

Metaphor Annotation. Exercise I Summary

Katja Markert

Heidelberg SS2019

- 1 Annotation Problems
- 2 Student suggestions for best-worst scaling
- 3 Do Dinh's approach to Best-worst scaling for metaphor annotation

Problems found by students: Innovative vs conventionalized

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*

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

- 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?

- 1 Annotation Problems
- 2 Student suggestions for best-worst scaling
- 3 Do Dinh's approach to Best-worst scaling for metaphor annotation

- 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

See <https://www.aclweb.org/anthology/D18-1171> Appendix for annotation guidelines.

	IAA	F ₁	avg assignment completion time
binary	0.38	0.75	1:39 min
scale	0.32	0.75	1:04 min
scale w/o met.	0.16	0.67	2:10 min
BWS	—	0.84	1:58 min

- 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)

- 1 Do Dinh, E.L., Wieland, H. and Gurevych, I., 2018. *Weeding out conventionalized metaphors: A corpus of novel metaphor annotations*. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (pp. 1412-1424).
- 2 Kiritchenko, S. and Mohammad, S.M., 2016. *Capturing Reliable Fine-Grained Sentiment Associations by Crowdsourcing and BestWorst Scaling*. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 811-817).