

Computational Semantics and Pragmatics

Autumn 2011

Raquel Fernández

Institute for Logic, Language & Computation
University of Amsterdam



What we have seen so far...

Recognising whether entailment holds is a core aspect of our ability to understand language.

- (1) Apple filed a lawsuit against Samsung for patent violation.
- (2) Samsung has been sued by Apple.

We have looked into some of the challenges involved in modelling the generic ability of recognising **textual entailment**.

- Knowledge required:
 - * syntax and compositional semantics (incl. active/passive relation)
 - * semantic relations between lexical items (e.g. *sell/buy*, *asphyxiate/kill*)
 - * reference resolution
 - * world knowledge
 - * ...

What we have seen so far...

We can model textual entailment in terms of **logical consequence**.

- representing the meaning of the target sentences and the required knowledge as logical formulas
- using automated reasoning tools (theorem proving and model building)
- problems: knowledge acquisition + undecidability

We can also develop a model using **shallow features**.

- extracting surface properties of the target sentences (seen as strings of words), e.g. length, word overlap, etc.
- computing semantic relatedness with WordNet (not a surface feature but not a logical method either).

We may also combine both types of approaches, as done e.g. by Bos & Markert (2005).

Plan for Coming Days

Recognising entailment relies on the ability to select the correct **senses** for the words in the target sentences or texts.

→ this is often left aside in approaches to RTE (*cf. HW1 ex. 2*)

- **Word sense disambiguation** (WSD): the task of determining which sense of a word is being used in a particular context.
 - * we will look into how to approach this task in a couple of weeks.
 - * HW1 ex. 4 – huge ambiguity, but context narrows it down!

Today: what are word senses really?

- Kilgariff's arguments for a distributional notion of word sense.
- Introduction to distributional semantic models (DSMs), aka vector space models (VSMs).

Next week:

- More on properties of DSMs and their evaluation.
- Lenci (2008): philosophical implications of DSMs.

“I don’t believe in word senses”

Adam Kilgarriff (1997) “I don’t believe in word senses”, *Computers and the Humanities*, 31:91-113.

- **Topic under investigation:** the paper tackles a 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.

Kilgarriff's 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.
- Typically, formal compositional semantic have an enumerative view of the lexicon: inventory of word senses or lexemes, plus a mapping between senses and forms. A rather crude notion of word meaning!

$[[bank_1]] = \{x \mid x \text{ is a slope of land adjoining a body of water}\} \quad f : D \rightarrow \{1, 0\}$
 $[[bank_2]] = \{x \mid x \text{ is a business establishment where money is kept}\} \quad f : D \rightarrow \{1, 0\}$

- 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 (homonymy); sense selection;
 - * related senses: polysemy; indeterminacy/vagueness; sense modulation

Kilgarriff's Motivation

Lexical ambiguity: one phonological form, several senses.

- **Homonymy** or *contrastive ambiguity*: accidental ambiguity between unrelated senses; one sense invalidates the other:

- (3) a. Mary walked along the **bank** of the river.
b. ABN-AMRO is the richest **bank** in the city.
- (4) a. Nadia's **plane** taxied to the **terminal**.
b. The central data storage device is served by multiple **terminals**.
c. He disliked the angular **planes** of his cheeks and jaw.

- **Polysemy** or *complementary ambiguity*: ambiguity between semantically related senses that overlap:



- (5) a. John crawled through the **window**.
b. The **window** is closed.
- (6) a. Mary painted the **door**.
b. Mary walked through the **door**
- (7) a. The **bank** raised its interest rates yesterday.
b. The store is next to the newly constructed **bank**.
- (8) a. The **farm** will fail unless we receive the subsidy promised.
b. To **farm** this land would be both foolish and without reward.

- Typically dictionary approach: different lexical entries for **homonymous** senses; **polysemous** senses grouped within one lexical entry.

bank¹  [bɑŋk]  [Show IPA](#)

–noun

1. a long pile or heap; mass: *a bank of earth; a bank of clouds.*
2. a slope or acclivity.
3. *Physical Geography* . the slope immediately bordering a stream course along which the water normally runs.

bank²  [bɑŋk]  [Show IPA](#)

–noun

1. an institution for receiving, lending, exchanging, and safeguarding money and, in some cases, issuing notes and transacting other financial business.
2. the office or quarters of such an institution.

<http://www.dictionary.com/>

- Given this theoretical distinction, it should be possible to classify pairs of examples as instances of 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.

Kilgarriff's Proposal

The author proposes to 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.

“I don’t believe in word senses”

Adam Kilgarriff (1997) “I don’t believe in word senses”, *Computers and the Humanities*, 31:91-113.

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.

Distributional Semantic Models or Vector Space Models

material based on slides by Marco Baroni and Stefan Evert

Distributional Semantic Models

DSMs are motivated by the so-called **Distributional Hypothesis**, which can be stated as follows:

The degree of semantic similarity between two linguistic expressions A and B is a function of the similarity of the linguistic contexts in which A and B can appear. [Z. Harris (1954) *Distributional Structure*]

- There are different types of DSMs, but they all assume a **general model of meaning**:
 - * the distribution of words in context plays a key role in characterising their semantic behaviour;
 - * word meaning depends, at least in part, on the contexts in which words are used \rightsquigarrow *usage-based perspective on meaning*
- DSMs make use of mathematical and computational techniques to turn the informal DH into empirically testable semantic models.

Main idea behind DSMs

- **Count** how many times each target word occurs in a certain context
- Build **vectors** out of (a function of) these context occurrence counts
- Measure the **distance** between vectors: similar words will have similar vectors

Context counts for target word **dog**:

The **dog** barked in the park.
The owner of the **dog** put him
on the leash since he barked.

bark	park	owner	leash
2	1	1	1

General Definition of DSMs

A distributional semantic model (DSM) is a co-occurrence matrix \mathbf{M} where rows correspond to *target terms* and columns correspond to *contexts* or *dimensions*.

	see	use	hear	...
boat	39	23	4	...
cat	58	4	4	...
dog	83	10	42	...

How do we go from *counts* to *vectors*?

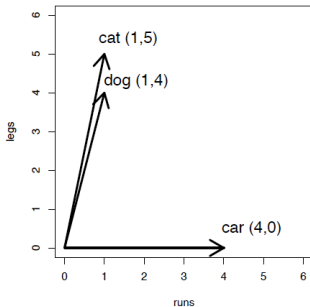
- Distributional vector of 'dog': $x_{dog} = (83, 10, 42, \dots)$
- Each value in the vector is a feature or dimension.

Vectors can be displayed in a *vector space*. This is easier to visualise if we look at two dimensions only, e.g. at two dimensional spaces.

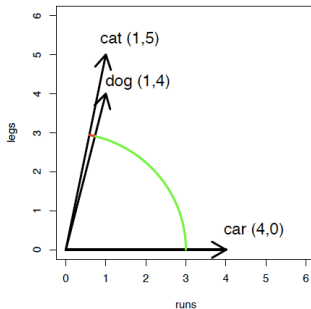
Vectors and Similarity

	run	legs
dog	1	4
cat	1	5
car	4	0

semantic space



semantic similarity as angle between vectors



Some DSM Parameters

- Target terms (rows) and dimensions (columns) can be word forms, lemmas, lemmas with POS tags, ...
 - * the minimum preprocessing required is tokenization
- Size of context where to look for occurrences:
 - * within a window of k words around the target
 - * within a particular linguistic unit:
 - ▶ a sentence
 - ▶ a paragraph
 - ▶ a turn in a conversation
 - ▶ a Webpage

Compare the effect of different term types and window sizes on lists of nearest neighbours with Web Infomap:

<http://clic.cimec.unitn.it/infomap-query/>

What's Next

Next week:

- More details about the properties of DSM, including how they can be evaluated.
- Discussion of the philosophical implications of DSMs based on:
 - * A. Lenci (2008) Distributional Semantics in Linguistic and Cognitive Research, in Lenci (ed.), *From context to meaning: Distributional models of the lexicon in linguistics and cognitive science*, special issue of the *Italian Journal of Linguistics*, 20(1):1-30.

⇒ **Homework #2** is on the website: due on **17 October 2011**