Cohesion-based Models for Detecting Figurative Language

Caroline Sporleder

MMCI / Computational Linguistics, Saarland University

csporled@coli.uni-sb.de

ILIKS 2010 Annual Meeting

November 30, 2010
Outline of the Talk

1. Processing Non-Literal Language
2. Cohesion-Based Idiom Detection
3. Combining Unsupervised and Supervised Learning
4. Conclusion
Processing Non-Literal Language
Types of Non-Literal Language

- idioms ("to play with fire", "to spill the beans", "to be under the weather")
- metaphors ("life is a journey")
- metonymy ("The White House announced", "the Crown’s attitude to the law")
- etc.
Non-Literal Language is Idiosyncratic

- is often semantically non-compositionally ("to be under the weather")
- can violate selectional restrictions ("to push one's luck")
- can have idiosyncratic semantic argument structures ("to break the ice with North Korea")

NLP Systems...

1. need to recognise non-literal language
2. need to process and analyse it appropriately
Non-Literal Language is Frequent

Trento is able to surprise. “Cittá in giardino” after the awakening of spring, its tasty products in the autumn period, or illuminated by the ”warm” Christmas market lights during Advent. Summer holidays in Trentino is the perfect time to get to know new friends, broaden your horizon or let your hair down. Trento is surrounded by beautiful mountains. The high craggy limestone buttresses and spires of the Dolomites soar above the conifer forests. The valleys, woodland, grassland, and small lakes between the peaks are breathtaking. This is why the Dolomite mountains are famous throughout the world: almost vertical walls, hundreds of metres high and the needles and towering rock of the very famous Torri del Vaiolet.
Non-Literal Language is Frequent

Trento is able to surprise. “Cittá in giardino” after the awakening of spring, its tasty products in the autumn period, or illuminated by the ”warm” Christmas market lights during Advent. Summer holidays in Trentino is the perfect time to get to know new friends, broaden your horizon or let your hair down. Trento is surrounded by beautiful mountains. The high craggy limestone buttresses and spires of the Dolomites soar above the conifer forests. The valleys, woodland, grassland, and small lakes between the peaks are breath taking. This is why the Dolomite mountains are famous throughout the world: almost vertical walls, hundreds of metres high and the needles and towering rock of the very famous Torri del Vaiolet.
Most previous research

- look them up in an electronically readable idiom dictionary, or
- use word association statistics to automatically extract idioms (type-based idiom classification)
Most previous research
- look them up in an electronically readable idiom dictionary, or
- use word association statistics to automatically extract idioms (type-based idiom classification)

Example

to be under the table

to be under the carpet

to be under the weather

to push one’s luck

to push one’s bike

to spill the milk

to spill the water

to spill the beans
Most previous research

- look them up in an electronically readable idiom dictionary, or
- use word association statistics to automatically extract idioms (type-based idiom classification)

Example

to be under the table ⇒ literal
to be under the carpet ⇒ literal
to be under the weather ⇒ non-literal
to push one’s luck ⇒ non-literal
to push one’s bike ⇒ literal
to spill the milk ⇒ literal
to spill the water ⇒ literal
to spill the beans ⇒ non-literal
Type-Based idiom classification

- doesn’t work for expressions which can have literal and idiomatic meaning
- doesn’t work for creative language use

Literal Usage

(1) *Somehow I always end up spilling the beans all over the floor and looking foolish when the clerk comes to sweep them up.*

(2) *Grilling outdoors is much more than just another dry-heat cooking method. It’s the chance to play with fire, satisfying a primal urge to stir around in coals.*
Type-Based idiom classification

- doesn’t work for expressions which can have literal and idiomatic meaning
- doesn’t work for creative language use

Literal Usage

1. *Somehow I always end up spilling the beans all over the floor and looking foolish when the clerk comes to sweep them up.*

2. *Grilling outdoors is much more than just another dry-heat cooking method. It’s the chance to play with fire, satisfying a primal urge to stir around in coals.*

⇒ **Idioms have to be recognised in discourse context!**
  (token-based classification)
Previous Approaches:

- Katz and Giesbrecht (2006): supervised machine learning (k-nn), vector space model
- Birke and Sarkar (2006): bootstrapping from seed lists
- Cook et al. (2007), Fazly et al. (to appear): unsupervised, predict non-literal if idiom is in canonical form (≈ dictionary form)

⇒ limited contribution of discourse context
How do you know whether an expression is used idiomatically?

Literal Usage

Grilling outdoors is much more than just another dry-heat cooking method. It’s the chance to **play with fire**, satisfying a primal urge to **stir around in coals**.
How do you know whether an expression is used idiomatically?

**Literal Usage**

*Grilling outdoors is much more than just another dry-heat cooking method. It’s the chance to play with fire, satisfying a primal urge to stir around in coals.*

Literally used expressions typically exhibit lexical cohesion with the surrounding discourse (e.g. participate in lexical chains of semantically related words).
How do you know whether an expression is used idiomatically?

Non-Literal Usage

Dissanayake said that Kumaratunga was "playing with fire" after she accused military’s top brass of interfering in the peace process. Kumaratunga has said in an interview she would not tolerate attempts by the army high command to sabotage her peace moves. A defence analyst close to the government said Kumaratunga had spoken a "load of rubbish" and the security forces would not take kindly to her disparaging comments about them.

Non-Literally used expressions typically do not participate in cohesive chains.
Literal Use without Lexical Chain

Chinamasa compared McGown’s attitude to morphine to a child’s attitude to playing with fire – a lack of concern over the risks involved.

Non-Literal Use with Lexical Chain

Saying that the Americans were "playing with fire" the official press speculated that the "gunpowder barrel" which is Taiwan might well "explode" if Washington and Taipei do not put a stop to their "incendiary gesticulations."

⇒ Both cases are relatively rare
Cohesion-Based Idiom Detection
Identifying Idiomatic Usage

Are there (strong) cohesive ties between the component words of the idiom and the context?

- Yes $\Rightarrow$ literal usage
- No $\Rightarrow$ non-literal usage

(cf. Hirst and St-Onge’s (1998) work on detecting malapropisms)

We need:

- a measure of semantic relatedness
- a method for modelling lexical cohesion
Modelling Semantic Relatedness (1)

Graph-based approach:

- compute the path between two concepts in a hierarchical organised lexicon (thesaurus, WordNet)
- **Assumption:** semantic distance between concepts correlates to path length

Distribution-based approach:

- compute co-occurrence vectors for target words in a corpus
- **Assumption:** related words occur in similar contexts
Graph-based approach:
- compute the path between two concepts in a hierarchical organised lexicon (thesaurus, WordNet)
- **Assumption**: semantic distance between concepts correlates to path length
- **Disadvantages**: limited coverage and restriction to a few relations (e.g. hypernymy)

Distribution-based approach:
- compute co-occurrence vectors for target words in a corpus
- **Assumption**: related words occur in similar contexts
Graph-based approach:

- compute the path between two concepts in a hierarchical organised lexicon (thesaurus, WordNet)

**Assumption:** semantic distance between concepts correlates to path length

**Disadvantages:** limited coverage and restriction to a few relations (e.g. hypernymy)

Distribution-based approach:

- compute co-occurrence vectors for target words in a corpus

**Assumption:** related words occur in similar contexts

**Disadvantage:** conflation of different word senses
We have to model non-classical relations (e.g. fire - coals, sweep up - spill, ice - freeze) and world knowledge (Wayne Rooney - ball).

⇒ distributional approaches better suited than WordNet-based ones
⇒ ideally, we need loads of up-to-date data

Normalised Google Distance (NGD) (Cilibrasi and Vitanyi, 2007)

- use search engine page counts (here: Yahoo) as proxies for word co-occurrence

\[
NGD(x, y) = \frac{\max\{\log f(x), \log f(y)\} - \log f(x, y)}{\log M - \min\{\log f(x), \log f(y)\}}
\]

\((x, y: \text{target words}, f(x) \text{ page count for } x, M: \text{total number of pages indexed})\)
Two methods:

- lexical chains
- cohesion graphs
Method 1: Lexical Chains

Literal Use

Dad had to **break the ice** on the chicken troughs so that they could get water.

similarity threshold \( t=0.5 \)
Method 1: Lexical Chains

**Literal Use**

Dad had to *break the ice* on the chicken troughs so that they could get water.

similarity threshold $t=0.5$

**Lexical Chains:**

- L1: *Dad*
Method 1: Lexical Chains

Literal Use

Dad had to **break** the **ice** on the chicken troughs so that they could get water.

similarity threshold $t=0.5$

**Lexical Chains:**

- $L1: \textit{Dad}$
Method 1: Lexical Chains

Literal Use
Dad had to break the ice on the chicken troughs so that they could get water.

similarity threshold $t=0.5$; $\text{sim}(\text{Dad}, \text{break})=0.2$

Lexical Chains:
- L1: Dad
Method 1: Lexical Chains

Literal Use

Dad had to **break** the ice on the chicken troughs so that they could get water.

similarity threshold $t=0.5$

Lexical Chains:

- L1: *Dad*
- L2: *break*
Literal Use
Dad had to break the ice on the chicken troughs so that they could get water.

similarity threshold $t=0.5$;

**Lexical Chains:**
- L1: Dad
- L2: break
Method 1: Lexical Chains

Literal Use

Dad had to break the ice on the chicken troughs so that they could get water.

similarity threshold $t=0.5$; $\text{sim(Dad,ice)}=0.1$

Lexical Chains:

- L1: Dad
- L2: break
Method 1: Lexical Chains

Literal Use

Dad had to **break the ice** on the chicken troughs so that they could get water.

similarity threshold $t=0.5$; $\text{sim}(\text{break, ice})=0.4$

Lexical Chains:

- L1: *Dad*
- L2: *break*
Method 1: Lexical Chains

Literal Use

Dad had to **break the ice** on the chicken troughs so that they could get water.

similarity threshold $t=0.5$

Lexical Chains:

- L1: *Dad*
- L2: *break*
- L3: *ice*
Literal Use
Dad had to *break the ice on the* chicken troughs so that they could get water.

similarity threshold $t=0.5$

**Lexical Chains:**
- L1: *Dad*
- L2: *break*
- L3: *ice*
Method 1: Lexical Chains

Literal Use

Dad had to break the ice on the chicken troughs so that they could get water.

similarity threshold $t=0.5$; $\text{sim}(\text{Dad}, \text{chicken})=0.1$

Lexical Chains:

- L1: $Dad$
- L2: $break$
- L3: $ice$
Literal Use
Dad had to break the ice on the chicken troughs so that they could get water.

similarity threshold $t=0.5; \text{sim(break, chicken)}=0.2$

Lexical Chains:
- L1: Dad
- L2: break
- L3: ice
Method 1: Lexical Chains

Literal Use
Dad had to break the ice on the chicken troughs so that they could get water.

similarity threshold $t=0.5$; $\text{sim}(\text{ice, chicken})=0.1$

Lexical Chains:
- L1: Dad
- L2: break
- L3: ice
Method 1: Lexical Chains

Literal Use

Dad had to **break the ice** on the **chicken** troughs so that they could get water.

similarity threshold $t=0.5$

Lexical Chains:

- L1: *Dad*
- L2: *break*
- L3: *ice*
- L4: *chicken*
Method 1: Lexical Chains

Literal Use

Dad had to break the ice on the chicken troughs so that they could get water.

similarity threshold $t=0.5$

Lexical Chains:
- L1: Dad
- L2: break
- L3: ice
- L4: chicken
Literal Use

Dad had to **break the ice** on the chicken **troughs** so that they could get water.

similarity threshold $t=0.5$; $\text{sim(Dad, troughs)}=0.1$

**Lexical Chains:**
- L1: *Dad*
- L2: *break*
- L3: *ice*
- L4: *chicken*
Method 1: Lexical Chains

Literal Use

Dad had to break the ice on the chicken troughs so that they could get water.

similarity threshold $t=0.5$; $\text{sim(\text{break, troughs})}=0.4$

Lexical Chains:

- L1: Dad
- L2: break
- L3: ice
- L4: chicken
Literal Use

Dad had to break the ice on the chicken troughs so that they could get water.

similarity threshold $t=0.5$; $\text{sim}(\text{ice}, \text{troughs})=0.55$

**Lexical Chains:**
- L1: Dad
- L2: break
- L3: ice
- L4: chicken
Method 1: Lexical Chains

Literal Use

Dad had to break the ice on the chicken troughs so that they could get water.

similarity threshold $t=0.5$; $\text{sim(chicken, troughs)}=0.7$

Lexical Chains:

- L1: Dad
- L2: break
- L3: ice
- L4: chicken
Literal Use

Dad had to break the ice on the chicken troughs so that they could get water.

similarity threshold $t=0.5$

Lexical Chains:

- L1: Dad
- L2: break
- L3: ice
- L4: chicken – troughs
Method 1: Lexical Chains

Literal Use

Dad had to break the ice on the chicken troughs so that they could get water.

similarity threshold $t=0.5$

**Lexical Chains:**

- L1: *Dad*
- L2: *break*
- L3: *ice*
- L4: *chicken – troughs*
Literal Use

Dad had to break the ice on the chicken troughs so that they could get water.

similarity threshold $t=0.5$; $\text{sim}(\text{Dad, water})=0.1$

Lexical Chains:

- L1: Dad
- L2: break
- L3: ice
- L4: chicken – troughs
Method 1: Lexical Chains

Literal Use

Dad had to break the ice on the chicken troughs so that they could get water.

similarity threshold $t=0.5; \ sim(break, water)=0.1$

Lexical Chains:

- L1: Dad
- L2: break
- L3: ice
- L4: chicken – troughs
Method 1: Lexical Chains

Literal Use
Dad had to break the ice on the chicken troughs so that they could get water.

similarity threshold $t=0.5$; $\text{sim(ice,water)}=0.8$

Lexical Chains:
- L1: Dad
- L2: break
- L3: ice
- L4: chicken – troughs
Method 1: Lexical Chains

Literal Use

Dad had to break the ice on the chicken troughs so that they could get water.

similarity threshold $t=0.5$; $\text{sim(chicken,water)}=0.4$

Lexical Chains:
- L1: Dad
- L2: break
- L3: ice
- L4: chicken – troughs
Method 1: Lexical Chains

Literal Use

Dad had to break the ice on the chicken troughs so that they could get water.

similarity threshold \( t = 0.5 \); \( \text{sim(troughs, water)} = 0.6 \)

Lexical Chains:
- L1: Dad
- L2: break
- L3: ice
- L4: chicken – troughs
Method 1: Lexical Chains

Literal Use

Dad had to break the ice on the chicken troughs so that they could get water.

similarity threshold $t=0.5$

Lexical Chains:

- L1: Dad
- L2: break
- L3: ice – water
- L4: chicken – troughs
Method 1: Lexical Chains

Literal Use

Dad had to **break the ice** on the chicken troughs so that they could get water.

similarity threshold $t=0.5$

**Lexical Chains:**

- L1: *Dad*
- L2: *break*
- L3: *ice – water*
- L4: *chicken – troughs*

⇒ **Literal!**
Drawbacks:

- one free parameter (similarity threshold $t$) for deciding when to put two words in the same chain
  $\Rightarrow$ needs to be optimised on an annotated data set (weakly supervised)

- approach is sensitive to chaining algorithm and parameter settings
Modelling Cohesion: Cohesion Graphs

Literal Use

Dad had to break the ice on the chicken troughs so that they could get water.

![Cohesion Graph Diagram]

Average connectivity = 0.34

with idiom: break ice

Caroline Sporleder
Detecting Figurative Language (19/32)
Literal Use

Dad had to **break the ice** on the chicken troughs so that they could get water.

with idiom:

avg. connectivity=0.34
Dad had to **break the ice** on the chicken troughs so that they could get water.

*with idiom:*

avg. connectivity = 0.34

*without idiom:*

avg. connectivity = 0.33
Dad had to break the ice on the chicken troughs so that they could get water.

- with idiom: avg. connectivity=0.34
- without idiom: avg. connectivity=0.33

⇒ Literal!
Data

- 17 idioms (mainly V+NP and V+PP) with literal and non-literal sense
- all (canonical form) occurrences extracted from a Gigaword corpus (3964 instances)
- five paragraphs context
- manually labelled as “literal” or “non-literal”
### Data (* = literal use is more common)

<table>
<thead>
<tr>
<th>expression</th>
<th>literal</th>
<th>non-literal</th>
<th>all</th>
</tr>
</thead>
<tbody>
<tr>
<td>back the wrong horse</td>
<td>0</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>bite off more than one can chew</td>
<td>2</td>
<td>142</td>
<td>144</td>
</tr>
<tr>
<td>bite one’s tongue</td>
<td>16</td>
<td>150</td>
<td>166</td>
</tr>
<tr>
<td>blow one’s own trumpet</td>
<td>0</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>bounce off the wall*</td>
<td>39</td>
<td>7</td>
<td>46</td>
</tr>
<tr>
<td>break the ice</td>
<td>20</td>
<td>521</td>
<td>541</td>
</tr>
<tr>
<td>drop the ball*</td>
<td>688</td>
<td>215</td>
<td>903</td>
</tr>
<tr>
<td>get one’s feet wet</td>
<td>17</td>
<td>140</td>
<td>157</td>
</tr>
<tr>
<td>pass the buck</td>
<td>7</td>
<td>255</td>
<td>262</td>
</tr>
<tr>
<td>play with fire</td>
<td>34</td>
<td>532</td>
<td>566</td>
</tr>
<tr>
<td>pull the trigger*</td>
<td>11</td>
<td>4</td>
<td>15</td>
</tr>
<tr>
<td>rock the boat</td>
<td>8</td>
<td>470</td>
<td>478</td>
</tr>
<tr>
<td>set in stone</td>
<td>9</td>
<td>272</td>
<td>281</td>
</tr>
<tr>
<td>spill the beans</td>
<td>3</td>
<td>172</td>
<td>175</td>
</tr>
<tr>
<td>sweep under the carpet</td>
<td>0</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>swim against the tide</td>
<td>1</td>
<td>125</td>
<td>126</td>
</tr>
<tr>
<td>tear one’s hair out</td>
<td>7</td>
<td>54</td>
<td>61</td>
</tr>
<tr>
<td>all</td>
<td>862</td>
<td>3102</td>
<td>3964</td>
</tr>
</tbody>
</table>
### Results

<table>
<thead>
<tr>
<th></th>
<th>$B_{\text{Maj}}$</th>
<th>$B_{\text{Rep}}$</th>
<th>Graph</th>
<th>$L_{C_d}$</th>
<th>$L_{C_o}$</th>
<th>$\text{Super}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acc</td>
<td>78.25</td>
<td>79.06</td>
<td>79.61</td>
<td>80.50</td>
<td>80.42</td>
<td>95.69</td>
</tr>
<tr>
<td>lit. Prec</td>
<td>-</td>
<td>70.00</td>
<td>52.21</td>
<td>62.26</td>
<td>53.89</td>
<td>84.62</td>
</tr>
<tr>
<td>lit. Rec</td>
<td>-</td>
<td>5.96</td>
<td>67.87</td>
<td>26.21</td>
<td>69.03</td>
<td>96.45</td>
</tr>
<tr>
<td>lit. $F_{\beta=1}$</td>
<td>-</td>
<td>10.98</td>
<td>59.02</td>
<td>36.90</td>
<td>60.53</td>
<td>90.15</td>
</tr>
</tbody>
</table>

- $B_{\text{Maj}}$: majority baseline, i.e., “non-literal” (cf. CForm classifier by Cook et al. (2007), Fazly et al. (to appear))
- $B_{\text{Rep}}$: predict “literal” if an idiom component word is repeated in the context
- Graph: cohesion graph
- $L_{C_d}$: lexical chains optimised on development set
- $L_{C_o}$: lexical chains optimised globally by oracle (upper bound for lexical chains)
- Super: supervised classifier (k-nn) using word overlap (leave-one-out)
### Results

<table>
<thead>
<tr>
<th></th>
<th>$B_{\text{Maj}}$</th>
<th>$B_{\text{Rep}}$</th>
<th>Graph</th>
<th>$L_{C_d}$</th>
<th>$L_{C_o}$</th>
<th>Super</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acc</td>
<td>78.25</td>
<td>79.06</td>
<td>79.61</td>
<td>80.50</td>
<td>80.42</td>
<td>95.69</td>
</tr>
<tr>
<td>lit. Prec</td>
<td>-</td>
<td>70.00</td>
<td>52.21</td>
<td>62.26</td>
<td>53.89</td>
<td>84.62</td>
</tr>
<tr>
<td>lit. Rec</td>
<td>-</td>
<td>5.96</td>
<td>67.87</td>
<td>26.21</td>
<td>69.03</td>
<td>96.45</td>
</tr>
<tr>
<td>lit. $F_{\beta=1}$</td>
<td>-</td>
<td>10.98</td>
<td>59.02</td>
<td>36.90</td>
<td>60.53</td>
<td>90.15</td>
</tr>
</tbody>
</table>

- **$B_{\text{Maj}}$**: majority baseline, i.e., “non-literal” (cf. CForm classifier by Cook et al. (2007), Fazly et al. (to appear))
- **$B_{\text{Rep}}$**: predict “literal” if an idiom component word is repeated in the context
- **Graph**: cohesion graph
- **$L_{C_d}$**: lexical chains optimised on development set
- **$L_{C_o}$**: lexical chains optimised globally by oracle (upper bound for lexical chains)
- **Super**: supervised classifier (k-nn) using word overlap (leave-one-out)
Results

<table>
<thead>
<tr>
<th></th>
<th>$B_{Maj}$</th>
<th>$B_{Rep}$</th>
<th>Graph</th>
<th>$LC_d$</th>
<th>$LC_o$</th>
<th>Super</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acc</td>
<td>78.25</td>
<td>79.06</td>
<td>79.61</td>
<td>80.50</td>
<td>80.42</td>
<td>95.69</td>
</tr>
<tr>
<td>lit. Prec</td>
<td>-</td>
<td>70.00</td>
<td>52.21</td>
<td>62.26</td>
<td>53.89</td>
<td>84.62</td>
</tr>
<tr>
<td>lit. Rec</td>
<td>-</td>
<td>5.96</td>
<td>67.87</td>
<td>26.21</td>
<td>69.03</td>
<td>96.45</td>
</tr>
<tr>
<td>lit. $F_{\beta=1}$</td>
<td>-</td>
<td><strong>10.98</strong></td>
<td><strong>59.02</strong></td>
<td><strong>36.90</strong></td>
<td><strong>60.53</strong></td>
<td><strong>90.15</strong></td>
</tr>
</tbody>
</table>

- $B_{Maj}$: majority baseline, i.e., “non-literal” (cf. CForm classifier by Cook et al. (2007), Fazly et al. (to appear))
- $B_{Rep}$: predict “literal” if an idiom component word is repeated in the context
- Graph: cohesion graph
- $LC_d$: lexical chains optimised on development set
- $LC_o$: lexical chains optimised globally by oracle (upper bound for lexical chains)
- Super: supervised classifier (k-nn) using word overlap (leave-one-out)
## Results

<table>
<thead>
<tr>
<th></th>
<th>$B_{Maj}$</th>
<th>$B_{Rep}$</th>
<th>Graph</th>
<th>$LC_d$</th>
<th>$LC_o$</th>
<th>Super</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acc</td>
<td>78.25</td>
<td>79.06</td>
<td>79.61</td>
<td>80.50</td>
<td>80.42</td>
<td>95.69</td>
</tr>
<tr>
<td>lit. Prec</td>
<td>-</td>
<td>70.00</td>
<td>52.21</td>
<td>62.26</td>
<td>53.89</td>
<td>84.62</td>
</tr>
<tr>
<td>lit. Rec</td>
<td>-</td>
<td>5.96</td>
<td>67.87</td>
<td>26.21</td>
<td>69.03</td>
<td>96.45</td>
</tr>
<tr>
<td>lit. $F_{\beta=1}$</td>
<td>-</td>
<td>10.98</td>
<td>59.02</td>
<td>36.90</td>
<td>60.53</td>
<td>90.15</td>
</tr>
</tbody>
</table>

- **$B_{Maj}$**: majority baseline, i.e., “non-literal” (cf. CForm classifier by Cook et al. (2007), Fazly et al. (to appear))
- **$B_{Rep}$**: predict “literal” if an idiom component word is repeated in the context
- **Graph**: cohesion graph
- **$LC_d$**: lexical chains optimised on development set
- **$LC_o$**: lexical chains optimised globally by oracle (upper bound for lexical chains)
- **Super**: supervised classifier (k-nn) using word overlap (leave-one-out)
### Results

<table>
<thead>
<tr>
<th></th>
<th>$B_{\text{Maj}}$</th>
<th>$B_{\text{Rep}}$</th>
<th>Graph</th>
<th>$LC_d$</th>
<th>$LC_o$</th>
<th>Super</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acc</td>
<td>78.25</td>
<td>79.06</td>
<td>79.61</td>
<td>80.50</td>
<td>80.42</td>
<td>95.69</td>
</tr>
<tr>
<td>lit. Prec</td>
<td>-</td>
<td>70.00</td>
<td>52.21</td>
<td>62.26</td>
<td>53.89</td>
<td>84.62</td>
</tr>
<tr>
<td>lit. Rec</td>
<td>-</td>
<td>9.56</td>
<td>67.87</td>
<td>26.21</td>
<td>69.03</td>
<td>96.45</td>
</tr>
<tr>
<td>lit. $F_{\beta=1}$</td>
<td>-</td>
<td>10.98</td>
<td>59.02</td>
<td>36.90</td>
<td>60.53</td>
<td>90.15</td>
</tr>
</tbody>
</table>

- $B_{\text{Maj}}$: majority baseline, i.e., “non-literal” (cf. CForm classifier by Cook et al. (2007), Fazly et al. (to appear))
- $B_{\text{Rep}}$: predict “literal” if an idiom component word is repeated in the context
- Graph: cohesion graph
- $LC_d$: lexical chains optimised on development set
- $LC_o$: lexical chains optimised globally by oracle (upper bound for lexical chains)
- Super: supervised classifier (k-nn) using word overlap (leave-one-out)
### Results

<table>
<thead>
<tr>
<th></th>
<th>$B_{Maj}$</th>
<th>$B_{Rep}$</th>
<th>Graph</th>
<th>$L_{C_d}$</th>
<th>$L_{C_o}$</th>
<th>Super</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acc</td>
<td>78.25</td>
<td>79.06</td>
<td>79.61</td>
<td>80.50</td>
<td>80.42</td>
<td>95.69</td>
</tr>
<tr>
<td>lit. Prec</td>
<td>-</td>
<td>70.00</td>
<td>52.21</td>
<td>62.26</td>
<td>53.89</td>
<td>84.62</td>
</tr>
<tr>
<td>lit. Rec</td>
<td>-</td>
<td>5.96</td>
<td>67.87</td>
<td>26.21</td>
<td>69.03</td>
<td>96.45</td>
</tr>
<tr>
<td>lit. $F_{\beta=1}$</td>
<td>-</td>
<td>10.98</td>
<td>59.02</td>
<td>36.90</td>
<td>60.53</td>
<td>90.15</td>
</tr>
</tbody>
</table>

- **$B_{Maj}$**: majority baseline, i.e., “non-literal” (cf. CForm classifier by Cook et al. (2007), Fazly et al. (to appear))
- **$B_{Rep}$**: predict “literal” if an idiom component word is repeated in the context
- **Graph**: cohesion graph
- **$L_{C_d}$**: lexical chains optimised on development set
- **$L_{C_o}$**: lexical chains optimised globally by oracle (upper bound for lexical chains)
- **Super**: supervised classifier (k-nn) using word overlap (leave-one-out)
<table>
<thead>
<tr>
<th></th>
<th>$B_{Maj}$</th>
<th>$B_{Rep}$</th>
<th>Graph</th>
<th>$L_{C_d}$</th>
<th>$L_{C_o}$</th>
<th>Super</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acc</td>
<td>78.25</td>
<td>79.06</td>
<td>79.61</td>
<td>80.50</td>
<td>80.42</td>
<td>95.69</td>
</tr>
<tr>
<td>lit. Prec</td>
<td>-</td>
<td>70.00</td>
<td>52.21</td>
<td>62.26</td>
<td>53.89</td>
<td>84.62</td>
</tr>
<tr>
<td>lit. Rec</td>
<td>-</td>
<td>5.96</td>
<td>67.87</td>
<td>26.21</td>
<td>69.03</td>
<td>96.45</td>
</tr>
<tr>
<td>lit. $F_{\beta=1}$</td>
<td>-</td>
<td>10.98</td>
<td>59.02</td>
<td>36.90</td>
<td>60.53</td>
<td>90.15</td>
</tr>
</tbody>
</table>

- $B_{Maj}$: majority baseline, i.e., “non-literal” (cf. CFom classifier by Cook et al. (2007), Fazly et al. (to appear))
- $B_{Rep}$: predict “literal” if an idiom component word is repeated in the context
- Graph: cohesion graph
- $L_{C_d}$: lexical chains optimised on development set
- $L_{C_o}$: lexical chains optimised globally by oracle (upper bound for lexical chains)
- Super: supervised classifier (k-nn) using word overlap (leave-one-out)
Results

<table>
<thead>
<tr>
<th></th>
<th>$B_{\text{Maj}}$</th>
<th>$B_{\text{Rep}}$</th>
<th>Graph</th>
<th>$L_{\text{C}_d}$</th>
<th>$L_{\text{C}_o}$</th>
<th>Super</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acc</td>
<td>78.25</td>
<td>79.06</td>
<td>79.61</td>
<td>80.50</td>
<td>80.42</td>
<td>95.69</td>
</tr>
<tr>
<td>lit. Prec</td>
<td>-</td>
<td>70.00</td>
<td>52.21</td>
<td>62.26</td>
<td>53.89</td>
<td>84.62</td>
</tr>
<tr>
<td>lit. Rec</td>
<td>-</td>
<td>5.96</td>
<td>67.87</td>
<td>26.21</td>
<td>69.03</td>
<td>96.45</td>
</tr>
<tr>
<td>lit. $F_{\beta=1}$</td>
<td>-</td>
<td>10.98</td>
<td>59.02</td>
<td>36.90</td>
<td>60.53</td>
<td>90.15</td>
</tr>
</tbody>
</table>

- $B_{\text{Maj}}$: majority baseline, i.e., “non-literal” (cf. CForm classifier by Cook et al. (2007), Fazly et al. (to appear))
- $B_{\text{Rep}}$: predict “literal” if an idiom component word is repeated in the context
- Graph: cohesion graph
- $L_{\text{C}_d}$: lexical chains optimised on development set
- $L_{\text{C}_o}$: lexical chains optimised globally by oracle (upper bound for lexical chains)
- Super: supervised classifier (k-nn) using word overlap (leave-one-out)
<table>
<thead>
<tr>
<th></th>
<th>$B_{Maj}$</th>
<th>$B_{Rep}$</th>
<th>Graph</th>
<th>$L_{C_d}$</th>
<th>$L_{C_o}$</th>
<th>Super</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acc</td>
<td>78.25</td>
<td>79.06</td>
<td>79.61</td>
<td>80.50</td>
<td>80.42</td>
<td>95.69</td>
</tr>
<tr>
<td>lit. Prec</td>
<td>-</td>
<td>70.00</td>
<td>52.21</td>
<td>62.26</td>
<td>53.89</td>
<td>84.62</td>
</tr>
<tr>
<td>lit. Rec</td>
<td>-</td>
<td>5.96</td>
<td>67.87</td>
<td>26.21</td>
<td>69.03</td>
<td>96.45</td>
</tr>
<tr>
<td>lit. $F_{\beta=1}$</td>
<td>-</td>
<td>10.98</td>
<td><strong>59.02</strong></td>
<td><strong>36.90</strong></td>
<td><strong>60.53</strong></td>
<td><strong>90.15</strong></td>
</tr>
</tbody>
</table>

- **$B_{Maj}$**: majority baseline, i.e., “non-literal” (cf. CForm classifier by Cook et al. (2007), Fazly et al. (to appear))
- **$B_{Rep}$**: predict “literal” if an idiom component word is repeated in the context
- **Graph**: cohesion graph
- **$L_{C_d}$**: lexical chains optimised on development set
- **$L_{C_o}$**: lexical chains optimised globally by oracle (upper bound for lexical chains)
- **Super**: supervised classifier (k-nn) using word overlap (leave-one-out)
Combining Unsupervised and Supervised Learning
Limitations

only cohesion is taken into account, other cues are ignored

Syntactic cues, local context

Gujral will meet Sharif on Monday and discuss bilateral relations, the Press Trust of India added. The minister said Sharif and Gujral would be able to break the ice over Kashmir.

Lexical cues, global context

Next week the two diplomats will meet in an attempt to break the ice between the two nations. A crucial issue in the talks will be the long-running water dispute.
Combining Unsupervised and Supervised Classification

Test Set

Supervised Classifier

Cohesion-based Classifier

Caroline Sporleder

Detecting Figurative Language (25/32)
Combining Unsupervised and Supervised Classification

Cohesion-based Classifier

apply

Test Set
Combining Unsupervised and Supervised Classification

Cohesion-based Classifier

apply

Test Set

Training Set

Caroline Sporleder
Detecting Figurative Language
Combining Unsupervised and Supervised Classification

Test Set

Training Set

Supervised Classifier

training

Caroline Sporleder
Detecting Figurative Language
Combining Unsupervised and Supervised Classification

Test Set

Supervised Classifier

Cohesion-based Classifier

Training Set

Test Set

Supervised Classifier

Cohesion-based Classifier

apply

training
Two-Stage Classification

1. apply unsupervised cohesion-based classifier to all examples
2. train supervised classifier on those examples about which unsupervised classifier was most confident
3. apply trained supervised classifier to remaining examples
Supervised classifier

Machine Learning Framework

- support vector machines

Features

- salient words for literal usage (300 most frequent words):

\[ sal_{lit}(w) = \frac{\log f_{lit}(w) \times i_{lit}(w)}{\log f_{nonlit}(w) \times i_{nonlit}(w)} \]

\((f_{lit}(w))\) is the frequency of \(w\) for literal usage; \(i_{lit}(w)\) is the number of literal usages which co-occur with word \(w\)

- related words (300 similar words)
- the 100 highest relatedness values
- the connectivity with and without the idiom
Further Extensions

Iterating

- extend training set iteratively

Boosting the literal class

- automatically extract non-canonical form examples (e.g., “rock the boat” → “rock the ship”)
- label them as literal
- adding them to the training set
<table>
<thead>
<tr>
<th>Model</th>
<th>Prec$_{lit}$</th>
<th>Rec$_{lit}$</th>
<th>F-Score$_{lit}$</th>
<th>Acc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>unsup.</td>
<td>50.04</td>
<td>69.72</td>
<td>58.26</td>
<td>78.38</td>
</tr>
<tr>
<td>combined</td>
<td>83.86</td>
<td>45.82</td>
<td>59.26</td>
<td>86.30</td>
</tr>
<tr>
<td>combined+boost</td>
<td>70.26</td>
<td>62.76</td>
<td>66.30</td>
<td>86.13</td>
</tr>
<tr>
<td>combined+it*</td>
<td>85.68</td>
<td>46.52</td>
<td>60.30</td>
<td>86.68</td>
</tr>
<tr>
<td>combined+boost+it*</td>
<td>71.86</td>
<td>66.36</td>
<td>69.00</td>
<td>87.03</td>
</tr>
<tr>
<td>Model</td>
<td>$\text{Prec}_{\text{lit}}$</td>
<td>$\text{Rec}_{\text{lit}}$</td>
<td>$\text{F-Score}_{\text{lit}}$</td>
<td>Acc.</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------------------</td>
<td>-----------------------------</td>
<td>------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>unsup.</td>
<td>50.04</td>
<td>69.72</td>
<td>58.26</td>
<td>78.38</td>
</tr>
<tr>
<td>combined</td>
<td>83.86</td>
<td>45.82</td>
<td>59.26</td>
<td>86.30</td>
</tr>
<tr>
<td>combined+boost</td>
<td>70.26</td>
<td>62.76</td>
<td>66.30</td>
<td>86.13</td>
</tr>
<tr>
<td>combined+it*</td>
<td>85.68</td>
<td>46.52</td>
<td>60.30</td>
<td>86.68</td>
</tr>
<tr>
<td>combined+boost+it*</td>
<td>71.86</td>
<td>66.36</td>
<td>69.00</td>
<td>87.03</td>
</tr>
<tr>
<td>Model</td>
<td>$\text{Prec}_{\text{lit}}$</td>
<td>$\text{Rec}_{\text{lit}}$</td>
<td>$\text{F-Score}_{\text{lit}}$</td>
<td>Acc.</td>
</tr>
<tr>
<td>---------------</td>
<td>-----------------------------</td>
<td>-----------------------------</td>
<td>-------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>unsup.</td>
<td>50.04</td>
<td>69.72</td>
<td>58.26</td>
<td>78.38</td>
</tr>
<tr>
<td>combined</td>
<td>83.86</td>
<td>45.82</td>
<td>59.26</td>
<td>86.30</td>
</tr>
<tr>
<td>combined+boost</td>
<td>70.26</td>
<td>62.76</td>
<td>66.30</td>
<td>86.13</td>
</tr>
<tr>
<td>combined+it*</td>
<td>85.68</td>
<td>46.52</td>
<td>60.30</td>
<td>86.68</td>
</tr>
<tr>
<td>combined+boost+it*</td>
<td>71.86</td>
<td>66.36</td>
<td>69.00</td>
<td>87.03</td>
</tr>
<tr>
<td>Model</td>
<td>Prec\textsubscript{lit}</td>
<td>Rec\textsubscript{lit}</td>
<td>F-Score\textsubscript{lit}</td>
<td>Acc.</td>
</tr>
<tr>
<td>----------------</td>
<td>--------------------------</td>
<td>--------------------------</td>
<td>-----------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>unsup.</td>
<td>50.04</td>
<td>69.72</td>
<td>58.26</td>
<td>78.38</td>
</tr>
<tr>
<td>combined</td>
<td>83.86</td>
<td>45.82</td>
<td>59.26</td>
<td>86.30</td>
</tr>
<tr>
<td>combined+boost</td>
<td>70.26</td>
<td>62.76</td>
<td>66.30</td>
<td>86.13</td>
</tr>
<tr>
<td>combined+it*</td>
<td>85.68</td>
<td>46.52</td>
<td>60.30</td>
<td>86.68</td>
</tr>
<tr>
<td>combined+boost+it*</td>
<td>71.86</td>
<td>66.36</td>
<td>69.00</td>
<td>87.03</td>
</tr>
<tr>
<td>Model</td>
<td>Prec$_{lit}$</td>
<td>Rec$_{lit}$</td>
<td>F-Score$_{lit}$</td>
<td>Acc.</td>
</tr>
<tr>
<td>--------------------</td>
<td>--------------</td>
<td>--------------</td>
<td>-----------------</td>
<td>-------</td>
</tr>
<tr>
<td>unsup.</td>
<td>50.04</td>
<td>69.72</td>
<td>58.26</td>
<td>78.38</td>
</tr>
<tr>
<td>combined</td>
<td>83.86</td>
<td>45.82</td>
<td>59.26</td>
<td>86.30</td>
</tr>
<tr>
<td>combined+boost</td>
<td>70.26</td>
<td>62.76</td>
<td>66.30</td>
<td>86.13</td>
</tr>
<tr>
<td>combined+it*</td>
<td>85.68</td>
<td>46.52</td>
<td>60.30</td>
<td>86.68</td>
</tr>
<tr>
<td>combined+boost+it*</td>
<td>71.86</td>
<td>66.36</td>
<td>69.00</td>
<td>87.03</td>
</tr>
</tbody>
</table>
Conclusion
Conclusion

- The **cohesive structure** of a text provides good cues for distinguishing literal and non-literal language.
- **Cohesion graphs** perform as well as lexical chains while being fully unsupervised.
- combination with a **second stage supervised classifier** leads to further performance gains.
Be able to handle the difficult cases!

Non-Literal Use with Lexical Chain

Saying that the Americans were "playing with fire" the official press speculated that the "gunpowder barrel" which is Taiwan might well "explode" if Washington and Taipei do not put a stop to their "incendiary gesticulations."