

Giving Ontological and Functional Structure to Speakers' Elicited Concepts

Sabine Schulte im Walde & Alissa Melinger

Computational Linguistics & Psycholinguistics
Saarland University, Saarbrücken

DGfS Annual Meeting 2005

AG 2: Lexical encoding of implicit information

Motivations: Implicit Semantic Relations

1. Semantic similarity for **Natural Language Processing**
 - task diversity, e.g. compositionality, verb classes
 - feature-based vector description
 - co-occurrence: window, syntactic functions
2. **Semantic priming** reflects semantic memory organisation
 - source: co-occurrence or world knowledge?
 - (non-objective) manipulation of relation types, function is ignored

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 ($1 < \text{freq} < 71,604$)
- 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
- Association input:
spontaneous, exhaustive, one word per line, capitalisation
- 30 sec. for each verb; 2 sec. break; total: ca. 30 min.

Web Experiment 2005

abnehmen

zunehmen

essen

Diät

Experiment Participants and Data

- Participants per data set: **between 44 and 54**
- 299 accepted data files:
 - native German speakers,
 - 80% of target verbs
- Expertise of participants: 166 experts vs. 132 non-experts
- Number of trials: 16,445
- Number of associations per target verb:
 - range 0-16, average: 5.16
- Number of single-word associations: **81,373**

Data Preparation

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

<i>klagen</i> 'complain'		
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
- 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 Analysis: Examples

	V	N	ADJ	ADV
Freq	19.863	48.905	8.510	1.268
Prob	25	62	11	2

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: GWN/Freq/Prob

	GermaNet	Freq	Prob
Synonymy	4.633	1.194	6
Antonymy	470	252	1
Hypernymy	9.281	1.979	10
Hyponymy	9.281	2.285	12
Cause	209	49	0
Entailment	8	0	0
Also see	2	0	0
No relation	-	10.509	55
Unknown relation	-	1.726	9

Window Approach for Semantic Relations

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

<i>window</i>	<i>pos (36%)</i>	<i>neg (64%)</i>
5	42	24
20	64	40
50	73	48

Window Examples for Semantic Relations

GWN-unrelated verbs in window of target verb:

◇ tagging:

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

◇ scene information:

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

Window Examples for Semantic Relations

GWN-unrelated verbs NOT in window of target verb:

◇ domain:

radeln 'bike' - *strampeln* 'pedal, kick' (1)

paddeln 'paddel' - *rudern* 'row' (22)

◇ world knowledge:

aufhören 'finish' - *stoppen* 'stop' (19)

initiieren 'initiate' - *anfangen* 'start' (21)

auftauen 'defrost' - *essen* 'eat' (8)

paddeln 'paddle' - *schwimmen* 'swim' (7)

Syntax-Semantic Functions of Nouns

- Source: statistical grammar model
- Verb valency:
 - ◇ 38 syntactic subcategorisation frames
 - ◇ PP information (case+preposition) → 178 frames
 - ◇ subcategorised nouns

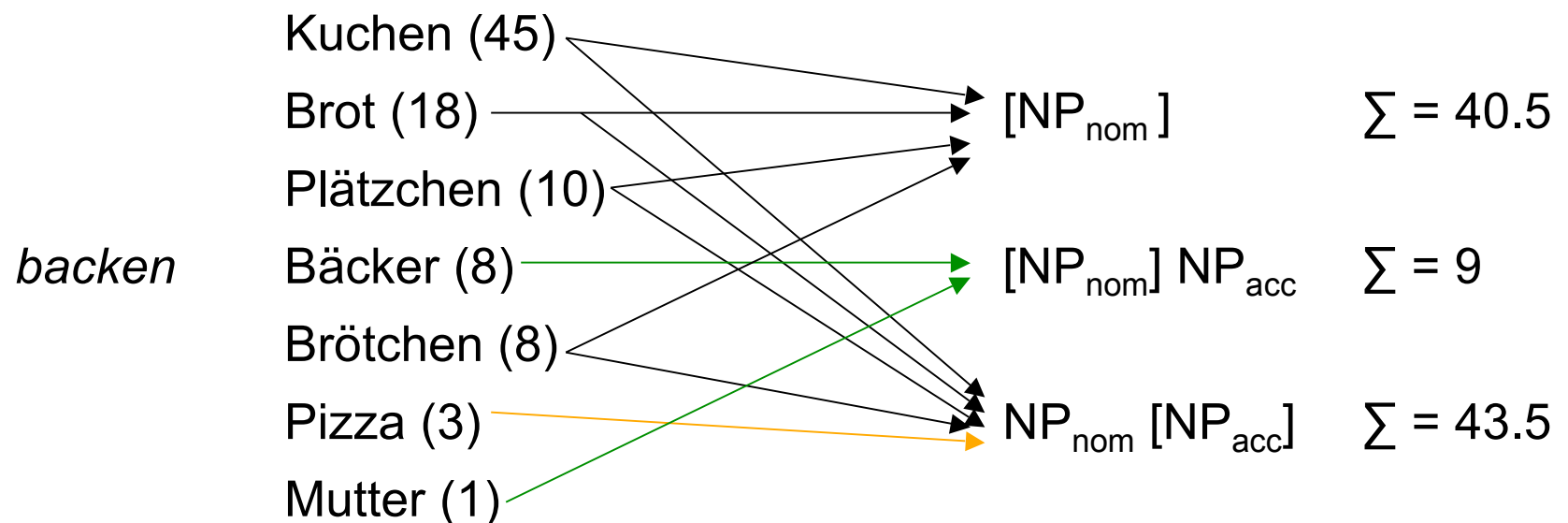
- Example: *backen* 'bake'

frames: NP_{nom}
 $NP_{nom} NP_{acc}$

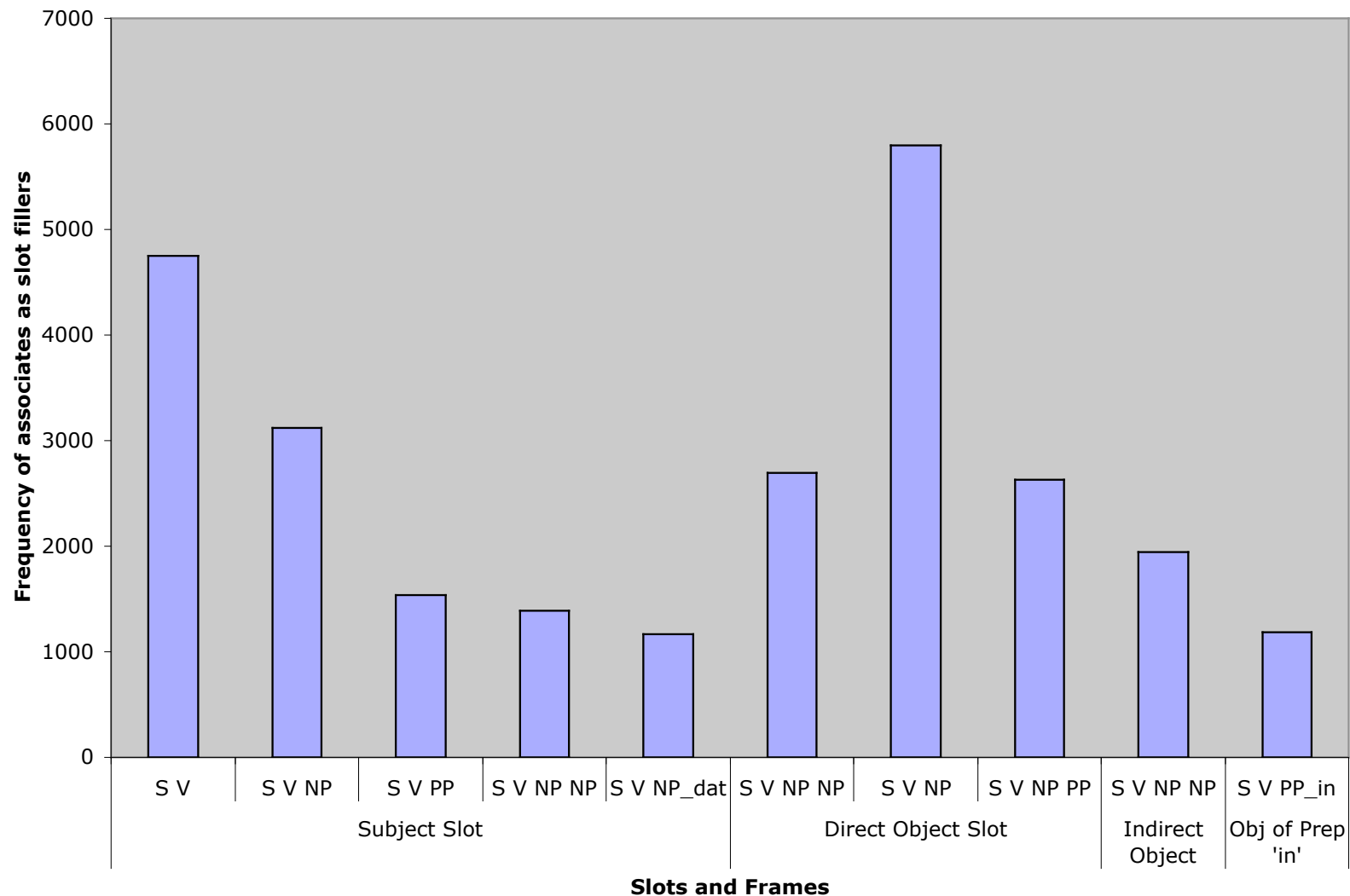
filler NP_{nom} [NP_{acc}]: *Brötchen* 'rolls'
Brot 'bread'
Kuchen 'cake' ...

Syntax-Semantic Functions: Analysis

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



Syntax-Semantic Frame Inspection



Window Approach for Linguistic Functions

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

<i>window</i>	<i>pos (55%)</i>	<i>neg (45%)</i>
5	83	13
20	93	30
50	96	40

Window Examples for Linguistic Functions

Functionally-unrelated nouns in window of target verb:

◇ **argument vs. adjunct:**

fahren 'drive' - *Alkohol* 'alcohol'

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'

Window Examples for Linguistic Functions

Functionally-unrelated nouns NOT in window of target verb:

◇ domain:

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

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

◇ world knowledge:

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

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

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

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

Summary and Outlook

- Ontological / functional structure for
36% / 55% of the target-response pairs
- Window analysis extends coverage to 67% / 73%
- Diversity of relations;
dependent on individual verbs, frequencies, verb classes?
- Why are associations not covered by lexical resources?
 - ◇ imperfect tools and resources
 - ◇ contribution of missing relations by world knowledge
- Application for semantic similarity features in NLP:
combination of syntactic functions and window