‘Deeper’ distributional semantics

Aurelie Herbelot

1Universität Potsdam
Department Linguistik

July 2012
Introduction

1 Producing distributions from the ERG

2 The semantics of adjectives
   - Adjective types
   - Obtaining adjective types from distributions

3 The semantics of quantifiers

4 Conclusion
‘Deeper’ distributional semantics

- Can we do linguistic analysis using distributions?
- Can we improve DELPH-IN tools and resources in the process?
Outline

1. Introduction

2. Producing distributions from the ERG

3. The semantics of adjectives
   - Adjective types
   - Obtaining adjective types from distributions

4. The semantics of quantifiers

5. Conclusion
The corpus

- wikiwoods, converted into DMRS format...
- ... and further processed to get ‘lemmatised’ links.

→ original_a ARG1 drummer_n
Pre-processing

- Nominalisations
- Compounds: fish \textit{compound\_rel} knife becomes fish\_knife
- Coordination: precision issue, we don’t know which predicates are distributive and which are collective.
Which relations?

- Adjective + noun
- Intransitive verb + ARG1
- Transitive verb + ARG1/ARG2
- Ditransitive verb + ARG1/ARG2/ARG3
- Adverb + verb
- Adverb + adjective
- Preposition + ARG1 (noun)/ARG2
- Preposition + ARG1 (verb, with dependents)/ARG2
- Poss_rel + ARG1/ARG2
- Coordination + ARG1/ARG2
Example: language

0.541816::other+than_p()+English_n
0.525895::English_n+as_p()
0.523398::English_n+be_v
0.48977::english_a
0.481964::and_c+literature_n
0.476664::people_n+speak_v
0.468399::French_n+be_v
0.463604::Spanish_n+be_v
0.463591::and_c+dialects_n
0.452107::grammar_n+of_p()
0.445994::foreign_a
0.445071::germanic_a
0.439558::German_n+be_v
0.436135::of_p()+instruction_n
0.435633::speaker_n+of_p()
0.423595::generic_entity_rel_+speak_v
0.42313::pron_rel_+speak_v
0.42294::colon_v+English_n
0.419646::be_v+English_n
0.418535::language_n+be_v
0.4159::and_c+culture_n
0.410987::arabic_a
0.408387::dialects_n+of_p()
0.399266::part_of_rel_+speak_v
0.397::percent_n+speak_v
0.39328::spanish_a
0.39273::welsh_a
0.391575::tonal_a
Problem

- Due to the weighting function (PMI), parts of fixed expressions and named entities are high up in the distribution.
- The cases related to named entities could be easily weeded out if named entity tagging was provided in the ERG parse.
The semantics of adjectives

Outline

1. Introduction

2. Producing distributions from the ERG

3. The semantics of adjectives
   - Adjective types
   - Obtaining adjective types from distributions

4. The semantics of quantifiers

5. Conclusion
Adjective types, Partee (1995)

- **Intersective:** carnivorous mammal
  \[ ||\text{carnivorous mammal}|| = ||\text{carnivorous}|| \cap ||\text{mammal}|| \]

- **Subsective:** skilful surgeon
  \[ ||\text{skilful surgeon}|| \subseteq ||\text{surgeon}|| \]

- **Non-subsective:** former senator
  \[ ||\text{former senator}|| \neq ||\text{former}|| \cap ||\text{senator}|| \]
  \[ ||\text{former senator}|| \not\subseteq ||\text{senator}|| \]
Integrating adjective types in the ERG

- The MRS of *skilful surgeon* shouldn’t be
  \[ l1: \text{skilful}(x) \]
  \[ l2: \text{surgeon}(x) \]
  ... because \( x \) is not ‘overall’ skilful.

- Similarly, the current MRSs for *former, fake*, etc. are semantically inappropriate.
The semantics of adjectives

Adjective types

Skilful

The skillful surgeon put Kim's head back on his shoulders.

[5 of 5 analyses; processing time: 0.44 seconds; 1087 edges]
The semantics of adjectives

Adjective types

Former

The former president badly needs a job.

[1 of 1 analysis; processing time: 0.06 seconds; 114 edges]
Extra complication

The semantics of *big city* should definitely be

\[ l1: \text{big}(x) \]
\[ l2: \text{city}(x) \]

... but lexically, there is more going on.

Distributional intersective composition misses out on:

*loud, underground, advertisement, crowd, Phantom of the Opera*...
Spotting non-intersective adjectives

- Hypothesis: the distributional meaning of non-intersective adjectives is not found in the phrases they appear in.
- That is... the cosine between skilful+surgeon° and skilful° should be fairly low.
Adjective distributions

- The nouns in ARG1 position?
- But then... no way to compare the distribution of the adjective with the distribution of an adjectival phrase.
- Instead: first assume all adjectives are intersective. Their semantic context is the semantic context of the nouns they modify.
Trying it out

- Looking at the 20 most frequent adjectives which occur with at least 10 different phrases of frequency >100.
- We record the average cosine between the adjective and the phrases it occurs in.
- Results:

  0.21287 late_a
  0.20550 old_a
  0.20047 large_a
  0.19687 former_a
  0.19649 original_a
  0.19338 early_a
  0.18843 small_a
  0.18591 only_a
  0.18134 national_a
  0.18046 general_a
  0.18000 high_a
  0.17931 american_a
  0.17749 great_a
  0.17717 same_a
  0.17277 main_a
  0.17113 good_a
  0.16459 other_a
  0.15379 several_a
  0.14607 new_a
  0.13859 current_a
Looking at individual phrases

- 0.333932 american_a+actor_n
- 0.109199 american_a+city_n
- 0.30784 early_a+1990s_n
- 0.116951 early_a+education_n
- 0.300824 former_a+member_n
- 0.0913057 former_a+champion_n
- 0.338689 good_a+friend_n
- 0.167788 good_a+man_n
### Different uses of a single adjective?

<table>
<thead>
<tr>
<th>Value</th>
<th>Value</th>
<th>Value</th>
<th>Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.263114</td>
<td>0.58895</td>
<td>0.368887</td>
<td>early_a+1970s_n</td>
</tr>
<tr>
<td>0.269555</td>
<td>0.600884</td>
<td>0.375395</td>
<td>early_a+1980s_n</td>
</tr>
<tr>
<td>0.30784</td>
<td>0.689216</td>
<td>0.365488</td>
<td>early_a+1990s_n</td>
</tr>
<tr>
<td>0.224138</td>
<td>0.446551</td>
<td>0.263564</td>
<td>early_a+age_n</td>
</tr>
<tr>
<td>0.0840708</td>
<td>0.212068</td>
<td>0.245176</td>
<td>early_a+attempt_n</td>
</tr>
<tr>
<td>0.216997</td>
<td>0.383286</td>
<td>0.253161</td>
<td>early_a+career_n</td>
</tr>
<tr>
<td>0.330545</td>
<td>0.328818</td>
<td>0.231219</td>
<td>early_a+century_n</td>
</tr>
<tr>
<td>0.154142</td>
<td>0.251523</td>
<td>0.237991</td>
<td>early_a+church_n</td>
</tr>
<tr>
<td>0.116951</td>
<td>0.239622</td>
<td>0.19837</td>
<td>early_a+education_n</td>
</tr>
<tr>
<td>0.130874</td>
<td>0.330921</td>
<td>0.199711</td>
<td>early_a+example_n</td>
</tr>
<tr>
<td>0.109178</td>
<td>0.187463</td>
<td>0.2937</td>
<td>early_a+form_n</td>
</tr>
<tr>
<td>0.233363</td>
<td>0.363116</td>
<td>0.345782</td>
<td>early_a+history_n</td>
</tr>
<tr>
<td>0.0373053</td>
<td>0.204327</td>
<td>0.13131</td>
<td>early_a+lead_n</td>
</tr>
<tr>
<td>0.25244</td>
<td>0.327949</td>
<td>0.313218</td>
<td>early_a+life_n</td>
</tr>
<tr>
<td>0.222114</td>
<td>0.342128</td>
<td>0.330715</td>
<td>early_a+period_n</td>
</tr>
<tr>
<td>0.123098</td>
<td>0.173442</td>
<td>0.201566</td>
<td>early_a+record_n</td>
</tr>
<tr>
<td>0.134532</td>
<td>0.343616</td>
<td>0.177605</td>
<td>early_a+reference_n</td>
</tr>
<tr>
<td>0.154835</td>
<td>0.363332</td>
<td>0.19154</td>
<td>early_a+settlement_n</td>
</tr>
<tr>
<td>0.161119</td>
<td>0.534706</td>
<td>0.159885</td>
<td>early_a+settler_n</td>
</tr>
<tr>
<td>0.121327</td>
<td>0.269149</td>
<td>0.25522</td>
<td>early_a+success_n</td>
</tr>
</tbody>
</table>
Clustering different adjective behaviours

Does the behaviour of adjectives differ depending on the type of noun they modify?

For each adjective, we cluster the nouns it modifies using three features:

- The distance of the adjective’s distribution to the phrase’s distribution
- The distance of the modified noun to the phrase’s distribution
- The distance of the adjective to the noun (distributions that are close indicate a high frequency of cooccurrence).
Examples

- **American**:  
  - student man **group organisation leader** (0.132, 0.1981, 0.2677)  
  - university school force community woman **music film culture history** (0.1857, 0.3172, 0.251)  
  - association society **musician artist author writer actress actor** (0.2754, 0.408, 0.3168)  
  - league **tribe ancestry population** (0.1156, 0.3579, 0.1735)  
  - team city version company game family life (0.12, 0.1871, 0.2044)
Examples

- Early:
  - career life age period century history year (0.2642, 0.3448, 0.2793)
  - education church record version (0.1362, 0.2417, 0.2259)
  - attempt success form (0.1049, 0.2229, 0.2647)
  - 1970s 1980s 1990s work (0.2802, 0.5696, 0.3748)
  - lead example reference settlement settler (0.1237, 0.3554, 0.172)
Examples

- Good:
  - actress actor school year (0.1431, 0.1996, 0.1765)
  - film album team player example (0.1938, 0.269, 0.2311)
  - friend (0.3387, 0.5939, 0.3378)
  - language (0.0279, 0.0936, 0.1774)
  - idea way man life place work thing record song (0.1683, 0.2201, 0.2877)
Good language

Latin is a good language for learning cases.
Examples

- High:
  - speed cost rank quality court rate mountain peak standard education (0.2047, 0.4547, 0.2503)
  - ground value degree position honour number point (0.18, 0.2847, 0.2571)
  - command priest street pressure frequency price award (0.1131, 0.3231, 0.1758)
  - commissioner risk rating percentage temperature score proportion concentration (0.1555, 0.4633, 0.185)
  - level school (0.4696, 0.6425, 0.4406)
First thoughts

- Don’t talk about intersective versus subsective/privative adjectives, but about intersective/subsective/privative *uses* of adjectives.
- Identify (semi-)fixed phrases (high school, high level): should be single lexical items??
- Adjectives with (mostly) flat distribution in the ‘difference’ space are *not* intersective.
- Low cosines between AN, A and N indicate anomaly in the semantics of an adjective (??)
Outline

1. Introduction
2. Producing distributions from the ERG
3. The semantics of adjectives
   - Adjective types
   - Obtaining adjective types from distributions
4. The semantics of quantifiers
5. Conclusion
Counting with distributions
Quantification and LC

- Because LC is entirely compatible with model-theoretic semantics, we can quantify in the usual way...
- ... and do more...

The heffalump

Heffalumps eat grass. They are striped and have a long tail, as well as a trunk.

**True or false:** All heffalumps are animals. Most heffalumps live underwater. Some heffalumps are blind. All heffalumps are blind.

- Impossible to calculate probabilities... this cannot be treated in a pure model-theoretic setting.
- But we have lexical information. This let us resolve cases of underspecified quantification like *Heffalumps live in forests.* (Some, most or all?)
Outline

1. Introduction
2. Producing distributions from the ERG
3. The semantics of adjectives
   - Adjective types
   - Obtaining adjective types from distributions
4. The semantics of quantifiers
5. Conclusion
We can get nice distributions out of wikiwoods.

It may be worth investigating ‘deeper’ lexical semantics issues under the microscope of distributions.

Classical problems like quantification have the potential of being resolved beyond the level of models and truth.

One day... integrate correct representations for adjectives in our grammars.

Disambiguate quantification in the parse?
Conclusion

- We can get nice distributions out of wikiwoods.
- It may be worth investigating ‘deeper’ lexical semantics issues under the microscope of distributions.
- Classical problems like quantification have the potential of being resolved beyond the level of models and truth.
- One day... integrate correct representations for adjectives in our grammars.
- Disambiguate quantification in the parse?
Conclusion

- We can get nice distributions out of wikiwoods.
- It may be worth investigating ‘deeper’ lexical semantics issues under the microscope of distributions.
- Classical problems like quantification have the potential of being resolved beyond the level of models and truth.
- One day... integrate correct representations for adjectives in our grammars.
- Disambiguate quantification in the parse?
Conclusion

- We can get nice distributions out of wikiwoods.
- It may be worth investigating ‘deeper’ lexical semantics issues under the microscope of distributions.
- Classical problems like quantification have the potential of being resolved beyond the level of models and truth.
- One day... integrate correct representations for adjectives in our grammars.
- Disambiguate quantification in the parse?
Conclusion

- We can get nice distributions out of wikiwoods.
- It may be worth investigating ‘deeper’ lexical semantics issues under the microscope of distributions.
- Classical problems like quantification have the potential of being resolved beyond the level of models and truth.
- One day... integrate correct representations for adjectives in our grammars.
- Disambiguate quantification in the parse?