RUHR UNIVERSITÄT BOCHUM RUB

LDSL Linguistic Data Science Lab

GerEO: A Large-Scale Resource on the Syntactic Distribution of German Experiencer-Object Verbs

Johanna M. Poppek, Simon Masloch and Tibor Kiss
{johanna.poppek, simon.masloch, tibor.kiss}@rub.de
Linguistic Data Science Lab, Ruhr-Universität Bochum

Features

How to Access

Open Access (CC-License)

github.com/Linguistic-Data-Science-Lab/German_EO_verbs

EO Verbs

Experiencer-object (EO) verbs: psych verbs whose experiencer argument is linked to the object position while the stimulus is linked to the subject position

(1) Das beeindruckt ihn. (EO acc)
    This impresses   him.ACC
(2) Das behagt   ihm. (EO dat)
    This pleases   him.DAT

The syntax and semantics of these verbs is heavily debated (see Rozwadowska et al., 2020 for a recent overview). The possibility of certain verbs to undergo certain argument structure alternations has often been used to argue for syntactic analyses.

Yet, existing corpus research on this domain in German is either restricted to a narrow question or quite limited in scope, there is **no larger corpus-based data or resource with annotated data**

Annotation

~10,300 annotated sentences
64 German EO verbs
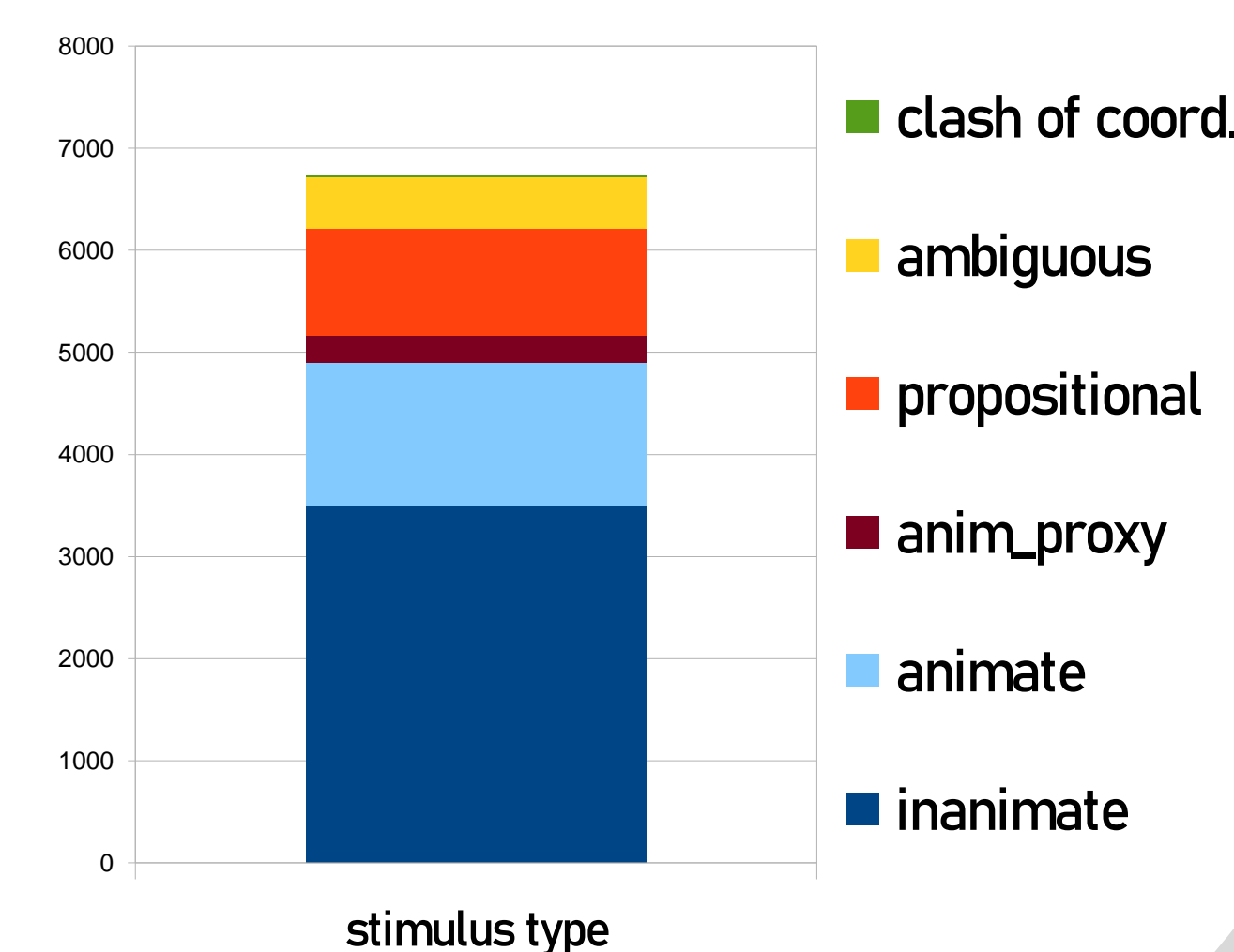Sentences randomly extracted from a newspaper corpus (*Neue Zürcher Zeitung*, NZZ 1993–1999)

Candidate verbs:
• Selected based on previous experimental and corpus studies
• Balanced on overall corpus frequency, case preference (dative and accusative), morphological structure, and perfect tense auxiliary selection preference
• Morphological minimal pairs of root and prefixed verbs, like *wundern* 'to wonder' and *verwundern* 'to astonish'

Annotated by three native speakers of German, annotation guidelines (Masloch et al., 2021) were developed with the MATTER-method (Pustejovsky and Stubbs, 2013)
Adjudication stage, each annotation verified by at least a simple majority

Distribution of Stimulus Types

clash of coord.
ambiguous
propositional
anim_proxy
animate
inanimate

stimulus type

Stimulus Type

Semantic category of the stimulus. Relevant because a proposed deviant syntactic behaviour is often argued to be absent on agentive usages of the verbs.

inanim	inanimate
anim	animate (persons, animals)
ambiguous	non-disambiguated pronouns etc.
anim_proxy	institutions, organisations
propositional	proposition-denoting clauses, some pronouns
clash_of_coordinations	coordinative stimulus with varying animacy

Stimulus PP

A possible stimulus PP – i.e. a PP whose internal argument refers to (part of) the stimulus causing the experiencer's emotion

(3) **Sie fielen durch eine unsichere
    they struck through a   insecure
    Fahrweise auf.**
    style.of.driving PRT
    'They attracted attention due to their insecure way of driving.'

Non-Psych Readings

Some verbs are ambiguous and also possess non-psych readings (example: *bewegen* ('to move').

Such examples are also annotated in the usual way in order to make a comparison possible.
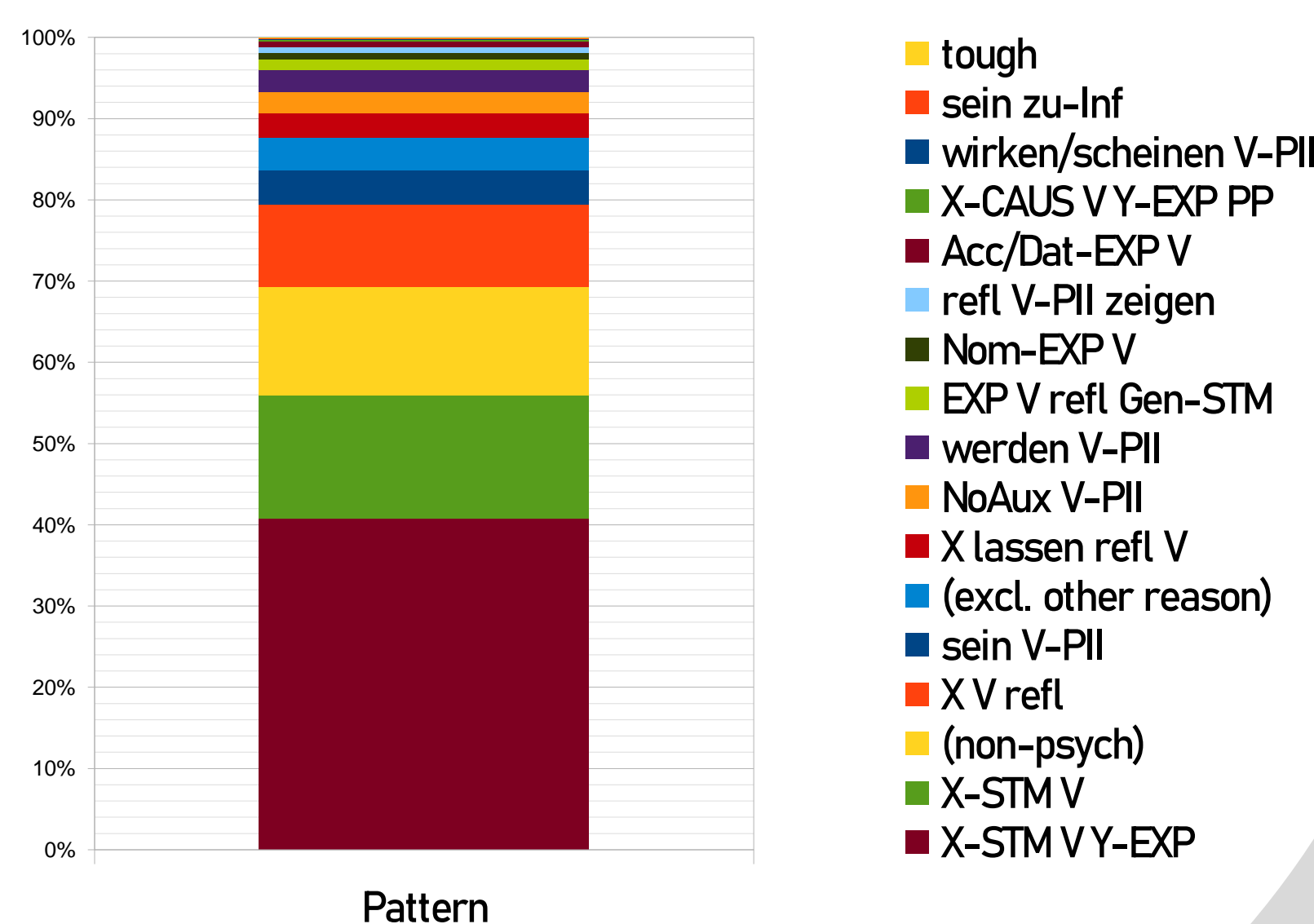
Syntactic Pattern

The core of the distributional annotation: Columns that each represent one syntactic pattern. There is huge variation between verbs in terms of the patterns they may enter and the observed frequency.

X–STM V Y–EXP   EO transitive: STM subject, EXP object
(4) Die Aura   der   Stararchitekten bezaubert neuerdings die   Welt.
    the.NOM aura.NOM the.GEN star.architects.GEN charms recently the.ACC world.ACC
    'The aura of star architects recently charms the whole world.'

X–STM V   EO intransitive, STM subject, EXP object dropped
(5) Die   musikalische Leistungen imponierten fast durchweg.
    the.NOM musical.NOM performances.NOM impressed almost throughout
    'The musical performances impressed almost without exception / the entire time.'

Acc/Dat-EXP V   no (phoric) subject
(6) Tut   mir   leid!
    feel.sorry.3SG me.DAT PRT
    'I'm sorry!'

werden V-PII   verbal passive
(7) Wer   soll eingeschüchtert werden?
    who.NOM shall intimidated   become
    'Who is to be intimidated?'

sein V-PII   stative passive
(8) Auch Staehelin   ist deprimiert.
    also Staehelin.NOM is depressed
    'Staehelin is also depressed.'

X V refl   reflexive pattern
(9) Der   Kongress amüsiert sich.
    the   congress amuses   REFL
    'The congress enjoys itself.'

X *lassen* refl V   *let*-reflexive construction
EXP V refl Gen-STM   reflexive pattern with genitive stimulus
Nom-EXP V   anticausative (with *ablaut*)
X-CAUS V Y-EXP PP   subject as causer with oblique object of emotion
(10) [...] der   [...] eine Henne [...] für seinen Hort begeistern konnte
     who.NOM   a.ACC hen.ACC for his   nest enthuse   could
     'who was able to make a hen like his nest'

NoAux V-PII   participle (V-PII)
refl V-PII *zeigen*   embedding under *sich zeigen* 'to show oneself'
wirken/scheinen V-PII   embedding under *wirken/scheinen* 'to appear'
sein zu-Inf   modal infinitive
tough-construction   *tough*-movement

Usage

• Review, develop, and falsify theoretical hypotheses
• Gold standard data for AI, modeling
• Aid the design of experiments: Detect properties of verbs that may cause interference effects (e.g. frequent alternative readings), search for sentences with specific properties to develop realistic stimuli (Börner et al., 2019)

Distribution of Patterns

tough
sein zu-Inf
wirken/scheinen V-PII
X-CAUS V Y-EXP PP
Acc/Dat-EXP V
refl V-PII zeigen
Nom-EXP V
EXP V refl Gen-STM
werden V-PII
NoAux V-PII
X lassen refl V
(excl. other reason)
sein V-PII
X V refl
(non-psych)
X-STM V
X-STM V Y-EXP

Pattern

Contact or Questions?

✉ johanna.poppek@rub.de
✉ simon.masloch@rub.de

https://ldsl.rub.de/research/projects/stm-subj

Verb	Case	Sample	Verb	Case	Sample
amüsieren 'to amuse'	acc	200	frustrieren 'to frustrate'	acc	125
anekeln 'to sicken'	acc	20	gefallen 'to like'	dat	200
ängstigen 'to frighten'	acc	109	genügen 'to suffice'	dat	200
anwidern 'to disgust'	acc	32	guttun 'to benefit, comfort'	dat	200
ärgern 'to anger'	acc	200	imponieren 'to impress'	dat	200
auffallen 'to strike'	dat	200	interessieren 'to interest'	acc	200
aufstoßen 'to strike (neg.)'	dat	200	irritieren 'to irritate, confuse'	acc	200
aufwühlen 'to stir up'	acc	117	langweilen 'to bore'	acc	200
ausreichen 'to suffice'	dat	200	leidtun 'to feel sorry'	dat	200
bedrücken 'to distress, depress'	acc	87	missfallen 'to displease'	dat	200
beeindrucken 'to impress'	acc	200	nahegehen 'to afflict, upset'	dat	29
befremden 'to alienate'	acc	177	nerven 'to bother'	acc	141
begeistern 'to thrill, enthuse'	acc	200	peinigen 'to tantalise'	acc	85
behagen 'to please'	dat	200	plagen 'to plague'	acc	200
bekümmern 'to concern, chagrin'	acc	60	provozieren 'to provoke'	acc	200
beruhigen 'to calm'	acc	200	quälen 'to torment'	acc	200
beschämen 'to shame'	acc	65	schmeicheln 'to flatter'	dat	200
beunruhigen 'to worry'	acc	200	schockieren 'to shock'	acc	200
bewegen 'to move'	acc	200	schwerfallen 'to find difficult'	dat	200
bezaubern 'to charm'	acc	138	stören 'to disturb, bother'	acc	200
deprimieren 'to depress'	acc	60	überfordern 'to be too much, overwhelm'	acc	200
einleuchten 'to be evident'	dat	200	verängstigen 'to frighten'	acc	36
einschüchtern 'to intimidate'	acc	200	verärgern 'to annoy'	acc	200
ekeln 'to disgust'	acc	23	verblüffen 'to flabbergast'	acc	200
empören 'to outrage'	acc	200	verschrecken 'to scare'	acc	58
entmutigen 'to discourage'	acc	200	verstören 'to distract, distress'	acc	52
entzücken 'to delight'	acc	190	verwirren 'to confuse'	acc	200
erfreuen 'to enjoy, delight'	acc	200	verwundern 'to astonish'	acc	200
erheitern 'to cheer, brighten'	acc	59	widerstreben 'to oppose, have an aversion'	dat	76
erschrecken 'to startle'	acc	200	wundern 'to wonder'	acc	200
faszinieren 'to fascinate'	acc	200	zermürben 'to demoralize, grind'	acc	152
freuen 'to please, be glad'	acc	200	zusetzen 'to badger, harass'	dat	200

Further Perspectives

• With 64 verbs, our resource already covers a larger portion of German EO verbs, however, we plan to expand both verb number and sample size in the future.
• GerEO allows to detect heterogeneous behaviour among the whole group of annotated EO verbs as well as supposed subclasses proposed in the literature. Databases of this kind may aid both theorising as well as experimental efforts.

Other Features

Features like *Control construction* and *Other stimulus adjuncts* (related to Stimulus PP)

Comment-Column with STM-, EXP-, sentence-related aspects (metaphors, linguistic variation, polysemy etc.)

References

Börner, A. K., Pieper, J., and Kiss, T. (2019). Corpus Data in Experimental Linguistics. In Proceedings of Linguistic Evidence 2018: Experimental Data Drives Linguistic Theory, Tübingen. University of Tübingen.
Masloch, S., Poppek, J. M., Robrecht, A., and Kiss, T. (2021). Syntactic pattern distribution analysis of experiencer-object psych verbs: An annotation manual. Studies in Linguistics and Linguistic Data Science, 4. https://ldsl.rub.de/media/pages/research/resources/sllds/a3914a6ac2f640260430/sllds_4.pdf.
Pustejovsky, J. and Stubbs, A. (2013). Natural Language Annotation for Machine Learning: A Guide to Corpus-Building for Applications. O'Reilly, Sebastopol.
Rozwadowska, B., Nowak, A., and Bondaruk, A. (2020). Psych verbs: Setting the scene. In Bożena Rozwadowska et al., editors, Beyond Emotions in Language: Psychological verbs at the interfaces, Pages 1–21. John Benjamins, Amsterdam.