A Laypeople Study on Terminology Identification across Domains and Task Definitions
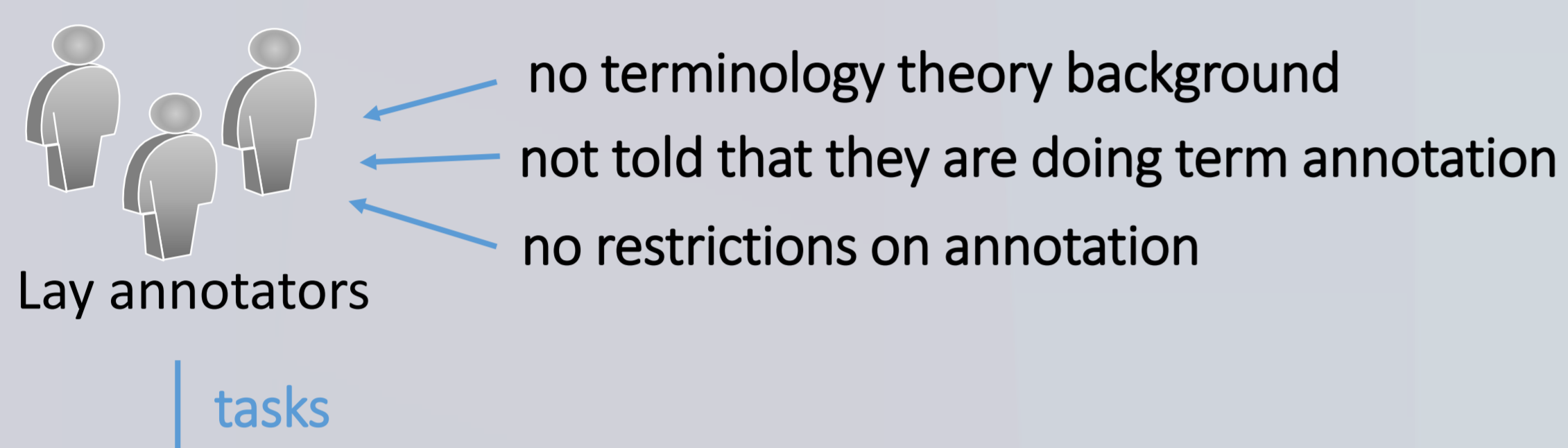
Anna Hätty
Robert Bosch GmbH

Sabine Schulte im Walde
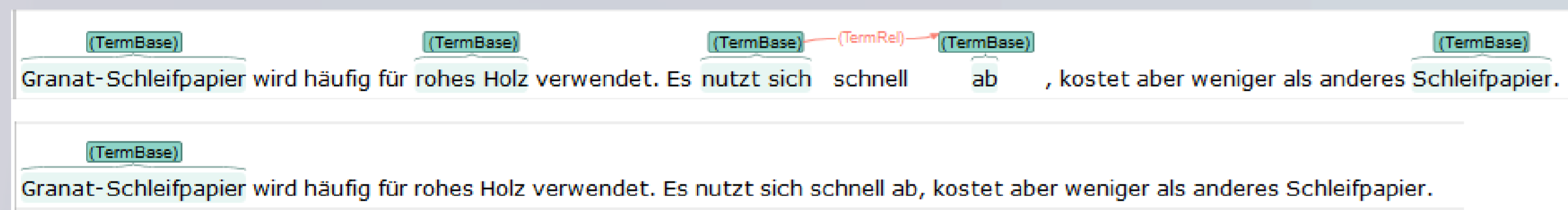University of Stuttgart

Motivation

- Neither **term annotation** nor **automatic term extraction** follows consistent rules
- → Estopà (2001): terminologists, domain experts, translators and documentalists are given the same task, and they select different kinds of terms

- One step back: How do laymen agree in a term annotation task?

Annotation Task

- no terminology theory background
- not told that they are doing term annotation
- no restrictions on annotation

Lay annotators

↓ tasks

- Highlight domain-specific phrases (**DS**)
- Create an index (**IND**)
- Define unknown words for creating a translation lexicon (**TR**)
- Create a glossary (**GL**)

Example for annotation in WebAnno:

Granat-Schleifpapier wird häufig für rohes Holz verwendet. Es nutzt sich schnell ab, kostet aber weniger als anderes Schleifpapier.

Granat-Schleifpapier wird häufig für rohes Holz verwendet. Es nutzt sich schnell ab, kostet aber weniger als anderes Schleifpapier.

Domains

- DIY
- Chess
- Cooking
- Hunting

Annotation Procedure

- Overall: 4 tasks * 7 annotators = 28 ("concordance")
- Per task: max of 7 annotations ("agreement")

Agreement across Tasks & Domains

- As expected: in absolute numbers, there are **more annotations** for broader term annotation tasks than for narrower ones

- Not as expected: **agreement** for broader term annotation tasks is higher than for narrower ones

- Across all tasks: agreement is similar for the same terms → laymen have an intuitive, common understanding about a term's domain specificity
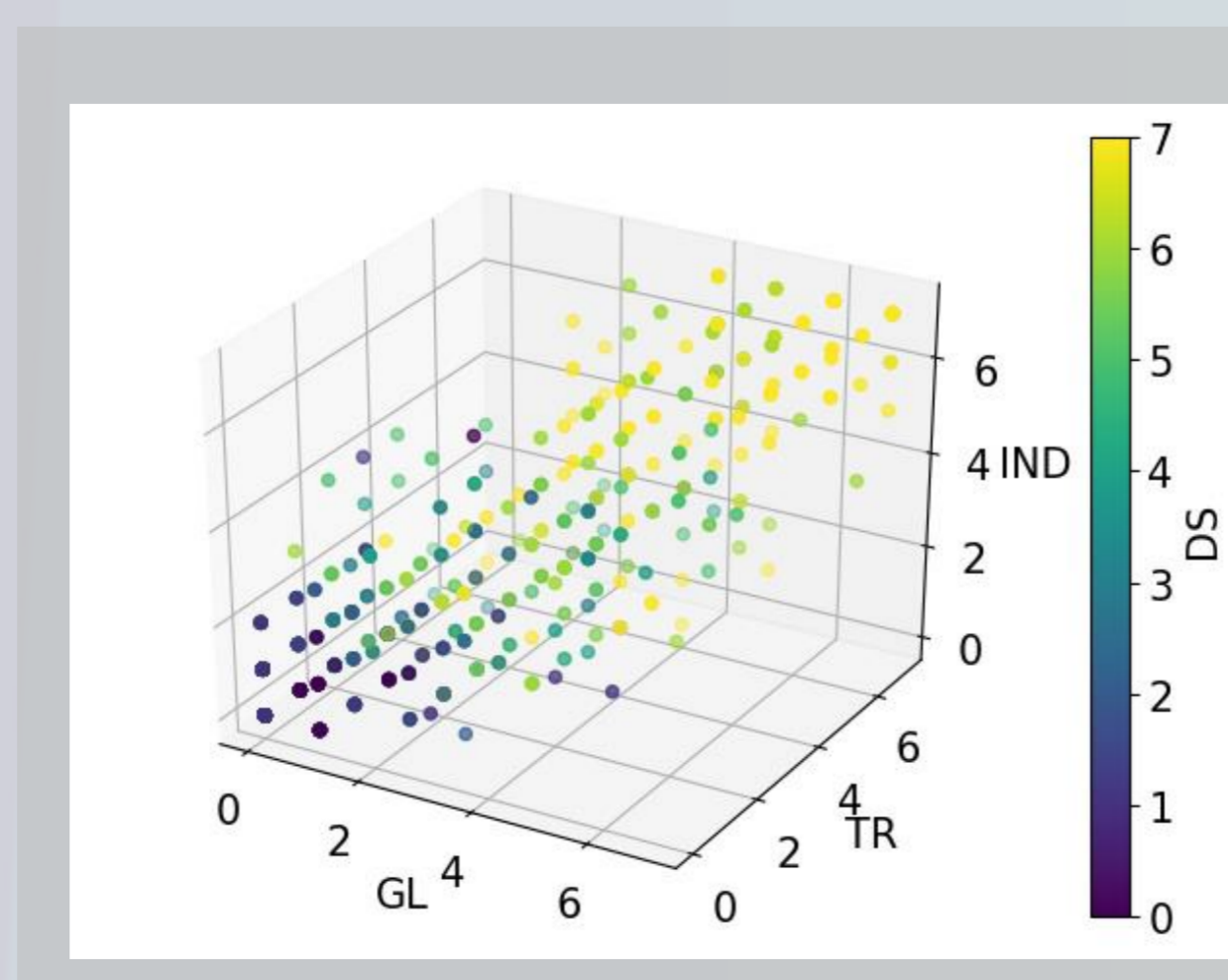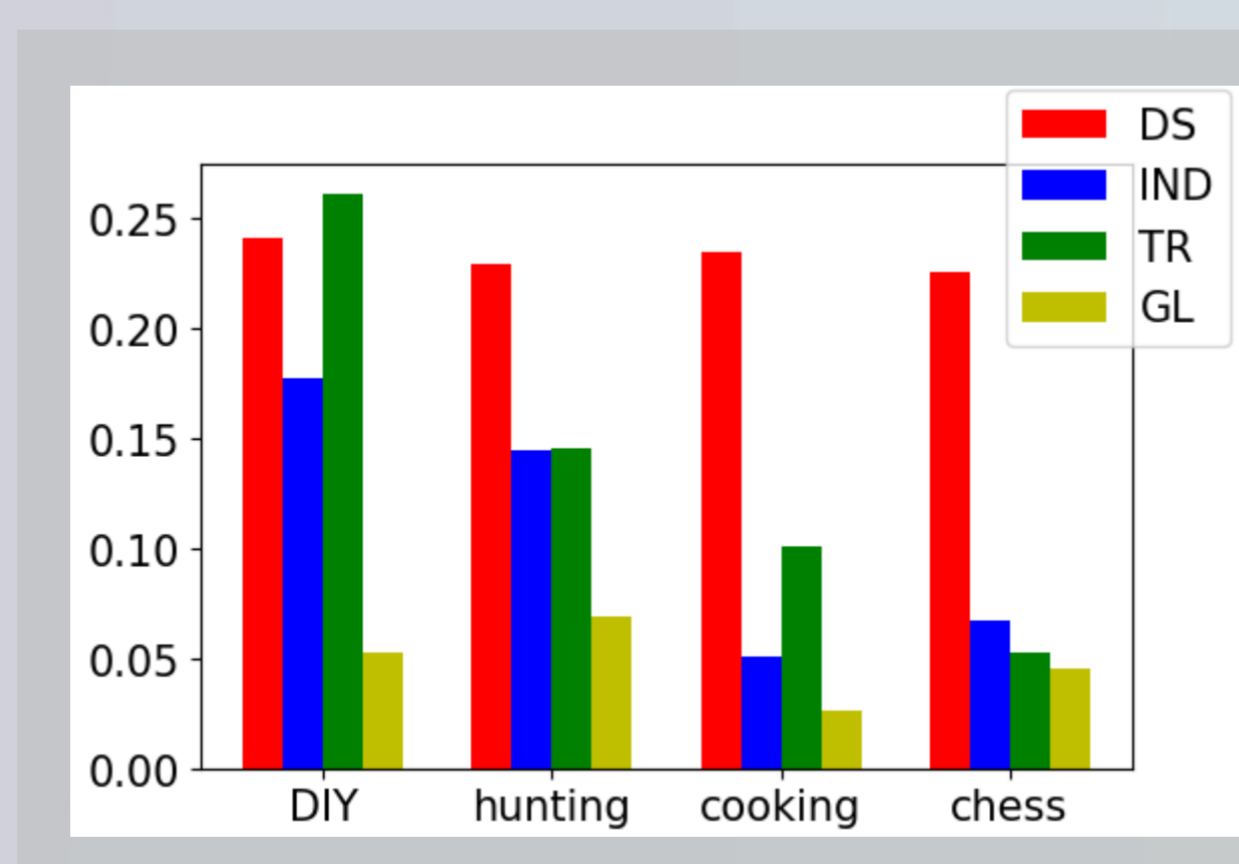

Word Classes

What constitutes a term? A noun? A verb?

- Terms are typically regarded as noun phrases
- Laymen accept **verbs** as terms, too, but with a low agreement
- Verbs are often used in MWEs: *Eigelb schaumig schlagen* ('beat the egg yolk until frothy').
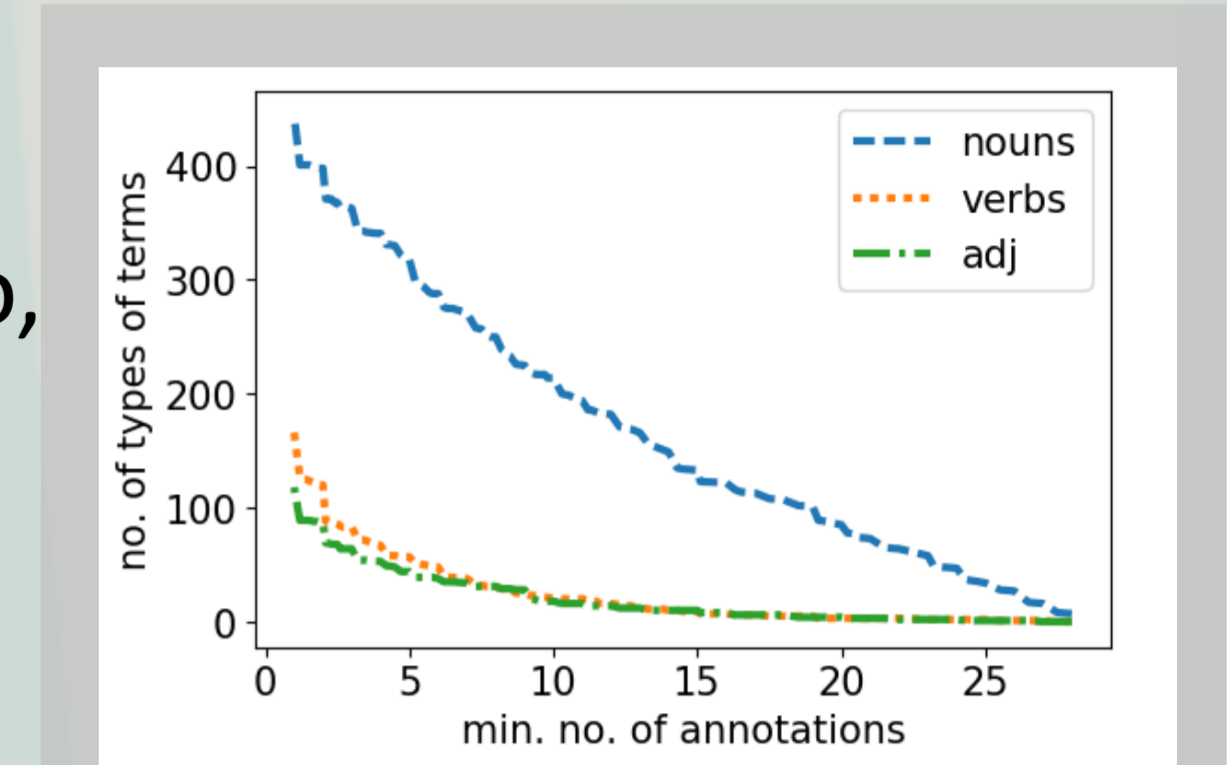
Complex Terms and Subterms

- High concordance for **compounds**, low concordance for **MWTs**
- But: for MWTs, concordance for **subterms** is higher with decreasing concordance for MWT → if unsure for the MWT, annotators rely on components
- Compounds: other way round

Ambiguity

"Fuchsschwanz"
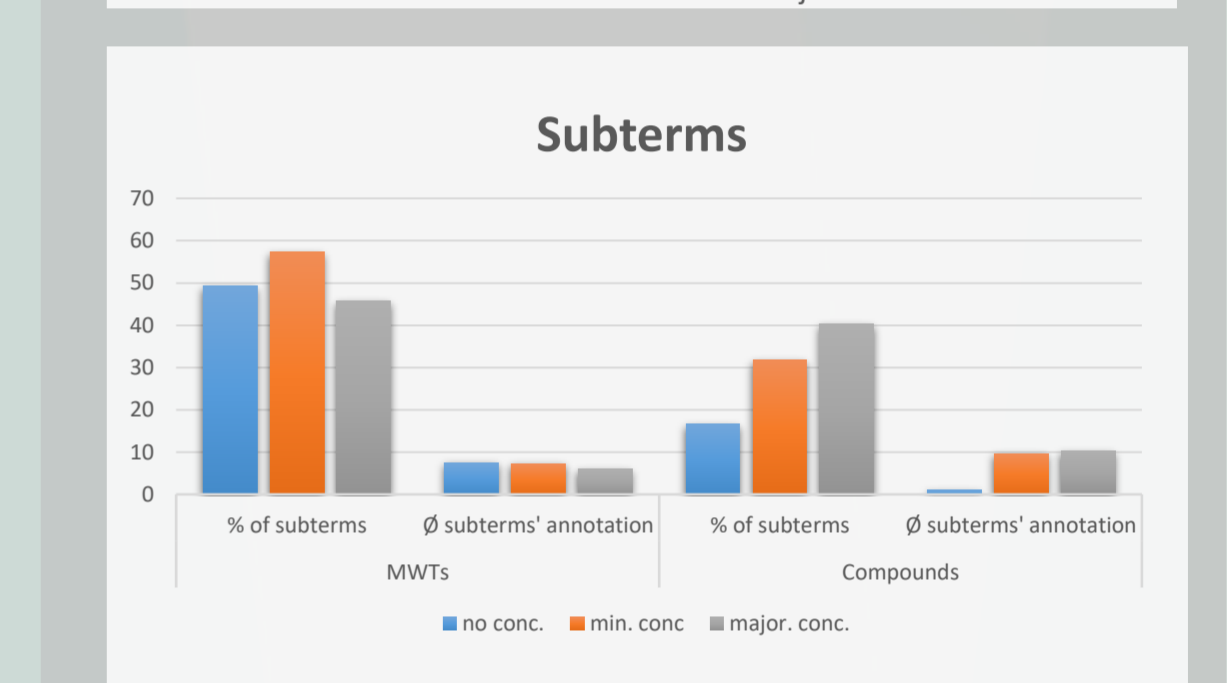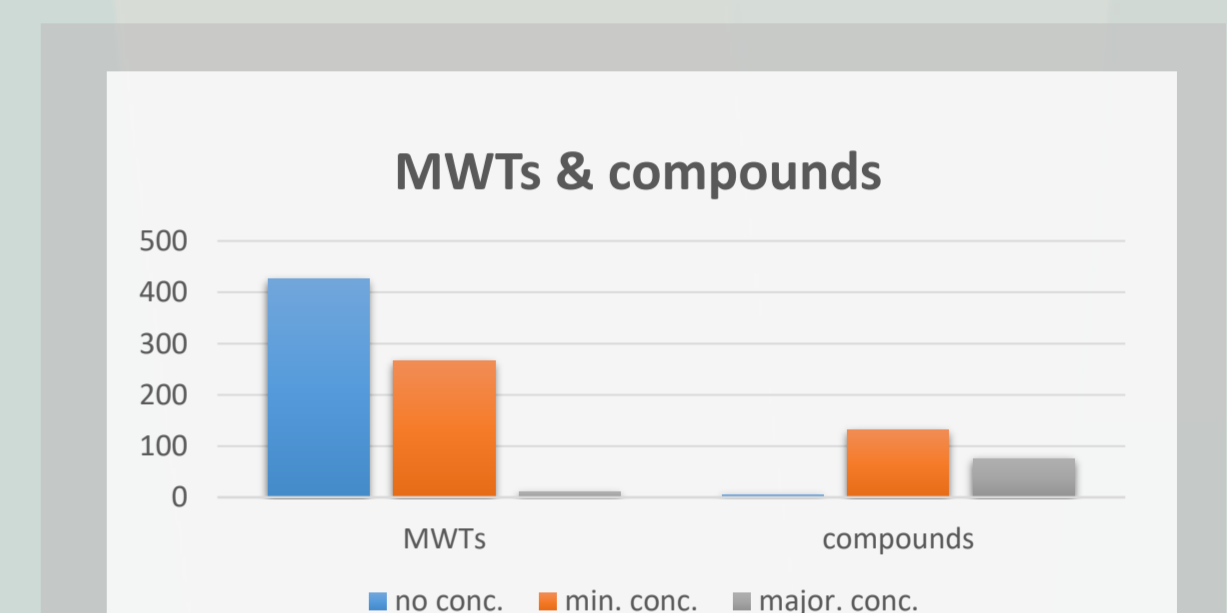
domain	DS	IND	TR	GL
hunting	5.32	3.74	4.12	3.44
chess	5.08	3.72	3.75	2.93

Average agreement on ambiguous terms

- General and domain-specific senses: *Wiktionary, Duden* and *Wikipedia*
- Results: ambiguous terms were often not selected (although often being highly specific)
- → either overseen or not considered relevant due to general language shape

Automatic Term Extraction

- Term ranking by annotator concordance compared to ranking by **hybrid term-candidate extractor** (Rösiger et al., 2016)
- → term extractors rank compounds and MWTs higher than the laypeople do

Conclusion

- Laypeople generally share a common notion of termhood

i. High inter-annotator variance for more specific tasks,
ii. little awareness of the degree of termhood of ambiguous terms and
iii. low agreement on MWTs with high reliance on subterms

- → show that laypeople's judgments deteriorate for specific and potentially unknown terms

References:
- Rosa Estopà. 2001. Les unités de signification spécialisées: élargissant l'objet du travail en terminologie. Terminology. 7(2):. 217–237.
- Ina Rösiger, Julia Bettinger, Johannes Schäfer, Michael Dorna, and Ulrich Heid. 2016. Acquisition of semantic relations between terms: How far can we get with standard NLP tools? In Proceedings of CompuTerm.
Pictures:
- https://www.tischlereicenter.eu/werkzeug/saegen/fuchsschwanz-nach-din-7244.html
- https://en.wikipedia.org/wiki/Red_fox#/media/File:Dogs,_jackals,_wolves,_and_foxes_(Plate_XXII).jpg

Universität Stuttgart

Institut für Maschinelle Sprachverarbeitung

BOSCH