

# Assessing Meaning Components in German Complex Verbs: A Collection of Source–Target Domains and Directionality

Sabine Schulte im Walde and Maximilian Köper and Sylvia Springorum

Institut für Maschinelle Sprachverarbeitung

Universität Stuttgart, Germany

{schulte,maximilian.koeper,sylvia.springorum}@ims.uni-stuttgart.de

## Abstract

This paper presents a collection to assess meaning components in German complex verbs, which frequently undergo meaning shifts. We use a novel strategy to obtain source and target domain characterisations via sentence generation rather than sentence annotation. A selection of arrows adds spatial directional information to the generated contexts. We provide a broad qualitative description of the dataset, and a series of standard classification experiments verifies the quantitative reliability of the presented resource. The setup for collecting the meaning components is applicable also to other languages, regarding complex verbs as well as other language-specific targets that involve meaning shifts.

## 1 Introduction

German particle verbs (PVs) are complex verb structures such as *anstrahlen* ‘to beam/smile at’ that combine a prefix particle (*an*) with a base verb (*strahlen* ‘to beam’). PVs represent a type of multi-word expressions, which are generally known as a “*pain in the neck for NLP*” (Sag et al., 2002). Even more, German PVs pose a specific challenge for NLP tasks and applications, because the particles are highly *ambiguous*; e.g., the particle *an* has a partitive meaning in *anbeißen* ‘to take a bite’, a cumulative meaning in *anhäufen* ‘to pile up’, and a topological meaning in *anbinden* ‘to tie to’ (Springorum, 2011). In addition, they often trigger *meaning shifts* of the base verbs (BVs), cf. Springorum et al. (2013); e.g., the PV *abschminken* with the BV *schminken* ‘to put on make-up’ has a literal meaning (‘to remove make-up’) and a shifted, non-literal meaning (‘to forget about something’).<sup>1</sup>

<sup>1</sup>We deliberately make use of the general term “meaning shift” in comparison to specific instances such as metaphor

With PVs representing a large and challenging class in the lexicon, their meaning components and their mechanisms of compositionality have received a considerable amount of interdisciplinary research interest. For example, a series of formal-semantic analyses manually classified German PVs (with particles *ab*, *an*, *auf*, *nach*) into soft semantic classes (Lechler and Roßdeutscher, 2009; Haselbach, 2011; Kliche, 2011; Springorum, 2011). Corpus studies and annotations demonstrated the potential of German PVs to appear in non-literal language usage, and to trigger meaning shifts (Springorum et al., 2013; Köper and Schulte im Walde, 2016b). Regarding computational models, the majority of existing approaches to PV meaning addressed the automatic prediction of German PV compositionality (Salehi et al., 2014; Bott and Schulte im Walde, 2015; Köper and Schulte im Walde, 2017b), in a similar vein as computational approaches for English PVs (Baldwin et al., 2003; Bannard, 2005; McCarthy et al., 2003; Kim and Baldwin, 2007; Salehi and Cook, 2013; Salehi et al., 2014). Only few approaches to German and English PVs have included the meaning contributions of the particles into the prediction of PV meaning (Bannard, 2005; Cook and Stevenson, 2006; Köper et al., 2016).

Overall, we are faced with a variety of interdisciplinary approaches to identifying and modelling the meaning components and the composite meanings of German PVs. Current and future research activities are however hindered by a lack of resources that go beyond PV–BV compositionality and can serve as gold standards for assessing

- (i) the meaning contributions of the notoriously ambiguous particles, and
- (ii) meaning shifts of PVs in comparison to their BVs.

and metonymy because non-literal language usage of PVs is not restricted to a specific type of meaning shift.

In this paper, we present a new collection for German PVs that aims to improve on this situation. The dataset includes 138 German BVs and their 323 existing PVs with particle prefixes *ab*, *an*, *auf*, *aus*. For all target verbs, we collected

1. sentences from 15 human participants across a specified set of domains, to address their ambiguity in context; and
2. spatial directional information (UP, DOWN, RIGHT, LEFT), also in context.

Meaning shifts are typically represented as a mapping from a rather concrete source-domain meaning to a rather abstract target-domain meaning (Lakoff and Johnson, 1980). For example, the abstract conceptual domain TIME may be illustrated in terms of the structurally similar, more concrete domain MONEY, enabling non-literal language such as *to save time* and *to spend time*. For German PVs, meaning shifts frequently take place when combining a BV from a concrete source domain with a particle (as in the *abschminken* example above, where the BV *schminken* is taken from the domain HUMAN BODY), resulting in a PV meaning (possibly among other meanings) related to an abstract target domain such as DESIRE.

Targeting the representation of meaning shifts with our collection, we specified source domains for the BVs (such as MENSCHLICHER KÖRPER 'HUMAN BODY') and target domains for the PVs (such as ZEIT 'TIME'). In this way, our dataset offers source–target domain combinations for assessing BV–PV meaning shifts across PVs and particle types. Our domains were taken from conceptual specifications in (Kövecses, 2002), which cluster semantically and encyclopedically related concepts to ensure a generally applicable set of domains involved in meaning shifts. The spatial directional information is captured through simple directional arrows and enables a view on spatial meaning components of particle types and PVs, which supposedly represent core meaning dimensions of PVs (Frassinelli et al., 2017).

While the collection focuses on German PVs, the representation of the meaning components (source and target domains, as well as directions) is language-independent. Therefore, the setup for collecting the meaning components that we present below should also be applicable to other languages, regarding complex verbs as well as regarding other language-specific targets that undergo meaning shifts.

## 2 Related Work

### PV Meaning Components and Classifications

So far, the most extensive manual resources regarding German PV meaning components rely on formal semantic research within the framework of Discourse Representation Theory (DRT), cf. Kamp and Reyle (1993). Here, detailed word-syntactic analyses and soft classifications were created for German PVs with the particles *auf* (Lechler and Roßdeutscher, 2009), *nach* (Haselbach, 2011), *ab* (Kliche, 2011), and *an* (Springorum, 2011).

### PV Compositionality

Most manual and computational research on PV meaning addressed the meaning of a PV through its degree of compositionality, for German as well as for English complex verbs. McCarthy et al. (2003) exploited various measures on distributional descriptions and nearest neighbours to predict the degree of compositionality of English PVs with regard to their BVs. Baldwin et al. (2003) defined Latent Semantic Analysis (LSA) models (Deerwester et al., 1990) for English PVs and their constituents, to determine the degree of compositionality through distributional similarity, and evaluated the predictions against various WordNet-based gold standards. Bannard (2005) defined the compositionality of an English PV as an entailment relationship between the PV and its constituents, and compared four distributional models against human entailment judgements. Cook and Stevenson (2006) addressed not only the compositionality but also the meanings of English particles and PVs. Focusing on the particle *up*, they performed a type-based classification using window-driven and syntactic distributional information about the PVs, particles and BVs. Kim and Baldwin (2007) combined standard distributional similarity measures with WordNet-based hypernymy information to predict English PV compositionality. Kühner and Schulte im Walde (2010), Bott and Schulte im Walde (2017) and Köper and Schulte im Walde (2017a) used unsupervised (soft) clustering and multi-sense embeddings to determine the degree of compositionality of German PVs. Salehi and Cook (2013) and Salehi et al. (2014) relied on translations into multiple languages in order to predict the degree of compositionality for English PVs. Bott and Schulte im Walde (2014) and Bott and Schulte im Walde (2015) explored and

compared word-based and syntax-based distributional models in the prediction of German PVs. Köper and Schulte im Walde (2017b) integrated visual information into a similar textual distributional model.

Altogether, most PV gold standards that are used for evaluation within the above approaches to compositionality rate the similarity between PV and BV, ignoring the contribution of the particle meaning. Exceptions to this is the gold standard by Bannard (2005), rating the entailment between the PV and its particle as well as between the PV and its BV. In addition, all PV gold standards are type-based, i.e., rating the compositionality for a PV type, rather than for PV senses in context.

**Spatial Meaning Components** The Grounding Theory indicates that the mental representation of a concept is built not only through linguistic exposure but also incorporating multi-modal information extracted from real-world situations, including auditory, visual, etc. stimuli (Barsalou, 1999; Glenberg and Kaschak, 2002; Shapiro, 2007). Spatial meaning plays an important role in grounding information. For example, Richardson et al. (2003) showed an interaction between spatial properties of verbs and their positions in language comprehension. Dudschig et al. (2012) and Kaup et al. (2012) demonstrated effects of typical locations of a word’s referent in language processing. Specifically for German PVs, Frassinelli et al. (2017) found spatial meaning (mis)matches for PVs with particles *an* and *auf*, when combining them with primarily vertical vs. horizontal BVs. The spatial information in our dataset provides an opportunity to further explore spatial meaning components in German BVs and PVs.

**Meaning Shift Datasets** Lakoff and Johnson (1980) and Gentner (1983) were the first to specify systematic conceptual mappings between two domains, within their theories of conventional metaphors and analogy by structure-mapping, respectively. In contrast, practical advice and projects on the actual annotation of source/domain categorisations or meaning shifts are sparse. The *Master Metaphor List (MML)* represents an extensive manual collection of metaphorical mappings between source and target domains (Lakoff et al., 1991) but from a practical point of view has been criticised for its incoherent levels of specificity and its lack of coverage by Lönneker-Rodman (2008),

who relied on the MML next to EuroWordNet when annotating a total of 1,650 French and German metaphor instances. Similarly, Shutova and Teufel (2010) used the source and target domains from the MML but relied only on a subset of the domains, which they then extended for their annotation purposes.

As to our knowledge, there is no previous dataset on meaning shifts of complex verbs, other than a smaller-scale collection developed in parallel by ourselves, which however focuses on analogies in meaning shifts rather than source–target domains (Köper and Schulte im Walde, 2018). Some datasets include non-literal meanings of verbs (Birke and Sarkar, 2006; Turney et al., 2011; Shutova et al., 2013; Köper and Schulte im Walde, 2016b), and the MML-based meaning shift annotations by Lönneker-Rodman (2008) and Shutova and Teufel (2010) also include verbs but are less target-specific than our work. In addition, while both Lönneker-Rodman (2008) and Shutova and Teufel (2010) asked their annotators to label words in their corpus data, we follow a different strategy and ask our participants to generate sentences according to domain-specific target senses.

### 3 Target Verbs, Domains, Directionalities

In this section, we describe our selections and representations of BV and PV targets (Section 3.1), the source and target domains (Section 3.2), and the directional arrows (Section 3.3).

#### 3.1 German Base and Particle Verbs

Based on the source domain descriptions by Kövecses (2002), cf. Section 3.2 below, we identified BVs which (i) supposedly belong to the respective source domain, and (ii) we expected to undergo meaning shifts when combined with one of our target particle types, as based on our linguistic expertise from previous work (see related work above).

All of the BVs were systematically combined with the four prefix particles *ab*, *an*, *auf*, *aus*, resulting in a total of 552 PVs. Since we did not want to include neologisms into our PV targets, we then checked the PV existence in the online version of the German dictionary *DUDEN*<sup>2</sup>. The final list of target PVs that were found in the dictionary comprised 323 verbs.

<sup>2</sup>[www.duden.de/suchen/dudenonline/](http://www.duden.de/suchen/dudenonline/)

Source Domains		Target Domains	
Menschlicher Körper	Human Body	Emotion und Gefühl	Emotion and Feeling
Gesundheit und Krankheit	Health and Illness	Wunsch und Sehnsucht	Desire
Tiere	Animals	Moral	Morality
Pflanzen	Plants	Gedanke	Thought
Gebäude und Konstruktion	Buildings and Construction	Gesellschaft und Nation	Society and Nation
Maschinen und Werkzeuge	Machines and Tools	Wirtschaft und Ökonomie	Economy
Spiele und Sport	Games and Sports	Menschliche Beziehungen	Human Relationships
Geld und Handel	Money and Economic Transaction	Kommunikation	Communication
Kochen und Essen	Cooking and Food	Zeit	Time
Hitze und Kälte	Heat and Cold	Leben und Tod	Life and Death
Licht und Dunkelheit	Light and Darkness	Religion	Religion
Kräfte	Forces	Ereignis und Handlung	Event and Action
Bewegung und Richtung	Movement and Direction		
Geräusch und Klang	Sound		

Table 1: Source and target domains.

### 3.2 Domains of Meaning Shifts

The *Master Metaphor List (MML)* provides the most extensive list of source–domain shift definitions but has been criticised for being incomplete regarding corpus annotations (Lönneker-Rodman, 2008; Shutova and Teufel, 2010), cf. Section 2. In addition, we found the MML and an extended subset as provided by Shutova and Teufel (2010) impractical to apply because the lists use too many categories that are based on too diverse motivations, such as event structures (e.g., change, causality, existence, creation) vs. event types (e.g., mental objects, beliefs, social forces).

Instead, our source and target domains were taken from specifications in (Kövecses, 2002), which we assumed to ensure a more stratified and generally applicable set of domains involved in meaning shifts. Table 1 lists all 13 source and 12 target domains by Kövecses (2002), including both the original English terms from Kövecses (2002) and the German translations that we used in our collection. Regarding the source domains, we added one domain to Kövecses’ original list, i.e., SOUND, which we expected to play a role in BV–PV meaning shifts (Springorum et al., 2013).

### 3.3 Spatial Directionality Arrows

According to Viberg (1983), spatial experience provides a cognitive structure for the concepts underlying language. Given that we focus on PVs with prepositional particles (*ab*, *an*, *auf*, *aus*), we assume that the particles are spatially grounded, similar to preposition meanings which indicate spatial fundamentals (Herskovits, 1986; Dirven, 1993) and structure space regarding location, orientation, and direction (Zwarts, 2017).

We decided to focus on directionality as a central function in space, and to use arrows as visual expressions of directional meaning, given that (i) visual expressions are supposedly analogous expressions in language and categorise meaning, cf. Tversky (2011); (ii) arrows are asymmetric lines that “fly in the direction of the arrowhead” and provide structural organisation (Heiser and Tversky, 2006; Tversky, 2011); and (iii) directed arrows provide a simple but unambiguous depictive expression for direction in space. Our selection of arrows uses the four basic directions

UP ↑                      LEFT ←  
DOWN ↓                    RIGHT →

## 4 Dataset<sup>3</sup>

In this section, we describe our collection of meaning components from three different perspectives: the instructions for annotators (Section 4.1), a broad qualitative description of the dataset (Section 4.2), and classification experiments to verify the quantitative value of the resource (Section 4.3).

### 4.1 Annotation Instructions

We randomly distributed BVs and PVs over lists with 35 verbs each. The annotators were asked

- (i) to choose one or more pre-defined semantic domain classes for each verb,
- (ii) to provide an example sentence to illustrate the class assignment, and
- (iii) to select an arrow that intuitively corresponds to the generated example sentence.

<sup>3</sup>The dataset is publicly available from [www.ims.uni-stuttgart.de/data/pv-bv-domains/](http://www.ims.uni-stuttgart.de/data/pv-bv-domains/).

<b>heulen</b>					
Menschliche Körper	<input checked="" type="checkbox"/>	Das Kind heult schon den ganzen Tag.	↑	<input checked="" type="checkbox"/>	←
Gesundheit/ Krankheit	<input type="checkbox"/>		↑	↓	←
Tiere	<input type="checkbox"/>		↑	↓	←
Pflanzen	<input type="checkbox"/>		↑	↓	←
Gebäude/ Konstruktion	<input type="checkbox"/>		↑	↓	←
Maschinen/ Werkzeuge	<input checked="" type="checkbox"/>	Die Maschine heult durch die Halle.	↑	↓	<input checked="" type="checkbox"/>
Spiele/ Sport	<input type="checkbox"/>		↑	↓	←
Geld/Handel	<input type="checkbox"/>		↑	↓	←
Kochen/Essen	<input type="checkbox"/>		↑	↓	←
Hitze/Kälte	<input type="checkbox"/>		↑	↓	←
Licht/ Dunkelheit	<input type="checkbox"/>		↑	↓	←
Kräfte	<input type="checkbox"/>		↑	↓	←
Bewegung/ Richtung	<input type="checkbox"/>		↑	↓	←
Klang/ Geräusch	<input checked="" type="checkbox"/>	Die Sirene heult sehr laut.	<input checked="" type="checkbox"/>	↓	←

Figure 1: Example annotation for the verb *heulen* ‘to howl’ with (i) a selection of three source domain classes, (ii) the corresponding three sentences, and (iii) the corresponding three arrows.

The classes (i.e., the source domains in the BV lists, and the target domains in the PV lists) were described by key words (e.g., the German equivalents of *appearance*, *growth*, *cultivation*, *care*, *use* for the source domain PFLANZEN ‘PLANTS’). Then, the annotators were provided one example annotation (cf. Figure 1 for the verb *heulen* ‘to howl’) before they started the annotation process.

## 4.2 Qualitative Description

The annotations enable multiple views into meaning components of the underlying BVs and PVs on a token basis. In the following, we provide selected analyses and interactions regarding domains and directions (Section 4.2.1) and non-literal language and meaning shifts (Section 4.2.2).

### 4.2.1 Analyses of Domains and Directions

Table 2 shows the total number of sentences that were generated by the participants, and the pro-

portions per domain. Similarly, Table 3 shows the proportions per arrow type across the generated sentences.

In total, we collected 2,933 sentences across the 138 BVs and the 14 source domains, and 4,487 sentences across the 323 PVs and the 12 target domains. We find a rather skewed distribution for the number of sentences per verb type, varying between 2–47 for BVs and 1–30 for PVs; still, the collection comprises  $\geq 10$  sentences per verb for 134 out of 138 BVs (97%), and for 277 out of 323 PVs (86%), as illustrated in the number of sentences per verb type in Figures 2 and 3.

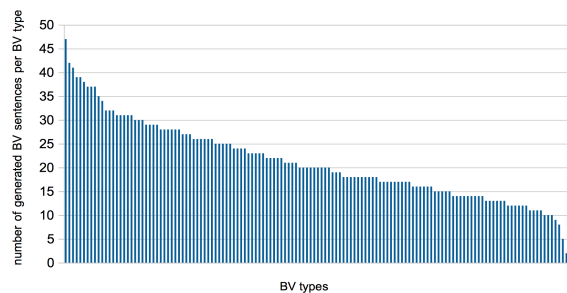


Figure 2: Number of generated sentences per BV.

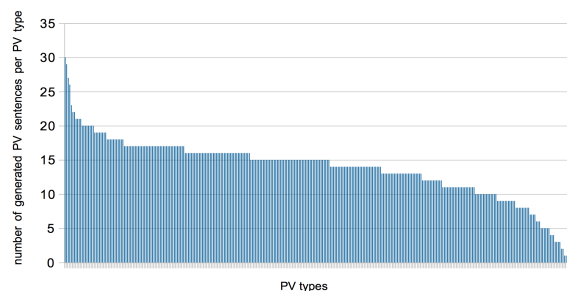


Figure 3: Number of generated sentences per PV.

The distribution of source domain sentences across domains ranges from a proportion of 3.41% for the domain FORCES up to 14.69% for the domain HUMAN BODY. The distribution of target domain sentences is more skewed, ranging from 0.47% for the domain RELIGION up to 33.88% for the domain EVENT/ACTION. Regarding directional information, we find a considerably low proportion of  $\approx 10\%$  for the left arrow ( $\leftarrow$ ), while the other three directions (up, down, right) received between 22% and 30%. Table 3 also shows that participants often chose more than one arrow for a specific generated sentence. We list those nine arrows and arrow combinations that were selected  $>50$  times in total, i.e., across BV and PV sentences.

Source Domains	No. of Sentences		Target Domains	No. of Sentences	
Human Body	431	14.69%	Event/Action	1,520	33.88%
Animals	322	10.98%	Economy	460	10.25%
Health/Illness	251	8.56%	Emotion/Feeling	452	10.07%
Machines/Tools	242	8.25%	Human Relationships	383	8.54%
Games/Sports	211	7.19%	Life/Death	365	8.13%
Cooking/Food	210	7.16%	Time	292	6.51%
Plants	207	7.06%	Thought	284	6.33%
Economic Transaction	190	6.48%	Communication	280	6.24%
Buildings/Construction	167	5.69%	Society/Nation	181	4.03%
Sound	165	5.63%	Desire	150	3.34%
Heat/Cold	156	5.32%	Morality	99	2.21%
Movement/Direction	154	5.25%	Religion	21	0.47%
Light/Darkness	127	4.33%			
Forces	100	3.41%			
<i>Total:</i>	2,933	100.00%	<i>Total:</i>	4,487	100.00%

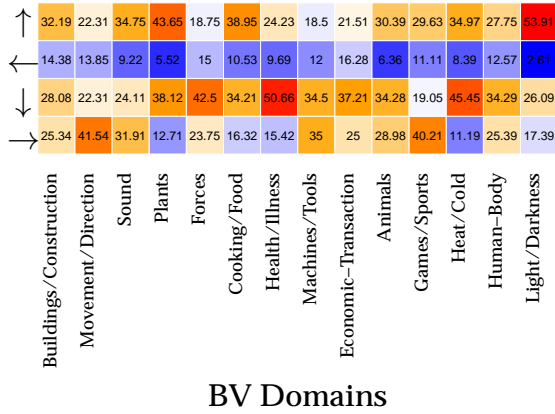
Table 2: Source and target domains: number and proportions of generated sentences per domain.

Source Domain Directions	No. of Sentences		Target Domain Directions	No. of Sentences	
↓	879	29.97%	→	1,300	28.97%
↑	782	26.66%	↓	1,218	27.15%
→	648	22.09%	↑	1,113	24.80%
←	270	9.21%	←	462	10.30%
↔	128	4.36%	↔	178	3.97%
↔↕	58	1.98%	↓→	52	1.16%
↕	50	1.70%	↑→	44	0.98%
↑→	16	0.55%	↕↔	28	0.62%
↓→	12	0.41%	↕	27	0.60%
<i>other combinations</i>	69	0.24%	<i>other combinations</i>	54	1.20%
<i>no choice</i>	21	0.72%	<i>no choice</i>	21	0.47%
<i>Total:</i>	2,933	100.00%	<i>Total:</i>	4,487	100.00%

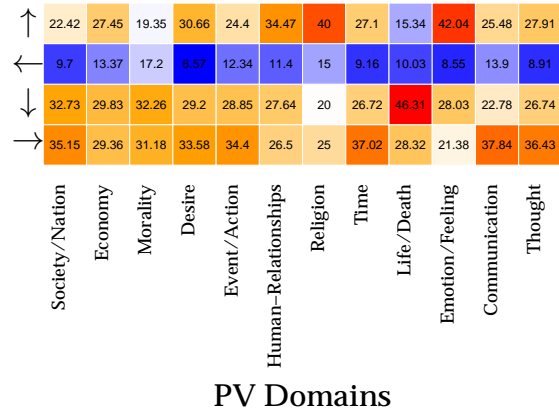
Table 3: Directional information: number and proportions of selected arrows and arrow combinations.

BV/PV	Domain	Direction	Sentence
BV	LIGHT/DARKNESS	↑	<i>Der Diamant <b>funkelt</b> im Licht.</i> 'The diamond sparkles in the light.'
BV	PLANTS	↓	<i>Die Blätter <b>fallen</b> von den Bäumen.</i> 'The leaves fall from the trees.'
BV	FORCES	←	<i>Er <b>bog</b> das Kupferrohr.</i> 'He bent the copper pipe.'
BV	ANIMALS	↔	<i>Die Bullen <b>fechten</b> miteinander.</i> 'The bulls fence with each other.'
BV	HEAT/COLD	↔↕	<i>Das Feuer <b>brennt</b> heiß.</i> 'The fire is burning hot.'
PV	MORALITY	↓	<i>Du solltest von deinem hohen Ross <b>absteigen</b>.</i> 'You should step down off your pedestal.'
PV	EMOTION/FEELING	↑	<i>Der Druck <b>wächst</b> kurz vor der Präsentation <b>an</b>.</i> 'The pressure increases shortly before the presentation.'
PV	HUMAN-RELATIONSHIPS	→	<i>Sie lässt ihn eiskalt <b>abblitzen</b>.</i> 'She turns him down cold-bloodedly.'
PV	LIFE/DEATH	↕	<i>Musst du mein ganzes Leben <b>aufwühlen</b>?</i> 'Do you have to chum up my whole life?'
PV	COMMUNICATION	↔	<i>Er <b>kauft</b> ihr die Lüge problemlos <b>ab</b>.</i> 'He believes her lie without any doubts.'

Table 4: Example BV and PV sentences with selected domains and directions.



(a) BV source domains and directionality.



(b) PV target domains and directionality.

Figure 4: Interaction of domains and directionality.

Figure 4 illustrates how source and target domains interact with the arrows as indicators of directionality. As in the overall picture in Table 3, the proportions for the direction LEFT are considerably lower than for the other directions, with few domains receiving up to 15–17%: FORCES and MONEY/ECONOMIC TRANSACTION in the source sentences, and MORALITY and RELIGION in the target sentences. The direction RIGHT is a very strong indicator of the source domains MOVEMENT/DIRECTION, GAMES/SPORTS, MACHINES/TOOLS and the target domains COMMUNICATION, TIME, THOUGHT, SOCIETY/NATION; the direction UP is a very strong indicator of the source domains LIGHT/DARKNESS, PLANTS and COOKING/FOOD and the target domains EMOTION/FEELING and RELIGION; the direction DOWN is a very strong indicator of the source domains HEALTH/ILLNESS, HEAT/COLD, FORCES, PLANTS and the target domain LIFE/DEATH; all of these strong indicators received proportions >35%. Table 4 presents example sentences for some BV and PV domain/arrow combinations.

Figure 5 breaks down the information on arrow directions across the four particle types. While the particles are notoriously ambiguous, we can see that across the PV target domain sentences three of the particle types (*ab*, *auf*, *aus*) show a predominant directional meaning, i.e., DOWN, UP, RIGHT, respectively. The particle *an* is more flexible in its directional meaning, which confirms prior assumptions (Frassinelli et al., 2017).

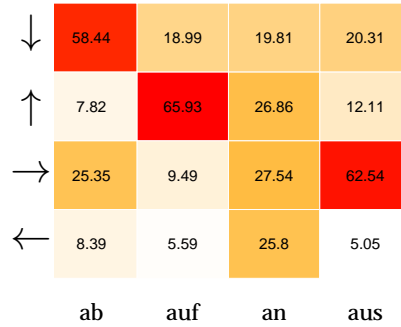


Figure 5: Directionality of particle types.

#### 4.2.2 Analyses of Meaning Shifts

We now take the first steps into analysing non-literal language and meaning shifts within our collection. We started out by assuming that “meaning shifts for German PVs frequently take place when combining a BV from a concrete source domain with a particle, resulting in a PV meaning (possibly among other meanings) related to an abstract target domain”. Consequently, the generated PV sentences are expected to (i) represent shifted, non-literal language meanings and to (ii) exhibit abstract meanings, both considerably more often than the generated BV sentences.

**(Non-)Literal BV/PV Language Usage** We asked three German native speakers to annotate the 2,933/4,487 BV/PV sentences with ratings on a 6-point scale [0,5], ranging from clearly literal (0) to clearly non-literal (5) language. Dividing the scale into two disjunctive ranges [0, 2] and [3, 5] broke down the ratings into binary decisions.

Table 5 shows the numbers and proportions of BV/PV sentences that were annotated as literal vs. non-literal language usage, distinguishing between full agreement (i.e., all annotators agreed on the binary category) and majority agreement (i.e., at least two out of three annotators agreed on the binary category). We can see that the proportions of non-literal sentences are indeed considerably larger for PVs than for BVs (14.8% vs. 3.2% for full agreement, and 29.5% vs. 14.8% for majority agreement), thus indicating a stronger non-literal language potential for German PVs in comparison to their BVs. Contrary to our assumptions, the participants in the generation experiment also produced a large number of literal sentences for PVs. In our opinion this indicates (a) the ambiguity of German PVs, which led participants to refer to literal as well as non-literal senses; and (b) that the presumably strongly abstract target domain definitions did not necessarily enforce non-literal senses.

		literal		non-literal	
BVs	full	2,443	83.3%	94	3.2%
	maj	2,674	91.2%	259	8.8%
PVs	full	2,174	48.5%	666	14.8%
	maj	3,150	70.2%	1,337	29.5%

Table 5: (Non-)literal language usage in generated BV/PV sentences.

**Abstractness in BV/PV Sentences** As meaning shifts typically take place as a mapping from a source to a target domain, where the target domain is supposedly more abstract than the source domain, we expect our sentences in the target domains to be more abstract than those in the source domains. Figure 6 shows that this is the case:

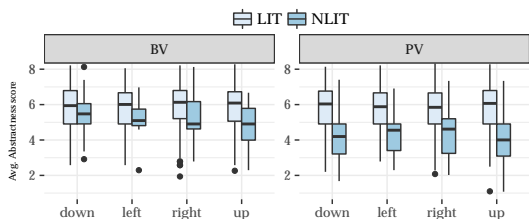


Figure 6: Average concreteness of nouns in BV/PV sentences, categorised by directionality.

Relying on abstractness/concreteness ratings of a semi-automatically created database (Köper and Schulte im Walde, 2016a), we looked up and averaged over the ratings of all nouns in a sentence.

The ratings range from 0 (very abstract) to 10 (very concrete). We can see that across directions the literal sentences are more concrete than the non-literal sentences. In addition, we can see that the differences in abstractness are much stronger for the PV target-domain sentences than for the BV source-domain sentences.

**Particle Meaning Shifts** Figure 7 once more illustrates preferences in arrow directions across the four particle types, but is –in contrast to Figure 5– restricted to the non-literal PV sentences (full agreement). For particles *ab* and *auf* we hardly find differences when specifying on non-literal language usage; for both *an* and *aus* we find an increase of DOWN meanings in non-literal language usage, which goes along with a decrease of LEFT meanings for *an* and a decrease of RIGHT meanings for *aus*. So within our collection we find some evidence for meaning shifts within PV types for the two particle types *an* and *aus* but not for *ab* and *auf*, which seem to stay with their predominant vertical meanings also in non-literal language.

↓	58.442	17.241	28.965	27.82
↑	7.143	73.399	25.517	13.534
→	25.974	4.434	26.897	53.383
←	8.442	4.926	18.621	5.263
	ab	auf	an	aus

Figure 7: Directionality of particle types restricted to non-literal sentences.

**Source–Target Domain Meaning Shifts** Figure 8 presents meaning shifts as strengths of relationships between source and target domains, when looking at only literal BV sentences and non-literal PV sentences. The cells in the heat map present the results of multiplying the target domain degrees of membership across all PVs with the source domain degrees of membership of their respective BVs. We applied positive pointwise mutual information (PPMI) weighting to avoid a bias towards popular classes. Examples of particularly strong combinations are PLANTS → TIME (e.g., *blühen* → *aufblühen*); and SOUND → COMMUNICATION (e.g., *bellen* → *anbellen*).



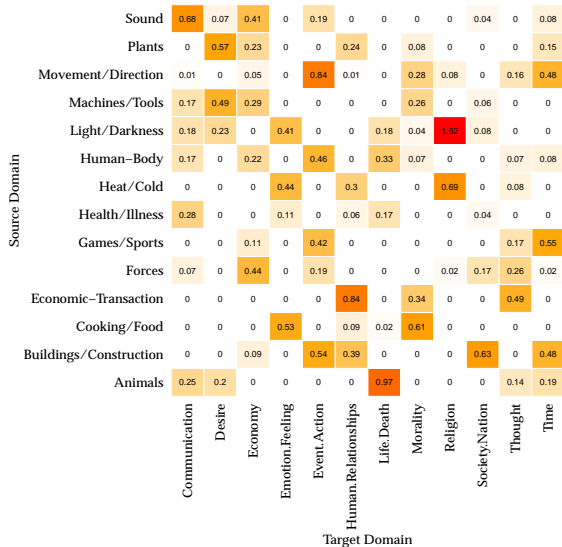


Figure 8: Source–target domain shifts.

### 4.3 Verification

While the previous section illustrated the value of the collection from a qualitative perspective, we also verified the information through computational approaches. We applied standard classifiers to predict source domains, target domains as well as directionality, given the underlying sentences. Our baseline is provided by *Majority*, which refers to the performance obtained by guessing always the largest class. For the target domains this majority provides a considerably high baseline with an accuracy of 33.95%, due to the very large class EVENT/ACTION. We therefore added a branch of experiments excluding this class (Target<sub>2</sub>).

As the most general set of features we used *Uni<sub>word</sub>*, a simple bag-of-words method where we counted how many times a certain unigram has been seen for a class. We implemented this method using Multinomial Naive Bayes. Similarly, we conducted experiments using *Uni<sub>lemma</sub>* instead of *Uni<sub>word</sub>*, which we expected to increase the chance of observing the unigram features.

*Affective* is a meaning-shift-related feature type. It relies on a range of psycholinguistic norms such as valency, arousal and concreteness/abstractness, which are supposedly salient features for meaning shifts and directions (Turney et al., 2011; Dudschig et al., 2015; Köper and Schulte im Walde, 2016b). We represented each sentence by providing an average affective score over all nouns, as taken from the semi-automatically created database by Köper and Schulte im Walde (2016a).

Finally we combined the above features (*Combination*). We relied on the affective norms, the lemma unigram features as well as the directionality information for domain prediction, or the domain information for directionality prediction.

Tables 6 and 7 present the accuracy results of classifying the generated sentences into domains and directionalities, respectively. According to the  $\chi^2$  test and  $p < 0.001$ , all our feature sets except for the affective norms in Table 7 outperform the baseline significantly, both individually and in combination. We thus conclude that also from a quantitative perspective the collection represents a valuable resource for complex verb meaning.

Feature Set	Method	Source	Target	Target <sub>2</sub>
Majority	Baseline	14.82	33.45	15.46
Affective	SVM	30.95	40.61	31.50
Uni <sub>word</sub>	Naive Bayes	54.15	43.40	42.60
Uni <sub>lemma</sub>	Naive Bayes	57.09	44.74	43.84
Combination	SVM	<b>60.74</b>	<b>49.87</b>	<b>45.46</b>

Table 6: Predicting domains.

Feature Set	Method	Source	Target
Majority	Baseline	34.09	31.74
Affective	SVM	40.93	35.63
Uni <sub>word</sub>	Naive Bayes	48.56	55.27
Uni <sub>lemma</sub>	Naive Bayes	<b>52.28</b>	<b>56.94</b>
Combination	SVM	49.18	55.93

Table 7: Predicting directionality.

## 5 Conclusion

We presented a new collection to assess meaning components in German complex verbs, by relying on a novel strategy to obtain source and target domain characterisations as well as spatial directional information via sentence generation rather than sentence annotation. A broad qualitative description of the dataset and a series of standard classification experiments assessed the reliability of the novel collection.

## Acknowledgments

The research was supported by the DFG Collaborative Research Centre SFB 732. In addition, we particularly thank our student researchers Ingrid Kasimir, Daniela Naumann and Florian Lux for their help in collecting and cleaning the dataset.

## References

- Timothy Baldwin, Colin Bannard, Takaaki Tanaka, and Dominic Widdows. 2003. An Empirical Model of Multiword Expression Decomposability. In *Proceedings of the ACL Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*. Sapporo, Japan, pages 89–96.
- Collin Bannard. 2005. Learning about the Meaning of Verb-Particle Constructions from Corpora. *Computer Speech and Language* 19:467–478.
- Lawrence W. Barsalou. 1999. Perceptual Symbol Systems. *Behavioral and Brain Sciences* 22:577–660.
- Julia Birke and Anoop 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*. Trento, Italy, pages 329–336.
- Stefan Bott and Sabine Schulte im Walde. 2014. Optimizing a Distributional Semantic Model for the Prediction of German Particle Verb Compositionality. In *Proceedings of the 9th International Conference on Language Resources and Evaluation*. Reykjavik, Iceland, pages 509–516.
- Stefan Bott and Sabine Schulte im Walde. 2015. Exploiting Fine-grained Syntactic Transfer Features to Predict the Compositionality of German Particle Verbs. In *Proceedings of the 11th Conference on Computational Semantics*. London, UK, pages 34–39.
- Stefan Bott and Sabine Schulte im Walde. 2017. Factoring Ambiguity out of the Prediction of Compositionality for German Multi-Word Expressions. In *Proceedings of the 13th Workshop on Multiword Expressions*. Valencia, Spain, pages 66–72.
- Paul Cook and Suzanne Stevenson. 2006. Classifying Particle Semantics in English Verb-Particle Constructions. In *Proceedings of the ACL/COLING Workshop on Multiword Expressions: Identifying and Exploiting Underlying Properties*. Sydney, Australia, pages 45–53.
- Scott Deerwester, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman. 1990. Indexing by Latent Semantic Analysis. *Journal of the American Society of Information Science* 41(6):391–407.
- René Dirven. 1993. Dividing up Physical and Mental Space into Conceptual Categories by Means of English Prepositions. In Zelinsky C. Wibbelt, editor, *The Semantics of Prepositions – From Mental Processing to Natural Language Processing*, Mouton de Gruyter, volume 3 of *Natural Language Processing*, pages 73–98.
- Carolin Dudschig, Irmgard de la Vega, and Barbara Kaup. 2015. What’s up? Emotion-specific Activation of Vertical Space during Language Processing. *Acta Psychologica* 156:143–155.
- Carolin Dudschig, Martin Lachmair, Irmgard de la Vega, Monica De Filippis, and Barbara Kaup. 2012. From Top to Bottom: Spatial Shifts of Attention caused by Linguistic Stimuli. *Cognitive Processes* 13:S151–S154.
- Diego Frassinelli, Alla Abrosimova, Sylvia Springorum, and Sabine Schulte im Walde. 2017. Meaning (Mis-)Match in the Directionality of German Particle Verbs. Poster at the 30th Annual CUNY Conference on Human Sentence Processing.
- Dedre Gentner. 1983. Structure-Mapping: A Theoretical Framework for Analogy. *Cognitive Science* 7:155–170.
- Arthur M. Glenberg and Michael P. Kaschak. 2002. Grounding Language in Action. *Psychonomic Bulletin and Review* 9(3):558–565.
- Boris Haselbach. 2011. Deconstructing the Meaning of the German Temporal Verb Particle “nach” at the Syntax-Semantics Interface. In *Proceedings of Generative Grammar in Geneva*. Geneva, Switzerland, pages 71–92.
- Julie Heiser and Barbara Tversky. 2006. Arrows in Comprehending and Producing Mechanical Diagrams. *Cognitive Science* 30:581–592.
- Anette Herskovits. 1986. *Language and Spatial Cognition: An Interdisciplinary Study of the Prepositions in English*. Studies in Natural Language Processing. Cambridge University Press, London.
- Hans Kamp and Uwe Reyle. 1993. *From Discourse to Logic*. Kluwer Academic Publishers.
- Barbara Kaup, Monica De Filippis, Martin Lachmair, Irmgard de la Vega, and Carolin Dudschig. 2012. When Up-Words meet Down-Sentences: Evidence for Word- or Sentence-based Compatibility Effects? *Cognitive Process* 13:S203–S207.
- Su Nam Kim and Timothy Baldwin. 2007. Detecting Compositionality of English Verb-Particle Constructions using Semantic Similarity. In *Proceedings of the 7th Meeting of the Pacific Association for Computational Linguistics*. Melbourne, Australia, pages 40–48.
- Fritz Kliche. 2011. Semantic Variants of German Particle Verbs with “ab”. *Leuvense Bijdragen* 97:3–27.
- Maximilian Köper and Sabine Schulte im Walde. 2016a. Automatically Generated Affective Norms of Abstractness, Arousal, Imageability and Valence for 350 000 German Lemmas. In *Proceedings of the 10th International Conference on Language Resources and Evaluation*. Portoroz, Slovenia, pages 2595–2598.
- Maximilian Köper and Sabine Schulte im Walde. 2016b. 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. San Diego, California, USA, pages 353–362.
- Maximilian Köper and Sabine Schulte im Walde. 2017a. Applying Multi-Sense Embeddings for German Verbs to Determine Semantic Relatedness and to Detect Non-Literal Language. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*. Valencia, Spain, pages 535–542.
- Maximilian Köper and Sabine Schulte im Walde. 2017b. Complex Verbs are Different: Exploring the Visual Modality in Multi-Modal Models to Predict Compositionality. In *Proceedings of the 13th Workshop on Multiword Expressions*. Valencia, Spain, pages 200–206.
- Maximilian Köper and Sabine Schulte im Walde. 2018. Analogies in Complex Verb Meaning Shifts: The Effect of Affect in Semantic Similarity Models. In *Proceedings of the 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. New Orleans, LA, USA. To appear.
- Maximilian Köper, Sabine Schulte im Walde, Max Kisselew, and Sebastian Padó. 2016. Improving Zero-Shot-Learning for German Particle Verbs by using Training-Space Restrictions and Local Scaling. In *Proceedings of the 5th Joint Conference on Lexical and Computational Semantics*. Berlin, Germany, pages 91–96.
- Zolzan Kövecses. 2002. *Metaphor: A Practical Introduction*. Oxford University Press, New York.
- Natalie Kühner and Sabine Schulte im Walde. 2010. Determining the Degree of Compositionality of German Particle Verbs by Clustering Approaches. In *Proceedings of the 10th Conference on Natural Language Processing*. Saarbrücken, Germany, pages 47–56.
- George Lakoff, Jane Espenson, and Alan Schwartz. 1991. Master Metaphor List. Technical Report.
- George Lakoff and Mark Johnson. 1980. *Metaphors we live by*. University of Chicago Press.
- Andrea Lechler and Antje Roßdeutscher. 2009. German Particle Verbs with "auf". Reconstructing their Composition in a DRT-based Framework. *Linguistische Berichte* 220:439–478.
- Birte Lönneker-Rodman. 2008. The Hamburg Metaphor Database Project: Issues in Resource Creation. *Language Resources and Evaluation* 42:293–318.
- Diana McCarthy, Bill Keller, and John Carroll. 2003. Detecting a Continuum of Compositionality in Phrasal Verbs. In *Proceedings of the ACL Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*. Sapporo, Japan, pages 73–80.
- Daniel C. Richardson, Michael J. Spivey, Lawrence W. Barsalou, and Ken McRae. 2003. Spatial Representations activated during Real-Time Comprehension of Verbs. *Cognitive Science* 27:767–780.
- Ivan A. Sag, Timothy Baldwin, Francis Bond, Ann Copestake, and Dan Flickinger. 2002. Multiword Expressions: A Pain in the Neck for NLP. In *Proceedings of the Conference on Intelligent Text Processing and Computational Linguistics*. Mexico City, Mexico.
- Bahar Salehi and Paul Cook. 2013. Predicting the Compositionality of Multiword Expressions Using Translations in Multiple Languages. In *Proceedings of the 2nd Joint Conference on Lexical and Computational Semantics*. Atlanta, GA, USA, pages 266–275.
- Bahar Salehi, Paul Cook, and Timothy Baldwin. 2014. Using Distributional Similarity of Multi-way Translations to Predict Multiword Expression Compositionality. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*. Gothenburg, Sweden, pages 472–481.
- Larry Shapiro. 2007. The Embodied Cognition Research Programme. *Philosophy Compass* 2(2):338–346.
- Ekaterina Shutova and Simone Teufel. 2010. Metaphor Corpus Annotated for Source – Target Domain Mappings. In *Proceedings of the 7th International Conference on Language Resources and Evaluation*. Valletta, Malta, pages 3255–3261.
- Ekaterina Shutova, Simone Teufel, and Anna Korhonen. 2013. Statistical Metaphor Processing. *Computational Linguistics* 39(2):301–353.
- Sylvia Springorum. 2011. DRT-based Analysis of the German Verb Particle "an". *Leuvense Bijdragen* 97:80–105.
- Sylvia Springorum, Jason Utt, and Sabine Schulte im Walde. 2013. Regular Meaning Shifts in German Particle Verbs: A Case Study. In *Proceedings of the 10th International Conference on Computational Semantics*. Potsdam, Germany, pages 228–239.
- Peter Turney, Yair Neuman, Dan Assaf, and Yohai Cohen. 2011. Literal and Metaphorical Sense Identification through Concrete and Abstract Context. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Edinburgh, UK, pages 680–690.
- Barbara Tversky. 2011. Visualizing Thought. *Topics in Cognitive Science* 3:499–535.
- Ake Viberg. 1983. The Verbs of Perception: A Typological Study. *Linguistics* 21(1):123–162.
- Joost Zwarts. 2017. Spatial semantics: Modeling the meaning of prepositions. *Language and Linguistics Compass* 11(5):1–20.