Evaluating the Relationships Instantiated by Semantic Associations of Verbs

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

Introduction

- Associative relatedness is the likelihood that a word is called to mind by another word.
  - It is assumed to reflect word co-occurrence probabilities rather than the organization of semantic memory. The co-occurrence assumption is supported by, e.g., correlations between associative strength and word-co-occurrence.
  - Free association tasks elicit a wide variety of association types, e.g., synonymy, category coordinates, personal recollections, etc. The heterogeneity of response types complicates the interpretation of experimental results based on associates.
- An in-depth investigation into the types of relations and functions that associates reflect, combined with a co-occurrence analysis, may help to assuage these concerns as well as provide a useful tool for future investigations.

Goals

- Characterize noun and verb associates of verb targets.
- Evaluate the co-occurrence assumption.

Associate Elicitation

- 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.
- 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 and a total of over 80,000 responses.
- Each response was quantified by its frequency of occurrence for any individual target, cf. Table 1.

klagen '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

Table 1. Most frequent responses for *klagen*.

Morpho-Syntactic Analysis

- Based on a machine-readable dictionary; each response word was classified according to its part-of-speech (POS).
- For ambiguous responses, e.g., *walk*, the response frequency was split over the possible POS tags.
- Table 2 provides the overall distribution of elicited concepts over POS.
- The average pattern varies with respect to individual verbs, frequencies, semantic class.

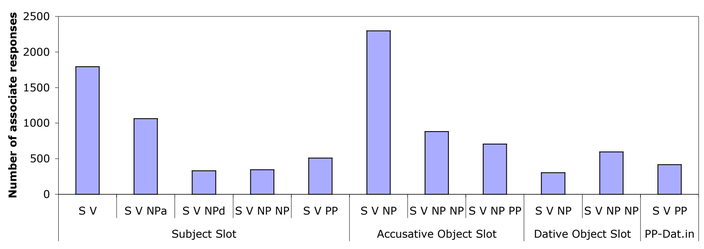
Table 2. Frequencies and proportions of all elicited concepts classified into the four major parts-of-speech.

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

Noun Responses: Functional Analysis

- Our quantitative grammar model contains empirical distributions of verbs for subcategorization frame types and nominal argument fillers (Schulte im Walde, 2003). The model is statistically trained on 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 associate 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" co-occurs as the direct object of the verb in transitive clauses AND as intransitive subject.
- Figure 1 shows which frame-slot combinations were most strongly represented by the noun associates.
  - 28% of all noun responses (26% of first responses) were argument fillers in our corpus. 11 frame-slot combinations were evoked by more than 1% of noun tokens.
  - There was a strong tendency for speakers to produce associates which are fillers of the direct object in the transitive frame or the subject in the intransitive frame. The high rate of subjects was surprising given that they are not particularly constraining.
  - Missing cases, beyond subcategorization:
    - *fliegen* "fly" ⇒ *Urlaub* "vacation"; *Flügel* "wings"; *Freiheit* "freedom"
    - *backen* "bake" ⇒ *Mehl* "flour"; *Ofen* "oven"; *Weihnachten* "Christmas"

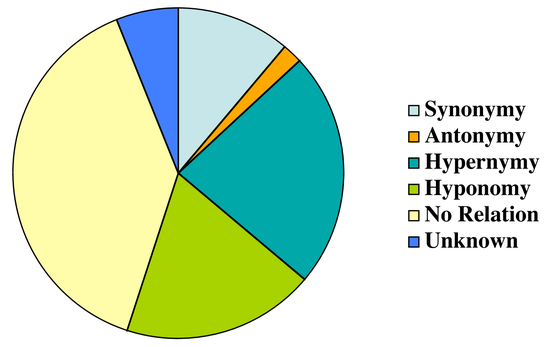
Figure 1. Frequency of associate responses occurring as verb argument fillers in 11 most frequent frame-slot combinations.



Verb Responses: Relational Analysis

- The lexical semantic taxonomy GermaNet organizes words into sets of synonyms. The sets are interconnected by semantic relations, including hypernomy/hyponymy, 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 2 shows the distribution of semantic relations between target verbs and the first response given to each target.
  - 42% of responses were hypernyms or hyponyms.
  - 39% of responses were not encoded as GermaNet semantic relation;
    - » non-classical relations:
      - causality: *schwitzen* "sweat" ⇒ *stinken* "stink"; *brauchen* "need" ⇒ *besorgen* "get"
      - implications: *setzen* "seat" ⇒ *sitzen* "sit"; *erfahren* "get to know" ⇒ *wissen* "know"
      - temporal order: *adressieren* "address" ⇒ *schreiben* "write" & *schicken* "send"
    - » some missing links in GermaNet:
      - antonymy: *erhitzen* "heat" ⇒ *abkühlen* "cool"; *aufhören* "finish" ⇒ *beginnen* "begin"
      - synonymy: *prüfen* ⇒ *testen* "check/test"; *fahren* ⇒ *wegfahren* "drive (away)"

Figure 2. Proportion of first elicited associate responses instantiating classical GermaNet semantic relations.



Co-occurrence Window Analysis

- We searched for target-response co-occurrence in three search windows of 5/20/50 words to the right and left of the target verb in our corpus. We distinguished positive and negative cases, with respect to the above analyses of the responses.
- Overall, only 54% of noun responses and 57% of verb responses co-occurred in the 20-word window. Considering only first responses, coverage increased to 69%/64% for nouns/verbs.
- Examples for negative noun cases which co-occur with the target in a 20-word window: *auftauen* "defrost" ⇒ *Wasser* "water"; *trocknen* "dry" ⇒ *Trockner* "dryer".

All Responses	Positive (28%)	Negative (72%)	All
	95%	37%	54%
First Response	Positive (26%)	Negative (74%)	All
	96%	48%	69%

Table 3. Window analysis for nouns (20-word window)

All Responses	Positive (37%)	Negative (63%)	All
	75%	46%	57%
First Response	Positive (55%)	Negative (45%)	All
	79%	46%	64%

Table 4. Window analysis for verbs (20-word window)

Conclusions

- This work examined the functional, ontological and co-occurrence relations of a large set of stimulus-response pairs.
  - Characterize associates: noun responses largely reflected non-argument concepts, and verb responses often reflected non-classical semantic relations.
  - Evaluate co-occurrence assumptions: 67% of first responses occurred in the co-occurrence window of ±20 words, 51% occurred in the co-occurrence window of ± 5 words.
- The analyses of noun and verb responses provides an objective characterization of associate relation types which can inform the interpretation of experimental results.
- Is the co-occurrence assumption still tenable, or do the analyses suggest that semantic memory may be tapped by associate elicitation? What is the appropriate baseline to use as an evaluation metric?
- Prior research focused on associates of noun targets. Do the above patterns generalize to nouns?
- Do different verb classes elicit different distributions of associates? Preliminary analyses suggest so. For example, verbs of creation, such as *bake*, elicited significantly more direct object fillers than aspectual verbs, such as *stop*; aspectual verbs elicited significantly more antonyms than creation verbs.