‘Deeper’ distributional semantics

Aurelie Herbelot

[1]Universität Potsdam
Department Linguistik

July 2012

Outline

‘Deeper’ distributional semantics

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

Outline

The corpus

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

```
<node nodeid='10011' cfrom='28' cto='36'><realpred lemma='original' pos='a' sense='1'/>
<sortinfo cvarsort='e' sf='prop' tense='untensed' mood='indicative'/></node>
<node nodeid='10012' cfrom='39' cto='46'><realpred lemma='drummer/nn' pos='u' sense='unknown'/>
<sortinfo cvarsort='x' pers='3' num='sg'/></node>
<link from='10011' to='10012'><rargname>ARG1</rargname><post>EQ</post></link>
```

$\rightarrow$ original_a ARG1 drummer_n

Pre-processing

- Nominalisations
- Compounds: fish *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.

Outline

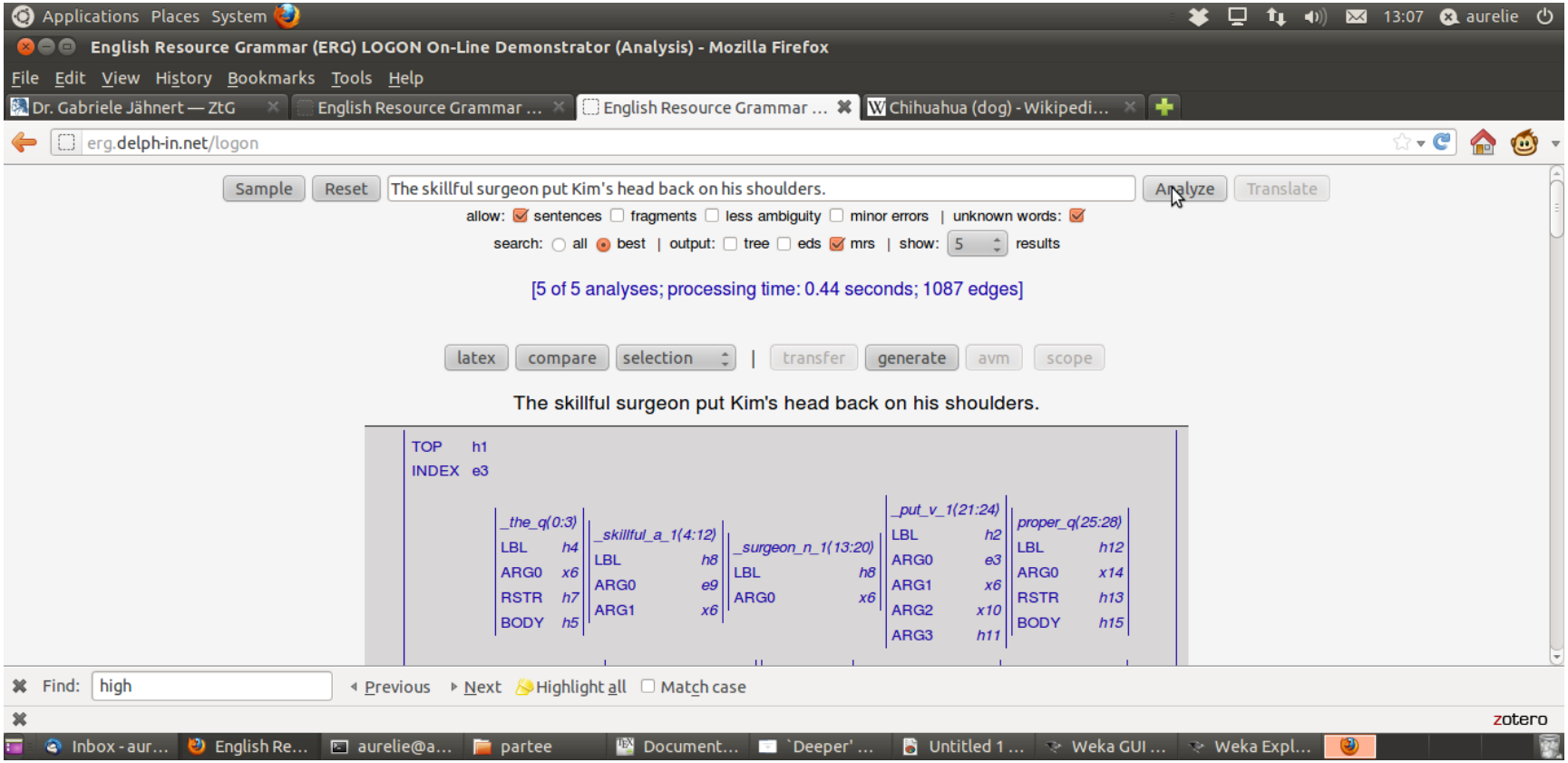
Adjective types, Partee (1995)

- **Intersective:** carnivorous mammal
  $||\text{carnivorous mammal}|| = ||\text{carnivorous}|| \cap ||\text{mammal}||$
- **Subsective:** skilful surgeon
  $||\text{skilful surgeon}|| \subseteq ||\text{surgeon}||$
- **Non-subsective:** former senator
  $||\text{former senator}|| \neq ||\text{former}|| \cap ||\text{senator}||$
  $||\text{former senator}|| \not\subseteq ||\text{senator}||$

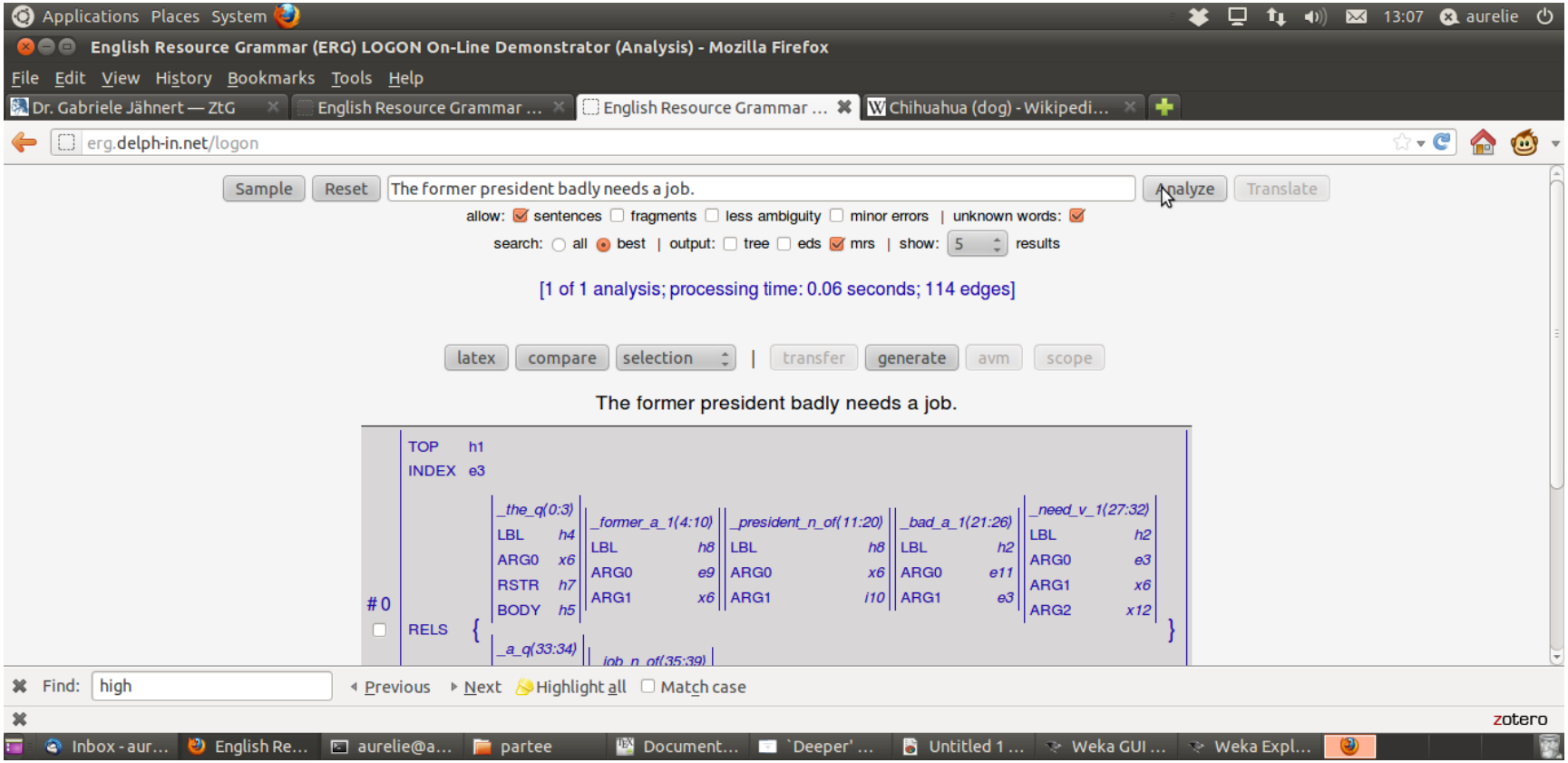
Integrating adjective types in the ERG

- The MRS of *skilful surgeon* shouldn't be
  l1:skilful(x)
  l2:surgeon(x)
  ... because *x* is not 'overall' skilful.
- Similarly, the current MRSs for *former*, *fake*, etc. are semantically inappropriate.

Skilful


Applications Places System
13:07 aurelie
English Resource Grammar (ERG) LOGON On-Line Demonstrator (Analysis) - Mozilla Firefox
File Edit View History Bookmarks Tools Help
Dr. Gabriele Jähnert — ZtG
English Resource Grammar ...
English Resource Grammar ...
Chihuahua (dog) - Wikipedi...
erg.delph-in.net/logon
Sample
Reset
The skillful surgeon put Kim's head back on his shoulders.
Analyze
Translate
allow: sentences fragments less ambiguity minor errors | unknown words:
search: all best | output: tree eds mrs | show: 5 results
[5 of 5 analyses; processing time: 0.44 seconds; 1087 edges]
latex
compare
selection
transfer
generate
avm
scope
The skillful surgeon put Kim's head back on his shoulders.
TOP h1
INDEX e3
_the_q(0:3)
LBL h4
ARG0 x6
RSTR h7
BODY h5
_skillful_a_1(4:12)
LBL h8
ARG0 e9
ARG1 x6
_surgeon_n_1(13:20)
LBL h8
ARG0 x6
_put_v_1(21:24)
LBL h2
ARG0 e3
ARG1 x6
ARG2 x10
ARG3 h11
proper_q(25:28)
LBL h12
ARG0 x14
RSTR h13
BODY h15
Find: high
Previous
Next
Highlight all
Match case
zotero
Inbox - aur...
English Re...
aurelie@a...
partee
Document...
`Deeper' ...
Untitled 1 ...
Weka GUI ...
Weka Expl...

Former


Applications Places System
13:07 aurelie
English Resource Grammar (ERG) LOGON On-Line Demonstrator (Analysis) - Mozilla Firefox
File Edit View History Bookmarks Tools Help
Dr. Gabriele Jähnert — ZtG
English Resource Grammar ...
English Resource Grammar ...
Chihuahua (dog) - Wikipedi...
erg.delph-in.net/logon
Sample
Reset
The former president badly needs a job.
Analyze
Translate
allow: sentences fragments less ambiguity minor errors | unknown words:
search: all best | output: tree eds mrs | show: 5 results
[1 of 1 analysis; processing time: 0.06 seconds; 114 edges]
latex
compare
selection
transfer
generate
avm
scope
The former president badly needs a job.
TOP h1
INDEX e3
0
RELS
_the_q(0:3)
LBL h4
ARG0 x6
RSTR h7
BODY h5
_former_a_1(4:10)
LBL h8
ARG0 e9
ARG1 x6
_president_n_of(11:20)
LBL h8
ARG0 x6
ARG1 i10
_bad_a_1(21:26)
LBL h2
ARG0 e11
ARG1 e3
_need_v_1(27:32)
LBL h2
ARG0 e3
ARG1 x6
ARG2 x12
_a_q(33:34)
job_n_of(35:39)
Find: high
Previous
Next
Highlight all
Match case
zotero
Inbox - aur...
English Re...
aurelie@a...
partee
Document...
`Deeper' ...
Untitled 1 ...
Weka GUI ...
Weka Expl...

Extra complication

- The semantics of *big city* should definitely be
  l1:big(x)
  l2: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$^\circ$ and skilful$^\circ$ 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:

.21287 late_a	.18000 high_a
.20550 old_a	.17931 american_a
.20047 large_a	.17749 great_a
.19687 former_a	.17717 same_a
.19649 original_a	.17277 main_a
.19338 early_a	.17113 good_a
.18843 small_a	.16459 other_a
.18591 only_a	.15379 several_a
.18134 national_a	.14607 new_a
.18046 general_a	.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?

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

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)

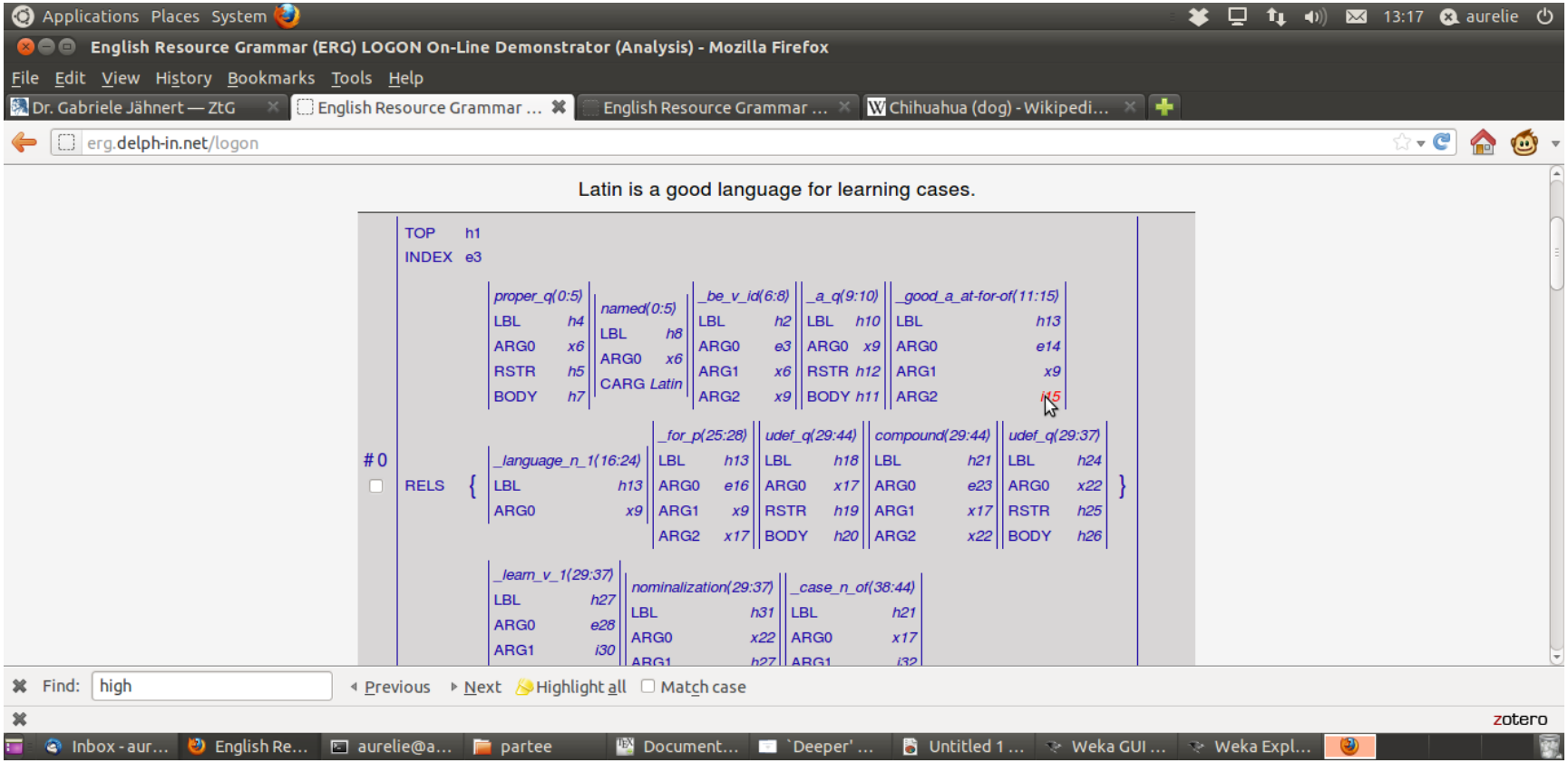
Examples

- 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



Examples

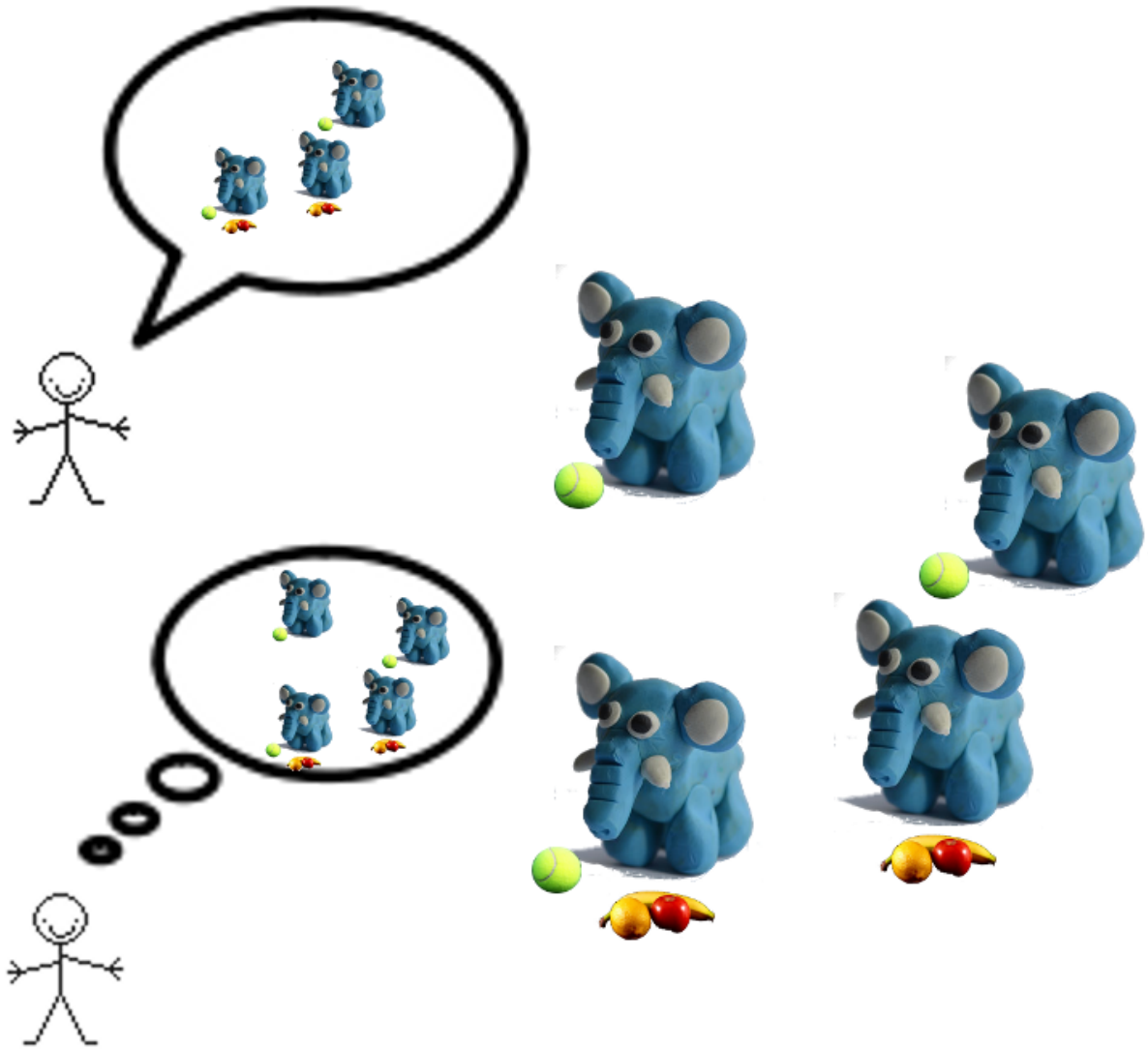
- High:
  - speed cost rank quality court rate mountain peak standard education (0.2047, 0.4547, 0.2503)
  - ground value degree position honour number point (0.18, 0.2847, 0.2571)
  - command priest street pressure frequency price award (0.1131, 0.3231, 0.1758)
  - commissioner risk rating percentage 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

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

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?