Identifying Semantic Relations and Functional Properties of Human Verb Associations

Sabine Schulte im Walde & Alissa Melinger
Saarland University, Germany
{schulte, melinger}@coli.uni-sb.de

Introduction

- Goal: use human verb associations for an investigation of verb properties
- Basis: collection of semantic associates evoked by German verbs in a web experiment
- *Semantic associate:* concept spontaneously called to mind by a stimulus word
- Task: identification and quantification of relationships between stimulus and response

1. Semantic verb relations: The notion of semantic verb relations is crucial for NLP tasks and applications: verb clustering, thesaurus extraction, word sense discrimination, summarisation, etc. Different applications incorporate different semantic verb relations, varying with respect to their demands. We suggest that *an analysis of human verb-verb associations may identify the range of semantic relations* which are crucial in NLP.
2. Functional noun properties: In data-intensive lexical semantics, words are commonly modelled by distributional vectors, and the relatedness of words is measured by vector similarity. The features in the distributional descriptions can be varied in nature. We assume that the *noun associates in our verb experiment are related to conceptual roles* of the respective verbs, and investigate their linguistic functions, to identify prominent roles and conditions for distributional verb descriptions.

Web Experiment

- 330 German verbs were selected.
  - Variety of semantic classes (loosely based on Levin, 1993).
  - Variety of verb frequencies (based on 35 million word corpus).
  - Random division into 6 presentation lists with 55 verbs each.
- 299 native German speakers participated, between 44 and 54 for each data set.
- Elicitation was conducted over the Internet.
- Participants had 30 secs per trail to provide as many responses as possible.
- In total, we collected data for 16,445 trials with over 80,000 responses.
- Each response was quantified by its frequency of occurrence for any individual target, cf. Table 1.

klagen ‘complain, moan, sue’		
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

Table 1. Most frequent responses for *klagen.*

Morpho-Syntactic Analysis:

- Based on an empirical, quantitative dictionary, each response was classified according to its part-of-speech (POS).
- For ambiguous responses such as *überlegen* (verb “think about” vs. adjective “superior”), the response frequency was split over the possible POS tags. 4% of all response types were ambiguous.
- Table 2 provides the overall distributions for response tokens over POS tags.
- The average pattern varies with respect to individual verbs, frequencies, semantic class.
- Correlation between POS and target verb frequency:
  verb frequency ↑ ~ ↑ verb/adverb responses, verb frequency ↑ ~ ↓ noun/adjective responses
- Relation between semantic class and POS? Aspectual verbs such as *aufhören* “stop” received significantly more verb responses and fewer noun responses than creation verbs such as *backen* “bake”, although the verb sets have comparable frequencies.

	Freq	Prob
V	19,863	25%
N	48,905	62%
ADJ	8,510	11%
ADV	1,268	2%

Table 2. POS distributions over all verbs and response tokens.

Semantic Verb Relations

- The lexical semantic taxonomy GermaNet organises words into sets of synonyms. The sets are interconnected by semantic relations like hypernymy, antonymy, etc.
- We looked up the semantic relation between the target and its responses (polysemy considered). The semantic relations were quantified by the target-response frequency.
- Figure 1 shows the distribution of relations between target verbs and verb responses. 54% of response tokens were not encoded as GermaNet semantic relation:
  - » missing links in GermaNet:
    antonymy:
    erhitzen “heat” ⇒ *abkühlen* “cool”
    synonymy:
    prüfen ⇒ *testen* “check/test”
  - » non-classical relations:
    cause / consequence:
    schwitzen “sweat” ⇒ *laufen* “run”
    schwitzen “sweat” ⇒ *stinken* “stink”
    implication:
    erfahren “get to know” ⇒ *wissen* “know”
    temporal order:
    adressieren “address” ⇒ *schreiben* “write” & *schicken* “send”

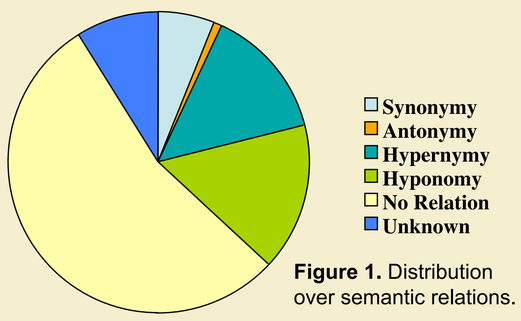
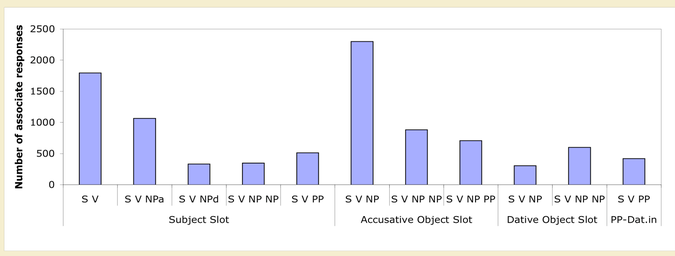


Figure 1. Distribution over semantic relations.

- Correlation between semantic relation and target verb frequency:
  verb frequency ↑ ~ ↑ synonyms/antonyms/hyponyms
- Relation between semantic class and instantiation of semantic relation types?
  Aspectual verbs such as *aufhören* “stop” received significantly more antonymic responses than creation verbs such as *backen* “bake”.
- Window look-up to determine the distance between two associated verbs:
  200 million words newspaper corpus, co-occurrence windows 5/20/50 words left/right;
  co-occurrence rates of 85/95/97% for paradigmatic (GermaNet) relations, i.e. 37% of data;
  61/74/99% for non-GermaNet relations, i.e. 63% of data.

Syntax-Semantic Noun Functions

- Our lexicalised grammar model contains empirical distributions of verbs for subcategorisation frame types and nominal argument fillers (Schulte im Walde, 2003). The model parameters were estimated in an unsupervised training procedure, using 35 million words of a large German newspaper corpus from the 1990s.
- We looked up the linguistic relationships between target verbs and noun associates. For each noun response, we evaluated whether the response co-occurs with the target verb as an argument filler in a particular frame (ambiguity considered). For example, the associate *Kuchen* “cake” in response to *backen* “bake” appears as the direct object of the verb in transitive frames AND as intransitive subject.
- 28% of all noun associates were identified with a subcategorisation function of the target verbs.
  11 frame-slot functions were evoked by the noun tokens with prob > 1%:



- 22% of the noun associates do not appear in the grammar model:
  lemmatisation in grammar: *Autorennen “car racing”* ⇒ lemma *Rennen* vs. full compound
  domain and size of corpus: *Grufties* “old people” ⇒ slang; *Plosiv* “plosive” ⇒ technical
- 50% missing cases, beyond subcategorisation:
  - *fliegen* “fly” ⇒ *Urlaub* “vacation”; *Flügel* “wings”; *Freiheit* “freedom”
  - *backen* “bake” ⇒ *Mehl* “flour”; *Ofen* “oven”; *Weihnachten* “Christmas”
- Relation between semantic class and alternation between linguistic functions?
  Creation verbs such as *backen* “bake” activate direct object slot fillers significantly more often than aspectual verbs such as *aufhören* “stop”.
- Window look-up to address scene-related information and world knowledge:
  200 million words newspaper corpus, co-occurrence windows 5/20/50 words left/right;
  22/69/75% of missing noun cases appear in the window ⇒ one quarter still missing.

Conclusions

- 20,000 verb associates:
  - 37% of the verb-verb pairs are specified as paradigmatic GermaNet relations. Verb-verb pairs with no GermaNet relation provide an empirical basis for detecting missing links in the taxonymy.
  - Non-classical verb-verb relations (such as temporal order, cause, consequence) are represented in a large proportion of the verb-verb pairs. They represent an excellent basis for defining an exhaustive set of non-classical relations.
  - The co-occurrence distance between two verbs varies with respect to their semantic relation; this information is useful in NLP applications, such as summarisation.
- 50,000 noun associates:
  - 28% of the verb-noun pairs refer to prominent frame-role combinations. The identified conceptual roles might contribute to distributional verb descriptions.
  - Window-based nouns also contribute to verb descriptions by encoding scene information, rather than intra-sentential functions. This finding supports the integration of window-based approaches into function-based approaches.
- Future work:
  - Establish a set of non-classical verb-verb relations.
  - Apply variations of verb feature descriptions to find dependencies between feature descriptions and verb relations.
  - Use morphological, syntactic and semantic properties of associates as indicators for semantic verb classes.