ReferenceNet
a semantic-pragmatic network for capturing reference relations

Piek Vossen, Marten Postma and Filip Ilievski
VU University Amsterdam, Netherlands
{piek.vossen, m.c.postma, f.ilievski}@vu.nl
What is a ReferenceNet?

Sets of word meanings used to make reference to the same instances in a similar topical context.

Given an instance in the world in a topical context, what is the set of word that can be used to make reference to it?

WordNet and word embeddings do not foresee in this.
Semantics in context

Sense and Reference:

1. Sense (Sinn): WordNet synsets and word embeddings capture variation in local contexts -> synonymy

2. Reference (Bedeutung): we make reference to the same entity or event across different local contexts, shifting roles and perspectives within a topical contexts and across different local contexts -> variation beyond synonymy (orthogonal but still consistent with the type hierarchy)
The paper

• Why do we need a ReferenceNet?

• What is a ReferenceNet?

• How to get a ReferenceNet?

• Some problems and results

• Future work
The paper

- Why do we need a ReferenceNet?
- What is a ReferenceNet?
- How to get a ReferenceNet?
- Some problems and results
- Future work
Word sense disambiguation & reference resolution are a joint problem.
Some model of the world

Sources  Signal

Mentions

The world
Identity, reference & perspective
ambiguity and variation
The paper

- Why do we need a ReferenceNet?
- What is a ReferenceNet?
- How to get a ReferenceNet?
- Some problems and results
- Future work
Model of a ReferenceNet

1. Collection of ReferenceSets, where each ReferenceSet consists of:
   1. Words and expressions used to make reference to the same instance (event or entity) in a topical context
   2. Synsets associated with these words and expressions in this context
   3. Topical context in which the reference relation was observed

2. ReferenceSets that abstract from the instance reference to instance type level
Investigators continue to look for suspects after one person was killed and four others were injured when gunfire erupted overnight at a bar in Homewood. Several witnesses, [...] They believe the gunman was not searched by the four security guards who left the business before police arrived.

Man Gets 15 - 30 Years For Deadly Shooting At Homewood Bar.
PITTSBURGH (KDKA) A man has pleaded guilty in a 2014 shooting that left four men injured and one dead. Cornell Poindexter, 30, appeared in court Monday and pleaded guilty to one count of 3rd degree murder, four counts of aggravated assault and one count of person not to possess a firearm. According to our partners at The Pittsburgh Post-Gazette, 23-year-old Corey Clark was originally accused of being the gunman, but those charges were dropped.
<ReferenceSet topic="gun-violence">
  <synset-ref corefcount="1" wid="pwn30:eng-00355365-v" iid="i23513"> <!-- kill -->
    <surface-form "tokencount="1">kill</surface-form>
  </synset-ref>
  <synset-ref corefcount="1" wid="pwn30:eng-00095280-a" iid="i500"> <!-- dead -->
    <surface-form "tokencount="1">dead</surface-form>
  </synset-ref>
  <synset-ref corefcount="1" wid="pwn30:eng-00045888-s" iid="i233"> <!-- deadly -->
    <surface-form "tokencount="=1">deadly</surface-form>
  </synset-ref>
  <synset-ref corefcount="1" wid="pwn30:eng-00220522-n" iid="i36562"> <!-- murder, slaying, execution -->
    <surface-form "tokencount="1">murder</surface-form>
  </synset-ref>
</ReferenceSet>

<ReferenceSet topic="gun-violence">
  <synset-ref corefcount="2" wid="pwn30:eng-00260470-v" iid="i23019"> <!-- hurt, injure -->
    <surface-form "tokencount="2">injure</surface-form>
  </synset-ref>
  <ReferenceSet topic="gun-violence">
    <synset-ref corefcount="1" wid="pwn30:eng-00123783-n" iid="i36591"> <!-- gunfire, gunshot -->
      <surface-form "tokencount="1">gunfire</surface-form>
    </synset-ref>
  </ReferenceSet>
</ReferenceSet>

<ReferenceSet topic="gun-violence">
  <synset-ref corefcount="2" wid="pwn30:eng-00225150-n" iid="i36591"> <!-- shooting -->
    <surface-form "tokencount="2">shooting</surface-form>
  </synset-ref>
  <synset-ref corefcount="1" wid="pwn30:eng-00123783-n" iid="i36562"> <!-- gunfire, gunshot -->
    <surface-form "tokencount="1">gunfire</surface-form>
  </synset-ref>
</ReferenceSet>

<ReferenceSet topic="gun-violence">
  <synset-ref corefcount="1" wid="pwn30:eng-00767826-n" iid="i39445"> <!-- assault -->
    <surface-form "tokencount="1">assault</surface-form>
  </synset-ref>
</ReferenceSet>
What is a ReferenceNet?
The paper

- Where did it start?
- Why do we need a ReferenceNet?
- What is a ReferenceNet?
- How to get a ReferenceNet?
- Some problems and results
- Future work
Police Officer Says He Was Fired (1) for Not Shooting (2). .. The officer could have fired (2) a shot, but he didn’t. That officer, Stephen Mader, now 26, was dismissed (1) weeks later by the Weirton, W.Va., police department.

The Weirton Police Department terminated (1) Mr. Mader’s employment because he chose not to use deadly force to shoot (2) and kill (2) and African-American man, who was suicidal, and whom Mr. Mader reasonably believed did not pose a risk of death or serious bodily injury,

Boxes are instances
Text-to-data drawbacks

• Entity and event detection is reasonable (F-scores 65-80%) although still not very good!

• **But** co-reference performance is very low both for humans and machines (F-scores 50-60%)

• Clustering methods compare surface forms of mentions and therefore generate ReferenceSets that tend to lack variation

• Manually annotated reference data is limited (Ilievski, Postma & Vossen 2016):
  • The volume of annotated coreference data is small (few thousand documents) —> rarely find multiple documents on the same incident
  • Evaluation sets lack variation because annotators are uncertain about coreference relations across different words (one referent per word, 1-2 variant words per referent)
Event coreference data

Table 1: Existing data sets with event-reference annotations for English text. Other data sets with event annotations such as Propbank, FactBank and TimeML are not listed because they lack (co-)reference relations.

<table>
<thead>
<tr>
<th>Name</th>
<th>Literature</th>
<th>Nr. of docs</th>
<th>Nr of event mentions</th>
<th>event mentions / doc</th>
<th>Nr of coreference clusters</th>
<th>event mentions / cluster</th>
<th>cross doc</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACE2005</td>
<td>Peng et al. 2016</td>
<td>599</td>
<td>5.268</td>
<td>8.79</td>
<td>4.046</td>
<td>1.30</td>
<td>NO</td>
</tr>
<tr>
<td>KBP2015</td>
<td>Mitamura et al. 2015</td>
<td>360</td>
<td>13.113</td>
<td>36.43</td>
<td>2.204</td>
<td>5.95</td>
<td>NO</td>
</tr>
<tr>
<td>OntoNotes</td>
<td>Pradhan et al. 2007</td>
<td>1.187</td>
<td>3.148</td>
<td>2.65</td>
<td>2.983</td>
<td>1.06</td>
<td>NO</td>
</tr>
<tr>
<td>IC</td>
<td>Hovy et al. 2013</td>
<td>65</td>
<td>2.665</td>
<td>41.00</td>
<td>1.300</td>
<td>2.05</td>
<td>NO</td>
</tr>
<tr>
<td>EECB</td>
<td>Lee et al. 2012</td>
<td>482</td>
<td>2.533</td>
<td>5.26</td>
<td>774</td>
<td>3.27</td>
<td>YES</td>
</tr>
<tr>
<td>ECB+</td>
<td>Cybulska &amp; Vossen 2014a</td>
<td>982</td>
<td>6.833</td>
<td>6.96</td>
<td>1.958</td>
<td>3.49</td>
<td>YES</td>
</tr>
<tr>
<td>MEANTIME</td>
<td>Minard et al. 2016</td>
<td>120</td>
<td>2.096</td>
<td>17.47</td>
<td>1.717</td>
<td>1.22</td>
<td>YES</td>
</tr>
<tr>
<td>EER</td>
<td>Hong et al. 2016</td>
<td>79</td>
<td>636</td>
<td>8.05</td>
<td>75</td>
<td>8.48</td>
<td>YES</td>
</tr>
<tr>
<td>RED</td>
<td>O'Gorman et al. 2016</td>
<td>95</td>
<td>8.731</td>
<td>91.91</td>
<td>2.390</td>
<td>3.65</td>
<td>YES</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>3,874</td>
<td>36.292</td>
<td>9.37</td>
<td>15.057</td>
<td>2.41</td>
<td></td>
</tr>
</tbody>
</table>
But people already register events as structured data!!!!
Thanks for your interest in our statistics! Unfortunately, we must limit result sets for the general queries.

<table>
<thead>
<tr>
<th>Incident Date</th>
<th>State</th>
<th>City Or County</th>
<th>Address</th>
<th># Killed</th>
<th># Injured</th>
<th>Operations</th>
</tr>
</thead>
<tbody>
<tr>
<td>September 18, 2017</td>
<td>Kentucky</td>
<td>Louisville</td>
<td>1600 block of Prentice St</td>
<td>1</td>
<td>0</td>
<td>View Incident</td>
</tr>
<tr>
<td>September 18, 2017</td>
<td>Ohio</td>
<td>Cincinnati</td>
<td>1800 block of Clarion Avenue</td>
<td>0</td>
<td>1</td>
<td>View Incident</td>
</tr>
<tr>
<td>September 18, 2017</td>
<td>Ohio</td>
<td>Cincinnati</td>
<td>1800 block of Rutland Avenue</td>
<td>1</td>
<td>1</td>
<td>View Incident</td>
</tr>
<tr>
<td>September 18, 2017</td>
<td>Montana</td>
<td>Dillon</td>
<td>104 N Pacific St</td>
<td>0</td>
<td>0</td>
<td>View Incident</td>
</tr>
<tr>
<td>September 18, 2017</td>
<td>Indiana</td>
<td>Clay (county)</td>
<td>I-70</td>
<td>0</td>
<td>0</td>
<td>View Incident</td>
</tr>
<tr>
<td>September 18, 2017</td>
<td>Kentucky</td>
<td>Louisville</td>
<td>5 23rd St</td>
<td>0</td>
<td>1</td>
<td>View Incident</td>
</tr>
<tr>
<td>September 18, 2017</td>
<td>Louisiana</td>
<td>Morgan City</td>
<td>2400 Tiger Dr</td>
<td>0</td>
<td>0</td>
<td>View Incident</td>
</tr>
<tr>
<td>September 18, 2017</td>
<td>Arizona</td>
<td>Phoenix</td>
<td>24th Ave and McDowell Rd</td>
<td>1</td>
<td>0</td>
<td>View Incident</td>
</tr>
<tr>
<td>September 18, 2017</td>
<td>Massachusetts</td>
<td>Boston</td>
<td>77 Maple St</td>
<td>0</td>
<td>1</td>
<td>View Incident</td>
</tr>
<tr>
<td>September 18, 2017</td>
<td>Massachusetts</td>
<td>Boston</td>
<td>Columbia and Camody</td>
<td>0</td>
<td>0</td>
<td>View Incident</td>
</tr>
<tr>
<td>September 18, 2017</td>
<td>Massachusetts</td>
<td>Boston</td>
<td>128 Zeigler St</td>
<td>0</td>
<td>0</td>
<td>View Incident</td>
</tr>
<tr>
<td>September 18, 2017</td>
<td>Massachusetts</td>
<td>Pittsfield</td>
<td>East Ave and Appleton Ave</td>
<td>0</td>
<td>0</td>
<td>View Incident</td>
</tr>
<tr>
<td>September 18, 2017</td>
<td>Nebraska</td>
<td>Omaha</td>
<td>15350 Shepard St</td>
<td>0</td>
<td>1</td>
<td>View Incident</td>
</tr>
</tbody>
</table>
Police Officer Says He Was Fired (1) for Not Shooting (2). .. The officer could have fired (2) a shot, but he didn’t. That officer, Stephen Mader, now 26, was dismissed (1) weeks later by the Weirton, W.Va., police department.

The Weirton Police Department terminated (1) Mr. Mader’s employment because he chose not to use deadly force to shoot (2) and kill an African-American man, who was suicidal, and whom Mr. Mader reasonably believed did not pose a risk of death or serious bodily injury,
From-data-to-text

• **Advantages:**

  • Structured information for establishing reference is already given (validate instead of annotate)
  
  • Not everything in a text needs annotation but just what matches the structured data
  
  • Everything that matches the structured data needs to be annotated
  
  • less conservative annotations, more (true) variation
  
  • faster, higher agreement
  
  • better starting point for cross-lingual interoperability
### Event registries

<table>
<thead>
<tr>
<th>Name</th>
<th>Source</th>
<th>Topic</th>
<th>Nr docs</th>
<th>Nr incidents</th>
<th>from</th>
<th>to</th>
<th>structured data</th>
<th>locations</th>
<th>Ref texts</th>
</tr>
</thead>
<tbody>
<tr>
<td>GunViolence</td>
<td><a href="http://gunviolencearchive.org/reports/">http://gunviolencearchive.org/reports/</a></td>
<td>gun violence</td>
<td>*70k</td>
<td>33,395</td>
<td>2012</td>
<td>present</td>
<td>fatalities, locations, time, participants, roles, weapon.</td>
<td>USA</td>
<td>news</td>
</tr>
<tr>
<td>FireIncidentReports</td>
<td><a href="https://www.firerescue1.com/incident-reports/">https://www.firerescue1.com/incident-reports/</a></td>
<td>fire disasters</td>
<td>644</td>
<td>644</td>
<td>2004</td>
<td>present</td>
<td>publishing time and location,</td>
<td>USA</td>
<td>reports</td>
</tr>
<tr>
<td>ASN incident database</td>
<td><a href="https://aviation-safety.net/database/">https://aviation-safety.net/database/</a></td>
<td>aircraft safety occurrences</td>
<td>31.723</td>
<td>21.149</td>
<td>1919</td>
<td>2017</td>
<td>fatalities, locations, time, and many other domain-specific data (aircrafts, crew, airports, etc.)</td>
<td>world</td>
<td>news, reports, social media</td>
</tr>
<tr>
<td>ASN Wikibase</td>
<td><a href="https://aviation-safety.net/wikibase/">https://aviation-safety.net/wikibase/</a></td>
<td>aircraft safety occurrences (wider set of aircrafts)</td>
<td>310.240</td>
<td>206.827</td>
<td>1905</td>
<td>2017</td>
<td>fatalities, locations, time, and many other domain-specific data (aircrafts, crew, airports, etc.)</td>
<td>world</td>
<td>news, reports, social media</td>
</tr>
<tr>
<td>Railwaysarchive</td>
<td><a href="http://www.railwaysarchive.co.uk/eventlisting.php">http://www.railwaysarchive.co.uk/eventlisting.php</a></td>
<td>railway accidents</td>
<td>4,520</td>
<td>9,015</td>
<td>1803</td>
<td>2017</td>
<td>casualties, locations, time, vehicle of operators</td>
<td>UK &amp; Ireland</td>
<td>news</td>
</tr>
<tr>
<td>Legible news</td>
<td><a href="https://legiblenews.com/">https://legiblenews.com/</a></td>
<td>multiple (science, sports, business and economics, armed conflicts and attacks, law and crime, politics and elections, arts and culture, international relations, disasters and accidents, etc.)</td>
<td>*20K</td>
<td>*15K</td>
<td>2014</td>
<td>present</td>
<td>NO</td>
<td>world</td>
<td>news</td>
</tr>
<tr>
<td>Global non-violent action DB</td>
<td><a href="http://nvdatabase.swarthmore.edu/browse">http://nvdatabase.swarthmore.edu/browse</a></td>
<td>social justice/protests</td>
<td>*5.895</td>
<td>1,179</td>
<td>1955</td>
<td>present</td>
<td>location, time</td>
<td>world</td>
<td>various</td>
</tr>
<tr>
<td>Techcrunch &amp; Crunchbase</td>
<td><a href="https://www.crunchbase.com/#/home/index">https://www.crunchbase.com/</a></td>
<td>technology companies</td>
<td>43,212</td>
<td>*500K</td>
<td>2010</td>
<td>2013</td>
<td>investment, product, acquisition, company name, management team, employees</td>
<td>world</td>
<td>news</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td></td>
<td>600K</td>
<td>781K</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Event registry</td>
<td><a href="http://eventregistry.org/search?type=events">http://eventregistry.org/search?type=events</a></td>
<td>general</td>
<td>185,191.253</td>
<td>6,399.988</td>
<td>2014</td>
<td>present</td>
<td>NO</td>
<td>world</td>
<td>news</td>
</tr>
<tr>
<td>EMM Event corpus</td>
<td><a href="http://labs.emm4u.eu/events.html">http://labs.emm4u.eu/events.html</a></td>
<td>Natural Disaster, Man-mad Disaster, Civic-Political Action, Crime or Violence, Military Action and Other</td>
<td>17,536</td>
<td>7,934</td>
<td>2009</td>
<td>2015</td>
<td>NO</td>
<td>world</td>
<td>news</td>
</tr>
</tbody>
</table>
Structured event data 2

DBpedia (Knuth et al., 2015; Elbassuoni et al., 2010), Wikidata (Vrandečić and Krötzsch, 2014), and YAGO2 (Hoffart et al., 2013)

<table>
<thead>
<tr>
<th>event class</th>
<th>wikidata ID</th>
<th># events</th>
</tr>
</thead>
<tbody>
<tr>
<td>explosion</td>
<td>Q179057</td>
<td>55</td>
</tr>
<tr>
<td>crime</td>
<td>Q83267</td>
<td>1.468</td>
</tr>
<tr>
<td>natural disaster</td>
<td>Q8065</td>
<td>1.160</td>
</tr>
<tr>
<td>accident</td>
<td>Q171558</td>
<td>2.126</td>
</tr>
<tr>
<td>sport competition</td>
<td>Q13406554</td>
<td>41.245</td>
</tr>
<tr>
<td>election</td>
<td>Q40231</td>
<td>8.178</td>
</tr>
<tr>
<td>referendum</td>
<td>Q43109</td>
<td>134</td>
</tr>
<tr>
<td>voting</td>
<td>Q189760</td>
<td>17.110</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>Q189760</strong></td>
<td><strong>71.476</strong></td>
</tr>
</tbody>
</table>
query for crimes

structured data for crime incident

a crime incident

link to Wikipedia

link to References

Reference text with mentions

Police in New Jersey have identified a child whose skeletal remains were found in 2005 as a New York City girl who was the victim of years of abuse. They charged three people, including her aunt and uncle, with covering up her death.

A tip received within the last two months enabled authorities to identify the victim in what came to be known as the "Baby Bones" case. A 9-year-old Jon-Niece Jones, of Harlem, Acting Monmouth County Prosecutor Christopher Gramiccioni said Wednesday.

The investigation found that the girl died at the Harlem home of her aunt, Likisha Jones, on Aug. 15, 2002, following years of abuse by the child's mother, Elisha Jones, the prosecutor said.
The paper

• Where did it start?
• Why do we need a ReferenceNet?
• What is a ReferenceNet?
• How to get a ReferenceNet?
• Some problems and results
• Future work
Experiment

• Documents from incidents within a single topical domain: gun violence.

• Manually annotated with events and (cross-)document coreference relations.

• Two automatic methods for event detection and coreference:
  
  • knowing which mentions report on the same incident (simulating data-to-text)
  
  • comparing all event mentions across all documents (simulating text-to-data)

• Compare ReferenceSets derived from the coreferential mentions (measure the false positive and false negatives)
Gun violence data set

- 510 documents annotated by two students in two months working 8 hours per week
- IAA = 70%
- 1,197 event instances with 6,942 event mentions for 250 incidents
- Event coreference based on the incident and the implication of the change:
  - **incident**: the incident as a whole is referred to, corresponding to an entry in the structured database.
  - **firing a gun**: the event of operating a gun without implying somebody got hit.
  - **hit**: somebody got hit as a result of shooting without implying death or injury.
  - **miss**: a gun was used but the bullet missed a person.
  - **injure**: somebody got injured as a result of being hit.
  - **die**: somebody died as a result of being hit.
Vigil held for teenage shot, UNK and killed, UNK in Killingly - WFSB 3 Connecticut (Published on: 2014-08-15)

Friends and family were on hand Friday night to remember the life of a teenager from Killingly, who was killed, UNK earlier this month.

A special vigil for 16-year-old Matthew Regula was held at 8 p.m. at Davis Park in Danielson.

Police said 22-year-old Kyle Carney, shot, UNK and killed, UNK Regula inside a home on Kenneth Drive on Aug. 5.

Carney is now facing charges after he is accused of "recklessly pointing a firearm in the direction of the victim when the rifle discharged, UNK." He was charged with second-degree manslaughter, assault and second-degree reckless endangerment, all UNK.

Carney reportedly told police the shooting was accidental.

Regula reportedly attended H.H. Ellis Regional Vocational Technical School in Danielson and was supposed to be starting his senior year in just a few weeks.

Carney, who has been previously arrested for disorderly conduct and possession of drugs, is expected to appear at Danielson Superior Court on Sept. 11. He has already been arraigned in this case, but not entered a plea.

Police said they know who the gun was registered to, however, they are not releasing those details.

Copyright 2014 WFSB (Meriden Corporation). All rights reserved.

Connecticut State Police: 16-year-old boy dies, UNK after accidental shooting, UNK (Published on: 2014-08-06)

Killingly (AP) — Connecticut State Police say a 16-year-old has died, UNK after being accidentally shot, UNK in the face by a man who was pointing a gun in his direction.

Police say they got a call for a shooting, UNK around 7 p.m. Tuesday at a home in Killingly. They found a boy unresponsive, UNK in an upstairs bedroom. He was taken to Day Kimball Hospital in Putnam, where he was later pronounced dead, UNK.

Authorities say 22-year-old Kyle Carney of Killingly had been pointing a rifle in the boy's direction when it accidentally discharged, UNK.

Carney was taken into custody and charged with manslaughter, assault and reckless endangerment, all UNK. He's being held on a $500,000 bond and was due to appear in court Wednesday.

The victim's name had not been released as of Wednesday morning.
Automatic within document
Event & coreference detection

- Extracts events with Mate tool (Johansson and Nugues 2008) trained on PropBank (Kingsbury and Palmer, 2002) and NomBank (Meyers et al., 2004)

- Assign synsets using WSD to content word: UKB (Agirre and Soroa 2009) and IMS (Zhong and Ng 2010)

- ReferenceSet-1 with lemma mentions and collect the synsets with highest cumulative WSD scores, e.g. threshold 80% of max cumulative score.

- Determine WordNet similarity (Leacock and Chodorow 1998) across lemma sets by comparing their top-scoring synsets, e.g. threshold is set to 1.5 similarity (empirically set)

- ReferenceSet-2 with similar lemmas, their top-scoring synsets, the lowest-common-subsumer (if any) and the hypernyms of the top-scoring synsets
Curry Bryson, the father of the 11-year-old who police say shot and killed a 3-year-old, appeared in court today for a hearing. ... Barney says it is not the charges against him that have torn his client apart. It is the fact Bryson's 11-year-old son is accused of shooting and killing 3-year-old Elijah Walker.
Within-document matching

Curry Bryson, the father of the 11-year-old who police say shot and killed a 3-year-old, appeared in court today for a hearing. ... Barney says it is not the charges against him that have torn his client apart. It is the fact Bryson's 11-year-old son is accused of shooting and killing 3-year-old Elijah Walker.

An 11-year-old Detroit boy has been charged with manslaughter in the fatal shooting of 3-year-old Elijah Walker.
Aggregate across documents

Proportional match of synsets and /or lemmas
Comparing ReferenceSets-2

Greedy Algorithm

With-incident-knowledge
Across reference texts

Without-incident-knowledge
Across all texts

ReferenceSet-3
Output

- **Manual** = all words referring to the same incident per type

- **With-incident-knowledge** = comparing documents belonging to the same incident

- **Without-incident-knowledge** = comparing all events across all documents

False negatives

False positives
Table 5: ReferenceSets at the event type level, derived from automatic annotation for 20 incidents on gun-violence

<table>
<thead>
<tr>
<th>Type</th>
<th>Manual</th>
<th>Reference set with-incident-knowledge</th>
<th>Reference set without-incident-knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>incident</td>
<td>[accident:39, murder:15, hunting:14, reckless:14, tragedy:9, happen:8, felony:7, manslaughter:5, what:5, homicide:4, shooting:4, assault:3, case:2, endanger:2, endangerment:2, that:2, violence:2, crime:1, event:1, go:1, mistake:1, situation:1]</td>
<td>[accident:3] [act:1] [action:1] [case:3] [crime:1] [happen:14, fact:1] [hunting:1] [manslaughter:1] [murder:1] [shootout:1] [tragedy:1] [victim:3]</td>
<td>[accident:3] [call:9, make:4, name:2, act:1, action:1, holler:1] [case:3] [crime:1] [happen:14, occur:2, fact:1] [hunt:2, hunter:1, hunting:1] [manslaughter:1, murder:1] [shootout:1] [tragedy:1] [victim:3]</td>
</tr>
<tr>
<td>use gun</td>
<td>[shooting:48, fire:25, discharge:16, go:12, shot:9, pull:7, gunman:6, gun:5, gunshot:4, firing:3, shoot:2, turn:2, accidental:1, act:1, action:1, handle:1, it:1, return:1, shootout:1, shotgun:1]</td>
<td>[discharge:3] [fire:5] [gun:1] [gunman:1] [address:1, deal:1, handle:1, speak:1] [pull:4] [return:3] [turn:3] [use:2, mother:1]</td>
<td>[fire:5, discharge:3, release:3, complete:2] [gun:1] [gunman:1] [pull:4, force:1] [return:3] [turn:3, grow:2, raise:2] [mother:6, use:2, bill:1]</td>
</tr>
<tr>
<td>Type</td>
<td>Manual</td>
<td>Reference set with-incident-knowledge</td>
<td>Reference set without-incident-knowledge</td>
</tr>
<tr>
<td>----------</td>
<td>-------------------------</td>
<td>---------------------------------------</td>
<td>------------------------------------------</td>
</tr>
<tr>
<td>incident</td>
<td>[accident:39, murder:15, hunting:14, reckless:14, case:3, crime:1, happen:14, fact:1] [hunting:1]</td>
<td>[accident:3] [act:1] [action:1]</td>
<td>[accident:3] [call:9, make:4, name:2, act:1, action:1, holler:1] [case:3] [crime:1]</td>
</tr>
<tr>
<td>injury</td>
<td>[wound:36, surgery:13, treat:5, injure:3, stable:3, injurious:2, send:2, bodily:1, critical:1, hit:1, hospitalize:1, hurt:1, injury:1, put:1, stabilize:1, unresponsive:1]</td>
<td>[send:7, post:4, message:1] [message:1, send:1] [treat:2] [wound:3] [hurt:3]</td>
<td>[send:6, carry:5, post:5, letter:1, message:1, transport:1] [handling:3, treat:2, deal:1, handle:1, manage:1] [wound:3] [hurt:3, back:2, suffer:2, support:2]</td>
</tr>
</tbody>
</table>
Results from paper

<table>
<thead>
<tr>
<th></th>
<th>Mentions (874 gold)</th>
<th>Lemmas (77 gold)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>with-i</td>
<td>without-i</td>
</tr>
<tr>
<td>r</td>
<td>20.25%</td>
<td>18.19%</td>
</tr>
<tr>
<td>p</td>
<td>59.80%</td>
<td>35.57%</td>
</tr>
<tr>
<td>f</td>
<td>30.26%</td>
<td>24.07%</td>
</tr>
</tbody>
</table>

- incident-based more fine-grained and shows less drift
- incident-based higher precision (+14 mentions, +16 lemmas)
- both have low recall:
  - Propbank and Nombank coverage that drive Mate's event detection
  - Wide interpretation of event: fatally, gunman, bodies, pull the trigger, use a gun, is not captured by any of the WordNet relations
Conclusions

• We introduced the notion of ReferenceSets filling a gap of topical coherence relations between synsets, frames and domains

• We motivated the need for coreference resolution, word sense disambiguation but also perspective analysis

• We described a method to acquire initial sets using a data-to-text method

• We provide evidence that this results in more precise and fine-grained sets compare to bottom-up coreference

• We demonstrated that recall is low due to the fact that WordNet does not capture these coreference relations
Future work

• We plan to build large data sets of references texts linked to event registries

• Include nominal and entity coreference in the analysis

• Bootstrap the extraction of ReferenceSets by analysing the reference sets per topic or domain

• Experiment with replacing explicit domains by topical domain repositories
Thank you for your attention

QUESTIONS?

The Reference Machine

Needs a ReferenceNet
The Reference Machine

Ultimate goal

• A machine that can observe the world

• Make reference to the world through language

• Taking into account semantic-pragmatic factors
  • shared background
  • perspective
  • motivation
  • (un)certainty

Needs a ReferenceNet
Using a wordnet to get from visual world to language

Beyond ImageNet
naarling, beroerling, ellendeling, etterbak, etterbuil, fielt, fluim, gemenerik, hond, hondenlul, kankerlijder, kelerelijder, kelerelijder, klerelijder, kloot, kloothommel, klootspiraal, klootzak, kwal, lamgat, lammeling, lamstraal, lamzak, lazersteen, lazerstraal, loeder, lul, lulhannes, lulletje, miesgasser, mispunt, onverlaat, paardelul, paardenlul, patjakker, pleurislijder, ploert, plurk, pokkenlijer, pokkenvent, pooler, rasploert, reptiel, rotzak, schoelje, schoft, serpent, smeerkap, stinker, teringlijder, tyfuslijder, vuilak, zakkenwasser, zwijn, zak, hondelul, etter, lelijkerd, smiecht, pokkenlijder, sekreet, stinkerd, individu

huichelaar, Januskop, draaikont, farizeeër, hypocriet, januskop, jezuïet, smoolentrekker, valsaard, valserik, veinzaard, veinzer

onruststoker, aanstoker, aanzetter, agitator, herrieschopper, onrustzaaier, oproerkraaier, opruier, paniekzaaier, provocateur, raddraaier, roervink, stemmingmaker, stokebrand, stoker, woelgeest