Sequence Labeling (I)

Brendan O'Connor
College of Information and Computer Sciences
University of Massachusetts Amherst
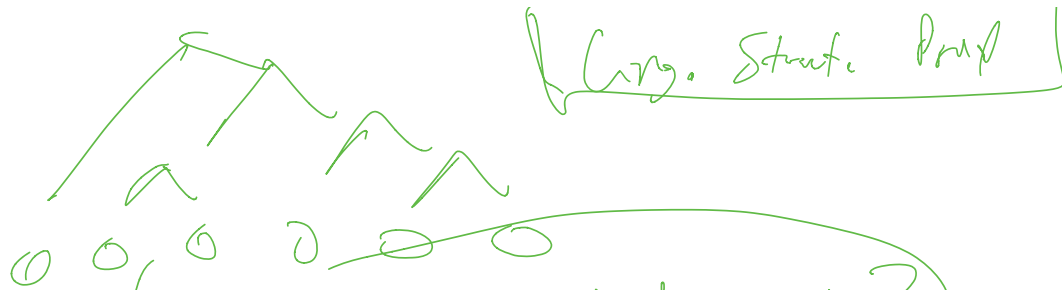
Lang. Struct. Pred.

Linguistics: What structs?
Psychology

Comp. Theory:
Automata &
Formal Langs.

Machine Learning
Struct. Pred.

- Sequence labeling tasks
- Latent variable Markovian models
  - Today: Hidden Markov model
  - Wed: Conditional Random Fields
-

CS 688
PGMS

- Sequence labeling: from $x_1..x_n$, predict tags $y_1..y_n$
- Named entity recognition:
an example of *span recognition*
  - BIO tags allow treatment as a sequence labeling problem

Part-of-Speech:

| NNP | NNP | NNP | NNP | , | DT | RB | JJ | JJ | NN | , | VBZ | VBN | IN | CD | IN | DT | NNP | POS | VBG |

1 Lt. Gen. H.R. McMaster, a widely respected military strategist, is seen as one of the Army 's leading

| NNS | . |

intellectuals.

Named Entity Recognition:

Person   Num   Org

1 Lt. Gen. H.R. McMaster, a widely respected military strategist, is seen as one of the Army's leading intellectuals.

O O B-PER I-PER O O   O   O   . . . . B-NUM

PER(2:4)

NUM(10:11)

3   http://nlp.stanford.edu:8080/corenlp/process

More span labeling tasks

- Syntactic chunking

encode as BIO

> "People tend to see risk primarily on that one dimension," says Timothy Kochis, natio

> But therein lies another aspect of investment risk : the hazard of shaping your portfo

> This is clearly not good news to all you people who sleep like babies every night, lull

> Risk wears many disguises, and investments that are low in one type of obvious ris
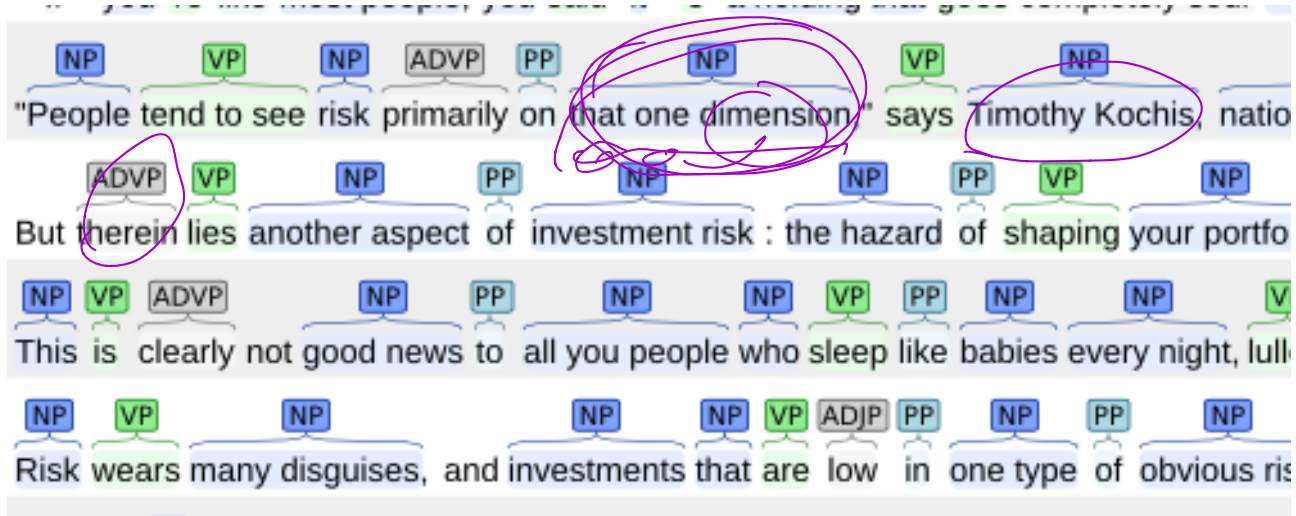
Next Week: Unreached syn spans!
Constituency Tree

http://brat.nlplab.org/examples.html

- Biological entities

Characterization of undifferentiated [spc human] [cell type ES cells] and differentiated EBs [anat] by antibodiesAll monoclon initially selected for their abilities to recognize recombinant proteins in direct ELISAs.
A subset were also tested by Western Blot analysis using recombinant proteins and cell lysate to con epitope.
The best clone was later screened for its applications for immunocytochemistry and flow cytometry us

[spc Human] [anatomy peripheral blood] [component platelets] were used for screening [spc mouse] [spc anti-human] [gene CD9] antibody.

[c line MCF-7] cells were used for screening [spc mouse] [spc anti-human] [gene E-Cadherin] and [gene PODXL] [gene or protein (podocalyxin-like)] an

[c line MG-63] cells were used for screening [spc mouse] [spc anti-human] [gene GATA1] [gene or protein (GATA binding protein 1)] antibody.

http://brat.nlplab.org/examples.html

What's a part-of-speech (POS)?

[handwritten: Grammar   Agreement   Word Order ...]

- Syntax = how words compose to form larger meaning-bearing units
- POS = syntactic categories for words
  - Approximately: You could substitute words within a class and have a syntactically valid sentence.
  - Give information how words can combine.

[handwritten: Bad subst: I saw the green   ☆ I saw the of.]
[handwritten: ADJ]
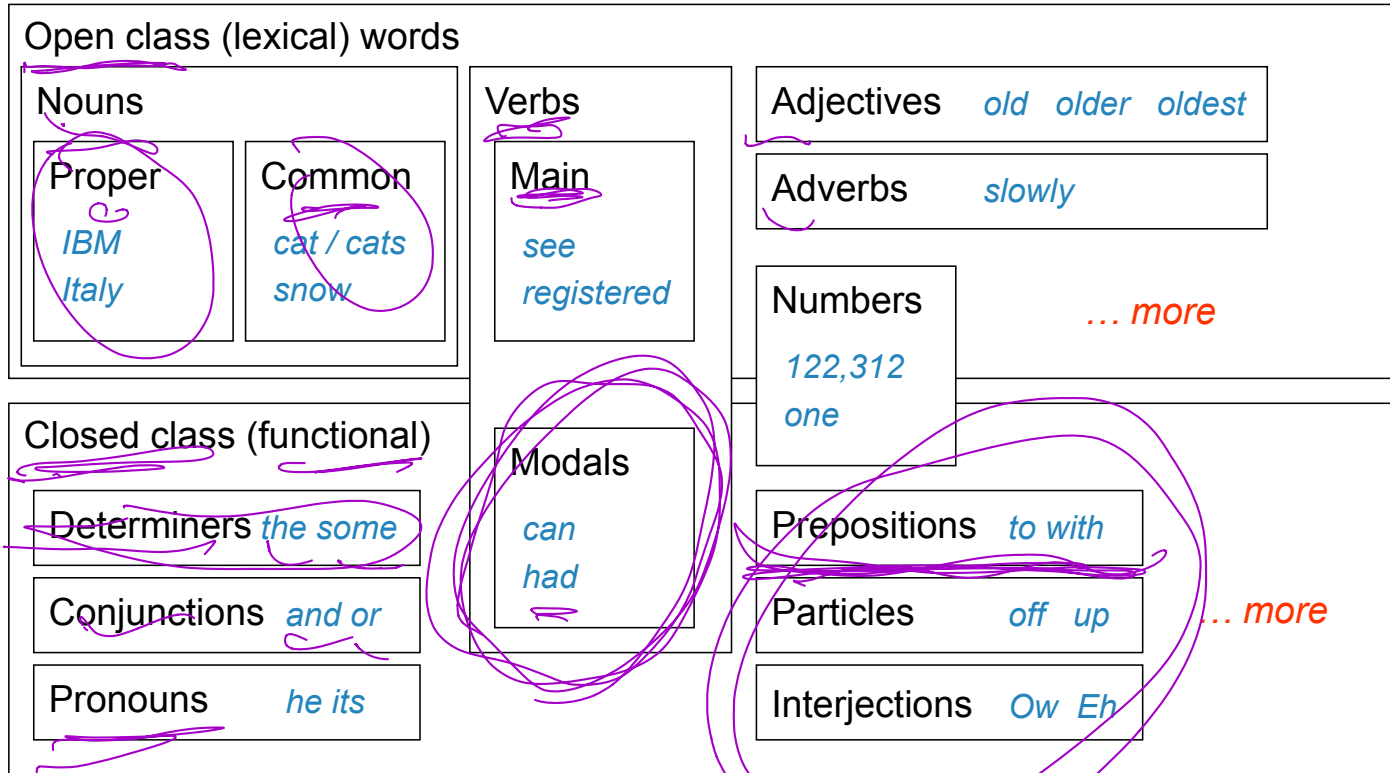*[handwritten: five   of   Prep]*

  - I saw the dog
  - I saw the cat
  - I saw the {table, sky, dream, school, anger, ...}

- (Phrasal/constituent categories generalize this idea. POS tags are constrained to single words.)

Open vs closed classes

Content

Open class (lexical) words

Nouns

Proper

IBM

Italy

Common

cat / cats

snow

Verbs

Main

see

registered

Modals

can

had

Adjectives *old older oldest*

Adverbs *slowly*

Numbers

122,312

one

… more

Closed class (functional)

Determiners *the some*

Conjunctions *and or*

Pronouns *he its*

Prepositions *to with*

Particles *off up*

Interjections *Ow Eh*

… more

higher freq

Why do we want POS?

- Useful for many syntactic and other NLP tasks.
  - Phrase identification ("chunking")
  - Named entity recognition
  - Full parsing
  - Sentiment
- Especially when there's a low amount of training data
- Rule-based methods to assemble candidate phrases for later downstream processing

POS patterns: sentiment

- Turney (2002): identify bigram phrases, from unlabeled corpus, useful for sentiment analysis.

Table 1. Patterns of tags for extracting two-word phrases from reviews.

	First Word	Second Word	Third Word (Not Extracted)
1.	JJ	NN or NNS	anything
2.	RB, RBR, or RBS	JJ	not NN nor NNS
3.	JJ	JJ	not NN nor NNS
4.	NN or NNS	JJ	not NN nor NNS
5.	RB, RBR, or RBS	VB, VBD, VBN, or VBG	anything

(plus PMI to sentiment seed words)

POS patterns: sentiment

- Turney (2002): identify bigram phrases, from unlabeled corpus, useful for sentiment analysis.

Table 1. Patterns of tags for extracting two-word phrases from reviews.

	First Word	Second Word	Third Word (Not Extracted)
1.	JJ	NN or NNS	anything
2.	RB, RBR, or RBS	JJ	not NN nor NNS
3.	JJ	JJ	not NN nor NNS
4.	NN or NNS	JJ	not NN nor NNS
5.	RB, RBR, or RBS	VB, VBD, VBN, or VBG	anything

PTB tagset

Table 2. An example of the processing of a review that the author has classified as *recommended*.[6]

Extracted Phrase	Part-of-Speech Tags	Semantic Orientation
online experience	JJ NN	2.253
low fees	JJ NNS	0.333
local branch	JJ NN	0.421
small part	JJ NN	0.053
online service	JJ NN	2.780
printable version	JJ NN	-0.705
direct deposit	JJ NN	1.288
well other	RB JJ	0.237
inconveniently located	RB VBN	-1.541
other bank	JJ NN	-0.850
true service	JJ NN	-0.732

(plus PMI to sentiment seed words)

9

POS patterns: simple noun phrases

- Quick and dirty noun phrase identification (Justeson and Katz 1995, Handler et al. 2016)
    - BaseNP = (Adj | Noun)* Noun
    - PP = Prep Det* BaseNP
    - NP = BaseNP PP*

Kleene Star

Regular Expression

Grammatical structure: Candidate strings are those multi-word noun phrases that are specified by the regular expression $((A \mid N)^+ \mid ((A \mid N)^*(NP)^?)(A \mid N)^*)N$,

Tag Pattern	Example
A N	*linear function*
N N	*regression coefficients*
A A N	*Gaussian random variable*
A N N	*cumulative distribution function*
N A N	*mean squared error*
N N N	*class probability function*
N P N	*degrees of freedom*

Table 5.2 Part of speech tag patterns for collocation filtering. These patterns were used by Justeson and Katz to identify likely collocations among frequently occurring word sequences.

Congressional bills
(Top terms, ranked by relative log-odds z-scores)

Uni. Dem.	and, deleted, health, mental, domestic, inserting, grant, programs, prevention, violence, program, striking, education, forensic, standards, juvenile, grants, partner, science, research
Uni. Rep.	any, offense, property, imprisoned, whoever, person, more, alien, knowingly, officer, not, united, intent, commerce, communication, forfeiture, immigration, official, interstate, subchapter
NPs Dem.	
NPs Rep.	

Congressional bills
(Top terms, ranked by relative log-odds z-scores)

Uni. Dem.	and, deleted, health, mental, domestic, inserting, grant, programs, prevention, violence, program, striking, education, forensic, standards, juvenile, grants, partner, science, research
Uni. Rep.	any, offense, property, imprisoned, whoever, person, more, alien, knowingly, officer, not, united, intent, commerce, communication, forfeiture, immigration, official, interstate, subchapter
NPs Dem.	mental health, juvenile justice and delinquency prevention act, victims of domestic violence, child support enforcement act of u.s.c., fiscal year, child abuse prevention and treatment act, omnibus crime control and safe streets act of u.s.c., date of enactment of this act, violence prevention, director of the national institute, former spouse, section of the foreign intelligence surveillance act of u.s.c., justice system, substance abuse criminal street gang, such youth, forensic science, authorization of appropriations, grant program
NPs Rep.	

Congressional bills
(Top terms, ranked by relative log-odds z-scores)

|	
Uni. Dem.	and, deleted, health, mental, domestic, inserting, grant, programs, prevention, violence, program, striking, education, forensic, standards, juvenile, grants, partner, science, research
Uni. Rep.	any, offense, property, imprisoned, whoever, person, more, alien, knowingly, officer, not, united, intent, commerce, communication, forfeiture, immigration, official, interstate, subchapter
NPs Dem.	mental health, juvenile justice and delinquency prevention act, victims of domestic violence, child support enforcement act of u.s.c., fiscal year, child abuse prevention and treatment act, omnibus crime control and safe streets act of u.s.c., date of enactment of this act, violence prevention, director of the national institute, former spouse, section of the foreign intelligence surveillance act of u.s.c., justice system, substance abuse criminal street gang, such youth, forensic science, authorization of appropriations, grant program
NPs Rep.	special maritime and territorial jurisdiction of the united states, interstate or foreign commerce, federal prison, section of the immigration and nationality act, electronic communication service provider, motor vehicles, such persons, serious bodily injury, controlled substances act, department or agency, one year, political subdivision of a state, civil action, section of the immigration and nationality act u.s.c., offense under this section, five years, bureau of prisons, foreign government, explosive materials, other person

How to build a POS tagger?

- Sources of information:

    - POS tags of surrounding words: syntactic context

    - The word itself

    - Features, etc.!
        - Word-internal information
        - Features from surrounding words
        - External lexicons
        - Embeddings, NN states

←HMM

←CRF
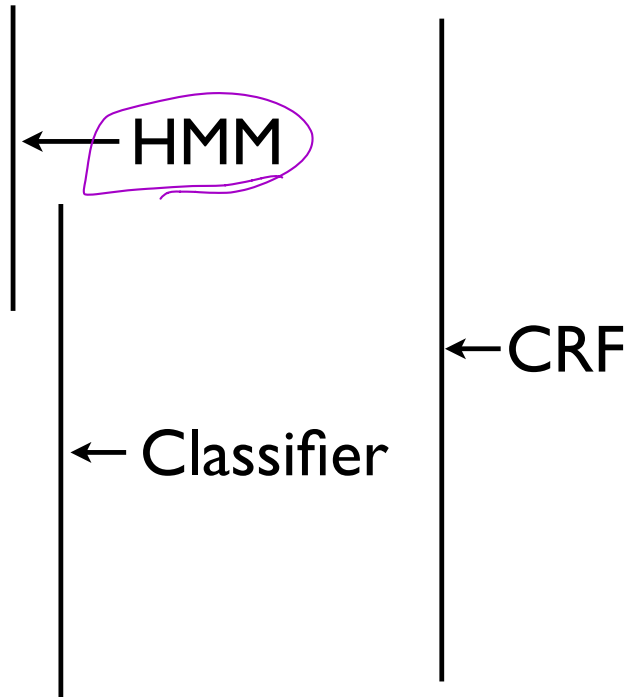
← Classifier

[BERT/ELMO may be sufficient alternatives to sharing contextual information?]

Sequence labeling

- Seq. labeling as classification:
  Each position *m* gets an independent classification, as a log-linear model.

$$\overset{\text{argmax}}{\scriptstyle y_m} \quad p(y_m \mid w_1 .. w_n)$$

$$\arg\max_y \theta^{\mathsf{T}} \mathbf{f}((\mathbf{w}, m), y)$$

proposed output tag at m

$$\begin{aligned}
\boldsymbol{f}((\boldsymbol{w} = they\ can\ fish, m = 1), \mathrm{N}) &= \langle they, \mathrm{N} \rangle \\
\boldsymbol{f}((\boldsymbol{w} = they\ can\ fish, m = 2), \mathrm{V}) &= \langle can, \mathrm{V} \rangle \\
\boldsymbol{f}((\boldsymbol{w} = they\ can\ fish, m = 3), \mathrm{V}) &= \langle fish, \mathrm{V} \rangle.
\end{aligned}$$

(extract tag features

$$\theta_{\langle they, N \rangle} = -2.2$$

Sequence labeling

- Seq. labeling as classification:
  Each position *m* gets an independent classification, as a log-linear model.

$$p(y_m \mid w_1..w_n)$$

$$\arg\max_y \theta^{\mathsf{T}} \mathbf{f}((\mathbf{w}, m), y)$$

$\mathbf{f}((\mathbf{w} = \textit{they can fish}, m = 1), \mathrm{N}) = \langle \textit{they}, \mathrm{N} \rangle$
$\mathbf{f}((\mathbf{w} = \textit{they can fish}, m = 2), \mathrm{V}) = \langle \textit{can}, \mathrm{V} \rangle$
$\mathbf{f}((\mathbf{w} = \textit{they can fish}, m = 3), \mathrm{V}) = \langle \textit{fish}, \mathrm{V} \rangle.$

- But syntactic (tag) context is sometimes necessary!

$$y_0 \rightarrow y_1 \rightarrow y_2 \rightarrow y_3 \rightarrow y_4 = \text{STOP}$$

$$w_1 \quad w_2 \quad w_3$$

$$P\left(\vec{w}, \vec{y}\right) = \prod_{t=1}^{M} P(y_t \mid y_{t-1}) \cdot P(w_t \mid y_t)$$

"Hidden"
Latent R.V.s
"H.V.s"

- Seq. labeling as log-linear ***structured prediction***

$$\hat{\boldsymbol{y}}_{1:M} = \operatorname*{argmax}_{\boldsymbol{y}_{1:M} \in \mathcal{Y}(\boldsymbol{w}_{1:M})} \boldsymbol{\theta}^\top \boldsymbol{f}(\boldsymbol{w}_{1:M}, \boldsymbol{y}_{1:M}),$$

- Example: the ***Hidden Markov model***

$$p(\mathbf{w}, \mathbf{y}) = \prod_t p(y_t \mid y_{t-1}) p(w_t \mid y_t)$$

- Efficiently supports operations via dynamic programming because of **local (Markovian) assumptions**
  - P(w): Likelihood (generative model)
    - Forward algorithm
  - P(y | w): Predicted sequence ("decoding")
    - Viterbi algorithm
  - P(y_m | w): Predicted tag marginals
    - Forward-Backward algorithm
    - Supports EM for unsupervised HMM learning

(handwritten annotations)

$y_1 \quad y_2 \quad y_3$

they can fish

M tokens

K tags

$|\mathcal{Y}(w_{1:M})| = M^K$

$\mathcal{Y}(w_{1:M}) = \{$

$(N,N,N),$

$(N,N,A),$

$\dots \}$

$\sum_y P(w,y)$