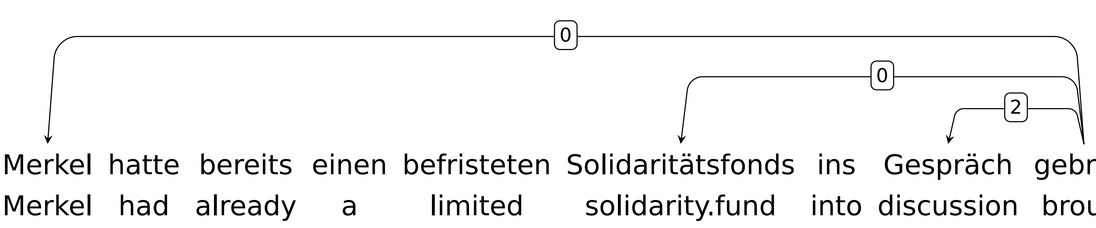
VerbCompoCor: A German Corpus with Compositionality Judgments for Verb-Dependent Pairs

Fabienne Cap[1], Rafael Ehren[2], Maximlian Köper[3], Timm Lichte[2], Sabine Schulte im Walde[3] & Heike Zinsmeister[4]

[1]Uppsala University, [2]University of Düsseldorf, [3]University of Stuttgart, [4]University of Hamburg

Overview

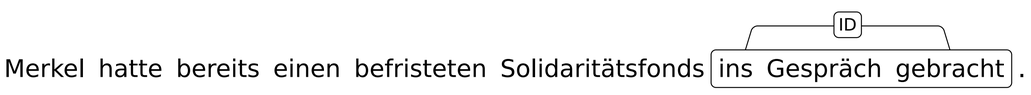
- 7,500 sentences from the German part of the PARSEME shared task corpus [6]
- Manually annotated judgments on the degree of compositionality of verb-dependent pairs
- 6-point scale from 0 (completely compositional) to 5 (completely non-compositional)
- IAA across three annotators: Fleiss' κ=0.354
- Annotation tool: WebAnno [4]
- Financial support by DGfS-CL, GSCL, SFB 732, SFB 991



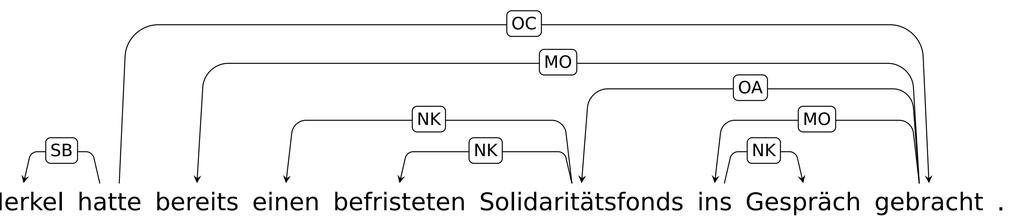
'Merkel already brought up a limited solidarity fund for discussion.'

Preprocessing

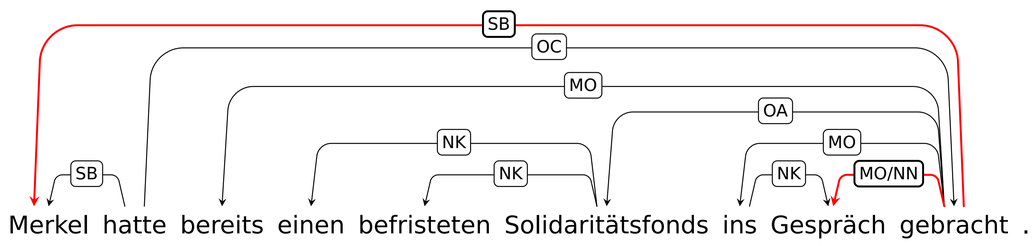
PARSEME shared task corpus [6] based on the shared task data from Bojar et al. [3]



1. Dependency parsing with the MATE parser [2]



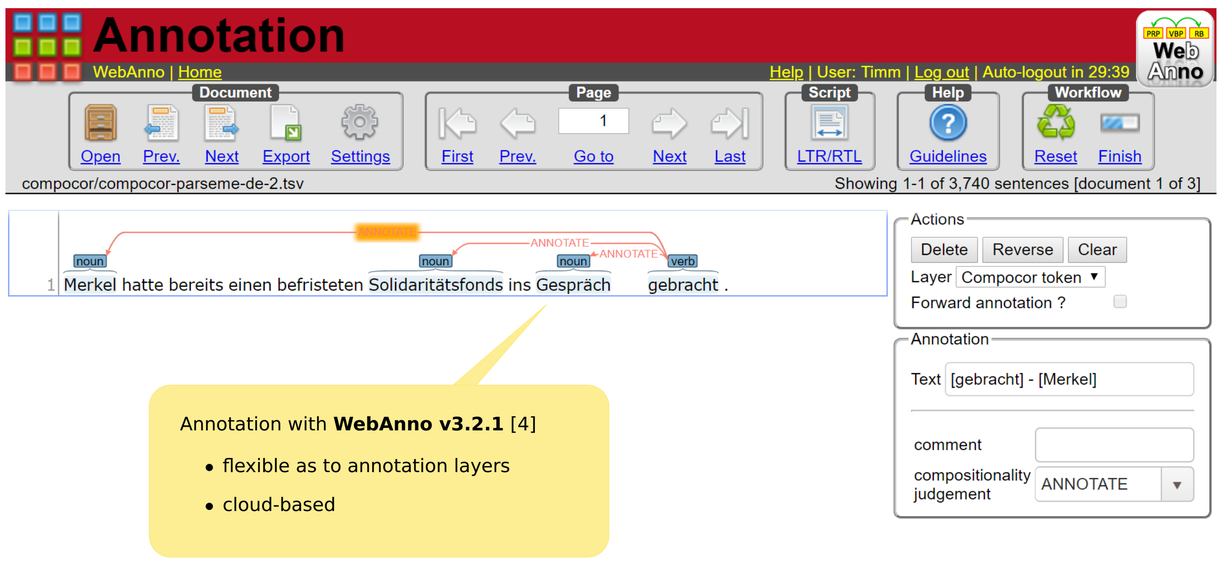
2. Extraction of subject-verb dependencies using the SubCat-Extractor [7]



3. Selection for annotation: verb-dependent pairs with nouns as dependents

Annotation procedure

- Annotation guidelines (condensed, original in German):
  1. Is the verb-dependent pair correct?
     YES → proceed to compositionality scoring in step 2
     NO → annotate "ERROR" and comment
  2. What compositionality score do you assign to the pair?
     compositional 0 1 2 3 4 5 non-compositional
  3. Supporting tests (adopted from the PARSEME shared task annotation guidelines):
     [CRAN] The verb-dependent pair contains a cranberry word.
     [LEX] Replacing one component with a semantically similar expression yields unexpected grammatical/semantic shifts.
     [MORPH]/[MORPH-SYNT]/[SYNT] Morphological/morphosyntactic/syntactic variation in one component yields unexpected grammatical/semantic shifts.
     [VFG] The verb contributes almost nothing to the overall expression.
     YES → choose a score between 1 and 5
     NO → choose 0 as a score
- Annotators: Rafael Ehren & Lea Vanessa Möllmann (Düsseldorf), Alexander Frey, Daniel Helbig, Glorianna Jagfeld & Florian Lux (Stuttgart)



Annotation with **WebAnno v3.2.1** [4]

- flexible as to annotation layers
- cloud-based

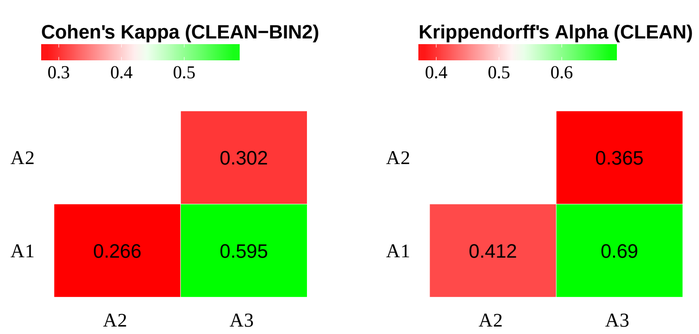
Annotation evaluation

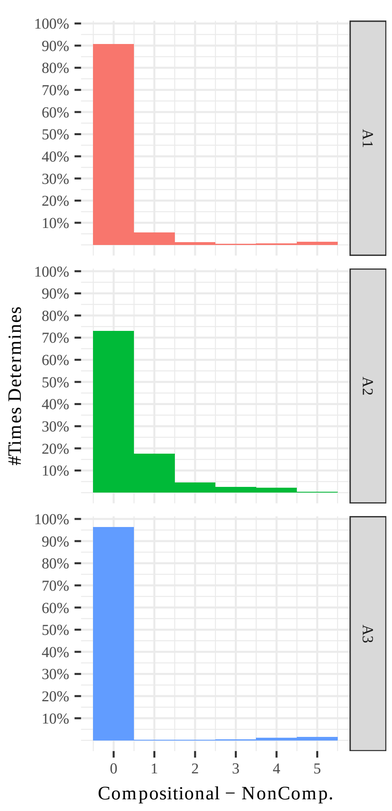
- Evaluation based on 3 full annotations
- 25,317 judgments per annotator
- 2,547 verb-dependent pairs were annotated with ERROR by at least one annotator
- Inter-annotator agreement (IAA):
  - Mean Spearman's ρ: 0.416
  - Fleiss' κ: 0.354 (binarized with $[0,1][2,3,4,5]$)

	Fleiss' κ	Krippenndorff's α
interval	0.127	0.478
binarized ($[0][1,2,3,4,5]$)	0.232	—
binarized ($[0,1][2,3,4,5]$)	0.354	—

  - big differences in pairwise IAA
- Distribution of scores:
  - highly skewed toward 0 (compositionality)

  ⇒ very strong baseline for computational models





Comparison to similar work

- PARSEME shared task corpus [6]
  - annotated categories of verbal multi-word expressions, rather than verb-dependent pairs: light-verb constructions, idioms, inherently reflexive verbs, verb-particle combinations
  - binary decision (rather than scale)
- Literal vs. non-literal language usage (binary decisions)
  - Birke & Sarkar [1]: Fleiss' κ=0.77 over two annotators (EN)
  - Tsvetkov et al. [8]: Fleiss' κ=0.75 / 0.78 over 5 / 6 annotators (EN / RU)
  - Köper & Schulte im Walde [5]: Fleiss' κ=0.70 over three annotators (DE)

Prospects

- Further annotations (ongoing)
- Cleansing (planned): e.g. majority vote
- Shared task at KONVENS 2018 (planned): prediction of degree of compositionality

References

[1] Birke, J. & A. Sarkar. 2006. A Clustering Approach for the Nearly Unsupervised Recognition of Nonliteral Language. In *Proceedings of the 11th conference of the european chapter of the acl*, 329–336. Trento, Italy. [2] Bohnet, B. 2010. Very high accuracy and fast dependency parsing is not a contradiction. In *Proceedings of the 23rd international Conference on Computational Linguistics (Coling 2010)*, 89–97. Bejing. [3] Bojar, O. et al. 2016. Findings of the 2016 conference on machine translation (wmt16). In *Proceedings of the first conference on machine translation (wmt16), volume 2: shared task papers*, 131–198. [4] Eckart de Castilho, R. et al. 2016. A web-based tool for the integrated annotation of semantic and syntactic structures. In *Proceedings of the workshop on language technology resources and tools for digital humanities (LT4DH)*, 76–84. Osaka, Japan. [5] Köper, M. & S. Schulte im Walde. 2016. Distinguishing Literal and Non-Literal Usage of German Particle Verbs. In *Proceedings of the conference of the north american chapter of the association for computational linguistics: human language technologies*, 353–362. San Diego, California, USA. [6] Savary, A. et al. 2017. The PARSEME shared task on automatic identification of verbal multiword expressions. In *Proceedings of the 13th Workshop on Multiword Expressions (MWE 2017)*, 31–47. Valencia, Spain. [7] Scheible, S. et al. 2013. A compact but linguistically detailed database for German verb subcategorisation relying on dependency parses from web corpora: Tool, guidelines and resource. In *Proceedings of the 8th web as corpus workshop (wac13)*, 63–72. [8] Tsvetkov, Y. et al. 2014. Metaphor Detection with Cross-Lingual Model Transfer. In *Proceedings of the 52nd annual meeting of the association for computational linguistics*, 248–258. Baltimore, Maryland.