ`	`ADJ`	`NOUN`	`PUNCT`

`My`	`apartment`	`was`	`usually`	`quiet`	`.`
`PRON`	`NOUN`	`AUX`	`ADV`	`ADJ`	`PUNCT`

AUX: Auxilary

An auxiliary is a function word that accompanies the lexical verb of a verb phrase and expresses grammatical distinctions not carried by the lexical verb, such as person, number, tense, mood, aspect, voice or evidentiality.

AUX: Auxilary

An auxiliary is a function word that accompanies the lexical verb of a verb phrase and expresses grammatical distinctions not carried by the lexical verb, such as person, number, tense, mood, aspect, voice or evidentiality. Much easier to understand by example:

```
He    was  drinking  .
PRON  AUX  VERB      PUNCT
That  's   it    .
PRON  AUX  PRON  PUNCT
Hopelessness  will  be   lost  .
NOUN          AUX   AUX  VERB  PUNCT
```

ADV: Adverb

Modifies verbs, adjectives or other adverbs

ADV: Adverb

Modifies verbs, adjectives or other adverbs Examples:

I	will	never	go	back	.
PRON	AUX	ADV	VERB	ADV	PUNCT
Beardies	are	actually	quite	delicate	.
NOUN	AUX	ADV	ADV	ADJ	PUNCT

CCONJ: Coordinating conjunction

Conjunction in which both parts are syntactically equal

CCONJ: Coordinating conjunction

Conjunction in which both parts are syntactically equal Examples:

```
Great  Wine  &      Service
ADJ    NOUN  CCONJ  NOUN
Outdated  but    not  bad
ADJ       CCONJ  ADV  ADJ
```

CCONJ: Coordinating conjunction

Conjunction in which both parts are syntactically equal Examples:

Great	Wine	&	Service
ADJ	NOUN	CCONJ	NOUN

Outdated	but	not	bad
ADJ	CCONJ	ADV	ADJ

Most frequent types: and, or, nor, either, neither, but, both

SCONJ: Subordinating conjunction

Conjunction in which one part is a *constituent* of the other

SCONJ: Subordinating conjunction

Conjunction in which one part is a *constituent* of the other
Examples:

```
Thanks  for    asking  .
NOUN    SCONJ  VERB    PUNCT
Even  if     you   line  up   .
ADV   SCONJ  PRON  VERB  ADP  PUNCT
```

SCONJ: Subordinating conjunction

Conjunction in which one part is a *constituent* of the other
Examples:

```
Thanks  for    asking  .
NOUN    SCONJ  VERB    PUNCT
Even  if     you   line  up   .
ADV   SCONJ  PRON  VERB  ADP  PUNCT
```

Most frequent types: before, if, like, for, of, as, in

PART: Particle

Function words that have to be combined with other words (and do not fall in another category):

- possessive marker
- negation particle

PART: Particle

Function words that have to be combined with other words (and do not fall in another category):

- possessive marker
- negation particle

Examples:

```
The   window   did   n't    break   !
DET   NOUN     AUX   PART   VERB    PUNCT
Follow   Katrina   's     path
VERB     PROPN     PART   NOUN
```

PART: Particle

Function words that have to be combined with other words (and do not fall in another category):

- possessive marker
- negation particle

Examples:

```
The   window   did   n't    break   !
DET   NOUN     AUX   PART   VERB    PUNCT
Follow   Katrina   's     path
VERB     PROPN     PART   NOUN
```

But:

```
It     's    easy   .
PRON   AUX   ADJ    PUNCT
```

Classes left out:

- NUM: [0-9]!
- PUNCT: . , - ” ’ ( ) ? ! ; : / ¿ ¡
- SYM: $ % ;) + @ as at, # as number
- INTJ: please, thanks, yes, well, lol, hey, ok
- X: misc.

www.menti.com/ 7727 3458

What about span labeling?

Named Entity Recognition

- Identify named entities in text
- Commonly with classes: person, organization, location and misc

What about span labeling?

Named Entity Recognition

- Identify named entities in text
- Commonly with classes: person, organization, location and misc

[Barack Obama]$_{PER}$ was born in [Hawaii]$_{LOC}$

What about span labeling?

Named Entity Recognition

- Identify named entities in text
- Commonly with classes: person, organization, location and misc

[Barack Obama]$_{PER}$ was born in [Hawaii]$_{LOC}$ First try:

```
| Barack | Obama | was |  born | in | Hawaii |
| PER    | PER   | O   |  O    | O  | LOC    |
```

What can go wrong?

Der var mange flotte præmier til de vindende hold , udsat af DEN DANSKE BANK TESS INN LYNGBY EL-service BIKUBEN Beboerbladet HER og ESSO Motorcenter .

Der var mange flotte præmier til de vindende hold , udsat af DEN DANSKE BANK TESS INN LYNGBY EL-service BIKUBEN Beboerbladet HER og ESSO Motorcenter .

Label tokens as beginning (B), inside (I), or outside (O) a **named entity:**

```
| Barack | Obama | was |  born | in | Hawaii |
| B-PER  | I-PER | O   | O     | O  | B-LOC  |
```

Also called BIO encoding (or IOB)

How to evaluate NER?

How to pick a main metric

Classes are equally important (as opposed to instances)

Y

The dataset is balanced

Y

Accuracy

N

Macro-f1

N

I'm interested in 1 class (often the case in binary)

Y

Class-f1

N

I have 1 label per instance OR < 3 classes

Y

Accuracy

N

Micro-f1

Micro-f1 with 2 classes == accuracy; but I prefer to use the simpler one for naming it

Weighted f1 is not in here as Rob thinks it is not a good fit for main metric

If Class-f1 is used, be clear about this (which class)!

- Is “The IT University of Copenhagen” more important than “University of Copenhagen”?

- Is “The IT University of Copenhagen” more important than “University of Copenhagen”?
- We use span-F1 instead:
  - Recall: How many of the total existing spans did we found?
  - Precision: How many of the spans we found are correct?

Words as features

Machine learning setup:

- Instances are described by features
- Features are transformed/interpreted to predict a label/value (input $\mapsto$ output)

Machine learning

size	distance	age	price
110	25	105	4,000,000
50	5	35	2,500,000

In language, we do not have numerical features!

- consider text classification (e.g. sentiment, language, topic)
- we can use words as features

Words as features

Convert to binary values, either present or not present:

	this	is	great	!	cool	<3	boring	a	waste	of	time	movie	just
this is great !	1	1	1	1									
cool <3					1	1							
a waste of time								1	1	1	1		
great movie			1									1	
boring just boring							1						1

Problems:

- Context
- Unknown words (commonly converted to [UNK])
- Tokenization (leads to sparsity, unknown words)
- Ambiguity

Lab: POS tagging

- You will annotate data yourself and compare to the annotation of a peer
- Next week, we will implement a POS tagger and use this data!
- There is Danish social media data, and an English alternative set in the repo

We will do a mini-experiment:

- I will also annotate the Danish data
- I will also train a POS-tagger
- We will see what is more important: linguistic knowledge (you) or the machine learning techniques (me)

Questions?

`www.menti.com/` 7727 3458