Distributional semantics for linguists

Lecture 2a: Distributional semantics vs. classical lexical semantics

Aurelie Herbelot

Universität Potsdam
Department Linguistik

ESSLLI 2012

Outline

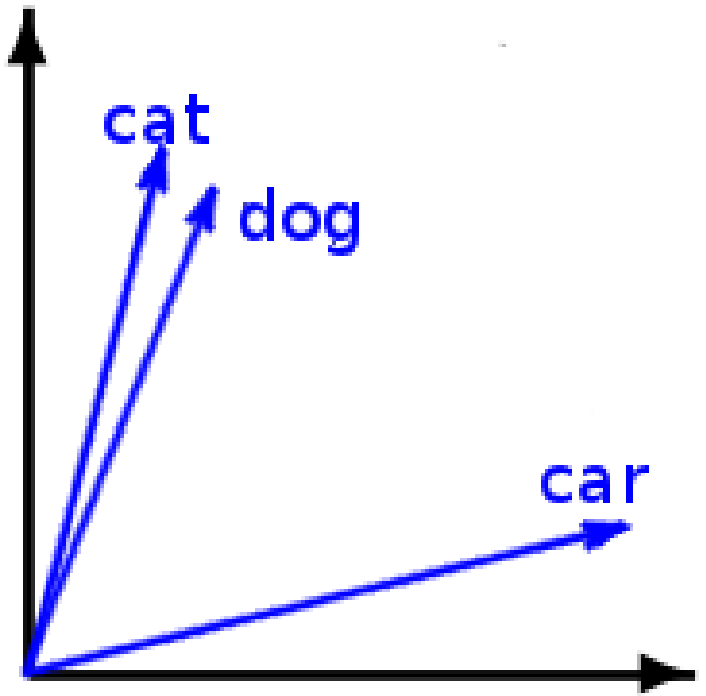
Overview

- How to calculate similarity in a distributional semantic space? What do we mean by ‘similarity’?
- A review of some standard lexical relations: synonymy, antonymy and hyponymy. How can they be translated in distributional terms?
- A summary of the (lack of) explanatory power in current distributional semantics theories.

Outline

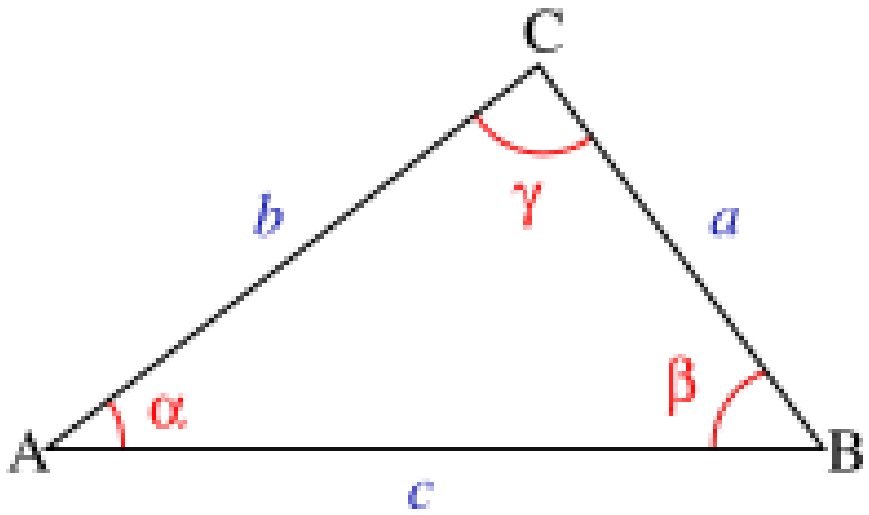
Calculating similarity in a distributional space

- Distributions are vectors, so distance can be calculated.



Some trigonometry

- Law of cosines: $c^2 = a^2 + b^2 - 2ab\cos\gamma$



Measuring similarity

- Cosine:

$$\frac{\sum v1_k * v2_k}{\sqrt{\sum v1_k^2} * \sqrt{\sum v2_k^2}} \quad (1)$$

- The cosine measure calculates the angle between two vectors and is therefore length-independent. This is important, as frequent words have longer vectors than less frequent ones.
- Other measures include Jaccard, Lin... (For an overview: see Weeds, 2004).

Some numbers

- The scale of similarity...
  - house – building 0.428354
  - gem – jewel 0.306866
  - capitalism – communism 0.294677
  - motorcycle – bike 0.29329
  - test – exam 0.269151
  - school – student 0.250291
  - singer – academic 0.168105
  - horse – farm 0.133888
  - man –accident 0.0885102
  - tree – auction 0.0234772
  - cat –county 0.00731196

Example

- Words most similar to *cat*, as chosen from the 5000 most frequent nouns in Wikipedia.

1 cat	0.276042 man	0.230953 person
0.4512 dog	0.275582 cow	0.229124 pet
0.357814 animal	0.264269 fox	0.228973 lizard
0.336883 rat	0.260912 girl	0.228406 chicken
0.331284 rabbit	0.26071 sheep	0.223872 monster
0.329772 pig	0.258142 boy	0.218094 people
0.309073 monkey	0.255272 elephant	0.216812 tiger
0.307839 bird	0.248803 deer	0.215497 mammal
0.302241 horse	0.247423 woman	0.212786 bat
0.296586 mouse	0.245761 fish	0.2122 duck
0.292734 wolf	0.243787 squirrel	0.209441 cattle
0.292047 creature	0.243725 dragon	0.208839 dinosaur
0.287286 human	0.243714 frog	0.207969 character
0.286601 goat	0.234795 baby	0.207257 kid
0.282235 snake	0.233694 child	0.206511 turtle
0.279406 bear	0.231072 lion	0.2049 robot

But what is similarity?

- In distributional semantics, very broad notion. Includes synonyms, near-synonyms, hyponyms, taxonomical siblings, antonyms, etc.
- The broad notion does correlate with a psychological reality. One of the favourite tests of the distributional semantics community is the calculation of correlation between a distributional similarity system and human judgments on the Miller & Charles (1991) test set.

Miller & Charles 1991

3.92 automobile-car	3.05 bird-cock	0.84 forest-graveyard
3.84 journey-voyage	2.97 bird-crane	0.55 monk-slave
3.84 gem-jewel	2.95 implement-tool	0.42 lad-wizard
3.76 boy-lad	2.82 brother-monk	0.42 coast-forest
3.7 coast-shore	1.68 crane-implement	0.13 cord-smile
3.61 asylum-madhouse	1.66 brother-lad	0.11 glass-magician
3.5 magician-wizard	1.16 car-journey	0.08 rooster-voyage
3.42 midday-noon	1.1 monk-oracle	0.08 noon-string
3.11 furnace-stove	0.89 food-rooster	
3.08 food-fruit	0.87 coast-hill	

- Miller & Charles experiment: re-run of Rubenstein & Goodenough (1965). Correlation coefficient = 0.97.

Distributional methods are discursive

- Distributions are a good conceptual representation if you believe that ‘the meaning of a word is given by its usage’.
- Corpus-dependent, culture-dependent, register-dependent. Example: similarity between *policeman* and *cop*: 0.232632.

Distributions are register-dependent

policeman
0.586482::ball_n+poss_rel
0.47911::and_c+civilian_n
0.424271::soldier_n+and_c
0.409217::and_c+soldier_n
0.384081::secret_a
0.370919::people_n+include_v
0.36834::corrupt_a
0.358544::uniformed_a
0.352538::uniform_n+poss_rel
0.349553::civilian_n+and_c
0.315058::iraqi_a
0.311442::lot_n+poss_rel
0.307535::chechen_a
0.303514::laugh_v
0.286281::and_c+criminal_n
0.285162::incompetent_a
0.284202::pron_rel_+shoot_v
0.279526::hat_n+poss_rel
0.276776::terrorist_n+and_c
0.272654::and_c+crowd_n
0.271465::military_a

cop
0.450031::crooked_a
0.448631::corrupt_a
0.439307::maniac_a
0.380065::dirty_a
0.373174::honest_a
0.357623::uniformed_a
0.350859::tough_a
0.327847::pron_rel_+call_v
0.320139::funky_a
0.317952::bad_a
0.29243::veteran_a
0.290737::and_c+robot_n
0.285521::and_c+criminal_n
0.279318::bogus_a
0.276689::talk_v+to_p()+pron_rel_
0.272944::investigate_v+murder_n
0.257574::on_p()+force_n
0.251643::parody_n+of_p()
0.249137::Mason_n+and_c
0.246172::pron_rel_+kill_v
0.246089::racist_a

Outline

Extension

- In set-theoretic semantics, the meaning or **extension** of *cat*, cat′, is the set of all cats in some world.
- Sets intersect, so the meaning of *black cat* is $\text{cat}'(x) \wedge \text{black}'(x)$, the intersection of the set of cats and the set of black things.
- Some entities will be in several sets.

Outline

The classical account

- Difference between full synonymy (*eggplant/aubergine*) and near-synonymy (*city, town*).
- The extensions of two full synonyms are identical sets. $\text{eggplant}' = \text{aubergine}'$
- The extensions of two near synonyms have a high (whatever that means...) overlap. i.e. with respect to a specific context, near-synonyms will often be substitutable.

Some facts about synonymy

- Near-synonymy is frequent, absolute synonymy relates to dialect etc. (*eggplant/aubergine*)
- Word sense assumptions affect synonymy assumptions.
- Language learners tend to assume non-synonymy.
  e.g., “labeling entities with distinct words leads infants to create representations of two distinct individuals” (Carey, 2009:p 277)

Near-synonymy and meaning acquisition

- Readers only need a few uses to obtain a working idea of a new word's meaning. (Rice, 1990)
- Hypothesis: understanding a new word (without definition) can be modelled by two-phase comparison:
  - initial approximation: e.g., *rancid* is similar to *off*
  - acquisition of differentiating information **characteristic contexts**: e.g., *rancid* tends to appear with fatty foods (or dairy foods, or . . .)
- People's beliefs about low-to-medium frequency words may differ but approximation is usually good enough for communication.

Are *frumpy* and *dowdy* synonyms?

Ann's intuition (pre data check): both negative, both refer to women/women's clothing, *dowdy* implies *dull*, *frumpy* implies *tasteless*.
BNC:

- frumpy: 17 total. 8 clothing, 9 people.
- dowdy: 73 total. 35% people, 10% clothing, 20% abstract, 15% location/organisation.
- Conjoined adjectives
  frumpy: *old* (twice), *new*
  dowdy: plain; solid; nondescript; gauche; second-rate; unkempt; unpleasant, stupid

slightly dowdy elegance — if there could be such a thing

Are *frumpy* and *dowdy* synonyms?

Ann's intuition (pre data check): both negative, both refer to women/women's clothing, *dowdy* implies *dull*, *frumpy* implies *tasteless*.

BNC:

- frumpy: 17 total. 8 clothing, 9 people.
- dowdy: 73 total. 35% people, 10% clothing, 20% abstract, 15% location/organisation.
- Conjoined adjectives
  frumpy: *old* (twice), *new*
  dowdy: plain; solid; nondescript; gauche; second-rate; unkempt; unpleasant, stupid

slightly dowdy elegance — if there could be such a thing

Are *frumpy* and *dowdy* synonyms?

Ann's intuition (pre data check): both negative, both refer to women/women's clothing, *dowdy* implies *dull*, *frumpy* implies *tasteless*.
BNC:

- frumpy: 17 total. 8 clothing, 9 people.
- dowdy: 73 total. 35% people, 10% clothing, 20% abstract, 15% location/organisation.
- Conjoined adjectives
  frumpy: *old* (twice), *new*
  dowdy: plain; solid; nondescript; gauche; second-rate; unkempt; unpleasant, stupid

slightly dowdy elegance — if there could be such a thing

Are *frumpy* and *dowdy* synonyms?

Ann's intuition (pre data check): both negative, both refer to women/women's clothing, *dowdy* implies *dull*, *frumpy* implies *tasteless*.
BNC:

- frumpy: 17 total. 8 clothing, 9 people.
- dowdy: 73 total. 35% people, 10% clothing, 20% abstract, 15% location/organisation.
- Conjoined adjectives
  frumpy: *old* (twice), *new*
  dowdy: plain; solid; nondescript; gauche; second-rate; unkempt; unpleasant, stupid

slightly dowdy elegance — if there could be such a thing

Are *frumpy* and *dowdy* synonyms?

Ann's intuition (pre data check): both negative, both refer to women/women's clothing, *dowdy* implies *dull*, *frumpy* implies *tasteless*.
BNC:

- frumpy: 17 total. 8 clothing, 9 people.
- dowdy: 73 total. 35% people, 10% clothing, 20% abstract, 15% location/organisation.
- Conjoined adjectives
  frumpy: *old* (twice), *new*
  dowdy: plain; solid; nondescript; gauche; second-rate; unkempt; unpleasant, stupid

slightly dowdy elegance — if there could be such a thing

Full synonymy and meaning acquisition

- Full synonyms are probably acquired differently from near-synonyms, generally by (relatively) explicit definition:

  The aubergine (eggplant) has to be one of my favourite vegetables.

- Full synonyms may be different vocalisations for the same concept (their lexemes share a single semantic functional web in the brain).
- Contrast with near-synonyms which are separate concepts.

Frequency and synonymy

- Speakers use the most frequent term in their experience to convey a particular idea (frequency assumed to correlate with strength of neural connections).
- More frequent words tend to have broader meanings (more 'senses' . . .)
- Two words of very different frequency are unlikely to cover exactly the same semantic space.
- Many words are of too low frequency for hearers to make reliable decisions about synonymy.

Synonymy: requisites for an ideal distributional account

- Distinguishing between near-synonyms and full synonyms.
- No hard line between near-synonyms and non-synonyms.
- Degree of synonymy between two lexemes will vary between individuals.

The distribution of synonyms

- Similarity between *egglant/aubergine*: 0.114024
  Relatively low cosine. Partly due to frequency (222 for *eggplant*, 56 for *aubergine*).
- Similarity between *policeman/cop*: 0.232632
  To be expected: *policeman* and *cop* are discursively very different.
- Similarity between *city/town*: 0.735319
- So... similarity does not tell us how to distinguish between full and near-synonymy.

Outline

The classical account

- Three basic types of antonymy:
  - gradable (opposite ends of a scale: *cold/hot*, modifiable with *very*, etc)
  - non-gradable (discrete opposition: *dead/alive*)
  - multiple (non-gradable, discontinuous scale: *lecturer, reader, professor*).
- In terms of extension: the same entity cannot be described as both X and its antonym Y in a given situation. i.e. for a micro-world corresponding to a situation where I drink tea, the tea cannot be in the set of cold things and in the set of hot things.

Distributions of antonyms

- Similarities between:
  - cold/hot 0.287398
  - dead/alive 0.242078
  - large/small 0.6783
  - colonel/general 0.333739

Identifying antonyms

- Antonyms have a high distributional similarity. It is hard to distinguish them from near-synonyms.
- The identification of antonyms usually requires some heuristics to be applied to pairs of highly similar distributions.
- For instance, it has been observed that antonyms are frequently coordinated while synonyms are not:
  - a selection of cold and hot drinks
  - wanted dead or alive
  - lectures, readers and professors are invited to attend

Outline

The classical account

- Relationship between a more general term and a more specific term (*dog/poodle*).
- The extension of the more general includes the extension of the more specific (all poodles are dogs).
- The intension of the more specific includes the intension of the more general (all that can be said about dogs can be said about poodles)... in an essentialist account (see penguins).

Distributions of hyponyms

- No clear inclusion relationship. The set of contexts recorded for *cat* and *animal* overlap, but they are by no means subsets.
- Kotlerman et al (2010), however, demonstrated that in general, if X is a hyponym of Y, features with high values in X tend to have a high value in Y.
- Baroni et al (2012) learn hyponymy from distributions for adjective-noun phrases (a black cat is a cat). But they do not report on the features used by the classifier.
- Similarity between *cat* and *animal*: 0.357814.

Outline

Issues

- There is no formal definition for the standard lexical relations in distributional semantics.
- The standard definitions rely on the idea of extension, but there is no obvious correspondence between the corpora used to produce distributions and the real world.

Outline

Conclusion

- A vague notion of similarity can be captured from distributions.
- Standard lexical relations tend to have medium to high similarity but there is very much variations, some of which due to frequency effects, some of which idiosyncratic.
- Identifying pairs of words in a specific lexical relation relies on heuristics.
- There is, in the standard distributional account, no formal specification for lexical relations.