Thirty Musts for Meaning Banking

Johan Bos & Lasha Abzianidze
1 Look at other meaning banks!
Damon showed me his stamp album.
Select public domain corpora!
3 Freeze the corpus before you start!
4 Work with raw texts in your corpus!

- words ≠ atoms of meaning
- tokenisation decisions can change
- never carry out annotation on tokenised text
- don’t think “syntax” can fix word boundaries

New York-based companies ...
5. Use stand-off annotation!

- keep annotations separate from original
- use character offsets of original (raw) text
- effective if various annotation layers are involved
Consider manual annotation!

- Many meaning banks are created with the help of a computational grammar (ERG, CCG)
- But there is an alternative: manually produce meanings (AMR)

(b/base-01
  :ARG0 (c/company)
  :Location (c/city
    :Name (n/name
      :Op1 “New”
      :Op2 “York”)))
Make a friendly annotation interface!

- Annotation can be fun – but also tedious
- Web-based, search facilities, statistics
- Sanity checker
Include an issue reporting system!

- Annotators will find bugs, raise questions, observe interesting phenomena.
- Provide means to record this valuable information.
- There are several existing issue reporting systems available (e.g., MANTIS).
Be careful with the crowd!

- Gamification: slow and hard to recruit new players
- Crowdsourcing: payment and dealing with cheaters
10 Profit from lexicalised grammars!

- A lexicalised grammar (e.g. CCG) gives an advantage in annotating syntactic structure.
- Thematic roles can be directly coupled to lexical entries (words).

Amy loves Bob.
NP (S\NP)/NP NP S\S
[Stim,Exp]
Don’t focus on just one language!

- Most meaning banks: one language (English)
- Risky!
- Meanings ought to be abstract (language neutral)
- Parallel corpora!
Try to use language-neutral tools!

- Tokenisers, taggers, parsers, boxers
- Resources tailored to a particular language ...
- ... prevent porting your method to other languages
### Apply normalisation to symbols!

<table>
<thead>
<tr>
<th>Expression</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>2pm</td>
<td>14:00</td>
</tr>
<tr>
<td>half past eight</td>
<td>08:30</td>
</tr>
<tr>
<td>five to ten</td>
<td>09:55</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Expression</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>forty five</td>
<td>45</td>
</tr>
<tr>
<td>two hundred</td>
<td>200</td>
</tr>
<tr>
<td>367</td>
<td>367</td>
</tr>
</tbody>
</table>
Limit underspecification!

• Underspecification hinders inference
• Provide most plausible interpretation

Tom got stuck in his sleeping bag.
Beware of annotation bias!

• Most plausible interpretation can also give unfair balance to stereotypes

• Examples:
  – gender
  – word sense
Use existing resources for word senses!
Apply symbol grounding!

- words → concepts
- named entities → wikification (AMR)
- toponyms → GPS coordinates
- visualisation of concepts or actions
- creating timelines
Adopt neo-Davidsonian events!

- *Davidson*  
  
  \(\text{eat}(e,x) \text{ eat}(e,x,y)\)

- *Hobbs*  
  
  \(\text{eat}(e,x,y,z)\)

- *Parsons*  
  
  \(\text{eat}(e) \text{ Agent}(e,x) \text{ Patient}(e,y)\)
Use existing role-labelling inventories!

- PropBank: small set (6) of universal roles
- VerbNet: larger set (ca. 25) of universal roles
- FrameNet: specific roles for each frame

Also needed: relations for prepositions, noun compounds, possessives
Treat role nouns differently!

- agent and recipient nouns are relational (worker, employee, victim, dentist)
- treating them as one-place predicates gives rise to contradictions
- AMR and PMB pursue this approach

(Mia’s husband is Vincent’s boss)
Beware of geopolitical entities!

- Names for GPEs can refer to locations, governments, sport squads, people
- Systematic polysemy: cities, states, countries, continents, etc.
- See ACE annotation guidelines
sentence meaning is about truth conditions

- *negation* plays a crucial role here

- proper treatment requires scope: $\neg(...)$

(concord, affixes, raising)
## Pay attention to compound words!

<table>
<thead>
<tr>
<th>Expression</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>hard hat</td>
<td>hard_hat.n.02(x)</td>
</tr>
<tr>
<td>peanut butter</td>
<td>peanut_buter.01(x)</td>
</tr>
<tr>
<td>old woman</td>
<td>old_woman.n.01(x)</td>
</tr>
<tr>
<td>secretary of state</td>
<td>secretary_of_state.n.01(x)</td>
</tr>
<tr>
<td>North and South Dakota</td>
<td>state.n.01(x) &amp; Name(x,North_Dakota) &amp; ...</td>
</tr>
<tr>
<td>Jingle Bells</td>
<td>song.n.01(x) &amp; Name(x,Jingle_Bells)</td>
</tr>
</tbody>
</table>

**Lucy in the Sky with Diamonds**

**Several Species of Small Furry Animals Gathered Together in a Cave and Grooving With A Pict**
23 Use inference tests in design!

- Amy is a smart woman → Amy is a woman
  - Amy is ten years old → Amy is old
  - Amy is ten years old → Amy is old
23 Use inference tests in design!

- Amy is a smart woman \(\text{yes}\rightarrow\) Amy is a woman
  Amy is ten years old \(\text{no}\rightarrow\) Amy is old

- Bob is unhappy \(\text{yes}\rightarrow\) Bob is not happy
  Bob is not happy \(\text{no}\rightarrow\) Bob is unhappy
23 Use inference tests in design!

- Amy is a smart woman \(\text{yes}\rightarrow\) Amy is a woman
  Amy is ten years old \(\text{no}\rightarrow\) Amy is old
- Bob is unhappy \(\text{yes}\rightarrow\) Bob is not happy
  Bob is not happy \(\text{no}\rightarrow\) Bob is unhappy
- Florence is the cultural capital of Italy \(\text{no}\rightarrow\)
  Florence is the capital of Italy
Divide and conquer!

- computing meaning is very hard
- many unsolved representation problems
  - aspect
  - information structure (focus)
  - presuppositions
  - generics
  - plurals
25 Put complex presuppositions on hold!

• some presuppositions triggers can be analysed as being anaphoric (definites, names)
• but more complex cases require copying or other sophisticated operations

(the crowd applauded again)
Respect elliptical expressions!

- Invisible, but omnipresent...
- Many different types of ellipsis

My hair is longer than Mary’s ø.
Tom eats fish, but his brother doesn’t ø.
ø Ho fame.
Think about generics!

- generics, habituals, kinds
- hard to model in first-order representations

A lion is strong.
Who invented the typewriter?
I play the piano.
Don’t try to be clever!

The fork is on the table.
Mary is an engineer.
John is the teacher who graded my essay.
30 Measure meaning discrepancies!

• Need to be able to compare meaning representations
• Current approaches based on Allen et al. 2008 (Smatch for AMR, Counter for DRS)
• We need to make “matching” more semantical
The End