Acquiring Naturalistic Concept Descriptions from the Web

Tony Veale, Yanfen Hao
School of Computer Science,
{Tony.Veale, Yanfen.Hao}@UCD.ie

UCD Creative Language Systems Group
Introduction

Concept representation is the building block in the knowledge representation.

Figurative language processing demands more naturalistic descriptions (common sense knowledge) to refer concepts. E.g., metaphor generation, metaphor comprehension.

Do lexicons have enough common sense knowledge? E.g., In WN, surgeon is “a physician who specializes in surgery”. Hypernyms: Surgeon is kind of doctor. Hyponyms: Neurosurgeon is kind of surgeon. Synonyms: {operating surgeon, sawbones}.

Our common sense tells us: Surgeons are delicate, skilled, accurate and careful. Supermodels are poised, skinny, lithe, pretty, and graceful.

How to find these common sense knowledge?
Using Similes to Identify Stereotypical Cultural Associations

- Similes / Comparisons reveal the most diagnostic features of a concept
  E.g., “as hot as the sun”, “as dry as sand”, “as wobbly as jelly”, “as sweet as pie”

- The most frequent similes characterize the most pivotal concepts / senses
  E.g., animal concepts (“lion”, “rat”, etc.) are frequently used in comparisons

- Unlike metaphors, similes have a standard, recognizable syntactic frame
  “as barren as a desert”, “as delicate as a surgeon”, “as stiff as a corpse”

- Detailed Knowledge-Representations can be gathered for individual concepts
  Example: surgeon = {delicate, sensitive, skilled, clinical, professional, …}
Sampling Comparisons/Similes from the WWW

Query-pattern #1: “as ADJ as a|an *” for all antonymous adjectives in WN
Query-pattern #2: “as * as a|an NOUN” for all nouns gathered with query #1

- 200 sampled snippets per query, to give 74,704 apparent simile instances
  42,618 unique simile types, linking 3769 adjectives to 9287 unique nouns

- Major Issues: Implicit/Local Context, Irony
  “as hairy as a bowling-ball”, “as sober as a Kennedy”

- Annotation:
  12,259 are bona-fide similes and 2796 are ironic similes
Ironic Comparisons/Similes from the WWW

Some Examples:

As {welcome, painless, appealing, pleasant, exciting, entertaining} as a root-canal
As subtle as a {sledgehammer, freight_train, anvil, axe, rhino, toilet_seat, ...}
As hefty as a {laptop, croissant}
As blind as a {referee, hawk}
As {muscular, epicurean, smart, straight, sturdy, weighty, ...} as a paper_clip
As rare as a {ham_sandwich, toaster, traffic_jam, monsoon, garbage_pickup}
As {bulletproof, scary, subversive} as a sponge_cake
As private as a {park_bench, town_hall, shopping_mall}

2796 unique adj:noun ironic simile types.
936 adjectives to 1417 nouns.

13% of all annotated simile instances. 18% of unique simile types

View on the Web: http://afflatus.ucd.ie/sardonicus/tree.jsp
After word sense assignment, we have 18,794 facet:feature tuples with 2032 different WordNet noun senses.
Stereotypical Frames: Web-Derived Attribute-Value Pairings

Frame-names used as anchor in Google queries

**peacock**
- Has_feather: brilliant
- Has_plumage: extravagant
- Has_strut: proud
- Has_tail: elegant
- Has_display: colorful
- Has_manner: stately

**lione**
- Has_gait: majestic
- Has_strength: magnificent
- Has_soul: noble
- Has_eyes: fierce
- Has_teeth: ferocious
- Has_roar: threatening
Emprirical Evaluation

* Finds 51,045 modifiers for 214 nouns

Google query to mine adjectival modifiers for a given Concept

* Google query to mine noun attributes for a given Concept

e.g., rocket = \{fast, powerful, speed, thrust, …\} vector space of 59,979 features
<table>
<thead>
<tr>
<th>cid</th>
<th>Entpy</th>
<th>Purty</th>
<th>body</th>
<th>crea</th>
<th>dise</th>
<th>fami</th>
<th>vehi</th>
<th>publ</th>
<th>feel</th>
<th>clot</th>
<th>buil</th>
<th>time</th>
<th>anim</th>
<th>frui</th>
<th>furn</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.000</td>
<td>1.000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0.087</td>
<td>0.941</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>16</td>
</tr>
<tr>
<td>2</td>
<td>0.106</td>
<td>0.923</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0.000</td>
<td>1.000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0.000</td>
<td>1.000</td>
<td>0</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0.000</td>
<td>1.000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>17</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0.321</td>
<td>0.750</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0.160</td>
<td>0.895</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>0.100</td>
<td>0.929</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0.000</td>
<td>1.000</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0.155</td>
<td>0.864</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>19</td>
</tr>
<tr>
<td>11</td>
<td>0.405</td>
<td>0.722</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>0.286</td>
<td>0.789</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

13-way clustering: $\{l2=9.58e+001\} [214 \text{ of } 214]$, Entropy: 0.133, Purity: 0.902

Comparing $V+H$:
- 7183 features

Comparing $A+P$:
- 59,979 features


Table 1: Clustering accuracy for experiment 1 (214 nouns, 13 WordNet semantic classes).

<table>
<thead>
<tr>
<th>Approach</th>
<th>Values only</th>
<th>Attr’s only</th>
<th>All (V + A)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Almu. + Poesio</td>
<td>71.96% (51045 vals)</td>
<td>64.02% (8934 attr)</td>
<td>85.51% (59979 v+a)</td>
</tr>
<tr>
<td>Naturalistic Descriptions</td>
<td>70.2% (2209 vals)</td>
<td>78.7% (4974 attr)</td>
<td>90.2% (7183 v+a)</td>
</tr>
</tbody>
</table>

Table 2: Clustering accuracy for experiment 2 (402 nouns, 21 WordNet semantic classes).

<table>
<thead>
<tr>
<th>Approach</th>
<th>Values only</th>
<th>Attr’s only</th>
<th>All (V + A)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Almu. + Poesio (no filtering)</td>
<td>56.7% (94989 vals)</td>
<td>65.7% (24178 attr)</td>
<td>67.7% (119167 v+a)</td>
</tr>
<tr>
<td>Almu. + Poesio (with filtering)</td>
<td>62.7% (51345 vals)</td>
<td>70.9% (12345 attr)</td>
<td>66.4% (63690 v+a)</td>
</tr>
<tr>
<td>Naturalistic Descriptions</td>
<td>64.3% (5547 vals)</td>
<td>54.7% (3952 attr)</td>
<td>69.85% (9499 v+a)</td>
</tr>
</tbody>
</table>
Conclusions

• Similes provide best clues to naturalistic descriptions of common concepts
  A large case-base of “natural” comparisons is easily acquired from the web

• Useful for Metaphor/Simile Processing On-Line   Afflatus.ucd.ie/aristotle
  Generate metaphors for arbitrary target concepts that highlight given features
Thanks