

# A corpus annotated for metaphor in German

Markus Egg and Valia Kordoni

Humboldt-Universität zu Berlin, Germany



LREC 2022, Marseille

## Introduction

- ▶ we are annotating a German corpus for **metaphor**
- ▶ the corpus includes material from **different registers** that displays register variation along a number of important dimensions
- ▶ first results show clear **interdependencies** between metaphor and register

## Theoretical background and previous work

- ▶ metaphors refer to an entity  $E_1$  that is **similar** to the referent of the literal interpretation  $E_2$  of the expression
  - ▷ metaphor is based on a **mapping** from the structure of a 'source domain' (including  $E_2$ ) onto a 'target domain' (with  $E_1$ ) (Lakoff 1993)
  - ▷ metaphorical expressions can be lexemes, phrases, or lexeme parts
- ▶ register is the influence of **situational and functional context** on intra-individual variation
  - ▷ it has been decomposed into features (Halliday and Hasan 1985)
  - ▷ it has been derived from linguistic phenomena (Biber 1988)
- ▶ some previous work
  - ▷ metaphor in specific registers, e.g., academic discourse (Beger 2015) or fiction (Dorst 2015)
  - ▷ correlating functions of metaphors with register features (Goatly 2011)
  - ▷ VU Amsterdam Metaphor Corpus with four registers (Steen et al. 2010)
  - ▷ Dipper et al. (2021) prepare a corpus of metaphors in German sermons

## A German corpus annotated for register

- ▶ the corpus comprises material from **different registers**
  - ▷ commentaries (incl. the Potsdam Commentary Corpus; Stede 2004)
  - ▷ speeches from the Parlamentsreden-Korpus (Blaette 2017)
  - ▷ sermons
  - ▷ light fiction
  - ▷ debates from competitions of 'Jugend debattiert'
- ▶ properties of the corpus
  - ▷ the subcorpora will comprise 100,000 words with 10-15% metaphorically relevant words (parallel to Steen et al. 2010 and Dorst 2015)
  - ▷ the subcorpora are **balanced** for important register dimensions

subcorpus	hierarchical/ equal	distant/ close	oral/ literal	dialogue/ monologue	reference	persuasion
speeches	E	D	L	M	+	o
sermons	H	C	L	M	o	+
commentaries	H	D	L	M	+	+
fiction	E	C	L	M	-	-
debates	E	D	O	D	o	+

Table 1: Register properties of the subcorpora

- ▶ the corpus is annotated with the INCEpTION tool (Klie et al. 2018)
- ▶ it is parsed with the Stanza package (Qi et al. 2020)

## The annotation guidelines

- ▶ starting point: 'Metaphor Identification Procedure-VU' (Steen et al. 2010; Herrmann, Woll, and Dorst 2019)
- ▶ we rely on lexical resources: the *Duden* and the *Digitales Wörterbuch der deutschen Sprache* (Digital dictionary of the German language)
  - ▷ **non-conventionalised** metaphors may not be listed in the resources
    - (1) der angepeilte ausgeglichene Bundeshaushalt für 2015  
'the envisaged balanced federal budget for 2015'
- ▶ **signals** of metaphoricity are annotated, like *wie* 'like' or *praktisch* 'in effect'
- ▶ we identify metaphorical expressions together with their **contexts** ('backgrounds'), which trigger the metaphorical interpretation
  - ▷ backgrounds and metaphorical expressions are related syntactically (e.g., head-complement or arguments of copulative verbs)
    - (2) in meinen Gedanken  
'in my thoughts'
  - ▷ no 1-1 relation between backgrounds and metaphors

## The annotation guidelines (ctd)

- ▶ **extended metaphors** reuse a mapping from source to target domain
  - (3) ... dass sich die gute Familienpolitik der letzten Jahre auszahlt, und dass sich das Geld, das wir investieren, lohnt  
'that the good politics for families of the last years is paying off, and that the money which we invest is worthwhile'
- ▶ new are **potential metaphors**, the deliberate combination of tokens of an expression with basic and metaphorical senses in the same text
  - (4) in dem Dunkel, in dem Wurzelbereich bei dem Weizen  
'in the dark zone, in the rhizosphere of the wheat'
  - (5) die dunkle Erde elterlicher Übermüdung  
'the dark soil of parental fatigue'
- ▶ some issues are **specific to German**, e.g. combinations of demonstratives and (potentially metaphorical) prepositions like *darin* 'that-in'
  - (6) dass darin der Schlüssel liegt  
'that the key is to be found (lit. 'lies') in there'

## First results

- ▶ non-conventionalised and extended metaphors show up predominantly in highly **persuasive** registers (sermons and commentaries)
  - ▷ there are fewer ones in debates, due to time pressure in oral discourse
- ▶ the expectation that **oral discourse** is low on metaphor (Steen et al. 2010) cannot be confirmed
  - ▷ previous low counts stem from conversations
- ▶ non-conventionalised and extended metaphors occur more frequently in registers with a **hierarchical difference** between authors and recipients
- ▶ **fiction** has very few metaphors, like in Steen et al. (2010)
- ▶ **sermons** have the highest metaphorical register marking
  - ▷ there are very many non-conventional metaphors
  - ▷ extended metaphor is a clear register marker for sermons
  - ▷ potential metaphors occur only in sermons (.41%)

subcorpus	conventionalised metaphor	non-conventionalised metaphor	extended metaphor
speeches	15.13%	.14%	.01%
sermons	10.14%	.24%	.29%
commentaries	11.34%	.24%	.10%
fiction	4.06%	.16%	.002%
debates	10.38%	.13%	.02%

Table 2: Metaphor counts for the subcorpora

## Acknowledgements

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1412, 416591334.

## References

- Beger, Anke (2015). "Metaphors in psychology genres. Counseling vs. academic lectures". In: *Metaphor in specialist discourse*. Ed. by Berenike Herrmann and Tony Berber Sardinha. Amsterdam: Benjamins, pp. 53–75.
- Biber, Douglas (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Blaette, Andreas (2017). *GermaParl. Corpus of Plenary Protocols of the German Bundestag*. TEI files. available at: <https://github.com/PolMine/GermaParlTEI>.
- Dipper, Stefanie et al. (2021). "Recognizing deliberate metaphors". Poster presented at DGfS-CL 21.
- Dorst, Aletta (2015). "More or different metaphors in fiction? A quantitative cross-register comparison". In: *Language and Literature* 24, pp. 3–22.
- Goatly, Andrew (2011). *The language of metaphors*. 2nd ed. Abingdon: Routledge.
- Halliday, Michael and Ruqaiya Hasan (1985). *Language, context and text*. Victoria: Deakin University Press.
- Herrmann, Berenike, Karola Woll, and Aletta Dorst (2019). "Linguistic metaphor identification in German". In: *Metaphor identification in multiple languages*. Ed. by Susan Nacey et al. Benjamins, pp. 113–135.
- Klie, Jan-Christoph et al. (2018). "The inception platform: Machine-assisted and knowledge-oriented interactive annotation". In: *Proceedings of COLING 2018: system demonstrations*, pp. 5–9.
- Lakoff, George (1993). "The contemporary theory of metaphor". In: *Metaphor and Thought*. Ed. by Andrew Ortony. 2nd. New York: Cambridge University Press, pp. 202–251.
- Qi, Peng et al. (2020). "Stanza: A Python Natural Language Processing Toolkit for Many Human Languages". In: *Proceedings of ACL 2020: System Demonstrations*.
- Stede, Manfred (2004). "The Potsdam Commentary Corpus". In: *ACL 2004 Workshop on Discourse Annotation*.
- Steen, Gerard et al. (2010). *A method for linguistic metaphor identification: from MIP to MIPVU*. Amsterdam: Benjamins.