Information Extraction (2)

Instructor: Thien Huu Nguyen

Based on slides from: Ralph Grishman
Reference/Coreference Resolution: Objective

- Identify all phrases which refer to the same real-word entity
  - first, within a single document (we will only focus on this in our class)
  - later, also across multiple documents
Terminology

*referent*: real-world object referred to

*referring expression* [mention]: a phrase referring to that object

Mary was hungry; she ate a banana.
Terminology

• *coreference*: two expressions referring to the same thing

   Mary was hungry; she ate a banana.

   antecedent  anaphor
   (prior expression) (following expression)

• So we also refer to process as *anaphora resolution*
Barack Hussein Obama II (born August 4, 1961) is the 44th and current President of the United States, and the first African American to hold the office. Born in Honolulu, Hawaii, Obama is a graduate of Columbia University and Harvard Law School, where he served as president of the Harvard Law Review. He was a community organizer in Chicago before earning his law degree. He worked as a civil rights attorney and taught constitutional law at the University of Chicago Law School from 1992 to 2004. He served three terms representing the 13th District in the Illinois Senate from 1997 to 2004, running unsuccessfully for the United States House of Representatives in 2000.
Types of referring expressions

- definite pronouns (he, she, it, ...)
- indefinite pronouns (one)
- definite NPs (the car)
- indefinite NPs (a car)
- names
Referring Expressions: pronouns

Definite pronouns: he, she, it, ...

• generally anaphoric
  – Mary was hungry; she ate a banana

• pleonastic (non-referring) pronouns
  – It is raining.
  – It is unlikely that he will come.

• pronouns can represent bound variables in quantified contexts:
  – Every lion finished its meal.
Referring Expressions: pronouns

Indefinite pronouns (one)

• refers to another entity with the same properties as the antecedent
  – Mary bought an iPhone 6.
  – Fred bought one too.
  – *Fred bought it too.

• can be modified
  – Mary bought a new red convertible.
  – Fred bought a used one.
    = a used red convertible
  (retain modifiers on antecedent which are compatible with those on anaphor)
Referring Expressions: pronouns

Reflexive pronouns (himself, herself, itself)

- used if antecedent is in same clause
  - I saw myself in the mirror.
Referring expressions: NPs

NPs with definite determiners ("the")

• reference to uniquely identifiable entity

• generally anaphoric
  – I bought a Ford Fiesta. The car is terrific.

• but may refer to a uniquely identifiable common noun
  – I looked at the moon
  – The president announced ...

• or a functional result
  – The sum of 4 and 5 is 9.
  – The price of gold rose by $4.
Referring expressions: NPs

NPs with indefinite determiners ("a")

- generally introduces a new ‘discourse entity’
- may also be generic:
  - A giraffe has a long neck.
Referring expressions: names

• subsequent references can use portions of name:

  – Fred Frumble and his wife Mary bought a house. Fred put up a hammock.
Complications

• Cataphora: pronoun referring to a following mention:
  – When _she_ entered the room, Mary looked around.

• Bridging anaphora: reference to related object
  – Entering the room, Mary looked at the _ceiling_.

• Zero anaphora: many languages allow subject omission, and some allow omission of other arguments (e.g., Japanese)
  – these can be treated as zero (implicit) anaphors
    • similar resolution procedures
  – some cases of bridging anaphora can be described in terms of PPs with zero anaphors:
    • IBM announced the appointment as Fred as president

• Non-NP anaphora: Pronouns can also refer to events or propositions:
  – Fred claimed that no one programs in Lisp. _That_ is ridiculous.

• Conjunctions and collective reference
Conjunctions and collective reference

• With a conjoined NP,
  ... Fred and Mary ...
we can refer to an individual (“he”, “she”) or the conjoined set (“the”)

• We can even refer to the collective set if not conjoined ...

  “Fred met Mary after work. They went to the movies.”
Resolving Pronoun Reference

• Constraints
• Preferences
• Hobbs Search
• Selectional preferences
• Combining factors
Pronouns: constraints

Pronoun must agree with antecedent in:

• animacy
  – We’re watching a movie. He likes it [*he = you and I]

• gender
  – Mary met Mr. and Mrs. Jones. She was wearing orange pants.
  – needs first-name dictionary
  – some nouns gender-specific: sister, ballerina

• number
  – some syntactically singular nouns can be referred to by a plural pronoun: “The platoon ... they”
Pronouns: preferences

Prefer antecedents that are

- recent
  - at most 3 sentences back

- salient
  - mentioned several times recently

- subjects

Recency and preference for subjects are often captured by Hobbs search order, a particular order for searching the current and preceding parse trees
Hobbs search order

- traverse parse tree containing anaphor, starting from anaphor
- traverse trees for preceding sentences, breadth first, left-to-right
  - incorporates subject precedence
- stop at first NP satisfying constraints
  - relatively simple strategy, competitive performance
Pronouns: selectional preferences

• Prefer antecedent that is more likely to occur in context of pronoun
  – Fred got a book and a coffee machine for his birthday. He read it the next day.
  – can get probabilities from a large (parsed) corpus
Pronouns: combining probabilities

\[ P = P ( \text{correct antecedent is at Hobbs distance } d) \times \]
\[ P (\text{pronoun} \mid \text{head of antecedent}) \times \]
\[ P (\text{antecedent} \mid \text{mention count}) \times \]
\[ P (\text{head of antecedent} \mid \text{context of pronoun}) \]

Ge, Hale, and Charniak 1998

83% success
Resolving names

• Generally straightforward: exact match or subsequence of prior name
  – some exceptions for locations
Resolving common noun phrases

• generally difficult

• typical strategies for resolving “the” + N:
  – look for prior NP with same head N
  – look for prior name including token N
    • “the New York Supreme Court” ... the court

• more ambitious: learn nouns used to refer to particular entities by searching for “name, N” patterns in a large corpus
  – “Lazard Freres, the merchant bank”
Types of models

• mention-pair model
  – train binary classifier: are two mentions coreferential?
  – to apply model:
    • scan mention in text order
      – link each mention to the closest antecedent classified +
      – link each mention to antecedent most confidently labeled +
    • cluster mentions
  – weak model of partially-resolved coreference

• entity-mention model
  – binary classifier: is a mention part of a partially-formed entity?
  – richer model: entity has features from constituent mentions
Diversity of approaches

Two recent systems show range of approaches:

• Stanford [CL 2013]
  – hand-coded rules
  – 10 passes over complete document, using rules of decreasing certainty

• Berkeley [EMNLP 2013]
  – classifier trained over large corpus with simple feature set
  – single pass

• No system does well on anaphoric NPs
Sieve-based, hand-coded system (Stanford)

- **sieve**: set of hand-coded rules, applied starting with the most precise rule
- Each rule is applied across the entire document (total of 10 passes)
- Rules reflect detailed linguistic analysis
- Most rules involve nominal anaphors; final pass (pass 10) resolves pronouns using agreement constraints
- Entity-centric model, uses information from all mentions gathered so far

All NPs, possessive pronouns, and named entity mentions are candidate mentions. Recall is more important than precision.

Lee et al., 2013: Deterministic Coreference Resolution Based on Entity-Centric, Precision-Ranked Rules
Shallow feature statistical system (Berkeley)

• statistical approach based on large annotated corpus (OntoNotes)

• mention-synchronous: single pass through document

• features make minimal reference to specific linguistic phenomena
  – large training corpus enables simple rules to capture most constraints

• Anaphoric nominals remain the weak point for all approaches. Durrett and Klein report that when an anaphoric mention is a nominal or name, their system identifies the proper antecedent less than 8% of the time.
Features for the mention-pair models

- **Unary features (valid of a single token)**
  - Token, lemma, part of speech
  - Salience

- **Binary features (valid of a pair of tokens)**
  - Number agreement (plural pronoun/plural NP)
  - Gender agreement
  - Sentence distance
  - Hobbs distance
  - Syntax: grammatical role
  - ...

Evaluation

- Coreference key is a set of links dividing the set of mentions into coreference classes
- System response has similar structure
- How to score response?
- MUC scorer
  - based on links ...
    recall error = how many links must be added to system response so that all members of a key set are connected by links
  - Does not give credit for correct singleton sets
Evaluation

• B-cubed metric:

  – Mention-based
  – For each mention m,
    \( r = \text{size of response set containing m} \)
    \( k = \text{size of key set containing m} \)
    \( i = \text{size of intersection of these sets} \)
    \( \text{recall}(m) = \frac{i}{k} \)
    \( \text{precision}(m) = \frac{i}{r} \)
  – Then compute average of recall, average of precision
Example

• Golden: 3 entity/coreference chains
  1. \{I, you, you, your, me, your, your, You\} (8 elements)
  2. \{you, your father, you, him, I, your father\} (6 elements)
  3. \{Obi-Wan, He\} (2 elements)
Example

- System output: 4 entity/coreference chains
  1. \{I, me, l\} (3 elements)
  2. \{you, you, you, your, you, your, your, you\} (8 elements)
  3. \{Obi-Wan, your father, your father\} (3 elements)
  4. \{He, him\}
Example

LUKE
I'll never join you!

VADER
If you only knew the power of the dark side. Obi-Wan never told you what happened to your father.

LUKE
He told me enough! It was you who killed him.

VADER
No. I am your father.

LUKE
No. No. That's not true! That's impossible!

VADER
Search your feelings. You know it to be true.

LUKE
No! No! No!

<table>
<thead>
<tr>
<th>Goldₖ ∩ Systemₖ</th>
<th>= 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goldₖ</td>
<td>= 8</td>
</tr>
<tr>
<td>Systemₖ</td>
<td>= 3</td>
</tr>
</tbody>
</table>
Example

I'll never join you!

LuKe

If you only knew the power of the dark side. Obi-Wan never told you what happened to your father.

Vader

He told me enough! It was you who killed him.

LuKe

No. I am your father.

Vader

No. No. That's not true! That's impossible!

LuKe

Search your feelings. You know it to be true.

Vader

No! No! No!

LuKe

I'll never join you!

Vader

If you only knew the power of the dark side. Obi-Wan never told you what happened to your father.

LuKe

He told me enough! It was you who killed him.

Vader

No. I am your father.

LuKe

No. No. That's not true! That's impossible!

Vader

Search your feelings. You know it to be true.

LuKe

No! No! No!

| Gold_i ∩ System_i | = 2 | | Gold_i | = 6 | | System_i | = 8 |
A Coherent Discourse

- A text is not a random collection of facts
- A text will tell a story, make an argument, ...
- This is reflected in the structure of the text and the connections between sentences
- Most of these connections are implicit, but a text without these connections is *incoherent*

Fred took an NLP course in the Spring.
He got a great job in June.

? Fred took an NLP course in the Spring.
He got a great cat in June.
A Coherent Discourse

• Criteria for coherence depend on type of text

• Most intensively studied for narratives
  – causal connections
  – temporal connections
  – scripts (conventional sequences)
Coherence and coreference

• Select anaphora resolution more consistent with coherence.
  Jack poisoned Sam. He died within a week. vs.
  Jack poisoned Sam. He was arrested within a week.

• How to do this in practice?
  – Collect from a large corpus a set of predicate/role pairs, such as:
    subject of poison -- subject of arrest
    object of poison -- subject of die.
  – Prefer anaphora resolution consistent with such pairs
Words and Senses

• Until now we have manipulated structures based on words.

• But if we are really interested in the meaning of sentences, we must consider the *senses* of words:
  • most words have several senses
  • frequently several words share a common sense
  • both are important for information extraction

• A word sense is a representation of one aspect of a word’s meaning.
Word senses

I’m going to the bank

- bank_1 = “financial institution”
- bank_2 = “sloping mound”
- bank_3 = “biological repository”
- bank_4 = “building where a bank_1 does its business”
Word senses

Verb

- **serve, function** (serve a purpose, role, or function) "The tree stump serves as a table"; "The female students served as a control group"; "This table would serve very well"; "His freedom served him well"; "The table functions as a desk"
- **serve** (do duty or hold offices; serve in a specific function) "He served as head of the department for three years"; "She served in Congress for two terms"
- **serve** (contribute or conduce to) "The scandal served to increase his popularity"
- **service, serve** (be used by; as of a utility) "The sewage plant served the neighboring communities"; "The garage served to shelter his horses"
- **serve** (help to some food; help with food or drink) "I served him three times, and after that he helped himself"
- **serve, serve up, dish out, dish up, dish** (provide (usually but not necessarily food)) "We serve meals for the homeless"; "She dished out the soup at 8 P.M."; "The entertainers served up a lively show"
- **serve** (devote (part of) one’s life or efforts to, as of countries, institutions, or ideas) "She served the art of music"; "He served the church"; "serve the country"
- **serve, serve well** (promote, benefit, or be useful or beneficial to) "Art serves commerce"; "Their interests are served"; "The lake serves recreation"; "The President’s wisdom has served the country well"
- **serve** (spend time in prison or in a labor camp) "He did six years for embezzlement"
- **serve, attend to, wait on, attend, assist** (work for or be a servant to) "May I serve you?"; "She attends the old lady in the wheelchair"; "Can you wait on our table, please?"; "Is a salesperson assisting you?"; "The minister served the King for many years"
- **serve, process, swear out** (deliver a warrant or summons to someone) "He was processed by the sheriff"
- **serve, do, answer, serve** (be sufficient; be adequate, either in quality or quantity) "A few words would answer"; "This car suits my purpose well"; "Will $100 do?"; "A ‘B’ grade doesn’t suffice to get me into medical school"; "Nothing else will serve"
- **serve** (do military service) "She served in Vietnam"; "My sons never served, because they are short-sighted"
- **serve, service** (mate with) "male animals serve the females for breeding purposes"
- **serve** (put the ball into play) "It was Agassi’s turn to serve"
Polysemy vs homophony

• Polysemy refers to phenomenon that one and the same word acquires different, though obviously related, meanings, often with respect to particular contexts.
  – The bank raised its interest rates yesterday.
  – The store is next to the newly constructed bank.
  – The bank appeared first in Italy in the Renaissance.

• Homophony refers to cases in which two words “accidentally” have the same phonological form
  – Mary walked along the bank of the river.
  – HarborBank is the richest bank in the city.
Zeugma (/ˈzoʊɡmə/)  

- Conjunction (“yoke”) of antagonistic readings; one test for whether word senses are distinct (often used intentionally to either confuse the reader or inspire them to think more deeply)

- The storm sank my boat.
- The storm sank my dreams.
- The storm sank my boat and my dreams.

- All over Ireland the farmers grew potatoes, barley, and bored.
Relationships between senses

• **Synonym**
  – Two senses of different words are synonyms of each other if their meaning is nearly identical.
  
  – Two words are never exactly the same in their meaning, distribution of use, dialect or other contexts in which they’re licensed.
  
  – Synonyms can be exchanged for each other without changing the truth conditions of a sentence.

<table>
<thead>
<tr>
<th>couch</th>
<th>sofa</th>
</tr>
</thead>
<tbody>
<tr>
<td>filbert</td>
<td>hazelnut</td>
</tr>
<tr>
<td>car</td>
<td>automobile</td>
</tr>
<tr>
<td>fair</td>
<td>impartial</td>
</tr>
<tr>
<td>fair</td>
<td>pale</td>
</tr>
</tbody>
</table>
Relationships between senses

- **Synonym**
  - Synonymy holds between word senses, not words
  - How *big* is that plane?
  - Would I be flying on a *large* or small plane?
  - Miss Nelson, for instance, became a kind of *big* sister to Benjamin
  - Miss Nelson, for instance, became a kind of *large* sister to Benjamin
Relationships between senses

- **Antonym (anotonymy)**
  - Two senses of different words are antonymous of each other if their meaning is nearly opposite
  - All aspects of meaning are nearly identical between antonyms, except one (very much like synonyms in this respect)

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>long</td>
<td>short</td>
<td>both describe length</td>
</tr>
<tr>
<td>big</td>
<td>little</td>
<td>both describe size</td>
</tr>
<tr>
<td>fast</td>
<td>slow</td>
<td>both describe speed</td>
</tr>
<tr>
<td>cold</td>
<td>hot</td>
<td>both describe temperature</td>
</tr>
<tr>
<td>dark</td>
<td>light</td>
<td>both describe luminescence</td>
</tr>
</tbody>
</table>
## Relationships between senses

### Hyponymy
- Sense A is a hyponym of sense B if A is a subclass of B
- Formally, entailment: for entity x, $A(x) \Rightarrow B(x)$
- Hyponymy is generally transitive

<table>
<thead>
<tr>
<th>hyponym/subordinate</th>
<th>hypernym/superordinate</th>
</tr>
</thead>
<tbody>
<tr>
<td>car</td>
<td>vehicle</td>
</tr>
<tr>
<td>mango</td>
<td>fruit</td>
</tr>
<tr>
<td>chair</td>
<td>furniture</td>
</tr>
<tr>
<td>dog</td>
<td>mammal</td>
</tr>
<tr>
<td>mammal</td>
<td>animal</td>
</tr>
</tbody>
</table>

hypo = “under” (e.g., hypothermia)
Relationships between senses

- Meronymy
  - Part-whole relations. A meronym is a part of a holonym.

<table>
<thead>
<tr>
<th>meronym</th>
<th>holonym</th>
</tr>
</thead>
<tbody>
<tr>
<td>leg</td>
<td>chair</td>
</tr>
<tr>
<td>wheel</td>
<td>car</td>
</tr>
<tr>
<td>car</td>
<td>automobile</td>
</tr>
</tbody>
</table>
WordNet

• large-scale database of lexical relations

• organized as graph whose nodes are synsets (synonym sets)
  • each synset consists of 1 or more word senses which are considered synonymous
  • fine-grained senses

• primary relation: hyponym / hypernym

• available on Web
  • along with foreign-language Wordnets
## Relations

<table>
<thead>
<tr>
<th>Relation</th>
<th>Also Called</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypernym</td>
<td>Superordinate</td>
<td>From concepts to superordinates</td>
<td>breakfast$^1 \rightarrow$ meal$^1$</td>
</tr>
<tr>
<td>Hyponym</td>
<td>Subordinate</td>
<td>From concepts to subtypes</td>
<td>meal$^1 \rightarrow$ lunch$^1$</td>
</tr>
<tr>
<td>Instance Hypernym</td>
<td>Instance</td>
<td>From instances to their concepts</td>
<td>Austen$^1 \rightarrow$ author$^1$</td>
</tr>
<tr>
<td>Instance Hyponym</td>
<td>Has-Instance</td>
<td>From concepts to concept instances</td>
<td>composer$^1 \rightarrow$ Bach$^1$</td>
</tr>
<tr>
<td>Member Meronym</td>
<td>Has-Member</td>
<td>From groups to their members</td>
<td>faculty$^2 \rightarrow$ professor$^1$</td>
</tr>
<tr>
<td>Member Holonym</td>
<td>Member-Of</td>
<td>From members to their groups</td>
<td>copilot$^1 \rightarrow$ crew$^1$</td>
</tr>
<tr>
<td>Part Meronym</td>
<td>Has-Part</td>
<td>From wholes to parts</td>
<td>table$^2 \rightarrow$ leg$^3$</td>
</tr>
<tr>
<td>Part Holonym</td>
<td>Part-Of</td>
<td>From parts to wholes</td>
<td>course$^7 \rightarrow$ meal$^1$</td>
</tr>
<tr>
<td>Substance Meronym</td>
<td></td>
<td>From substances to their subparts</td>
<td>water$^1 \rightarrow$ oxygen$^1$</td>
</tr>
<tr>
<td>Substance Holonym</td>
<td></td>
<td>From parts of substances to wholes</td>
<td>gin$^1 \rightarrow$ martini$^1$</td>
</tr>
<tr>
<td>Antonym</td>
<td></td>
<td>Semantic opposition between lemmas</td>
<td>leader$^1 \Leftrightarrow$ follower$^1$</td>
</tr>
<tr>
<td>Derivationally Related Form</td>
<td></td>
<td>Lemmas w/same morphological root</td>
<td>destruction$^1 \Leftrightarrow$ destroy$^1$</td>
</tr>
</tbody>
</table>

*Figure 17.2* Noun relations in WordNet.
## Synsets

<table>
<thead>
<tr>
<th>synset</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>mark, grade, score</td>
<td>a number or letter indicating quality</td>
</tr>
<tr>
<td>scratch, scrape, scar, mark</td>
<td>an indication of damage</td>
</tr>
<tr>
<td>bell ringer, bull's eye, mark, home run</td>
<td>something that exactly succeeds in achieving its goal</td>
</tr>
<tr>
<td>chump, fool, gull, mark, patsy, fall guy, sucker, soft touch, mug</td>
<td>a person who is gullible and easy to take advantage of</td>
</tr>
<tr>
<td>mark, stigma, brand, stain</td>
<td>a symbol of disgrace or infamy</td>
</tr>
</tbody>
</table>
Synsets

- *S:* (n) *victim, dupe* (a person who is tricked or swindled)
- *S:* (n) *person, individual, someone, somebody, mortal, soul* (a human being) "there was too much for one person to do"
- *S:* (n) *organism, being* (a living thing that has (or can develop) the ability to act or function independently)
- *S:* (n) *living thing, animate thing* (a living (or once living) entity)
- *S:* (n) *whole, unit* (an assemblage of parts that is regarded as a single entity) "how big is that part compared to the whole?"; "the team is a unit"
- *S:* (n) *object, physical object* (a tangible and visible entity; an entity that can cast a shadow) "it was full of rackets, balls and other objects"
- *S:* (n) *physical entity* (an entity that has physical existence)
- *S:* (n) *entity* (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

Hypernyms of {chump, fool, gull, patsy, fall guy, sucker, soft touch, mug} synset
WordNet

- WordNet encodes human-judged measures of similarity. Learn distributed representations of words that respect WordNet similarities (Faruqui et al. 2015).
- By indexing word senses, we can build annotated resources on top of it for word sense disambiguation (WSD).
- Semcor: 200K+ words from Brown corpus tagged with Wordnet senses.

<table>
<thead>
<tr>
<th>original</th>
<th>It urged that the city take steps to remedy this problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>lemma sense</td>
<td>It urge$^1$ that the city$^2$ take$^1$ step$^1$ to remedy$^1$ this problem$^2$</td>
</tr>
<tr>
<td>synset number</td>
<td>It urge$^{2:32:00}$ that the city$^{1:15:01}$ take$^{2:41:04}$ step$^{1:04:02}$ to remedy$^{2:30:00}$ this problem$^{1:10:00}$</td>
</tr>
</tbody>
</table>

http://web.eecs.umich.edu/~mihalcea/downloads/semcor/semcor3.0.tar.gz
“All-word” Word Sense Disambiguation

“Only a relative handful of such reports was received.”

- For all content words in a sentence, resolve each token to its sense in a fixed sense inventory (e.g., WordNet).

- Methods:
  - Dictionary methods (Lesk)
  - Supervised (machine learning)
  - Sem-supervised (bootstrapping)
Dictionary methods

• Predict the sense for a given token that has the highest overlap between the token’s context and sense’s dictionary gloss.

The boat washed up on the river bank.

<table>
<thead>
<tr>
<th align="left">bank&lt;sup&gt;1&lt;/sup&gt;</th>
<th align="left">Gloss:</th>
<th align="left">a financial institution that accepts deposits and channels the money into lending activities</th>
</tr>
</thead>
<tbody>
<tr>
<td align="left"></td>
<td align="left">Examples:</td>
<td align="left">“he cashed a check at the bank”, “that bank holds the mortgage on my home”</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th align="left">bank&lt;sup&gt;2&lt;/sup&gt;</th>
<th align="left">Gloss:</th>
<th align="left">sloping land (especially the slope beside a body of water)</th>
</tr>
</thead>
<tbody>
<tr>
<td align="left"></td>
<td align="left">Examples:</td>
<td align="left">“they pulled the canoe up on the bank”, “he sat on the bank of the river and watched the currents”</td>
</tr>
</tbody>
</table>
Lesk Algorithm

```python
function SIMPLIFIED LESK(word, sentence) returns best sense of word

best-sense ← most frequent sense for word
max-overlap ← 0
context ← set of words in sentence
for each sense in senses of word do
    signature ← set of words in the gloss and examples of sense
    overlap ← COMPUTEOVERLAP(signature, context)
    if overlap > max-overlap then
        max-overlap ← overlap
        best-sense ← sense
    end
return(best-sense)
```

- Extension (Basile et al. 2014): measure similarity between gloss $g = \{g_1, \ldots, g_G\}$ and context $c = \{c_1, \ldots, c_C\}$ as cosine similarity between sum of distributed representations.
Supervised WSD

- We have labeled training data; let’s learn from it
  - Decision trees (Yarowsky 1994)
  - Naive Bayes, log-linear classifiers, support vector machines (Zhong and Ng 2010)
  - Bidirectional LSTM (Raganato et al. 2017)
- Typical features
  - Collocational: words in specific positions before/after the target word to be disambiguated (e.g., one word before and after)
  - Bag-of-words: words in window around target (without encoding specific position)
  - part of speech tagging, lemmatization, syntactic parsing (headwords, dependency relations)

Can we apply naïve bayes for this? What’s the problem?
## Supervised WSD

<table>
<thead>
<tr>
<th></th>
<th>Dev</th>
<th>Test Datasets</th>
<th>Concatenation of All Test Datasets</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SE07</td>
<td>SE2</td>
<td>SE3</td>
</tr>
<tr>
<td>BLSTM</td>
<td>61.8</td>
<td>71.4</td>
<td>68.8</td>
</tr>
<tr>
<td>BLSTM + att.</td>
<td>62.4</td>
<td>71.4</td>
<td><strong>70.2</strong></td>
</tr>
<tr>
<td>BLSTM + att. + LEX</td>
<td>63.7</td>
<td><strong>72.0</strong></td>
<td>69.4</td>
</tr>
<tr>
<td>BLSTM + att. + LEX + POS</td>
<td><strong>64.8</strong></td>
<td><strong>72.0</strong></td>
<td>69.1</td>
</tr>
<tr>
<td>Seq2Seq</td>
<td>60.9</td>
<td>68.5</td>
<td>67.9</td>
</tr>
<tr>
<td>Seq2Seq + att.</td>
<td>62.9</td>
<td>69.9</td>
<td>69.6</td>
</tr>
<tr>
<td>Seq2Seq + att. + LEX</td>
<td>64.6</td>
<td>70.6</td>
<td>67.8</td>
</tr>
<tr>
<td>Seq2Seq + att. + LEX + POS</td>
<td>63.1</td>
<td>70.1</td>
<td>68.5</td>
</tr>
<tr>
<td>IMS</td>
<td>61.3</td>
<td>70.9</td>
<td>69.3</td>
</tr>
<tr>
<td>IMS+emb</td>
<td><strong>62.6</strong></td>
<td><strong>72.2</strong></td>
<td><strong>70.4</strong></td>
</tr>
<tr>
<td>Context2Vec</td>
<td>61.3</td>
<td>71.8</td>
<td>69.1</td>
</tr>
<tr>
<td>Lesk&lt;sub&gt;ext&lt;/sub&gt;+emb</td>
<td><em>56.7</em></td>
<td>63.0</td>
<td>63.7</td>
</tr>
<tr>
<td>UKB&lt;sub&gt;gloss&lt;/sub&gt; w2w</td>
<td>42.9</td>
<td>63.5</td>
<td>55.4</td>
</tr>
<tr>
<td>Babelfy</td>
<td>51.6</td>
<td><em>67.0</em></td>
<td>63.5</td>
</tr>
<tr>
<td>MFS</td>
<td>54.5</td>
<td>65.6</td>
<td><em>66.0</em></td>
</tr>
</tbody>
</table>

Raganato et al. 2017
We can extend the notion of disambiguating individual words to cover multi-word terms and names.

– Wikipedia comes closest to providing an inventory of such concepts: people, places, classes of objects, ....

– This has led to the process of Wikification: linking the phrases in a text to Wikipedia articles (also called entity linking)

• annual evaluation (for names) as part of NIST Text Analysis Conference
# Entity linking

**Michael Jordan (disambiguation)**

From Wikipedia, the free encyclopedia

Michael Jordan (born 1963), American basketball player and businessman

Michael or Mike Jordan el cirujano may also refer to:

## People

<table>
<thead>
<tr>
<th>People</th>
</tr>
</thead>
<tbody>
<tr>
<td>[edit]</td>
</tr>
</tbody>
</table>

## Sports

<table>
<thead>
<tr>
<th>Sports</th>
</tr>
</thead>
<tbody>
<tr>
<td>[edit]</td>
</tr>
</tbody>
</table>

- Michael Jordan (footballer) (born 1986), English goalkeeper
- Mike Jordan (racing driver) (born 1958), English racing driver
- Mike Jordan (baseball, born 1863) (1863–1940), baseball player
- Mike Jordan (cornerback) (born 1992), American football cornerback
- Michael Jordan (offensive lineman), American football offensive lineman
- Michael-Hakim Jordan (born 1977), American professional basketball player
- Michal Jordán (born 1990), Czech ice hockey player

## Other people

<table>
<thead>
<tr>
<th>Other people</th>
</tr>
</thead>
<tbody>
<tr>
<td>[edit]</td>
</tr>
</tbody>
</table>

- Michael B. Jordan (born 1987), American actor
- Michael I. Jordan (born 1956), American researcher in machine learning and artificial intelligence
- Michael Jordan (insolvency baron) (born 1931), English businessman
- Michael Jordan (Irish politician), Irish Farmers' Party TD from Wexford, 1927–1932
- Michael H. Jordan (1936–2010), American executive for CBS, PepsiCo, Westinghouse
- Michael Jordan (mycologist), English mycologist
Entity linking

- Entity linking is often cast as a learning to rank problem: given a mention $x$, some set of candidate entities $\mathcal{Y}(x)$ for that mention, and context $c$, select the highest scoring entity from that set.

$$\hat{y} = \arg \max_{y \in \mathcal{Y}(x)} \Psi(y, x, c)$$

- We can also use these scores to compute a probability distribution over the candidate entities (i.e., via softmax). We can then train the models based by optimizing the likelihood function.
The scoring function

\[ \Psi(y, x, c) \]

- **Feature-based approach**
  
  \[
  \text{feature} = f(x, y, c) \\
  \text{string similarity between } x \text{ and } y \\
  \text{popularity of } y \\
  \text{NER type}(x) = \text{type}(y) \\
  \text{cosine similarity between } c \text{ and Wikipedia page for } y
  \]

  \[
  \Psi(y, x, c) = f(x, y, c) \beta
  \]

- **Deep learning to learn the similarity between the context and the Wikipedia page for the entity candidate (e.g., using CNN to learn the representation vectors for both the context and the Wikipedia page from which the cosine similarity can be computed)**
Local vs Global Disambiguation

• Local disambiguation
  – each mention (word, name, term) in an article is disambiguated separately based on context (other words in article)

• Global disambiguation:
  – take into account coherence of disambiguations across document
  – optimize sum of local disambiguation scores plus a term representing coherence of referents
    • coherence reflected in links between Wikipedia entries
  – relative importance of prominence, local features, and global coherence varies greatly
Using Coherence

Wikipedia entries

Texas Rangers (lawmen)
Texas Rangers (baseball team)
NY Yankees (baseball team)

document:

... the Texas Rangers defeated the New York Yankees ...
Using Coherence

Wikipedia entries

- Texas Rangers (lawmen)
- Texas Rangers (baseball team)
- NY Yankees (baseball team)

document: ...
... the Texas Rangers defeated the New York Yankees ...

links in Wikipedia
... the Texas Rangers defeated the New York Yankees ...
Entity linking

Nguyen et al., 2016
Supervised vs. Semi-supervised

• problem: training some classifiers (such as WSD) needs lots of labeled data
  – supervised learners: all data labeled

• alternative: semi-supervised learners
  – some labeled data ("seed") + lots of unlabeled data
Bootstrapping: a semi-supervised learner

Basic idea of bootstrapping:

• start with a small set of labeled seeds $L$ and a large set of unlabeled examples $U$

repeat

• train classifier $C$ on $L$
• apply $C$ to $U$
• identify examples with most confident labels; remove them from $U$ and add them (with labels) to $L$
Bootstrapping WSD

Premises:

• one sense per discourse (document)

• one sense per collocation
example

“bass” as fish or musical term
example

- bass
  - catch bass
  - play bass
- bass
  - catch bass
  - play bass
example

- label initial examples

- bass
  - fish
  - catch bass

- bass
  - music
  - play bass

- bass
  - play bass

- bass
  - catch bass
example

- label other instances in same document

- bass
  - fish
  - catch bass
- bass
  - music
  - play bass
- bass
  - music
  - play bass
- bass
  - fish
  - catch bass
example

- learn collocations: catch ... → fish; play ... → music
example

- label other instances of collocations

- bass
  - fish
  - catch bass
    - fish
  - play bass
    - fish

- bass
  - music
  - play bass
    - music
Using WordNet

- Simplest measures of semantic similarity based on WordNet: path length:

  ![Diagram showing the hierarchy of mammals, felines, apes, cats, tigers, gorillas, and humans with path lengths for comparison.]

  longer path → less similar

  mammals

  felines     apes
  cats  tigers gorillas  humans

Using WordNet

- path length ignores differences in degrees of generalization in different hyponym relations:

```
mammals
  /\      
cats  people
```

*a cat’s view of the world (cats and people are similar)*
Information Content

- $P(c) =$ probability that a word in a corpus is an instance of the concept (matches the synset $c$ or one of its hyponyms) (computed based on a corpus)

- Information content of a concept
  $$IC(c) = -\log P(c)$$

- If $LCS(c_1, c_2)$ is the lowest common subsumer of $c_1$ and $c_2$, the IC distance between $c_1$ and $c_2$ is
  $$IC(c_1) + IC(c_2) - 2 IC(LCS(c_1, c_2))$$

- Variants: Resnik Similarity, Jiang-Conrath Similarity

http://www.nltk.org/howto/wordnet.html