Unsupervised Recognition of Literal and Non-Literal Use of Idiomatic Expressions

Caroline Sporleder and Linlin Li

MMCI / Computational Linguistics, Saarland University

{csporled,linlin}@coli.uni-sb.de

EACL, Athens
April 3, 2009
Why is Non-Literal Language a Problem?

Examples of Non-Literal Language

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.
Why is Non-Literal Language a Problem?

Examples of Non-Literal Language

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-Literal Expressions (idioms, metaphors etc.) . . .

- occur frequently in language
- often behave idiosyncratically
- have to be recognised automatically to be analysed and interpreted in an appropriate way

Caroline Sporleder, Linlin Li
Recognition of Literal and Non-Literal Use of Idioms
Dealing with Idioms

Most previous research:

- automatic idiom extraction methods (type-based classification)

But:

- doesn’t work for creative language use
- potentially idiomatic expressions can be used in literal sense

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.*
Dealing with Idioms

Most previous research:

- automatic idiom extraction methods (type-based classification)

But:

- doesn’t work for creative language use
- potentially idiomatic expressions can be used in literal sense

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?

<table>
<thead>
<tr>
<th>Literal Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>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.</em></td>
</tr>
</tbody>
</table>

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.
Limitations of the Cohesion-Based Approach

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
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**:
  - lexical chains
  - cohesion graphs
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: target words, M: total number of pages indexed)
Literal Use

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

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

Four Lexical Chains:
- Chain 1: *Dad*
- Chain 2: *break*
- Chain 3: *ice – water*
- Chain 4: *chicken – troughs*
Literal Use

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

Four Lexical Chains:

- Chain 1: Dad
- Chain 2: break
- Chain 3: ice – water
- Chain 4: chicken – troughs

⇒ Literal!
Modelling: Lexical Chains

Drawbacks:

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

- approach is sensitive to chaining algorithm and parameter settings
Dad had to break the ice on the chicken troughs so that they could get water.
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
Modelling Cohesion: Cohesion Graphs

Literal Use

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

\[ \text{with idiom:} \]
\[ \text{avg. connectivity}=0.34 \]

\[ \text{without idiom:} \]
\[ \text{avg. connectivity}=0.33 \]

⇒ Literal!
Experiments

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” (862 instances) or “non-literal” (3102 instances)
## 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>$B_{Maj}$</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)
## 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>$B_{Maj}$</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>$L_{C_d}$</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>-</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)
### 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)
<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>$\text{Super}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Acc</strong></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><strong>lit. Prec</strong></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><strong>lit. Rec</strong></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><strong>lit. F_{\beta=1}</strong></td>
<td>-</td>
<td>10.98</td>
<td>59.02</td>
<td>36.90</td>
<td><strong>60.53</strong></td>
<td><strong>90.15</strong></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)
## 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)
### 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><strong>Acc</strong></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)
<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)
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.
Future Work

- bootstrapping by combining supervised and unsupervised techniques
- more sophisticated graphical models
- apply to other cases of non-literal language (e.g., spontaneously created metaphors)