Evaluating the Relationships Instantiated by Semantic Associates of Verbs

Alissa Melinger and **Sabine Schulte im Walde**
Psycholinguistics and Computational Linguistics
Saarland University
Saarbrücken, Germany

Abstract

This work is concerned with an investigation of semantic associations. We performed an elicitation task where native speakers were asked to spontaneously list semantic associations for German verbs. The elicited conceptual knowledge was then given ontological structure based on codes from the psycholinguistic taxonomy GermaNet as well as linguistic functions obtained from statistical corpus parses. The investigation is directed towards discovering and specifying the structural and conceptual types of verb associations, stemming from an interest in better characterizing semantic associates.

Introduction

In psycholinguistic research, a fundamental goal is to understand how linguistic meaning is represented and accessed in the course of language understanding and production. To this end, several paradigms have been developed over the years to investigate the nature of semantic or conceptual networks. Early results demonstrated that word recognition is sped up by the prior presentation of a semantically related word. For example, the recognition of the target word *bread* is faster when it is preceded by the semantic associate *butter* compared to the unrelated *doctor* (Meyer and Schvaneveldt, 1971). Likewise, unrelated primes which share conceptual neighbours with the target can also influence response times, e.g., *winter* facilitates the recognition of *swim*, mediated via *summer* (Seidenberg et al., 1984). These results implicate an intricate network of semantic relations that are activated in the course of word recognition, cf. Collins and Loftus (1975).

In the literature on semantic priming, several types of relationships can be distinguished, such as semantic associates (bread-butter), unassociated category type relations (pig-dog), functional relations (broom-floor), etc. Associative relatedness reflects 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, cf. Plaut (1995); McKoon and Ratcliff (1992). This assumption is supported by observed correlations between associative strength and word co-occurrence in large language corpora (Spence and Owens, 1990). These findings have therefore been taken as an argument against the use of associates in investigations of semantic memory, e.g., McRae et al. (1997).

A further argument against the use of associates in the investigation of semantic memory is that free association elicits a wide variety of association types, including synonyms (*sofa - couch*), category coordinates (*sofa - chair*), words with high transitional probabilities in text corpora (*private - property*), personal recollections (*bicycle - dad*), super- and subordinate relations (*sofa - furniture*), etc. (McRae and Boisvert, 1998; McKoon and Ratcliff, 1992). This heterogeneity of response type makes it difficult to determine what aspects of meaning might be relevant to the priming effect.

These objections to the use of association norms raise interesting questions. Specifically, exactly what types of relations are evoked by the free association task and, if associations reflect co-occurrence frequencies, do associates correspond to particular functional roles of the target verb? To address these questions, we combine an investigation of semantic associates with lexical resources from computational linguistics, with a first goal of determining the semantic relations and linguistic functions of speakers' elicited concepts with respect to the target verb.

In the following section we present a method for eliciting associated concepts which will serve as the data source for the feature exploration to follow. The elicitation procedure asked participants to provide their associations to German verbs.

Associate Elicitation Method

Participants: 298 native German speakers participated in the elicitation procedure. They received no monetary compensation but one individual was randomly selected to receive a 25Euro gift certificate for Amazon.

Materials: 330 verbs were selected for the study. They were drawn from a variety of semantic classes including verbs of self-motion (e.g., *gehen* 'walk', *schwimmen* 'swim'), transfer of possession (e.g., *kaufen* 'buy', *kriegen* 'receive'), cause (e.g., *verbrennen* 'burn', *reduzieren* 'reduce'), experiencing (e.g., *lachen* 'laugh', *hassen* 'hate', *überraschen* 'surprise'), communication (e.g., *reden* 'talk', *beneiden* 'envy'), etc. Drawing verbs from different categories was intended only to ensure that the elicitation covered a wide variety of verb types; the inclusion of any verb into any particular verb class was achieved in part with reference to prior verb classification work, e.g., Levin (1993) but also on intuitive

grounds. It is not critical for the subsequent analyses. The frequencies of the verbs were checked using a German statistical grammar trained on 35 million words (Schulte im Walde, 2003). The 330 verbs were divided into 6 separate presentation lists of 55 verbs each. Each list contained verbs from each grossly defined semantic class. The verbs were also divided such that the lists had equivalent overall frequency distributions.

Procedure: The elicitation study was administered electronically over the Internet. The program was compatible with most browsers and platforms. When participants loaded the elicitation page, they were first asked for their biographical information, such as linguistic expertise, age and regional accent. Next, the participant was presented with instructions for the elicitation study and an example item set. Participants clicked on an 'ok' button to indicate that they had understood the instructions and that they were ready to proceed.

Each trial consisted of a verb presented in a box at the top of the screen. Below the verb was a series of data input lines where participants could type their associations. They were instructed to type at most one word per line and, following German grammar, to distinguish nouns from other parts of speech with capitalisation. Participants had 30 sec. per verb to type as many associations as they could. After this time limit, the program automatically advanced to the next trial. There was a 2 sec. pause between trials to prevent response spillover between trials. In total, we collected data for 16,445 trials; each trial elicited an average of 5.16 associate responses with a range of 0-16. In sum we collected over 80,000 non-unique target-response pairs.

Once the study began, it could not be stopped or paused, nor could participants return to prior trials with the 'back' button. At the end of the study the data was automatically saved to an individually named file and e-mailed to the first author.

Data Preparation: Each completed data set contains the background information of the participant, followed by the list of target verbs. Each target verb is paired with a list of associations, in the order in which the participant provided the associates. For the analyses to follow, we pre-processed all data sets in the following way: For each target verb, we quantified over all responses in the study, disregarding the participant's background and the order of the associates. Table 1 lists the most frequent responses for the verbs *abhauen* 'walk off', and *klagen* 'complain'.

Linguistic Analysis of Elicited Concepts

The verb associations were investigated on three linguistic issues. We were interested in the type of relationship typical associates established with the target verb: whether verb responses refer to particular semantic relations (such as synonyms, antonyms, hypernyms), and whether noun responses are typical argument holders of verb valency. To address these questions, we conducted the following three analyses:

Table 1: Most frequently provided responses (and their response frequencies) for two sample target verbs.

abhauen			*klagen*		
Flucht	escape	12	Gericht	court	19
weglaufen	run away	12	jammern	moan	18
Angst	fear	10	weinen	cry	13
fliehen	escape	10	Anwalt	lawyer	11
wegrennen	run away	9	Richter	judge	9
rennen	run	6	Klage	complaint	7
flüchten	escape	6	Leid	suffering	6
schnell	quickly	6	Trauer	mourning	6

1. In a preliminary step, we distinguished the responses with respect to the major part-of-speech tags: nouns, verbs, adjectives and adverbs.

2. For each response classified as a verb, we looked up the semantic relation between the target and response using the lexical taxonomy GermaNet (Kunze, 2000).

3. For each response classified as a noun, we investigated the kinds of linguistic functions that are realized by the associate with respect to the target verb. The analysis is based on an empirical grammar model.

Morpho-Syntactic Analysis on Part-of-Speech Tags

Each associate of the target verb was assigned its (possibly ambiguous) part-of-speech. The assignment was based on a machine-readable dictionary with information on word forms, parts-of-speech tags and lemmas. Whenever a word in the dictionary is morphologically ambiguous with respect to its part-of-speech or lemma, it constitutes a separate dictionary entry. Originally, the dictionary distinguished approximately 50 morpho-syntactic categories, but we only considered the major categories verb (V), noun (N), adjective (ADJ) and adverb (ADJ), disregarding case, number and gender features. Ambiguities between these categories arise e.g., in the case of nominalized verbs (e.g., *Rauchen* 'smoke', *Vergnügen* 'please/pleasure'), where the participant could have intended either a verb or noun, or in the case of past participles (e.g., *verschlafen* 'slept/sleep') or infinitives (e.g., *überlegen* 'consider/superior'), where the participant could have intended either a verb or an adjective.

Having assigned part-of-speech tags to the associates, we can distinguish and quantify the morpho-syntactic categories of the responses. When the response was non-ambiguous, the unique part-of-speech received the total target-response frequency; when the response was ambiguous, the target-response frequency was split over the possible part-of-speech tags. As the result of this first analysis, we can specify the frequency and probability distributions for the part-of-speech tags for each verb and also in total. Table 2 presents the total numbers and specific verb examples. Participants produced noun associates in the clear majority of instances, 62%; verbs are given in 25% of the responses, adjectives in 11%, adverbs almost never (2%). This average pattern varies, of course, with respect to specific verbs.

Table 2: Total frequencies and proportions of all elicited concepts classified into the four major parts-of-speech. Proportions for specific target examples also provided.

	V	N	ADJ	ADV
Total Freq	19.863	48.905	8.510	1.268
Total Prob	25%	62%	11%	2%
aufhören 'stop'	49%	39%	4%	6%
aufregen 'be upset'	22%	54%	21%	0%
backen 'bake'	7%	86%	6%	1%
bemerken 'realize'	52%	31%	12%	2%
dünken 'seem'	46%	30%	18%	1%
flüstern 'whisper'	19%	43%	37%	0%
nehmen 'take'	60%	31%	3%	2%
radeln 'bike'	8%	84%	6%	2%
schreiben 'write'	14%	81%	4%	1%
wundern 'be surprised'	30%	35%	31%	1%

Semantic Relations of Verb Associates

To determine which types of relationships are typically instantiated between target and response verbs, we used the lexical semantic taxonomy GermaNet (Kunze, 2000), the German counterpart of WordNet (Fellbaum, 1998). The lexical database is inspired by psycholinguistic research on semantic memory. The resource organizes nouns, verbs, adjectives and adverbs into classes of synonyms (synsets), which are connected by lexical and conceptual relations. The GermaNet version from October 2001 contains 6,904 verbs and defines the semantic relations *synonymy, antonymy, hypernymy/hyponymy, entailment, cause,* and *also see* between verbs or verb synsets. (*Also see* is an underspecified association which captures relationships other than the preceding standard ones. For example, *sparen* 'save' is related to *haushalten* 'budget' by *also see*.) The hypernym-hyponym relation imposes a multi-level hierarchical structure on the taxonomy. Words with several senses are assigned to multiple classes.

Based on the GermaNet relations, we could distinguish between different kinds of verb associations elicited from speakers. For example, the response *hetzen* for *hasten* (both meaning: 'rush') are synonyms of each other, but the response *bewegen* 'move' is a hypernym for verbs such as *rennen* 'run', *rollen* 'roll', *fließen* 'float'. With these distinctions, we can identify the relations established by the verb concepts evoked by the target verbs.

Our analysis proceeded as follows. For each verb associate, we looked up the semantic relation between the target and response verbs as coded in GermaNet: For each pair of target and response verbs, we looked up whether any semantic relation is defined between any of the synsets the verbs belong to. For example, if the target verb *rennen* is in synsets a and b, and the response verb *bewegen* is in synsets c and d, we determined whether there is any semantic relation between the synsets a and c, a and d, b and c, b and d. Two verbs belonging to the same synset are synonymous. The semantic relations are quantified by the target-response pair frequencies, e.g., if 12 participants provided the association *bewegen* 'move' for *rennen* 'run', the hypernymy relation is quantified by the token frequency 12. If the target and the response verb are both in GermaNet, but there is no relation between their synsets, then the verbs do not bear any kind of semantic relation. If either of them is not in GermaNet, we cannot make any statement about the verb-verb relationship. Nine percent of our data falls into this unknown category. Table 3 shows the number of semantic relations encoded in the 2001 GermaNet version, and the token frequencies and probabilities of their instantiations in our data. For example, there are 19,424 cases of hypernymy-hyponymy defined between the verbs in GermaNet. Among our target-verb response pairs, 2,807 corresponded to this defined set of related verbs, which accounts for 14% of all our verb responses. Again, the distributions vary with respect to the individual verbs. For example, the aspectual verb *aufhören* 'stop' was mostly associated with antonyms such as *anfangen* 'begin', and *weitermachen* 'go on', and hypernyms such as *enden* 'end'; *schreiben* 'write' was mainly associated with hyponyms such as *tippen* 'type', *aufschreiben* 'write down' and *kritzeln* 'scribble'; *aufregen* 'be upset' was mainly associated with synonyms such as *ärgern* 'be angry' and *nerven* 'annoy'.

Table 3 shows that a remarkable number of the verb-verb associations elicited in our study (54%) do not correspond to a semantic relation defined in GermaNet. This failure to capture elicited verb relations stems from two main sources; on the one hand, we find a larger variety of verb relations among the associates than the classical relations defined in GermaNet (e.g., implication, causality, temporal relation, see the General Discussion for more details), and on the other hand work on the GermaNet taxonomy is not yet finished.

Table 3: Total frequency of each semantic relation type in GermaNet, token frequency of each relation in our data set and the proportion of our data captured by each relation type.

	GermaNet	Freq	Prob
Synonymy	4,633	1,194	6%
Antonymy	571	252	1%
Hypernymy	19,424	2,807	14%
Hyponymy	19,424	3,016	16%
Cause	236	49	0%
Entailment	15	0	0%
Also see	2	0	0%
No relation	-	10,509	54%
Unknown cases	-	1,726	9%

A more detailed inspection of the semantic relations provides some insight into target verb properties. For example, target verbs with synonym associates are rather high frequency verbs (and therefore conceptually more general), such as *bekommen* 'receive', *gehen* 'go', *laufen* 'run'; target verbs with antonym associations tend to be aspectual or change of state verbs, such as *anfangen* 'begin', *einfrieren* 'freeze', *schmelzen* 'melt'. Target verbs with hypernym associates tend to be rather specific, such as *eintüten* 'bag', *hüpfen* 'hop', *schlurfen* 'scuffle', while target verbs with hyponym associates

tend to be rather general, such as *denken* 'think', *sagen* 'say', *wahrnehmen* 'observe'; target verbs with cause associates are transfer and change of state verbs, such as *formen* 'form', *legen* 'place', *töten* 'kill'. So far, these insights are based rather on intuitive comparisons; correlation analyses are planned to further investigate the generalizability of these impressions. However, it appears that the obtained ontological data could be very useful for conducting controlled priming studies in which different types of associative relations are contrasted.

The analysis with GermaNet is consistent with the view that normed associates reflect word co-occurrence frequencies. Indeed, if many unrelated verb-verb pairs reflect implications, cause/effect and temporally linked events, we may well find that verb-verb pairs often co-occur in texts, for example in adjacent clauses. To investigate this possibility, we returned to our 35 million word corpus and searched for target-response co-occurrences in three search windows of 5/20/50 words to the left and right of the target word. Note that this is a weak estimate of co-occurrence as a target-response pairs need only co-occur one time in the corpus to positively contribute to the analysis. We also evaluated co-occurrence likelihood for just the first response provided to each verb, rather than the entire set of responses. Table 4 shows the percentage of verb responses that co-occur with their respective target verbs in the moderate window of 20 words. For responses which were captured by GermaNet (positive cases), 75% of all verb responses appeared in the search window. For responses not expressing GermaNet relations (negative cases), 46% were found in the search window. Thus, 43% of all verb associate responses were not found in the search window. Furthermore, a full third of the verb responses were not captured by either the analysis with GermaNet or the search window. As Table 4 shows, this pattern persists if we consider only the first response provided to each target rather than the entire response set, which includes idiosyncratic singleton responses. Even the strongest associates reflect GermaNet relations only 55% of the time and they occur in the search window only 64% of the time.

Table 4: Percentage of verb responses captured (positive) and not captured (negative) by GermaNet found in a 20 word search windows.

All Responses	positive (37%)	negative (63%)	all
	75%	46%	57%
First Response	positive (55%)	negative (45%)	all
	79%	46%	64%

This result presents a challenge to those researchers who hold that associate elicitation reflects word-form co-occurrences. However, the majority of research into semantic memory and word recognition has investigated noun-noun relationships, e.g., Spence and Owens (1990); Moss et al. (1995); McRae et al. (1997); Meyer and Schvaneveldt (1971). It could be assumed that verb-noun pairs would co-occur textually if associates reflect typical argument fillers of the verbs. We investigate this point in the next section.

Syntax-Semantic Functions of Noun Associates

We investigated the kinds of linguistic functions that are realized by noun associates of the target verbs. This analysis utilizes a German statistical grammar framework: Schulte im Walde (2003) developed a context-free grammar for German, with the goal of obtaining reliable lexical information on verbs. Work concentrated on defining linguistic structures which are relevant to lexical verb information, especially subcategorisation. The manually defined grammar was trained by lexicalized parameter estimation, using 35 million words of a large German newspaper corpus from the 1990s. The resulting grammar model contains quantitative information on lexicalized linguistic functions, and head-head relationships.

With respect to verb subcategorisation, the empirical grammar contains frequency distributions of verbs for 178 subcategorisation frame types, including prepositional phrase information and frequency distributions of verbs for nominal argument fillers. For example, the verb *backen* 'bake' appeared 80 times with an intransitive frame and 109 times with a transitive frame, subcategorising for a direct object. With a total corpus frequency of 240, this corresponds to 33% for the intransitive and 45% for the transitive frame. The most frequent nouns subcategorized as direct object in the transitive frame are *Brötchen* 'rolls' (37%), *Brot* 'bread' (17%), *Kuchen* 'cake' (14%), *Plätzchen* 'cookies' (8%), and *Waffel* 'waffle' (5%).

We used the grammar information to look up the syntactic relationships which exist between a target verb and a response noun. For example, the noun associates *Kuchen* 'cake', *Brot* 'bread', *Plätzchen* 'cookies' and *Brötchen* 'rolls' associated with *backen* 'bake' appeared not only as the verb's direct objects (as illustrated above), but also as intransitive subjects; *Pizza* appeared only as a direct object, and *Bäcker* 'baker', *Bäckerei* 'bakery' and *Mutter* 'mother' appeared only as transitive subjects. The verb-noun relationships which were found in the grammar were then quantified by the verb-noun association frequency, and divided by the number of different relationships found in the grammar for the specific lexeme pair (to account for the ambiguity represented by multiple relationships). For example, the noun *Kuchen* was elicited 45 times as response to *bake*, the grammar contains the noun both as direct object and as intransitive subject for that verb, so both functions were assigned a frequency of 22.5.

We then accumulated the verb-association frequencies for all nouns with a specific relationship, e.g., for the intransitive subjects, we summed over the empirical association evidence for *Kuchen*, *Brot*, *Plätzchen*, *Brötchen*. The result is frequency and probability distributions for the linguistic functions for each target verb, i.e. for each verb we can determine which linguistic functions were ac-

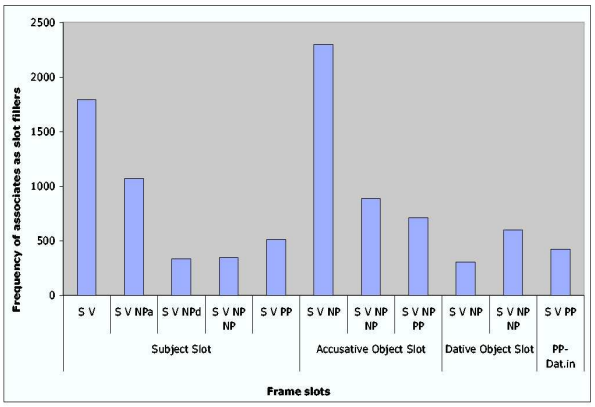

Figure 1: Frequencies of associates as slot fillers.

tivated by how many nouns. Abstracting over the verbs provides the distributions for the general case, i.e. it provides an empirical measure of the linguistic functions of our noun associations.

Examining the overall frequency distribution for linguistic relationships, we discovered that only 11 frame-slot combinations were represented by more than 1% of the noun tokens: subjects in the intransitive, transitive (with direct object, indirect object, or prepositional phrase) and ditransitive frames; the direct object slot in the transitive and ditransitive frames as well as in the direct object plus PP frame; the indirect object in the ditransitive frames and the object of the preposition *Dat:in* for dative (locative) 'in' frames. The frequency proportions are illustrated in Figure 1, with the x-axis referring to the frame-slot combinations and the y-axis to the association frequencies. As this Figure shows, there was a strong tendency for speakers to produce associates which are fillers of either the direct object of the transitive frame or the subject of the intransitive frame. The overlap of noun associates with corpus-based verb preferences illustrates that to a certain extent speakers had conceptual roles for the target verbs in mind when they provided the associates.

As with the verb responses, we now investigate whether the noun responses typically co-occur with the target verb. We apply the same window analysis described for the verb responses, looking again at all noun responses and only the first provided response in separate analyses. Table 5 shows that, contrary to the received view, only half of the noun responses co-occur with their target verb in the 20 word window. Furthermore, for those responses which do not correspond to an argument role filler, only 37% of the responses are found in the search window. The coverage of the search window improves somewhat when only the first noun response for each target verb is considered, as shown in Table 5. Here, overall coverage increases to 69%, and half of the non-argument noun responses are found in the 20 word window. However, 31% of the strongest associates are not in the search window which contradicts the claim that association norms reflect word co-occurrence frequencies.

Table 5: Percentage of all noun responses, both encoding a grammatical function (positive) and not (negative) found in a 20 word search window.

All responses	positive (28%)	negative (72%)	all
	95%	37%	54%
First response	positive (26%)	negative (74%)	all
	96%	48%	69%

Discussion and Outlook

This paper was concerned with an investigation of spontaneous semantic associations. Our aim was to identify which conceptual roles are captured by speakers' elicited concepts. To this end, we used existing lexical resources to determine semantic relations and linguistic functions of response words with respect to the target verbs. The analyses resulted in ontological and functional structure for approximately 40%/30% of the target-response pairs. Additionally, an examination of the co-occurrence of target-response pairs in a large corpus of written German revealed that the received wisdom about what normed associates reflect may be wrong. The insights into the kinds of related concepts elicited by free association norms should prove useful for researchers interested in further distinguishing types of relatedness in a non-ad hoc fashion.

Insight into the nature of normed associations can also be gleaned from an examination of which kinds of responses are not captured by our analyses. For example, do the majority of missing links in our GermaNet analysis correspond to classic semantic relations (which are not yet instantiated), or more interestingly, does free association produce non-classic semantic relations? As already mentioned, some missing links in our data refer to causal relations between verbs (e.g., the target verb *abstürzen* 'crash' evokes the associate *fallen* 'fall', *schwitzen* 'sweat' evokes *stinken* 'stink'), implications (e.g., *setzen* 'seat' → *sitzen* 'sit'), or synonyms/hypernyms of infrequent verbs (e.g., *glucksen* 'chortle' → *lachen* 'laugh', *paddeln* 'paddle' → *rudern* 'row'). A large number of missing relations refer to a temporal order of states and events, (e.g., *adressieren* 'address' → *schreiben* 'write' and *schicken* 'send', *abstürzen* 'crash' → *klettern* 'climb'). Thus, while some relations could be integrated into the current GermaNet framework (excluding from consideration the potential purview of the *also see* relation), many fall outside the scope of traditional semantic relations.

Matching noun associates with conceptual roles in the statistical grammar only covered 28% of all elicited nominal associations (quantified by response frequency) and only 26% of the first responses. For the remaining 72% of all responses, the grammar does not provide linguistic functions. On the one hand, this is due to the fact

that the grammar is trained on newspaper data, and therefore biased to use newspaper-related words, subcategorisation frames, and conceptual roles.[1] More importantly, the conceptual roles of the noun associates are obviously not restricted to arguments of the target verbs. For example, frequent nouns for the verb *backen* 'bake' are *Ofen* 'oven' (referring to the typical device for baking), *Mehl* 'flour' (referring to a typical substance for baking), *Weihnachten* 'Christmas' (referring to a typical occasion for baking), and *Teig* 'dough' (referring to a typical stage of the baking product). These kinds of noun associates fulfill conceptual roles which are not captured by subcategorisation.

Although our analyses do not provide complete coverage even of our strongly associated pairs, we do not view our results as directly conflicting with prior findings which showed that associates co-occur in texts. Rather, our analyses provide another measure which suggests a more conservative relationship between associates and lexical co-occurrence. For example, prior work focused on noun-noun associations (Spence and Owens, 1990; McKoon and Ratcliff, 1992), while we examined both verb-verb and verb-noun pairs. Spence and Owens' prior results were also based on an investigation of a small set of optimized stimulus-response pairs consisting largely of near synonyms (e.g., house-home) and noun pairs which conjoin into single NPs (e.g., bread & butter). For such a small optimized set of associates, it is not surprising that they found their pairs co-occurred more often than unrelated words. In our study, we examined the functional, ontological and co-occurrence relations of all stimulus-response pairs provided by our participants. Our results point to the possibility that prior findings do not generalize to other associates. In sum, our results are not directly comparable to prior approaches but do point to a very different conclusion, namely that lexical co-occurrence and association norms do not index the same relationship.

To conclude, our analysis provides a detailed breakdown of the types of relations that are evoked by target verbs during a association elicitation task. Furthermore, contrary to the commonly held view that speakers produce associates which co-occur with the target word in linguistic contexts, our analyses reveal that the majority of responses did not occur in the extended linguistic context. This finding poses a challenge to the view that associative priming effects are driven by spreading activation between commonly co-occurring lexical items rather than due to spreading activation at the conceptual level between related concepts.

References

Collins, A. M. and Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82:407–428.

Fellbaum, C., editor (1998). *WordNet – An Electronic Lexical Database*. Language, Speech, and Communication. MIT Press, Cambridge, MA.

Kunze, C. (2000). Extension and use of GermaNet, a lexical-semantic database. In *Proceedings of the 2nd International Conference on Language Resources and Evaluation*, pages 999–1002, Athens, Greece.

Levin, B. (1993). *English Verb Classes and Alternations*. The University of Chicago Press.

McKoon, G. and Ratcliff, R. (1992). Spreading activation versus compound cue accounts of priming: Mediated priming revisited. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 18:1155–1172.

McRae, K. and Boisvert, S. (1998). Automatic semantic similarity priming. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 24:558–572.

McRae, K., de Sa, V. R., and Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meanings. *Journal of Experimental Psychology: General*, 126:99–130.

Meyer, D. and Schvaneveldt, R. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90:227–235.

Moss, H., Ostrin, R., Tyler, L., and Marslen-Wilson, W. (1995). Accessing different types of lexical semantic information: Evidence from priming. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 21:863–883.

Plaut, D. (1995). Semantic and associative priming in a distributed attractor network. In *Proceedings of the 17th Annual Conference of the Cognitive Science Society*, pages 37–42.

Roland, D. and Jurafsky, D. (1998). How verb subcategorization frequencies are affected by corpus choice. In *Proceedings of the 17th International Conference on Computational Linguistics*, Montreal, Canada.

Schulte im Walde, S. (2003). *Experiments on the Automatic Induction of German Semantic Verb Classes*. PhD thesis, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.

Seidenberg, M. S., Waters, G. S., Sanders, M. E., and Langer, P. (1984). Pre- and post-lexical loci of contextual effects on word recognition. *Memory and Cognition*, 12:315–328.

Spence, D. P. and Owens, K. C. (1990). Lexical co-occurrence and association strength. *Journal of Psycholinguistic Research*, 19:317–330.

[1] Roland and Jurafsky (1998) showed that word senses and subcategorisation frames strongly vary with respect to the corpus.