Using Syntactic Clauses for Analysing Conflict Coverage

Wouter van Atteveldt, Tamir Sheafer, Shaul Shenhav, Yair Fogel-Dror

CLIN 2015
http://vanatteveldt.com
Outline

1. Framing conflict
2. Grammatical analysis: Clauses
3. Explorative Results
4. Conclusions
Framing Conflict

- Framing: selectively emphasising and aspect of an issue to promote
  - A problem definition / causal interpretation
  - A moral evaluation
  - A treatment recommendation
Framing as emphasis

- Co-occurrence between issue/actor and "aspect" as framing
  - Muslims and terrorism
  - Immigrants and crime/suffering
  - BP and problem/solution

- (e.g. Ruigrok/Van Atteveldt 2008, Vliegenthart/Roggebrand 2007, Schultz et al. 2013, ... )
Framing conflict situations

- Framing of conflict as 'just war'
- Focus on negative aspects of opponent:
  - Aggressor (military means)
  - Evil acts (terrorism, suppression, crimes)
  - Cause of the problem
- Focus on positive aspects of self:
  - Limited response to achieve specific end
  - Acting in self defense (solution of the problem)
Framing the 2009 Gaza war

- **Israel:**
  - Hamas rocket fire and terrorism as problem
  - Military action in order to end terrorism

- **Palestinians / Hamas:**
  - Israeli occupation/invasion as problem
  - Focus on civilian casualties

- **Research question:**
  Is international coverage of Gaza war biased?
Co-occurrence in conflict situation

- **Problem:**
  Both frames employ the same actors, issue aspects
- Israel occurs together with Hamas in 90% (88% within 30 words)
- Who is shooting whom?
Outline

1. Framing Conflict
2. Grammatical analysis: Clauses
3. Explorative results
4. Conclusions
Grammatical analysis

- Who is agent/patient of aggression?
- Co-occurrence is undirected
- Word order not reliable
  - (esp. outside English)
- → Grammatical analysis
Clauses as Semantic Role Labeling

• Who does what to whom?
  • Semantic Role Labeling / verb subframe identification
  • Semaphor: recall too low, does too much
    • better "off the shelf" tools?

• Identifying verbs and agents is easy
• Differentiating between object, predicate and other roles is hard
  • He shot a rocket at him
  • He shot him at dawn

• Clauses: subject/predicate tuples
  • ("Old" meaning of predicate)
Source Detection

- Quoted/paraphrased sources key journalistic device
  - Citation bias
- Who is framing the news?
- Extracting quotes from syntax is surprisingly easy
Method

- Rule based approach
- Syntactic Parsing
- Source detection
- Clause identification
- Extraction as tokens
Parsing

According to Mary, she was attacked by John

(a) English

Volgens Marie werd ze door Jan aangevallen

(b) Dutch

Figure 1. Parse tree of example sentences
Source Detection

\[ \text{<speech verb> \quad \text{[source]} \quad \text{[quote]}} \]

\[ \text{[source], according to [source]} \]

\[ \text{[quote], according to [source]} \]

**Figure 2.** Syntax patterns for finding English quotes and paraphrases
Source Detection

Figure 3. Syntax patterns for finding Dutch quotes and paraphrases
Clauses

According to Mary,
she was attacked by John

(a) English

Volgens Marie
werd ze door Jan aangevallen

(b) Dutch

Figure 4. Parse tree of example sentences
Extraction

- Extract information as annotations on tokens
  - "Annotated bag of words"
- Each role includes all descendants until new role found
  - Sources include clauses, not vice versa
  - No relations between clauses
Extraction

Table 1

<table>
<thead>
<tr>
<th>Word</th>
<th>Lemma</th>
<th>POS</th>
<th>ID</th>
<th>Role</th>
<th>ID</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 According</td>
<td>accord</td>
<td>VBG</td>
<td>1</td>
<td>source</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 to</td>
<td>to</td>
<td>TO</td>
<td>1</td>
<td>source</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Mary</td>
<td>Mary</td>
<td>NNP</td>
<td>1</td>
<td>source</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 ,</td>
<td>,</td>
<td>,</td>
<td>1</td>
<td>quote</td>
<td>1</td>
<td>predicate</td>
</tr>
<tr>
<td>5 she</td>
<td>she</td>
<td>PRP</td>
<td>1</td>
<td>quote</td>
<td>1</td>
<td>predicate</td>
</tr>
<tr>
<td>6 was</td>
<td>be</td>
<td>VBD</td>
<td>1</td>
<td>quote</td>
<td>1</td>
<td>predicate</td>
</tr>
<tr>
<td>7 attacked</td>
<td>attack</td>
<td>VBN</td>
<td>1</td>
<td>quote</td>
<td>1</td>
<td>predicate</td>
</tr>
<tr>
<td>8 by</td>
<td>by</td>
<td>IN</td>
<td>1</td>
<td>quote</td>
<td>1</td>
<td>subject</td>
</tr>
<tr>
<td>9 John</td>
<td>John</td>
<td>NNP</td>
<td>1</td>
<td>quote</td>
<td>1</td>
<td>subject</td>
</tr>
</tbody>
</table>
Validation

• 250 randomly selected sentences containing aggression and Israel and/or Hamas
• Manually coded on
  • Is there aggression?
  • Aggressor? (Hamas/Palestinians, Israel, both, neither)
  • Victim? (Hamas/Palestinians, Israel, both, neither)
  • Attributed source? (Hamas, Israel, other, none)
• Precision/recall for aggressor-victim pairs, sources
• Comparses to word-order baseline
  (with speech verbs as source indicator)
Validation: results

<table>
<thead>
<tr>
<th>Method</th>
<th>Clauses</th>
<th></th>
<th></th>
<th>Sources</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pr</td>
<td>Re</td>
<td>F1</td>
<td>Pr</td>
<td>Re</td>
<td>F1</td>
<td></td>
</tr>
<tr>
<td>Syntactic rules</td>
<td>.70</td>
<td>.72</td>
<td>.71</td>
<td>.95</td>
<td>.61</td>
<td>.74</td>
<td></td>
</tr>
<tr>
<td>Word-order baseline</td>
<td>.36</td>
<td>.35</td>
<td>.35</td>
<td>.5</td>
<td>.62</td>
<td>.55</td>
<td></td>
</tr>
</tbody>
</table>
Outline

1. Framing Conflict
2. Grammatical analysis: Clauses
3. Explorative results
4. Conclusions
Explorative results

- 2009 Gaza war
- US and Chinese news media
  - LexisNexis, English language articles
  - US: 6.5k articles, Chinese: 2.2k
  - Total 800k clauses
- Simple dictionary for 'Israel'/'Hamas' and 'aggression'
Explorative results: The 2009 Gaza war

Who does what? (According to who)

Table 2
*Mentions of Hamas and Israel in different roles in US and Chinese media* (column percentages)

<table>
<thead>
<tr>
<th></th>
<th>Hamas</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>U.S.</td>
<td>China</td>
<td>U.S.</td>
<td>China</td>
</tr>
<tr>
<td>All clauses</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source</td>
<td>6.7</td>
<td>10.6</td>
<td>12.0</td>
<td>11.0</td>
</tr>
<tr>
<td>Subject</td>
<td>45.2</td>
<td>38.8</td>
<td>45.3</td>
<td>50.0</td>
</tr>
<tr>
<td>Predicate</td>
<td>48.1</td>
<td>50.5</td>
<td>42.7</td>
<td>39.0</td>
</tr>
<tr>
<td>N</td>
<td>34837</td>
<td>4188</td>
<td>80121</td>
<td>12768</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Hamas</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>U.S.</td>
<td>China</td>
<td>U.S.</td>
<td>China</td>
</tr>
<tr>
<td>Clauses containing aggression</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source</td>
<td>5.2</td>
<td>7.3</td>
<td>12.3</td>
<td>12.3</td>
</tr>
<tr>
<td>Subject</td>
<td>49.1</td>
<td>33.1</td>
<td>65.4</td>
<td>70.3</td>
</tr>
<tr>
<td>Predicate</td>
<td>45.7</td>
<td>59.5</td>
<td>22.3</td>
<td>17.4</td>
</tr>
<tr>
<td>N</td>
<td>3298</td>
<td>477</td>
<td>9621</td>
<td>1645</td>
</tr>
</tbody>
</table>
What is being done?

- Select words in predicates
- Compare Israeli subject vs Hamas subject
  - chi-squared overrepresentation
- Occurrence and simple co-occurrence in predicate
Explorative results: The 2009 Gaza war

(Overrepresented predicate words, subject Israel, US Media)
Explorative results: The 2009 Gaza war

(Overrepresented predicate words, subject Hamas, US Media)
Explorative results: The 2009 Gaza war

(Overrepresented pred. words, subject Israel, Chinese Media)
(Overrepresented pred. words, subject Hamas, Chinese Media)
Explorative results: The 2009 Gaza war

Dyadic words

- Who does what to whom?
- Frequent words on dyadic edges
  - actor in subject → actor in predicate
- EU, US, Israel and Palestine
Explorative results: The 2009 Gaza war

Israel → Palestine
US → Israel
### Intro

Explorative results: The 2009 Gaza war

### EU → Israel

<table>
<thead>
<tr>
<th>Condemn</th>
<th>Plead</th>
<th>Meet</th>
<th>End</th>
<th>Call</th>
<th>Work</th>
<th>Urge</th>
<th>Would</th>
<th>Protest</th>
</tr>
</thead>
</table>

---

Using Syntactic Clauses for Analysing Conflict Coverage

[http://vanatteveldt.com](http://vanatteveldt.com)
Conclusion: Clauses

- Useful subset of semantic role labeling
- Plausible and rich results
- Many more substantive analyses possible
- (Relatively) easy to use
  - Extraction standardized and fully coded in R
  - Yields an annotated list of tokens
  - → Connect with existing tools
    - R, topic models, dictionary, network analysis, ...
Batteries included

- This presentation:
  http://vanatteveldt.com/clin2015

- Source code:
  http://github.com/vanatteveldt/rsyntax
- http://vanatteveldt.com/learningr
- https://rawgit.com/vanatteveldt/learningr/master/clauses.html
- http://github.com/amcat/amcat-r
- http://github.com/kasperwelbers/corpus-tools