

# Association norms for German noun compounds and their constituents

Sabine Schulte im Walde · Susanne R. Borgwaldt

© Psychonomic Society, Inc. 2015

**Abstract** We present a collection of association norms for 246 German depictable compound nouns and their constituents, comprising 58,652 association tokens distributed over 26,004 stimulus–associate pair types. Analyses of the data revealed that participants mainly provided noun associates, followed by adjective and verb associates. In corpus analyses, co-occurrence values for compounds and their associates were below those for nouns in general and their associates. The semantic relations between compound stimuli and their associates were more often co-hyponymy and hypernymy and less often hyponymy than for associations to nouns in general. Finally, we found a moderate correlation between the overlap of the associations to compounds and their constituents and the degree of semantic transparency. These data represent a collection of associations to German compound nouns and their constituents that constitute a valuable resource concerning the lexical semantic properties of the compound stimuli and the semantic relations between the stimuli and their associates. More specifically, the norms can be used for stimulus selection, hypothesis testing, and further research on morphologically complex words. The norms are available in text format (utf-8 encoding) as supplemental materials.

**Keywords** Associations · Semantic relations · Compound nouns · Complex words

**Electronic supplementary material** The online version of this article (doi:10.3758/s13428-014-0539-y) contains supplementary material, which is available to authorized users.

S. Schulte im Walde (✉)  
Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart,  
Pfaffenwaldring 5B, 70569 Stuttgart, Germany  
e-mail: schulte@ims.uni-stuttgart.de

S. R. Borgwaldt  
Germanistisches Seminar, Universität Siegen, Siegen, Germany

In this article, we introduce a new data collection of association norms for German noun compounds and their constituents. 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. In experimental psychology, association norms have—for example—been extensively used to conduct studies of semantic priming to investigate (among other things) word recognition, knowledge representation and semantic processes (see McNamara, 2005, for a review of methods, issues, and findings).

We collected associates to German noun compounds because we believe that the associations are a valuable resource for cognitive and computational linguistics research. On the basis of an existing collection of German noun compounds (von der Heide & Borgwaldt, 2009), we therefore gathered associates for the compounds and also for their constituents (e.g., *Ahornblatt/Ahorn/Blatt* “maple leaf/maple/leaf”). The data were collected via the crowdsourcing interface Amazon Mechanical Turk (AMT). We perform detailed analyses of the collection, regarding the parts of speech of the associate responses, and the co-occurrence and syntactic patterns as well as the semantic relations between the stimulus–associate pairs. We also predicted the degree of semantic transparency of the compounds, as based on a simple association overlap. The analyses are compared to those of an earlier collection (Schulte im Walde, Borgwaldt, & Jauch, 2012), in which associates to a superset of our compound and constituent stimuli were gathered in a more controlled Web experiment.

The association norms can be used as a lexical semantic resource concerning the target stimuli—that is, the compound nouns and their constituents. The data should be relevant for research on the lexical semantic properties of the compound stimuli—for example, the semantic relations between the stimuli and their associates—and the degree of semantic

relatedness between the compound stimuli and their constituents—that is, the degree of semantic transparency (alternatively *compositionality*).

In this article, we first provide an overview of association norms in general terms ([Previous work on association norms](#)) and introduce the German compound and constituent targets the norms rely on ([Noun compounds](#)), before we describe the collection and analyses of the noun compound association norms. The final part of the article summarizes and discusses the results.

## Previous work on association norms

### Collections of association norms

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, Armstrong, Milroy, & Piper, 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 underlay the association norms from the University of South Florida (Nelson, McEvoy, & Schreiber, 2004), who developed a stimulus–associate network for more than 20 years, starting in 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 (Melinger & Weber, 2006; Russell, 1970; Russell & Meseck, 1959; Schulte im Walde, Melinger, Roth, & Weber, 2008), Dutch (De Deyne & Storms, 2008; Lauteslager, Schaap, & Schievels, 1986), French (Ferrand & Alario, 1998), Spanish (Fernandez, Diez, Alonso, & Beato, 2004; Macizo, Gómez-Ariza, & Bajo, 2000), Portuguese (Comesaña, Fraga, Moreira, Frade, & Soares, 2014), and across languages (Kremer & Baroni, 2011), as well as for different populations of speakers, such as adults versus children (Hirsh & Tree, 2001; Macizo et al., 2000), and for words with various degrees of emotion (John, 1988) or homographs (French & Richards, 1992). Although some of the norms occasionally contain compounds, as far as we know, no specific collection has focused on associations for compounds and their constituents yet, other than our own previous collection for a superset of the compound and constituent stimuli (Schulte im Walde et al., 2012).

### Analyses of association norms

In parallel to the interest in collecting association norms, researchers have analyzed association data in order to get

insight into semantic memory. The following paragraphs provide an overview of these analyses, starting with theoretical considerations on relationships between stimuli and responses in association norms, and progressing toward analyses of collected norms.

Clark (1971) identified relations between stimulus words and their associates on a theoretical basis, not with respect to collected association norms. He categorized stimulus–associate relations into sub-categories of paradigmatic and syntagmatic relations, such as synonymy and antonymy, selectional preferences, and so forth. Heringer (1986) concentrated on syntagmatic associations to a small selection of 20 German verbs. He asked his subjects to provide question words as associates (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 Palermo and Jenkins norms, and co-occurrence counts in a window of 250 characters in the 1-million-word Brown corpus. Church and Hanks (1990) were the first to apply information-theoretic measures to corpus data in order to predict word associations. However, they did not rely on or evaluate their findings against existing association data, but rather concentrated on the usage of the measure for lexicographic purposes. Rapp (2002) combined 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. 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 a similarity measure. A qualitative inspection revealed a strong overlap of very similar words with human associations, and applying the associations to solve 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 the association database. In addition to his above contributions, his article also provided an illustration of how strongly the co-occurrence distance between target stimuli and their associates was related to the respective numbers of responses to the stimuli in the association norms.

Work by Fellbaum and colleagues in the 1990s focused on human judgments concerning the semantic relationships between verbs. Fellbaum and Chaffin (1990) asked participants in an experiment to provide associates to verbs. Their work concentrated on verb–verb relations and therefore explicitly required verb responses to the verb stimuli. The resulting verb–verb pairs were manually classified into five

predefined semantic relations. Fellbaum (1995) investigated the relatedness between antonymous verbs and nouns and their co-occurrence behavior. 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 frequencies that are much higher than chance. Last but not least, the WordNet organization of the various parts-of-speech largely relies on psycholinguistic evidence (Fellbaum, 1998).

On the basis of the associates to German nouns and verbs that were collected by Schulte im Walde et al. (2008; cf. *Collections of association norms* above), they performed detailed analyses at the syntax–semantics interface. Guida (2007) replicated most of their analyses on verb association norms for Italian verbs. Roth and Schulte im Walde (2008) extended the co-occurrence analysis for noun associations in Schulte im Walde et al. (2008) and explored whether dictionary and encyclopedic information provided more world knowledge about associations than corpus co-occurrence. They found that the information in the three resource types complemented each other. Schulte im Walde and Melinger (2008) performed a more in-depth analysis of the co-occurrence distributions of the noun associations in Schulte im Walde et al. (2008). Schulte im Walde et al. (2012) performed a part-of-speech analysis on previously collected associates for compound nouns and their constituents.

From a more applied point of view in the field of computational linguistics, Melinger, Schulte im Walde, and Weber (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. Schulte im Walde (2008) relied on the associations to German verbs described by Schulte im Walde et al. (2008) to determine salient features for automatic semantic verb classification.

## Noun compounds

Compounds are morphologically complex words, coined by two or more simple words. Our focus of interest is on German noun compounds (see Fleischer & Barz, 2012, for a detailed overview, and Klos, 2011, for a recent exploration), such as *Ahornblatt* “maple leaf,” *Feuerwerk* “fireworks,” *Nähmaschine* “sewing machine,” *Obstkuchen* “fruit cake,” and *Rotkohl* “red cabbage,” in which the grammatical head (in German, the rightmost constituent) is a noun and the modifier can belong to various parts of speech.

More specifically, we are interested in the degrees of *semantic transparency* of German noun compounds—that is, the relation between the meaning of the whole compound (e.g., *butterfly*) and the meaning of its parts (e.g., *butter*,

*fly*)—which has been studied intensively by psycholinguists, in order to find out how compound words are cognitively processed and represented in the mental lexicon. There is ongoing debate about whether morphologically complex words are stored and processed as single units (Butterworth’s full listing approach (Butterworth, 1983)), whether they are decomposed into their morphemes (Taft, 2004; Taft & Forster, 1975) or can be accessed both ways—as whole forms and componentially, via their constituent morphemes (*dual-route models*; see, e.g., Baayen & Schreuder, 1999; Caramazza, Laudanna, & Romani, 1988)—and which variables predict processing behavior. The majority of studies in this area have investigated *morphological decomposition* during *compound comprehension* (but see, e.g., Lüttmann, Zwitserlood, Böhl, and Bölte, 2011, for evidence of *morphological composition* during *compound production*).

Factors that have been found to influence the cognitive processing and representation of compounds include orthographic variables like the number of letters (Bertram & Hyönä, 2003) or the presence of hyphens or interword spaces (Bertram, Kuperman, Baayen, & Hyönä, 2011), frequency-based measures as the frequencies of the compound and its constituents (e.g., Janssen, Bi, & Caramazza, 2008; van Jaarsveld & Rattink, 1988) and the morphological family size—that is, the number of compounds that share a constituent (de Jong, Feldman, Schreuder, Pastizzo, & Baayen, 2002), variables relating to morphological complexity, such as the number of morphemes or the existence of linking elements (Krott, Schreuder, Baayen, & Dressler, 2007), and semantic variables, such as the relationship between compound modifier and head—that is, a *teapot* is a *pot FOR tea*, and a *snowball* is a *ball MADE OF snow* (Gagné & Spalding, 2009).

With some researchers (e.g., Longtin, Segui, & Hallé, 2003; Marslen-Wilson, Tyler, Waksler, & Older, 1994) arguing that morphological decomposition happens only in semantically transparent polymorphemic words and not in semantically opaque ones, one variable that might be particularly important for the processing of compounds is their compositionality/semantic transparency. For example, studies by Sandra (1990) and Zwitserlood (1994) showed that the meanings of the constituents of semantically transparent compounds (e.g., *dog* and *house* in *doghouse*) were activated during processing, whereas the meanings of the constituents of opaque compounds (e.g., *butter* and *fly* in *butterfly*) were not activated.

Although interrater agreement about compounds’ perceived semantic transparency is generally rather high (Reddy, McCarthy, & Manandhar, 2011; Roller, Schulte im Walde, & Scheible, 2013), little research has investigated how the semantic transparency is assessed. One assumption is that the degree of a compound’s semantic transparency should also be reflected in its association patterns. If a compound is classified as opaque (e.g., *butterfly*), one would assume that

the associations to the whole compound show less overlap with the associations to the components (*butter*, *fly*), than in the case of a transparent compound (e.g., *doghouse*), as during the processing of (partially) opaque compounds the opaque constituents might be less activated at the semantic level. This is what Libben's (1994, 1998) "automatic progressive parsing and lexical excitation (APPLE) model of morphological parsing" predicts: It assumes that compound words are represented at three levels: a stimulus level, a lexical level (purely morphological) and a conceptual level (semantic). A (partially) opaque compound such as *strawberry* is decomposed at the lexical level into *straw* and *berry*. At the conceptual level, however, only the semantically transparent constituent *berry* is represented and can accordingly generate associations. This representation difference might, for example, explain the observed dissociation between constituent repetition priming effects (Libben, Gibson, Yoon, & Sandra, 2003; Zwitserlood, 1994) and semantic priming effects (Sandra, 1990; Zwitserlood, 1994) for opaque compounds.

In sum, our collection of association norms for compounds and their constituents is mainly aimed to be of use for researchers on compound processing. It provides a resource for stimulus selection in experimental studies and also allows researchers to examine properties of associations to compounds and their constituents in detail, in order to gain more insight into the way meanings of compounds are processed and/or represented in the mental lexicon.

## Experiment

Associations are commonly obtained by presenting *target stimuli* to the participants in an experiment, who then provide *associate responses*—that is, words that are spontaneously called to mind by the stimulus words. The quantification of the resulting stimulus–associate pairs (i.e., how often a certain associate is provided for a certain stimulus) is called an *association norm*. In the following sections, we describe the collection of our associates to German noun compounds.

## Method

**Material** The target compounds and constituents were based on the selection of noun compounds by von der Heide and Borgwaldt (2009). They created a set of 450 concrete, depictable German noun compounds that they grouped into four transparency classes: compounds that are transparent with regard to both constituents (TT; e.g., *Ahornblatt* "maple leaf"), compounds that are opaque with regard to both constituents (OO; e.g., *Löwenzahn* "lion + tooth → dandelion"), compounds that are transparent with regard to the modifier but opaque with regard to the head (TO; e.g., *Feuerzeug* "fire + stuff → lighter"), and compounds that are opaque with regard

to the modifier but transparent with regard to the head (OT; e.g., *Fliegenpilz* "fly + mushroom → toadstool").<sup>1</sup> In total, the four classes contained 220 instances of TT, 126 instances of OT, 79 instances of TO, and 25 instances of OO.

The 450 noun compounds from von der Heide and Borgwaldt (2009) were categorized according to the morphological category of the modifier (AN, adjective–noun compound; NN, noun–noun compound; PN, preposition–noun compound; VN, verb–noun compound; MN, noun compound, in which the modifier is morphologically motivated by multiple classes; unique, noun compound with unique modifier). The categorization was performed by consensus decision of four computational linguists.

In our associate collection, we only used 246 of the 450 noun compounds from von der Heide and Borgwaldt (2009)—that is, 237 bimorphemic noun–noun compounds and nine noun compounds in which the modifier is unique, a so-called "cranberry morpheme" (such as *him* in *Himbeere* "him + berry → raspberry"), with no meaning by itself. Each compound had exactly two simple constituents. The compound set comprised 106 instances of TT, 37 instances of TO, 87 instances of OT, and 16 instances of OO. We restricted the target set because the subset of the two-part noun–noun compounds was most relevant to our research. Appendix A provides the complete list of our noun compounds.

In total, our materials comprised 571 targets. The total number of target stimuli was less than  $3 \times 246$  because some of the compounds share constituents. We first divided the stimuli randomly into four separate packages, making sure that there were similar numbers of compounds in each of the four packages, and that both constituents of each compound were in the same package as the compound. In this way, constituents that are shared by several compounds might appear in more than one part, but we could collect the associates to the compounds and their constituents in the four packages independently of each other. Taking the multiple occurrences of some constituents into account, the four packages contained 173/169/173/172 targets, respectively. The four packages were then each randomly divided into eight parts containing 21–22 targets in a random order. To control for spammers and to identify non-native speakers of German, we also included three German fake compound nouns into each of the batches, in random positions of the lists. The list of fake nouns is (in alphabetical order): *Analigzerbruch*, *Armmoder*, *Brattlider*, *Bulkerzagen*, *Engschogen*, *Fennhoder*, *Harmweg*, *Luderschwege*, *Malligwohmer*, *Pillinrugen*, *Quetpfluge*, *Tropebuhle*, *Wierzverkuhr*, and *Zogschucht*.

<sup>1</sup> The original classification into TT, TO, OT, and OO was performed by the two authors (Claudia von der Heide and Susanne Borgwaldt). Transparency ratings were collected in a follow-up step, using a 1–7 scale.



**Procedure** The experiment was performed via Amazon Mechanical Turk.<sup>2</sup> When an AMT worker chose one of our batches, the worker was presented with 24 or 25 noun compounds (including 21 or 22 real and three fake compounds) in 24/25 *HITs*, respectively. The setup of the experiment is shown by Fig. 1 in Appendix B, with translations in red font. The actual collection of the associates was performed as is shown in Figs. 2 and 3: Each trial (represented by a *HIT*) provided a brief description of the experiment, an example item with potential responses, and a single target (one of the noun compounds, or one of the constituents, or a fake noun). Below the target were three data input lines where participants could type their associates. They were instructed to type at most one word per line and, following German grammar, to distinguish nouns from other parts of speech with capitalization. Below the three input lines was a box that participants were asked to check if they did not know the word. The  $4 \times 8 = 32$  batches were completed within 1 to 26 days each, and the whole collection was finished within 3 months.

**Participants** The participants were AMT workers. We asked for 25 native German speakers per target, and paid USD 0.02 per trial (i.e., for up to three associates). We only accepted participants who identified the fake nouns correctly and who in addition had an overall approval rate of at least 95 % (after the experiment was completed). These checks ensured that we received associates from native German speakers only. Over all trials, we had participants with 146 different worker IDs, who provided associates for 1 up to 683 stimuli.

## Data

For each stimulus, we had between 2 and 120 participants. Because the participants could provide between zero and three associates per target, the actual number of associates per stimulus varied between 6 and 356. All but the unique constituents received  $\geq 48$  associates. In total, we collected 58,652 associates from 20,333 trials, an average of 2.88 associates per trial. The 58,652 associate tokens are distributed over 26,004 association types.

## Previous experiment

This section reports on a previous experiment to collect associates to the compound nouns and their constituents from von der Heide and Borgwaldt (2009), as described by Schulte im Walde et al. (2012). This experiment was our first attempt to collect associates to compound nouns and their constituents, and was stopped after one year because the incoming data

stagnated. The present experiment described in Experiment was then set up to continue the collection with a more specific focus on noun–noun compounds, still using identical collection instructions and procedures.

The reasons why we report on the earlier experiment are the following.

- The previous associates were collected for a superset of the noun compounds that were used in the present study. The two collections can be exploited independently (taking into account that they were collected in different ways) or together (making use of a richer set of associations for the intersection of the noun–noun compounds and their constituents, as well as exploiting an overall larger stimulus set).
- Moreover, we were interested in whether we could find differences between the two sets of association norms, since they were collected in different ways.

## Method

**Material** We used 442 compound nouns and constituents from the original selection by von der Heide and Borgwaldt (2009). In total, our materials comprised 996 targets—that is, 442 compounds and 554 constituents. The stimuli were divided randomly into 12 separate experimental lists of 83 nouns each.

**Procedure** The experiment was administered over the Internet and announced by e-mails to colleagues and friends. Participants were first provided a brief description of the experiment and asked for their biographical information, such as linguistic expertise, age and regional dialect (see Fig. 4, with a translated version in Fig. 5). Next, the participant was presented with the written instructions for the experiment and an example item with potential responses. In the actual experiment, one of the 12 experimental lists was chosen randomly. Each of the 83 trials in the experimental list consisted of a single word presented in a box at the top of the screen. The word was either one of the noun compounds or one of the constituents. If a compound constituent was not a noun, the base form was nominalized by starting it with a capital letter. For example, the verbal modifier *fahren* in *Fahrplan* “to ride + schedule  $\rightarrow$  time table”) was represented as *Fahren*, the adjectival modifier *blau* in *Blaubeere* “blue + berry  $\rightarrow$  blueberry”) was represented as *Blau*. The experiment ran for approximately 1 year.

The order of the target words was random for each data set and each participant. Below the target were three data input lines where participants could type their associates. They were instructed to type at most one word per line and, following German grammar, to distinguish nouns from other parts of speech with capitalization. Below the three input lines was a

<sup>2</sup> [www.mturk.com](http://www.mturk.com)

**Table 1** Numbers of association norms

Experiment	Stimuli	Responses	
		Types	Tokens
Web experiment	All noun compounds and their constituents	28,238	47,249
	<i>Noun–noun compounds</i> and their constituents	15,600	26,397
AMT experiment	<i>Noun–noun compounds</i> and their constituents	26,004	58,652
Both experiments	All noun compounds and their constituents	47,523	106,693
	<i>Noun–noun compounds</i> and their constituents	34,560	85,049

box that participants were asked to check if they did not know the word.

**Participants** A group of 268 participants took part in the experiment. Of these, 225 claimed that their L1 was German; for 19 of them, German was not their L1; one of the participants claimed to have grown up bilingual; and 23 did not provide information about their L1. One hundred twenty four of the individuals identified themselves as having a linguistic background, and 112 rated themselves as linguistic novices; 32 participants did not provide information about their linguistic background.

#### Data

For each experimental list, we had between 14 and 28 participants. Because the participants could provide between zero and three associates per target, the number of participants per stimulus varied between 10 and 36, and the number of associates per stimulus varied between 6 and 74. In total, we collected 47,249 associates from 17,128 trials, an average of 2.76 associates per trial. The 47,249 associate tokens were distributed over 28,238 association types. In 861 trials, the participants did not provide any associate, out of which 327 targets were explicitly checked as not being known by the participants.

#### Overall results and analyses

Quantifying overall responses in the two experiments resulted in a total of 47,523/106,693 stimulus–associate types/tokens. Table 1 summarizes the numbers of association norms with regard to the present study, the previous study, and the union of the two. We distinguish between associations to all noun compound stimuli and their constituents (as in the previous experiment) and associations to the subset of 246 noun–noun compounds and their constituents (which were used in the present experiment and were already part of the previous experiment). Even though our previous experiment ran for approximately 1 year and the present experiment for only three months, we collected more associates in the present experiment.

Tables 2, 3, and 4 provide examples of the associates to three compounds and their constituents, in each case listing the ten strongest (i.e., most frequently provided) associates. We selected three examples rather than a single one, to observe (i) the effect of the degree of transparency of the compound on the associates to the compounds versus the constituents, as well as (ii) the effect of monosemous versus polysemous stimuli with regard to the semantics of the associates. Note that *Fliegenpilz* is less transparent than *Ahornblatt* (at least with respect to its modifier), so that the associates of the

**Table 2** Most frequent responses to the compound *Ahornblatt* “maple leaf” and its constituents

<i>Ahornblatt</i> “maple leaf”			<i>Ahorn</i> “maple”			<i>Blatt</i> “leaf/sheet (of paper)”		
Kanada	“Canada”	26	Baum	“tree”	39	Baum	“tree”	29
Baum	“tree”	23	Sirup	“syrup”	30	Papier	“paper”	27
Herbst	“autumn”	15	Kanada	“Canada”	22	grün	“green”	10
Sirup	“syrup”	12	Blatt	“leaf”	16	schreiben	“write”	7
Laub	“leaves”	7	Blätter	“leaves”	6	Pflanze	“plant”	4
Wald	“forest”	4	Herbst	“autumn”	4	Laub	“leaves”	3
rot	“red”	4	Wald	“forest”	3	Herbst	“autumn”	3
bunt	“colorful”	3	grün	“green”	2	weiß	“white”	3
grün	“green”	3	rot	“red”	2	Blume	“flower”	3
zackig	“jagged”	2	Ahornsirup	“maple syrup”	2	Wald	“forest”	2

**Table 3** Most frequent responses to the compound *Fliegenpilz* “toadstool” and its constituents

<i>Fliegenpilz</i> “toadstool”			<i>Fliege</i> “fly/bowtie”			<i>Pilz</i> “mushroom”		
giftig	“poisonous”	34	nervig	“annoying”	11	Wald	“forest”	32
rot	“red”	26	Insekt	“insect”	9	giftig	“poisonous”	13
Wald	“forest”	18	summen	“buzz”	9	sammeln	“collect”	9
Gift	“poison”	6	Sommer	“summer”	6	Fliegenpilz	“toadstool”	6
Punkte	“dots”	5	Klatsche	“flap”	6	Champignon	“mushroom”	5
weiß	“white”	4	lästig	“annoying”	5	Suppe	“soup”	4
Märchen	“tale”	2	Fliegenklatsche	“fly flap”	5	essen	“eat”	3
rot-weiß	“red-white”	2	Tier	“animal”	5	essbar	“edible”	3
gepunktet	“dotted”	2	klein	“small”	5	Schimmel	“mold”	3
Vergiftung	“poisoning”	2	Krawatte	“tie”	4	Champignons	“mushrooms”	3

compound *Fliegenpilz* and the modifier *Fliege* differ more strongly. Note also that two of the nouns are polysemous: *Fliege* and *Blatt*. For both nouns, we collected associates to two senses: In the case of *Blatt*, associates were given to the plant sense “leaf” as well as to the paper sense “sheet (of paper)”; *Fliege* evoked associates to the animal sense “fly” as well as to the clothes sense “bowtie.”

In the following subsections, we present a series of analyses of the association norms with regard to the following stimulus properties and stimulus–associate relations:

- a *morpho-syntactic analysis*, looking into the parts of speech of the associates,
- a *distributional analysis*, looking into the co-occurrence of stimuli and associates,
- a *syntactic analysis*, looking into the dependency paths between stimuli and associates, and
- a *semantic relation analysis*, looking into the stimulus–associate semantic relations.

In all of the analyses, we will pay attention to the effects of our specific set of stimuli, compounds and their constituents, and explore which syntactic and semantic properties are specific to these stimuli. We conclude our analyses with a study regarding the compounds’ semantic transparency by investigating the correlation between association overlap (of compounds and constituents) and an existing set of compound–constituent ratings.

### Morpho-syntactic analysis

In the morpho-syntactic analysis, each response to the stimuli was assigned its—possibly ambiguous—part of speech (*pos*). The results provide insight into the relevance of predominant part-of-speech categories. Similar morpho-syntactic analyses have been performed by Guida (2007), Schulte im Walde et al. (2008), and Schulte im Walde et al. (2012).

As a resource for the pos assignment, we relied on the lemmatized and pos-tagged frequency list from the *SdeWaC* corpus (Faaß & Eckart, 2013), a cleaned version of the German Web corpus *deWaC* created by the *WaCky* group (Baroni, Bernardini, Ferraresi, & Zanchetta, 2009). The *SdeWaC* contains approximately 880 million word tokens. We disregarded fine-grained distinctions such as case, number, and gender features, and considered only the major categories verb (V), noun (N), and adjective (ADJ). A fourth category, “OTHER,” comprises all other part-of-speech categories, such as adverbs, prepositions, particles, interjections, conjunctions, and so forth. Ambiguities between the categories arose, for instance, when the experiment participant could have been referring to either an adjective or a (noncapitalized) noun, such as *fett* “fat.”<sup>3</sup>

Having assigned part-of-speech tags to the associates, we were able to distinguish and quantify the morpho-syntactic categories of the associates. In nonambiguous situations, the unique part of speech received the total stimulus–associate strength. For example, *Herd* “cooker” was provided as an associate of the stimulus *Pfanne* “pan” by 11 participants. Our pos resource contained *Herd* in the corpus only as a noun. So *Pfanne* received a contribution of all 11 mentions for a noun pos. In ambiguous situations, the stimulus–associate frequency was split over the possible part-of-speech tags according to the pos proportions in the frequency list. For example, *fett* “fat” was provided as an associate of the stimulus *Pfanne* by three participants. Our pos resource contained *Fett* 10,780 times in the corpus as a noun, and 493 times as an adjective. So, with regard to *Fett*, *Pfanne* received a contribution of  $3 \times 10,780 / (10,780 + 493) = 2.87$  nouns, and of  $3 \times 493 / (10,780 + 493) = 0.13$  adjectives.

The output of this analysis was frequency distributions of the part-of-speech tags for each stimulus individually, and also

<sup>3</sup> Despite our instructions, some participants failed to use capitalization, leading to ambiguity.

**Table 4** Most frequent responses to the compound *Schlittenhund* “sledge dog” and its constituents

<i>Schlittenhund</i> “sledge dog”			<i>Schlitten</i> “sledge”			<i>Hund</i> “dog”		
Schnee	“snow”	24	Winter	“winter”	34	bellen	“bark”	46
Husky	“husky”	21	Schnee	“snow”	32	Katze	“cat”	33
Alaska	“Alaska”	11	Kinder	“children”	9	Haustier	“pet”	20
Winter	“winter”	10	kalt	“cold”	8	Tier	“animal”	18
kalt	“cold”	7	Spaß	“fun”	7	Freund	“friend”	17
Eskimo	“Eskimo”	5	rodeln	“sledge”	5	treu	“faithful”	13
Schlitten	“sledge”	4	fahren	“ride”	5	Leine	“leash”	10
rennen	“run”	4	Hügel	“hill”	4	Maus	“mouse”	9
Eis	“ice”	4	Kufen	“skids”	3	beißen	“bite”	8
ziehen	“pull”	3	Hund	“dog”	2	Hundehütte	“kennel”	6

as a sum over all stimuli. Table 5 presents the total numbers over all associate response tokens, across the two experiments. Overall, the participants provided noun associates for the clear majority of token instances, 69 %–73 %; adjectives were given for 12 %–16 % of the associates, and verbs for 11 %–13 %. There are slight differences across the two experiments and when considering all compounds versus only the noun–noun compounds, but overall the proportions are very similar in this table and in comparison to the same analyses in previous work regarding associations to noun stimuli (Schulte im Walde et al., 2008). This is the case even if we only consider the associations to the compound stimuli.

#### Co-occurrence analysis

In this analysis, we examined whether the co-occurrence assumption held for our association norms—that is, what percentage of the associates were found in co-occurrence with the stimulus words in a corpus. The co-occurrence hypothesis assumes that associations are related to the textual co-occurrence of the stimulus–associate pairs. The hypothesis has been confirmed in many previous studies, among them Miller (1969), Spence and Owens (1990), Fellbaum (1995), Schulte im Walde and Melinger (2008), Schulte im Walde et al. (2008), and Schulte im Walde and Müller (2013).

As association norms, we used all of the experiment data, and our study again relied on the *SdeWaC* corpus. This time, no parsing information was required, since we only checked how often the associates co-occurred within windows of 5 and 20 words (to the left and to the right).<sup>4</sup> The analysis was again token-based. Table 6 presents the results. The columns show the proportions of pairs that co-occur with a co-occurrence strength of 1 (i.e., at least once), 2 (i.e., at least twice), and so

forth. The rows distinguish between windows of sizes 5 and 20. For example, the associate *Wald* “forest,” which was provided by 32 participants for the stimulus *Pilz* “mushroom,” was found 76 times in a window of five words from the stimulus. The 32 tokens thus appear in all columns. In contrast, the associate *Husky* “husky,” which was provided by 21 participants for the stimulus *Schlittenhund* “sledge dog,” was found ten times in a window of five words from the stimulus. The ten tokens thus appear in the columns with co-occurrence strengths  $\geq 1$ , 2, 5, and 10.

For each co-occurrence range, we present two lines, one for all stimulus–associate pairs, and the second when restricting the stimuli to the compounds. Note that the *all-stimuli* condition was restricted to the 246 noun–noun compounds and their nominal constituents (i.e., disregarding nonnominal stimuli), so that we could compare the co-occurrence analyses to previous work on nominal stimuli and thus compare the co-occurrence information for (i) noun–noun compounds and their nominal constituents (all stimuli), (ii) the nominal compound stimuli only, and (iii) nouns in general (cf. Schulte im Walde et al., 2008).

We can see that the co-occurrence assumption clearly holds for our norms: 73 % of the associates appeared within a five-word window of the respective stimuli at least once, and more than half (53 %) of the associates appeared at least five times in a five-word window of the respective stimuli. However, we also notice that the co-occurrence values are below those for German nouns and their associates in general: As compared to earlier studies (Schulte im Walde et al., 2008), in which we performed the co-occurrence check on the associates of 409 depictable German nouns (only including a few compound nouns) and found 84 % of the stimulus–associate pairs in a 20-word window at least once, in the present analysis we found a value of only 79 % (and with regard to compound–associate pairs, only 67 %). The difference is stronger than it first appears, because the earlier analyses were performed on a 200-million word German newspaper corpus from the 1990s

<sup>4</sup> Note that the sentences in the *SdeWaC* corpus are sorted alphabetically, so the window co-occurrence refers to 5/20 words to the left and right BUT within the same sentence.



**Table 5** Numbers and percentages of associate parts of speech

Experiment	Stimuli	Freq/Prop	N	ADJ	V	OTHER
Web experiment	All noun compounds and their constituents	Freq	33,322	6,835	5,264	1,827
		Prop	71 %	14 %	11 %	4 %
	<i>Noun–noun compounds</i> and their constituents	Freq	18,263	4,197	2,971	967
		Prop	69 %	16 %	11 %	4 %
AMT experiment	<i>Noun–noun compounds</i> and their constituents	Freq	43,104	7,162	7,365	1,021
		Prop	73 %	12 %	13 %	2 %
Both experiments	<i>Noun–noun compounds</i> and their constituents	Freq	61,367	11,359	10,337	1,986
		Prop	72 %	14 %	12 %	2 %
	<i>Noun–noun compounds</i> only	Freq	23,002	5,246	3,972	845
		Prop	70 %	16 %	12 %	3 %

(so the co-occurrence source was much smaller than the 880-million word corpus SdeWaC). In order to make the two analyses more comparable, we repeated our co-occurrence check for the compound norms on the 200-million word corpus. For the general noun norm data in Schulte im Walde et al. (2008), we found 73 %/84 % of the stimulus–associate pairs at least once in windows 5 and 20, whereas we found only 57 %/68 % of the stimulus–associate pairs in the present analysis (taking compound and constituent stimuli into account), and with regard to compound–associate pairs we found only 38 %/53 %.<sup>5</sup> This comparison clearly shows that although the co-occurrence assumption holds for our norms, the corpora offer less information on compounds than on simple nouns.

#### Syntactic dependency paths

This analysis went beyond pure co-occurrence and investigates the syntactic dependency paths between the stimuli and their associates. The analysis would tell us whether any syntactic relationships were specifically strong as triggers for associates. For example, we do not know a priori whether responses refer to associates that are in a modifier relationship with the stimuli (such as *grün* “green” for *Salat* “salad”) or in conjunction (such as *Katze* “cat” for *Hund* “dog”), and so forth. To obtain insight into which syntactic dependencies might trigger associates to our compound and constituent stimuli, we checked, for all stimulus–associate pairs, whether and which dependency paths exist between the stimuli and the associates.

The analysis relied on dependency parses of the SdeWaC corpus (see above) by Bohnet’s MATE dependency parser (Bohnet, 2010). For each occurrence of a stimulus in the

parsed corpus, we first checked for each associate whether it appeared in the same sentence. In the positive case, we determined the shortest dependency path between the stimulus and the associate. The analysis was again token-based: That is, we took the strength between the stimulus–associate pairs into account—how often the associate was provided for a stimulus.

Table 7 presents the ten most frequent paths that exist between the stimuli and the associates, accompanied by a stimulus–associate pair example and by the total frequency, indicating how often the paths appear over all stimulus–associate pairs. The paths start with the part of speech of the stimulus and end with the part of speech of the associate response. The tag “NN” refers to common nouns; “ADJA” to attributive adjectives, “APPR(ART)” to (portmanteau) prepositions, “KON” to conjunctions, and any part of speech starting with “VV” to a main verb.

The semantics of the paths (at least of those that are similarly short as those in the examples) is quite obvious. For example, “NN/KON/NN” refers to the stimulus and the associate appearing as two conjuncts (such as *Hund* “dog” and *Katze* “cat”); “NN/ADJA” to an attributive adjective that directly modifies the noun stimulus (such as *grün* “green”

**Table 6** Corpus co-occurrence of stimulus–associate pairs

		Co-Occurrence Strength					
		1	2	5	10	20	50
Window 5	All stimuli	73 %	65 %	53 %	44 %	34 %	23 %
	Compound stimuli	58 %	46 %	30 %	20 %	11 %	4 %
Window 20	All stimuli	79 %	72 %	62 %	53 %	43 %	31 %
	Compound stimuli	67 %	56 %	41 %	29 %	18 %	9 %

<sup>5</sup> Schulte im Walde and Müller (2013) performed a detailed comparison of stimulus–associate corpus co-occurrences across corpora, varying the corpus sizes and corpus domains.

**Table 7** Most frequent dependency paths between stimuli and associates

Path	Example Stimulus–Associate Pair	Total Frequency
NN/NN	<i>Tasse</i> “cup”– <i>Kaffee</i> “coffee”	3,377
NN/APPR/NN	<i>Hütte</i> “shed”– <i>Holz</i> “wood”	2,357
NN/ADJA	<i>Salat</i> “salad”– <i>grün</i> “green”	3,343
NN/KON/NN	<i>Hund</i> “dog”– <i>Katze</i> “cat”	2,235
NN/VVFIN	<i>Buch</i> “book”– <i>lesen</i> “read”	1,163
NN/VVFIN/APPR/NN	<i>Kokosnuss</i> “coconut”– <i>Palme</i> “palm”	949
NN/VVFIN/NN	<i>Maulwurf</i> “mole”– <i>Hügel</i> “hill”	941
NN/APPR/VVFIN/NN	<i>Papierkorb</i> “bin”– <i>Abfall</i> “trash”	888
NN/APPR/VVFIN	<i>Brille</i> “glasses”– <i>watch</i> “sehen”	661
NN/APPRART/NN	<i>Hut</i> “hat”– <i>Kopf</i> “head”	610

for *Salat* “salad”), “NN/APPR/NN” to a syntactic construction in which either the stimulus or the associate depends on a preposition that itself depends on the other noun (such as *Hütte/aus/Holz* “shed/made from/wood”). One could specify the paths in more detail (e.g., by adding the dependency direction or by specifying the prepositional heads), or generalize more about some aspects (e.g., by summarizing over the various forms of the verbs). However, the present version seems appropriate for a first impression of the syntactic dependencies between the stimuli and their associates.

Table 8 lists the five strongest examples for the five most prominent syntactic paths.<sup>6</sup> The path strength is that portion of the association strength for which the respective path between the stimulus and the associate was found in the parsed corpus (since there might have been other paths between the stimulus–associate pair). For example, 74.39 % of the dependency paths between *Tasse* “cup” and *Kaffee* “coffee” were NN/NN, so given an association strength of 36, the path strength is  $36 \times .7439 = 26.78$ .

In order to look into the path instances, we checked on the actual corpus parses of the sentences in which stimuli and associates appeared with the given paths. In most cases, our initial intuitions about the semantics of the paths were confirmed. For example, the NN/ADJA cases all refer to adjectival noun modifiers; the NN/KON/NN cases refer to conjuncts (often representing collocations) such as *Pfeil und Bogen* “bow and arrow,” *Hund und Katze* “dog and cat,” and *Kaffee und Kuchen* “coffee and cake”; and the NN/VVFIN cases refer to nominal complements of verbs (such as subjects, as in *Hund-bellen* “dog barks” and *Telefon-klingeln* “phone rings,” or direct objects, as in *Brief-schreiben* “write letter”). The NN/NN cases,

however, refer to measure constructions such as *Tasse Tee/Kaffee* “a cup of tea/coffee,” or genitive constructions such as *Dach eines Hauses* “roof of a house,” which we had not predicted in advance. Looking into actual sentences also helps to specify the NN/APPR/NN cases, by investigating the prepositional heads. For example, there are instances of *Zeit zwischen . . . Uhr und . . . Uhr* “time between . . . o’clock and . . . o’clock,” and *Zeit von . . . Uhr bis . . . Uhr* “time from . . . o’clock until . . . o’clock”; *Thermoskanne mit Tee* “thermos with tea,” and *Tee aus der Thermoskanne* “tea from the thermos”; *Lampe auf dem Nachttisch* “lamp on the bed table”; *Hütte aus Holz* “hut made out of wood”; *Wasser auf die Mühle* “water onto the mill” (note that the latter example is part of a collocation, meaning “grist to the mill,” encouragement). In the NN/APPR/NN cases, a refinement of the paths with prepositions types would obviously be a useful extension.

In sum, the dependency paths provide interesting and diverse insights into salient syntactic dependencies between the stimuli and the associates. To our knowledge, none of the previous related work had explored the syntactic paths in association norms. Rather, Schulte im Walde et al. (2008) explored parse tuples for noun–verb pairs to identify potential syntactic dependencies between stimulus nouns and associate verbs, and between stimulus verbs and associate nouns. The present analysis provides a broader spectrum of syntactic information, such that not only noun–verb pairs are involved. Although the above information comprises both compound and constituent stimuli, the strongest syntactic paths for only compounds strongly overlap with this larger set.

### Semantic relations

The final analysis concerned types of semantic relationships (such as synonymy) that hold between the stimulus words and the associate responses. This analysis provides insight into which semantic relations the experiment’s participants might have had in mind when they provided associates. We relied on the German WordNet, *GermaNet* (Kunze, 2000), to explore the semantic information, and used GermaNet version 6.0, which was the latest version when the study started, comprising 69,594 synsets and 93,407 lexical units. We considered the paradigmatic relations *synonymy*, *antonymy*, *hypernymy*, *hyponymy*, and *co-hyponymy*. The synonymy information was based on the 69,594 synsets; in addition, GermaNet 6.0 contains a total of 74,945 hypernymy relations, and 1,587 antonymy relations.

The analysis was again token-based—that is, incorporating the strengths between stimuli and their associates. So, if a relationship between a stimulus and an associate was found, it was instantiated by the strength between them. For example, if the associate *Obst* “fruit” was provided 12 times for the stimulus *Apfel* “apple,” the strength of the semantic relation *hypernymy* that combined the two lexemes in GermaNet was

<sup>6</sup> Note that for ambiguous stimuli and associates, only the translation that refers to the sense in the respective context is provided.

**Table 8** Examples of stimulus–associate pairs and their dependency paths

Stimulus	Associate	Association Strength	Path	Path Strength
<i>Tasse</i> “cup”	<i>Kaffee</i> “coffee”	36	NN/NN	26.78
<i>Tasse</i> “cup”	<i>Tee</i> “tea”	35		19.89
<i>Dach</i> “roof”	<i>Haus</i> “house”	67		18.50
<i>Blatt</i> “sheet”	<i>Papier</i> “paper”	27		17.78
<i>Kanne</i> “pot”	<i>Kaffee</i> “coffee”	40		17.20
<i>Uhr</i> “o’clock”	<i>Zeit</i> “time”	48	NN/APPR/NN	16.04
<i>Thermoskanne</i> “thermos”	<i>Tee</i> “tea”	34		10.66
<i>Nachttisch</i> “bed table”	<i>Lampe</i> “lamp”	21		10.50
<i>Hütte</i> “shed”	<i>Holz</i> “wood”	29		10.26
<i>Mühle</i> “mill”	<i>Wasser</i> “water”	16		9.29
<i>Jeans</i> “jeans”	<i>blau</i> “blue”	48	NN/ADJA	26.57
<i>Licht</i> “light”	<i>hell</i> “bright”	54		25.62
<i>Salat</i> “salad”	<i>grün</i> “green”	32		23.00
<i>Ziegelstein</i> “brick”	<i>rot</i> “red”	28		22.84
<i>Messer</i> “knife”	<i>scharf</i> “sharp”	28		21.15
<i>Bogen</i> “bow”	<i>Pfeil</i> “arrow”	52	NN/KON/NN	35.26
<i>Hund</i> “dog”	<i>Katze</i> “cat”	33		17.31
<i>Kuchen</i> “cake”	<i>Kaffee</i> “coffee”	21		14.98
<i>Tisch</i> “table”	<i>Stuhl</i> “chair”	31		13.60
<i>Mann</i> “man”	<i>Frau</i> “woman”	30		11.97
<i>Hund</i> “dog”	<i>bellen</i> “bark”	46	NN/VVFIN	20.52
<i>Telefon</i> “phone”	<i>klingseln</i> “ring”	22		18.76
<i>Hahn</i> “cock”	<i>krähen</i> “craw”	26		18.09
<i>Brief</i> “letter”	<i>schreiben</i> “write”	41		13.11
<i>Buch</i> “book”	<i>lesen</i> “read”	84		11.83

instantiated by 12. If another hypernym of *Apfel* (such as *Frucht* “fruit”) was found among the associates, the stimulus–associate strength was added to the existing hypernym strength. So, on the one hand, we determined the strengths between individual stimulus–associate pairs, and on the other hand, we could summarize the overall associations for a certain stimulus (and also for all stimuli), in order to identify the strengths of the various semantic relations with regard to a specific stimulus, and also over all stimuli. Table 9 presents the 20 strongest semantic relationships between individual compound stimuli and their associates.

In total, we found semantic relation information for 426 of our 571 stimuli (75 %). If we only look at the compound stimuli, the proportion is much lower, at 59 %, covering associations to 145 out of the 246 compounds. It is very impressive, however, that all of the stimuli are included in GermaNet. That is, GermaNet covers not only the simple nouns in our data set, but also all of the compounds. The missing proportions are due to nonexistent relationships between the stimuli and the associates; for this, the coverage was lower for the compounds than for the whole set of stimuli, which is not surprising.

Table 10 shows the proportions of GermaNet relations that were identified for the stimulus–associate pairs. The numbers in columns 2 and 3 correlate with the total proportions of semantic relation instances that are coded in GermaNet, so there are no specifically strong semantic relations among the stimulus–associate pairs. Looking at the associations to the compound stimuli in columns 4 and 5, however, demonstrates that with regard to the compound stimuli, it is more intuitive, and thus easier, to generate hypernyms and co-hyponyms than hyponyms, since the compounds are often very specific. There are only 26 hyponyms to the compound stimuli in total, with the most prominent examples being *Taschentuch–Tempo* and *Kreditkarte–VISA*, which appear among the top 20 pairs in Table 9. In sum, we found an interesting difference in the semantic relations of the stimulus–associate pairs when looking at the compounds in comparison to all stimulus nouns.

#### Associations and semantic transparency

Our main motivation to collect the association norms for the German noun compounds and their constituents was that the

**Table 9** Examples of semantic relations between stimulus–associate pairs

Stimulus	Associate	Semantic Relation	Strength
<i>Seezunge</i> “sole”	<i>Fisch</i> “fish”	hypemym	31
<i>Ohrring</i> “earring”	<i>Schmuck</i> “jewelry”	hypemym	24
<i>Mohrrübe</i> “carrot”	<i>Karotte</i> “carrot”	synonym	22
<i>Schlittenhund</i> “sledge dog”	<i>Husky</i> “husky”	co-hyponym	21
<i>Armbrust</i> “crossbow”	<i>Pfeil</i> “arrow”	co-hyponym	19
<i>Eisenbahn</i> “railway”	<i>Zug</i> “train”	synonym	17
<i>Wirbelsäule</i> “spine”	<i>Knochen</i> “bone”	hypemym	16
<i>Truthahn</i> “turkey cock”	<i>Pute</i> “turkey hen”	co-hyponym	14
<i>Meerschweinchen</i> “guinea pig”	<i>Haustier</i> “pet”	hypemym	14
<i>Federball</i> “badminton”	<i>Badminton</i> “badminton”	co-hyponym	14
<i>Taschentuch</i> “tissue”	<i>Tempo</i> “tissue brand name”	hyponym	13
<i>Bildschirm</i> “monitor”	<i>Fernseher</i> “television”	co-hyponym	13
<i>Kreditkarte</i> “credit card”	<i>VISA</i> “VISA”	hyponym	11
<i>Bullauge</i> “bull”s eye”	<i>Fenster</i> “window”	hypemym	11
<i>Wäscheklammer</i> “clothespin”	<i>Wäscheleine</i> “clothes line”	co-hyponym	10
<i>Skistock</i> “ski stick”	<i>Ski</i> “ski”	co-hyponym	10
<i>Seemann</i> “seaman”	<i>Matrose</i> “sailor”	co-hyponym	10
<i>Fingerhut</i> “thimble”	<i>Nadel</i> “needle”	co-hyponym	10
<i>Brombeere</i> “blackberry”	<i>Himbeere</i> “raspberry”	co-hyponym	10
<i>Schornstein</i> “chimney”	<i>Kamin</i> “chimney”	synonym	9

associations provide insight into the semantic properties of the compounds and their constituents, and should therefore represent a valuable resource for cognitive and computational linguistics research on semantic transparency, and for lexical semantics in general. More specifically, we are interested in the degree of semantic transparency of the compound nouns with regard to their constituents, and are currently investigating whether the degree of overlap of associations is indeed an indicator of the semantic relatedness between compounds and their constituents. The examples in Tables 2, 3, and 4 give a first impression that this might indeed be the case, and in the following discussion we will provide more evidence for our hypothesis.

**Table 10** Percentages of semantic relations among stimulus–associate pairs according to GermaNet

Semantic Relation	Found in GermaNet			
	All Stimuli		Compound Stimuli	
co-hyponymy	2,410	43 %	408	48 %
hyponymy	1,324	23 %	71	8 %
hypemymy	1,249	22 %	260	31 %
synonymy	646	11 %	113	13 %
antonymy	38	1 %	0	0 %

We applied a measure suggested by Schulte im Walde et al. (2012) that relies on a simple association overlap to predict the degree of semantic transparency of the experimental compound nouns: That is, we used the proportion of shared associations of a compound and its constituent with respect to the total number of associations of the compound. The degree of semantic transparency of a compound noun was calculated with respect to each constituent of the compound. As an example of the calculation, when considering the ten most frequent responses of the compound noun *Ahornblatt* “maple leaf” and its constituents, provided in Table 2, the compound noun received a total of 99 associate tokens, out of which it shared 87 with the first constituent *Ahorn* “maple,” and 52 with the second constituent *Blatt* “leaf.” Thus, the predicted degrees of semantic transparency were  $87/99 = .88$  for *Ahornblatt*–*Ahorn*, and  $52/99 = .53$  for *Ahornblatt*–*Blatt*. These predicted degrees of semantic transparency were compared against the mean transparency judgments collected by von der Heide and Borgwaldt (2009),<sup>7</sup> using the Spearman rank-order correlation coefficient  $\rho$ . Our hypothesis was that the larger the overlap of the associations of a compound and a

<sup>7</sup> Von der Heide and Borgwaldt (2009) collected transparency ratings for their 450 compounds: 30 native German speakers were asked to rate the semantic transparency of the compounds with respect to each of their constituents, on a scale from 1 (*opaque*) to 7 (*transparent*). For more details on the ratings, see Schulte im Walde, Müller, and Roller (2013).



constituent, the stronger would be the degree of compound–constituent transparency.

The resulting correlation values are  $\rho = .6227$  (taking both constituent types into account),  $\rho = .6128$  for the compound–modifier pairs, and  $\rho = .6547$  for the compound–head pairs. The simple association overlap measure therefore exceeds moderate correlation values, going up to .6547. Looking at the two experiments separately, the results are slightly better for the AMT norms than for the standard Web data, and thus confirm the semantic usefulness of both experimental setups. In addition, the overlap measure is stronger for heads than for modifiers.

This case study shows that the associations provide insight into the semantic properties of the compounds (and their constituents) that should be useful for models of compounds’ semantic transparency. Specifically, the data seem to indicate that associations to compound nouns comprise associations to both the head and modifier; that is, they reflect the meaning components of both constituents. Since the overlap between constituents and compounds correlates with transparency ratings, we can conclude that associations to compound nouns include fewer associations to their opaque than to their transparent constituents. That is, for a transparent compound like *Schlittenhund* “sledge dog,” we would expect to find associations to both components *Schlitten* “sledge” and *Hund* “dog,” whereas in a partially opaque compound like *Fliegenpilz* “toadstool,” the associations might be more related to the transparent constituent *Pilz* “mushroom” than to the opaque constituent *Fliege* “fly/bowtie.” This result is in accordance with Libben’s APPLE model (1994, 1998), which assumes that opaque constituents are not conceptually activated during processing.

## Conclusion

In the present study, we have presented a collection of association norms for German noun compounds and their constituents. The collection was conducted via crowdsourcing and was compared to an earlier Web experiment. The norms are aimed to be of use for researchers on compound processing, containing association norms not only for compounds but also for their constituents, in order to provide a resource for stimulus selection in experimental studies (e.g., Sandra, 1990). Second, by examining properties of the associations for compounds and their constituents in detail, our data could provide more insight into the way in which the meanings of compounds are being processed and/or represented in the mental lexicon—that is, they could complement the results of online experiments.

In total, the new association norms for 246 noun–noun compounds and their constituents comprise 58,652 associates (tokens), distributed over 26,004 stimulus–associate pairs (types). The quantitative contribution of the crowdsourcing experiment is slightly larger than that of the previous Web experiment (58,652 vs. 47,249 stimulus–associate tokens), even though it ran for only a quarter of the time of the previous study.

We analyzed the stimulus–associate pairs of both the present and the previous experiments in four ways, performing (1) a morpho-syntactic analysis, (2) a distributional analysis, (3) a syntactic analysis, and (4) a semantic relation analysis. Concerning the first of these, a part-of-speech analysis of the associate responses, and the third, a dependency path analysis of the stimulus–associate pairs, there were no noticeable differences when comparing the associations of the compounds and the associations of nouns in general. Concerning the second analysis, the window co-occurrence of stimulus–associate pairs, and the fourth, the semantic relations between stimuli and associates, we demonstrated:

- Although we could confirm the co-occurrence hypothesis for our compound nouns, the corpus coverage pointed to the caveat that compounds are typically covered worse by corpus data than are simple nouns. This demonstrates the importance of a reasonable morphological annotation of compounds, and that we have considerably less corpus information for compound nouns than for nouns in general.
- The semantic relations between associates and compound stimuli differ from those between associates and nouns in general. Not surprisingly, we found fewer hyponyms among the associates, but more hypernyms and co-hyponyms.

Finally, the overlap between associations to compounds and to their constituents correlated moderately with semantic transparency ratings obtained from human raters (von der Heide & Borgwaldt, 2009), providing additional insight into how the meanings of complex words are related to the meanings of their parts. In summary, the associations could in general represent a valuable resource concerning the lexical semantic properties of the compound stimuli and the semantic relations between the stimuli and their associations, and—more specifically—they could be useful in future cognitive and computational linguistic research on semantic transparency.

**Author note** This work was supported by Heisenberg Fellowship SCHU-2580/1 from the Deutsche Forschungsgemeinschaft (Sabine Schulte im Walde). We also thank Ronny Jauch, who performed parts of the analyses in [Overall results and analyses](#) for his diploma thesis (Jauch, 2012), and the anonymous reviewers for their valuable comments.

## Appendix A

In the following, we provide a list of our subset of bi-morphemic noun compounds from von der Heide & Borgwaldt (2009), including the compound constituents, and the categorisation label for „NN“ (noun-noun compound) vs. „unique“ (noun compound with unique modifier).

Ahornblatt	Ahorn	Blatt	NN	Eisstadion	Eis	Stadion	NN
Armband	Arm	Band	NN	Eiswürfel	Eis	Würfel	NN
Armbrust	Arm	Brust	NN	Ellenbogen	Elle	Bogen	NN
Aschenbecher	Asche	Becher	NN	Erdbeere	Erde	Beere	NN
Bahnhof	Bahn	Hof	NN	Erdnuss	Erde	Nuss	NN
Bananenschale	Banane	Schale	NN	Espressomaschine	Espresso	Maschine	NN
Bärlauch	Bär	Lauch	NN	Federball	Feder	Ball	NN
Basketball	Basket	Ball	NN	Federboa	Feder	Boa	NN
Baumhaus	Baum	Haus	NN	Feldsalat	Feld	Salat	NN
Bettwäsche	Bett	Wäsche	NN	Feuerwerk	Feuer	Werk	NN
Bierfass	Bier	Fass	NN	Feuerzeug	Feuer	Zeug	NN
Bildschirm	Bild	Schirm	NN	Fieberthermometer	Fieber	Thermometer	NN
Billardtisch	Billiard	Tisch	NN	Filzstift	Filz	Stift	NN
Bleistift	Blei	Stift	NN	Fingerhut	Finger	Hut	NN
Blockflöte	Block	Flöte	NN	Fleischwolf	Fleisch	Wolf	NN
Blumenkohl	Blume	Kohl	NN	Fliegenklatsche	Fliege	Klatsche	NN
Blumenstrauß	Blume	Strauß	NN	Fliegenpilz	Fliege	Pilz	NN
Blumenvase	Blume	Vase	NN	Flohmarkt	Floh	Markt	NN
Brautkleid	Braut	Kleid	NN	Fotoalbum	Foto	Album	NN
Briefkasten	Brief	Kasten	NN	Fotoapparat	Foto	Apparat	NN
Briefmarke	Brief	Marke	NN	Fußball	Fuß	Ball	NN
Briefpapier	Brief	Papier	NN	Fußleiste	Fuß	Leiste	NN
Briefträger	Brief	Träger	NN	Gesangsbuch	Gesang	Buch	NN
Brillenetui	Brille	Etui	NN	Gewächshaus	Gewächs	Haus	NN
Brombeere	Brom	Beere	unique	Glockenspiel	Glocke	Spiel	NN
Bullauge	Bulle	Auge	NN	Golfball	Golf	Ball	NN
Büroklammer	Büro	Klammer	NN	Gummibärchen	Gummi	Bärchen	NN
Cowboyhut	Cowboy	Hut	NN	Gummiente	Gummi	Ente	NN
Cowboystiefel	Cowboy	Stiefel	NN	Gummistiefel	Gummi	Stiefel	NN
Dachboden	Dach	Boden	NN	Gürteltier	Gürtel	Tier	NN
Dachfenster	Dach	Fenster	NN	Haarreifen	Haar	Reifen	NN
Dachstuhl	Dach	Stuhl	NN	Hahnenfuß	Hahn	Fuß	NN
Dominostein	Domino	Stein	NN	Halskette	Hals	Kette	NN
Eidechse	Eid	Echse	unique	Handcreme	Hand	Creme	NN
Eisbär	Eis	Bär	NN	Handschuh	Hand	Schuh	NN
Eisberg	Eis	Berg	NN	Handtasche	Hand	Tasche	NN
Eisenbahn	Eisen	Bahn	NN	Handtuch	Hand	Tuch	NN
				Haselnuss	Hasel	Nuss	NN
				Hausboot	Haus	Boot	NN
				Heidelbeere	Heide	Beere	NN
				Heuhaufen	Heu	Haufen	NN
				Himbeere	Him	Beere	unique
				Hirschkäfer	Hirsch	Käfer	NN
				Hollywoodschaukel	Hollywood	Schaukel	NN
				Hornbrille	Horn	Brille	NN
				Hufeisen	Huf	Eisen	NN

Hundehütte	Hund	Hütte	NN	Nachttisch	Nacht	Tisch	NN
Hüttenkäse	Hütte	Käse	NN	Nadelkissen	Nadel	Kissen	NN
Jägerzaun	Jäger	Zaun	NN	Nagelfeile	Nagel	Feile	NN
Jeanshemd	Jeans	Hemd	NN	Nagellack	Nagel	Lack	NN
Jeansjacke	Jeans	Jacke	NN	Nasenbär	Nase	Bär	NN
Kaffeemaschine	Kaffee	Maschine	NN	Nashorn	Nase	Horn	NN
Kaffeemühle	Kaffee	Mühle	NN	Nilpferd	Nil	Pferd	NN
Kaffeepad	Kaffee	Pad	NN	Nudelholz	Nudel	Holz	NN
Kettensäge	Kette	Säge	NN	Nummernschild	Nummer	Schild	NN
Kleiderschrank	Kleider	Schrank	NN	Obstkuchen	Obst	Kuchen	NN
Knoblauch	Knob	Lauch	unique	Ohrring	Ohr	Ring	NN
Kokosnuss	Kokos	Nuss	NN	Papierkorb	Papier	Korb	NN
Kopfkissen	Kopf	Kissen	NN	Pfannkuchen	Pfanne	Kuchen	NN
Kopfsalat	Kopf	Salat	NN	Pfauenfeder	Pfau	Feder	NN
Kreditkarte	Kredit	Karte	NN	Pfeffermühle	Pfeffer	Mühle	NN
Kreissäge	Kreis	Säge	NN	Postbote	Post	Bote	NN
Kronkorken	Krone	Korken	NN	Postkarte	Post	Karte	NN
Kronleuchter	Krone	Leuchter	NN	Pudelmütze	Pudel	Mütze	NN
Kuckucksuhr	Kuckuck	Uhr	NN	Radkappe	Rad	Kappe	NN
Kuhfladen	Kuh	Fladen	NN	Reetdach	Reet	Dach	NN
Kulturbeutel	Kultur	Beutel	NN	Regenbogen	Regen	Bogen	NN
Kürbiskern	Kürbis	Kern	NN	Regenmantel	Regen	Mantel	NN
Lachsschinken	Lachs	Schinken	NN	Regenrinne	Regen	Rinne	NN
Latzhose	Latz	Hose	NN	Regenschirm	Regen	Schirm	NN
Leberwurst	Leber	Wurst	NN	Ringfinger	Ring	Finger	NN
Lederhose	Leder	Hose	NN	Rittersporn	Ritter	Sporn	NN
Lichtschalter	Licht	Schalter	NN	Rucksack	Ruck	Sack	NN
Löwenzahn	Löwe	Zahn	NN	Sandburg	Sand	Burg	NN
Luftmatratze	Luft	Matratze	NN	Sanduhr	Sand	Uhr	NN
Luftpumpe	Luft	Pumpe	NN	Schachbrett	Schach	Brett	NN
Maisfeld	Mais	Feld	NN	Schildkröte	Schild	Kröte	NN
Maiskolben	Mais	Kolben	NN	Schlauchboot	Schlauch	Boot	NN
Marienkäfer	Maria	Käfer	NN	Schlittenhund	Schlitten	Hund	NN
Maßstab	Maß	Stab	NN	Schlüsselbund	Schlüssel	Bund	NN
Maulwurf	Maul	Wurf	NN	Schneckenhaus	Schnecke	Haus	NN
Mausefalle	Maus	Falle	NN	Schneeball	Schnee	Ball	NN
Meerschweinchen	Meer	Schweinchen	NN	Schneemann	Schnee	Mann	NN
Mettwurst	Mett	Wurst	NN	Schnittlauch	Schnitt	Lauch	NN
Mikadostäbchen	Mikado	Stäbchen	NN	Schornstein	Schom	Stein	unique
Milchshake	Milch	Shake	NN	Schulbuch	Schule	Buch	NN
Mohrrübe	Mohr	Rübe	NN	Schwertfisch	Schwert	Fisch	NN
Motorhaube	Motor	Haube	NN	Seehund	See	Hund	NN
Motorrad	Motor	Rad	NN	Seemann	See	Mann	NN
Mülleimer	Müll	Eimer	NN	Seerose	See	Rose	NN
Mülltonne	Müll	Tonne	NN	Seestern	See	Stern	NN
Mundharmonika	Mund	Harmonika	NN	Seezunge	See	Zunge	NN

Seifenblase	Seife	Blase	NN	Vogelhaus	Vogel	Haus	NN
Seilbahn	Seil	Bahn	NN	Vogelkäfig	Vogel	Käfig	NN
Sessellift	Sessel	Lift	NN	Walnuss	Wal	Nuss	unique
Skistock	Ski	Stock	NN	Wäscheklammer	Wäsche	Klammer	NN
Sonnenblume	Sonne	Blume	NN	Wasserfall	Wasser	Fall	NN
Sonnenbrille	Sonne	Brille	NN	Wasserhahn	Wasser	Hahn	NN
Sonnencreme	Sonne	Creme	NN	Wassermelone	Wasser	Melone	NN
Sonnenschirm	Sonne	Schirm	NN	Wasserwaage	Wasser	Waage	NN
Sonnenuhr	Sonne	Uhr	NN	Weihnachtsbaum	Weihnachten	Baum	NN
Spiegelei	Spiegel	Ei	NN	Werwolf	Wer	Wolf	unique
Spinnennetz	Spinne	Netz	NN	Windlicht	Wind	Licht	NN
Stachelbeere	Stachel	Beere	NN	Windmühle	Wind	Mühle	NN
Stacheldraht	Stachel	Draht	NN	Wintermantel	Winter	Mantel	NN
Stereoanlage	Stereo	Anlage	NN	Wirbelsäule	Wirbel	Säule	NN
Sternschnuppe	Stern	Schnuppe	NN	Wollschal	Wolle	Schal	NN
Strandkorb	Strand	Korb	NN	Würfelzucker	Würfel	Zucker	NN
Straßenbahn	Straße	Bahn	NN	Zahnbürste	Zahn	Bürste	NN
Strohhalbm	Stroh	Halm	NN	Zahnkrone	Zahn	Krone	NN
Strumpfhose	Strumpf	Hose	NN	Zahnpasta	Zahn	Pasta	NN
Suppenteller	Suppe	Teller	NN	Zahnrad	Zahn	Rad	NN
Tannenzapfen	Tanne	Zapfen	NN	Zahnseide	Zahn	Seide	NN
Taschenbuch	Tasche	Buch	NN	Zahnspange	Zahn	Spange	NN
Taschenlampe	Tasche	Lampe	NN	Zeitschrift	Zeit	Schrift	NN
Taschenmesser	Tasche	Messer	NN	Ziegelstein	Ziegel	Stein	NN
Taschentuch	Tasche	Tuch	NN	Zitronenpresse	Zitrone	Presse	NN
Teddybär	Teddy	Bär	NN	Zollstock	Zoll	Stock	NN
Teebeutel	Tee	Beutel	NN	Zuckerhut	Zucker	Hut	NN
Teekanne	Tee	Kanne	NN	Zuckerwatte	Zucker	Watte	NN
Teelicht	Tee	Licht	NN				
Teelöffel	Tee	Löffel	NN				
Teetasse	Tee	Tasse	NN				
Telefonbuch	Telefon	Buch	NN				
Telefonhörer	Telefon	Hörer	NN				
Telefonzelle	Telefon	Zelle	NN				
Tennisball	Tennis	Ball	NN				
Tennisschläger	Tennis	Schläger	NN				
Thermoskanne	Thermo	Kanne	unique				
Tintenfisch	Tinte	Fisch	NN				
Toilettenpapier	Toilette	Papier	NN				
Truthahn	Trut	Hahn	unique				
Türklinke	Tür	Klinke	NN				
Vanilleeis	Vanille	Eis	NN				
Visitenkarte	Visite	Karte	NN				

## Appendix B

In the following, we provide the experiment setups for the association collection via AMT in the current study, and for the association collection through a standard web experiment in the previous study (Schulte im Walde et al., 2012). The AMT setup of the current collection is shown by Fig. 1, with translations in red font. The actual collection of the associations was performed as shown in Fig. 2, translated by Fig. 3. The web collection of the associations was performed as shown in Fig. 4, translated by Fig. 5.



Describe your HIT to Workers

Title

*Associations to complex German nouns (Part 1)*  
Assoziationen zu zusammengesetzten deutschen Nomen (Teil 1)

Description

*Write down what spontaneously comes to mind.*  
Schreibe auf, was Dir spontan zu einem Nomen-Kompositum einfällt.

Keywords

kompositum, compound, semantik, semantics, nomen, noun, assoziation, association, spontan, spontaneous

☐ This project may contain potentially explicit or offensive content, for example, nudity. [\(See details\)](#)

Setting up your HIT

Reward per assignment

\$ 0.02

Tip: Consider how long it will take a Worker to complete each task. A 30 second task that pays \$0.05 is a \$6.00 hourly wage.

Number of assignments per HIT

25

How many unique Workers do you want to work on each HIT?

Time allotted per assignment

10 Minutes

Maximum time a Worker has to work on a single task. Be generous so that Workers are not rushed.

HIT expires in

7 Days

Maximum time your HIT will be available to Workers on Mechanical Turk.

Results are automatically approved in

7 Days

After this time, all unreviewed work is approved and Workers are paid.

Advanced

Worker requirements:

Customize Worker Requirements...

Specify ALL the qualifications Workers must meet to work on your HITs:

-- Select -- equal to 0 clear

(+ Add another criterion (up to 5)

Only Workers who qualify to do my HITs can preview my HITs.

☐ Yes ☒ No

Fig. 1 AMT setup

## Assoziationen zu deutschen Nomen

In diesem Experiment geht es um **Nomen-Assoziationen**.

*Aufgabe:* Wir präsentieren Nomen, und Du schreibst Wörter auf, die Dir **spontan** dazu einfallen. Pro Nomen stehen bis zu drei Felder zur Verfügung.

*Beispiel:* Woran denkst Du spontan bei dem Wort **Schnee**?

*Winter, kalt, Schlitten*

Wichtige Hinweise:

- Nur für **deutsche Muttersprachler - Native speakers of German** only.
- Bitte nicht lange nachdenken, sondern **spontan** und **zügig** antworten.
- Bitte nur **ein Wort pro Zeile** schreiben.
- Bitte **alle 24 HITs** bearbeiten.
- Wir haben als Test einige Wörter in den Daten versteckt, die von Muttersprachlern als "erfunden" erkannt werden sollten. In einem solchen Fall bitte die Auswahl "Kenne ich nicht!" treffen.

*Assoziationen zu  $\{word\}$  :*


☐ Ich kenne das Nomen " $\{word\}$ " nicht!

**Fig. 2** Background for AMT experiment

## **Associations to German Nouns**

This experiment is about **associations to nouns**.

*Task:* You will see nouns, and we ask you to list words that **spontaneously** come to mind. For each noun you will be provided three lines in which to list associations.

*Example:* What comes to mind when you read the word **snow**?

*winter, cold, sledge*

Important notes:

- The task is for **native speakers of German** only.
- Please respond **spontaneously** and **quickly**.
- Please only write **one word per line**.
- Please do **all 24 HITs**.
- As test cases to control for spammers, we hid some non-existing words in the data that you should identify as such. In these cases, please check the "I don't know!" button.

***Associations to  $\${word}$  :***


☐ I don't know the noun " $\${word}$ "!

**Fig. 3** Background for AMT experiment, translated into English

# Nomen-Assoziationen 2010

## Idee und Aufgabe

In diesem Experiment geht es um **Wortassoziationen**. Wir präsentieren Nomen, und Sie schreiben Wörter auf, die Ihnen **spontan** dazu einfallen. Pro Nomen stehen Ihnen bis zu drei Felder für Assoziationen zur Verfügung.

*Beispiel:*

Woran denken Sie spontan bei dem Wort **Schnee**?

*Winter, kalt, Schlitten*

Bevor das Experiment losgeht, können Sie nachfolgend Angaben zu Ihrer Person machen. Betätigen Sie den Knopf "Weiter", um auf eine Beispielseite mit Instruktionen zu kommen. *Bitte beachten Sie:* Wenn Sie einmal angefangen haben, sollten Sie das Experiment nicht unterbrechen.

## Persönliche Angaben

Die folgenden Angaben sind freiwillig. Es ist allerdings hilfreich für die Auswertung des Experiments, wenn Sie uns die Informationen zur Verfügung stellen, besonders bei den mit \* gekennzeichneten Feldern. Die Daten werden selbstverständlich nur zur Auswertung des Experiments verwendet.

Name

Alter

\*Muttersprache

\*In welchem Land sind Sie aufgewachsen?

\*In welcher Region sind Sie aufgewachsen?

\*Haben Sie Kenntnisse in Linguistik (nicht aus der Schule)?

- ☐ ja  
☐ nein

Beruf

Email

Weiter

Fig. 4 Background to previous association experiment in Schulte im Walde et al. (2012).



# Noun Associations 2010

## Idea and Task

This experiment is about **word associations**. You will see nouns, and we ask you to list words that **spontaneously** come to mind. For each noun you will be provided three lines in which to list associations.

*Example:*

What comes to mind when you read the word **snow**?

*winter, cold, sledge*

Before the experiment starts, we kindly ask you for some personal information. Then press the button "continue" to proceed to an example page with instructions. *Please note:* Once you have started the experiment, please do not pause until finished.

## Personal Information

For evaluating the associations of the study, it would be very helpful for us to have as much background about the participants as possible, especially regarding fields marked by \*. We will not use the data for anything other than evaluating the study.

Name

Age

\*Native Language

\*In which country did you grow up?

\*In which region did you grow up?

\*Do you have a background in linguistics (professional or academic)?

- ☐ yes  
☐ no

Profession

Email

Fig. 5 Background to previous association experiment, translated into English

## References

- Baayen, R. H., & Schreuder, R. (1999). War and peace: Morphemes and full forms in a noninteractive activation parallel dual-route model. *Brain and Language*, 68, 27–32.
- Baroni, M., Bernardini, S., Ferraresi, A., & Zanchetta, E. (2009). The WaCky Wide Web: A collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation*, 43, 209–226.
- Bertram, R., & Hyönä, J. (2003). The length of a complex word modifies the role of morphological structure: Evidence from reading short and long Finnish compounds. *Journal of Memory and Language*, 48, 615–634.
- Bertram, R., Kuperman, V., Baayen, R. H., & Hyönä, J. (2011). The hyphen as a segmentation cue in triconstituent compound processing: It's getting better all the time. *Scandinavian Journal of Psychology*, 52, 530–544.
- Bohnet, B. (2010). Top accuracy and fast dependency parsing is not a contradiction. In *Proceedings of the 23rd International Conference on Computational Linguistics* (pp. 89–97). Stroudsburg, PA: Association for Computational Linguistics.
- Butterworth, B. (1983). Lexical representation. In B. Butterworth (Ed.), *Language production* (Vol. 2, pp. 257–294). New York, NY: Academic Press.
- Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition*, 28, 297–332.
- Church, K., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16, 22–29.
- Clark, H. (1971). Word associations and linguistic theory. In J. Lyons (Ed.), *New horizons in linguistics* (pp. 271–286). Baltimore, MD: Penguin.
- Comesaña, M., Fraga, I., Moreira, A. J., Frade, C. S., & Soares, A. P. (2014). Free associate norms for 139 European Portuguese words for children from different age groups. *Behavior Research Methods*, 46, 564–574. doi:10.3758/s13428-013-0388-0
- De Deyne, S., & Storms, G. (2008). Word associations: Norms for 1,424 Dutch words in a continuous task. *Behavior Research Methods*, 40, 198–205. doi:10.3758/BRM.40.1.198
- de Jong, N. H., Feldman, L. B., Schreuder, R., Pastizzo, M., & Baayen, R. H. (2002). The processing and representation of Dutch and English compounds: Peripheral morphological and central orthographic effects. *Brain and Language*, 81, 555–567. doi:10.1006/brln.2001.2547
- Faaß, G., & Eckart, K. (2013). SdeWaC—A corpus of parsable sentences from the web. In *Proceedings of the International Conference of the German Society for Computational Linguistics and Language Technology* (pp. 61–68). Darmstadt, Germany: Gesellschaft für Sprachtechnologie und Computerlinguistik.
- Fellbaum, C. (1995). Co-occurrence and antonymy. *Lexicography*, 8, 281–303.
- Fellbaum, C. (1998). *WordNet—An electronic lexical database*. Cambridge, MA: MIT Press.
- Fellbaum, C., & Chaffin, R. (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* (pp. 420–427). Stroudsburg, PA: Association for Computational Linguistics.
- Fernandez, A., Diez, E., Alonso, M. A., & Beato, M. S. (2004). Free-association norms for the Spanish names of the Snodgrass and Vanderwart pictures. *Behavior Research Methods, Instruments, & Computers*, 36, 577–584. doi:10.3758/BF03195604
- Ferrand, L., & Alario, F.-X. (1998). Normes d'associations verbales pour 366 noms d'objets concrets. *L'Année Psychologique*, 98, 659–709. doi:10.3406/psy.1998.28564
- Fleischer, W., & Barz, I. (2012). *Wortbildung der deutschen Gegenwartssprache* (4th ed.). Berlin, Germany: de Gruyter.
- French, C., & Richards, A. (1992). Word association norms for a set of threat/neutral homographs. *Cognition and Emotion*, 6, 65–87.
- Gagné, C., & Spalding, T. (2009). Constituent integration during the processing of compound words: Does it involve the use of relational structures? *Journal of Memory and Language*, 60, 20–35.
- Guida, A. (2007). *The representation of verb meaning within lexical semantic memory: Evidence from word associations* (Unpublished master's thesis). Università degli studi di Pisa, Pisa, Italy.
- Heringer, H. J. (1986). The verb and its semantic power: Association as the basis for valence. *Journal of Semantics*, 4, 79–99.
- Hirsh, K., & Tree, J. (2001). Word association norms for two cohorts of British adults. *Journal of Neurolinguistics*, 14, 1–44.
- Janssen, N., Bi, Y., & Caramazza, A. (2008). A tale of two frequencies: Determining the speed of lexical access for Mandarin Chinese and English compounds. *Language and Cognitive Processes*, 23, 1191–1223.
- Jauch, R. (2012). *Empirische Analysen von Assoziationen und distributionelle Modellierung der Kompositionalität von deutschen Nomen-Komposita* (Unpublished M. Sc. Thesis). Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.
- John, C. H. (1988). Emotionality ratings and free-association norms of 240 emotional and non-emotional words. *Cognition and Emotion*, 2, 49–70.
- Kiss, G., Armstrong, C., Milroy, R., & Piper, J. (1973). An associative thesaurus of English and its computer analysis. In A. Aitken, R. Bailey, & N. Hamilton-Smith (Eds.), *The computer and literary studies* (pp. 153–165). Edinburgh, UK: Edinburgh University Press.
- Klos, V. (2011). *Komposition und Kompositionalität*. Berlin, Germany: Walter de Gruyter.
- Kremer, G., & Baroni, M. (2011). A set of semantic norms for German and Italian. *Behavior Research Methods*, 43, 97–109. doi:10.3758/s13428-010-0028-x
- Krott, A., Schreuder, R., Baayen, R. H., & Dressler, W. (2007). Analogical effects on linking elements in German compound words. *Language and Cognitive Processes*, 22, 25–57.
- Kunze, C. (2000). Extension and use of GermaNet, a lexical-semantic database. In *Proceedings of the 2nd International Conference on Language Resources and Evaluation* (pp. 999–1002). Stroudsburg, PA: Association for Computational Linguistics.
- Lauteslager, M., Schaap, T., & Schievels, D. (1986). *Schriftelijke Woordassociatienormen voor 549 Nederlandse Zelfstandige Naamwoorden*. Lisse, The Netherlands: Swets & Zeitlinger.
- Libben, G. (1994). How is morphological decomposition achieved? *Language and Cognitive Processes*, 9, 369–391.
- Libben, G. (1998). Semantic transparency in the processing of compounds: Consequences for representation, processing, and impairment. *Brain and Language*, 61, 30–44.
- Libben, G., Gibson, M., Yoon, Y. B., & Sandra, D. (2003). Compound fracture: The role of semantic transparency and morphological headedness. *Brain and Language*, 84, 50–64. doi:10.1016/S0093-934X(02)00520-5
- Longtin, C.-M., Segui, J., & Hallé, P. (2003). Morphological priming without morphological relationship. *Language and Cognitive Processes*, 18, 313–334. doi:10.1080/01690960244000036
- Lüttmann, H., Zwitserlood, P., Böhl, A., & Bölte, J. (2011). Evidence for morphological composition at the form level in speech production. *Journal of Cognitive Psychology*, 23, 818–836.
- Macizo, P., Gómez-Ariza, C., & Bajo, M. (2000). Associative norms of 58 Spanish words for children from 8 to 13 years old. *Psicológica*, 21, 287–300.
- Marslen-Wilson, W., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review*, 101, 3–33. doi:10.1037/0033-295X.101.1.3
- McNamara, T. (2005). *Semantic priming: Perspectives from memory and word recognition*. New York, NY: Psychology Press.

- Melinger, A., Schulte im Walde, S., & Weber, A. (2006a). 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"* (pp. 41–48). Stroudsburg, PA: Association for Computational Linguistics.
- Melinger, A., & Weber, A. (2006). *Database of noun associations for German*. Retrieved from [www.coli.uni-saarland.de/projects/nag/](http://www.coli.uni-saarland.de/projects/nag/)
- Miller, G. (1969). The organization of lexical memory: Are word associations sufficient? In G. Talland & N. Waugh (Eds.), *The pathology of memory* (pp. 223–237). New York, NY: Academic Press.
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, 36, 402–407. doi:10.3758/BF03195588
- Palermo, D., & Jenkins, J. (1964). *Word association norms: Grade school through college*. Minneapolis, MN: University of Minnesota Press.
- Rapp, R. (2002). The computation of word associations: Comparing syntagmatic and paradigmatic approaches. In *Proceedings of the 19th International Conference on Computational Linguistics*. Stroudsburg, PA: Association for Computational Linguistics.
- Reddy, S., McCarthy, D., & Manandhar, S. (2011). An empirical study on compositionality in compound nouns. In *Proceedings of the 5th International Joint Conference on Natural Language Processing* (pp. 210–218). Chiang Mai, Thailand.
- Roller, S., Schulte im Walde, S., & Scheible, S. (2013). The (un)expected effects of applying standard cleansing models to human ratings on compositionality. In *Proceedings of the 9th Workshop on Multiword Expressions* (pp. 32–41). Stroudsburg, PA: Association for Computational Linguistics.
- Roth, M., & Schulte im Walde, S. (2008). Corpus co-occurrence, dictionary and wikipedia entries as resources for semantic relatedness information. In *Proceedings of the 6th International Conference on Language Resources and Evaluation* (pp. 1852–1859). European Language Resources Association (ELRA).
- Russell, W. (1970). The complete German language norms for responses to 100 words from the Kent-Rosanoff word association test. In L. Postman & G. Keppel (Eds.), *Norms of word association* (pp. 53–94). New York, NY: Academic Press.
- Russell, W., & Meseck, O. (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.
- Sandra, D. (1990). On the representation and processing of compound words: Automatic access to constituent morphemes does not occur. *Quarterly Journal of Experimental Psychology*, 42A, 529–567. doi:10.1080/14640749008401236
- Schulte im Walde, S. (2008). Human associations and the choice of features for semantic verb classification. *Research on Language and Computation*, 6, 79–111.
- Schulte im Walde, S., Borgwaldt, S., & Jauch, R. (2012). Association norms of German noun compounds. In *Proceedings of the 8th International Conference on Language Resources and Evaluation* (pp. 632–639). European Language Resources Association (ELRA).
- Schulte im Walde, S., & Melinger, A. (2008). An in-depth look into the co-occurrence distribution of semantic associates. *Italian Journal of Linguistics*, 20, 89–128.
- Schulte im Walde, S., Melinger, A., Roth, M., & Weber, A. (2008). An empirical characterisation of response types in German association norms. *Research on Language and Computation*, 6, 205–238.
- Schulte im Walde, S., & Müller, S. (2013). Using web corpora for the automatic acquisition of lexical-semantic knowledge. *Journal for Language Technology and Computational Linguistics*, 28, 85–105.
- Schulte im Walde, S., Müller, S., & Roller, S. (2013). Exploring vector space models to predict the compositionality of German noun–noun compounds. In: *Proceedings of the 2nd Joint Conference on Lexical and Computational Semantics* 255–265. Stroudsburg, PA: Association for Computational Linguistics.
- Spence, D. P., & Owens, K. C. (1990). Lexical co-occurrence and association strength. *Journal of Psycholinguistic Research*, 19, 317–330.
- Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *Quarterly Journal of Experimental Psychology*, 57A, 745–765.
- Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, 14, 638–647. doi:10.1016/S0022-5371(75)80051-X
- van Jaarsveld, H., & Rattink, G. (1988). Frequency effects in the processing of lexicalized and novel nominal compounds. *Journal of Psycholinguistic Research*, 17, 447–473.
- von der Heide, C., & Borgwaldt, S. (2009). Assoziationen zu Unter-, Basis- und Oberbegriffen: Eine explorative Studie. In R. Vogel & S. Sahel (Eds.), *Proceedings of the 9th Norddeutsches Linguistisches Kolloquium* (pp. 51–74). Bielefeld, Germany: Universität Bielefeld.
- Zwitserslood, P. (1994). The role of semantic transparency in the processing and representation of Dutch compounds. *Language and Cognitive Processes*, 9, 341–368. doi:10.1080/01690969408402123