Coreference evaluation

- Ideally, downstream tasks
- Direct coref evaluation: compare mention clusterings

```
gold
[Harry Potter]1 was a [wizard]1.
[Voldemort]2 shook [his]1 hand.
```

```
prediction examples
[Harry Potter]A was a [wizard]B.
```

```
[Harry Potter]A was a [wizard]A.
```

```
[Harry Potter]A was a [wizard]A.
```

```
[Harry Potter]A was a [wizard]A.
```

- Two possible metrics
  - 1. Pairwise P/R/F
  - 2. B-cubed P/R/F
Pairwise prec/rec

- Entity clustering as a classification task among all mention pairs

\[
\text{prec} = \frac{\text{Num TP links}}{\text{Num pred links}} \quad \text{rec} = \frac{\text{Num TP links}}{\text{Num gold links}}
\]

Gold entity 1  Gold entity 2

- \( \text{prec} = \frac{4}{4} \quad \text{rec} = \frac{4}{4} \)
- \( \text{prec} = \frac{4}{10} \quad \text{rec} = \frac{4}{4} \)
- \( \text{prec} = \frac{2}{5} \quad \text{rec} = \frac{2}{4} \)
- \( \text{prec} = \frac{0}{0} = 100\% \quad \text{rec} = \frac{0}{4} \)
**B³ prec/rec**

- Mention-averaged versions
  - Prec: pick a mention at random.
  - Expected proportion of predicted coreferents are actually coreferent?

Gold entity 1  Gold entity 2

\[
\text{prec} = \frac{3/3 + 3/3 + 3/3 + 2/2 + 2/2}{5} \quad \text{rec} = \frac{3/3 + 3/3 + 3/3 + 2/2 + 2/2}{5}
\]

\[
\text{prec} = \frac{3/5 + 3/5 + 3/5 + 2/5 + 2/5}{5} \quad \text{rec} = \frac{3/3 + 3/3 + 3/3 + 2/2 + 2/2}{5}
\]

\[
\text{prec} = \frac{2/2 + 2/2 + 1/3 + 2/3 + 2/3}{5} \quad \text{rec} = \frac{2/3 + 2/3 + 1/3 + 1/2 + 1/2}{5}
\]

\[
\text{prec} = \frac{1/1 + 1/1 + 1/1 + 1/1 + 1/1 + 1/1}{5} \quad \text{rec} = \frac{1/3 + 1/3 + 1/3 + 1/2 + 1/2}{5}
\]
Precision/Recall tradeoffs

- F-score: a very common but somewhat meaningless way to combine precision and recall into one number
  - My prec went up but rec went down. Was it a good change? Without downstream eval, can only answer this in an arbitrary way.
  - Uses a harmonic mean, so if one number is really low, the whole thing is bad

- Making P/R tradeoffs
  - Rule-based system: Match on anything? Only match if gender and number match?
  - Machine-learning system: say confidence must be at least $t$ in order to select a match. Tune $t$?

\[ F = \frac{2 \times P \times R}{P + R} \]
Semantics and Lexical semantics
Levels of meaning

• 1. Lexical semantics
  • The meaning of words

• 2. Compositional semantics
  • How do word meanings combine to create the meaning of sentences/utterances?

• 3. Discourse/pragmatics
  • How do meanings combine with context to create meaning for an entire text / conversation?

• Entities / coreference: a discourse problem over relatively “simple” concepts (entities...)
Word senses

• What about word meanings that aren’t specific entities?
• A single word (word form) can have different senses.
  • I went to my bank today.
  • I saw the bank of a river.
• A language dictionary identify these as different senses, e.g. bank#1, bank#2
• What is a word meaning?
  One arguable definition:
  a reference to a set of things in the world
Word sense relations

• Synonymy: bidirectional entailment under substitution (in simple contexts...)
  • I drank cocoa \iff I drank hot chocolate

• Hypernymy: directional entailment under substitution (in simple contexts...)
  • I went to my bank \implies I went to my financial institution

• These relations hold at the sense, not word, level
  • I saw the river bank \neq \implies I saw the river financial institution
Taxonomies

- Hierarchies of concepts or word meanings
  - For a class, a list of things within that class
- Simple example: word/concept lists
  - List of all baseball teams in the US
  - List of all sports teams in the US
Taxonomy: complex example

CLASSIFICATION OF ANIMALS

This is the grouping together of animals with similar characteristics. Animals can be classed as either vertebrates or invertebrates.

VERTEBRATES

These are animals that have a backbone.

Reptiles
- Have dry scaly skin.
- Lay eggs on dry land.
- Are cold blooded.
  (Snake, Crocodile)

Fish
- Have scales on their bodies.
- Have gills for breathing.
- Are cold blooded.
  (Shark, Tuna)

Amphibians
- Have moist slimy skin.
- Lay eggs in water.
- Are cold blooded.
  (Frog, Newt)

Birds
- Have feathers and wings.
- Have beaks and lay eggs.
- Are warm blooded.
  (Wren, Swan)

Mammals
- Have fur or hair.
- Feed young on milk.
- Are warm blooded.
  (Cow, Human)

INVERTEBRATES

These are animals that do not have a backbone.

Protozoa
- Single cell organisms
  all microscopic.

Flatworms
- Simple and soft bodied.
  (Tape worm, Flukes)

Annelid Worms
- Segmented bodies.
  (Earthworm, Leech)

Echinoderms
- Spiny sea creatures.
  (Starfish, Sea urchin)

Coelenterates
- Hard external skeleton
  and jointed limbs.

Arthropods
- Soft bodied, stinging cells.
  (Jellyfish, Sea anemone)

Molluscs
- Soft bodied, most have shells.
  (Snails, Limpet)

Arachnids
- Eight legs, two body parts,
  no antennae.
  (Spider, Scorpion)

Crustaceans
- Mostly sea creatures.
  Many legs and two sets of antennae.
  (Crab, Lobster)

Insects
- Wings, six legs, three body parts, one pair of antennae.
  (Bee, Ladybird)

Myriapods
- Many legs and body segments.
  (Centipede, Millipede)
Taxonomy: complex example
Wordnet

- One of the biggest lexical resources out there.
- Hand-curated database of word senses for English.
- Each concept ("synset") has
  - A set of words it corresponds to (one-many relationship)
    - A word/lexeme is a (lemma, POS) pair
    - A wordsense is a (lemma, POS, concept) tuple
  - Synonym undirected links b/w concepts
  - Hypernym/Hyponym directed links b/w concepts
- Hyponyms are a DAG (not quite a tree)
Wordnet

<table>
<thead>
<tr>
<th>POS</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun</td>
<td>117,097</td>
</tr>
<tr>
<td>Adjective</td>
<td>22,141</td>
</tr>
<tr>
<td>Verb</td>
<td>11,488</td>
</tr>
<tr>
<td>Adverb</td>
<td>4,601</td>
</tr>
</tbody>
</table>

http://wordnetweb.princeton.edu/perl/webwn
What’s Wordnet good for?

- Going beyond individual words to more general meanings
  - Especially helpful if you can’t use supervised ML (no or little training data)
- But WN doesn’t always have the right granularity or coverage -- it’s hand-built, for better or worse
- WN trick: only use most-frequent-sense
Hyponyms of “person”
7588 total -- with MFS restriction. [from Michael Heilman]

- vintager
- matrisib
- horseback rider
- ceo
- seeker
- fieldhand
- radiologist
- captain
- moujik
- research director
- damsels
- nibbler
- nailer
- nude person
- seismologist
- oddball
- prankster
- radiotherapist
- nebraskan
- cupbearer
- psychic
- accompanist
- plagiariser
- timberman
- photographer's model
- lombard
- debaser
- courtier
- dutch uncle
- schlemiel
- dizygotic twin
- mental case
- matriarch
- vocalist
- internist
- transplanter
- techie
- sniffler
- marrano
- first baseman
- government man
- child prodigy
- athenian
- hospital chaplain
- dominatrix
- bibliopole
- hombre
- east indian
- ballet master
- bad person
- rock 'n' roll musician
- flack catcher
- telephonner
- dominus
- cheater
- groveler
- accomplice
- herb doctor
- schoolfriend
- preteen
- gastronome
- concierge
- shogun
- flutist
- bottom dog
- imperialist
- emir
- libeler
- manichaean
- abnegator
- cousin-german
- masorite
- trouble maker
- villainess
- rajpoot
- calapooya
- overlord
- bank guard
- tumbler
- polycarp
- radiographer
- slave owner
- stick-in-the-mud
- audile
- deadbeat
- maltman
- jeweler
Word sense disambiguation

- Say you have a concept database -- has word forms paired with concept entries
- Given “I saw the bank”, is it bank#1 or bank#2?
  - “Most frequent sense” baseline
- Given text “Michael Jordan”, is it...

- For an entity DB, this is called “entity linking”
  A form of cross-document coreference
- In all cases: context features are critical
From hierarchy to attributes

• Can a single hierarchy capture all semantic attributes?

• Borges’ “Celestial Emporium of Benevolent Knowledge”
  Animals are divided into ...
  • (a) those that belong to the Emperor,
  • (b) embalmed ones,
  • (c) those that are trained,
  • (d) suckling pigs,
  • (e) mermaids,
  • (f) fabulous ones,
  • (g) stray dogs,
  • (h) those that are included in this classification,
  • (i) those that tremble as if they were mad,
  • (j) innumerable ones,
  • (k) those drawn with a very fine camel brush,
  • (l) others,
  • (m) those that have just broken a flower vase,
  • (n) those that resemble flies from a distance.