Metaphor Annotation. Exercise I Summary

Katja Markert

Heidelberg SS2019
Overview

1. Annotation Problems

2. Student suggestions for best-worst scaling

3. Do Dinh’s approach to Best-worst scaling for metaphor annotation
Students annotated *I have a dream* speech and compared to another annotator. Method used was the MIP method.

Innovative metaphors easier to agree on than conventionalized ones

1. ? *x finds himself*
2. ? *go down in history*
3. ? *great American*
4. ? *dream*
5. ? *demonstration* (established terms originating from metaphorical usage)
6. *manacles of segregation*
7. *lonely island of poverty*
Metaphors are harder to recognise and agree on if wider context is needed to understand them.

- *I have a dream*
- *chains, torture* → could be meant literally or metaphorically
Is the word the right unit of annotation? Start and end of metaphors. Extended metaphors.

1. *great beacon light*
2. *lonely island of poverty*
3. *seared in the flames of withering injustice*
4. *daybreak . . . night*
Problems: Other

- Lexicon dependence
- Difference/Delineation to other figurative types: for example pars-pro-toto *The negro*
- Native speaker vs. non-native speaker
Overview

1. Annotation Problems

2. Student suggestions for best-worst scaling

3. Do Dinh’s approach to Best-worst scaling for metaphor annotation
Ines Reinig’s suggestion

1. Annotators read a definition of metaphor (with examples?)

2. Four samples are presented in random order to a candidate with same POS. Highlight the word to be annotated. Example:

   - I am happy to join with you today in what will go down in history as the greatest **demonstration** for freedom in the history of our nation.
   - Five score years ago, a great American, in whose symbolic **shadow** we stand today, signed the Emancipation Proclamation.
   - This momentous decree came as a great beacon **light** of hope millions of Negro slaves.
   - One hundred **years** later, the Negro lives on a lonely island of poverty.

3. Annotators are asked to choose the most and least metaphorical unit.

   Leave “mathematical design” $2 \times n$ samples etc the same. Leave agreement computation the same.
Discussion points: Task definition

- Rely on an intuitive understanding of “metaphor” or figure of speech?
- Short definition such as *A metaphor is a word or group of words that is used to describe one thing in terms of another*
- Usage of other criteria: *How much does the contextual meaning differ from the basic/oldest meaning?*
- Usage of other criteria: *How abstract is the meaning in context?*
- Usage of other criteria: *Rank the terms from first to last by how much their usage in text is from base meaning? → needs a lexicon*
Discussion points: Context and Annotation Unit

- How much context to give? Ambiguous metaphors?
- Should the annotation unit be a whole sentence or an individual word?
- Should the annotation units to be compared be of the same POS? If so, why?
- Should the annotation units to be compared be randomly selected?
Other discussion points

- Randomness of 4-tuples?
- Artificial data or real data?
Overview

1. Annotation Problems

2. Student suggestions for best-worst scaling

3. Do Dinh’s approach to Best-worst scaling for metaphor annotation
Overall idea

- Take VU Amsterdam corpus that has metaphors already marked!
- Only for the **already marked content-word metaphors**, conduct metaphor novelty annotation, i.e. identify the most novel and the most conventional metaphors
- Experiment with four annotation methods: binary, scale annotation on four-point scale, scale annotation (no metaphor), best-worst scaling
## Results

<table>
<thead>
<tr>
<th>Method</th>
<th>IAA</th>
<th>F₁</th>
<th>avg assignment completion time</th>
</tr>
</thead>
<tbody>
<tr>
<td>binary</td>
<td>0.38</td>
<td>0.75</td>
<td>1:39 min</td>
</tr>
<tr>
<td>scale</td>
<td>0.32</td>
<td>0.75</td>
<td>1:04 min</td>
</tr>
<tr>
<td>scale w/o met.</td>
<td>0.16</td>
<td>0.67</td>
<td>2:10 min</td>
</tr>
<tr>
<td>BWS</td>
<td>—</td>
<td>0.84</td>
<td>1:58 min</td>
</tr>
</tbody>
</table>
For next time

- Students with 6 LP: read Tsvetkov et al and post two technical questions (always to presenter and me)
- Students with 8 LP: read Bulat et al and post two technical questions (always to presenter and me)