

# Sentence Generation and Compositionality of Systematic Neologisms of German Particle Verbs

Sylvia Springorum   Sabine Schulte im Walde   Antje Roßdeutscher

Institut für Maschinelle Sprachverarbeitung

Universität Stuttgart

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

**Compositionality of German particle verbs** German particle verbs (PVs) are highly productive combinations of a base verb and a prefix particle. Concerning their semantics, there is an ongoing discussion whether the meaning of German particle verbs is in general compositional or not. For example, Kratzer (2003) claimed that German PVs are idiosyncratic; this stands in opposition to the semantic analyses by Lechler and Roßdeutscher (2009), Kliche (2011), among others. who demonstrated that each particle has several different readings which however form regular patterns depending on the contexts. Our position is in-between: we agree that not every PV composition is transparent, but with a fine-grained sub-lexical analysis and taking analogy and meaning shift mechanisms into account, the majority of combinations can be explained by patterns.

Our research focuses on how speakers of German combine particle senses with base verb senses. Questions which come along with this focus are: (i) how applicable, (ii) how available and (iii) how common or prototypical is a semantic pattern of a meaning composition?

**Goal of this study** This study presents preliminary insights into an ongoing experiment for German PVs, where the participants generate sentences with attested PVs and also with yet not attested formations which we call systematic Neologisms of German Particle Verbs (neoPV). A PV is a neoPV if it is not listed in the Duden dictionary <sup>1</sup> and if it is not attested in the German web corpus *SdeWac* (Faaß *et al.*, 2010). The main assumption for the experiment is that if PVs are compositional and productive, neoPVs should have at least one understandable meaning. If neoPVs are given an interpretation by way of generating sentences with them, the idea of a rule based interpretation is hard to deny. In the following, we first describe the experiment to collect the neoPV data, and then perform a quantitative and qualitative description of the preliminary results, over all data and focusing on specific subsets.

**Experiment** The experiment is running with Amazon Mechanical Turk (AMT). The subjects are presented with a PV and two tasks: first, they are asked to provide a rating of 0-3 whether the PV is known or unknown or how familiar they are with it. Then, they have to generate at least one sentence using the PV, such that the sentences illustrate the verb meaning. After the generation, the subjects have the opportunity to mark a checkbox, if they feel it was difficult to generate a sentence for the particular PV.

The data comprise a total of 125 PVs: Five different particles (*ab*, *an*, *auf*, *aus*, *nach*) were combined with verbs from five different semantic verb classes: (1) DE-ADJECTIVAL e.g. *kürzen* 'shorten', (2) ACHIEVMENT/ACCOMPLISHMENT e.g. *finden* 'find', (3) PHYSICAL PROCESS e.g. *stricken* 'knit', (4) MENTAL PROCESS e.g. *denken* 'think', and (5) STATE e.g. *lieben* 'love'. The chosen base verbs (BVs) were balanced for corpus frequencies in the *SdeWac*.

---

<sup>1</sup>The PVs were looked up in the online version of the dictionary: [www.Duden.de](http://www.Duden.de)

**Results** Table 1 shows for each BV class (column 1) the amount of so far generated sentences (column 2), the familiarity of a PV (column 3), the difficulty, which refers to the previously mentioned checkbox option (column 4), the percentage of these 'difficult' cases of the unknown PVs and the amount of neoPVs in the presented data (last column).

So far, we collected 1,470 sentences, out of which 863 contain a neoPV unknown to the subject. In 230 of these unknown cases (26.65%), the subject claimed that it was difficult to think of a sentence. Comparing across the semantic BV classes, the unknown PHYSICAL PROCESS verbs seem to be easier to handle than the other BVs (only 20.71% were difficult). This finding fits with our expectation, that it should be easier to apply a particle meaning to verbs with a homogeneous event structure than to verbs either coming with a result or a state by themselves, because particles often contribute result and state to BVs.

BV Class	Sentences	Unknown	Difficult	% of Diff.	neoPV
All	1470	863	230	26.65	81
DE-ADJECTIVAL	303	199	57	28.64	19
ACHIEV./ACCOMP.	310	134	37	27.61	12
PHYSICAL PROCESS	305	140	29	20.71	15
MENTAL PROCESS	301	165	48	29.09	16
STATE	251	225	59	26.22	22

Table 1: Quantification of current results.

**Results for the BV *stricken*** Taking a closer look at the PHYSICAL PROCESS verb *stricken* (Table 2), we have collected 49 sentences so far, with 26 PV ratings as 'unknown' and 5 ratings as 'difficult'. The attested verbs with this BV are *aufstricken* and *anstricken*, and the neoPVs are *abstricken*, *ausstricken* and *nachstricken*. Surprisingly, the attested verb *anstricken* was judged as unknown in 5 of the 9 sentences and the also attested *aufstricken* was even judged as unknown in 9 of 11 sentences. On the contrary, we also find the reverse case, where for the neoPV *abstricken* only 3 of the 8 sentences were marked as 'unknown'.

Verbs	Sentences	Unknown	Difficult	neoPV
All	49	26	5	3
<i>ab-</i>	8	3	1	+
<i>an-</i>	9	5	2	-
<i>auf-</i>	11	9	1	-
<i>aus-</i>	10	8	1	+
<i>nach-</i>	11	3	-	+

Table 2: Quantification of results for the BV *stricken*.

**Results for the BV *abstricken*** Table 3 shows the distribution of *ab* readings in sentences with *abstricken* in relation to whether the PV was known or unknown. For the known verbs, in one case *ab* was given the COPY reading as in *Ein Bild abmalen* 'to copy a picture', in the other case it is not clear which reading was used. The COPY reading also occurred for one unknown case. We also found sentences, where *ab* has the END OF SUPPORT reading, implying an end of a contact relation as in *Knopf von einer Hose abreißen* 'to rip off the button of a trouser', a QUANTIFICATION reading as in *die Aufgaben abarbeiten* 'to complete a task step by step' and a TERMINATION reading as in *das Baby abstillen* 'to wean the baby'. The TERMINATION reading was used 3 times, and example 1 is one of the sentences. The adjective *letzten* 'last' in this case shows that there must be a semantics which terminates the *stricken* event.

- (1) *Die letzten Maschen müssen abgestrickt werden. Das ist nicht schwer.*

Reading	Known	Unknown	Difficult
END OF SUPPORT		1	1
COPY	1	1	
QUANTIFICATION		1	
TERMINATION	3		
Undef	1		

Table 3: Distribution of particle readings for *abstricken*.

'The last stitches have to be cast off. This is not difficult'

**Results for PVs with particle *ab*** Table 4 shows the distribution of the readings of the particle *ab* over all 5 PHYSICAL PROCESS BVs. The most common readings are the END OF SUPPORT and the QUANTIFICATION reading occurring in around 10 sentences each. So it seems that these two readings tend to be more prototypical than the others.

Another very interesting point is the metaphorical use of a neoPV which occurred several times with different particles and BVs. This means in general, that people are not only able to compose the meaning of a PV but also to embed the resulting concept in another domain and in an understandable way. The PVs in the Sentences 2 is not literal. In 2 we have the abstract object *Arbeitsstelle* 'job' which was mentally attached to some future life plans, but since the job interview did not go well, the job has to be mentally detached. This was expressed by the PV *abnageln* '[ab] + to nail', an abstract END OF SUPPORT *ab* and an abstract interpretation of *nageln* 'to nail'. We see the modal context together with the dative as an evidence, that the construction is analogous to the existing metaphorical reading of *sich etwas abschminken können/müssen* 'to get something out the head/literally: to be able to remove make up' in example 3.

- (2) *Das Vorstellungsgespraech lieft gar nicht gut, die neue Arbeitsstelle kann ich mir wohl abnageln.*  
'The job interview didn't go well, I have to get the new job out of my head.'
- (3) *Wenn du weiter so verschwederisch lebst, kannst du dir die Reise abschminken.*  
'If you keep on living lavishly then you can get the travel out of your head.'

Reading	<i>abnageln</i>	<i>abstricken</i>	<i>abrühren</i>	<i>abschaukeln</i>	<i>abschlafen</i>	Sum
END OF SUP.	5	1	1	1		8
QUANT.	3	1		1	5	10
COVER	1					1
TERMINATION		3		1		4
COPY		2				2
USE UP				1	1	2
MIX			4			4
METAPHOR	1		1	2		4
Total	15	8	9	7	10	

Table 4: Distribution of readings for *ab* + PHYSICAL PROCESS BVs.

Even though the experiment is not finished yet we showed that (i) there are several interpretations of the PVs and the particles and some of them seem to be more difficult (ii) that not all readings have to be available to everyone, cf. example with *anstricken*, which is attested, but was rated in more than half of the sentences as unknown and (iii) we have PV readings which were used more often than others (cf. the QUANTIFICATION reading). We also found metaphorical and therefore non-prototypical PV usages, like in the case of *abnageln*.

## References

- Faaß G., Heid, U., and Schmid, H. (2010). Design and Application of a Gold Standard for Morphological Analysis: SMOR in Validation. In *Proceedings of the 7th International Conference on Language Resources and Evaluation*, pages 803–810, Valletta, Malta.
- Kliche, F. (2011). Semantic Variants of German Particle Verbs with *ab*. *Leuvense Bijdragen*, (97).
- Kratzer, A. (2003). *The Event Argument and the Semantics of Verbs*. <http://semanticsarchive.net/Archive/GU1NWM4Z/>.
- Lechler, A. and Roßdeutscher, A. (2009). Analysing German Verb-Particle Construction with *auf* in a DRT-based Framework. Technical Report 4, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.