Mechanisms of Meaning
Autumn 2010

Raquel Fernández
Institute for Logic, Language & Computation
University of Amsterdam
• We will discuss the following two papers:

• We will discuss the homework to be done in the coming couple of weeks (recall that the following two classes are cancelled.)
• **Topic under investigation**: polysemous adjective-noun combinations.

• **Approach**: probabilistic model of the polysemous meanings of adjective-noun combinations which acquires such meanings from corpus-based data.

• **Motivation**: according to GL, the adjective binds the telic role of the noun, but theoretical models do not give an exhaustive list of the events a noun can be related to, nor have anything to say about the likelihood of possible interpretations.

⇒ *Example from M. van Lambalgen’s guest lecture in LoLaCo course*
A nice example from Michiel van Lambalgen’s guest lecture “Logic in a Neuroscience Lab” in the LoLaCo MoL course on Sept 13:

- I began the novel in early december.
  - [reading, writing]

- The pupils were very attentive and fascinated by the story.
  - [explaining]

- However some found dictation extremely boring.
  - [dictating]
Proposition: The meaning of adjective-noun combinations can be paraphrased using a verb that instantiates the telic role of the noun. Given an adjective-noun combination, the proposed model exploits the likelihood of any verb to be modified by the adjective/adverb and to take the noun as argument to propose a ranking of possible meanings.

Evaluation and Results: The results obtained with the probabilistic model are compared against human judgements. The output of the model correlates significantly with human intuitions and performs consistently better than a baseline model.
• **Topic under investigation:** This is a more theoretical paper that tackles foundational issues. How adequate are current [1997] accounts of “word sense”?

• **Motivation:** The problem of Word Sense Disambiguation (WSD) takes for granted the notion of “word sense”. However, existing accounts of such a notion do not seem to be well-founded.

• **Proposal:** Word senses as clusters of usage instances extracted from corpus evidence. Importantly, clusters (senses) are domain- and task-dependent – in the abstract (independently of a particular purpose) they do not exist.
Motivation: What are the problems with existing accounts of word senses according to the author?

• **Fact**: there is a one-to-many relation between word forms and senses.

• How are the different senses of a word related to one another? The common assumption is that there are basically two options (dif. terms):
  
  * unrelated senses: ambiguity; sense selection; (homonymy’)
  * related senses: polysemy; indeterminacy/vagueness; sense modulation

• Given this theoretical distinction, it should be possible to classify pairs of examples as instances if either ambiguity or polysemy.

• However, there isn’t a set of criteria or tests that allows us to reliably make such classification (⇝ what are the problems Kilgarriff points out?)

• Semantic judgements are problematic; psycholinguistic findings may help us out...

• ...but this does not seem to be enough to provide a solid theoretical grounding for the above distinction.
Proposal: switch from subjective to objective methods; from introspective judgements to contexts.

∗ Extract concordances for a word (occurrences in context, with the key word aligned)

Part of a concordance for ‘handbag’ in the British National Corpus (BNC):

They might a Cartier watch or a Chanel handbag. It is the Rolls-Royce of pens; prices could cost you money! If you carry a handbag, make sure it has a secure clasp or zip dog worth his salt would bite open a handbag to get to the chocolate? " It's pieces and could n't even find a handbag. Of course now they're recalling told by men) are about hanging a handbag on the pulled-out choke or never being which could nevertheless fit into a handbag. The magazine circulated his trousers hit him over the head with a handbag and he launched himself instantly at and a baby inadvertently left in a handbag at Victoria Station left-luggage office in and out of expensive perfume and handbag shops. One or two ruined houses could

You can extract concordances from several English corpora here: http://corpus.leeds.ac.uk/protected/query.html

∗ Divide them into clusters corresponding to senses – the inventory of senses will depend on the rationale behind the clustering process.
Conclusions:

- The basic units to characterize word meaning are occurrences of words in context.
- Word senses are reduced to abstractions over clusters of word usages.
- The rationale behind clustering is domain dependent: word senses can only be defined relative to a set of interests.
Homework for Coming Weeks

- **Homework 1**: Summary of the CSL talk on Wednesday.
- **Homework 2**: Semantic annotation exercise.
- **Homework 3**: Next topic starting on Monday 11 October: psychological theories of concepts and word meaning.
  
  * Readings: selected chapters from Murphy (2002) *The Big Book of Concepts*
  * Student presentations: need to decide who presents what.
Homework 1

- Attend the talk by Stefan Evert on “Distributional Semantic Models” [Computational Linguistics Seminar on Wed 22, 4pm].
- Write a summary of the talk. It should include two parts
  - an objective summary of the contents of the talk where you do not give your opinion, and
  - a critical comment where you do give your opinion.
- Practical matters:
  - Minimum 1 page; maximum 2 pages.
  - Sent to me via email (raquel.fernandez@uva.nl) as a PDF attachment with your name (e.g. raquel-summary.pdf)
  - Due on Monday 27 September.
Semantic Annotation Exercise:
Adapted from an exercise designed by Gemma Boleda; Computational Lexical Semantics (ESSLLI 2009).

• Hands-on exercise on semantics judgements regarding one type of semantic relation
• Task: decide, for each sentence in a data set, whether two nouns bear the semantic relation Content-Container.

* The <e1>apples</e1> are in the <e2>basket</e2>.
  Content-Container(e1, e2) = true

* The <e1>silver</e1> <e2>ship</e2> usually carried silver bullion bars, but sometimes the cargo was gold or platinum.
  Content-Container(e1, e2) = true

* Summer was over and he knew that the <e1>climate</e1> in the <e2>forest</e2> would only get worse.
  Content-Container(e1, e2) = false
Semantic Annotation Exercise: Instructions

• Download the data set and the guidelines from the course website (further examples of positive and negative instances in the guidelines).

• Read the definition of the semantic relation carefully, and annotate the data set according to it:
  * create a text file or a spreadsheet file;
  * make sure you use one line per item in the data set;
  * use the label true if the relation holds and false if it doesn’t.

• Your annotation file will look like this:

  true
  true
  false
  ...

  Each line corresponds to one sentence in the data set. Use only one label per line (do not include the sentence number).

• Name your annotation file with your name (e.g. raquel-annotation.txt)

• Due on Monday 4 October, sent via email as an attachment (text, excel or open office format).
Semantic Annotation Exercise: Instructions

- Do it independently without discussing among yourselves!!
- There are no “correct” and “incorrect” answers.
- We will calculate the *inter-annotator agreement* among yourselves and with respect to a *gold standard* in class.
- Make a note of those examples where you were doubtful between true and false. What was the problem?
- In those cases where you chose *false*, which semantic relation would have been appropriate? Here are a few possibilities:
  - Cause-Effect (e.g., virus-flu)
  - Instrument-Agency (e.g., laser-printer)
  - Product-Producer (e.g., honey-bee)
  - Origin-Entity (e.g., rye-whiskey)
  - Theme-Tool (e.g., soup-pot)
  - Part-Whole (e.g., wheel-car)
- We will discuss this in the next class.

- **Chapter 2: Typicality and the Classical View of Categories**
  * we need a volunteer to present this on Monday 11 October

- **Chapter 3: Theories**
  * we need a volunteer to present this on Monday 18 October

- **Chapter 11: Word Meaning**
  * we need a volunteer to present this on Monday 18 October

**Course Evaluation:**

- Homework 25%
- Presentations of readings 25% (1 or 2 presentations per head)
- Final paper + presentation 50%