L113 Word Meaning and Discourse Understanding
Session 5: Figurative Language and Sentiment

Simone Teufel

Natural Language and Information Processing (NLIP) Group

UNIVERSITY OF CAMBRIDGE

Simone.Teufel@cl.cam.ac.uk

2011/2012

1 Phenomenology
   - Logical Metonymy
   - Regular Metonymy
   - Metaphor
   - Idioms

2 Automatic Approaches
   - Logical Metonymy
   - Regular Metonymy
   - Metaphor

3 Semantic Orientation of Adjectives
   - Antonymy
   - Linguistic tests for complementaries and antonymy type
   - Linguistic vs. natural polarity

4 Automatic Detection of Sem. Orientation
   - Hatzivassiloglou and McKeown
   - Turney (PMI Method)
Types of Figurative Language

- **Hyperbole** (*mile-high ice cream cone)*,
- **Irony, Humour** (*beauty is in the eye of the beer-holder)*
- **Metonymy**
  - Creative: *The ham sandwich is waiting for his check.*
  - Regular: *All eyes were on Germany, but Berlin seemed unwilling to lead the Union.*
  - Logical: *a fast plane*
- **Metaphor**
  - *He shot down all my arguments.*
- **Simile**
  - *She is like a rose.*
- **Idiom**
  - *He has a bee in his bonnet.*

Logical Metonymy

- Due to Pustejovsky (1991, 1995)
- Additional meaning arises for particular verb-noun and adjective-noun combinations in a systematic way
- Verb (or adjective) semantically selects for an event-type argument, but syntactically selects for a noun.
- The event is however predictable from the semantics of the noun.

Examples:

- *Mary finished her beer.*
  - *Mary finished drinking her beer.*
- *easy problem*
  - *difficult language*
- *good cook*
- *good soup*
Metonymy

- Creative metonymy is hard to recognise automatically, because it depends on the understanding of the entire situation. AI bottleneck of knowledge representation.

- Regular metonymy follows schemes:
  - LOCATION-FOR-EVENT: After Lockerbie, people were more careful about saying that.

- Very frequent phenomenon in language

Metaphor

Express one concept/situation in terms of another concept/situation (including all other participants, properties and events of that situation).

FEELINGS are LIQUIDS:
- A simple phone call had managed to stir up all these feelings.
- Now here I was, seething with anger
- is a kind of pressure valve for the release of pent-up nervous energy
- ...provide an outlet for creativity ... Just ignore the turbulent feelings and turn your attention towards ...

ARGUMENT is WAR:
- Parties go into battle about how high to push the bar for skills
- Villagers launch fight to save their primary school from closure
Conceptual Metaphor Theory

- Due to Lakoff and Johnson (1980)
- Mapping between two cognitive domains
- Source and target domains
- Usually, source domain is more concrete/evocative

![Diagram of conceptual metaphor between war and argument domains]

```
+-- war
|   | ammunition
|   | shoot down
|   | peace offering
|   | win
|   | defense
|   | retreat
|   | battle
|   | attack
+-- SOURCE DOMAIN: WAR

+-- arguments
|   | evidence
|   | disagree
|   | heatedly
|   | rationally
|   | aggressively
+-- TARGET DOMAIN: ARGUMENT

- respond
- discuss
- agree
- persuade

```

Mixed Metaphor

Combination of two incompatible metaphorical mappings:

- *If we can hit that bullseye then the rest of the dominoes will fall like a house of cards... Checkmate.*
  Zapp Brannigan (Futurama)

- *it would somehow bring the public school system crumbling to its knees.*

- *biting the hand that rocks the cradle*
- *He took it like a fish out of water.*
- *He wanted to get out from under his father’s coat strings.*
- *She’s been burning the midnight oil at both ends.*
**Dead metaphor:** The image that the metaphor invokes has been established in the language, i.e., is now contained in the “lexicon”. Creative, situational figurative images are excluded.

- *I simply cannot grasp this idea.*
- *This really made an impression on me.*

Often not perceived as metaphor.

**Idioms**

- Minimal semantic constituents which consist of more than one word.
- Definition: the meaning of an idiom cannot be inferred as a compositional function of the meaning of its parts.

  - *pull somebody’s leg*
  - *be off one’s rocker*

**Syntactic Variability Tests:**

- ?*Arthur has a bee, apparently, in his bonnet.* (insertion)
- ?*Arthur kicked the large bucket.* (modification)
Idioms: crosslingual issues

Level of translatability of idiom into another language is unpredictable.

- “donner sa langue au chat” (give your tongue to the cat)
- “appeller un chat un chat” (call a cat a cat)

Idiom or dead metaphor? Rephrasing Test

If rephrasing results in similar semantics, the multi-word entity is not a semantic constituent (thus a dead metaphor, not an idiom).

**Dead metaphors:**

- *They tried to sweeten the pill.* ≈ *They tried to sugar the medicine.*
- *We shall leave no stone unturned in our search for the culprit.*
  ≈
  *We shall look under every stone in our search for the culprit.*

**Idioms:**

- *John pulled his sister’s leg* ≠ *John tugged at his sister’s leg*
- *Arthur kicked the bucket* ≠ *Arthur tipped over the water recepticle*
Logical Metonymy: Lapata and Lascarides (2003)

- **a fast** \{ landing? \}
  \{ taxiing? \}
  \{ flying? \}
- **I enjoyed** \{ reading? \}
  \{ writing? \}
  \{ eating? \}

What is missing for full automatic recognition is the implicit verb (\textit{fly(ing)} and \textit{read(ing)}).

Cooccurrences of \textit{plane–fly} and \textit{fly–fast} and \textit{like–reading} and \textit{read–book} in corpus can give us the answer.

But: conditioning on both associations at the same time will result in data sparseness.

Therefore: probabilistic model used separates the two associations.

Verbs:

\[ P(e, o, v) = \frac{f(v, e)f(o, e)}{f(e)N} \]

Adjectives:

\[ P(a, e, n, rel) = \frac{f(rel, e, n)f(a, e)}{f(e)N} \]

<table>
<thead>
<tr>
<th>Frequency: verbs modified by \textit{fast}.</th>
<th>Frequency: verbs taking \textit{plane} as argument.</th>
</tr>
</thead>
<tbody>
<tr>
<td>f(fast, e)</td>
<td>f(fast, e)</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>go</td>
<td>29</td>
</tr>
<tr>
<td>grow</td>
<td>28</td>
</tr>
<tr>
<td>beat</td>
<td>27</td>
</tr>
<tr>
<td>run</td>
<td>16</td>
</tr>
<tr>
<td>rise</td>
<td>14</td>
</tr>
<tr>
<td>travel</td>
<td>13</td>
</tr>
<tr>
<td>move</td>
<td>12</td>
</tr>
<tr>
<td>come</td>
<td>11</td>
</tr>
<tr>
<td>drive</td>
<td>8</td>
</tr>
<tr>
<td>get</td>
<td>7</td>
</tr>
</tbody>
</table>
Markert and Nissim (06):

- Supervised learning problem: country and organisation names are classified as metonymical or not
- Manually annotate large training corpus (1,000 examples of each from the BNC)
- Good human agreement
- Use grammatical information as features
- Roughly 20% of country names are used metonymically, and 33% of organisation names.

Countries:

- *Or have you forgotten that America did once try to ban alcohol and look what happened!*
- *At one time there were nine tenants there who went to America.*

Organisations:

- *BMW and Renault sign recycling pact.*
- *How I bought my first BMW.*
Metonymy: Features and results

Features:
- Grammatical function (subj, premod, gen, obj, PP, pred, subjpassive, iobj, other)
- Number, definiteness of determiner
- Lexical head

Results:
- 87% correct for country names (EMNLP 2002 paper)
- 76% correct for organisations (IWCS 2005 paper)

Automatic Approaches to Metaphor Recognition

- Selectional restrictions of metaphorically used word in literal interpretation are violated (Wilks 79)
- is-a metaphors violate WN-hyponymy relation: all the world is a stage (Krishnakumaran and Zhu, 2007)
- Or use manually created metaphor-specific knowledge bases (Martin 1980; Narayanan 1999; Barnden and Lee 2002).
SLIPNET (Veale and Hao 2008) relates two concepts via definitions, allowing for deletions, insertions and substitutions. Goal: to find a connection between source and target concepts. Example:

*Make-up is a Western Burqa*

\[ \text{make-up} \Rightarrow \]
- typically worn by women
- expected to be worn by women
- must be worn by women
- must be worn by Muslim women

\[ \text{burqa} \Leftarrow \]

Metaphor Recognition (Shutova et al. 2010)

- Start from seed set including a potentially metaphorical verb
- Model possible target domain \( \rightarrow \) cluster its arguments and subject
- Most “abstract” cluster corresponds to target concept cluster
- Model possible source domain \( \rightarrow \) cluster the verbs that go with these arguments

<table>
<thead>
<tr>
<th><strong>Target concept cluster</strong></th>
<th><strong>Source domain cluster</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>desire hostility anxiety passion excitement doubt fear anger curiosity enthusiasm impulse instinct emotion feeling suspicion rage</td>
<td>gulp drain stir empty pour sip spill swallow drink pollute seep flow drip purify ooze pump bubble splash ripple simmer boil tread</td>
</tr>
</tbody>
</table>

\[ \text{stir excitement} \rightarrow \text{swallow anger} \]
\[ \text{cast doubt} \rightarrow \text{spark enthusiasm} \]
Metaphor Interpretation by literal paraphrase

Input: A *carelessly leaked* report
Output: A *carelessly disclosed* report

- Find lexically similar candidates for replacement (standard distributional semantics approach)
- Use a Resnik-type selectional restriction filter to filter out metaphorical expressions (those that have low selectional restriction strength), so that only literal ones are left over.

\[ A_R(v, c) = \frac{1}{S_R(v)} P(c|v) \log \frac{P(c|v)}{P(c)} \]
Logical Metonymy can be solved by individual associations of implicit verb with explicitly mentioned lexical items.

Problem with Lapata/Lascarides (2003): word senses all conflated.

Regular Metonymy can be solved by supervised classification with features similar to supervised WSD.

Metaphors can be recognised by seed clustering and paraphrased by lexical similarity and selectional restrictions.

Shutova et al.’s system: precision is high (~80%), but recall is very low (0.25%).
- There are different kinds of opposites: complementaries and antonyms.
- Semantic orientation: degree of positiveness/negativeness.
- Many antonyms have opposite semantic orientation. Exceptions: *verbose*—*terse*.
Complementaries

Complementaries between them exhaustively divide some conceptual domain into mutually exclusive compartments. Antonyms don’t.

*neither–nor* test:

- ? Mary is neither married nor is she single.
- It’s neither hot nor cold today.

They are also not gradable:

- ? extremely true – extremely safe
- ? more pregnant than most – longer than some
- ? moderately female – moderately clean

*light–heavy* and *hot–cold* do not behave in the same way:

- This box is light, but it’s heavier than that one.
- ? Today it’s cold, but hotter than yesterday.

What is going on? *hot* seems to mean *hot, and to a larger degree*, whereas *heavier* seems to mean *of greater weight*.

- hotter is a true comparative of *hot*
- heavier is a pseudo-comparative of *heavy/1*, and a true comparative of *heavy/2*
Antonyms 2: How-adj questions

Are they possible for both antonyms?
Compare long–short:
- How long is it?
- ? How short is it?
with hot–cold:
- How cold is it?
- How hot is it?

Antonyms 3: Impartiality of how-adj questions

Does one of the questions imply something about your presuppositions?
hot–cold:
- How cold is it? \(\rightarrow\) committed
- How hot is it? \(\rightarrow\) committed

clean–dirty:
- How clean was the room? \(\rightarrow\) impartial
- How dirty was the room? \(\rightarrow\) committed
Antonyms: Three types

- **good–bad** is an example of an **overlapping** antonym.
  - Overlapping antonyms are evaluative, and thus carry semantic orientation in our sense.
- **hot–cold** is an example of an **equipollent** antonym.
  - Equipollent antonyms are often correlated with sensory perceptions.
- **long–short** is an example of a **polar** antonym.
  - Polar antonyms show the greatest level of abstraction, but are neutral/descriptive.

Linguistic polarity vs. natural polarity

- Can we predict from the linguistic form which one of the antonyms is more positive?
  - Prediction: the more salient antonym often has a positive polarity.
- **Test 1**: The antonym that can be paraphrased as the other one plus a negative prefix is the less salient one.
- **Test 2**: The more salient antonym is associated with “more” properties:
  - *Something is dead when there is no life present.*
  - *Something is alive when there is no deadness present.*
  - *Something is clean when there is no dirt present.*
  - *Something is dirty when there is no cleanness present.*
  - *Something is clean when there is no dirt present.*
  - *Something is dirty when there is no cleanness present.*
**Test 3:** The more salient antonym yields the impartial interpretation in the how-adj question.

In the case of verbs:
- Antonymy in verbs often concerns directional actions, and reversionary actions (Cruse, chapter 10)
- The salient antonym is the one that results in “increased entropy” (undress, dismount, disarrange, unscrew, unpack...)

Hatzivassiloglou and McKeown’s (1997) algorithm classifies adjectives into those with positive or negative semantic orientation:

- **Semantic Polarity of an adjective:**
  - **Direction:** In which direction does the referent deviate from the norm in its semantic field?
  - **Evaluative:** Is this good or bad?

If we know that two adjectives relate to the same property (e.g., *hot* and *cold*) but have different orientations they are usually antonyms.
Idea

- In coordinations, these facts result in constraints on the semantic orientation:

  (1)

  a. The tax proposal was **simple and well-received** by the public.
  b. The tax proposal was **simplistic but well-received** by the public.
  c. The tax proposal was **simplistic and well-received** by the public.

- *but* combines adjectives of opposite orientation; *and* adjectives of the same orientation

- This indirect information can be exploited using a corpus.

Algorithm

- Extract all coordinated adjectives from corpus
- Classify each extracted adjective pair as same or different orientation
- This results in graph with same or different links between adjectives
- Cluster into two orientations, placing as many words of the same orientation as possible into the same subset
- Cluster with higher overall frequency is labelled positive
- Evaluate against independently orientation-annotated gold standard set (1336 most frequent adjectives; 657 positive, 679 negative)
Coordinated adjectives

- Extract from POS tagged WSJ (21 million words) adjective pairs coordinated by *and, or, but, either-or, neither-nor*
- This results in 15048 adjective pairs (token); 9296 (type)
- Number of those where orientation of both partners is known (via gold standard): 4024 (token); 2748 (type)
  - *and* is most reliable same-orientation predictor, particularly in predicative position (85%), this drops to 70% in appositive position.
  - *but* has 31% same-orientation.

Classifier

- Features:
  - Type of coordination
  - Type of modification (attributive, predicative, appositive, resultative (“Bill laughed himself hoarse”))
  - Number of modified noun (singular or plural)
- Simple derivational morphological analysis suggests additional different orientations: Out of the labelled adjectives, 97% of morphologically related pairs (102) have different orientation
- Log-linear regression model with linear predictor; best classifier achieves 82%
- Baseline: always predict same-orientation: 79%
- But-rule: different if seen with *but*, same-orientation otherwise: 82%
Clustering adjectives with same orientation

- Interpret classifier’s P(same-orientation) as dissimilarity value.
- Perform non-hierarchical clustering via Exchange Method
- Start from random partition, locate the adjective which reduces the cost \( c \) most if moved.

\[
c = \sum_{i=1}^{2} \left( \frac{1}{|C_i|} \sum_{x,y \in C_i, x \neq y} d(x, y) \right)
\]

- Repeat until no movements can improve the cost; overall dissimilarity cost is now minimised.
- At final iteration, move any adjective which violates the following constraint:

\[
\frac{1}{|\bar{C}| - 1} \sum_{y \in \bar{C}, x \neq y} d(x, y) < \frac{1}{|\bar{C}|} \sum_{y \in \bar{C}} d(x, y)
\]

Labelling Clusters as Positive or Negative

- In antonym pairs, the one which is semantically unmarked is also in most cases the positive one.
- Semantically unmarked ones should occur overall more frequently \( \rightarrow \) group with overall higher frequency count gets labelled as positive.
Results

- Dependent on how sparse the test set is, results between 78% and 92% correct.
- Baselines: MFC 51% negative.
- Classified as positive: bold, decisive, disturbing, generous, good, honest, important, large, mature, patient, peaceful, positive, proud, sound, stimulating, straightforward, strange, talented, vigorous, witty.
- Classified as negative: ambiguous, cautious, cynical, evasive, harmful, hypocritical, inefficient, insecure, irrational, irresponsible, minor, outspoken, pleasant, reckless, risky, selfish, tedious, unsupported, vulnerable, wasteful.

Discussion

Strengths:
- Fully unsupervised, algorithm starts from nothing.
- Convincing results.

Weaknesses:
- Analysis of isolated adjectives, not phrases.
- Needs large corpus in order to contain enough coordinated adjectives.
- Clustering algorithm is not optimal (problem is NP-hard); it is a steepest-descending hill climbing method, which is at least guaranteed to converge (but might run algorithm repeatedly with different start partitions).
Turney’s 2002 method

- Determine semantic orientation of phrases, not just single adjectives
- Single adjectives do not always carry full orientation; context is needed. *unpredictable plot* vs. *unpredictable steering*
- Unsupervised method based on distributional semantics
- Assign a numerical ranking indicating strength of orientation
- Use search engine hits to estimate semantic orientation of a phrase

**Idea**

- If an adjectival phrase has a positive semantic orientation, it will appear more frequently in the intermediate vicinity of known positive adjectives, and vice versa.
- Measure an adjective’s tendency to appear in positive or negative vicinity via PMI-IR
  - Pointwise mutual information determines similarity of a pair of phrases
  - Use IR to quantify effect
- Measure success indirectly via classification of entire reviews
PMI and SO

\[ \text{PMI}(\text{word}_1, \text{word}_2) = \log_2 \left( \frac{P(\text{word}_1, \text{word}_2)}{P(\text{word}_1)P(\text{word}_2)} \right) \]

- **Semantic Orientation:**
  \[ \text{SO}(\text{phrase}) = \text{PMI}(\text{phrase, excellent}) - \text{PMI} \ (\text{phrase, poor}) \]

- Counts are calculated via search engine hits
- Altavista’s NEAR operator – window of 10 words

Therefore:

\[ \text{SO}(\text{phrase}) = \log_2 \left( \frac{\text{hits}(\text{phrase NEAR excellent})\text{hits(poor)}}{\text{hits}(\text{phrase NEAR poor})\text{hits(excellent)}} \right) \]

**Results:** indirectly via classification of documents

- 74% accuracy on classifying 410 reviews from Epinions
- 66% accuracy on movie reviews

An example:

<table>
<thead>
<tr>
<th>Word</th>
<th>PMI</th>
</tr>
</thead>
<tbody>
<tr>
<td>little difference</td>
<td>-1.615</td>
</tr>
<tr>
<td>clever tricks</td>
<td>-0.040</td>
</tr>
<tr>
<td><strong>programs such</strong></td>
<td><strong>0.117</strong></td>
</tr>
<tr>
<td>possible moment</td>
<td>-0.668</td>
</tr>
<tr>
<td>unethical practices</td>
<td>-8.484</td>
</tr>
<tr>
<td>old man</td>
<td>-2.566</td>
</tr>
<tr>
<td>other problems</td>
<td>-2.748</td>
</tr>
<tr>
<td>probably wondering</td>
<td>-1.830</td>
</tr>
<tr>
<td>virtual monopoly</td>
<td>-2.050</td>
</tr>
<tr>
<td>other bank</td>
<td>-0.850</td>
</tr>
<tr>
<td>extra day</td>
<td>-0.286</td>
</tr>
<tr>
<td><strong>direct deposits</strong></td>
<td><strong>5.771</strong></td>
</tr>
<tr>
<td>online web</td>
<td>1.936</td>
</tr>
<tr>
<td>cool thing</td>
<td>0.395</td>
</tr>
<tr>
<td>very handy</td>
<td>1.349</td>
</tr>
<tr>
<td>lesser evil</td>
<td>-2.288</td>
</tr>
</tbody>
</table>

Total: -1.218. Rating: Not recommended.
Discussion

Strengths:
- Fully unsupervised
- Nominal context makes adjective semantics more interpretable

Weaknesses:
- No direct evaluation of SO provided
- Very simple model
- Requires many searches (too many without API)
- NEAR no longer supported
- Results depend substantially on lexical items chosen, but choice largely unmotivated

Reading

- Cruse (1986), chapters 9 and 11.3;