Distributional semantics for linguists: 2b

Ann Copestake and Aurélie Herbelot

Computer Laboratory, University of Cambridge
and
Department Linguistik, Universität Potsdam

August 2012
Session 2b: Outline

Polysemy and sense induction

Collocation and multiword expressions

Some linguistic applications of distributional semantics
  Adjective and binomial ordering
  English compound noun relations
Outline.

Polysemy and sense induction

Collocation and multiword expressions

Some linguistic applications of distributional semantics
  Adjective and binomial ordering
  English compound noun relations
Polysemy

Distribution for *pot*, obtained from Wikipedia.

0.566454::melt_v
0.442374::pron_rel_+smoke_v
0.434682::of_p()+gold_n
0.40773::porous_a
0.401654::of_p()+tea_n
0.39444::player_n+win_v
0.393812::money_n+in_p()
0.39317::amount_n+in_p()
0.329211::ceramic_a
0.326387::hot_a
0.323321::boil_v
0.313404::bowl_n+and_c
0.306324::ingredient_n+in_p()
0.301916::plant_n+in_p()
0.298764::simmer_v
0.292397::pot_n+and_c
0.284539::bottom_n+of_p()
0.28338::of_p()+flower_n
0.279412::of_p()+water_n
0.278914::food_n+in_p()
0.262501::pron_rel_+heat_v
0.260375::size_n+of_p()
0.25511::pron_rel_+split_v
0.254363::of_p()+money_n
0.2535::of_p()+culture_n
0.249626::player_n+take_v
0.246479::in_p()+hole_n
0.244051::of_p()+soil_n
0.243797::city_n+become_v
Polysemy

- Distribution for *drug*, obtained from Wikipedia.

0.608869::and_c+alcohol_n 0.397089::of_p()+abuse_n
0.510397::alcohol_n+and_c 0.39542::war_n+on_p()
0.464624::or_c+substance_n 0.393311::dose_n+of_p()
0.462777::alcohol_n+or_c 0.386679::metabolism_n+of_p()
0.451267::over-the-counter_a 0.369514::and_c+crime_n
0.451249::inflammatory_a 0.36857::effect_n+poss_rel
0.448604::food_n+and_c 0.366681::of_p()+choice_n
0.445496::addictive_a 0.365335::and_c+substance_n
0.428868::and_c+prostitution_n 0.364455::drug_n+be_v
0.42017::illegal_a 0.360401::anti-_a
0.41921::recreational_a 0.359099::generic_a
0.417316::have_v+side_effect_n 0.358552::overdose_n+of_p()
0.408879::like_p()+Me_n 0.358029::treatment_n+with_p()
0.402512::side_effect_n+of_p() 0.35767::prostitution_n+and_c
0.400139::intravenous_a 0.35661::diabetic_a
Polysemy

- Distribution for *soft*, obtained from Wikipedia.

- 0.624533::plump_a
- 0.624433::drink_n
- 0.609981::plumage_n
- 0.588074::fluffy_a
- 0.547627::uneven_a
- 0.540281::silky_a
- 0.51885::palate_n
- 0.50562::tissue_n
- 0.477878::spine_n+and_c
- 0.453215::colourful_a
- 0.444027::hand-off_n
- 0.413344::pretzel_n
- 0.40609::call_n+be_v
- 0.388752::Cell_n
- 0.387858::feather_n
- 0.387565::and_c+tail_n
- 0.379231::become_v+and_c
- 0.377516::paste_n
- 0.373097::ray_n
- 0.372154::spot_n
- 0.367734::coral_n
- 0.362632::dorsal_a
- 0.361666::reboot_n
- 0.359202::acidic_a
- 0.358819::texture_n
- 0.358372::and_c+snack_n
- 0.352847::beer_n+and_c
- 0.348029::erosion_n+of_p()
- 0.346968::fleshy_a
- 0.344807::porn_n
Sense induction

Normally, single point in vector space represents all uses.

- Sense induction: cluster contexts and associate new instances with a cluster (contrast word sense disambiguation, where prior list of word senses).
- Different senses for each word (contrast topic clustering, where words are associated with a global set of topics).
- Early work by Neill (2002): automatically discovers ‘seed’ words which discriminate between clusters.
- Clusters are more discrete for homonyms compared to general polysemy: some uses in between senses?
- Current applications tend not to distinguish senses.
- More on Thursday on regular polysemy.
Outline.

Polysemy and sense induction

Collocation and multiword expressions

Some linguistic applications of distributional semantics

  Adjective and binomial ordering

  English compound noun relations
Multiword expressions (MWEs)

- ‘words with spaces’: e.g., *ad hoc* (in English!)
- non-decomposable: e.g., *kick the bucket*
- decomposable but non-compositional: e.g., *cat out of the bag* (meaning ‘secret out of hiding place’)
- idioms of encoding/collocations: e.g., *heavy shower*

MWEs and distributions:

- MWEs might be expected to obscure distributional meaning.
- But: ranking of contexts by PMI very similar to techniques for finding MWEs!
- and higher associations suggest lower compositionality.
Magnitude adjectives and non-physical-solid nouns. (Copestake, 2005)

Distributional data from the British National Corpus (100 million words)

<table>
<thead>
<tr>
<th></th>
<th>importance</th>
<th>success</th>
<th>majority</th>
<th>number</th>
<th>proportion</th>
<th>quality</th>
<th>role</th>
<th>problem</th>
<th>part</th>
<th>winds</th>
<th>support</th>
<th>rain</th>
</tr>
</thead>
<tbody>
<tr>
<td>great</td>
<td>310</td>
<td>360</td>
<td>382</td>
<td>172</td>
<td>9</td>
<td>11</td>
<td>3</td>
<td>44</td>
<td>71</td>
<td>0</td>
<td>22</td>
<td>0</td>
</tr>
<tr>
<td>large</td>
<td>1</td>
<td>1</td>
<td>112</td>
<td>1790</td>
<td>404</td>
<td>0</td>
<td>13</td>
<td>10</td>
<td>533</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>high</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>92</td>
<td>501</td>
<td>799</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>90</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>major</td>
<td>62</td>
<td>60</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>272</td>
<td>356</td>
<td>408</td>
<td>1</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>big</td>
<td>0</td>
<td>40</td>
<td>5</td>
<td>11</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>79</td>
<td>79</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>strong</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>8</td>
<td>0</td>
<td>3</td>
<td>132</td>
<td>147</td>
<td>0</td>
</tr>
<tr>
<td>heavy</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>198</td>
</tr>
</tbody>
</table>
### Adjectives: selected examples.

**BNC frequencies:**

<table>
<thead>
<tr>
<th></th>
<th>number</th>
<th>proportion</th>
<th>quality</th>
<th>problem</th>
<th>part</th>
<th>winds</th>
<th>rain</th>
</tr>
</thead>
<tbody>
<tr>
<td>large</td>
<td>1790</td>
<td>404</td>
<td>0</td>
<td>10</td>
<td>533</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>high</td>
<td>92</td>
<td>501</td>
<td>799</td>
<td>0</td>
<td>3</td>
<td>90</td>
<td>0</td>
</tr>
<tr>
<td>big</td>
<td>11</td>
<td>1</td>
<td>0</td>
<td>79</td>
<td>79</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>heavy</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>198</td>
</tr>
</tbody>
</table>

**Acceptability judgements:**

<table>
<thead>
<tr>
<th></th>
<th>number</th>
<th>proportion</th>
<th>quality</th>
<th>problem</th>
<th>part</th>
<th>winds</th>
<th>rain</th>
</tr>
</thead>
<tbody>
<tr>
<td>large</td>
<td></td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>high</td>
<td></td>
<td></td>
<td>*</td>
<td>?</td>
<td></td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>big</td>
<td></td>
<td></td>
<td>?</td>
<td></td>
<td></td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>heavy</td>
<td></td>
<td>?</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>
Magnitude adjective distribution.

- Investigated the distribution of heavy, high, big, large, strong, great, major with the most common co-occurring nouns in the BNC.
- Nouns tend to occur with up to three of these adjectives with high frequency and low or zero frequency with the rest.
- 50 nouns in BNC with the extended use of heavy with frequency 10 or more, 160 such nouns with high. Only 9 with both: price, pressure, investment, demand, rainfall, cost, costs, concentration, taxation
- Clusters: e.g., weather precipitation nouns with heavy. Note heavy shower (weather, not bathroom).
Hypotheses about distribution.

- ‘abstract’ heavy, high, big, large, strong, great, major all denote magnitude (in a way that can be made formally precise)
- distribution differences due to collocation, soft rather than hard constraints
- adjective-noun combination is semi-productive
- denotation and syntax allow heavy esteem etc, but speakers are sensitive to frequencies, prefer more frequent phrases with ‘same’ meaning
## Adjective similarities

<table>
<thead>
<tr>
<th></th>
<th>high</th>
<th>heavy</th>
<th>big</th>
<th>large</th>
<th>strong</th>
<th>major</th>
</tr>
</thead>
<tbody>
<tr>
<td>high</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>heavy</td>
<td>0.22</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>big</td>
<td>0.26</td>
<td>0.22</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>large</td>
<td>0.40</td>
<td>0.30</td>
<td>0.45</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>strong</td>
<td>0.30</td>
<td>0.29</td>
<td>0.30</td>
<td>0.34</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>major</td>
<td>0.31</td>
<td>0.20</td>
<td>0.44</td>
<td>0.45</td>
<td>0.32</td>
<td>-</td>
</tr>
</tbody>
</table>
Outline.

Polysemy and sense induction

Collocation and multiword expressions

Some linguistic applications of distributional semantics
  Adjective and binomial ordering
  English compound noun relations
Applications of distributional semantics

- Many applications in natural language processing: e.g., improving search, processing scientific text, sentiment analysis.
- Also applications in philosophy and sociolinguistics: e.g., Herbelot, von Redecker and Müller (2012) ‘Distributional techniques for philosophical enquiry’ (gender studies and intersectionality).
- Today (very briefly)
  - Adjective and binomial ordering
  - Compound noun relations
- Logical metonymy and sense extension (Thursday)
Adjective and binomial ordering

- gigantic striped box not striped gigantic box
- brandy and soda not soda and brandy, run and hide
- some pairs are irreversible
- rare and novel phrases may be irreversible (sake and grapefruit, armagnac and blackcurrant)
- ordering principles partially semantic
- lots of discussion in literature about gendered examples: e.g., boy and girl
Adjective and binomial ordering: approaches

- adjective (pre-nominal modifier) ordering fairly well studied in CL: data-driven approaches, but still unseen pairs of adjectives. Back-off techniques include positional probabilities (later).
- binomial ordering less studied in CL (but Copestake and Herbelot, 2011)
- Benor and Levy (2006) corpus-based investigation of binomials
  - models include explicit semantic features, based on prior literature
  - e.g., Iconicity and Power
Mixed drinks: Iconicity or Power?

The Gin and Bitters cocktail is made from Gin and Angostura bitters, and served in a chilled cocktail glass.

Gin and Bitters Ingredients
- 3 oz Gin
- 1 tsp Angostura Bitters

Gin and Bitters Instructions
- Add the bitters to a cocktail glass.
- Swirl it around until the glass is fully coated.
- Fill with gin, and enjoy at room temperature.
Binomials and gender

- Male terms tend to precede female (for humans).
- e.g., *men and women* (85%), *boys and girls* (80%), *male and female* (91%) (% from Google ngram).
- Also personal names: e.g., *James and Sarah* (82%).
- Exceptions: *father and mother* (51%), *mothers and fathers* (67%), *ladies and gentlemen* (97%).
- B+L take gender as an example of the Power feature.
- BUT: possible phonological effects (female names tend to have more syllables than male).
- Animal terms often don’t show a clear order: e.g., *stallion and mare* (50%), *stallion and broodmare* (54%), *ram and ewe* (50%), *sow and boar* (51%).
Binomials and gender

- Male terms tend to precede female (for humans).
  - e.g., *men and women*(85%), *boys and girls*(80%), *male and female*(91%) (% from Google ngram).
- Also personal names: e.g., *James and Sarah*(82%).
- Exceptions: *father and mother*(51%), *mothers and fathers*(67%), *ladies and gentlemen*(97%).
- B+L take gender as an example of the Power feature.
- BUT: possible phonological effects (female names tend to have more syllables than male).
- Animal terms often don’t show a clear order: e.g., *stallion and mare*(50%), *stallion and broodmare*(54%), *ram and ewe*(50%), *sow and boar*(51%).
Binomials and gender

- Male terms tend to precede female (for humans).
- e.g., *men and women* (85%), *boys and girls* (80%), *male and female* (91%) (% from Google ngram).
- Also personal names: e.g., *James and Sarah* (82%).
- Exceptions: *father and mother* (51%), *mothers and fathers* (67%), *ladies and gentlemen* (97%).
- B+L take gender as an example of the Power feature.
- BUT: possible phonological effects (female names tend to have more syllables than male).
- Animal terms often don’t show a clear order: e.g., *stallion and mare* (50%), *stallion and broodmare* (54%), *ram and ewe* (50%), *sow and boar* (51%).
Analogical approach to binomial and adjective ordering

- our hypothesis: humans maintain order of known examples, order unseen by semantic similarity with seen
- essentially same model for binomials and adjectives
- baseline is to use positional probabilities (Malouf 2000)
- \( a \prec b \)
  
  \[
  \begin{align*}
  & \text{if } C(a \text{ and } b) > C(b \text{ and } a) \\
  \text{or } & C(a \text{ and } b) = C(a \text{ and } b) \\
  \text{and } & C(a \text{ and } b,C(\text{and } a) > C(b \text{ and } b,C(\text{and } a)
  \end{align*}
  \]

  and conversely for \( b \prec a \)

- e.g., if \textit{tea and biscuits} is known, prefer \textit{tea and scones} over \textit{scones and tea}
Adjective and binomial ordering: Kumar (2012)

- Same type of model used for adjectives and binomials: unseen cases ordered by k-nearest neighbour comparison to seen examples using distributional similarity.
- e.g., if ordering *coffee, cake* compare to all known binomials A and B based on similarities A:coffee, A:cake, B:coffee, B:cake, decide on basis of closest match (best k around 6 or 7).
- Distributions from unparsed WikiWoods data: significantly better than using positional probabilities.
- Expect further improvement using phonological features in addition.
Compound noun relations

- **cheese knife**: knife for cutting cheese
- **steel knife**: knife made of steel
- **kitchen knife**: knife characteristically used in the kitchen

Automatic disambiguation:

- Syntactic parsers can’t distinguish: \( N1(x), N2(y), \) compound\((x,y)\)
- One approach: human annotation of compounds, use distributional techniques to compare unseen to seen examples.
Compound noun relation schemes

- Ó Séaghdha, 2007: BE, HAVE, INST, ACTOR, IN, ABOUT: (with subclasses) LEX: lexicalised, REL: weird, MISTAG: not a noun compound.
  - Based on Levi (1978)
  - Considerable experimentation to define a usable scheme: some classes very rare (therefore not annotated reliably)
  - Annotation of 1400 examples from BNC by two annotators.
Compound noun relation learning

- BE
  - pine tree
  - tuna fish
  - steel knife

- LEX
  - home secretary

- company president
- tuberculosis
- machine learning

- honey bee
- have
- pork pie
- car door
- pine cone
- fairy tale
- crime investigation

- IN
  - forest hut
  - midnight mass

- ABOUT
  - cheese knife
  - rice cooker

- ACTOR
- LEX

- machine learning

Compound noun relation learning

- BE
  - pine tree
  - tuna fish
  - steel knife

- ACTOR
  - company president
  - forest hut

- IN
  - midnight mass

- ABOUT
  - fairy tale
  - crime investigation

- HAVE
  - pork pie
  - car door
  - pine cone

- INST
  - cheese knife
  - rice cooker
  - machine learning

- LEX

- squirrel pasty?
Squirrels and pasties
Compound noun relation learning

- Ó Séaghdha, 2008 (also Ó Séaghdha and Copestake, forthcoming)
- Treat compounds as single words: doesn’t work!
- Constituent similarity: compounds $x_1 \ x_2$ and $y_1 \ y_2$, compare $x_1$ vs $y_1$ and $x_2$ vs $y_2$.
  - *squirrel* vs *pork*, *pasty* vs *pie*
- Relational similarity: **sentences** with $x_1$ and $x_2$ vs sentences with $y_1$ and $y_2$.
  - *squirrel is very tasty, especially in a pasty* vs *pies are filled with tasty pork*
- Comparison using **kernel methods**: allows combination of kernels.
- Best accuracy: about 65% (slightly lower than agreement between annotators) using combined kernels.
Summary

- Both applications described depend on using distributional similarity to match known cases: a type of analogical reasoning.
- Known examples may be explicitly annotated (this approach to compounds) or based on observation (adjectives and binomials).
- Techniques can be simple (k-nearest neighbours) or more complex (Ó Séaghdha’s use of kernel methods).
- Range of other possible applications — we will return to some of these on Thursday.