Detecting and Processing Figurative Language in Discourse

Caroline Sporleder

Computational Linguistics & Digital Humanities
Trier University

sporledc@uni-trier.de

NODALIDA 2013
May 23, 2013
Oslo*

Oslo, the Viking city, the culture city, the winter capital, the city of rolling green hills and spectacular fjords is the capital city of Norway, aptly credited with many adjectives. The city breathes adventure with rich history pulsing through its veins. Soak in the beauty of the Vigeland Park and feel the wind in your hair as you ski down the jewelled snow clad slopes of Oslo. To make the most of your holidays in Oslo, we have listed the top highlights of Oslo such as the top landmarks, top bar and must do activities in Oslo. We have also painted a picture of the communication, transport system for your easier navigation in Oslo.

* Source: http://www.oslo.com/v/city_info/
Oslo

Oslo, the Viking city, the culture city, the winter capital, the city of rolling green hills and spectacular fjords is the capital city of Norway, aptly credited with many adjectives. The city breathes adventure with rich history pulsing through its veins. Soak in the beauty of the Vigeland Park and feel the wind in your hair as you ski down the jewelled snow clad slopes of Oslo. To make the most of your holidays in Oslo, we have listed the top highlights of Oslo such as the top landmarks, top bar and must do activities in Oslo. We have also painted a picture of the communication, transport system for your easier navigation in Oslo.

* Source: http://www.oslo.com/v/city_info/
Apple’s Tim Cook defends tax strategy in Senate*

The technology giant Apple has been defending itself against accusations that it’s avoided paying tax on tens of billions of dollars in profits.

Chief executive Tim Cook told a US Senate committee Apple paid all the taxes it owed, complying with both the law, and the spirit of the law.

The appearance comes just a day after the same panel branded Apple’s complex structure the ”Holy Grail of tax avoidance”.

On Monday, the Subcommittee said Apple had used ”a complex web of offshore entities” to avoid paying billions of dollars in US income taxes. But it said there was no indication the firm acted illegally.

Apple has a cash stockpile of $145bn, but the Senate committee said $102bn of this was held offshore.

* Source: BBC news, 21.5.2013
Apple’s Tim Cook defends tax strategy in Senate*

The technology giant Apple has been defending itself against accusations that it’s avoided paying tax on tens of billions of dollars in profits.

Chief executive Tim Cook told a US Senate committee Apple paid all the taxes it owed, complying with both the law, and the spirit of the law. . . .

The appearance comes just a day after the same panel branded Apple’s complex structure the ”Holy Grail of tax avoidance” . . . .

On Monday, the Subcommittee said Apple had used ”a complex web of offshore entities” to avoid paying billions of dollars in US income taxes. But it said there was no indication the firm acted illegally.

Apple has a cash stockpile of $145bn, but the Senate committee said $102bn of this was held offshore.

* Source: BBC news, 21.5.2013
Why bother?

Figurative expressions . . .

- are frequent
- can cause problems for grammatical processing (parsing, semantic role labelling)
- provide a challenge for natural language understanding (esp. non-lexicalised expressions)
Main Types of Figurative Language

Metaphor

implicit comparison of two superficially unrelated concepts/entities
(*tax haven, web of offshore entities, head of a company*)

Idioms

sequence of words whose meaning is not (entirely) compositional
(*break the ice, get one’s feet wet, by and large*)
Main Types of Figurative Language

Metaphor
implicit comparison of two superficially unrelated concepts/entities
(tax haven, web of offshore entities, head of a company)

Idioms
sequence of words whose meaning is not (entirely) compositional
(break the ice, get one’s feet wet, by and large)

⇒ This talk’s focus is on idioms and on detection (rather than interpretation)
Type-Based Detection
aka “idiom extraction” (Baldwin & Kim, 2010)

- match target expression against an electronically readable idiom dictionary, or
- use word association statistics to classify target expression into ‘idiom’ or 'not idiom'

⇒ identify target expression out of context
Type vs. Token-Based Detection (1)

Type-Based Detection
aka “idiom extraction” (Baldwin & Kim, 2010)

- match target expression against an electronically readable idiom dictionary, or
- use word association statistics to classify target expression into ‘idiom’ or ’not idiom’

⇒ identify target expression out of context

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
Token-Based Detection
aka “idiom identification” (Baldwin & Kim, 2010)

- many idioms share their form with literal, compositional phrases

⇒ need to be detected in context

Literal vs. Non-Literal Usage

1. Dad had to break the ice on the chicken troughs so that they could get water. [GIGA NYT200008]

2. If you’ve just moved to a new area a good way to break the ice for you and your child is a parent and toddler group. [BNC AAY]
Outline of the Talk

1. Case Study: Idioms in Context
2. Cohesion-Based Idiom Detection
3. Including Further Linguistic Cues
4. Conclusion
Case Study: Idioms in Context
Idiom Annotated Corpora

**IDIX**

**IDIoms In conteXt** [Sporleder et al., LREC 2010]

- manually disambiguated occurrences of potentially idiomatic expressions
- BNC add-on
- so far 5,836 annotated instances, 78 expressions
- avg. 75 instances per expression, ranging from one (lower the bar) to 540 (ring the/a bell)
- 93.19% inter-annotator agreement, $K=.87$

**Similar Resources Include:**

- VNC-Tokens Dataset (Cook et al., 2008)
- German token disambiguated data (Fritzinger et al., 2010)
- Japanese token disambiguated data (Hashimoto & Kawahara, 2008)
Potentially Idiomatic Expressions in IDIX (1)

Literal Usage

(1) Dad had to break the ice on the chicken troughs so that they could get water. [GIGA NYT200008]

Non-Literal Usage

(2) If you've just moved to a new area a good way to break the ice for you and your child is a parent and toddler group. [BNC AAY]
Meta-Linguistic Usage
(3) It has long been recognised that expressions such as to pull someone’s leg, to have a bee in one’s bonnet, to kick the bucket [...] etc. are semantically peculiar. [BNC FAC]

Mixed Literal Non-Literal Usage
(4) Left holding the baby, single mothers find it hard to fend for themselves. [BNC CRA]
(5) You’re like a restless bird in a cage. When you get out of the cage, you’ll fly very high.

(BNC FR6)

(6) Political prudence and the dangers of a frontal attack on the Church restrained them to the sale of common lands and the abolition of civil entails, ’pulling up by the roots the tree which bears such bitter fruits’.

(BNC FB7)
Multiple Non-Literal Cases

“topics dealt with”

(7) The ground covered by both books is, in the early stages, fairly similar.

“distance moved”

(8) Only 5ft 8ins tall and under 10st in weight, he covers the ground quicker than anyone I have recently seen, with the exception of one he resembles, South Africa’s Jonty Rhodes.

Literal Usage

(9) The most reliable standby for climbing or simply to cover the ground, has to be ivy.
Figure: Percentage of labels in the currently annotated data set
Cohesion-Based Idiom Detection
Previous Approaches:

- Fotherfill and Baldwin (2012, 2011): supervised classifier
- 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. (2009): unsupervised, predict non-literally if idiom is in canonical form (≈ dictionary form)
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.
Identifying Idiomatic Usage [Sporleder & Li, EACL 2009]

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

- Yes ⇒ literal usage
- No ⇒ 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
- **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: 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 $t = 0.5$; $\text{sim}(\text{Dad,ice}) = 0.1$

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

**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*
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
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
Method 1: Lexical Chains

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

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

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

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
Method 1: Lexical Chains

Literal Use

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

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

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

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

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

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 – troughs
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 \( t = 0.5 \); \( \text{sim(Dad, water)} = 0.1 \)

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 \( t=0.5 \); \( \text{sim(break, water)}=0.1 \)

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

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 $t=0.5$; $\text{sim(throughs,water)}=0.6$

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
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.

with idiom: break ice

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
Literal Use

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

\[
\begin{array}{cccc}
\text{Dad} & 0.1 & 0.3 & 0.1 \\
0.1 & 0.7 & 0.1 \\
0.3 & 0.1 & 0.6 \\
0.7 & 0.1 & 0.4 \\
0.1 & 0.6 & 0.4 \\
\end{array}
\]

avg. connectivity=0.34

**with idiom:**

avg. connectivity=0.33

\[\Rightarrow \text{Literal!}\]
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”
## Experiments

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>
<tr>
<td></td>
<td>$B_{Maj}$</td>
<td>$B_{Rep}$</td>
<td>Graph</td>
</tr>
<tr>
<td>--------</td>
<td>-----------</td>
<td>-----------</td>
<td>---------</td>
</tr>
<tr>
<td>Acc</td>
<td>78.25</td>
<td>79.06</td>
<td>79.61</td>
</tr>
<tr>
<td>lit. Prec</td>
<td>-</td>
<td>70.00</td>
<td>52.21</td>
</tr>
<tr>
<td>lit. Rec</td>
<td>-</td>
<td>5.96</td>
<td>67.87</td>
</tr>
<tr>
<td>lit. F$_{\beta=1}$</td>
<td>-</td>
<td>10.98</td>
<td>59.02</td>
</tr>
</tbody>
</table>

- **$B_{Maj}$**: majority baseline, i.e., “non-literal” (cf. CForm classifier by Cook et al. (2007), Fazly et al. (2009))
- **$B_{Rep}$**: 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)
### 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>
</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>
</tr>
<tr>
<td>lit. Prec</td>
<td>-</td>
<td>70.00</td>
<td>52.21</td>
<td>62.26</td>
<td>53.89</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>
</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>
</tr>
</tbody>
</table>

- **$B_{Maj}$**: majority baseline, i.e., “non-literal” (cf. CForm classifier by Cook et al. (2007), Fazly et al. (2009))
- **$B_{Rep}$**: 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)
### 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>
</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>
</tr>
<tr>
<td><em>lit. Prec</em></td>
<td>-</td>
<td>70.00</td>
<td>52.21</td>
<td>62.26</td>
<td>53.89</td>
</tr>
<tr>
<td><em>lit. Rec</em></td>
<td>-</td>
<td>5.96</td>
<td>67.87</td>
<td>26.21</td>
<td>69.03</td>
</tr>
<tr>
<td><em>lit. $F_{\beta=1}$</em></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>
</tr>
</tbody>
</table>

- **$B_{Maj}$**: majority baseline, i.e., “non-literal” (cf. CForm classifier by Cook et al. (2007), Fazly et al. (2009))
- **$B_{Rep}$**: 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)
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>
</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>
</tr>
<tr>
<td>lit. Prec</td>
<td>-</td>
<td>70.00</td>
<td>52.21</td>
<td>62.26</td>
<td>53.89</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>
</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>
</tr>
</tbody>
</table>

- $B_{Maj}$: majority baseline, i.e., “non-literal” (cf. CForm classifier by Cook et al. (2007), Fazly et al. (2009))
- $B_{Rep}$: 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)
<table>
<thead>
<tr>
<th></th>
<th>$B_{Maj}$</th>
<th>$B_{Rep}$</th>
<th>Graph</th>
<th>$LC_d$</th>
<th>$LC_o$</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>
</tr>
<tr>
<td>$lit. Prec$</td>
<td>-</td>
<td>70.00</td>
<td>52.21</td>
<td>62.26</td>
<td>53.89</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>
</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>
</tr>
</tbody>
</table>

- **$B_{Maj}$**: majority baseline, i.e., “non-literal” (cf. CForm classifier by Cook et al. (2007), Fazly et al. (2009))
- **$B_{Rep}$**: 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)
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>
</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>
</tr>
<tr>
<td>lit. Prec</td>
<td>-</td>
<td>70.00</td>
<td>52.21</td>
<td>62.26</td>
<td>53.89</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>
</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>
</tr>
</tbody>
</table>

- $B_{Maj}$: majority baseline, i.e., “non-literal” (cf. CForm classifier by Cook et al. (2007), Fazly et al. (2009))
- $B_{Rep}$: 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)
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>
</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>
</tr>
<tr>
<td>lit. Prec</td>
<td>-</td>
<td>70.00</td>
<td>52.21</td>
<td>62.26</td>
<td>53.89</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>
</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>
</tr>
</tbody>
</table>

- $B_{Maj}$: majority baseline, i.e., “non-literal” (cf. CForm classifier by Cook et al. (2007), Fazly et al. (2009))
- $B_{Rep}$: 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)
<table>
<thead>
<tr>
<th></th>
<th>$B_{Maj}$</th>
<th>$B_{Rep}$</th>
<th>Graph</th>
<th>$LC_d$</th>
<th>$LC_o$</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>
</tr>
<tr>
<td>lit. Prec</td>
<td>-</td>
<td>70.00</td>
<td>52.21</td>
<td>62.26</td>
<td>53.89</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>
</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>
</tr>
</tbody>
</table>

- **$B_{Maj}$**: majority baseline, i.e., “non-literal” (cf. CForm classifier by Cook et al. (2007), Fazly et al. (2009))
- **$B_{Rep}$**: 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)
## Results

<table>
<thead>
<tr>
<th></th>
<th>B\textsubscript{Maj}</th>
<th>B\textsubscript{Rep}</th>
<th>Graph</th>
<th>L\textsubscript{C}\textsubscript{d}</th>
<th>L\textsubscript{C}\textsubscript{o}</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>
</tr>
<tr>
<td>lit. Prec</td>
<td>-</td>
<td>70.00</td>
<td>52.21</td>
<td>62.26</td>
<td>53.89</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>
</tr>
<tr>
<td>lit. F\textsubscript{β=1}</td>
<td>-</td>
<td>10.98</td>
<td>59.02</td>
<td>36.90</td>
<td>60.53</td>
</tr>
</tbody>
</table>

- **B\textsubscript{Maj}**: majority baseline, i.e., “non-literal” (cf. CForm classifier by Cook et al. (2007), Fazly et al. (2009))
- **B\textsubscript{Rep}**: predict “literal” if an idiom component word is repeated in the context
- **Graph**: cohesion graph
- **L\textsubscript{C}\textsubscript{d}**: lexical chains optimised on development set
- **L\textsubscript{C}\textsubscript{o}**: lexical chains optimised globally by oracle (upper bound for lexical chains)
Including Further Linguistic Cues
Beyond Cohesion

Cohesion is only one cue, there may be others

- cohesion involving figurative usage
- local context (prepositions, modifiers)
- selectional preferences
- syntax (argument structure, coordination)
- modality, negation
- typography
Cohesion with Non-Literal Meaning

Non-Literal Usage

“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.

Literal Usage

Meanwhile in Germany, the cold penetrated Cologne cathedral, where worshippers had to break the ice on the frozen holy water in the font.
Indicator Words . . . can be misleading

- In the documentary the producer **literally spills the beans** on the real deal behind the movie production.
- The new philosophy is blatantly permissive and **literally passes the buck** to the House’s other committees.

Prepositions, modifiers work better

- The wiki includes a page of tasks suitable for those **just getting their feet wet**.
- Would the visit of the minister help **break the ice** between India and Pakistan?
Dudayev repeated his frequent warnings that Russia was playing with fire.

Edwards usually manages to break the ice with the taciturn monarch.
Target expression as a dependent of a preposition

- Ross headed back last week to Washington to brief president Bill Clinton on the Hebron talks after achieving a breakthrough **in breaking the ice** in the Hebron talks by arranging an Arafat-Netanyahu summit.

Co-ordination with the target expression

- They may **break the ice** and **fall through**.
US defender Alexi Lalas twice went close to forcing an equaliser, first with a glancing equaliser from a Paul Caligiuri free kick and then from a Wynalda corner when Prunea dropped the ball \[on the ground\] only \[for Tibor Selyme to kick frantically clear\].

“Clinton dropped the ball \[on this\],” said John Parachini.
If the context contains modals or negation this may (?) shift the probabilities of idiomatic vs. literal usage...

**Modals**

- The visit *may break the ice* between India and Pakistan.
- Dad *had to break the ice* on the chicken troughs.

**Negation**

- If the pond is frozen please *do not break the ice* as this can be deadly to your fish.
Do consider “getting your feet wet” online, using some of the technology that is now available to us.
### Supervised, idiom-specific models [Li & Sporleder, Coling 2010]

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>cohesion</td>
<td>90.42</td>
<td>76.44</td>
<td>82.85</td>
<td>93.36</td>
</tr>
<tr>
<td>loc. Contxt</td>
<td>76.51</td>
<td>88.34</td>
<td>82.00</td>
<td>91.86</td>
</tr>
<tr>
<td>select. Pref</td>
<td>76.49</td>
<td>88.22</td>
<td>81.94</td>
<td>91.84</td>
</tr>
<tr>
<td>Syntax</td>
<td>76.30</td>
<td>86.13</td>
<td>80.92</td>
<td>91.48</td>
</tr>
<tr>
<td>All</td>
<td>89.84</td>
<td>77.06</td>
<td>82.96</td>
<td>93.36</td>
</tr>
</tbody>
</table>
### Supervised, generic models  [Li & Sporleder, Coling 2010]

<table>
<thead>
<tr>
<th>Feature</th>
<th>Avg. literal</th>
<th></th>
<th></th>
<th>Avg. Acc.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prec.</td>
<td>Rec.</td>
<td>F-Score</td>
<td>Acc.</td>
</tr>
<tr>
<td>cohesion</td>
<td>82.53</td>
<td>60.86</td>
<td>70.06</td>
<td>89.08</td>
</tr>
<tr>
<td>loc.Contxt</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>79.01</td>
</tr>
<tr>
<td>select.Pref</td>
<td>62.45</td>
<td>20.00</td>
<td>30.30</td>
<td>80.69</td>
</tr>
<tr>
<td>Syntax</td>
<td>50.83</td>
<td>59.88</td>
<td>54.99</td>
<td>79.42</td>
</tr>
<tr>
<td>All</td>
<td>89.59</td>
<td>65.77</td>
<td>73.22</td>
<td>89.90</td>
</tr>
</tbody>
</table>
Many features don’t generalise across expressions
requires annotated examples for each target expression
this is infeasible

\( \Rightarrow \text{Can we do better (e.g. bootstrapping)?} \)
**Two-Stage Classification** [Li & Sporleder, EMNLP 2009]

1. apply unsupervised cohesion-based classifier to all examples
2. train supervised classifier on those examples about which unsupervised classifier was most confident
   focus on cohesion features
3. apply trained supervised classifier to remaining examples
Combining Unsupervised and Supervised Classification (2)

Cohesion-based Classifier

apply

Test Set

Training Set

Caroline Sporleder  Figurative Language (42/52)
Combining Unsupervised and Supervised Classification

- Test Set
- Supervised Classifier
- Cohesion-based Classifier
- Training Set

Caroline Sporleder
Figurative Language (42/52)
Combining Unsupervised and Supervised Classification

Test Set
Supervised Classifier
Cohesion-based Classifier
apply
Training Set

Supervised Classifier
apply
training

Test Set

Training Set
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\textsubscript{lit}</th>
<th>Rec\textsubscript{lit}</th>
<th>F-Score\textsubscript{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>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(_{\text{lit}})</td>
<td>Rec(_{\text{lit}})</td>
<td>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&lt;sub&gt;lit&lt;/sub&gt;</td>
<td>Rec&lt;sub&gt;lit&lt;/sub&gt;</td>
<td>F-Score&lt;sub&gt;lit&lt;/sub&gt;</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>
</tbody>
</table>
Conclusion
Summary

Engineering Perspective

- The cohesive structure of a text provides good cues for distinguishing literal and non-literal language.
- Local context, syntax and selectional preferences provide further cues but don’t generalise well.
- Cohesion also makes it possible to detect truly creative usages (one-off’s) [Li & Sporleder, NAACL 2010].
- It is possible to go far without supervision.

(Corpus) Linguistic Perspective

- Need to learn more about how figurative expressions are used in context.
  - Multiple metaphors/idioms in a text.
  - Role of modality, negation etc.
Idiom Interpretation

- finding paraphrases [Berzak, MSc 2011]
- determining semantic argument structure

Assigning semantic roles

- The moment *the penny dropped* on Andrew Lansley’s *NHS reforms*...
- ...when *the symmetry penny dropped* on the rebus form...

Linlin Li, Alexis Palmer, Yevgeni Berzak, Philip Gorinski, Xaver Koch, Todd Shore
Timothy Baldwin and Su Nam Kim.  
Multiword expressions.  

Yevgeni Berzak.  
Sense-driven paraphrase acquisition for idiomatic expressions.  

Julia Birke and Anoop Sarkar.  
A clustering approach for the nearly unsupervised recognition of nonliteral language.  

Rudi L. Cilibrasi and Paul M. B. Vitanyi.  
The Google similarity distance.  
ISSN 1041-4347.  
doi: http://dx.doi.org/10.1109/TKDE.2007.48.

Paul Cook, Afsaneh Fazly, and Suzanne Stevenson.  
Pulling their weight: Exploiting syntactic forms for the automatic identification of idiomatic expressions in context.  

Paul Cook, Afsaneh Fazly, and Suzanne Stevenson.  
The VNC-Tokens Dataset.  
Afsaneh Fazly, Paul Cook, and Suzanne Stevenson.  
Unsupervised type and token identification of idiomatic expressions.  

Richard Fothergill and Timothy Baldwin.  
Fleshing it out: A supervised approach to MWE-token and MWE-type classification.  

Richard Fothergill and Timothy Baldwin.  
Combining resources for MWE-token classification.  

Fabienne Fritzinger, Marion Weller, and Ulrich Heid.  
A survey of idiomatic preposition-noun-verb tuples on token level.  

Chikara Hashimoto and Daisuke Kawahara.  
Construction of an idiom corpus and its application to idiom identification based on WSD incorporating idiom-specific features.  

Graeme Hirst and David St-Onge.  
Lexical chains as representations of context for the detection and correction of malapropisms.  
Graham Katz and Eugenie Giesbrecht.  
Automatic identification of non-compositional multi-word expressions using latent semantic analysis.  

Linlin Li and Caroline Sporleder.  
Contextual idiom detection without labelled data.  

Linlin Li and Caroline Sporleder.  
Using Gaussian Mixture Models to detect figurative language in context.  

Linlin Li and Caroline Sporleder.  
Linguistic cues for distinguishing literal and non-literal usage.  

Caroline Sporleder and Linlin Li.  
Unsupervised recognition of literal and non-literal use of idiomatic expressions.  

Caroline Sporleder, Linlin Li, Philip Gorinski, and Xaver Koch.  
Idioms in context: The IDIX corpus.  
In The seventh international conference on Language Resources and Evaluation (LREC), 2010.