Mechanisms of Meaning
Autumn 2010

Raquel Fernández

Institute for Logic, Language & Computation
University of Amsterdam
Plan for Today

During the next few days, we’ll concentrate on *Distributional Semantic Models*.

- Stefan Evert’s talk: applications of DSMs in Computational Linguistics and technical details of their construction.
- Today we’ll look into the main theoretical claims behind DSMs.
  - Historical notes
  - Key features: theoretical assumptions and their implications
  - Some shortcomings

- Main reference for today’s lecture:

Distributional Semantic Models

As you know well by now, 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* ]

- DSMs make use of mathematical and computational techniques to turn the informal DH into empirically testable semantic models.
- There are different types of DSMs (given the variety of methods available), 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 ⇝ *usage-based perspective on meaning*

- The aim of today’s lecture is to try to understand better what’s behind this general model and what its potential problems are.
Historical Notes
Origins of Distributional Semantics

• Currently, distributional semantics is especially popular in computational linguistics.

• However, its origins are grounded in the linguistic tradition:
  * American structural linguistics during the 1940s and 50s, especially the figure of Zellig Harris (influenced by Sapir and Bloomfield).

• Harris proposed the method of *distributional analysis* as a scientific methodology for linguistics:
  * introduced for phonology, then methodology for all linguistic levels.

• Structuralists don’t consider meaning an *explanans* in linguistics: too subjective and vague a notion to be methodologically sound.
  * linguistic units need to be determined by formal means: by their distributional structure.

• Harris goes one step farther and claims that *distributions* should be taken as an *explanans* for meaning itself.
  * only this can turn semantics into a proper part of the linguistic science.
Beyond Structuralism

Some traditions that developed after Structuralism are critical of DS:

- **Generative linguistics**: focus on I-language — internalised competence of ideal speakers — and dismissal of language use.
- **Formal semantics**: model-theoretic and referential tradition, focus on denotational semantics; meaning is anchored in the world, not language-internal.

In contrast, other traditions embrace DS:

- **Corpus linguistics and lexicography**: distributional semantics is the main methodological principle for semantic analysis.
- **Psychology**: Contextual Hypothesis by Miller and Charles (1991) distributions as a way to explain cognitive semantic representations and how they are built by learners.
Some traditions that developed after Structuralism are critical of DS:

- **Generative linguistics**: focus on I-language — internalised competence of ideal speakers — and dismissal of language use.
- **Formal semantics**: model-theoretic and referential tradition, focus on denotational semantics; meaning is anchored in the world, not language-internal.

In contrast, other traditions embrace DS:
- **Corpus linguistics and lexicography**: distributional semantics is the main methodological principle for semantic analysis. *召回我们讨论过的Kiilgarriff的论文。
- **Psychology**: *Contextual Hypothesis* by Miller and Charles (1991) distributions as a way to explain cognitive semantic representations and how they are built by learners.
Key Features of Distributional Semantics
Again, the main general assumption behind DSMs is that word meaning depends on the contexts in which words are used.

There are three main aspects that characterise distributional semantic representations and make them very different from representations in lexical and formal semantics. They are:

• inherently context-based and hence context-dependent
  * the linguistic contexts in which words are observed enter into their semantic constitution;

• inherently distributed and dynamic
  * meaning derives from the way a word interacts with different contexts (dimensions) - from its global distributional history, which is constantly evolving;

• inherently quantitative and gradual
  * meaning is represented in terms of statistical distribution in various linguistic contexts.
Other important aspects linked to DSMs:

- **Use of linguistic corpora**: Currently DS is corpus-based, however DS ≠ corpus linguistics: the DH is not by definition restricted to linguistic context
  * but current corpus-based methods are more advanced than available methods to process extra-linguistic context.
  * corpus-based methods allow us to investigate how linguistic context shapes meaning.

- **Use of statistical techniques**: Statistical and mathematical techniques are key tools for DS:
  * used to create an abstract contextual representation over usages;
  * formal and empirically testable semantics models.
Where does DS stand within the nativist vs. empiricist debate?

- **Nativism**: part of our *language faculty* is innate. The human brain comes equipped with a limited set of choices; when learning, children select the correct options using their parents’ speech, in combination with the context.

- **Empiricism**: emphasis on *learning* from usage. There isn’t an innate language structure, but general and perhaps language-specific learning capabilities part of our cognitive apparatus.
Essence of Distributional Semantics (III)

Where does DS stand within the nativist vs. empiricist debate?

- **Nativism**: part of our *language faculty* is innate. The human brain comes equipped with a limited set of choices; when learning, children select the correct options using their parents’ speech, in combination with the context.

- **Empiricism**: emphasis on *learning* from usage. There isn’t an innate language structure, but general and perhaps language-specific learning capabilities part of our cognitive apparatus.

- Lenci points out that DS is indeed empiricist, but not inherently anti-nativist:
  - some DSMs extract meaning features from raw data;
  - others may include higher level information such as syntax.
The core assumption behind DSMs — once more: *word meaning depends on the contexts in which words are used* — can be interpreted in different ways.

- **Key issue**: is the hypothesized dependency between contexts (word distributions) and semantics (word meaning) simply a *correlation* or is there a *causal* relation between them?

Answers to this question give rise to two versions of the Distributional Hypothesis, which differ on the status they assign to contextual representations:

⇒ “Weak” DH vs. “Strong” DH
The Weak Distributional Hypothesis

- Only assumes a \textit{correlation} between semantic content and contextual distributions:
  \begin{itemize}
  \item by examining distributions and exploiting their correlation with semantics, we get at a better understanding of lexical meaning;
  \item word meaning (whatever this might is) determines the distribution of words in context;
  \item we can try to uncover semantic content by inspecting a significant number of distributions.
  \end{itemize}

- This weak version of the hypothesis is compatible with different research programmes:
  \begin{itemize}
  \item find paradigmatic classes of e.g. verbs
  \item empirical foundations to Generative Lexicon notions
  \item method to derive conceptual schemata, mentioned by Murphy
  \end{itemize}
The Strong Distributional Hypothesis

• Assumes that distributions have a *causal role* in the creation of semantic representations at a cognitive level:
  * the distributional behaviour of a word is a way to *explain* its semantic content;
  * the environments where a word appears have an effect on its cognitive semantic representation.

• Evidence for this strong version of the hypothesis comes from the possibility of modelling psychological phenomena with distributional representations, such as:
  * human similarity judgements
  * child lexical development
  * semantic priming
Some Shortcomings of DSMs
The strong version of the DH is committed to the cognitive plausibility of DSMs. However, some core aspects of semantics are not satisfactorily treated by these models:

- *semantic relations and lexical entailment*
- *compositionality*
- *reference, symbol grounding and embodiment*

This has raised criticisms: skeptics point out that whatever distributions can tell us about a word, this cannot be its meaning.

**Key issue:** Do this these weak points depend on features of current models, or are they inherent to the essence of the DH?
Knowing the meaning of a word involves recognising the validity of inferences that hold between sentences that include that word:

(1) a. Google *bought* a new company $\rightarrow$ Google *purchased* a new company
    b. Ann drives a *car* $\rightarrow$ Ann drives a *vehicle*

(2) a. Google *purchased* a new company $\rightarrow$ Google *bought* a new company
    b. Ann drives a *vehicle* $\nleftrightarrow\;$ Ann drives a *car*
    c. Ann drives a *car* $\nleftrightarrow\;$ Ann drives a *van*
       Ann drives a *van* $\nleftrightarrow\;$ Ann drives a *car*

• (1a)-(2a) synonymy  (1b)-(2b) hyponymy  (2c) co-hyponymy

DSMs can recognise the semantic relatedness between these words:

Web Infomap [http://clic.cimec.unitn.it/infomap-query/]
Digression: how does Infomap work?

In short: it represents words as vectors encoding information about their distribution in a corpus, and models semantic similarity between words by comparing their vectors.

- it selects a set of content words as dimensions (1000 most frequent words in a corpus);
- it counts how often a word occurs with each of the content words within a particular window of text;
- it uses singular value decomposition to reduce the number of dimensions and select the most informative ones (ca. 100);
- it compares the resulting vectors by measuring their cosine similarity (angle between two vectors) to find the nearest neighbours of a given word.

You can find a more detailed, but still easy to follow description of the Infomap algorithm here: http://infomap-nlp.sourceforge.net/doc/algorith...
However, DSMs cannot represent the fact that distributionally similar words may be similar *under very different semantic respects*.  

- They place words in a common semantic space and measure the distances among them to account for their similarity.  
- The distance between words in a semantic space is a *symmetric relation*:  
  - DSMs are good at recognising synonyms (e.g. TOEFL synonym task), and at modelling semantic similarity judgements and categorisation,  
  - but cannot capture asymmetric relations such as hyperonymy.  

⇒ Semantic inference lies beyond the current capabilities of DSMs  
  - is this a sign of the explanatory limits of the DH or a shortcoming of current models?
• DS is concerned with \textit{lexical meaning} – compositionality is typically not the focus (as with most lexical semantic theories).

• However, arguably any semantic theory should be able to explain how the meaning of a complex expression can be built up from the meanings of its components.

• Can DSMs provide a satisfactory account of compositionality?

• What is needed is a way to \textit{compose distributional information}. But this is not straightforward:

  * Landauer & Dumais (2007) propose \textit{vector summation}: the distributional meaning of \textit{‘the dog bit the man’} is the sum of the vectors of each of the words in the sentence.
Compositionality

• DS is concerned with *lexical meaning* – compositionality is typically not the focus (as with most lexical semantic theories).

• However, arguably any semantic theory should be able to explain how the meaning of a complex expression can be built up from the meanings of its components.

• Can DSMs provide a satisfactory account of compositionality?

• What is needed is a way to *compose distributional information*. But this is not straightforward:
  * Landauer & Dumais (2007) propose vector summation: the distributional meaning of ‘the dog bit the man’ is the sum of the vectors of each of the words in the sentence.
  * But this can’t distinguish it from ‘The man bit the dog’!
Compositionality

• DS is concerned with *lexical meaning* – compositionality is typically not the focus (as with most lexical semantic theories).

• However, arguably any semantic theory should be able to explain how the meaning of a complex expression can be built up from the meanings of its components.

• Can DSMs provide a satisfactory account of compositionality?

• What is needed is a way to *compose distributional information*. But this is not straightforward:
  * Landauer & Dumais (2007) propose *vector summation*: the distributional meaning of ‘the dog bit the man’ is the sum of the vectors of each of the words in the sentence.
  * But this can’t distinguish it from ‘The man bit the dog’!

• Recent approaches adopt more sophisticated models of vector composition that include syntactic dependencies.

• Compositionality is an open issue in current DSM research.
DSM are often refused as plausible cognitive models of meaning for two reasons:

- They are regarded as *ungrounded* symbolic representations. As such they fall under the “symbol grounding problem” (Harnad 1990) and the “Chinese Room argument” (Searle 1980)

  * Chinese Room: thought experiment by John Searle against the possibility of strong AI (Turing test).

  * To the extent that DSMs claim that meaning can be derived by pure symbol manipulation, without direct reference to the world, they are subject to this problem.
Reference and Embodiment

The second reason why DSMs are refused as cognitive models is tied to the *Embodied Cognition Hypothesis*.

- According to the ECH, conceptual representations are grounded in the sensory-motor systems.
  - Concepts/meanings are not amodal, formal symbols but perceptual symbols represented within the perceptual systems we acquire them.
  - Embodied simulations proposed by Barsalou: knowing the meaning of the word ‘turtle’ implies being able to re-enact our perceptual experiences with turtles.
  - Some findings in neuroscience back up these claims.

- To the extent that distributional contextual representations are *not embodied*, they are not cognitively plausible according to this line of research. Linguistic distributions are seen as a *product* of embodied conceptualisation.
Is it right to assume that DS cannot play any substantial role in a cognitive explanation of meaning? According to A. Lenci:

• It doesn’t seem to make sense to completely reduce meaning and concepts to representations grounded in sensory modalities;
• as it probably doesn’t make sense to reduce everything to symbol manipulation.

⇒ Both aspects play a role in the processes leading to meaning formation.

• There is a growing trend of proposing dual models that combine embodied and distributional information.

• These issues are not limited to DSM research: what processes create meaning remains a big open question in philosophy and cognitive science.
Summing Up

• **DH**: Word meaning depends on the context in which words are used.

• We can distinguish between a *weak* and a *strong* version of the DH, depending on the status assigned to contextual representations.

• Regardless of their theoretical status, distributional semantic representations are:
  * inherently *context-based* and hence *context-dependent*
  * inherently *distributed* and *dynamic*
  * inherently *quantitative* and *gradual*

• This makes them interesting candidates for representing meaning.

• However, DSMs still face serious shortcomings:
  * *semantic relations and lexical entailment*
  * *compositionality*
  * *reference and embodiment*
Next Week

- We’ll discuss two recent papers on DSMs:
  - B. Murphy, M. Baroni and M. Poesio (2009) *EEG responds to conceptual stimuli and corpus semantics*, in *Proceedings of EMNLP*. [to be presented by Holger]

- Please do read the papers beforehand!