Research on Language and Computation **00:** 1–38, 2008.

An Empirical Characterisation of Response Types in German Association Norms**

SABINE SCHULTE IM WALDE (schulte@ims.uni-stuttgart.de)
Institute for Natural Language Processing, University of Stuttgart, Germany

ALISSA MELINGER (a.melinger@dundee.ac.uk)
School of Psychology, University of Dundee, Scotland, UK

MICHAEL ROTH (mroth@coli.uni-sb.de)
Computational Linguistics, Saarland University, Saarbrücken, Germany

ANDREA WEBER (aweber@coli.uni-sb.de)
Psycholinguistics, Saarland University, Saarbrücken, Germany

March 6, 2008

Abstract. This article presents a study to distinguish and quantify the various types of semantic associations provided by humans, to investigate their properties, and to discuss the impact that our analyses may have on NLP tasks. Specifically, we concentrate on two issues related to word properties and word relations: (1) We address the task of modelling word meaning by empirical features in data-intensive lexical semantics. Relying on large-scale corpus-based resources, we identify the contextual categories and functions that are activated by the associates and therefore contribute to the salient meaning components of individual words and across words. As a result, we discuss conceptual roles and present evidence for the usefulness of co-occurrence information in distributional descriptions. (2) We assume that semantic associates provide a means to investigate the range of semantic relations between words and contexts, and we provide insight into which types of semantic relations are treated as important or salient by the speakers of the language.

Key words: association norms, semantic associates, semantic relations, data-intensive semantics, distributional features, corpus co-occurrence, lexical resources

1. Motivation

A collection of semantic associates is used as the basis for an empirical characterisation of verb and noun properties. We define *semantic associates* here as those words spontaneously called to mind by a stimulus word, and assume that these evoked words reflect highly salient linguistic and conceptual features of the stimulus word. Given this assumption, identifying the types of information provided by speakers and distinguishing and quantifying the relationships between stimulus and response can serve a number of purposes for creating NLP resources and defining and applying NLP techniques.

** The original publication is available at www.springerlink.com, doi 10.1007/s11168-008-9048-4.

Within this article, we concentrate on two issues related to word properties and word relations. First, we address the task of *modelling word meaning by empirical features*: in data-intensive lexical semantics, it is necessary to empirically define and induce features that (a) capture the various meaning aspects of the words to be described, and (b) can be obtained automatically from corpus-data, in order to be able to determine the similarity or dissimilarity of words, sentences, paragraphs, or even documents. Progressing from the word level to the document level, examples for this task are: clustering of similar words (Pereira et al., 1993; Lin, 1998a; Merlo and Stevenson, 2001; Schulte im Walde, 2006), word sense discrimination (Schütze, 1998), the identification of multi-word expressions (Lin, 1999) and their decomposability (Baldwin et al., 2003), anaphora resolution (Poesio et al., 2002), and text indexing (Deerwester et al., 1990), among others. Generally, semantic features are not readily available.[1] Following the *distributional hypothesis*, namely that 'each language can be described in terms of a distributional structure, i.e., in terms of the occurrence of parts relative to other parts' (Harris, 1968), distributional descriptions have been applied to model aspects of word meaning. Specifically, contextual features such as words co-occurring in a document, in a context window, or with respect to a word-word relationship, such as syntactic structure, syntactic and semantic valency, etc. have been used. However, these prior investigations of distributional similarity have either focused on a specific word-word relation to induce features (such as Pereira et al. (1993), Rooth et al. (1999) referring to a direct object noun for describing verbs, and similarly Curran (2003) referring to subjects and direct objects), or used any dependency relation detected by the chunker or parser (such as Lin (1998a), McCarthy et al. (2003)). Little effort has been spent on investigating the eligibility of the various types of features. An exception to this are Padó and Lapata (2007) who started out with all syntactic functions obtained from a dependency parser, but allowed their representation to adopt the selection of functions with respect to their applications. We assume that semantic associates provide a useful means to identify the contextual functions that might be relevant to empirical feature descriptions, by examining which functions are activated by the associates and therefore contribute to the salient meaning components of individual words and across words.

Second, there are tasks where the notion of similarity goes beyond discriminating degrees of similarity: for many NLP resources and applications, it is crucial to define and use *semantic relations* between words or contexts. These tasks include the creation of lexical taxonomies (Fellbaum, 1998) and ontologies (Maedche and Staab, 2000; Navigli and Velardi, 2004; Kavalek and Svatek, 2005), anaphora resolution (Vieira and Poesio, 2000; Ji et al., 2005), text understanding, e.g., with respect to interpreting noun compounds (Lapata, 2002; Girju et al., 2007), question answering (Moldovan and Novischi, 2002; Girju, 2003; Girju et al., 2006), and textual entailment (Geffet and Dagan, 2005; Tatu and Moldovan, 2005). For example, when resolving bridging definite descriptions, Vieira and Poesio (2000) relied on the semantic relations between antecedent and anaphor (such as *house ... the build-*

ing), which were only partly covered by WordNet; Ji et al. (2005) exploited seven ACE relation types (such as 'physical': *a town south of Salzburg*) in co-reference prediction: if both the anaphor and a candidate antecedent participated in a semantic relation, the relation partners were checked on co-reference, in addition to the anaphor and antecedent themselves. Up to date, the research on semantic relations has either concentrated on a small set of relation types, such as hypernymy (Hearst, 1998), subject- vs. object-hood in noun compounds (Lapata, 2002), causal relation (Girju, 2003), or part-whole relation (Berland and Charniak, 1999; Girju et al., 2006), or developed an individual scheme of semantic relations. Accordingly, Girju et al. (2007) remarked with respect to semantic relations between nominals, that 'there is little consensus on the relation sets and algorithms for analysing semantic relations, and it seems unlikely that any single scheme could work for all applications'. For example, Nastase (2003) presented a two-level hierarchy for classifying noun-modifier relations, Moldovan et al. (2004) and Girju et al. (2005) proposed a classification with 35 nominal semantic relations, and Chklovski and Pantel (2004) addressed five different verb-verb relations. Pantel and Pennacchiotti (2006) are an exception in that they refrained from defining semantic relation types, but presented a bootstrapping algorithm that takes a few seed instances of a particular semantic relation as input and iteratively learns surface patterns to extract more instances. In conclusion, there is still the need for resources that support the specification of the range of relations. We assume that semantic associates provide a useful means to investigate the range of semantic relations; we concentrate on inter-categorical semantic relations, i.e., we investigate verb-verb and noun-noun relations. As above, by examining the types of relations that are captured by semantic associations, we can gain insights into which types of semantic relations are treated as important or salient by the speakers of the language.

The basis for the current investigation is provided by a collection of semantic associates evoked by German verbs and nouns. A series of analyses are performed on this database, to explore the relationships between the stimulus and the response words. Each analysis is motivated by its potential NLP uses, and the analyses are based on available resources with respect to the semantic investigation. As manually linking each stimulus-associate pair to a particular relationship would be time-intensive and subjective, we rely on large-scale lexicographic databases and on empirical, corpus-based resources that have the potential to characterise the associations.

Our work is in line with recent discussions that relate the computational modelling of language to human data, cf. Daelemans (2006). I.e., we argue that language data as collected from humans represents an excellent, if not an optimal, source of information about language properties within the computational modelling of language, given that the data are gathered with materials and methods that are appropriate for the respective purpose.

2. Data Collection and Preparation

This section introduces our methods for collecting human associations to German verbs and nouns (Sections 2.1 and 2.2),[2] and a distributional representation of the data as stimulus-associate type frequencies (Section 2.3).

2.1. ASSOCIATES OF VERB STIMULI

The data collection of verb associations was performed as a web experiment, which asked native speakers to provide associations to German verbs.

2.1.1. *Material*

330 verbs were selected for the experiment. 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. *hassen* 'hate', *überraschen* 'surprise'), communication (e.g. *reden* 'talk', *beneiden* 'envy'), etc. Selecting verbs from different categories was only intended to ensure that the experiment covered a wide variety of verb types; the inclusion of any verb in any particular verb class was done in part with reference to prior verb classification work (e.g., Levin (1993)) but also on intuitive grounds. The stimulus verbs were divided randomly into 6 separate experimental lists of 55 verbs each. The lists were balanced for class affiliation and frequency ranges (0, 100, 500, 1000, 5000), such that each list contained verbs from each grossly defined semantic class, and had equivalent overall verb frequency distributions. The frequencies of the verbs were determined by a 35 million word newspaper corpus, cf. Section 3.2; the verbs showed corpus frequencies between 1 and 71,604.

2.1.2. *Procedure*

The experiment was administered over the Internet. Participants were first presented with written instructions for the experiment and an example item with potential responses. In the actual experiment, each trial consisted of a verb presented in a box at the top of the screen. All stimulus verbs were presented in the infinitive. 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.[3] Participants had 30 seconds per verb to type as many associations as they could. After this time limit, the program automatically advanced to the next verb.

2.1.3. *Participants*

299 native German speakers participated in the experiment, between 44 and 54 for each data set.

2.1.4. *Data*

In total, we collected 79,480 associate tokens distributed over 39,254 different response types. Each trial elicited an average of 5.16 associate responses with a range of 0-16. Each completed data set contains the list of stimulus verbs, paired with a list of associations in the order in which the participant provided them. Considering the first responses only, the norm comprises 15,788 tokens over 7,425 types. Participants continued to provide new response types at a fairly consistent rate across all response positions.

2.2. ASSOCIATES OF NOUN STIMULI

The data collection of noun associations was performed as an offline experiment, which asked native speakers to provide up to three associations to German nouns. The target objects were presented in two forms: Lexical stimuli consisted of the written name of the noun targets; pictorial stimuli consisted of the written names accompanied by black and white line drawings of the referred-to objects.

2.2.1. *Material*

409 German nouns referring to picturable objects were chosen as target stimuli. To ensure broad coverage, target objects represented a variety of semantic classes including animals (e.g. *Affe* 'monkey', *Schwein* 'pig'), plants (e.g. *Tulpe* 'tulip', *Baum* 'tree'), professions (e.g. *Lehrerin* 'teacher', *Jäger* 'hunter'), furniture (e.g. *Stuhl* 'chair', *Bett* 'bed'), vehicles (e.g. *Flugzeug* 'plane', *Zug* 'train'), tools (e.g. *Hammer* 'hammer', *Besen* 'broom'), etc.. For the pictorial version of the experiment, simple black and white line drawings of target stimuli were drawn from several sources, including the data by (Snodgrass and Vanderwart, 1980) and the picture database from the Max Planck Institute for Psycholinguistics in the Netherlands.

2.2.2. *Procedure*

The 409 target stimuli were divided randomly into three separate questionnaires consisting of approximately 135 nouns each. Each questionnaire was printed in two formats: target objects were either presented as pictures together with their preferred name (to ensure that associate responses were provided for the desired lexical item), or the name of the target objects was presented without a representative picture accompanying it. Next to each target stimulus three lines were printed on which participants could write up to three semantic associate responses for the stimulus. The order of stimulus presentation was individually randomised for each participant. Participants were instructed to give one associate word per line, for a maximum of three responses per trial. No time limits were given for responding, though participants were told to work swiftly and without interruption. Each version of the questionnaire was filled out by 50 participants, resulting in a maximum

of 300 data points for any given target stimulus (50 participants $\times$ 2 presentation modes $\times$ 3 responses).

2.2.3. *Participants*

300 German participants, mostly students from Saarland University, received either course credit or monetary compensation for filling out the questionnaire.

2.2.4. *Data*

In total, we collected 116,714 associate tokens distributed over 31,035 different response types; 39,727 associate tokens over 11,389 types, when considering the first responses only. Collected associate responses were entered into a database with the following additional information: For each target stimulus we recorded a) whether it was presented as a picture or in written form, and b) whether the name was a homonym (and thus likely to elicit semantic associates for multiple meanings). For each response type provided by a participant, we coded a) the order of the response, i.e., first, second, third, b) the part-of-speech of the response,[4] and c) whether the response was related to the intended, depicted meaning of the stimulus or to an alternative meaning. The database is freely accessible (Melinger and Weber, 2006).

2.3. DISTRIBUTIONAL REPRESENTATION

For the analyses to follow, we pre-processed all data sets in the following way: For each stimulus word, we quantified over all responses in the experiment, disregarding the order in which associates were provided and, for noun stimuli, the presentation type of the questionnaire. The result is a frequency distribution for the stimulus words, providing frequencies for each response type. The responses were not distinguished according to polysemic senses of the stimuli.

To illustrate the frequency distribution, Table I lists the 10 most frequent responses for the polysemous verb *klagen* ‘complain, moan, sue’ and Table II lists the 10 most frequent responses for the polysemous noun *Schloss* ‘caste, lock’.

3. Resources for Data Investigation

This section introduces the manual and empirical resources that contributed to the characterisation of the association norms: a) a German newspaper corpus, b) a statistical grammar model that was trained on the corpus data, and c) the semantic taxonomy *WordNet* and its German counterpart *GermaNet*.

Table I. Associate frequencies for example stimulus verb.

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 II. Associate frequencies for example stimulus noun.

Schloss 'castle, lock'		
Schlüssel	'key'	51
Tür	'door'	15
Prinzessin	'princess'	8
Burg	'castle'	8
sicher	'safe'	7
Fahrrad	'bike'	7
schließen	'close'	7
Keller	'cellar'	7
König	'king'	7
Turm	'tower'	6

3.1. CORPUS DATA

A German newspaper corpus from the 1990s was used for co-occurrence analyses between verb/noun stimuli and associate responses. The corpus contains approximately 200 million words of newspaper text from *Frankfurter Rundschau*, *Stuttgarter Zeitung*, *VDI-Nachrichten*, *die Tageszeitung (taz)*, *German Law Corpus*, *Donaukurier*, and *Computerzeitung*. In addition to the co-occurrence analyses, the corpus was used as training data for the statistical grammar model described in Section 3.2.

3.2. STATISTICAL GRAMMAR MODEL

Some of the quantitative data in the analyses to follow are derived from an empirical grammar model based on a German context-free grammar which paid specific attention to verb subcategorisation (Schulte im Walde, 2002). The grammar was lexicalised, and the parameters of the probabilistic version were estimated in an unsupervised training procedure, using 35 million words of the above German newspaper corpus. The trained grammar model provides empirical frequencies for word forms, part-of-speech tags and lemmas, and quantitative information on lexicalised rules and syntax-semantics bi-lexical head-head co-occurrences.

3.3. WORDNET / GERMANET

WordNet is a lexical semantic taxonomy developed at Princeton University (Miller et al., 1990; Fellbaum, 1998). The lexical database was inspired by psycholinguistic research on human lexical memory. The resource organises English nouns, verbs, adjectives and adverbs into classes of synonyms *(synsets)*, which are connected by lexical and conceptual relations. Words with several senses are assigned to multiple classes. The idea of WordNet has been transferred to other languages than English. The University of Tübingen is developing the German version of WordNet, *GermaNet* (Hamp and Feldweg, 1997; Kunze, 2000; Kunze, 2004), which we use for our data.

The GermaNet version from October 2001 contains 7,825 verbs and defines the paradigmatic semantic relations *synonymy, antonymy, hypernymy, hyponymy* as well as the non-paradigmatic relations *entailment, cause,* and *also see* between verbs or verb synsets. (*Also see* is an underspecified relation, which captures relationships other than the preceding ones. For example, *sparen* 'save' is related to *haushalten* 'budget' by *also see*.) With respect to nouns, the GermaNet version contains 36,601 nouns and defines the paradigmatic semantic relations *synonymy, antonymy, hypernymy, hyponymy, holonymy, meronymy* as well as the non-paradigmatic relation *also see* between nouns or noun synsets.

4. Linguistic Analyses of Experimental Data

This section represents the main body of the article, providing a series of analyses that investigate step-wise the modelling of word meaning by empirical features: namely, a morpho-syntactic analysis in Section 4.1, an analysis of the syntax-semantic functions of the noun (stimuli/associates) with respect to the verb (associates/stimuli) in Section 4.2, and a co-occurrence analysis of the stimulus-response pairs in Section 4.3. A second series of analyses explores the semantic relations between stimuli and associates, in Section 4.4. Each analysis is performed separately for the verb stimuli and their associations, and for the noun stimuli and their associations. All of our analyses reported in this paper were based on response tokens; however, we also performed the respective type analyses, and they showed

the same overall pictures. Finally, as the association norms for verbs and nouns were collected in independent studies and therefore differ to some extent, this section uses all response tokens collected, but the following Section 5 provides the analyses only for the first responses, in order make the norms and analysis results more directly comparable.

Each analysis is structured in the same way: first, we introduce the motivation from Natural Language Processing, discussing why the respective analysis is relevant for NLP purposes; second, we present the analysis; third, we interpret the analysis' results.

4.1. MORPHO-SYNTACTIC ANALYSIS

The morpho-syntactic analyses of the response tokens distinguish and quantify the part-of-speech categories of the associate responses. On the one hand, this analysis can be considered as a preparatory step for the analyses to follow. In addition, the results will provide insight into the relevance of predominant part-of-speech categories with respect to meaning aspects. This knowledge is important in NLP tasks whenever words are represented by a choice of features that are supposed to model the word meaning, usually with the goal of determining the similarity or dissimilarity of words.

For example, the *vector space model* (Salton et al., 1975) uses words in documents to describe the contents of the respective documents. The model was originally designed for information retrieval (Salton and McGill, 1983), and has been generalised to describe not only documents, but also smaller structural units such as queries in question answering and individual words by co-occurring words. Often, the co-occurring words are restricted to content words, to certain part-of-speech categories, or even to a subset of words from a certain part-of-speech. With respect to a local perspective (i.e., co-occurrence within the near neighbourhood, such as the same sentence, or even the same phrase), the vector space model is related to the above mentioned *distributional hypothesis* and therefore the vector space model forms the basis for distributional descriptions.

Variants of the vector space model have been used in Latent Semantic Analysis for text indexing (Deerwester et al., 1990), word similarity (Landauer and Dumais, 1997) and particle verb compositionality (Baldwin et al., 2003; McCarthy et al., 2003); in NLP tasks and applications including word sense discrimination (Schütze, 1998), anaphora resolution (Poesio et al., 2002), thesaurus extraction (Lin, 1998b; Lin, 1999), general models of semantic similarity (Lin, 1998a; Sahlgren, 2006; Schulte im Walde, 2006; Padó and Lapata, 2007); and also in psycholinguistic models of semantic priming (Lund et al., 1995; Lowe and McDonald, 2000; Vigliocco et al., 2004).

4.1.1. *Analyses*

4.1.2. *Associates of Verb Stimuli*

Each response to the stimulus verbs was assigned its – possibly ambiguous – part-of-speech (*POS*) by our empirical grammar dictionary, cf. Section 3.2. Originally, the dictionary distinguished approx. 50 morpho-syntactic categories, but we disregarded fine-grained distinctions such as case, number and gender features and considered only the major categories verb (V), noun (N), adjective (ADJ) and adverb (ADV). A fifth category 'OTHER' comprises all other part-of-speech categories such as particles, interjections, conjunctions, etc. Ambiguities between the categories arose e.g. in the case of nominalised verbs (such as *Rauchen* 'smoke', *Vergnügen* 'please/pleasure'), where the experiment participant could have been referring either to a verb or a noun, or in the case of past participles (such as *verschlafen* 'sleep' or 'drowsy') and infinitives (such as *überlegen* 'think about' or 'superior'), where the participant could have been referring either to a verb or an adjective). In total, 3.6% of all response types were ambiguous between multiple part-of-speech tags.

Having assigned part-of-speech tags to the responses, we were able to distinguish and quantify the morpho-syntactic categories of the responses. In non-ambiguous situations, the unique part-of-speech received the total stimulus-response frequency; in ambiguous situations, the stimulus-response frequency was split over the possible part-of-speech tags. The output of this analysis is the frequency distributions of the part-of-speech tags for each verb individually, and also as a sum over all verbs. Table III presents the total numbers and specific verb examples. Participants provided noun associates in the clear majority of token instances, 62%; verbs were given in 25% of the responses, adjectives in 11%, adverbs almost never (2%). The table also shows that the POS distributions vary across the semantic classes of the verbs. For example, aspectual verbs, such as *aufhören* 'stop', received more verb responses, $t(12)=3.11$, $p<.01$, and fewer noun responses, $t(12)=3.84$, $p<.002$, than creation verbs, such as *backen* 'bake'.

4.1.3. *Associates of Noun Stimuli*

In contrast to the analysis of the verb data, the part-of-speech categories of the associate responses to noun stimuli were entered manually into the association database, cf. Section 2.2. The coding distinguished the three major categories verbs (V), nouns (N), adjectives (ADJ), and in addition proper names (PN). As for the associations to verbs, a fifth category 'OTHER' comprises all other part-of-speech categories such as particles, interjections (such as *igitt* 'ugh' for food nouns), numbers, and onomatopoeia (such as *wau-wau* 'woof-woof' for *Dackel* 'dachshund'). As in the verb analysis, we then determined the frequency distributions of the part-of-speech tags for each noun individually, and also as a sum over all nouns. Table IV presents the total numbers and specific noun examples.

Table III. Part-of-speech tag distributions of verb responses.

	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%
bedrohen 'threaten'	12%	75%	12%	0%
bemerken 'realise'	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%

Table IV. Part-of-speech tag distributions of noun responses.

	ADJ	N	PN	V
TOTAL FREQ	19,075	80,419	3,147	13,905
TOTAL PROB	16%	69%	3%	12%
Ananas 'pineapple'	45%	51%	3%	1%
Daumen 'thumb'	15%	71%	1%	11%
Esel 'donkey'	45%	42%	4%	6%
Hamburger 'hamburger'	14%	57%	24%	5%
Kopf 'head'	5%	89%	0%	6%
Löffel 'spoon'	6%	86%	0%	8%
Mund 'mouth'	11%	65%	0%	34%
Schildkröte 'turtle'	50%	44%	3%	3%
Tempel 'temple'	13%	58%	24%	5%
Telefon 'telephone'	4%	53%	2%	41%
Wecker 'alarm clock'	22%	42%	0%	36%
Zwiebel 'onion'	15%	54%	0%	31%

As for the verb stimuli, participants provided noun associates in the clear majority of token instances, 69%; adjectives were given in 16% of the responses, verbs in 12%, and proper names in 3%. Again, the table also shows that the POS distributions vary with respect to the individual noun stimuli. For example, nouns referring to food or animals enforced a stronger usage of adjectives, such as *Ananas*

– *gelb, süß, lecker* 'pineapple – yellow, sweet, tasty', or *Schildkröte – langsam, alt, grün* 'turtle – slow, old, green' than other nouns, $t(407)=51.3$, $p<.001$. Similarly, nouns referring to natural objects evoked more adjectives, $t(407)=46.8$, $p<.001$, and fewer noun responses, $t(407)=6.5$, $p<.02$, than nouns referring to man-made objects.

4.1.4. *Interpretation*

The morpho-syntactic analyses demonstrate that nouns play a major role among verb and noun features. This insight corresponds to the predominant use of nominal features in distributional descriptions that address the semantic modelling of words for various purposes as well as frequency of occurrence in corpora. However, the analyses also showed that the relevance of the part-of-speech categories with respect to meaning aspects varies according to the semantic class of the word to model. We conclude that nouns are important for distributional descriptions, but other features than nouns should also be relevant in modelling word meaning. This insight should have an impact on the choice of feature categories in distributional representations; restricting the categories to nominal features restricts the feature sets to those features that are relevant for the average of words, but they do not necessarily cover the meaning aspects of all semantic word classes.

4.2. SYNTAX-SEMANTIC NOUN FUNCTIONS

The analyses in this section continue exploring the eligibility of various types of features for modelling word meaning, now concentrating on the conceptual roles of nouns. As explained in the Introduction, most previous work on distributional similarity that used nominal features within distributional descriptions has either focused on a specific word-word relation to induce features (such as Pereira et al. (1993), Rooth et al. (1999), Curran (2003)), or used any dependency relation detected by the chunker or parser (such as Lin (1998a), McCarthy et al. (2003), Korhonen et al. (2003), Schulte im Walde (2006)). Little effort has been spent on investigating the eligibility of the various types of nominal features. Even though the use of the distributional features depends on the respective applications, we aim to identify prominent roles for distributional verb descriptions by evaluating which functional roles are highlighted by verb-noun pairs. For these analyses, we assume that the noun responses to verb stimuli and verb responses to noun stimuli relate to conceptual roles required by the verbs. Thus, we investigate the linguistic functions that are realised by the response nouns with respect to the stimulus verbs, and by the stimulus nouns with respect to the response verbs. The analyses are based on our empirical grammar model, cf. Section 3.2.

4.2.1. *Analyses*

4.2.2. *Associates of Verb Stimuli*

With respect to verb subcategorisation, the empirical grammar model offers 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 240 times in our training corpus. In 80 of these instances it was parsed as intransitive, and in 109 instances it was parsed as transitive subcategorising for a direct object. The most frequent nouns subcategorised for as direct objects in the grammar model were *Brötchen* 'rolls', *Brot* 'bread', *Kuchen* 'cake', *Plätzchen* 'cookies', and *Waffel* 'waffle'. We used the grammar information to look up the syntactic relationships which existed between a stimulus verb and a response noun. For example, the nouns *Kuchen* 'cake', *Brot* 'bread', *Pizza* and *Mutter* 'mother' were produced in response to the stimulus verb *backen* 'bake'. The grammar look-up told us that *Kuchen* 'cake' and *Brot* 'bread' appeared not only as the verb's direct objects (as illustrated above), but also as intransitive subjects; *Pizza* only appeared as a direct object, and *Mutter* 'mother' only appeared as transitive subject. The verb-noun relationships which were found in the grammar were quantified by the verb-noun association frequency, taking into account the number and proportions of different relationships (to incorporate the ambiguity represented by multiple relationships). For example, the noun *Kuchen* was elicited 45 times in response to *backen*; the grammar contained the noun both as direct object and as intransitive subject for that verb. Of the total association frequency of 45 for *Kuchen*, 15 would be assigned to the direct object of *backen*, and 30 to the intransitive subject if the empirical grammar evidence for the respective functions of *backen* were one vs. two thirds.

In a following step, we accumulated the association frequency proportions with respect to a specific relationship, e.g., for the direct objects of *backen* 'bake' we summed over the frequency proportions for *Kuchen*, *Brot*, *Plätzchen*, *Brötchen*, etc. The final result was a frequency distribution over linguistic functions for each stimulus verb, i.e., for each verb we determined which linguistic functions were activated by how many noun associates. For example, the most prominent functions for the inchoative-causative verb *backen* 'bake' were the transitive direct object (8%), the intransitive subject (7%) and the transitive subject (4%); for the object-drop verb *schreiben* 'write' we found 11% for the transitive direct object, 3% and 4% for the intransitive and the transitive subject, respectively, and evidence for the writing instrument (the PP headed by *mit* 'with' in various frames with a total of 10%).

By generalising over all verbs, we discovered that only 10 frame-slot combinations were linked to at least 1% of the noun tokens:[5] subjects in the intransitive frame and the transitive frame (with accusative/dative object, or prepositional phrase); the accusative object slot in the transitive, the ditransitive frame and the direct object plus PP frame; the dative object in a transitive and ditransitive frame,

and the prepositional phrase headed by *Dat:in* (dative (locative) 'in'). The frequency and probability proportions[6] are illustrated in Table V; the function is indicated by a slot within a frame (with the relevant slot in bold font); 'S' is a subject slot, 'AO' an accusative object, 'DO' a dative object, and 'PP' a prepositional phrase.

Table V. Associates as nominal slot fillers.

	Function	Freq	Prob
S	**S** V	1,792	4%
	S V AO	1,040	2%
	S V DO	265	1%
	S V PP	575	1%
AO	S V **AO**	3,124	6%
	S V **AO** DO	824	2%
	S V **AO** PP	653	1%
DO	S V **DO**	268	1%
	S V AO **DO**	468	1%
PP	S V **PP-Dat:in**	487	1%
Total (of these 10)		9,496	19%
Total found in grammar		13,527	28%
Unknown verb or noun		10,964	22%
Unknown function		24,250	50%

4.2.3. *Associates of Noun Stimuli*

Parallelling the preceding analysis, we checked whether any of the noun-verb relationships were found in our statistical grammar model. In the positive cases, the relationships were quantified by the noun-verb association frequency, again taking into account the number and proportions of the various grammar functions. The most prominent functions are listed in Table VI.

The table shows that – to a large extent – the most prominent functions for the noun-verb pairs are the same as for the verb-noun pairs: subjects in the intransitive frame and the transitive frame (with accusative object, or prepositional phrase); the accusative object slot in the transitive, the ditransitive, and the AO plus PP frame; the dative object in the transitive and the ditransitive frame.

Table VI. Stimuli as nominal slot fillers.

	Function	Freq	Prob
S	**S** V	1,095	8%
	S V AO	300	2%
	S V PP	406	3%
	S V C-2	103	1%
	S V INF	71	1%
AO	S V **AO**	1,480	11%
	S V **AO** DO	206	1%
	S V **AO** PP	218	2%
DO	S V **DO**	144	1%
	S V AO **DO**	99	1%
PP	S V **PP-Dat:auf**	263	2%
	S V **PP-Dat:in**	193	1%
Total (of these 12)		4,578	33%
Total found in grammar		5,661	41%
Unknown verb or noun		1,505	11%
Unknown function		6,712	48%

4.2.4. *Interpretation*

In total, only 28% of all verb-noun pairs (i.e., noun responses to verb stimuli) and 41% of all noun-verb pairs (i.e., verb responses to noun stimuli) were identified by the statistical grammar as a filler for any slot in any of the 178 identified frames (which corresponds to a total of 592 frame-slot combinations).[7] The majority of pairs was not found as slot fillers: 22/11% of the stimulus-associate pairs (marked as 'unknown verb or noun' in Tables V and VI) were missing because either the verb or the noun did not appear in the grammar model at all. These cases were due to (i) lemmatisation in the empirical grammar dictionary, where noun compounds such as *Autorennen* 'car racing' were lemmatised by their lexical heads, creating a mismatch between the full compound and its head; (ii) multi-word expressions among the associates, like *Zähne putzen* 'brush teeth' or *frisch machen* 'refresh'; (iii) domain of the training corpus, which underrepresented slang responses like *Grufties* 'old people' and *lümmeln* 'loll', dialect expressions such as *Ausstecherle* 'cookie-cutter' and *heimfahren* 'go home', as well as technical expressions such as *Plosiv* 'plosive'; and (iv) size of the corpus data: the whole newspaper corpus of 200 million words contained more than 99% of the stimuli and the associates in the two analyses; the 35 million word partition on which the grammar model was

trained contained still more than 99% of the verb stimuli/associates, but only 78% of the noun associates to the verb stimuli, and only 90% of the noun stimuli.

The 50/48% of the nouns/verbs which are marked as 'unknown function' in Tables V and VI were present in the grammar but did not fill subcategorised-for linguistic functions; clearly the conceptual roles of the noun associates were not restricted to the subcategorisation of the stimulus verbs.

Although direct object and subject roles are prominent among the verb-noun relationships, they are also highly frequent in the grammar model as a whole. In fact, across all possible frame-slot combinations, we found an extremely strong correlation between the frequency of a frame-slot combination in the grammar model and the number of responses that link to that frame-slot combination in our data, $r(592)=.87$, $p<.001$ for the noun responses to verbs, and $r(592)=.92$, $p<.001$ for the verb responses to nouns. Thus, the direct object and subject roles are not over-represented in our data; they are represented proportionate to their frequency in the grammar. Therefore, we cannot conclude from the tables that specific functions within distributional representations are dominant – as we had hypothesised originally – and should be recommended.

Furthermore, contrary to our initial assumptions, the majority of nouns in verb-noun pairs did not reflect grammatical functions of the respective verbs. In part what was or was not covered by the grammar model can be characterised as an argument/adjunct contrast. The grammar model distinguishes argument and adjunct functions, and only arguments are included in the verb subcategorisation and were therefore found as linguistic functions. Adjuncts such as the instrument *Pinsel* 'brush' for *bemalen* 'paint', *Pfanne* 'pan' for *erhitzen* 'heat', or clause-internal information such as *Aufmerksamkeit* 'attention' for *bemerken* 'notice' and *Musik* 'music' for *feiern* 'celebrate' were not found. Similarly, verbs provided as associates for their respective instruments, e.g. *trocknen* 'dry' for *Handtuch* 'towel', *biegen* 'bend' for *Zange* 'pincer', or providing world knowledge, e.g. *streichen* 'paint' for *Klebeband* 'tape', *schlafen* 'sleep' for *Kissen* 'cushion', *riechen* 'smell' for *Nase* 'nose' were also not found. These nouns fulfil scene-related roles or represent world knowledge, and were not captured by subcategorisation in the grammar model. The analyses therefore illustrated that the noun stimuli/responses were not restricted to verb subcategorisation role fillers, and that clause-internal adjuncts as well as clause-external, scene-related information or world knowledge should also play a role when using nominal features in distributional descriptions of word meaning.

4.3. CO-OCCURRENCE ANALYSIS

The motivation for the last set of analyses on word meaning features arose from our syntax-semantics analyses in the previous section, which demonstrated that there were verb-noun pairs within the association norms which might co-occur in local contexts even if they were not related by a subcategorisation function.

In more general terms, we were interested in the role of co-occurrence information within an empirical distributional description. It is commonly assumed that human associations reflect word co-occurrence probabilities, cf. McKoon and Ratcliff (1992), Plaut (1995); this assumption was supported by observed correlations between associative strength and word co-occurrence in language corpora (Spence and Owens, 1990; Rapp, 1996). Our analyses examined whether the co-occurrence assumption holds for our (much larger) German association data, i.e., which proportion of the associations were found in co-occurrence with the stimulus words. A positive outcome of these analyses might encourage the use of low-level co-occurrence information in corpus-based word descriptions.

4.3.1. *Analyses*

4.3.2. *Associates of Verb Stimuli*

The analysis used our complete newspaper corpus, 200 million words, and checked whether the associate responses occurred in a window of 20 words to the left or to the right of the relevant stimulus word.[8] We determined the co-occurrence strength between the stimulus verbs and their associations, the results are presented in Table VII. The 'all' row shows the percentage of associate responses that were found in co-occurrence with their stimulus verbs at least once, or twice, or 3/5/10/20/50 times. The co-occurrence proportions are rather high, especially when taking into account the restricted domain of the corpus. For example, for a co-occurrence strength of 3 we find two thirds of the associations covered by the 20-word window in the corpus data.

Table VII. Verb-association co-occurrence in 20-word window.

	Co-Occurrence Strength						
POS	1	2	3	5	10	20	50
all	77	70	66	59	50	40	27
V	79	71	67	60	50	41	29
N	76	69	66	59	50	40	27
ADJ	77	69	64	57	45	36	22
ADV	91	88	85	80	72	62	50

The following rows are specified for their POS, verbs 'V', nouns 'N', adjectives 'ADJ', and adverbs 'ADV'. The proportions of verb, noun and adjective responses which were found in co-occurrence with their stimulus verbs are very similar to

the overall proportions. The 'ADV' co-occurrence strengths stand out in Table VII: they represent only 2% of all response tokens, but the analysis shows they exhibit a much stronger co-occurrence behaviour to the verbs than the other POS. Chi-squared analyses confirm that adverbs co-occurred with target verbs at least once in the 20-word window more than is proportionally expected, $\chi^2(1)=16.93$, $p<.001$. No other comparisons (of this co-occurrence strength) were significantly different from their expected frequencies.

4.3.3. *Associates of Noun Stimuli*

The co-occurrence analysis for the associates of noun stimuli was conducted exactly as for the verbs. Table VIII presents the results. Again, the proportions of verb, noun and adjective responses which were found in co-occurrence with their stimulus nouns are very similar to the overall proportions, with the verb proportions above, $\chi^2(1)=9.15$, $p<.003$, and the adjective proportions slightly below, $\chi^2 < 1$, the overall co-occurrence values. Verbs occurred more often then expected. Furthermore, all co-occurrence values are between 6-9% above the co-occurrence values of the verb analysis.

Table VIII. Noun-association co-occurrence in 20-word window.

	Co-Occurrence Strength						
POS	1	2	3	5	10	20	50
all	84	77	72	64	52	38	23
V	88	82	77	69	57	44	28
N	84	78	72	65	53	39	23
ADJ	83	76	70	63	50	36	20

4.3.4. *Interpretation*

Our analyses showed that the co-occurrence assumption holds for our German association data, to a large extent: 77/84% of our response tokens were covered at least once in a 20-word window of the stimulus words, approximately two thirds were covered at least three times, and even approximately 40% were covered at least 20 times. These results suggest that co-occurrence information is an integral component for empirical descriptions of word properties, an important insight since co-occurrence information is essentially less expensive (because no high-level preprocessing is necessary) and therefore easier to obtain than annotated data. Thus

co-occurrence information could be especially valuable for languages with few NLP resources available.

Furthermore, comparing the co-occurrence strength of nominal responses with the proportions of the nouns that were found as subcategorised by the respective verbs (cf. Tables V and VI) demonstrates once more that verb subcategorisation accounts only for a part of the nominal responses, and therefore only for a subset of the verb concepts represented by nouns; but more general scene-related information beyond the clause level is captured by corpus co-occurrence.[9]

Examples of associations that did not appear in co-occurrence with the respective stimulus verbs are *nass* 'wet' for *nieseln* 'drizzle', *lecker* 'yummy' for *mampfen* 'munch', *Wasser* 'water' for *auftauen* 'defrost', *Freude* 'joy' for *überraschen* 'surprise', or *Verantwortung* 'responsibility' for *leiten* 'guide'. Correspondingly, examples of associations that did not appear in co-occurrence with the respective stimulus nouns are *gelb* 'yellow' for *Ananas* 'pineapple', *kalt* 'cold' for *Iglu* 'igloo', *Überraschung* 'surprise' for *Geschenk* 'present', *Weihnachten* 'Christmas' for *Walnuß* 'walnut', *Physik* 'physics' for *Magnet* 'magnet', and *Herbst* 'autumn' for *Drachen* 'kite'. These associations reflect world knowledge rather than clause-internal/-external scene-related information, and are therefore not expected to be found in the immediate context of the stimuli at all. These cases pose an interesting challenge to empirical models of word meaning. Higher-order co-occurrence models as suggested by e.g. Lemaire and Denhière (2006) should be able to capture more of these cases than our first-order co-occurrence model, but we believe that it is not surprising that world knowledge is not entirely captured by corpus data. The association analyses illustrated that, as a consequence, empirical features that model world knowledge are missing in distributional word meaning descriptions.

Finally, comparing the overall co-occurrence strength of associates with those of specific part-of-speech categories demonstrates that the co-occurrence information for some categories is more easily available than for others. For example, the verb association analysis showed that adverbs play a major role for verbs in the corpus proximity. We can explain this finding by the token-type ratio of adverbs within a corpus: even though adverbs are an open class in German, the absolute number of adverb types in a corpus is much lower than those of verbs, nouns, and adjectives, but at the same time the token-per-type ratio is higher, and the grammatical restrictions within German clause structure are lower. So there is a high prior probability to find one or more adverbs in the vicinity of a verb, especially within a large co-occurrence window. However, adverbs that appear in a large corpus distance to a verb are not very likely to contribute to the meaning of the verb, but rather to the meaning of the verb in the respective clause. To sum up, the relatively large proportions of adverbs in a large co-occurrence window are not expected to contribute greatly to a meaning representation of verbs, but such a meaning representation could capitalise from adverbs in a close vicinity of verbs, for which the proportion of co-occurrence is also enormously high. Thus, we suggest to restrict adverb co-occurrence to small window sizes. (Schulte im

Walde and Melinger, 2008) confirmed these insights on a more general basis for all four POS types, with respect to the verb-association data: They showed that the co-occurrence distributions across the POS types correspond to a large extent to the prior distributions of verbs and V/N/ADJ/ADVs. Establishing a baseline estimated by the co-occurrence rate of unrelated words reversed the picture in Table VII: adverb and noun responses were found in co-occurrence with the verb stimuli less than on average, and verb and adjective responses were found more than on average. Taking these findings into account once more emphasises the importance of including other features besides nominal features into distributional descriptions.

4.4. SEMANTIC RELATIONS

In contrast to the previous analyses of associates, this final set of analyses is concerned with the types of relationships between the stimulus words and the associate responses. As illustrated in the Introduction, defining and using semantic relations is crucial for many NLP resources, tasks and applications, including lexical taxonomies (Fellbaum, 1998) and ontologies (Maedche and Staab, 2000; Navigli and Velardi, 2004; Kavalek and Svatek, 2005), anaphora resolution (Vieira and Poesio, 2000; Ji et al., 2005), text understanding (Lapata, 2002; Girju et al., 2007), question answering (Moldovan and Novischi, 2002; Girju, 2003; Girju et al., 2006), and textual entailment (Geffet and Dagan, 2005; Tatu and Moldovan, 2005). However, most work on semantic relations so far has concentrated on a small set of relation types, such as hypernymy (Hearst, 1998), subject- vs. object-hood in noun compounds (Lapata, 2002), causal relation (Girju, 2003), or part-whole relation (Berland and Charniak, 1999; Girju et al., 2006), or developed an individual scheme of semantic relations that was relevant for the respective purposes, e.g., (Nastase, 2003; Chklovski and Pantel, 2004; Moldovan et al., 2004; Girju et al., 2005). Thus, there is still the need for resources with a potential to support the specification of the range of relations.

This paper performs an analysis on verb-verb and noun-noun relations,[10] as derived from the association norms. We suggest that an analysis of human associations may identify the range of semantic relations which are crucial in NLP applications; stimulus-response pairs not covered by an existing taxonomy can help to detect missing links, and provide an empirical basis for defining additional relations.

4.4.1. *Analyses*

4.4.2. *Associates of Verb Stimuli*

We looked up the semantic relation between each verb stimulus and verb response using the lexical semantic taxonomy GermaNet, cf. Section 3.3, to distinguish between the different kinds of responses to verb stimuli elicited from speakers. Our analysis proceeded as follows. For each pair of stimulus and response verbs,

we looked up whether any kind of semantic relation was defined between any of the synsets the verbs belong to. For example, if the stimulus verb *rennen* 'run' was in synsets a and b, and the response verb *bewegen* 'move' was in synsets c and d, we determined whether there was 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 were considered synonymous; two verbs sharing a common hypernym were considered co-hyponyms. The semantic relations were then quantified by the stimulus-response verb frequencies, e.g. if 12 participants provided the association *bewegen* for *rennen*, the hypernymy relation was quantified by the frequency 12. Table IX shows the number of semantic relations encoded in our GermaNet version,[11] and the frequencies and probabilities of our response tokens found among them.[12] For example, there were 19,424 verb-verb instances (belonging to 9,275 verb synset combinations) where GermaNet defines a hypernymy-hyponymy relation between their synsets; for 1,343 of our verb-verb pairs (verb response tokens with respect to stimulus verbs) we found a direct hypernymy relation among the GermaNet definitions, which accounts for 9% of all our verb responses; for 540 cases we found an indirect hypernymy relation (i.e., with a path length >1 between the synsets). If the stimulus and the response verb were both in GermaNet, but there was no relation between their synsets, then the verbs did not bear any kind of semantic relation ('no relation'), according to GermaNet's current status. If either of them was not in GermaNet, we could not make any statement about the verb-verb relationship ('unknown cases').

Table IX. Semantic verb-verb relations.

	GermaNet		Freq	Prob
Synonymy		4,633	762	4%
Antonymy	226	571	209	1%
Hypernymy	9,275	19,424	1,343	7%
(indirect)			540	3%
Hyponymy	9,275	19,424	1,702	9%
(indirect)			514	3%
Co-Hyponymy	55,122	102,018	2,232	12%
(indirect)			1,517	8%
Cause	95	236	40	0%
Entailment	8	15	0	0%
Also see	1	2	0	0%
Total found in GermaNet			8,859	47%
No relation			7,841	41%
Unknown cases			2,207	12%

The distribution of stimulus-response relations is correlated with stimulus verb frequency. The proportion of associate responses captured by the respective relations of synonym, antonym and hyponym increases as a function of stimulus verb frequency, $r(323)=.147$ for synonymy, $r(328)=.341$ for antonymy and $r(328)=.243$ for hyponymy (all $p<.01$); the proportion of hypernym relations is not correlated with verb frequency. The distribution of relations also varies by verb class. For example, aspectual stimulus verbs like *aufhören* 'stop' received significantly more antonymic responses like *anfangen* 'begin' or *weitermachen* 'go on' than creation verbs such as *backen* 'bake', $t(12)=3.44$, $p<.05$.

An interesting piece of information is provided by the verb-verb pairs for which we did not find a relationship in GermaNet. The minority of such cases (12%) is due to either the stimulus or the response not coded in GermaNet. Most of these cases are based on part-of-speech confusion because the participants did not consistently use capitalisation, cf. Section 2; e.g. the non-capitalised noun *wärme* 'warmth' was classified as a verb because it is the imperative of the verb *wärmen* 'warm'; in addition, the responses unknown to GermaNet include (low-frequent) particle verbs, which are highly productive in German.

A remarkable number of verb-verb associations (41%) did not show any kind of semantic relation according to GermaNet despite both verbs appearing in the taxonomy. A detailed manual inspection of the data revealed that on the one hand, this is partly due to the GermaNet taxonomy not being finished; we found verb associates such as *weglaufen* 'run away' for *abhauen* 'walk off', or *untersuchen* 'examine' for *analysieren* 'analyse' where we assume (near) synonymy not yet coded in GermaNet; or *gehen* 'leave' for *bleiben* 'stay', *frieren* 'be cold' for *schwitzen* 'sweat' where we assume antonymy not yet coded in GermaNet. However, a large proportion of the "no relation" associations represent instances of verb-verb relations not targeted by GermaNet. For example, *adressieren* 'address' was associated with the temporally following *schicken* 'send'; *schwitzen* 'sweat' was associated with a consequence *stinken* 'stink' and with a cause *laufen* 'run'; *erfahren* 'get to know' with the implication *wissen* 'know'.

4.4.3. *Associates of Noun Stimuli*

As in the verb analysis, we looked up the semantic relations between the noun stimuli and noun responses in GermaNet. The results of the analysis are displayed in Table X.

The distribution of stimulus-response relations is again correlated with stimulus noun frequency, but the relationship between noun frequency and semantic relations is different for nouns than it was for verbs (where the proportion of associate responses captured by synonymy, antonymy and hyponymy increased as a function of stimulus verb frequency). Specifically, as the frequency of the noun target increased, so did the number of associates that reflect a direct hypernym, $r(409)=.190$, $p<.000$, a direct or indirect hyponym, $r(409)=.230$, $p<.000$ for direct and $r(409)=.101$, $p<.05$ for all, a direct co-hyponym, $r(409)=.135$, $p<.006$, or a

Table X. Semantic noun-noun relations.

	GermaNet		Freq	Prob
Synonymy		18,992	533	1%
Antonymy	478	1,553	33	0%
Hypernymy	30,707	82,685	1,387	2%
(indirect)			2,365	3%
Hyponymy	30,708	82,829	714	1%
(indirect)			289	0%
Co-Hyponymy	302,755	575,585	3,584	4%
(indirect)			2,964	4%
Holonymy	3,995	8,625	579	1%
(indirect)			102	0%
Meronymy	3,998	8,625	1,171	1%
(indirect)			224	0%
Also see	892	2,670	84	0%
Total found in GermaNet			14,028	17%
No relation			52,814	66%
Unknown cases			13,543	17%

direct meronym, $r(409)=.134$, $p<.007$, relationship. No relationship was negatively related. As with the verb frequency by relationship correlations, the amount of variance captured by any one relationship is very small, with 5% being the strongest relationship observed for the nouns. The distribution of relations did not vary as regularly as they did for the verb relations. Comparing the types of relations elicited by noun stimuli referring to naturally occuring objects vs. noun stimuli referring to artifacts, only the number of direct hypernyms differed significantly, $t(354)=2.04$, $p<.05$. Specifically, naturally-occurring target nouns such as *Kind* 'child' and *Berg* 'mountain' elicited more hypernymic associates than man-made target nouns such as *Spiegel* 'mirror' or *Messer* 'knife'. Comparing manipulable man-made objects such as tools and musical instruments to non-manipulable man-made objects such as furniture and building parts revealed no differences in the distribution of relation types.

Again, we take a deeper look into why noun-noun pairs are not covered by GermaNet relations. The 17% unknown cases are due to mistakes in capitalisation (e.g., *Einkaufen* 'shop' is a nominalised verb and therefore could not be found as a noun in GermaNet), missing regional expressions (e.g., *Weck* 'roll'), proper names (e.g., *Moses*), and noun compounds. 66% of the noun-noun associations did not show any kind of semantic relation according to GermaNet despite both nouns appearing in the taxonomy. This is partly due to the GermaNet taxonomy

not being finished; of this, however, we found only few instances, such as *Schiff* 'ship' for *Anker* 'anchor', *Stachel* 'spike' for *Kaktus* 'cactus', or *Bein* 'leg' for *Spinne* 'spider', where we assume holonymy/meronymy not yet coded in GermaNet. Assuming that noun compounds might play a role, a post-check on noun-noun combinations looked up whether (a) the association represents a compound noun including the stimulus, or (b) the compound consisting of the two nouns was found encoded in GermaNet, and showed that we actually found 12% of the noun-noun pairs as compounds, e.g., *Honig* 'honey' as association to *Melone* 'melon' (→*Honigmelone* 'cantaloupe'), *Obst* 'fruit' as association to *Schale* 'bowl' (→*Obstschale* 'fruit bowl'), and *Stange* 'pole' as association to *Gardine* 'curtain' (→*Gardinenstange* 'curtain pole'). This insight is interesting, because noun compounds are an extremely productive class in German, and therefore only a small part of them can be assumed to be captured by the taxonomy, so we had not expected to find such a large proportion of stimulus-associate pairs among them. Finally, a large proportion of the "no relation" associations represents instances of noun-noun relations not targeted by GermaNet. For example, *Kamel* 'camel' was provided a proto-typical location *Wüste* 'desert', and similarly *Krankenschwester* 'nurse' and *Krankenhaus* 'hospital', *Wiege* 'cradle' and *Baby*, and *Jäger* 'hunter' and *Wald* 'forest'. In addition, *Gans* 'goose' was provided a typical event *Weihnachten* 'Christmas', and similarly *Schlitten* 'sledge' and *Schnee* 'snow', and *Eule* 'owl' and *Nacht* 'night'.

4.4.4. *Interpretation*

We identified paradigmatic WordNet relations for 47% of the verb-verb pairs and 17% of the noun-noun pairs. Covering about half of the verb-verb pairs demonstrates that the major paradigmatic semantic relations play an important role among the verb responses to target verbs. For the noun-noun pairs, the relative proportions of the major semantic relations are much lower. We believe, however, that this does *not* mean that the major relations are not important among the noun responses to target nouns. The absolute numbers of noun-noun relations are actually higher than those for the verb-verb relations. Rather, the major semantic relations are complemented by another type of major relation between the nouns, i.e., noun compounding. Recall that we found 12% of the noun-noun pairs among noun compounds (in addition to the 17% among the paradigmatic relations), which changes the 'total found' vs. 'no relation' proportions from 17/66% to 29/54%, even though one cannot expect GermaNet to cover the highly productive class of noun compounds. Thus, the actual proportion of noun-noun pairs representing noun compounds is expected to be considerably higher than 12%, and the actual proportion of noun-noun pairs covered by either paradigmatic relations or noun compounds is expected to be similar to the proportion of verb-verb pairs covered by paradigmatic relations only.

About half of the verb-verb and noun-noun stimulus-associate pairs could not be related via the taxonomy. This result demonstrates a) that there are missing

links in GermaNet, and our association data provides a useful starting point to enhance the taxonomy; and – more importantly – b) that relations other than those coded in GermaNet (such as temporal order, cause, and consequence for verb-verb pairs, and condition, instrument, and result for noun-noun pairs) are represented in a large proportion of the stimulus-associate pairs. These data could be subsumed under *also see* relations or *evocation* (cf. the related work in Section 6) in GermaNet, but it is obvious that one would prefer more fine-grained distinctions. We are specifically interested in those cases that are not covered by GermaNet, because we expect that human associations cover the range of possible semantic relations to a large extent. We therefore believe that they represent an excellent basis for defining an exhaustive set of semantic relations, as alternative to e.g. text-based relations (Morris and Hirst, 2004; Beigman Klebanov and Shamir, 2006), cf. Section 6 on related work. Such a definition is independent of the question whether the data should enhance GermaNet or not. Our intention to use GermaNet as a resource to identify semantic relations between verb-verb and noun-noun pairs was to investigate which proportions of the respective target-associate pairs are covered by the resource, and which existing relations are important. Additionally, we hoped to illustrate that there are semantic relations among the target-associate pairs that are not covered by the WordNet family, and that those relations represent a considerable proportion of the target-associate pairs. These missing relations still have to be defined, assuming that other than the mostly paradigmatic relations in WordNet are relevant for NLP uses.

5. Analyses of First Responses only

Early procedures for eliciting associates allowed participants to supply multiple responses to each stimulus. However, more recent protocols have opted for a discrete elicitation procedure, in which only a single response is provided. The shift towards a discrete procedure was partly due to concerns about association chain effects, i.e., that the nth response is associated to the $(n-1)$th response rather than the stimulus, and that association chaining would contaminate the later responses (McEvoy and Nelson, 1982). For example, given a target word *storm*, a first response could be *lightning* and a second response could be *Zeus*, which is arguably more related to *lightning* than it is to the target word *storm*. Additionally, investigations into the reliability of associations and the explanatory power of modelling behavioural data have shown that the first response is at least sufficient and possibly superior to subsequent responses (McEvoy and Nelson, 1982; Nelson et al., 2000).

However, Schulte im Walde and Melinger (2008) demonstrated within an in-depth analysis of co-occurrence distributions that – if corpus co-occurrence is an index of association response relevance – the rank $n + 1$ responses are as closely related to the targets as they are to their respective rank n responses. Therefore, depending on the goals of the project, multiple responses could provide a richer picture of the semantics of the target word by indexing additional meaning com-

ponents, and the fact that they are also related to the preceding responses only highlights the extent to which semantic knowledge is a network of inter-related nodes (cf. Lund and Burgess (1996)). This section nevertheless provides the results of the analyses from Section 4 for the first responses only, thus addressing the concern about chaining, along with noteworthy differences which we highlight.

- **Morpho-Syntactic Analysis:** Comparing the part-of-speech distributions of the first responses with the overall responses in Tables III and IV, we find very similar results: 34/57/8/1% are V/N/ADJ/ADV responses to verbs, and 14/71/3/12% are ADJ/N/PN/V responses to nouns. The only difference in the POS distributions concerns the POS type of the responses that is identical to that of the stimuli: Both the proportion of verb responses to verb stimuli and noun responses to noun stimuli increased.

- **Syntax-Semantic Noun Functions:** Concerning the syntactic functions of the noun responses to verbs and the verb responses to nouns, cf. Tables V and VI, the respective analyses of the first responses show a similar overall picture, with one difference: The grammar functions capture larger proportions of the responses, in both cases: 38% vs. 28% of the noun responses to verbs, and 49% vs. 41% of the verb responses to nouns are covered by our grammar. Accordingly, the overall proportions of cases with unknown verbs, nouns, or functions decreased: the proportions of cases with 'unknown verbs or nouns' went down from 22% to 18% for the noun responses to verbs; for the verb responses to nouns, this proportion did not change; the proportions of cases with known words but 'unknown functions' went down from 50% to 44% (noun responses to verbs) and from 48% to 40% (verb responses to nouns).

- **Co-Occurrence Analysis:** As for the syntax-semantic analysis, this analysis also captures larger proportions of the stimulus-response tokens than Tables VII and VIII: Using a co-occurrence window of 20 words as before, Table XI lists the co-occurrence proportions across all parts-of-speech *all*, for responses to verbs in the upper row and responses to nouns in the lower row. In comparison, the respective co-occurrence proportions over all responses were between 77% (strength 1) and 27% (strength 50) for responses to verbs, and between 84% and 23% for responses to nouns.

- **Semantic Relations:** Again, the analysis on the first responses showed a similar picture but covered larger proportions than Tables IX and X: For 64% of the verb-verb pairs and 22% of the noun-noun pairs (in comparison to 47/17%), we found semantic relations in GermaNet. Accordingly, the proportions of 'unknown cases' decreased from 12% to 9% (verb-verb pairs) and from 17% to 16% (noun-noun pairs), and the proportions of cases with 'no relation' decreased from 41% to 27% (verb-verb pairs) and from 66% to 62% (noun-noun pairs).

Table XI. Verb/Noun-association co-occurrence in 20-word window.

	Co-Occurrence Strength						
POS	1	2	3	5	10	20	50
V-*all*	84	79	76	68	60	50	36
N-*all*	88	82	78	72	59	46	29

Summarising the above analyses, the concerns about chaining effects are only partly justified. Using all responses vs. only the first response from the association norms does not result in great differences. Rather, the overall distributions of part-of-speech categories, syntax-semantic functions, co-occurrence, and semantic relations behave similarly with respect to each other. This insight is in line with the above mentioned co-occurrence analyses (Schulte im Walde and Melinger, 2008) indicating that the rank $n + 1$ responses are not solely semantically related to the rank n responses but also to the targets. Nevertheless, the proportions covered by the individual analyses for first responses are all above those for all responses, indicating that an $n + 1$ response is potentially related to any previous response in the association chain, thus the relation to the target is less strong.

Last but not least, the analyses demonstrated that the differences in the analyses with respect to the verb vs. noun association norms were not caused by the different methods we used for the collection. If the methods had caused the differences between the response analyses, then the analysis results should have been changed when comparing the first response analyses only, as those analyses made the two norms more comparable. The overall insights (and differences between the norms) remained similar, though, indicating that differences in the analyses with respect to the norms mainly correspond to the differences between responses to verbs vs. nouns.

6. Related Work

There are three lines of research which are related to central issues within this article: the collection of association norms (Section 6.1), analyses of association norms (Section 6.2), and human judgements on specific tasks that are concerned with semantic relatedness (Section 6.3).

6.1. COLLECTIONS OF ASSOCIATION NORMS

Association norms have a long tradition in psycholinguistic research. They have been used for more than 30 years to investigate semantic memory, making use of

the implicit notion that associates reflect meaning components of words. One of the first collections of word association norms was done by Palermo and Jenkins (1964), comprising associations for 200 words. The *Edinburgh Association Thesaurus* (Kiss et al., 1973) was a first attempt to collect association norms on a larger scale, and also to create a network of stimuli and associates, starting from a small set of stimuli derived from the Palermo and Jenkins norms. A similar motivation underlied the association norms from the University of South Florida (Nelson et al., 1998), who grew a stimulus-associate network over more than 20 years, from 1973. Their goal was to obtain the "largest database of free associations ever collected in the United States available to interested researchers and scholars". More than 6,000 participants produced nearly three-quarters of a million responses to 5,019 stimulus words. Smaller sets of association norms have also been collected for example for German (Russell and Meseck, 1959; Russell, 1970), Dutch (Lauteslager et al., 1986), French (Ferrand and Alario, 1998) and Spanish (Fernández et al., 2004) as well as for different populations of speakers, such as adults vs. children (Hirsh and Tree, 2001). Association norms have been used extensively in experimental psychology to conduct studies using the variations on the semantic priming technique to investigate, among other things, word recognition, knowledge representation and semantic processes (see (McNamara, 2005) for a review of methods, issues, and findings). Our own data which was the basis for this article concerned association norms for German verbs and nouns. The *Database of Noun Associations for German* (Melinger and Weber, 2006) can be accessed online.

6.2. ANALYSES OF ASSOCIATION DATA

In parallel to the interest in collecting association norms, researchers have analysed association data in order to get insight into semantic memory and – more specifically – issues concerning semantic relatedness. The following paragraphs provide an overview of these analyses, starting with theoretical considerations on relationships between stimuli and responses in association norms (not actually based on collected data), and progressing towards analyses of collected norms, which eventually got together with research in (Statistical) Natural Language Processing.

Clark (1971) identified relations between stimulus words and their associations on a theoretical basis, not with respect to collected association norms. He categorised stimulus-association relations into sub-categories of paradigmatic and syntagmatic relations, such as synonymy and antonymy, selectional preferences, etc. Heringer (1986) concentrated on syntagmatic associations to a small selection of 20 German verbs. He asked his subjects to provide question words as associations (e.g., *wer* 'who', *warum* 'why'), and used the responses to investigate the valency behaviour of the verbs. Spence and Owens (1990) showed that associative strength and word co-occurrence are correlated. Their investigation was based on 47 pairs of semantically related concrete nouns, as taken from the *Word Association Norms* (Palermo and Jenkins, 1964), and their co-occurrence counts in a window

of 250 characters in the 1-million-word Brown corpus. Church and Hanks (1989) were the first to apply information-theoretic measures to corpus data in order to predict word association norms. However, they did not rely on or evaluate against existing association data, but rather concentrated on the usage of the measure for lexicographic purposes. Rapp (2002) brought together research questions and methods from the above previous work: He developed corpus-based approaches to predict paradigmatic and syntagmatic associations, relying on the 100-million word BNC corpus (BNC, 1995). Concerning paradigmatic associations, he computed word association as the similarity of context vectors, applying the *City block distance* (also known as *Manhattan distance*, or L_1 *norm*) as similarity measure. A qualitative inspection revealed a strong overlap of strongly similar words with human associations, and a quantitative evaluation on the TOEFL test resulted in an accuracy of 69%. Concerning syntagmatic associations, he demonstrated that the word with the strongest co-occurrence to a target word (and filtered by a log-likelihood test) corresponded to the first human association of the respective target word in 27 out of 100 cases. Rapp's work used the *Edinburgh Association Thesaurus* as association database. In addition to his above contributions, his paper also provided an illustration of how strongly the co-occurrence distance between target stimuli and their associations was related to the respective number of responses to the stimuli in the association norms. These results complement our co-occurrence analyses in Section 4.3, providing an additional indicator of the co-occurrence hypothesis that includes the corpus distance.

Work by Fellbaum in the 1990s focused on human judgements concerning the semantic relationships between verbs. Similarly to our association experiment, Fellbaum and Chaffin (1990) asked participants in an experiment to provide associations to verbs. However, their work concentrated on verb-verb relations and therefore explicitly required verb responses to the verb stimuli. Also differently to our work, they restricted their stimuli to only 28 verbs; the resulting verb-verb pairs were manually classified into five pre-defined semantic relations. Fellbaum (1995) investigated the relatedness between antonymous verbs and nouns and their co-occurrence behaviour. Within that work, she searched the Brown corpus for antonymous word pairs in the same sentence, and found that regardless of the syntactic category, antonyms occur in the same sentence with much higher-than-chance frequencies. Last but not least, the WordNet organisation of the various parts-of-speech does rely on psycholinguistic evidence to a large extent (Fellbaum, 1998).

Our own work, of course, is also closely related to this article: Parts of the analyses have been published before, with Schulte im Walde and Melinger (2005) investigating the associates to verbs, Roth (2006) performing analyses on the associates to nouns, and Schulte im Walde et al. (2007) bringing the verb and noun analyses together, to investigate their usefulness with respect to distributional descriptions. Furthermore, concerning the analyses of verb associates, Schulte im Walde and Melinger (2008) performed a more in-depth analysis of the co-occurrence distri-

butions of stimulus-response pairs. Guida (2007) can be considered as a first piece of cross-linguistic work. She replicated most of our analyses on verb association norms for Italian verbs. Finally, Melinger et al. (2006) took the noun associations as input to a soft clustering approach, in order to predict noun ambiguity, and to discriminate the various noun senses of ambiguous stimulus nouns.

6.3. HUMAN DATA ON RELATEDNESS FOR NLP PURPOSES

In the following, we present a selection of work that collected human judgements on semantic relatedness, other than the "classical association norms" covered above.

On the one hand, there is an enormous number of approaches that used human judgements on semantic relatedness for the development and/or the assessment of linguistic resources and methods. It is impossible to cover the wealth of methods and data, so we just pick two examples: McCarthy et al. (2003) collected human rankings on the semantic relatedness of word pairs, because they were interested in the semantic similarity of particle verbs with respect to their base verbs, to evaluate models of particle verb compositionality. Similarly, Gurevych et al. (2007) collected human rankings across part-of-speech word pairs, and used them as gold standard semantic relatedness data within Information Retrieval experiments.

However, more closely related to this article are approaches that collected data similar to human associations on a more general basis referring to semantic relatedness. The following selection of work is restricted to recent approaches where the human data as well as the types of relatedness are most similar to our association data. As alternative source for semantic relations, researchers have used texts and asked readers of the texts to detect (and partly define) pairs or sets of words that are semantically related. For example, Morris and Hirst (2004) performed a study on lexical semantic relations that ensure text cohesion, as based on human labels of semantic text relations. Their relations were not restricted to combinations of certain morpho-syntactic categories but across categories, and included e.g. descriptive noun-adjective pairs (such as *professors/brilliant*), or stereotypical relations (such as *homeless/drunk*). Beigman Klebanov also worked text-based: Beigman Klebanov and Shamir (2006) investigated how well readers do agree on which items in a text are lexically cohesive, and why (i.e., based on which semantic relations); Beigman Klebanov (2006) continued on this work, investigated form-based clues to lexical cohesion in text, and modelled the text relations by various WordNet similarity measures. Differently to our work and the work by Morris and Hirst, Beigman Klebanov and Shamir did ask their experiment participants to identify word pairs that are semantically related, but they did not ask them to classify the relationships between the words.

Sinopalnikova (2004) collected word associations for Russian, across part-of-speech categories. As in our study, she analysed the corpus co-occurrence of the stimulus-response pairs. Using a window of 10 words, she found co-occurrence for 36% of the association data. In addition, she performed an intuitive analysis

of the various types of relationships between stimuli and associates. She demonstrated that the association norm contains a larger variety than typically covered by a thesaurus, and thus suggested word association norms as a useful means for constructing NLP resources, such as WordNet.

Boyd-Graber et al. (2006) performed a large-scale study on *evocation*, a semantic relation similar to association. They define evocation as "how much one concept evokes or brings to mind the other". 20 Princeton undergraduates rated evocation in 120,000 pairs of WordNet synsets, drawn randomly from all synset pairs defined from a set of 1,000 "core" synsets (as compiled by the WordNet group); as a relationship between concepts, evocation is a relation between meanings as expressed by synsets, not a relation between words; furthermore, it is not always a symmetrical relation (e.g., between the synsets *dollar–green*). The goal of the study was to enhance WordNet by adding a cross-part-of-speech kind of underspecified relation (and its strength).

Finally, Schulte im Walde (2008) relied on the verb association data within this article, and investigated whether the human associations to verbs could help to identify salient features within an automatic induction of semantic verb classes.

7. Summary and Conclusions

This article presented a study to identify, distinguish and quantify the various types of semantic associations provided by humans, and to illustrate their usage for NLP purposes. Within a first series of analyses we investigated the morpho-syntactic categories and the contextual functions that are represented by the associates with respect to the experiment stimuli. We demonstrated that nouns play a major role among the content word categories; this finding supports the predominant usage of noun features in distributional word representations. In addition, we showed that there is an extremely strong correlation between the frame-slot combinations in a grammar model and frame-slot combinations activated by our data; no linguistic functions could be considered to be prominent to represent conceptual nominal roles for verbs. A final analysis illustrated that clearly the noun associations are not restricted to verb subcategorisation role fillers, and that clause-internal adjuncts as well as clause-external, scene-related information or world knowledge should also play a role as features: we showed that the co-occurrence assumption holds for our German association data, to a large extent. These results suggest co-occurrence information for an appropriate usage in empirical descriptions of word properties, an important insight since co-occurrence information is essentially less expensive (because no high-level pre-processing such as parsing is necessary), and therefore easier to obtain – especially in languages with few NLP resources available – than annotated data.

A second part of the paper investigated the semantic relations between experiment stimuli and associates. We identified relationships as represented in GermaNet relations for 47/17% of the verb-verb and noun-noun pairs; the most com-

mon relationships refer to variants of the *is-a* relation, capturing hypernymy, hyponymy, and co-hyponymy. However, the major proportion of the stimulus-associate pairs could not be related via GermaNet. This result demonstrates that a) there are missing links in GermaNet, and our association data provides a useful starting point to enhance the taxonomy; and b) relations other than those coded in GermaNet (such as temporal order, cause, and consequence for verb-verb pairs, and condition, instrument, and result for noun-noun pairs) are represented in a large proportion of the stimulus-associate pairs. These data could be subsumed under *also see* relations in GermaNet, but it is obvious that one would prefer more fine-grained distinctions. We are specifically interested in those cases, because we expect that human associations cover the range of possible semantic relations to a large extent, and we believe that they represent an excellent basis for defining an exhaustive set, as alternative to e.g. text-based relations (Morris and Hirst, 2004; Beigman Klebanov and Shamir, 2006). One line of future work therefore concerns the identification and definition of the various verb-verb and noun-noun relations within our association norms (as a first step), and then to use the classified pairs in order to learn distributional cues to distinguish among the various relationships. This knowledge concerning the diversity of semantic relations is a crucial ingredient for NLP resources and applications, such as taxonomy and ontology learning, thesaurus extraction, and summarisation, so it is important to identify an exhaustive set of relations.

A further contribution of this article concerns the generalisation of the above insights. For each of the individual analyses, we demonstrated that the results are to a large extent correlated with the semantic classes of the stimulus words, and/or with their corpus frequencies. This observation is important with respect to both our goals: we infer that distributional feature descriptions as well as semantic relationships as described above will only cover the "average" of word meaning aspects. If one is concerned with specifying word properties and word-word relations with respect to individual words, the semantic class and the frequency range of that word should be taken into account. For example, even though nouns are clearly the dominant part-of-speech with respect to our analyses, for some semantic classes adjectives or adverbs play a major role. This insight should have an impact on the choice of features to describe word meaning aspects, suggesting to go beyond the pre-dominant categories and contextual functions within the selection of word features.

A final remark concerns a caveat with respect to our analyses: There are, of course, more aspects of the association norms than those covered by our analyses, and there are more resources that could be used for such analyses. This article was restricted to issues concerning a) which analyses might provide insight into the selection and acquisition of features to describe word meaning aspects, word similarities, and word relationships, and b) which resources are appropriate and available with respect to these analyses. Additionally, with respect to the noun stimuli used in our study, we included only depictable, and therefore concrete,

objects with minimal ambiguity. The extent to which the patterns of association would be mirrored for abstract or polysemous stimuli is unclear.

In conclusion, we believe that the association norms have contributed to the understanding of issues in computational linguistics. Even though the data represent a collection of word-word associations on a limited scale, they have proven useful to get insight into the computational modelling of words and word features. There is even more potential within the norms, which e.g. will allow us to address representational and distributional requirements with respect to the modelling of polysemy in future work.

Notes

[1]Few resources are semantically annotated and provide semantic information off-the-shelf (such as *FrameNet* (Fillmore et al., 2003) and *PropBank* (Palmer et al., 2005)).

[2]The association norms for verbs and nouns were originally collected in independent studies with different goals; as a consequence they differ somewhat in the methods used for data collection.

[3]Despite these instructions, some participants failed to use capitalisation, leading to some ambiguity. Similarly, the participants provided some multi-word expressions.

[4]As in the responses to the verb stimuli, we found that not all participants used capitalisation regularly to distinguish nouns from the other parts-of-speech. The hand-coding for the database therefore required a forced choice of POS to be made. If a participant appeared to use capitalisation regularly (i.e., never capitalised responses that could not possibly be nouns, such as adjectives) then we based the POS coding on this distinguishing characteristic. If the participant did not use capitalisation regularly, then the response was classified as the more frequent usage of the word.

[5]This cut-off is arbitrary and only for illustrative purposes.

[6]Throughout the article, the probabilities within the analyses are proportions of the absolute frequencies for a certain relationship over all response tokens.

[7]Note that these proportions are not restricted to the 10/12 functions which cover at least 1% of the stimulus-associate pairs, but include all functions found in the grammar (*'*total found in grammar').

[8]The original analyses in Schulte im Walde and Melinger (2005) and Roth (2006) used various window sizes; the 20-word window was considered the most illustrative case for covering a local context that goes beyond the clause boundaries.

[9]Note, however, that the 28/41% subcategorised nouns can only be compared indirectly with the 76/88% co-occurring nouns/verbs, because the former rely on only 35 million of the 200 million word corpus.

[10]The analyses in this section are restricted to target-associate pairs with identical part-of-speech, as they are the most prominent relations in WordNet.

[11]Note that the GermaNet relations are provided for relation types (in contrast to our analysis citing tokens).

[12]Note that the number of encoded relations in GermaNet differs strongly across relation types, which influenced the number of tokens that could potentially be found.

References

Baldwin, T., C. Bannard, T. Tanaka, and D. Widdows: 2003, 'An Empirical Model of Multiword Expression Decomposability'. In: *Proceedings of the ACL-2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*. Sapporo, Japan, pp. 89–96.

Beigman Klebanov, B.: 2006, 'Measuring Semantic Relatedness Using People and WordNet'. In: *Proceedings of the joint Conference on Human Language Technology and the North American Chapter of the Association for Computational Linguistics*. New York City, NY, pp. 13–17.

Beigman Klebanov, B. and E. Shamir: 2006, 'Reader-based Exploration of Lexical Cohesion'. *Language Resources and Evaluation* **40**(2), 109–126.

Berland, M. and E. Charniak: 1999, 'Finding Parts in Very Large Corpora'. In: *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*. Maryland, MD, pp. 57–64.

BNC: 1995, 'British National Corpus'. URL http://www.hcu.ox.ac.uk/BNC/.

Boyd-Graber, J., C. Fellbaum, D. Osherson, and R. Schapire: 2006, 'Adding Dense, Weighted Connections to WordNet'. In: *Proceedings of the Third Global WordNet Meeting*. Jeju Island, Korea.

Chklovski, T. and P. Pantel: 2004, 'VerbOcean: Mining the Web for Fine-Grained Semantic Verb Relations'. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Barcelona, Spain.

Church, K. W. and P. Hanks: 1989, 'Word Association Norms, Mutual Information, and Lexicography'. In: *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*. Vancouver, Canada, pp. 76–83.

Clark, H. H.: 1971, 'Word Associations and Linguistic Theory'. In: J. Lyons (ed.): *New Horizon in Linguistics*. Penguin, Chapt. 15, pp. 271–286.

Curran, J.: 2003, 'From Distributional to Semantic Similarity'. Ph.D. thesis, Institute for Communicating and Collaborative Systems, School of Informatics, University of Edinburgh.

Daelemans, W.: 2006, 'A Mission for Computational Natural Language Learning'. In: *Proceedings of the 10th Conference on Computational Natural Language Learning*. New York City, NY, pp. 1–5.

Deerwester, S., S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman: 1990, 'Indexing by Latent Semantic Analysis'. *Journal of the American Society of Information Science* **41**(6), 391–407.

Fellbaum, C.: 1995, 'Co-Occurrence and Antonymy'. *Lexicography* **8**(4), 281–303.

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

Fellbaum, C. and R. Chaffin: 1990, 'Some Principles of the Organization of Verbs in the Mental Lexicon'. In: *Proceedings of the 12th Annual Conference of the Cognitive Science Society of America*.

Fernández, A., E. Diez, M. A. Alonso, and M. S. Beato: 2004, 'Free-Association Norms for the Spanisch Names of the Snodgrass and Vanderwart Pictures'. *Behavior Research Methods, Instruments and Computers* **36**(3), 577–583.

Ferrand, L. and F.-X. Alario: 1998, 'French Word Association Norms for 366 Names of Objects'. *L'Annee Psychologique* **98**(4), 659–709.

Fillmore, C. J., C. R. Johnson, and M. R. Petruck: 2003, 'Background to FrameNet'. *International Journal of Lexicography* **16**, 235–250.

Geffet, M. and I. Dagan: 2005, 'The Distributional Inclusion Hypotheses and Lexical Entailment'. In: *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*. Ann Arbor, MI.

Girju, R.: 2003, 'Automatic Detection of Causal Relations for Question Answering'. In: *Proceedings of the ACL Workshop on Multilingual Summarization and Question Answering – Machine Learning and Beyond*. Sapporo, Japan.

Girju, R., A. Badulescu, and D. Moldovan: 2006, 'Automatic Discovery of Part-Whole Relations'. *Computational Linguistics* **32**(1), 83–135.

Girju, R., D. Moldovan, M. Tatu, and D. Antohe: 2005, 'On the Semantics of Noun Compounds'. *Journal of Computer Speech and Language* **19**(4). Special Issue on Multiword Expressions.

Girju, R., P. Nakov, V. Nastase, S. Szpakowicz, P. Turney, and D. Yuret: 2007, 'SemEval-2007 Task 04: Classification of Semantic Relations between Nominals'. In: *Proceedings of the 4th International Workshop on Semantic Evaluations*. Prague, Czech Republic, pp. 13–18.

Guida, A.: 2007, 'The Representation of Verb Meaning within Lexical Semantic Memory: Evidence from Word Associations'. Master's thesis, Universit degli studi di Pisa.

Gurevych, I., C. Müller, and T. Zesch: 2007, 'Electronic Career Guidance based on Semantic Relatedness'. In: *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*. Prague, Czech Republic.

Hamp, B. and H. Feldweg: 1997, 'GermaNet – a Lexical-Semantic Net for German'. In: *Proceedings of the ACL Workshop on Automatic Information Extraction and Building Lexical Semantic Resources for NLP Applications*. Madrid, Spain, pp. 9–15.

Harris, Z.: 1968, 'Distributional Structure'. In: J. J. Katz (ed.): *The Philosophy of Linguistics*, Oxford Readings in Philosophy. Oxford University Press, pp. 26–47.

Hearst, M.: 1998, 'Automated Discovery of WordNet Relations'. In (Fellbaum, 1998).

Heringer, H. J.: 1986, 'The Verb and its Semantic Power: Association as the Basis for Valence'. *Journal of Semantics* **4**, 79–99.

Hirsh, K. W. and J. Tree: 2001, 'Word Association Norms for two Cohorts of British Adults'. *Journal of Neurolinguistics* **14**(1), 1–44.

Ji, H., D. Westbrook, and R. Grishman: 2005, 'Using Semantic Relations to Refine Coreference Decisions'. In: *Proceedings of the joint Conference on Human Language Technology and Empirial Methods in Natural Language Processing*. Vancouver, Canada, pp. 17–24.

Kavalek, M. and V. Svatek: 2005, 'A Study on Automated Relation Labelling in Ontology Learning'. In: P. Buitelaar, P. Cimiano, and B. Magnini (eds.): *Ontology Learning and Population*, Vol. 123 of *Frontiers in Artificial Intelligence*. IOS Press.

Kiss, G., C. Armstrong, R. Milroy, and J. Piper: 1973, 'An Associative Thesaurus of English and its Computer Analysis'. In: *The Computer and Literary Studies*. Edinburgh University Press.

Korhonen, A., Y. Krymolowski, and Z. Marx: 2003, 'Clustering Polysemic Subcategorization Frame Distributions Semantically'. In: *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*. Sapporo, Japan, pp. 64–71.

Kunze, C.: 2000, 'Extension and Use of GermaNet, a Lexical-Semantic Database'. In: *Proceedings of the 2nd International Conference on Language Resources and Evaluation*. Athens, Greece, pp. 999–1002.

Kunze, C.: 2004, 'Semantische Relationstypen in GermaNet'. In: S. Langer and D. Schnorbusch (eds.): *Semantik im Lexikon*, Vol. 479 of *Tübinger Beiträge zur Linguistik*. Tübingen: Gunter Narr Verlag, pp. 162–178.

Landauer, T. K. and S. T. Dumais: 1997, 'A Solution to Plato's Problem: the Latent Semantic Analysis Theory of Acquisition, Induction and Representation of Knowledge'. *Psychological Review* **104**(2), 211–240.

Lapata, M.: 2002, 'The Disambiguation of Nominalisations'. *Computational Linguistics* **28**(3), 357–388.

Lauteslager, M., T. Schaap, and D. Schievels: 1986, *Schriftelijke Woordassociatienormen voor 549 Nederlandse Zelfstandige Naamworden.* Swets and Zeitlinger.

Lemaire, B. and G. Denhière: 2006, 'Effects of High-Order Co-Occurrences on Word Semantic Similarity'. *Current Psychology Letters – Behaviour, Brain and Cognition* **18**(1).

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

Lin, D.: 1998a, 'Automatic Retrieval and Clustering of Similar Words'. In: *Proceedings of the 17th International Conference on Computational Linguistics*. Montreal, Canada.

Lin, D.: 1998b, 'Extracting Collocations from Text Corpora'. In: *Proceedings of the First Workshop on Computational Terminology*. Montreal, Canada.

Lin, D.: 1999, 'Automatic Identification of Non-compositional Phrases'. In: *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*. Maryland, MD, pp. 317–324.

Lowe, W. and S. McDonald: 2000, 'The Direct Route: Mediated Priming in Semantic Space'. In: *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*. Philadelphia, PA, pp. 675–680.

Lund, K. and C. Burgess: 1996, 'Producing High-Dimensional Semantic Spaces from Lexical Co-Occurrence'. *Behavior Research Methods, Instruments, and Computers* **28**(2), 203–208.

Lund, K., C. Burgess, and R. A. Atchley: 1995, 'Semantic and Associative Priming in High-Dimensional Semantic Space'. In: *Proceedings of the 17th Annual Conference of the Cognitive Science Society of America*. pp. 660–665.

Maedche, A. and S. Staab: 2000, 'Discovering Conceptual Relations from Text'. In: *Proceedings of the 14th European Conference on Artificial Intelligence*. Berlin, Germany.

McCarthy, D., B. Keller, and J. Carroll: 2003, 'Detecting a Continuum of Compositionality in Phrasal Verbs'. In: *Proceedings of the ACL-SIGLEX Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*. Sapporo, Japan.

McEvoy, C. L. and D. L. Nelson: 1982, 'Category Name and Instance Norms for 106 Categories of Various Sizes'. *American Journal of Psychology* **95**, 581–634.

McKoon, G. and R. Ratcliff: 1992, 'Spreading Activation versus Compound Cue Accounts of Priming: Mediated Priming Revisited'. *Journal of Experimental Psychology: Learning, Memory and Cognition* **18**, 1155–1172.

McNamara, T. P.: 2005, *Semantic Priming: Perspectives from Memory and Word Recognition*. New York: Psychology Press.

Melinger, A., S. Schulte im Walde, and A. Weber: 2006, 'Characterizing Response Types and Revealing Noun Ambiguity in German Association Norms'. In: *Proceedings of the EACL Workshop 'Making Sense of Sense: Bringing Computational Linguistics and Psycholinguistics together'*. Trento, Italy, pp. 41–48.

Melinger, A. and A. Weber: 2006, 'Database of Noun Associations for German'. URL: www.coli.uni-saarland.de/projects/nag/.

Merlo, P. and S. Stevenson: 2001, 'Automatic Verb Classification Based on Statistical Distributions of Argument Structure'. *Computational Linguistics* **27**(3), 373–408.

Miller, G. A., R. Beckwith, C. Fellbaum, D. Gross, and K. J. Miller: 1990, 'Introduction to WordNet: An On-Line Lexical Database'. *International Journal of Lexicography* **3**(4), 235–244.

Moldovan, D., A. Badulescu, M. Tatu, D. Antohe, and R. Girju: 2004, 'Models for the Semantic Classification of Noun Phrases'. In: *Proceedings of the HLT-NAACL Computational Lexical Semantics Workshop*. Boston, MA, pp. 60–67.

Moldovan, D. and A. Novischi: 2002, 'Lexical Chains for Question Answering'. In: *Proceedings of the 19th International Conference on Computational Linguistics*. Taipei, Taiwan.

Morris, J. and G. Hirst: 2004, 'Non-Classical Lexical Semantic Relations'. In: *Proceedings of the HLT Workshop on Computational Lexical Semantics*. Boston, MA.

Nastase, V. A.: 2003, 'Semantic Relations across Syntactic Levels'. Ph.D. thesis, School of Information Technology and Engineering, University of Ottawa.

Navigli, R. and P. Velardi: 2004, 'Learning Domain Ontologies from Document Warehouses and Dedicated Web Sites'. *Computational Linguistics* **30**(2), 151–179.

Nelson, D., C. McEvoy, and T. Schreiber: 1998, 'The University of South Florida Word Association, Rhyme, and Word Fragment Norms'.

Nelson, D. L., C. L. McEvoy, and S. Dennis: 2000, 'What is Free Association and what does it measure?'. *Memory and Cognition* **28**, 887–899.

Padó, S. and M. Lapata: 2007, 'Dependency-based Construction of Semantic Space Models'. *Computational Linguistics* **33**(2), 161–199.

Palermo, D. and J. Jenkins: 1964, *Word Association Norms: Grade School through College*. Minneapolis: University of Minnesota Press.

Palmer, M., D. Gildea, and P. Kingsbury: 2005, 'The Proposition Bank: An annotated Resource of Semantic Roles'. *Computational Linguistics* **31**(1), 71–106.

Pantel, P. and M. Pennacchiotti: 2006, 'Espresso: Leveraging Generic Patterns for Automatically Harvesting Semantic Relations'. In: *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics*. Sydney, Australia, pp. 113–120.

Pereira, F., N. Tishby, and L. Lee: 1993, 'Distributional Clustering of English Words'. In: *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics*. Columbus, OH, pp. 183–190.

Plaut, D. C.: 1995, 'Semantic and Associative Priming in a Distributed Attractor Network'. In: *Proceedings of the 17th Annual Conference of the Cognitive Science Society*, Vol. 17. pp. 37–42.

Poesio, M., T. Ishikawa, S. Schulte im Walde, and R. Viera: 2002, 'Acquiring Lexical Knowledge for Anaphora Resolution'. In: *Proceedings of the 3rd Conference on Language Resources and Evaluation*, Vol. IV. Las Palmas de Gran Canaria, Spain, pp. 1220–1224.

Rapp, R.: 1996, *Die Berechnung von Assoziationen*, Vol. 16 of *Sprache und Computer*. Georg Olms Verlag.

Rapp, R.: 2002, 'The Computation of Word Associations: Comparing Syntagmatic and Paradigmatic Approaches'. In: *Proceedings of the 19th International Conference on Computational Linguistics*. Taipei, Taiwan.

Rooth, M., S. Riezler, D. Prescher, G. Carroll, and F. Beil: 1999, 'Inducing a Semantically Annotated Lexicon via EM-Based Clustering'. In: *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*. Maryland, MD.

Roth, M.: 2006, 'Relationen zwischen Nomen und ihren Assoziationen'. Studienarbeit. Institut für Computerlinguistik und Phonetik, Universität des Saarlandes.

Russell, W. A.: 1970, 'The complete German Language Norms for Responses to 100 Words from the Kent-Rosanoff Word Association Test'. In: L. Postman and G. Keppel (eds.): *Norms of Word Association*. New York: Academic Press, pp. 53–94.

Russell, W. A. and O. Meseck: 1959, 'Der Einfluss der Assoziation auf das Erinnern von Worten in der deutschen, französischen und englischen Sprache'. *Zeitschrift für Experimentelle und Angewandte Psychologie* **6**, 191–211.

Sahlgren, M.: 2006, 'The Word-Space Model: Using Distributional Analysis to Represent Syntagmatic and Paradigmatic Relations between Words in High-Dimensional Vector Spaces'. Ph.D. thesis, Stockholm University.

Salton, G. and M. McGill: 1983, *Introduction to Modern Information Retrieval*. New York: McGraw-Hill.

Salton, G., A. Wong, and C.-S. Yang: 1975, 'A Vector Space Model for Automatic Indexing'. *Communications of the ACM* **18**(11), 613–620.

Schulte im Walde, S.: 2002, 'A Subcategorisation Lexicon for German Verbs induced from a Lexicalised PCFG'. In: *Proceedings of the 3rd Conference on Language Resources and Evaluation*, Vol. IV. Las Palmas de Gran Canaria, Spain, pp. 1351–1357.

Schulte im Walde, S.: 2006, 'Experiments on the Automatic Induction of German Semantic Verb Classes'. *Computational Linguistics* **32**(2), 159–194.

Schulte im Walde, S.: 2008, 'Human Associations and the Choice of Features for Semantic Verb Classification'. *Research on Language and Computation*. To appear.

Schulte im Walde, S. and A. Melinger: 2005, 'Identifying Semantic Relations and Functional Properties of Human Verb Associations'. In: *Proceedings of the joint Conference on Human Language Technology and Empirial Methods in Natural Language Processing*. Vancouver, Canada, pp. 612–619.

Schulte im Walde, S. and A. Melinger: 2008, 'An In-Depth Look into the Co-Occurrence Distribution of Semantic Associates'. *Italian Journal of Linguistics. Special Issue on "From Context to Meaning: Distributional Models of the Lexicon in Linguistics and Cognitive Science"*. To appear.

Schulte im Walde, S., A. Melinger, M. Roth, and A. Weber: 2007, 'Which Distributional Functions are Crucial to Word Meaning: An Investigation of Semantic Associates'. In: C. Kunze, L.

Lemnitzer, and R. Osswald (eds.): *Proceedings of the GLDV Workshop on Lexical Semantic and Ontological Resources*, Vol. 336-3 of *Informatik-Berichte FernUniversität Hagen*. Tübingen, Germany, pp. 109–118.

Schütze, H.: 1998, 'Automatic Word Sense Discrimination'. *Computational Linguistics* **24**(1), 97–123. Special Issue on Word Sense Disambiguation.

Sinopalnikova, A.: 2004, 'Word Association Thesaurus as a Resource for Building WordNet'. In: *Proceedings of the 2nd International WordNet Conference*. Brno, Czech Republic, pp. 199–205.

Snodgrass, J. G. and M. Vanderwart: 1980, 'A Standardized Set of 260 Pictures: Norms for Name Agreement, Image Agreement, Familiarity, and Visual Complexity'. *Journal of Experimental Psychology: Human Learning and Memory* **6**, 174–215.

Spence, D. P. and K. C. Owens: 1990, 'Lexical Co-Occurrence and Association Strength'. *Journal of Psycholinguistic Research* **19**, 317–330.

Tatu, M. and D. Moldovan: 2005, 'A Semantic Approach to Recognizing Textual Entailment'. In: *Proceedings of the joint Conference on Human Language Technology and Empirial Methods in Natural Language Processing*. Vancouver, Canada, pp. 371–378.

Vieira, R. and M. Poesio: 2000, 'An Empirically-Based System for Processing Definite Descriptions'. *Computational Linguistics* **26**(4), 539–593.

Vigliocco, G., D. Vinson, W. Lewis, and M. Garrett: 2004, 'Representing the Meanings of Object and Action Words: The Featural and Unitary Semantic Space Hypothesis'. *Cognitive Psychology* **48**, 422–488.