

Identifying Semantic Relations and Functional Properties of Human Verb Associations

Sabine Schulte im Walde
(joint work with Alissa Melinger)

Computational Linguistics
Saarland University, Saarbrücken

*Universitat Pompeu Fabra
May 23, 2005*

Motivation

Basis: Human verb associations

Task: Characterisation of verb and noun responses

- NLP tasks vary in usage of semantic relations, e.g. thesaurus extraction, summarisation
⇒ Identification of semantic relations between verbs
- Data-intensive lexical semantics:
words \Leftrightarrow distributional vectors,
relatedness of words \Leftrightarrow vector similarity
⇒ Identification of nominal features to describe verbs

Experiment Material

- 330 German verbs
- Variety of semantic verb classes, possible ambiguity:
 - » **self-motion**: *gehen* 'walk', *schwimmen* 'swim'
 - » **cause**: *verbrennen* 'burn', *reduzieren* 'reduce'
 - » **experiencing**: *lachen* 'laugh', *überraschen* 'surprise'
 - » **communication**: *erzählen* 'tell', *klagen* 'complain'
 - » **body**: *schlafen* 'sleep', *abnehmen* 'lose weight'
- Variety of frequency ranges (1 < freq < 71,604)
- Random distribution: 6 data sets à 55 verbs,
balanced for class affiliation and frequency ranges

Experiment Procedure

- Web experiment over Internet
- Bibliographic information:
linguistic experience, age, regional accent, profession
- Instructions and example page
- Experiment page for each verb
- 30 sec. for each verb; 2 sec. break; total: ca. 30 min.
- Association input:
spontaneous, exhaustive, one word per line, capitalisation

schneien

kalt

rodeln

Schneemann

weiß

dämmern

Experiment Participants and Data

- 299 accepted data files:
native German speakers, 80% of target verbs
- Expertise of participants: 166 experts vs. 132 non-experts
- Participants per data set: **between 44 and 54**
- Number of trials: 16,445
- Number of associations per target verb:
range 0-16, average: 5.16
- All associations: **81,373 tokens with 18,884 types**
(first) **15,780 tokens with 4,856 types**

Data Preparation

1. Lexicon look-up
2. (Semi-automatic)
Data correction
3. Quantification
over responses

<i>klagen</i> 'complain, moan'		
Gericht	'court'	19
jammern	'moan'	18
weinen	'cry'	13
Anwalt	'lawyer'	11
Richter	'judge'	9
Klage	'complaint'	7
Leid	'suffering'	6
Trauer	'mourning'	6
Klagemauer	'Wailing Wall'	5
laut	'noisy'	5

Linguistic Analysis of Experiment Data

- Preference for **morpho-syntactic category** of responses?
→ distinguish major parts-of-speech:
nouns, verbs, adjectives, adverbs
- Encoding of particular **semantic relations**?
→ look up relation between **target and response verb**:
GermaNet (Kunze, 2000/2004)
- Typical **argument holders** of verb valency?
→ investigate **linguistic functions realised by nouns**:
empirical grammar model (Schulte im Walde, 2003)

Statistical Grammar Model

- Lexicalised probabilistic context-free grammar (Charniak, 1995; Carroll and Rooth, 1998)
- 35 million words of German newspaper corpora
- Unsupervised training by *EM-Algorithm* (Baum, 1972)
- Robust statistical parser *LoPar* (Schmid, 2000)
- Corpus-based quantitative lexical information: word frequencies, linguistic functions, head-head relations

Morpho-Syntactic Analysis of Responses

- Source: machine-readable quantitative dictionary
- Dictionary information:
word forms, part-of-speech tags, lemmas, frequencies
- Ambiguous part-of-speech tags;
examples: *Vergnügen* `please/pleasure' (V/N)
überlegen `think about/superior' (V/ADJ)
- Result: distinction and quantification of morpho-syntactic categories of responses

Morpho-Syntactic Distribution

	V	N	ADJ	ADV	
Freq	19.863	48.905	8.510	1.268	TOKEN
Prob	25	62	11	2	all assoc
Freq	5.355	8.838	1.178	199	TOKEN
Prob	34	57	8	1	first assoc

Morpho-Syntactic Distribution Examples

	V	N	ADJ	ADV
Total Prob	25	62	11	2
<i>aufhören</i> 'stop'	49	39	4	6
<i>aufregen</i> 'be upset'	22	54	21	0
<i>backen</i> 'bake'	7	86	6	1
<i>bemerken</i> 'realise'	52	31	12	2
<i>dünken</i> 'seem'	46	30	18	1
<i>flüstern</i> 'whisper'	19	43	37	0
<i>nehmen</i> 'take'	60	31	3	2
<i>radeln</i> 'bike'	8	84	6	2
<i>schreiben</i> 'write'	14	81	4	1

Semantic Relations between Verbs

- Semantic relations between **target and response verbs**
- Source: lexical semantic taxonomy **GermaNet** (GWN)
- Organisation of verbs, nouns, adjectives, adverbs
- Classes of synonyms: synsets
- Ambiguous words: assignment to multiple classes
- Lexical and conceptual relations between synsets:
antonymy, hypernymy, entailment, cause, etc.

Semantic Relations between Verbs

- Synonymy: target and response verb in common synset
- Other semantic relations:
 - look up GermaNet semantic relations between
 - » target verb synsets
 - » response verb synsets
- Quantification of target-response relation:
association frequency
- No relation: target and response verb in GWN, no relation
- Unknown relation: response verb not in GWN

Semantic Relations: Distributions

	SYN	ANT	HYPER	HYPO	CAUSE	ENTAIL	ALSO	NONE	?	
Freq	1,194	252	2,807	3,016	49	0	0	10,509	1,726	TO
Prob	6	1	14	16	0	0	0	54	9	all
Freq	622	110	1,281	1,023	11	0	0	2,184	314	TO
Prob	11	2	23	19	0	0	0	39	6	first

Semantic Relations: Examples

- **Synonymy:** *heulen* 'cry' - *weinen* 'cry' (43)
laufen 'run' - *rennen* 'run' (30)
sinken 'sink' - *untergehen* 'sink' (10)
- **Antonymy:** *auftauen* 'defreeze' - *einfrieren* 'freeze' (7)
emigrieren 'emigrate' - *immigrieren* 'immigrate' (7)
zunehmen 'gain weight' - *abnehmen* 'lose weight' (15)
- **Hypernymy:** *begeistern* 'be enthusiastic' - *freuen* 'be glad' (10)
geschehen 'occur' - *passieren* 'happen' (30)
umbringen 'kill' - *töten* 'kill' (43)
- **Hyponymy:** *achten* 'respect' - *respektieren* 'respect' (10)
wahrnehmen 'perceive' - *sehen* 'see' (35)
wenden 'turn' - *umdrehen* 'turn around' (21)
- **Cause:** *bekommen* 'receive' - *haben* 'have' (6)
legen 'lay' - *liegen* 'lie' (8)
zeigen 'show' - *sehen* 'see' (5)

No Semantic Relation in GermaNet

- *abstürzen* 'crash'
fallen 'fall' (12), *klettern* 'climb' (2)
- *schwitzen* 'sweat'
stinken 'stink' (8), *laufen* 'run' (5), *frieren* 'be cold' (2)
- *adressieren* 'address'
schreiben 'write' (15), *schicken* 'send' (6)
- *backen* 'bake'
essen 'eat' (6), *schmecken* 'taste' (2), *kneten* 'knead' (2)

Window Approach for Semantic Relations

- Corpus data from statistical grammar model
- Window (left+right): 5/20/50 words, excluding symbols
- Basis: association **tokens**

<i>window</i>	<i>pos (37%)</i>	<i>neg (63%)</i>	<i>all</i>
5	57	29	39
20	75	46	57
50	82	54	64

Window Approach for GWN Relations

Relation	window: 5		window: 20		window: 50	
Synonymy	1,194	652 55%	923 77%	1,018	85%	
Antonymy	252	200 79%	209 83%	213	85%	
Hypernymy	2,807	1,564 56%	2,103 75%	2,254	80%	
Hyponymy	3,016	1,695 56%	2,232 74%	2,463	82%	
Cause	49	39 80%	41 84%	42	86%	
Entailment	0	0 0%	0 0%	0	0%	
Also see	0	0 0%	0 0%	0	0%	

Missing Semantic Relations

GermaNet: NO / Window: YES

» tagging:

auftauen 'defreeze' - *wärme* 'warmth'
erhitzen 'heat' - *topf* 'pot'

» missing links in GermaNet:

fahren 'drive' - *webfahren* 'drive away'
erhitzen 'heat' - *abkühlen* 'cool'

» non-classical semantic relation/scene information:

fliegen 'fly' - *starten* 'start'
erhitzen 'heat' - *essen* 'eat'
beenden 'stop' - *abgeben* 'hand in'

Missing Semantic Relations

GermaNet: NO / Window: NO

» domain:

radeln 'bike' - *strampeln* 'pedal', 'struggle' (1)
paddeln 'paddel' - *rudern* 'row' (22)

» non-classical semantic relation/scene information:

aufhören 'finish' - *stoppen* 'stop' (19)
initiiieren 'initiate' - *anfangen* 'start' (21)
auftauen 'defrost' - *essen* 'eat' (8)
paddeln 'paddle' - *schwimmen* 'swim' (7)

Semantic Relations: Summary

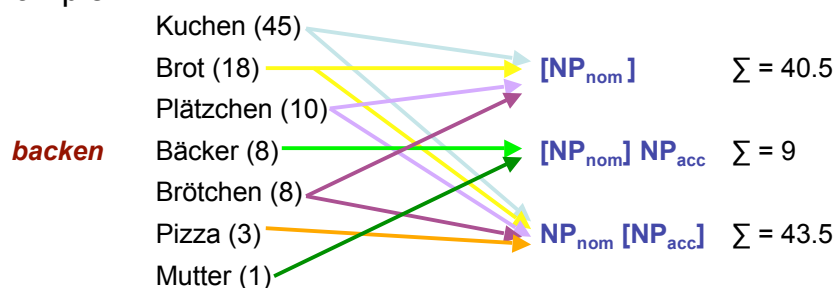
- Characterisation of verb-verb associations
- Proportion and distinction of 'classical' verb relations
- Distinction of verb-verb co-occurrences in corpus window
- Detection of missing links in GermaNet
- Non-classical verb relations as major part of associations
- Associations as basis for defining non-classical relations

Syntax-Semantic Functions of Nouns

- Source: statistical grammar model
- Verb valency:
 - » 38 syntactic subcategorisation frames
 - » plus PP information (case+preposition) → 178 frames
 - » subcategorised nouns
- Example: *backen* 'bake'
 - » frames: NP_{nom}
 $NP_{nom} NP_{acc}$...
 - » filler examples for NP_{nom} [NP_{acc}]: *Brötchen* 'rolls'
Brot 'bread'
Kuchen 'cake' ...

Syntax-Semantic Functions: Analysis

- Typical conceptual roles which speakers have in mind
- Look up syntactic relationships between verb and nouns
- Example:



Syntax-Semantic Functions: Distributions

Function		TOKEN (all)		TOKEN (first)	
S	S V	1,793	4	482	5
	S V AO	1,065	2	252	3
	S V DO	330	1	119	1
	S V AO DO	344	1	83	1
	S V PP	510	1	108	1
	S V Clause _{FIN}	178	0	60	1
	S V Clause _{INF}	204	0	59	1
	AO	S V AO	2,298	5	520
S V AO DO		882	2	242	3
S V AO PP		706	1	190	2
S V AO Refl		171	0	45	1
DO	S V DO	302	1	55	1
	S V AO DO	597	1	151	2
PP	S V PP:in _{Dat}	418	1	123	1
	S V PP:auf _{Dat}	180	0	44	1
Unknown noun		10,663	22	1,477	17
Unknown function		24,536	50	3,974	45

Sabine Schulte im Walde - UPF, May 2005

25

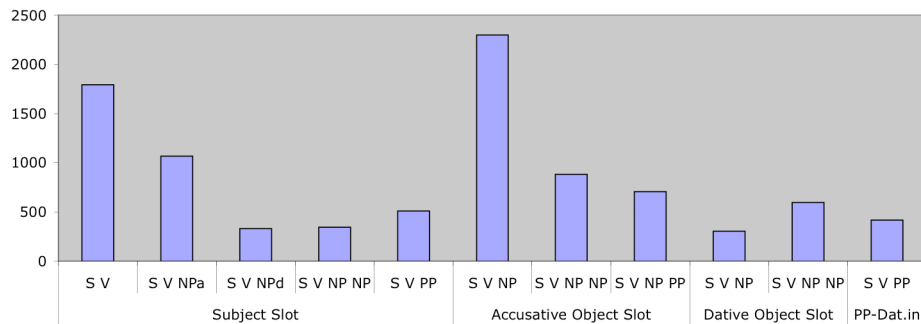
Syntax-Semantic Functions: Examples

- *ankommen* 'arrive' — S V — *Zug* 'train' (13)
- *adressieren* 'address' — S V AO PP — *Brief* 'letter' (41)
- *anvertrauen* 'confide' — S V AO DO — *Geheimnis* 'secret' (26)
- *anvertrauen* 'confide' — S V AO DO — *Freund* 'friend' (8)
- *gleiten* 'slide' — S V PP:durch_{Akk} — *Luft* 'air' (6)
- *knattern* 'crackle' — S V — *Motor* 'motor' (13)
- *schließen* 'close' — S V AO — *Tür* 'door' (38)
- *trocknen* 'dry' — S V AO — *Wäsche* (32)
- *lehren* 'teach' — S V PP:in_{Dat} — *Schule* 'school' (34)
- *kochen* 'cook' — S V PP:auf_{Dat} — *Herd* 'cooker' (14)
- *schreiben* 'write' — S V PP:mit_{Dat} — *Stift* 'pen' (26)
- *dienen* 'serve' — S V DO — *König* 'king' (4)

Sabine Schulte im Walde - UPF, May 2005

26

Syntax-Semantic Frame Inspection



No Linguistic Function in Grammar

50%

- *backen* 'bake'
Ofen 'oven' (19), *Mehl* 'flour' (17), *Weihnachten* 'Xmas' (15)
- *fliegen* 'fly'
Urlaub 'vacation' (11), *Flügel* 'wings' (9)
- *anfangen* 'begin'
Start 'start' (14), *Motivation* 'motivation' (3)
- *enden* 'end'
Feierabend 'week-day evening' (4), *Rente* 'retirement' (2)

Window Approach for Linguistic Functions

- Corpus data from statistical grammar model
- Window (left+right): 5/20/50 words, excluding symbols
- Basis: association **tokens**

<i>window</i>	<i>pos (28%)</i>	<i>neg (72%)</i>	<i>all</i>
5	87	19	39
20	95	37	54
50	97	47	62

Missing Linguistic Functions

Grammar function: NO / Window: YES

» lemmatisation and tagging:

composita: Übermacht, Zeitspanne, Autorennen;
erhitzen 'heat' - *Diskussion* 'discussion'

» argument vs. adjunct:

enden 'end' - *Belohnung* 'reward'
erhitzen 'heat' - *Pfanne* 'pan'

» scene information (in a different clause?):

fahren 'drive' - *Stau* 'traffic-jam'
beginnen 'begin' - *Erfahrung* 'experience'
trocknen 'dry' - *Anstrengung* 'effort'

Missing Linguistic Functions

Grammar function: NO / Window: NO

» domain:

radeln 'bike' - *Oma* 'grand-mom' (1)

stoppen 'stop' - *Plosiv* 'plosive' (1)

» scene information/world knowledge:

trocknen 'dry' - *Trockner* 'dryer' (11)

rudern 'row' - *Kraft* 'strength' (6)

radeln 'bike' - *Sonne* 'sun' (8)

auftauen 'defrost' - *Wasser* 'water' (14)

Linguistic Functions: Summary

- Characterisation of verb-noun associations
- Properties of nouns represent **conceptual roles** of verbs
- **Scene information** in addition to subcategorisation
- **Co-occurrence counts** to supplement argument counts
- Usage of roles for **distributional verb descriptions**

Outlook

- Characterisation of **non-classical** semantic verb relations
- Use verb-noun knowledge:
empirical verb descriptions to induce verb relations
- Variations of **verb feature descriptions** to find dependencies between verb descriptions and relations
- **Application** of verb relations and verb descriptions
- **Properties of verb classes**: semantic classes, experiencer classes, unergative-unaccusative distinction