

# G<sub>h</sub>oSt-PV: A Representative Gold Standard of German Particle Verbs

Stefan Bott, Nana Khvtisavrishvili, Max Kisselew, Sabine Schulte im Walde

Institut für Maschinelle Sprachverarbeitung – Universität Stuttgart  
 {stefan.bott|nana.khvtisavrishvili|max.kisselew|schulte}@ims.uni-stuttgart.de

## Goal

Creation of a Gold Standard for the study of German Particle Verbs (PVs) which should:

- Represent levels of compositionality
- Be reasonable sized
- Be randomly sampled
- Be balanced

## Motivation

German particle verbs:

- Especially frequent in German
- A highly productive paradigm; it frequently produces neologisms
- Different levels of lexical semantic compositionality
- In some syntactic paradigms particle verbs are separable:

(1) a. Das Kind **sah** seine Mutter **an**.  
 The child looked his/her mother an-PRT.  
 'The child looked at his/her mother.'

b. .... dass das Kind seine Mutter **an|sah**.  
 ... that the child his/her mother an-PRT|looked.  
 '... that the child looked at his mother.'

## German Particle Verbs: Levels of Compositionality

FULLY COMPOSITIONAL: e.g. *an|leuchten* (to illuminate); *an* expresses directionality (among other senses)

(2) Peter **leuchtete** das Bild mit der Lampe **an**.  
 Peter shined the picture with the lamp an-PRT.  
 'Peter illuminated the picture with the lamp.'

SEMI-COMPOSITIONAL: e.g. *ab|segnen* (to approve); a meaning shift occurred from *segnen* (to bless) to this PV. Such PVs are usually part of a productive paradigm: *ab|segnen* patterns with verbs like *ab|nicken* and *ab|zeichnen*.

(3) Der Chef **segnete** die Pläne **ab**.  
 The boss blessed the plans ab-PRT.  
 'The boss approved the plans.'

NON-COMPOSITIONAL: e.g. *nach|schlagen* (to look up (e.g. a reference) or to consult (e.g. a dictionary)); the BV *schlagen* means to beat.

(4) Stella **schlug** das Wort im Wörterbuch **nach**.  
 Stella beat the word in-the dictionary nach-PRT.  
 'Stella looked up the word in the dictionary.'

## Desired Properties of the Gold Standard

- *Random selection*: In order to avoid bias we wanted to obtain a random sample of all existing PVs.
- *Scalar judgments on compositionality*: The degree of compositionality falls on a continuum from fully compositional to non-compositional.
- *Balanced over frequency bands*: Both very frequent and very sparse PVs tend to present special problems (Bott and Schulte im Walde, 2014).
  - high-frequency items: strongly lexicalization and ambiguity
  - low-frequency items: data sparseness issues

- *Different ambiguity levels*: Polysemy is a factor which influences both human ratings and automatic computational assessment.

- *Selection of particles*:

- PVs with de-prepositional verb particles.
- 11 particles: *an, auf, aus, nach, ab, zu, ein, über, unter, um, durch*
- High tendency towards particle ambiguity and abstract readings.

## Gold Standard Creation

### Compilation of a full list of PVs

We looked for combinations of verbs and particles which occurred both

- written together as one word and
- syntactically separated, relying on a dependency-parsed version of the *SdeWaC* corpus.

PROBLEMS:

- Verbs look accidentally like PVs: *zupfen* (to pluck/pick) is not a PV with *zu*.
- Lemmatization and parsing errors (e.g. prepositions may be interpreted as particles in syntactically separated cases).
- PVs may be confounded with prefix verbs; some verbs have homophones as prefix- and particle verbs.

### Verb Selection Process

- Random selection
- Balanced over 3 frequency bands (tertiles computed per particle)

### Cleaning

Problematic entries were excluded:

- Prefix verbs and PVs with homophone prefix verbs
- Possibly non-existing verbs
- Extremely high frequent and low frequent verbs (20 at each extreme per particle)

### Collection of Ratings

- Over Amazon Mechanical Turk
- German native speakers only (filtered with bogus test items)
- Without given context
- Each item was rated by 16.14 raters, average (min 7)
- Rating was done on a scale from 1 to 7

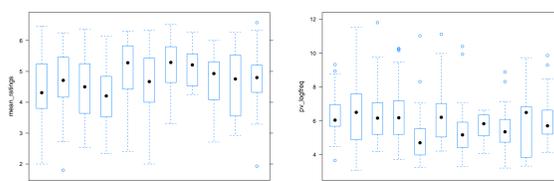


Figure 1: Mean Ratings and log frequencies of particle verbs across particle types.

PV	PV freq	freq band	ambig. band	no. raters	mean rating	std dev	prop. synt. sep.	prop. non-sep.
abkratzen	39.80	M	AG3	14	5.29	2.52	0.16	0.84
absegnen	23.38	H	A1	14	4.07	1.90	0.09	0.91
anleuchten	6.37	L	A1	20	5.95	1.50	0.62	0.38
anstiften	7.92	M	A2	15	1.80	0.86	0.17	0.83
aufhorchen	74.58	H	A1	29	4.55	1.97	0.16	0.84
aufschneiden	43.31	H	AG3	14	6.07	1.73	0.32	0.68
ausreizen	19.35	M	A2	29	3.62	2.13	0.07	0.93
durchrosten	9.66	M	A1	14	6.29	0.73	0.31	0.69
einstampfen	33.34	H	A1	14	4.07	2.06	0.15	0.85
nachschicken	22.81	H	A1	15	6.00	1.07	0.29	0.71
nachtragen	3.97	L	A2	15	4.47	2.03	0.21	0.79
umplanen	14.44	M	A2	15	4.93	1.83	0.10	0.90
zukneifen	8.53	M	A2	14	4.71	1.77	0.33	0.67
zulegen	4.00	L	AG3	14	3.86	2.07	0.29	0.71

Table 1: Sample entries from the gold standard.

## Information Included in the Gold Standard

The gold standard contains 400 German particle verbs

- PV lemma
- Harmonic mean of PV corpus frequencies across four corpora
- The PV frequency band (low, mid, high)
- The PV level of ambiguity (ambiguities of 1, 2, 3 or greater than 3)
- The number of human ratings for the PV
- The mean compositionality rating for each PV
- The standard deviation of ratings among raters, as a measure of agreement
- The proportions of syntactically separated and syntactically non-separated appearances of the PV

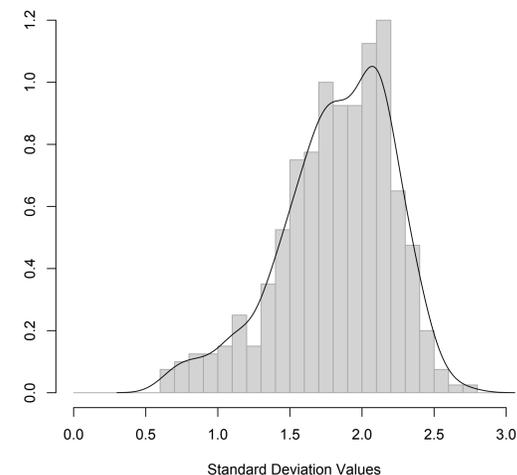


Figure 2: Histogram of the distribution and approximate density of standard deviation values for compositionality ratings across PVs. (StdDev approximates inter-annotator agreement per item.)

## Availability

Available under a Creative Commons ShareAlike non-commercial 4.0 License.  
<http://www.ims.uni-stuttgart.de/data/ghost-pv>

## Acknowledgments

The research was supported by the DFG Research Grant SCHU 2580/2 (Stefan Bott) and the DFG Heisenberg Fellowship SCHU-2580/1 (Sabine Schulte im Walde).

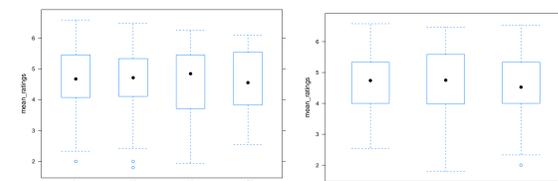


Figure 3: Mean compositionality ratings across ambiguity levels and frequency bands.