One grand goal of artificial intelligence is to understand what people mean when they talk.

But how do we know if we succeeded?

Joseph Weizenbaum wrote the computer program Eliza in 1969 to demonstrate how easily people can be fooled into thinking that machines have some deep understanding.

young woman: Men are all alike.
eliza: In what way?
young woman: They’re always bugging us about something specific or other.
eliza: Can you think of a specific example?
young woman: Well, my boyfriend made me come here.
eliza: Your boyfriend made you come here?
young woman: He says I’m depressed much of the time.
eliza: I am sorry to hear you are depressed.
young woman: It’s true. I’m unhappy.
eliza: Do you think coming here will help you not to be unhappy?

Online demo: http://www.manifestation.com/neurotoys/eliza.php3
A Concrete Goal

- We would like to build
  - a machine that answers questions in natural language.
  - may have access to knowledge bases
  - may have access to vast quantities of English text

- Basically, a smarter Google

- This is typically called **Question Answering**

Why is lexical semantics important for building such a system?

---

Semantics

- To build our QA system we will need to deal with issues in **semantics**, i.e., meaning.

- Lexical semantics: the meanings of individual words (next few lectures)

- Sentential semantics: how word meanings combine (after that)
  - Who did what to whom; when, how, why . . .

- Consider some examples to highlight problems in lexical semantics

---

Example Question

- **Question**
  When was Barack Obama born?

- Text available to the machine
  Barack Obama was born on August 4, 1961

- This is easy.
  - just phrase a Google query properly:
    "Barack Obama was born on *"
  - syntactic rules that convert questions into statements are straight-forward

---

Example Question (2)

- **Question**
  What plants are native to Scotland?

- Text available to the machine
  A new chemical plant was opened in Scotland.

- What is hard?
  - words may have different meanings (**senses**)
  - we need to be able to disambiguate between them
Example Question (3)

- Question
  Where did David Cameron go on vacation?

- Text available to the machine
  David Cameron spent his holiday in Cornwall

- What is hard?
  - words may have the same meaning (synonyms)
  - we need to be able to match them

Example Question (4)

- Question
  Which animals love to swim?

- Text available to the machine
  Polar bears love to swim in the freezing waters of the Arctic.

- What is hard?
  - words can refer to a subset (hyponym) or superset (hypernym) of the concept referred to by another word
  - we need to have database of such A is-a B relationships, called an ontology

Example Question (5)

- Question
  What is a good way to remove wine stains?

- Text available to the machine
  Salt is a great way to eliminate wine stains

- What is hard?
  - words may be related in other ways, including similarity and gradation
  - we need to be able to recognize these to give appropriate responses

Example Question (6)

- Question
  Did Poland reduce its carbon emissions since 1989?

- Text available to the machine
  Due to the collapse of the industrial sector after the end of communism in 1989, all countries in Central Europe saw a fall in carbon emissions. Poland is a country in Central Europe.

- What is hard?
  - we need to do inference
  - a problem for sentential, not lexical, semantics
WordNet

- Some of these problems can be solved with a good ontology, e.g., WordNet

- WordNet (English) is a hand-built resource containing 117,000 synsets: sets of synonymous words (See http://wordnet.princeton.edu/)

- Synsets are connected by relations such as
  - hyponym/hypernym (IS-A: chair-furniture)
  - meronym (PART-WHOLE: leg-chair)
  - antonym (OPPOSITES: good-bad)

- globalwordnet.org now lists wordnets in over 50 languages (but variable size/quality/licensing)

Word Sense Ambiguity

- One word form, same category, but more than one sense (homonyms):
  
  I put my money in the bank. vs. He rested at the bank of the river.

- More generally, words can have multiple (related or unrelated) senses

- Words often exhibit sense ambiguities that fall into (semi-)predictable patterns (polysemy): see next slides (from Hugh Rabagliati in PPLS).

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Participating Senses</th>
<th>Example Sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animal for fur</td>
<td>Mink, chinchilla, rabbit, beaver, raccoon*, alpaca*, crocodile*</td>
<td>The mink drank some water / She likes to wear mink</td>
</tr>
<tr>
<td>Animal/Object for personality</td>
<td>Chicken, sheep, pig, snake, star*, rat*, doll*</td>
<td>The chicken drank some water / He is a chicken</td>
</tr>
<tr>
<td>Animal for meat</td>
<td>Chicken, lamb, fish, shrimp, salmon*, rabbit*, lobster*</td>
<td>The chicken drank some water / The chicken is tasty</td>
</tr>
<tr>
<td>Artifact for activity</td>
<td>Shower, bath, sauna, baseball,</td>
<td>The shower was leaking / The shower was relaxing</td>
</tr>
<tr>
<td>Body part for object part</td>
<td>Arm, leg, hand, face, back*, head*, foot*, shoulder*, lip*,</td>
<td>John’s arm was tired / The arm was reupholstered</td>
</tr>
<tr>
<td>Building for people</td>
<td>Church, factory, school, airplane,</td>
<td>The church was built 20 years ago / The church sang a song</td>
</tr>
<tr>
<td>Complement Coercion</td>
<td>Begin, start, finish, try</td>
<td>John began reading the book / John began the book</td>
</tr>
<tr>
<td>Container for contents</td>
<td>Bottle, can, pot, pan, bowl*, plate*, box*, bucket*</td>
<td>The bottle is made of steel / He drank half of the bottle</td>
</tr>
<tr>
<td>Word for question</td>
<td>Price, weight, speed</td>
<td>The price of the coffee was low / He asked the price of the coffee</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Participating Senses</th>
<th>Example Sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure for Ground</td>
<td>Window, door, gate, goal</td>
<td>The window is broken / The cat walked through the window</td>
</tr>
<tr>
<td>Grinding</td>
<td>Apple, chair, fly</td>
<td>The apple was tasty / There is apple all over the table</td>
</tr>
<tr>
<td>Instrument for action</td>
<td>Hammer, brush, shovel, tape, lock*, bicycle*, comb*, saw*</td>
<td>The hammer is heavy / She hammered the nail into the wall</td>
</tr>
<tr>
<td>Instance of an entity for kind</td>
<td>Tennis, soccer, cat, dog, class*, dinner*, chair*, table*</td>
<td>Tennis was invented in England / Tennis was fun today</td>
</tr>
<tr>
<td>Location / Place at location</td>
<td>Bench, land, floor, ground, box*, bottle*, jail*</td>
<td>The bench was made of pine / The coach bench the player</td>
</tr>
<tr>
<td>Object for placing at goal</td>
<td>Water, paint, salt, butter, frame*, dress*, oil*</td>
<td>The water is cold / He watered the plant.</td>
</tr>
<tr>
<td>Object for taking from source</td>
<td>Milk, dust, weed, peel, pit*, skin*, juice*</td>
<td>The milk tastes good / He milked the cow.</td>
</tr>
<tr>
<td>Material for artifact</td>
<td>Tin, iron, china, glass, linen*, rubber*, nickel*, fur*</td>
<td>Watch out for the broken glass / He filled the glass with water</td>
</tr>
<tr>
<td>Occupation for role in action</td>
<td>Boss, nurse, guard, tutor</td>
<td>My boss is nice / He bossed me around</td>
</tr>
</tbody>
</table>
Some lessons to draw

- Words are typically semantically ambiguous
- There’s a lot of regularity (and hence predictability) in the range of senses a word can take
- Those senses also influence the word’s syntactic behaviour
- But all regularities admit (arbitrary) exceptions
- Word senses can be productive, making a dictionary model inadequate
- But it’s a dominant model in CL these days, and works quite well in lots of cases.

How many senses?

- 5 min. exercise: How many senses does the word interest have?
How many senses?

- How many senses does the word interest have?
  - She pays 3% interest on the loan.
  - He showed a lot of interest in the painting.
  - Microsoft purchased a controlling interest in Google.
  - It is in the national interest to invade the Bahamas.
  - I only have your best interest in mind.
  - Playing chess is one of my interests.
  - Business interests lobbied for the legislation.

- Are these seven different senses? Four? Three?

- Also note: distinction between polysemy and homonymy not always clear!

Lumping vs. Splitting

- For any given word, lexicographer faces the choice:
  - Lump usages into a small number of senses? or
  - Split senses to reflect fine-grained distinctions?

WordNet senses for interest

- S1: a sense of concern with and curiosity about someone or something, Synonym: involvement
- S2: the power of attracting or holding one’s interest (because it is unusual or exciting etc.), Synonym: interestingness
- S3: a reason for wanting something done, Synonym: sake
- S4: a fixed charge for borrowing money; usually a percentage of the amount borrowed
- S5: a diversion that occupies one’s time and thoughts (usually pleasantly), Synonym: pastime, pursuit
- S6: a right or legal share of something; a financial involvement with something, Synonym: stake
- S7: (usually plural) a social group whose members control some field of activity and who have common aims, Synonym: interest group
Synsets and Relations in WordNet

- **Synsets** ("synonym sets", effectively senses) are the basic unit of organization in WordNet.
  - Each synset is specific to nouns (.n), verbs (.v), adjectives (.a, .s), or adverbs (.r).
  - Synonymous words belong to the same synset: \texttt{car} \texttt{1} (\texttt{car.n.01}) = \{car, auto, automobile\}.
  - Polysemous words belong to multiple synsets: \texttt{car} \texttt{1} vs. \texttt{car} \texttt{4} = \{car, elevator car\}. Numbered roughly in descending order of frequency.

- Synsets are organized into a **network** by several kinds of relations, including:
  - **Hypernymy** (Is-A): hyponym \{\texttt{ambulance}\} is a kind of hypernym \texttt{car} \texttt{1}
  - **Meronymy** (Part-Whole): meronym \{\texttt{air bag}\} is a part of holonym \texttt{car} \texttt{1}

Using WordNet

- NLTK provides an excellent API for looking things up in WordNet:

```python
>>> from nltk.corpus import wordnet as wn
>>> wn.synsets('car')
[Synset('car.n.01'), Synset('car.n.02'), Synset('car.n.03'), Synset('car.n.04'), Synset('cable_car.n.01')]
>>> wn.synset('car.n.01').definition()
'u \text{'a motor vehicle with four wheels; usually propelled by an internal combustion engine}''
>>> wn.synset('car.n.01').hypernyms()
[Synset('motor_vehicle.n.01')]
```

- (WordNet uses an obscure custom file format, so reading the files directly is not recommended!)

Coverage in WordNet

- Online stats:
  - 155k unique strings, 118k unique synsets, 207k pairs
  - nouns have an average 1.24 senses (2.79 if excluding monosemous words)
  - verbs have an average 2.17 senses (3.57 if excluding monosemous words)

- Too fine-grained?

- WordNet is a snapshot of the English lexicon, but by no means complete.
  - E.g., consider **multiword expressions** (including noncompositional expressions, idioms): hot dog, take place, carry out, kick the bucket, but not take a break, stress out, pay attention
  - Neologisms: hoodie, facepalm
  - Names: Microsoft
  - Predictable but novel uses of known words: Badger is a delicacy in Mongolia.
Different sense = different translation

- Another way to define senses: if occurrences of the word have different translations, these indicate different sense

- Example interest translated into German
  - Zins: financial charge paid for load (WordNet sense 4)
  - Anteil: stake in a company (WordNet sense 6)
  - Interesse: all other senses

- Other examples might have distinct words in English but a polysemous word in German.

Word sense disambiguation (WSD)

- For many applications, we would like to disambiguate senses
  - we may be only interested in one sense
  - searching for chemical plant on the web, we do not want to know about chemicals in bananas

- Task: Given a sense ambiguous word, find the sense in a given context

- Popular topic, data driven methods perform well

WSD as classification

- Given a word token in context, which sense (class) does it belong to?

- We can train a supervised classifier, assuming sense-labeled training data:
  - She pays 3% interest/INTEREST-MONEY on the loan.
  - He showed a lot of interest/INTEREST-CURIOSITY in the painting.
  - Playing chess is one of my interests/INTEREST-HOBBY.

- SensEval and later SemEval competitions provide such data
  - held every 1-3 years since 1998
  - provide annotated corpora in many languages for WSD and other semantic tasks

What kind of classifier?

- Lots of options available:
  - Naïve Bayes, MaxEnt (see Lecture 7)
  - Decision lists (see J&M, 20.2.2)
  - Decision trees (see any ML textbook)
Naïve Bayes for WSD

\[
\hat{s} = \arg \max_{s \in S} P(s | \vec{f})
\]

\[
\approx \arg \max_{s \in S} P(s) \prod_{j=1}^{n} P(f_j | s)
\]

- Naïve Bayes requires estimates of:
  - The prior probability of each class (sense)
  - The probability of each feature given each class
- These can be estimated from the training data.
- But what features to use? (Same question for other classifiers!)

Simple features

- Directly neighboring words (and/or their lemmas)
  - interest paid
  - rising interest
  - lifelong interest
  - interest rate
  - interest piqued
- Any content words in a 50 word window
  - pastime
  - financial
  - lobbied
  - pursued

More features

- Syntactically related words
- Syntactic role in sense
- Topic of the text
- Part-of-speech tag, surrounding part-of-speech tags

Of course, with NB we have the usual problem with correlated features. MaxEnt doesn’t assume they are independent.

Evaluation

- Extrinsic: test as part of IR, QA, or MT system
- Intrinsic: evaluate classification accuracy or precision/recall against gold-standard senses
- Baseline: choose the most frequent sense (sometimes hard to beat)
Issues with WSD

- Not always clear how fine-grained the gold-standard should be
- Difficult/expensive to annotate corpora with fine-grained senses
- Classifiers must be trained separately for each word
  - Hard to learn anything for infrequent or unseen words
  - Requires new annotations for each new word
  - Motivates unsupervised and semi-supervised methods (see J&M 20.5, 20.10)

Semantic Classes

- Other approaches, such as named entity recognition and supersense tagging, define coarse-grained semantic categories like PERSON, LOCATION, ARTIFACT.
- Like senses, can disambiguate: APPLE as ORGANIZATION vs. FOOD.
- Unlike senses, which are refinements of particular words, classes are typically larger groupings.
- Unlike senses, classes can be applied to words/names not listed in a lexicon.

Named Entity Recognition

- Recognizing and classifying proper names in text is important for many applications. A kind of information extraction.
- Different datasets/named entity recognizers use different inventories of classes.
  - Smaller: PERSON, ORGANIZATION, LOCATION, MISCELLANEOUS
  - Larger: sometimes also PRODUCT, WORK_OF_ART, HISTORICAL_EVENT, etc., as well as numeric value types (TIME, MONEY, etc.)
- NER systems typically use some form of feature-based sequence tagging, with features like capitalization being important.
- Lists of known names called gazetteers are also important.

Supersenses in WordNet

- The supersense tagging goes beyond NER to cover all nouns and verbs.

```
N:Tops
N:Act
N:Animal
N:Artifact
N:Attribute
N:Body
N:Cognition
N:Communication
N:Event
N:Feeling
N:Food
N:Group
N:Location
N:Motive

N:Object
N:Person
N:Phenomenon
N:Plant
N:Possession
N:Process
N:Quantity
N:Relation
N:Shape
N:State
N:Substance
N:Time

V:Cognition
V:Communication
V:Competition
V:Consumption
V:Contact
V:Creation
V:Emotion
V:Motion
V:Perception
V:Possession
V:Social
V:Stative
V:Change
```

- Lists of known names called gazetteers are also important.
• In order to support technologies like question answering, we need ways to reason computationally about meaning. **Lexical semantics** addresses meaning at the word level.
  – Words can be ambiguous, sometimes with related meanings (**polysemy**), and other times with unrelated meanings (**homonymy**).
  – Different words can mean the same thing (**synonymy**).

• Computational lexical databases, notably WordNet, organize words in terms of their meanings.
  – **Synsets** and relations between them such as hypernymy and meronymy.

**Summary (2)**

• **Word sense disambiguation** is the task of choosing the right sense for the context.
  – Classification with contextual features
  – Relying on dictionary senses has limitations in granularity and coverage

• **Semantic classes**, as in NER and supersense tagging, are a coarser-grained representation for semantic disambiguation and generalization.

**Next Lecture: Distributional lexical semantics**

• What can we learn about a word’s meaning from “the company it keeps”?

• What do we do if our thesaurus is incomplete?

• Distributional lexical semantics is about learning word meaning from the contexts in which words appear