Foundations of Natural Language Processing
Lecture 20b
Lexical Semantics:
The Generative Lexicon

Alex Lascarides

School of informatics

So far

- NL and its use relies on commonsense inference and hence on lexical semantics
- Relations among word meanings
  - Hyponym, hypernym, antonym, synonym, meronym

  influence NL understanding

Now:

- Computational lexical resources
- Word meanings can be productive and (partly) predictable
- How do we represent that?

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.
  I like playing *squash* vs. I like drinking *squash*

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

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

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

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

Pattern	Participating Senses	Example Sentences
Place for an event	Vietnam, Korea, Waterloo, Iraq	It is raining in *Vietnam* / John was shot during *Vietnam*
Place for an institution	White House, Washington, Hollywood, Pentagon, Wall Street*, Supreme Court	The *White House* is being repainted / The *White House* made an announcement
Plant for food or material	Corn, broccoli, coffee, cotton, lettuce*, eggs*, oak*, pine*	The large field of *corn* / The *corn* is delicious
Portioning	Water, beer, jam	She drank some *water* / She bought three *waters*
Publisher for product	Newspaper, magazine, encyclopedia, Wall Street Journal*, New York Times*,	The *newspaper* is badly printed / The *newspaper* fired three employees
Artist for product	Writer, artist, composer, Shakespeare, Dickens*, Mozart*, Picasso*	The *writer* drank a lot of wine / The *writer* is hard to understand
Object for contents	Book, CD, DVD, TV*, magazine*, newspaper*	The heavy, leather- bound *book* / The *book* is funny.
Visual Metaphor	Beam, belt, column, stick, bug*, leaf*	Most of the weight rests on the *beam* / There was a *beam* of light

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 (like WordNet) inadequate
- But it's a dominant model in CL these days, and works quite well in lots of cases.

How many senses?

- Exercise: How many senses does the word interest have?

How many senses?

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 regular polysemy and homonymy not always clear!

Lexicography requires data

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

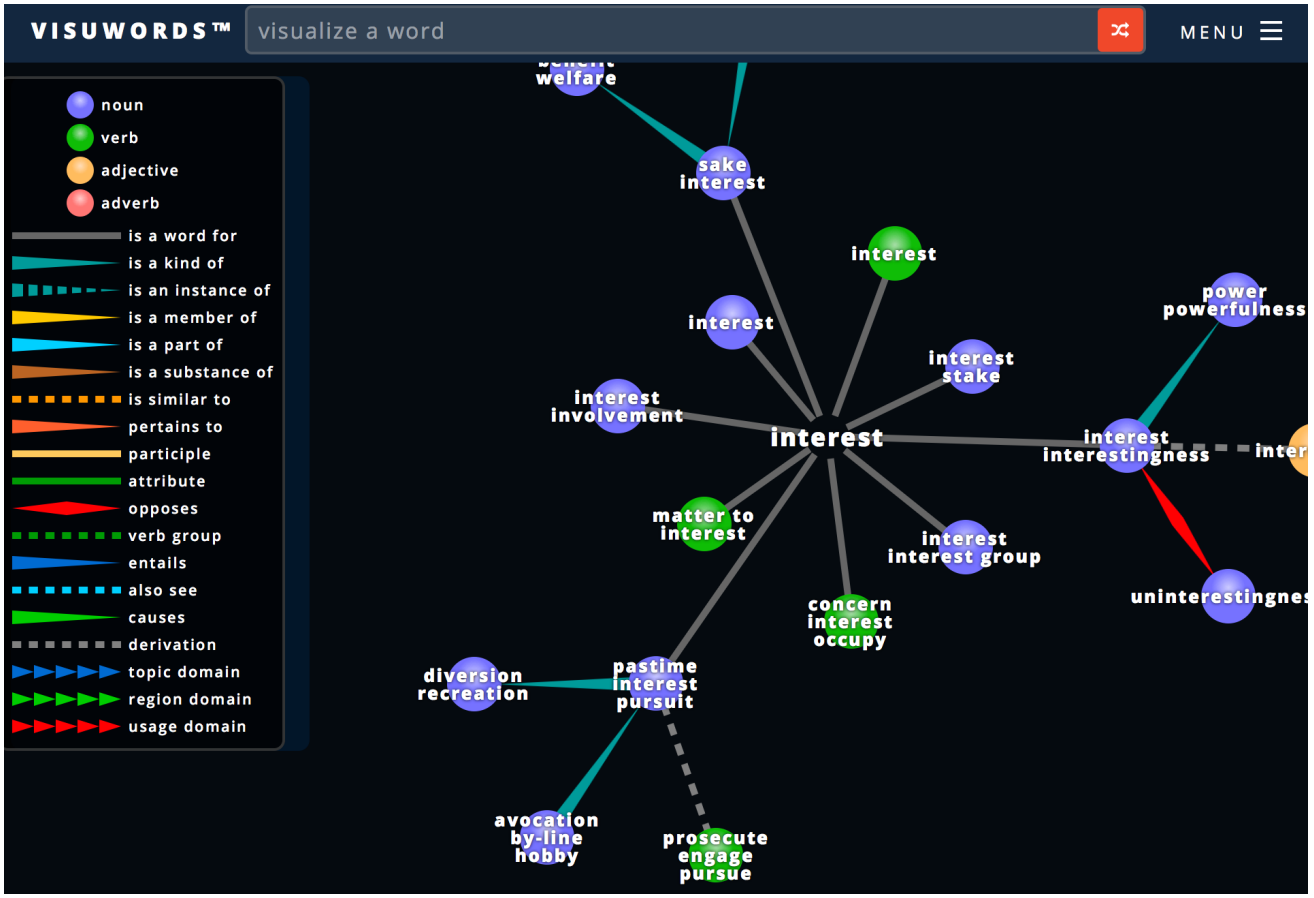
- S1: a sense of concern with and curiosity about someone or something, Synonym: involvement
- S2: the power of attracting or holding one's interest (because it is unusual or exciting etc.), Synonym: interestingness
- S3: a reason for wanting something done, Synonym: sake
- S4: a fixed charge for borrowing money; usually a percentage of the amount borrowed
- S5: a diversion that occupies one's time and thoughts (usually pleasantly), Synonyms: 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: car^1 (`car.n.01`) = {car,auto,automobile}.
  - Polysemous words belong to multiple synsets: car^1 vs. car^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 {ambulance} is a kind of hypernym car^1
  - **Meronymy** (Part-Whole): meronym {air bag} is a part of holonym car^1

Visualizing WordNet



Using WordNet

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

```
>>> 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'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 exluding monosemous words)
  - verbs have an average 2.17 senses (3.57 if exluding 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 are in WordNet, 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 (ambiguous) word in German.

Summary

- Words are often sense ambiguous
- So lexical resources associate a single word form to several senses
- A word sense can be a product of a lexical generalisation that applies to a whole class of words (e.g., animal words, sound words. . .)
- That is, some word senses are predictable
  and the lexicon is productive
  making a 'dictionary' model (like WordNet) inadequate.
- Nevertheless, WordNet is very useful!