

# Identification, Quantitative Description, and Preliminary Distributional Analysis of German Particle Verbs



Sabine Schulte im Walde  
Computational Linguistics  
Saarland University, Saarbrücken, Germany  
schulte@coli.uni-sb.de

## Characteristics - A Challenge for SNLP

**Morphology** affix preposition + base verb

*an* + *fangen*  
*ein* + *führen*  
*vor* + *stellen*

**Syntax** separability of particle verbs

V-first: *Fängt* Kai mit den Hausaufgaben *an*?  
*Stellt* Anna sich ihre Zukunft *vor*?

V-second: Kai *fängt* mit den Hausaufgaben *an*.  
Anna *stellt* sich ihre Zukunft *vor*.

V-final: ..., weil Kai mit den Hausaufgaben *anfängt*.  
..., dass Anna sich ihre Zukunft *vorstellt*.

## Semantics

transparency  $\longleftrightarrow$  opacity

*aufstellen* \ *stellen* – put (up)    *einfallen* \ *fallen* – cross one's mind, invade (in)    *anfangen* – begin  
*stellen* – put (up)    *fallen* – fall (in)    *fangen* – catch

selectional restrictions (examples cf. Lüdeling, 2001)

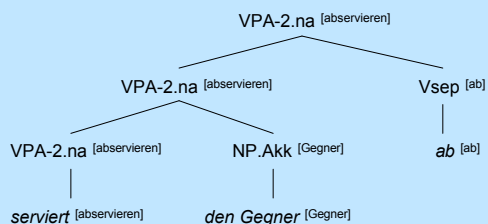
\*[NP<sub>nom</sub> Der Prinz] *stellt* [NP<sub>acc</sub> die Flasche].  
[NP<sub>nom</sub> Der Prinz] *stellt* [NP<sub>acc</sub> die Flasche] [PP<sub>dir</sub> auf den Tisch].  
[NP<sub>nom</sub> Der Prinz] *stellt* [NP<sub>acc</sub> die Flasche] *ab*.

## Identification

### Statistical Grammar Model

- head-lexicalised probabilistic context-free grammar for German (Charniak, 1997; Schulte im Walde, 2003)
- left-corner parser *LoPar* (Schmid, 2000)
- 35 million words of German newspaper corpus

**Strategy** statistical learning of particle verbs and their behaviour



## Quantitative Lexical Description for Particle and Base Verbs

Subcategorisation Frames prominent: n, na (plus adjuncts)

<i>ankommen</i> (freq: 1,831)				<i>kommen</i> (freq: 43,270)			
n	38.82	np:Dat.bei	1.50	np	34.18	np:Dat.zu	11.21
x	16.12	nap:Dat.in	1.40	n	28.78	np:Akk.in	5.35
na	10.56	np:Akk.auf	1.02	na	8.05	xp:Dat.zu	5.33
ns-w	5.76			x	5.65	np:Dat.aus	4.77
ns-2	4.63			xp	5.53	np:Dat.von	2.67
<i>einsetzen</i> (freq: 3,390)				<i>setzen</i> (freq: 7,545)			
na	40.29	nap:Dat.in	4.33	na	25.87	np:Akk.auf	13.66
nap	16.44	nap:Dat.bei	2.81	nap	19.09	nap:Dat.in	5.03
nr	15.87	npr:Akk.für	2.63	np	16.50	nap:Akk.auf	2.80
n	10.86	nap:Dat.zu	1.76	nad	12.28	nap:Dat.mit	1.86
nad	4.71	nap:Akk.für	1.35	n	10.13	nap:Akk.in	1.66
<i>umbringen</i> (freq: 683)				<i>bringen</i> (freq: 12,249)			
na	53.60	nap:Dat.in	5.43	na	42.71	nap:Akk.in	9.58
nr	19.36	nap:Dat.nach	1.54	nap	31.65	nap:Dat.zu	4.86
nap	12.23	npr:Dat.in	1.23	nad	11.47	nap:Akk.auf	4.44
nad	3.20			n	3.21	nap:Dat.mit	4.17
nas-2	1.97			nd	2.05	nap:Dat.in	3.41

frame types: nominative (n), accusative (a) and dative (d) noun phrases, reflexive pronouns (r), prepositional phrases (p), expletive es (x), non-finite clauses (i), and finite clauses (s-2 for verb second, s-w for indirect *wh*-questions, s-dass, s-ob)

## References

- Timothy Baldwin and Aline Villavicencio, 2002. Extracting the Unextractable: A Case Study on Verb Particles. In *Proceedings of the Sixth Conference on CoNLL*.
- Erich and Hildegard Bullitta, 2003. *Wörterbuch der Synonyme und Antonyme*. Fischer Taschenbuch Verlag.
- Eugene Charniak, 1997. *Statistical Parsing with a Context-Free Grammar and Word Statistics*. In *Proceedings of the 14th National Conference on Artificial Intelligence*.
- Nicole Dehé, Ray Jackendoff, Andrew McIntyre, and Silke Urban, 2002. *Verb-Particle Explorations*. Mouton de Gruyter.
- Anke Lüdeling, 2001. *On German Particle Verbs and Similar Constructions in German*. Dissertations in Linguistics. CSLI Publications.
- Helmut Schmid, 2003. *LoPar: Design and Implementation*. Arbeitspapiere des Sonderforschungsbereichs 340, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.
- Sabine Schulte im Walde, 2003. *Experiments on the Automatic Induction of German Semantic Verb Classes*. PhD thesis, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.

## Selectional Noun Preferences

<i>Botschaft</i>	'message'	17.17	10.22
<i>Zug</i>	'train'	11.14	14.78
<i>Flüchtling</i>	'refugee'	7.13	28.07
<i>Film</i>	'movie'	5.18	13.32
<i>Spende</i>	'donation'	5.00	6.61
<i>Brief</i>	'letter'	4.72	22.18
<i>Buch</i>	'book'	4.60	13.78

transparent particle verb:  
*ankommen* vs. *kommen*  
→ n

opaque particle verb:  
*umbringen* vs. *bringen*  
→ na

<i>Mensch</i>	'human'	13.63	<i>Ergebnis</i>	'result'	111.86
<i>Frau</i>	'woman'	10.24	<i>Erfolg</i>	'success'	73.39
<i>Kind</i>	'child'	9.54	<i>Geld</i>	'money'	54.29
<i>Mann</i>	'man'	7.46	<i>Problem</i>	'problem'	52.44
<i>Vater</i>	'father'	5.97	<i>Vorteil</i>	'advantage'	47.58
<i>Million</i>	'million'	4.91	<i>Opfer</i>	'victim'	44.34
<i>Leute</i>	'people'	4.87	<i>Entscheidung</i>	'decision'	39.92

## Semantic Similarity

### Distributional Description and Similarity

- association of German verbs with frequency distributions
- 38 subcategorisation frame types excluding pp-specification
- nominal fillers for n, nr, na, nad, nar
- cosine distance compares particle verbs, base verbs, synonyms

### Semantic Class Examples

	<i>ankommen</i>		<i>einsetzen</i>	
1.	<i>kommen</i>	0.50	<i>benutzen</i>	0.37
2.	<i>erscheinen</i>	0.42	<i>verwenden</i>	0.36
3.	<i>daherkommen</i>	0.41	<i>engagieren</i>	0.32
4.	<i>anrollen</i>	0.40	<i>handeln</i>	0.31
5.	<i>herkommen</i>	0.38	<i>anwenden</i>	0.27
6.	<i>einlaufen</i>	0.24	<i>schicken</i>	0.26
7.	<i>landen</i>	0.22	<i>eintreten</i>	0.25
8.	<i>nähern</i>	0.21	<i>einschalten</i>	0.24
...				
21.			<i>setzen</i>	0.10