

German norms for semantic typicality, age of acquisition, and concept familiarity

Astrid Schröder · Teresa Gemballa · Steffie Ruppin ·
Isabell Wartenburger

Published online: 22 December 2011
© Psychonomic Society, Inc. 2011

Abstract The present study introduces the first substantial German database with norms for semantic typicality, age of acquisition, and concept familiarity for 824 exemplars of 11 semantic categories, including four natural (ANIMALS, BIRDS, FRUITS, and VEGETABLES) and five man-made (CLOTHING, FURNITURE, VEHICLES, TOOLS, and MUSICAL INSTRUMENTS) categories, as well as PROFESSIONS and SPORTS. Each category exemplar in the database was collected empirically in an exemplar generation study. For each category exemplar, norms for semantic typicality, estimated age of acquisition, and concept familiarity were gathered in three different rating studies. Reliability data and additional analyses on effects of semantic category and intercorrelations between age of acquisition, semantic typicality, concept familiarity, word length, and word frequency are provided. Overall, the data show high inter- and intrastudy reliabilities, providing a new resource tool for designing experiments with German word materials. The full database is available in the supplementary material of this file and also at www.psychonomic.org/archive.

Keywords Semantic typicality · Age of acquisition · Concept familiarity · Norm data · Semantic categories · German database · Lexical–semantic processing · Exemplar generation

Electronic supplementary material The online version of this article (doi:10.3758/s13428-011-0164-y) contains supplementary material, which is available to authorized users.

A. Schröder (✉) · T. Gemballa · S. Ruppin · I. Wartenburger
Center of Excellence Cognitive Sciences Neurocognition
of Language/Neurolinguistics,
University of Potsdam,
Karl-Liebknecht-Str. 24-25,
14476 Potsdam, Germany
e-mail: astrid.schroeder@uni-potsdam.de

Introduction

Among the variables known to affect the speed and accuracy of lexical–semantic processing in both healthy and brain-damaged participants are *semantic typicality*, *age of acquisition*, and *concept familiarity*. In most studies, these variables have been assessed empirically by asking participants to estimate the age at which they learned a word (*age of acquisition*), how familiar a concept is in a person's individual experience (*concept familiarity*), or how well it represents a particular semantic category (*semantic typicality*). Other variables that affect word processing are intrinsic to each word and can be determined directly from its surface structure (e.g., their *word length*, in terms of number of syllables, phonemes, or letters). Furthermore, variables such as *word frequency* are determined by counting the frequency of occurrence of words in large language corpora (for German: CELEX—Baayen, Piepenbrock, & van Rijn, 1993; or the dlexDB database—Heister, Würzner, Bubenzer, Pohl, Hanneforth, Geyken, & Kliegl, 2011).

The use of category norms for selecting stimuli for experimental investigations has had a long tradition in cognitive psychology. In her pioneering work on the internal structure of semantic categories, Rosch (1975) used examples from different semantic categories that had been directly generated by speakers in a norming study by Battig and Montague (1969; see also Van Overschelde, Rawson, & Dunlosky, 2004, for an updated and extended version of these norms). For these empirically generated category exemplars, Rosch collected norms for estimated within-category typicality using a 7-point scale (see also Uyeda & Mandler, 1980, for an extension of Rosch's typicality norms). As such, *semantic typicality* reflects the degree to which a concept (e.g., *penguin*, *robin*) is

representative of a given category (e.g., BIRDS; Rosch & Mervis, 1975). It has been shown that typicality influences semantic-processing performance in online categorization or semantic decision tasks for both healthy (Holmes & Ellis, 2006; Morrison & Gibbons, 2006; Rips, Shoben, & Smith, 1973) and aphasic (Kiran, Ntourou, & Eubank, 2007; Kiran & Thompson, 2003a; Stanczak, Waters, & Caplan, 2006) processing. Semantic typicality has also been shown to influence processing speed in picture naming in healthy adults (Dell'Acqua, Lotto, & Job, 2000; Holmes & Ellis, 2006), as well as picture-naming accuracy in patients suffering from different neuropsychological disorders (Laiacona, Luzzatti, Zonca, Guarnaschelli, & Capitani, 2001; Woollams, Cooper-Pye, Hodges, & Patterson, 2008). Furthermore, within the framework of the “complexity account of treatment efficacy” (Thompson, 2007), it has been suggested that the treatment of aphasic word-finding difficulties is possibly more effective when targeting atypical items during treatment (Kiran & Thompson, 2003b). Variances in typicality are also reflected in differential neurophysiological responses, as atypical items have been shown to increase the N400 component in healthy participants (e.g., Heinze, Munte, & Kutas, 1998; Monetta, Tremblay, & Joannette, 2003; Núñez-Peña & Honrubia-Serrano, 2005; Stuss, Picton, & Cerri, 1988).

Age of acquisition refers to the age at which a word was learned. There are several assumptions as to why age of acquisition might affect word processing. One is that earlier-acquired concepts might build the basis for the acquisition of later concepts; hence, they might be more connected and/or more often used. The greater use of early-acquired concepts can also be described as a higher cumulative frequency of their associated words or by different frequency trajectories (Zevin & Seidenberg, 2002, 2004). Another assumption is that “different” or “better” learning mechanisms are available at early ages based on specific biological foundations such as brain plasticity (see Hernandez & Li, 2007, for a review). Although some studies have used objective measures of age of acquisition with data from children who were asked to name pictures (Álvarez & Cuetos, 2007; Morrison, Chappell, & Ellis, 1997; Pind, Jónsdóttir, Gossurardóttir, & Jónsson, 2000), most researchers have used subjective measures of estimated age of acquisition judged retrospectively by adult participants. In these estimates, age of acquisition has usually been rated on a 7-point scale (after Gilhooly & Logie, 1980). Since the original studies of Carroll and White (1973a) and Gilhooly and Logie (1980), norm data of estimated age of acquisition have been collected for a number of different languages (e.g., Alario & Ferrand, 1999; Bonin, Peereman, Malardier, Méot, & Chalard, 2003; Cameirão & Vicente, 2010; Cortese &

Khanna, 2008; Dell'Acqua et al., 2000; Dimitropoulou, Duñabeitia, Blitsas, & Carreiras, 2009; Ghyselinck, De Moor, & Brysbaert, 2000; Izura, Hernández-Muñoz, & Ellis, 2005; Khanna & Cortese, 2011; Manoiloff, Artstein, Canavoso, Fernández, & Segui, 2010; Marques, Fonseca, Morais, & Pinto, 2007; Nishimoto, Miyawaki, Ueda, Une, & Takahashi, 2005; Pind et al., 2000; Ruts, De Deyne, Ameel, Vanpaemel, Verbeemen, & Storms, 2004; Sirois, Kremin, & Cohen, 2006; Tsaparina, Bonin, & Méot, 2011). Overall, those ratings seem to be consistent, as indexed by high intra- and intergroup reliability measures with high correlations between the rating scores within the group of participants and between the rating scores for the same words used in different studies with different participant populations. In addition, a number of studies have found high correlations between estimated and objective age of acquisition, measured as the age by which children can read words (Carroll & White, 1973a) or as the age by which children can name pictures (e.g., Morrison et al., 1997; Pind et al., 2000; Schröder, Kauschke, & De Bleser, 2004). However, it has also been shown that objective and estimated age of acquisition differ, in that the subjective measure based on adult estimates is more influenced by word frequency and the familiarity of concepts than by objective age-of-acquisition data (Morrison et al. 1997). Nevertheless, as objective and estimated age of acquisition values are highly correlated, adult estimates are regarded as being adequate measures of age of acquisition (Morrison et al., 1997).

In general, words acquired earlier in life are processed faster or more accurately than words acquired later in life in various language-processing tasks. Age of acquisition has been discussed as an important variable at the lexical processing level, where it affects the speed of processing in word recognition (Baumgaertner & Tompkins, 1998; Turner, Valentine, & Ellis, 1998) and picture-naming (e.g., Barry, Morrison, & Ellis, 1997; Carroll & White, 1973b; Chalard & Bonin, 2006; Cuetos, Ellis, & Alvarez, 1999; Hodgson & Ellis, 1998; Johnston & Barry, 2006; Morrison & Ellis, 1995) tasks. In support of semantic hypotheses (e.g., the “semantic locus” theory of Brysbaert, Van Wijnendaele, & De Deyne, 2000), age of acquisition also affects the semantic system, with faster responses for earlier- than for later-acquired words in various semantic tasks (e.g., Brysbaert et al., 2000; Cortese & Khanna, 2007; De Deyne & Storms, 2007; Ghyselinck, Custers, & Brysbaert, 2004; Morrison & Gibbons, 2006). However, some studies have not reported an influence of age of acquisition on semantic processing (e.g., Catling & Johnston, 2006; Morrison, Ellis, & Quinlan, 1992), especially when the items used were controlled for semantic typicality (Holmes & Ellis, 2006). Age of acquisition influences the speed and the

accuracy of picture naming in normal aging (Morrison, Hirsh, Chappell, & Ellis, 2002), as well as affecting word processing in patients suffering from different neuropsychological conditions, with words acquired early being better preserved than words acquired later (e.g., Cuetos, Herrera, & Ellis, 2010; De Bleser & Kauschke, 2003; Gerhand & Barry 2000; Lambon Ralph, Graham, Ellis, & Hodges, 1998; Nickels & Howard, 1995; see Ellis, *in press*, for an overview). Interestingly, the data from event-related potentials (ERP) and functional magnetic resonance imaging (fMRI) studies have suggested that words acquired early versus later may be represented differently in the brain (Cuetos, Barbón, Urrutia, & Domínguez, 2009; Fiebach, Friederici, Müller, von Cramon, & Hernandez, 2003).

The term *familiarity* has been used in the literature in the senses of both *lexical* familiarity with the word form (subjective frequency/subjective familiarity; Balota, Pilotti, & Cortese, 2001; Gernsbacher, 1984; Gilhooly & Logie, 1980; Stadthagen-Gonzalez & Davis, 2006) and familiarity with the *concept* of an object (e.g., Snodgrass & Vanderwart, 1980). Snodgrass and Vanderwart defined familiarity as “the degree to which you come in contact with or think about the concept” (p. 183). These or similar instructions, which explicitly cover not only the familiarity of the word form but also the usage of an item, have been used in several rating studies in which the familiarity of objects has been rated by participants after presentation of pictures (e.g., Alario & Ferrand, 1999; Bonin et al., 2003; Cuetos et al., 1999; Genzel, Kerkhoff, & Scheffter, 1995; Morrison et al., 1997; Snodgrass & Vanderwart, 1980) as well as of words (Izura et al., 2005). In contrast to word frequency and age of acquisition, it is less clear how concept familiarity and lexical retrieval in picture naming are related. An influence of familiarity on lexical retrieval in picture naming, with better processing of highly familiar words, has been found in some studies (e.g., Cuetos et al., 1999; Snodgrass & Yuditsky, 1996), but not in others (Bonin et al., 2003; Ellis & Morrison, 1998). In neuropsychological research, familiarity is regarded as a variable that influences semantic processing. In patients with acquired semantic-processing disorders, highly familiar items seem to be protected better against loss than are less familiar items, leading to better performance with highly familiar items in lexical retrieval and comprehension tasks (e.g., Funnell & De Mornay Davies, 1996; Hirsh & Funnell, 1995; Lambon Ralph et al., 1998; Woollams et al., 2008).

Within the study of so-called *category-specific* semantic disorders, it has been shown that some “category-specific” effects may arise due only to material–intrinsic differences in word frequency, concept familiarity, or age of acquisition (Cappa, Frugoni, Pasquali, Perani, & Zorat,

1998; Funnell & De Mornay Davies, 1996; Funnell & Sheridan, 1992; Stewart, Parkin, & Hunkin, 1992). In some studies, sets of items from animate categories (e.g., ANIMALS, FRUITS, VEGETABLES) were less frequent or were acquired earlier than items from inanimate categories (e.g., FURNITURE, TOOLS, MUSICAL INSTRUMENTS; Funnell & De Mornay Davies, 1996; Howard, Best, Bruce, & Gatehouse, 1995). In line with this, it has been shown in normative studies that items from different semantic categories may vary in their mean ratings, with ANIMALS rated as being acquired relatively early and highly typical, yet rated as relatively low in familiarity (Izura et al., 2005). Likewise, Snodgrass and Vanderwart (1980) found that the items from different semantic categories could be grouped by their significant differences in their mean familiarity ratings, with items from the category of ANIMALS, together with BIRDS and MUSICAL INSTRUMENTS, getting the lowest familiarity ratings, and items from the categories of FURNITURE and KITCHEN UTENSILS, together with BODY PARTS, gaining the highest familiarity ratings. Hence, controlling variables when designing experiments is particularly important when performance is assessed for different semantic categories.

During the last decade, several normative databases for age of acquisition, semantic typicality, and concept familiarity have been collected in different languages. Cross-linguistic comparisons have shown that one needs to be careful when using norms from one language in another, because culture-specific differences may arise not only with respect to name agreement, but also with regard to the conceptual familiarity of objects (Cuetos et al., 1999; Dell’Acqua et al., 2000; Sanfeliu & Fernandez, 1996). In addition, comparisons of correlations between variables have shown that although there is considerable overlap in the types of correlations, the magnitudes of these relations vary across studies. For this reason, it has been suggested that normative data should be collected for each language separately (Bonin et al., 2003).

In sum, language-specific norm data on typicality, age of acquisition, and familiarity are needed for selecting items in research on healthy and impaired language processing. In recent years, some extensive German databases of more than 2,000 words have been published for imageability, concreteness, emotional valence, and arousal (Lahl, Göritz, Pietrowsky, & Rosenberg, 2009; Vö, Conrad, Kuchinke, Urton, Hofmann, & Jacobs, 2009; Vö, Jacobs, & Conrad, 2006). By contrast, German databases with rather limited lists of items exist with norms for familiarity, visual complexity, and age of acquisition (i.e., norms for $N = 244$ – 255 items from Snodgrass & Vanderwart, 1980, are provided in Genzel et al., 1995, and Schröder et al., 2004). In addition, German norms for typicality are—to our knowl-

edge—not yet available. Therefore, the present study had two main objectives:

- First, to provide substantial German norm data for *semantic typicality*, *age of acquisition*, and *concept familiarity* for a large number of words from various semantic categories. Despite the existence of such databases in other languages, there is no such instrument in German.
- Second, to investigate the characteristics of the present database in terms of an analysis of its intra- and interstudy reliabilities, the degree of intercorrelations between variables, and differences in rating scores with regard to different semantic categories.

Method

Four different studies were conducted for developing the current database of German norms for the semantic typicality, age of acquisition, and concept familiarity of 824 exemplars of 11 semantic categories (ANIMALS, BIRDS, FRUITS, VEGETABLES, CLOTHING, FURNITURE, VEHICLES, TOOLS, MUSICAL INSTRUMENTS, PROFESSIONS, and SPORTS). First, all category exemplars were collected in an *exemplar generation study*. Subsequently, three different rating studies were conducted to gather German norm data for the *semantic typicality* (Rating Study 1), *age of acquisition* (Rating Study 2), and *concept familiarity* (Rating Study 3) of the collected items. All materials were presented in German. For the present purpose, the closest English equivalent was chosen to describe the data set.

Participants

Table 1 lists the overall characteristics of the 160 participants who took part in the exemplar generation study and the three rating studies. The specific characteristics of the participants in the different rating studies are listed in the separate subsections for the four studies. All participants gave signed consent for participation and were monolingual native speakers of German. Some of the participants were enrolled in university degree programs and received course credit for their participation. The participants took part in only one study; that is, there was no overlap of participants across the four different studies.

Selection of stimuli: Exemplar generation study

Participants, materials, and procedure A group of 20 participants (15 female, 5 male) took part in the exemplar generation study. Participants were provided with a booklet

containing a list of 11 category labels (VEGETABLES, VEHICLES, TOOLS, CLOTHING, FURNITURE, SPORTS, BIRDS, FRUITS, ANIMALS, PROFESSIONS, and MUSICAL INSTRUMENTS¹). Each category label was presented on a separate sheet of paper. Participants were asked by written instructions to write down as many examples as they could think of for each semantic category. No time limit was given to complete the task.

Data analysis All responses were considered for further analyses. For each item, its generation frequency (number of participants listing that item) was coded. Adaptation of the raw data was kept to a minimum in order to keep a wide range of category exemplars that should be rated for semantic typicality, age of acquisition, and familiarity in the rating studies. Items that were judged by two independent raters as not belonging to the depicted category, as well as homographs (e.g., *kiwi*: BIRD, FRUIT; *horn*: MUSICAL INSTRUMENT, PART OF AN ANIMAL) were eliminated. Singular and plural forms of the same lemma were merged, as well as synonyms. For the category of PROFESSIONS, each item was coded in its singular male word form. In the case of synonyms, the term generated by the majority of participants was selected. Items were regarded as synonyms (e.g., German: *Grapefruit*, *Pampelmuse*; English: *grapefruit*, *shaddock*) only if they were coded as such in a German online database of the University of Leipzig (Biemann, Bordag, Heyer, Quasthoff, & Wolff, 2004, <http://wortschatz.uni-leipzig.de>). All other items with minimal semantic differences (e.g., German: *Stöckelschuh*, *Pumps*; English: *stiletto*, *pumps*) remained in the set. Items listed for both the categories of ANIMALS and BIRDS (i.e., three items: *duck*, *parrot*, and *chicken*), as well as superordinates (e.g., *wildcat*, *cat of prey*) and subordinates (e.g., *kitchen table*, *dining table*) remained in the set and were rated for their within-category typicality in the typicality rating study (1,123 exemplars).

Rating Study 1: Semantic typicality

Participants, materials, and procedure A group of 20 participants (15 female, 5 male) took part in the semantic

¹ The categories selected for the exemplar generation study are categories commonly used in linguistic, cognitive, and neuropsychological research (Kiran & Thompson, 2003b; Rosch, 1975; Uyeda & Mandler, 1980). Unlike in Uyeda and Mandler's study, our category of ANIMALS was not restricted to four-footed animals. This was done to get a substantial list of a wide range of ANIMALS from different subcategories (e.g., insects, birds, reptiles, or four-footed animals). With the exception of BIRDS, only basic-level categories were chosen for the exemplar generation study. Nevertheless, the category of BIRDS was included because it is one of the most mentioned categories chosen by Rosch, and items from this category could serve as measures of interstudy reliability in the typicality rating study.

Table 1 Age and years of education of participants in the four studies

Study	Participants	Age	Years of Education		
	<i>N</i>	<i>M</i> (<i>SD</i>)	Range	<i>M</i> (<i>SD</i>)	Range
Exemplar generation study	20	38.5 (15.25)	20–70	12.00 (1.26)	10–13
Rating Study 1: Typicality	20	45.05 (17.42)	23–69	12.05 (1.05)	10–13
Rating Study 2: Age of acquisition	60	40.38 (17.02)	20–69	11.67 (1.46)	8–13
Rating Study 3: Familiarity	60	41.86 (16.22)	20–68	12.37 (1.06)	9–13

typicality rating study. A total of 1,123 exemplars of the 11 categories collected in the exemplar generation study were included in the typicality rating. Following Rosch (1975), items were presented block-wise within their corresponding categories. Two lists with different randomizations (appearance of categories and items within their categories) were presented. The participants were asked to rate the typicality of the category exemplar on a 7-point scale from 1 (*very good example of the category/typical*) to 7 (*bad example of the category/atypical*; see Appendix A and B for the specific instructions). In addition, participants could indicate if they did not know the item (*unfamiliar*) or if they thought that the item was not a member of the requested category (*not a category member*).

Data analysis Items that were judged either as being unfamiliar or as not being a category member by 25% (5/20) or more of the participants were removed from the item set ($n = 63$ items). In addition, items that showed a high variability in judgments, resulting in standard deviations greater than ± 2 , were also removed from the item set ($n = 264$ items). The final set of items consisted of 870 items that were included in the age-of-acquisition rating and familiarity rating studies.

Rating Study 2: age of acquisition

Participants, materials, and procedure A group of 60 participants (35 female, 25 male) participants took part in the age-of-acquisition rating study. The 870 words rated for typicality were divided into three lists of items ($n = 290$ items each). Items from the 11 semantic categories were equally distributed across the three lists. Items in each list (and within each category) did not differ in terms of typicality (t test for unrelated samples, all $ps > .1$). Following Gilhooly and Logie (1980), participants were asked to indicate on a 7-point scale when they thought they had learned the words. At the top of each page, the 7-point scale was explained, in which 1 = 0–2 years, 2 = 3–4 years, 3 = 5–6 years, 4 = 7–8 years, 5 = 9–10 years, 6 = 11–12 years, 7 = 13 years or older. An additional column (*item unknown*) was added (after Marques et al., 2007; see Appendix A and B for the specific instructions), and each of the three lists was rated by 20 new participants. Items

were presented in blocks² within their corresponding categories. Two lists with different randomizations (appearance of categories and items within their categories) were presented.

Data analysis Items that were judged as being unknown by 25% (5/20) or more of the participants were removed from the item set ($n = 22$ items). In addition, items that showed a high variability in judgments, resulting in standard deviations greater than ± 2 , were also removed from the item set ($n = 5$ items).

Rating Study 3: Concept familiarity

Participants, materials, and procedure A group of 60 participants (31 female, 29 male) took part in the familiarity rating study. The 870 words rated for typicality in Rating Study 1 were divided into three lists of items ($n = 290$ items each). Items from the 11 semantic categories were distributed equally across the three lists. Items in each list (and within each category) did not differ in terms of typicality (t test for unrelated samples, all $ps > .1$). Following Snodgrass and Vanderwart (1980), participants were asked to estimate the degree to which they thought about or came in contact with a concept, using a 5-point scale ranging from 1 (*very unfamiliar*) to 5 (*very familiar*). Care was taken to make sure that the estimate had been attributed to the concept itself and not the word (see Appendix A and B for specific instructions). Each of the three lists was rated by 20 new participants. Items were presented together with its category label. Two lists with different randomizations of items across categories (with no more than two items from the same semantic category appearing subsequently) were presented.

² Note that the block-wise presentation for ratings of semantic typicality and age of acquisition might result in an overall activation of the category, obscuring the results for individual items. However, this procedure was adapted from Rosch (1975) and intended to make the (retrospective) decisions easier without introducing heavy category-switch costs. Both intra- and cross-study correlations (see below) indicate that the mode of presentation might not have a strong influence on the estimates.

Data analysis Items that were judged as unfamiliar by 25% (5/20) or more of the participants were removed from the item set ($n = 8$ items). None of the items remaining in the item set showed high variability in judgments that resulted in standard deviations greater than ± 2 .

Characteristics of the final database

Finally, the data from the exemplar generation study and the three rating studies were subsumed into a single database. The database consisted of 824 German nouns that were exemplars from 11 semantic categories (ANIMALS, BIRDS, FRUITS, VEGETABLES, CLOTHING, FURNITURE, VEHICLES, TOOLS, MUSICAL INSTRUMENTS, PROFESSIONS, and SPORTS). Each semantic category included between 40 and 193 exemplars that were generated in the exemplar generation study and rated for *semantic typicality*, *age of acquisition*, and *concept familiarity* by 20 different participants. For each category exemplar, its exemplar generation frequency (number of participants listing that item in the exemplar generation study) is provided in the database. In addition, norms for semantic typicality, age of acquisition, and familiarity are provided. Furthermore, for each word in the database, measures of *word length* (number of phonemes, number of syllables) and *word frequency* (normalized lemma frequency per million and logarithmic normalized lemma frequency) are given. All frequency values given in the database were taken from the German dLexDB database (www.dlexdb.de; Heister et al., 2011), which is based on the reference corpus of the German language compiled by the Digital Dictionary of the German Language (DWDS) with a size of about 100 million words (tokens) and 2.3 million distinct words (types).³ The full database can be downloaded from www.springerlink.com.

Results

Reliability

The intrastudy reliability of the data was tested by computing split-half correlations of the mean rating values for two different lists of randomized items. The results showed high intrastudy reliabilities for all three rating studies, with strong correlations between the mean rating values of the two lists of randomized items (typicality

³ dLexDB was considered to be more useful for the present study because it entails many more entries relevant to the present study than does the CELEX database (Baayen et al., 1993). For the present database, only 299 entries in the CELEX database would have been available, whereas the analysis with frequency values taken from dLexDB could be done on a subset of 747 items.

rating, $r = .87$; age-of-acquisition rating, $r = .92$; familiarity rating, $r = .79$). Interstudy reliability was examined by carrying out cross-study correlations on the variables in common on subsets of identical items included in other, comparable databases. A database was included if there was an overlap of about 100 or more items in both databases.⁴ For the measures of semantic typicality, items were only included if they were estimated in relation to the same category in the comparable studies.⁵ Table 2 depicts the results for the cross-study correlations on subsets of identical items—precisely, 2 other German studies and 15 other studies carried out in 10 different languages. Overall, there were highly significant correlations across studies for all three ratings carried out in the present study. The strongest correlations were found for measures of age of acquisition and conceptual familiarity carried out in 2 other German studies by Schröder et al. (2004) and by Genzel et al. (1995). For ratings obtained in studies from other languages, moderate to strong correlations were shown for the estimates of semantic typicality, age of acquisition, and concept familiarity (see Table 2). Negative values were simply due to the fact that in some studies, rating scales opposite to the ones used in our studies were used, with high values on the rating scale representing lower values of the estimated variables, and vice versa.

Intercorrelations between variables

Table 3 provides the intercorrelations of the German norms collected for semantic typicality, age of acquisition, and concept familiarity in the present study. In addition, the correlations of these variables with measures of word frequency and word length are also reported.⁶

The correlational analyses showed that all variables were significantly correlated with each other (all $ps < .01$). The two measures of word length (i.e., number of phonemes and syllables) were strongly correlated. In addition, there

⁴ Four other databases were considered for analyses but were not included because of insufficient overlap of items: British English—Stadthagen-Gonzalez and Davis (2006), $n = 59$ identical items; French—Bonin et al. (2003), $n = 34$ items; Italian—Barca, Burani, and Arduino (2002), $n = 76$ items; Portuguese—Cameirão and Vicente (2010), $n = 49$ items.

⁵ For example, *knife* was rated for semantic typicality within the category of TOOLS in our study, whereas in the study by Dell'Acqua et al. (2000), it was rated within the category of HOUSEHOLD ITEMS. As a result, *knife* was included in the correlational analyses for age of acquisition and familiarity, but not for semantic typicality.

⁶ For word frequency, all values were taken from the German dLexDB database (www.dlexdb.de; Heister et al., 2011). For all analyses, normalized logarithmic lemma frequencies (\log_{10}) were used. We did not compute correlations with German norms for imageability (Võ et al., 2006) and concreteness (Lahl et al., 2009) because of insufficient overlap of identical items in the databases ($n = 130$ items, Lahl et al., 2009; $n = 155$ items, Võ et al., 2006).

Table 2 Correlations of the variables in common between the ratings of the present study and other studies (Pearson's r)

Language	Study	TYP	AOA	FAM
German	Schröder et al. (2004)	–	.93 ⁽¹⁰⁶⁾	–
German	Genzel et al. (1995)	–	–	.85 ⁽¹⁰⁷⁾
American	Cortese & Khanna (2008)	–	.51 ⁽¹⁴⁰⁾	–
American	Rosch (1975)	.60 ⁽²²⁰⁾	–	–
American	Snodgrass & Vanderwart (1980)	–	–	.66 ⁽¹⁰⁸⁾
American	Uyeda & Mandler (1980)	.55 ⁽¹⁷²⁾	–	–
British English	Morrison et al. (1997)	–	.65 ⁽¹¹¹⁾	.72 ⁽¹¹¹⁾
Dutch/Flemish	Ruts et al. (2004)	–.65 ⁽¹⁶⁸⁾	.81 ⁽²²¹⁾	–
French	Alario & Ferrand (1999)	–	.69 ⁽¹⁵⁰⁾	.73 ⁽¹⁵⁰⁾
Greek	Dimitropoulou et al. (2009)	–	.67 ⁽¹¹⁴⁾	–
Icelandic	Pind et al. (2000)	–	.55 ⁽¹¹⁹⁾	.64 ⁽¹¹⁹⁾
Italian	Dell'Acqua et al. (2000)	–.74 ⁽⁹³⁾	.68 ⁽¹²⁵⁾	–.71 ⁽¹²⁵⁾
Japanese	Nishimoto et al. (2005)	–	.51 ⁽¹⁴⁶⁾	.66 ⁽¹⁴⁶⁾
Spanish	Sanfeliu & Fernandez (1996)	–	–	.59 ⁽¹⁰⁷⁾
Spanish	Izura et al. (2005)	–.63 ⁽¹¹⁹⁾	.65 ⁽¹¹⁹⁾	.58 ⁽¹¹⁹⁾
Spanish (Argentinean)	Manoiloff et al. (2010)	–	.61 ⁽¹⁵¹⁾	.67 ⁽¹⁵¹⁾
Russian	Tsaparina et al. (2011)	–	.69 ⁽¹⁰⁸⁾	.81 ⁽¹⁰⁸⁾

TYP = typicality, AOA = age of acquisition, FAM = familiarity. All correlations significant at the $p < .01$ level (two-tailed). The numbers of identical items used for the analyses are given in parentheses. Dashes indicate that data were not available

were moderate correlations between age of acquisition, familiarity, typicality, and word frequency, showing that the words in the database that were acquired early also tended to be more familiar, more typical, and more frequent than words acquired later. Furthermore, word length (e.g., in syllables) was moderately correlated with age of acquisition and word frequency, as well as being weakly but still significantly correlated with semantic typicality and concept familiarity (see Table 3).

Effects of semantic category

To analyze an effect of semantic category, one-way analyses of variance were used to compare the mean ratings of each item on each of the variables (semantic typicality, age of acquisition, and concept familiarity). Table 4 provides a description of the mean rating values for the 11 semantic categories in each of the rating studies.

Rating Study 1: Semantic typicality The analysis revealed a main effect of semantic category on the typicality ratings, $F(10, 813) = 7.33, p < .001, \eta^2 = .08$. Pairwise comparisons between the mean rating values in the different semantic categories showed that items in the categories of VEGETABLES, FRUITS, and BIRDS were rated as more typical than were items from the categories of CLOTHES, SPORTS, and VEHICLES, which were rated as more atypical (Tukey's HSD, all $ps < .05$).

Rating Study 2: Age of acquisition The analysis revealed a main effect of semantic category on the age-of-acquisition ratings, $F(10, 813) = 21.97, p < .001, \eta^2 = .21$. Pairwise comparisons between the mean rating values in the different semantic categories (see Table 4) showed that items in the category of ANIMALS were rated as being acquired earlier than were items in any of the other categories (Tukey's HSD, all $ps < .01$), except for FURNITURE and VEHICLES. By

Table 3 Correlations among semantic typicality, age of acquisition, concept familiarity, word frequency, and word length ($N = 824$ items)

	TYP	AOA	FAM	FREQ (log)	PHON	SYL
TYP	1					
AOA	.50	1				
FAM	–.59	–.58	1			
FREQ (log)	–.42	–.57	.44	1		
PHON	.24	.47	–.25	–.51	1	
SYL	.18	.48	–.20	–.44	.83	1

TYP = typicality, AOA = age of acquisition, FAM = familiarity, FREQ (log) = logarithmic word frequency, PHON = word length in number of phonemes, SYL = word length in number of syllables. All correlations significant at $p < .01$ (two-tailed)

Table 4 German database with norms for semantic typicality, age of acquisition, and concept familiarity: Mean ratings of items in each category

Semantic Category	Items ($N=824$)	TYP (scale 1–7)		AOA (scale 1–7)		FAM (scale 1–5)	
		M	SD	M	SD	M	SD
Animals	109	2.72	1.08	3.24	1.17	2.96	0.67
Birds	68	2.43	0.66	3.98	1.12	2.73	0.55
Vegetables	51	2.51	0.92	4.31	1.44	3.67	0.57
Fruits	42	2.42	1.08	4.13	1.51	3.70	0.60
Clothing	77	3.28	1.33	3.98	1.40	3.25	0.86
Vehicles	44	3.51	1.70	3.59	1.28	3.10	0.77
Furniture	51	2.99	1.38	3.57	1.21	3.58	0.87
Musical instruments	40	2.99	1.13	4.47	1.04	2.52	0.60
Tools	51	3.16	1.11	4.63	1.15	2.90	0.77
Sports	98	3.38	1.23	4.88	1.16	2.60	0.59
Professions	19	3.03	0.83	5.09	1.37	2.94	0.61

TYP = typicality, AOA = age of acquisition, FAM = familiarity

contrast, words in the category of PROFESSIONS were rated as being acquired later than were words in any of the other categories (Tukey's HSD, all $ps < .01$), except for MUSICAL INSTRUMENTS, TOOLS, and SPORTS.

Rating Study 3: Concept familiarity The analysis of variance revealed a main effect of semantic category on the familiarity ratings, $F(10, 813) = 21.12$, $p < .001$, $\eta^2 = .21$. Pairwise comparisons between the mean rating values in the different semantic categories showed that items in the categories of VEGETABLES, FRUITS, and FURNITURE had relatively high familiarity ratings and were rated as more familiar than were items in any of the other categories (Tukey's HSD, all $ps < .01$). By contrast, items in the category of MUSICAL INSTRUMENTS were rated as less familiar than were items in any of the other categories (Tukey's HSD, all $ps < .01$) except for BIRDS and SPORTS.

Discussion

The present study had two aims: (a) to provide a large German database containing norms for *semantic typicality*, *age of acquisition*, and *concept familiarity* for German nouns from numerous semantic categories, and (b) to provide a descriptive analysis of the database that included an examination of the intra- and interstudy reliabilities, an analysis of intercorrelations between the estimated variables, word frequency, and word length, and the distributions of the mean rating scores for different semantic categories.

To meet the first goal, of establishing a large German database with norms for semantic typicality, age of acquisition, and concept familiarity, we collected a large list of exemplars of 11 semantic categories that were directly generated by native speakers of German in an

exemplar generation study. For each of these 824 category exemplars, norms for semantic typicality, age of acquisition, and concept familiarity were gathered. In addition, values of word frequency taken from the German lexical database dlexDB (Heister et al., 2011) and measures of word length (number of phonemes, number of syllables) were included in the database.

Second, we characterized our database by providing analyses of inter- and intrastudy reliabilities. To obtain a measure of intrastudy reliability for each rating study, split-half Pearson's r s were computed for the participants rating the two lists of randomized items. Overall, the data showed high intrastudy reliability, with scores of $r = .87$ for the typicality rating, $r = .79$ for the familiarity rating, and $r = .92$ for the age-of-acquisition rating. The reliability scores obtained in the present study for the familiarity and typicality ratings are somewhat lower than those reported previously (e.g., for typicality, $r = .90$ or higher in Ruts et al., 2004, and Rosch, 1975; for familiarity, $r = .92$, Izura et al., 2005), whereas the intrastudy correlations obtained for the age-of-acquisition ratings were similar to those reported in the literature (e.g., $r = .98$, Gilhooly & Logie, 1980; $r = .88$, Izura et al., 2005). It is difficult to interpret these findings, as a number of studies have not reported split-half reliabilities at all (e.g., Dell'Acqua et al., 2000; Sirios et al., 2006; Snodgrass & Vanderwart, 1980). Overall, all three rating studies reached high interrater reliabilities, as all correlations were quite strong and reached statistical significance (all $ps < .01$).

The analysis of interstudy reliability revealed moderate to strong correlations between the data from this study and studies carried out in American English (Cortese & Khanna, 2008; Rosch, 1975; Snodgrass & Vanderwart, 1980; Uyeda & Mandler, 1980), British English (Morrison et al., 1997), Dutch (Ruts et al., 2004), French (Alario & Ferrand, 1999), German (Genzel

et al. 1995; Schröder et al., 2004), Greek (Dimitropoulou et al., 2009), Icelandic (Pind et al., 2000), Italian (Dell'Acqua et al., 2000), Japanese (Nishimoto et al., 2005), Spanish (Izura et al., 2005; Manoiloff et al., 2010; Sanfeliu & Fernandez, 1996), and Russian (Tsaparina et al., 2011), showing further evidence for the reliability of the obtained data.

For semantic typicality, no cross-linguistic comparisons between different studies are reported in the literature. We conducted five cross-study correlations with the typicality ratings for overlapping, identical items used in our study and in the studies of Rosch (1975), Uyeda and Mandler (1980), Ruts et al. (2004), Izura et al. (2005), and Dell'Acqua et al. (2000) in order to validate the present database further. The results of these correlations showed moderate correlations of $r = .55$ (Uyeda & Mandler, 1980), $r = .60$ (Rosch, 1975), $r = .63$ (Izura et al., 2005), $r = -.65$ (Ruts et al., 2004), and $r = -.74$ (Dell'Acqua et al., 2000). Dell'Acqua et al. carried out their typicality rating on the basis of a set of line drawings, whereas the items used by Izura et al. were selected by their lexical availability (produced in a category-fluency task within 2 min; Izura et al., 2005, p. 387). The other three studies (Uyeda & Mandler, 1980; Rosch, 1975; Izura et al., 2005) used items for their typicality ratings, which were generated in exemplar generation studies within 30 s. It might be possible that the items used in those studies have a higher production frequency and higher semantic typicality than do the items used in our study (produced without time limit). To explore this issue further, we compared the set of $n = 220$ overlapping items in our study and the study by Rosch.⁷ Overall, we think that selected items from our study and the study by Rosch are quite comparable in terms of production frequency. In our study, we included every category exemplar, even if it was produced by only 1 of the participants (corresponding to a production frequency of 0.05%). Rosch (p.197) stated that she included all items that had been produced by 10 ($n = 2.3\%$ of the participants) or more subjects in the Battig and Montague (1969) study, as well as items that were produced “by fewer subjects in the Battig and Montague norms.” For example, for the category of FURNITURE, she included the items *magazine rack*, *closet*, and *fan*, which were only listed by 3 (0.7%), 2 (0.45%), and 1 (0.23%) participant(s) in the Battig and Montague study. Thus, in both our and Rosch's studies some of the included items had relatively high (e.g., *bed*) and low (e.g., *newspaper/magazine rack*) production frequencies. In addition, the two sets did not differ in terms of their distributions of rated typicality (for our study, $M = 2.42$, $SD = 1.09$, range 1.00–5.74, $Mdn = 2.25$; for Rosch,

$M = 2.36$, $SD = 1.01$, range = 1.02–5.90, $Mdn = 2.25$), which makes it unlikely that the moderate correlation of $r = .60$ between our and Rosch's data was due to a general difference in the mean typicality of the items included in the analysis.

In sum, we cannot rule out the possibility that the differences in study design (i.e., the time limit) could contribute to the moderate correlations of the overlapping items. However, as the reported correlations are comparable for all studies, despite the use of different study designs, we think that the differences shown in the magnitudes of the correlation coefficients were also influenced by cultural and/or linguistic differences across the different studies. That is, some items that are more typical representatives of a given category in one culture may not be as representative in another culture (e.g., *potato* is rated as being a relatively typical representative of the category of VEGETABLES in German, whereas *papaya* seems to be a relatively typical representative of the category of FRUITS for most of the American raters⁸).

For *age of acquisition* and *concept familiarity*, the comparison of the variables in common on subsets of identical items across studies showed that the highest correlations were obtained for the measures of age of acquisition and concept familiarity in the present study and in two other studies conducted in German (age of acquisition, $r = .93$; concept familiarity, $r = .85$). For all of the studies carried out in other languages, moderate to strong correlations were obtained (age of acquisition, all r s between .51 and .81; concept familiarity, all r s between .58 and .81; see Table 2). These results mirror those from other studies in which comparable analyses have been performed (e.g., Alario & Ferrand, 1999; Dell'Acqua et al., 2000; Nishimoto et al., 2005; Sanfeliu & Fernandez, 1996; Tsaparina et al., 2011). Whereas most of the ratings on concept familiarity reported in the literature are carried out with pictures as input stimuli, our study was carried out with words (note that it would have been very difficult or impossible to find nonambiguous pictures for some of the words of the exemplar generation study, especially for those that were not very typical in a given semantic category or that belonged to the categories PROFESSIONS or SPORTS). Furthermore, the correlational analyses indicated that the differences in magnitudes of the correlations obtained in the present study are not attributable to differences in the input modalities. In fact, the highest cross-study correlations of concept familiarity were

⁷ The study by Rosch (1975) was selected for this comparison because of the high number of overlapping items (see Table 2).

⁸ A closer look at the items rated for semantic typicality by the American (Rosch, 1975) and German (this study) participants revealed several item-specific differences in the semantic typicality ratings (e.g., mean typicality ratings for *potato*: German, 1.37; American, 2.89; for *papaya*: German, 4.15; American, 2.58; for *bowling*: German, 4.00; American, 2.18).

shown with two other studies carried out with pictures ($r = .85$, Genzel et al., 1995; $r = .81$, Tsaparina et al., 2011). At the same time, moderate correlations were found with overlapping items from two Spanish studies, both carried out with pictures ($r = .59$, Sanfeliu & Fernandez, 1996) or words ($r = .58$, Izura et al., 2005). These results suggest that the input modality does not seem to influence the familiarity ratings and that the participants of our study were rating the object concept and not the specific word form of each exemplar given.

Overall, the pattern of significant cross-study correlations obtained in the present study indicates that the items in our database share most aspects of age of acquisition, semantic typicality, and concept familiarity across different cultures and languages. Objects that are common in one culture may not be as common in another culture (Sanfeliu & Fernandez, 1996), and words that are acquired early in one language may be acquired later in another language, especially when they differ in morphological complexity, word frequency, or word length. In sum, the analyses of intra- and interstudy correlations have provided further evidence for the reliability of our data. Importantly, as several differences across different languages occurred, ratings of semantic typicality, age of acquisition, and concept familiarity should be carried out for each language separately.

The results of the intercorrelational analyses of the variables investigated in the present study showed that semantic typicality, age of acquisition, and concept familiarity, as well as word frequency, were moderately correlated with each other. It is assumed that this finding reflects a natural correlation of these variables (e.g., typical exemplars of a semantic category tend to be more familiar and more frequent in adult language and to be acquired earlier during childhood than atypical exemplars). By nature, age of acquisition and frequency are correlated, as highly frequent words are learned earlier in life and are more central (De Deyne & Storms, 2008; Morrison et al., 1997). Especially when age of acquisition is estimated retrospectively by adults, the correlation of estimated age of acquisition with word frequency and concept familiarity is high (e.g., Morrison et al., 1997). However, frequency and age of acquisition are not interchangeable with each other, as age of acquisition still significantly accounts for variance in performance, even when other variables such as frequency are controlled for (for recent data, see, e.g., Brysbaert & Cortese, 2010). Note that the influence of age of acquisition (and its relation to other variables, such as imageability) depends highly on the task applied (e.g., written naming vs. lexical decision; Cortese & Khanna, 2007) or the “transparency” or “regularity” of the input and output variables, as suggested by computational models (Zevin & Seidenberg, 2002). In sum, our findings

provide further evidence that the degree of natural correlations needs to be taken into account when interpreting any effects seen in experimental investigations of language processing, and further research needs to investigate the independent contributions of each of these variables on different aspects of language processing in various experimental tasks.

Whereas the present study has been conducted with a group of participants of a relatively wide age range (20–70 years), most of the norming studies mentioned above were conducted with young participants, mainly college students. However, Hodgson and Ellis (1998) suggested that the age of acquisition and familiarity of certain objects can be different for participants of different age groups, because they may have encountered the objects at different stages in their lives. In line with this theory, Sirois et al. (2006) found age-related differences in familiarity ratings, with older participants (60–85 years) judging the familiarity of 388 pictured objects as being more familiar than did young (18–39 years) or middle-aged (40–59 years) participants. In the same study, young participants estimated that they had learned the words corresponding to those pictures earlier than did the middle-aged and older participants. De Deyne and Storms (2007) suggested that differences in age-of-acquisition ratings may occur especially for words introduced in recent decades (e.g., EXOTIC FRUITS, such as *mango*). In line with this, the age-of-acquisition ratings of elderly raters (61–85 years) for some specific words (e.g., *robot*, *television*, *lime*) differed essentially from those of younger participants in a study by Cuetos, Samartino and Ellis (2011). To investigate whether there was a tendency for any age effects in our data, we conducted several separate analyses for the groups of younger (20–40 years) and elderly participants (41–70 years, $n = 10$ participants each). The results of the correlational analyses showed that there was a high overlap between the ratings of both age groups ($r = .82$ for the typicality rating, $r = .85$ for the age-of-acquisition rating, and $r = .77$ for the familiarity rating). In an analysis of variance, we found no main effects of age for either the typicality [$F(1, 18) = 0.00$, $p = .986$, $\eta^2 = .00$] or the age-of-acquisition [$F(1, 58) = 2.75$, $p = .103$, $\eta^2 = .02$] rating. However, for the familiarity rating, a significant effect of age [$F(1, 58) = 5.12$, $p = .027$, $\eta^2 = .04$]⁹ was found. Given our relatively small sample size, we are careful about interpreting these results. Still, these data indicate that further studies should explore possible age differences to avoid over- or underestimating any effects shown.

⁹ For the age-of-acquisition and familiarity ratings, the pool of items was divided into three lists, so that each item was rated by 20 participants; thus, each participant rated only 33% of the whole item set. For this reason, the analysis across participants was done with the mean rating values of the 11 semantic categories.

The findings of the present study add to recent research showing that ratings of age of acquisition, semantic typicality, and concept familiarity may differ for items from different semantic categories (Izura et al., 2005; Snodgrass & Vanderwart, 1980).¹⁰ As in the study by Izura et al., items in our database in the category of ANIMALS were estimated as being highly typical and being acquired earlier than items in most of the other categories. Similarly, we were able to replicate some findings of other studies, with items in the category of FURNITURE being rated as highly familiar (Izura et al., 2005; Snodgrass & Vanderwart, 1980) and items in the categories of BIRDS and MUSICAL INSTRUMENTS being rated as relatively low in concept familiarity (Snodgrass & Vanderwart, 1980). It is possible that this result reflects only random differences in the distributions of category exemplars in our database, with some categories being more widely dispersed (note that all items generated in the exemplar generation study were included in the database, even if they were generated only by 1 or 2 participants). However, these replicated findings suggest that differences in semantic typicality, age of acquisition, and concept familiarity may be inherent to certain semantic categories. It is therefore important to control for age of acquisition, semantic typicality, and concept familiarity when designing experiments for research on semantic processing. Future studies should explore these effects in greater detail and in various populations, such as different age groups or clinical populations.

Conclusion

The present study provides the first substantial German database of 824 nouns from 11 semantic categories, with norms for semantic typicality, age of acquisition, concept familiarity, word frequency, and word length that can be used by researchers from different scientific fields. Overall, high inter- and intrastudy reliabilities were shown. The results revealed that items in different semantic categories might vary with regard to semantic typicality, age of acquisition, and concept familiarity, indicating that it is important to control for these variables when designing experiments for psycho- or neurolinguistic research. In sum, the present database increases the pool of available German norm data and will serve as an important tool for selecting stimuli in the research of healthy lexical–semantic processing and in the assessment and rehabilitation of patients suffering from lexical–semantic impairments.

¹⁰ To our knowledge, no other studies have included an analysis of the influence of semantic category on the respective ratings.

Author note We thank Romy Råling for her help in establishing the database. Astrid Schröder gained support from the sponsorship for young academics/postdoc researchers from the Faculty of Human Sciences, Potsdam University. The research of Isabell Wartenburger is supported by the Stifterverband für die Deutsche Wissenschaft (Claussen-Simon-Stiftung).

Appendix A

Instructions for the typicality rating

In accordance with Rosch (1975), the specific instructions were as follows (translated):

This study addresses the question of what we have in mind when we use words that refer to different categories. Let's take the category of FLOWERS as an example. If you were asked to name a typical flower, you would probably say *rose* or *tulip*, but probably not *thistle*. The same is true for other kinds of categories, such as DOGS. A *German Shepherd* would probably be named more often as a typical dog than a *Pekinese*. Why is this the case? A *German Shepherd* is obviously a very typical dog, whereas a *Pekinese* is a less typical kind of a dog. Please note that this kind of judgment has nothing to do with our own preferences—for example, whether you like *Pekingese* more than *German Shepherds*. In this questionnaire, you will be asked to judge how well an example of a category represents its whole category. On the top of the page, you will find the name of the category. For each category, different members will be listed below the category name. For each member of the category, you should judge on a 7-point scale how well it represents the category. A “1” means that this is a very good example of the category. A “4” means that this is a possible example of a category. A “7” means that this is a bad example of the category. In the table below, you will see possible estimations of typicality for some members of the category of TOYS. For example, one member of the category TOYS is *puppet*. If you feel that *puppet* is a very good example of the category of TOYS, you would put a “1” for *puppet*. By contrast, if you feel that *weapon* is a bad example of the category of TOYS, you would put “7.” A *jumprope* would possibly get a “4,” as it is a possible example of the category of TOYS, not too good and not too bad. Use the other numbers of the 7-point scale to indicate intermediate judgments. Don't worry about why you feel that a thing is or isn't a good example of the category. And don't worry about whether it's just you or people in general who feel that way. Don't think too long about an answer, just mark it the way you see

it. If you feel that the listed item is not a category member at all, or if you don't know that item, please choose "no member" or "unknown."

Instructions for the age-of-acquisition rating

In accordance with Schröder et al. (2004), the specific instructions for the age-of-acquisition rating were as follows (translated):

Please try to estimate as precisely as possible the age at which you think you learned the following words together with their meanings, and indicate below when you think you first produced the word. It is not important whether you were able to use the word entirely correctly or not by then. Please try to use the whole scale from 1 to 7, with 1 = 0–2 years, 2 = 3–4 years, 3 = 5–6 years, 4 = 7–8 years, 5 = 9–10 years, 6 = 11–12 years, and 7 = 13 years or older. If you don't know an item, please mark "word unknown."

Instructions for the familiarity rating

Following Snodgrass and Vanderwart (1980), the specific instructions for the familiarity rating were as follows (translated):

In the following, you will see a list of different items (objects, sports, professions). Please try to judge the "familiarity" of each item—that is, the degree to which you come in contact with or think about the concept in your daily life. Please use the following 5-point scale in order to judge how familiar or common—or unfamiliar or uncommon—an object, kind of sport, or profession is for you. Please note that you should judge the familiarity of each item individually. Please also note that you don't judge whether you have yet encountered a specific word, but the familiarity of the concept itself. Don't think too long about an answer, but try to judge as quickly as possible whether the mentioned objects, sports, or professions are more or less familiar to you. Please try to use the whole scale from 1 to 5, where 1 = *not familiar*, 2 = *a little familiar*, 3 = *moderately familiar*, 4 = *very familiar*, and 5 = *highly familiar*.

Appendix B

Original instructions for the typicality rating (German)

Diese Studie beschäftigt sich mit der Frage wie typisch ein Wort für eine Kategorie nach unseren Vorstellungen ist.

Nehmen wir zur Verdeutlichung ein Beispiel aus der Kategorie BLUMEN. Auf die Frage nach einer typischen BLUME würde wahrscheinlich automatisch als erstes die *Rose*, *Tulpe* oder *Nelke* fallen, aber kaum jemand würde als erstes die *Distel* nennen. Ähnliches zeigt sich bei der Kategorie HUNDE. Ein *Schäferhund* würde öfter aufgezählt werden als ein *Pekinese*. Warum ist das so? Ein *Schäferhund* ist scheinbar nach unseren Vorstellungen ein sehr typischer Vertreter der Kategorie Hunde, ein *Pekinese* eher nicht. Beachten Sie bitte, dass die Bewertung darüber, wie typisch etwas für eine Kategorie ist, nicht von unseren Vorlieben abhängen sollte; es spielt daher keine Rolle, ob man z.B. *Pekinesen* viel lieber mag als *Schäferhunde*. Auf diesem Fragebogen sollen Sie entscheiden, wie gut ein Vertreter einer Kategorie die gesamte Kategorie repräsentiert. Am oberen linken Seitenanfang befindet sich der Name der Kategorie. Darunter sind verschiedene Mitglieder dieser Kategorie aufgelistet. Jedes Mitglied soll mit Hilfe einer 7-Punkte -Skala bewertet werden. Bewertet wird, wie gut ein Mitglied die jeweilige Kategorie vertritt. Eine „1“ bedeutet „sehr guter Repräsentant der Kategorie.“ Eine „4“ bedeutet mittelguter Repräsentant dieser Kategorie, und eine „7“ bedeutet schlechter Repräsentant dieser Kategorie. In der untenstehenden Tabelle sind einige mögliche Einschätzungen für die Kategorie SPIELZEUG aufgelistet. Wenn Sie der Meinung sind, dass *Puppe* ein sehr guter Repräsentant der Kategorie SPIELZEUG ist, würde *Puppe* auf der Punkteskala den Wert 1 bekommen. Wenn Sie hingegen meinen, dass *Waffe* ein sehr schlechter Repräsentant für die Kategorie SPIELZEUG ist, würde es den Wert 7 erhalten. *Springseil* könnte den Punktwert 4 erhalten, da es für Sie eventuell ein mittelgutes Beispiel für die Kategorie SPIELZEUG darstellt, nicht sehr gut aber auch nicht schlecht. Die anderen Punktwerte auf der Skala können benutzt werden, um Zwischenstufen anzuzeigen. Machen Sie sich keine Gedanken darüber, warum Sie denken, dass etwas ein gutes bzw. kein gutes Beispiel für eine Kategorie ist. Und denken Sie nicht darüber nach, ob es Ihre persönliche Meinung ist, oder ob diese Meinung allgemein geteilt wird. Überlegen Sie nicht zu lange, und bewerten Sie jedes Mitglied einfach so, wie Sie es für richtig empfinden. Wenn Sie ein Wort nicht kennen, oder es für Sie kein Mitglied einer Kategorie ist, so geben Sie *kein Mitglied* bzw. *unbekannt* an.

Original instructions for the age-of-acquisition rating (German)

Versuchen Sie so genau wie möglich einzuschätzen, in welchem Alter Sie jedes der folgenden Wörter zusammen mit seiner Bedeutung gelernt und selbst erstmalig in der gesprochenen Form verwendet haben. Es ist dabei nicht von Bedeutung, ob das Wort vollkommen fehlerfrei verwendet wurde. Bitte kreuzen Sie das Kästchen an, das

Ihrer Meinung nach dem eigenen produktiven Erwerbssalter am besten entspricht. Versuchen Sie hierbei, jeweils die volle Skala von 1–7 zu benutzen: 1 = 0–2 Jahre, 2 = 3–4 Jahre, 3 = 5–6 Jahre, 4 = 7–8 Jahre, 5 = 9–10 Jahre, 6 = 11–12 Jahre, 7 = 13 Jahre. Wenn Sie ein Wort nicht kennen, kreuzen Sie bitte „Wort unbekannt“ an.

Original instructions for the familiarity rating (German)

Im Folgenden sehen Sie eine Liste mit verschiedenen Begriffen, die entweder Objekte oder Sportarten oder Berufe darstellen. Für diese Objekte, Sportarten und Berufe sollen Sie einschätzen, wie „vertraut“ diese Ihnen sind, d.h. wie häufig Sie damit in Ihrem täglichen Leben in Kontakt kommen oder darüber nachdenken. Um einzuschätzen, wie vertraut bzw. gewöhnlich oder nicht vertraut, d.h. ungewöhnlich ein bestimmtes Objekt, ein Beruf oder eine Sportart für Sie ist, verwenden Sie bitte die Skala von 1–5. Hierbei gelten folgende Zuordnungen: 1: *gar nicht vertraut*, 2: *wenig vertraut*, 3: *mittelmäßig vertraut*, 4: *sehr vertraut*, 5: *hoch vertraut*. Wenn Sie einen Begriff gar nicht kennen, so kreuzen Sie bitte „nicht bekannt“ an. Bitte beachten Sie, dass Sie die Vertrautheit individuell für sich selbst einschätzen. Bitte beachten Sie auch, dass Sie nicht einschätzen, ob Sie ein bestimmtes Wort bereits kennen, sondern wie vertraut diese in Bezug auf ihren täglichen Alltag für Sie sind. Bitte denken Sie nicht zu lange über eine Antwort nach, sondern entscheiden relativ zügig „nach Gefühl,“ ob Ihnen die Begriffe eher vertraut oder eher nicht vertraut sind. Versuchen Sie dabei, die gesamte Skala von 1–5 zu verwenden.

References

- Alario, F.-X., & Ferrand, L. (1999). A set of 400 pictures standardized for French: Norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition. *Behavior Research Methods, Instruments, & Computers*, *31*, 531–552. doi:10.3758/BF03200732
- Álvarez, B., & Cuetos, F. (2007). Objective age of acquisition norms for a set of 328 words in Spanish. *Behavior Research Methods*, *39*, 377–383. doi:10.3758/BF03193006
- Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database (CD-ROM)*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.
- Balota, D. A., Piloti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & Cognition*, *29*, 639–647. doi:10.3758/BF03200465
- Barca, L., Burani, C., & Arduino, L. S. (2002). Word naming times and psycholinguistic norms for Italian nouns. *Behavior Research Methods, Instruments, & Computers*, *34*, 424–434. doi:10.3758/BF03195471
- Barry, C., Morrison, C. M., & Ellis, A. W. (1997). Naming the Snodgrass and Vanderwart pictures: Effects of age of acquisition, frequency, and name agreement. *Quarterly Journal of Experimental Psychology*, *50A*, 560–585. doi:10.1080/783663595
- Battig, W. F., & Montague, W. E. (1969). Category norms of verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology*, *80*(3, Pt. 2), 1–46. doi:10.1037/h0027577
- Baumgaertner, A., & Tompkins, C. A. (1998). Beyond frequency: Predicting auditory word recognition in normal elderly adults. *Aphasiology*, *12*, 601–617. doi:10.1080/02687039808249561
- Biemann, C., Bordag, S., Heyer, G., Quasthoff, U., & Wolff, C. (2004). Language-independent methods for compiling monolingual lexical data. In A. Gelbukh (Ed.), *Computational linguistics and intelligent text processing* (pp. 217–228). Berlin: Springer.
- Bonin, P., Peerean, R., Malardier, N., Méot, A., & Chalard, M. (2003). A new set of 299 pictures for psycholinguistic studies: French norms for name agreement, image agreement, conceptual familiarity, visual complexity, image variability, age of acquisition, and naming latencies. *Behavior Research Methods, Instruments, & Computers*, *35*, 158–167. doi:10.3758/BF03195507
- Brybaert, M., & Cortese, M. J. (2010). Do the effects of subjective frequency and age of acquisition survive better word frequency norms? *Quarterly Journal of Experimental Psychology*, *64*, 545–559. doi:10.1080/17470218.2010.503374
- Brybaert, M., Van Wijnendaele, I., & De Deyne, S. (2000). Age-of-acquisition effects in semantic processing tasks. *Acta Psychologica*, *104*, 215–226. doi:10.1016/S0001-6918(00)00021-4
- Cameirão, M. L., & Vicente, S. G. (2010). Age-of-acquisition norms for a set of 1.749 Portuguese words. *Behavior Research Methods*, *42*, 474–480. doi:10.3758/BRM.42.2.474
- Cappa, S. F., Frugoni, M., Pasquali, P., Perani, D., & Zorati, F. (1998). Category-specific naming impairment for artefacts: A new case. *Neurocase*, *4*, 391–397.
- Carroll, J. B., & White, M. N. (1973a). Age-of-acquisition norms for 220 picturable nouns. *Journal of Verbal Learning and Verbal Behavior*, *12*, 563–576. doi:10.1016/S0022-5371(73)80036-2
- Carroll, J. B., & White, M. N. (1973b). Word frequency and age of acquisition as determiners of picture-naming latency. *Quarterly Journal of Experimental Psychology*, *25*, 85–95. doi:10.1080/14640747308400325
- Catling, J. C., & Johnston, R. A. (2006). The effects of age of acquisition on an object classification task. *Visual Cognition*, *13*, 968–980. doi:10.1080/13506280544000138
- Chalard, M., & Bonin, P. (2006). Age-of-acquisition effects in picture naming: Are they structural and/or semantic in nature? *Visual Cognition*, *13*, 864–883. doi:10.1080/13506280544000084
- Cortese, M. J., & Khanna, M. M. (2007). Age of acquisition predicts naming and lexical-decision performance above and beyond 22 other predictor variables: An analysis of 2,342 words. *Quarterly Journal of Experimental Psychology*, *60*, 1072–1082. doi:10.1080/17470210701315467
- Cortese, M. J., & Khanna, M. M. (2008). Age of acquisition ratings for 3,000 monosyllabic words. *Behavior Research Methods*, *40*, 791–794. doi:10.3758/BRM.40.3.791
- Cuetos, F., Barbón, A., Urrutia, M., & Domínguez, A. (2009). Determining the time course of lexical frequency and age of acquisition using ERP. *Clinical Neurophysiology*, *120*, 285–294. doi:10.1016/j.clinph.2008.11.003
- Cuetos, F., Ellis, A., & Alvarez, B. (1999). Naming times for the Snodgrass and Vanderwart pictures in Spanish. *Behavior Research Methods*, *31*, 650–658. doi:10.3758/BF03200741
- Cuetos, F., Herrera, E., & Ellis, A. W. (2010). Impaired word recognition in Alzheimer's disease: The role of age of acquisition. *Neuropsychologia*, *48*, 3329–3334. doi:10.1016/j.neuropsychologia.2010.07.017
- Cuetos, F., Samartino, T., Ellis, A. E. (2011). Age acquisition norms from elderly Spanish people: Characteristics and the prediction of word recognition performance in Alzheimer's disease. Retrieved 05-04-2011, from www.uv.es/psicologica/preprints/Cuetos2.pdf

- De Bleser, R., & Kauschke, C. (2003). Acquisition and loss of nouns and verbs: Parallel or divergent patterns? *Journal of Neurolinguistics*, *16*, 213–229. doi:10.1016/s0911-6044(02)00015-5
- De Deyne, S., & Storms, G. (2007). Age-of-acquisition differences in young and older adults affect latencies in lexical decision and semantic categorization. *Acta Psychologica*, *124*, 274–295. doi:10.1016/j.actpsy.2006.03.007
- De Deyne, S., & Storms, G. (2008). Word associations: Network and semantic properties. *Behavior Research Methods*, *40*, 213–231. doi:10.3758/BRM.40.1.213
- Dell'Acqua, R., Lotto, L., & Job, R. (2000). Naming times and standardized norms for the Italian PD/DPSS set of 266 pictures: Direct comparisons with American, English, French, and Spanish published databases. *Behavior Research Methods, Instruments, & Computers*, *32*, 588–615. doi:10.3758/BF03200832
- Dimitropoulou, M., Duñabeitia, J., Blitsas, P., & Carreiras, M. (2009). A standardized set of 260 pictures for Modern Greek: Norms for name agreement, age of acquisition, and visual complexity. *Behavior Research Methods*, *41*, 584–589. doi:10.3758/BRM.41.2.584
- Ellis, A. W. (in press). The acquisition, retention and loss of vocabulary in aphasia, dementia and other neuropsychological conditions. In M. Faust (Ed.), *Handbook of the neuropsychology of language*. Oxford: Blackwell.
- Ellis, A. W., & Morrison, C. M. (1998). Real age-of-acquisition effects in lexical retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 515–523. doi:10.1037/0278-7393.24.2.515
- Fiebach, C. J., Friederici, A. D., Müller, K., von Cramon, D. Y., & Hernandez, A. E. (2003). Distinct brain representations for early and late learned words. *NeuroImage*, *19*, 1627–1637. doi:10.1016/s1053-8119(03)00227-1
- Funnell, E., & De Mornay Davies, P. (1996). JBR: A reassessment of concept familiarity and a category-specific disorder for living things. *Neurocase*, *2*, 461–474. doi:10.1080/13554799608402422
- Funnell, E., & Sheridan, J. (1992). Categories of knowledge? Unfamiliar aspects of living and nonliving things. *Cognitive Neuropsychology*, *9*, 135–153. doi:10.1080/02643299208252056
- Genzel, S., Kerkhoff, G., & Scheffter, S. (1995). PC-gestützte Standardisierung des Bildmaterials von Snodgrass & Vanderwart (1980). *Neurolinguistik*, *9*, 41–53.
- Gerhand, S., & Barry, C. (2000). When does a deep dyslexic make a semantic error? The roles of age-of-acquisition, concreteness, and frequency. *Brain and Language*, *74*, 26–47. doi:10.1006/brln.2000.2320
- Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General*, *113*, 256–281. doi:10.1037/0096-3445.113.2.256
- Ghyselinck, M., Custers, R., & Brysbaert, M. (2004). The effect of age of acquisition in visual word processing: Further evidence for the semantic hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 550–554. doi:10.1037/0278-7393.30.2.550
- Ghyselinck, M., De Moor, W., & Brysbaert, M. (2000). Age-of-acquisition ratings for 2816 Dutch four- and five-letter nouns. *Psychologica Belgica*, *40*, 77–98.
- Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation*, *12*, 395–427. doi:10.3758/BF03201693
- Heinze, H.-J., Munte, T.-F., & Kutas, M. (1998). Context effects in a category verification task as assessed by event-related brain potential (ERP) measures. *Biological Psychology*, *47*, 121–135. doi:10.1016/s0301-0511(97)00024-0
- Heister, J., Würzner, K.-M., Bubenzer, J., Pohl, E., Hanneforth, T., Geyken, A., & Kliegl, R. (2011). dlexDB—Eine lexikalische Datenbank für die psychologische und linguistische Forschung. *Psychologische Rundschau*, *62*, 10–20. doi:10.1026/0033-3042/a000029
- Hernandez, A. E., & Li, P. (2007). Age of acquisition: Its neural and computational mechanisms. *Psychological Bulletin*, *133*, 638–650. doi:10.1037/0033-2909.133.4.638
- Hirsh, K. W., & Funnell, E. (1995). Those old, familiar things: Age of acquisition, familiarity and lexical access in progressive aphasia. *Journal of Neurolinguistics*, *9*, 23–32. doi:10.1016/0911-6044(95)00003-8
- Hodgson, C., & Ellis, A. W. (1998). Last in, first to go: Age of acquisition and naming in the elderly. *Brain and Language*, *64*, 146–163. doi:10.1006/brln.1998.1960
- Holmes, S. J., & Ellis, A. W. (2006). Age of acquisition and typicality effects in three object processing tasks. *Visual Cognition*, *13*, 884–910. doi:10.1080/13506280544000093
- Howard, D., Best, W., Bruce, C., & Gatehouse, C. (1995). Operativity and animacy effects in aphasic naming. *International Journal of Language & Communication Disorders*, *30*, 286–302. doi:10.3109/13682829509021443
- Izura, C., Hernández-Muñoz, N., & Ellis, A. W. (2005). Category norms for 500 Spanish words in five semantic categories. *Behavior Research Methods*, *37*, 385–397. doi:10.3758/BF03192708
- Johnston, R. A., & Barry, C. (2006). Age of acquisition and lexical processing. *Visual Cognition*, *13*, 789–845. doi:10.1080/13506280544000066
- Khanna, M., & Cortese, M. (2011). Age of acquisition estimates for 1,208 ambiguous and polysemous words. *Behavior Research Methods*, *43*, 89–96. doi:10.3758/s13428-010-0027-y
- Kiran, S., Ntourou, K., & Eubank, M. (2007). The effect of typicality on online category verification on inanimate category exemplars in aphasia. *Aphasiology*, *21*, 844–866. doi:10.1080/02687030600743564
- Kiran, S., & Thompson, C. K. (2003a). Effect of typicality on online category verification of animate category exemplars in aphasia. *Brain and Language*, *85*, 441–450. doi:10.1016/S0093-934X(03)00064-6
- Kiran, S., & Thompson, C. K. (2003b). The role of semantic complexity in treatment of naming deficits: Training semantic categories in fluent aphasia by controlling exemplar typicality. *Journal of Speech, Language, and Hearing Research*, *46*, 608–622. doi:10.1044/1092-4388(2003)048
- Lahl, O., Göritz, A. S., Pietrowsky, R., & Rosenberg, J. (2009). Using the World-Wide Web to obtain large-scale word norms: 190,212 ratings on a set of 2,654 German nouns. *Behavior Research Methods*, *41*, 13–19. doi:10.3758/BRM.41.1.13
- Laiacona, M., Luzzatti, C., Zonca, G., Guarnaschelli, C., & Capitani, E. (2001). Lexical and semantic factors influencing picture naming in aphasia. *Brain and Cognition*, *46*, 184–187. doi:10.1016/s0278-2626(01)80061-0
- Lambon Ralph, M. A. L., Graham, K. S., Ellis, A. W., & Hodges, J. R. (1998). Naming in semantic dementia—What matters? *Neuropsychologia*, *36*, 775–784. doi:10.1016/s0028-3932(97)00169-3
- Manoiloff, L., Artstein, M., Canavoso, M., Fernández, L., & Segui, J. (2010). Expanded norms for 400 experimental pictures in an Argentinean Spanish-speaking population. *Behavior Research Methods*, *42*, 452–460. doi:10.3758/BRM.42.2.452
- Marques, J. F., Fonseca, F. L., Morais, A. S., & Pinto, I. A. (2007). Estimated age of acquisition norms for 834 Portuguese nouns and their relation with other psycholinguistic variables. *Behavior Research Methods*, *39*, 439–444. doi:10.3758/BF03193013
- Monetta, L., Tremblay, T., & Joanette, Y. (2003). Semantic processing of words, cognitive resources and N400: An event-related potentials study. *Brain and Cognition*, *53*, 327–330. doi:10.1016/s0278-2626(03)00136-2

- Morrison, C. M., Chappell, T. D., & Ellis, A. W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *Quarterly Journal of Experimental Psychology*, *50A*, 528–559. doi:10.1080/027249897392017
- Morrison, C. M., & Ellis, A. W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 116–133. doi:10.1037/0278-7393.21.1.116
- Morrison, C. M., Ellis, A. W., & Quinlan, P. T. (1992). Age of acquisition, not word frequency, affects object naming, not object recognition. *Memory & Cognition*, *20*, 705–714. doi:10.3758/BF03202720
- Morrison, C. M., & Gibbons, Z. C. (2006). Lexical determinants of semantic processing speed. *Visual Cognition*, *13*, 949–967. doi:10.1080/13506280544000129
- Morrison, C. M., Hirsh, K. W., Chappell, T., & Ellis, A. W. (2002). Age and age of acquisition: An evaluation of the cumulative frequency hypothesis. *European Journal of Cognitive Psychology*, *14*, 435–459. doi:10.1080/09541440143000159
- Nickels, L., & Howard, D. (1995). Aphasic naming: What matters? *Neuropsychologia*, *33*, 1281–1303. doi:10.1016/0028-3932(95)00102-9
- Nishimoto, T., Miyawaki, K., Ueda, T., Une, Y., & Takahashi, M. (2005). Japanese normative set of 359 pictures. *Behavior Research Methods*, *37*, 398–416. doi:10.3758/BF03192709
- Núñez-Peña, I., & Honrubia-Serrano, L. (2005). N400 and category exemplar associative strength. *International Journal of Psychophysiology*, *56*, 45–54. doi:10.1016/j.ijpsycho.2004.09.006
- Pind, J., Jónsdóttir, H., Gossurardóttir, H., & Jónsson, F. (2000). Icelandic norms for the Snodgrass and Vanderwart (1980) pictures: Name and image agreement, familiarity, and age of acquisition. *Scandinavian Journal of Psychology*, *41*, 41–48. doi:10.1111/1467-9450.00169
- Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, *12*, 1–20. doi:10.1016/s0022-5371(73)80056-8
- Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, *104*, 192–233. doi:10.1037/0096-3445.104.3.192
- Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573–605. doi:10.1016/0010-0285(75)90024-9
- Ruts, W., De Deyne, S., Ameel, E., Vanpaemel, W., Verbeemen, T., & Storms, G. (2004). Dutch norm data for 13 semantic categories and 338 exemplars. *Behavior Research Methods, Instruments, & Computers*, *36*, 506–515. doi:10.3758/BF03195597
- Sanfeliu, M., & Fernandez, A. (1996). A set of 254 Snodgrass–Vanderwart pictures standardized for Spanish: Norms for name agreement, image agreement, familiarity, and visual complexity. *Behavior Research Methods*, *28*, 537–555. doi:10.3758/BF03200541
- Schröder, A., Kauschke, C., & De Bleser, R. (2004). Messungen des Erwerbsalters für konkrete Nomina. *Neurolinguistik*, *17*, 107–138.
- Sirois, M., Kremin, H., & Cohen, H. (2006). Picture-naming norms for Canadian French: Name agreement, familiarity, visual complexity, and age of acquisition. *Behavior Research Methods*, *38*, 300–306. doi:10.3758/BF03192781
- Snodgrass, J. G., & Vanderwart, M. (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. doi:10.1037/0278-7393.6.2.174
- Snodgrass, J. G., & Yuditsky, T. (1996). Naming times for the Snodgrass and Vanderwart pictures. *Behavior Research Methods, Instruments, & Computers*, *28*, 516–536. doi:10.3758/BF03200540
- Stadthagen-Gonzalez, H., & Davis, C. J. (2006). The Bristol norms for age of acquisition, imageability, and familiarity. *Behavior Research Methods*, *38*, 598–605. doi:10.3758/BF03193891
- Stanczak, L., Waters, G., & Caplan, D. (2006). Typicality-based learning and generalisation in aphasia: Two case studies of anomia treatment. *Aphasiology*, *20*, 374–383. doi:10.1080/02687030600587631
- Stewart, F., Parkin, A. J., & Hunkin, N. M. (1992). Naming impairments following recovery from herpes simplex encephalitis: Category-specific? *Quarterly Journal of Experimental Psychology*, *44A*, 261–284. doi:10.1080/02724989243000037
- Stuss, D. T., Picton, T. W., & Cerri, A. M. (1988). Electrophysiological manifestations of typicality judgment. *Brain and Language*, *33*, 260–272. doi:10.1016/0093-934x(88)90068-5
- Thompson, C. K. (2007). Complexity in language learning and treatment. *American Journal of Speech-Language Pathology*, *16*, 3–5. doi:10.1044/1058-0360(2007)002
- Tsaparina, D., Bonin, P., & Méot, A. (2011). Russian norms for name agreement, image agreement for the colorized version of the Snodgrass and Vanderwart pictures and age of acquisition, conceptual familiarity, and imageability scores for modal object names. *Behavior Research Methods*. doi:10.3758/s13428-011-0121-9
- Turner, J. E., Valentine, T., & Ellis, A. W. (1998). Contrasting effects of age of acquisition and word frequency on auditory and visual lexical decision. *Memory & Cognition*, *26*, 1282–1291. doi:10.3758/BF03201200
- Uyeda, K. M., & Mandler, G. (1980). Prototypicality norms for 28 semantic categories. *Behavior Research Methods & Instrumentation*, *12*, 587–595. doi:10.3758/BF03201848
- Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language*, *50*, 289–335. doi:10.1016/j.jml.2003.10.003
- Vö, M. L.-H., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M. (2009). The Berlin Affective Word List Reloaded (BAWL-R). *Behavior Research Methods*, *41*, 534–538. doi:10.3758/BRM.41.2.534
- Vö, M. L.-H., Jacobs, A. M., & Conrad, M. (2006). Cross-validating the Berlin Affective Word List. *Behavior Research Methods*, *38*, 606–609. doi:10.3758/BF03193892
- Woollams, A. M., Cooper-Pye, E., Hodges, J. R., & Patterson, K. (2008). Anomia: A doubly typical signature of semantic dementia. *Neuropsychologia*, *46*, 2503–2514. doi:10.1016/j.neuropsychologia.2008.04.005
- Zevin, J. D., & Seidenberg, M. S. (2002). Age of acquisition effects in word reading and other tasks. *Journal of Memory and Language*, *47*, 1–29. doi:10.1006/jmla.2001.2834
- Zevin, J. D., & Seidenberg, M. S. (2004). Age-of-acquisition effects in reading aloud: Tests of cumulative frequency and frequency trajectory. *Memory & Cognition*, *32*, 31–38. doi:10.3758/BF03195818