Word associations: Norms for
1,424 Dutch words in a continuous task

SIMON DE DEYNE AND GERT STORMS
University of Leuven, Leuven, Belgium

This study describes the collection of a large set of word association norms. In a continuous word association task, norms for 1,424 Dutch words were gathered. For each cue, three association responses were obtained per participant. In total, an average of 268 responses were collected for each cue. We investigated the relationship with similar procedures, such as discrete association tasks and exemplar generation tasks. The results show that the use of a continuous task allows the study of weaker associations in comparison with a discrete task. The effects of the continuous tasks were investigated for set size and the availability characteristics of the responses, measured through word frequency, age of acquisition, and imageability. Finally, we compared our findings to those of a semantically constrained version of the association task in which participants generated responses within the domain of a semantic category. Results of this comparison are discussed. The Appendix cited in this article is available at www.psychonomic.org/archive.

The study and use of word associations has been widespread in many fields of psychology since its conception as an academic discipline (Boring, 1950; Deese, 1965). In the late nineteenth century, pioneers such as Galton and Wundt started systematic investigations with psychometric, classificatory studies (Boring, 1950). Later, research on associations was used in clinical studies of human pathologies and intelligence, to be adopted by behaviorists and finally reinstated by modern cognitive psychologists who study language and memory (Cramer, 1968). Language and memory researchers are interested in word associations because, from a theoretical point of view, they are necessary at the center of understanding the organization of word knowledge. Although the exact interpretation of word associations is certainly not a closed chapter in cognitive psychology or psychology in general, it is generally agreed that word associations reflect our lexical knowledge acquired through world experience by means of words and the relationships between them (Nelson, McEvoy, & Schreiber, 2004). Furthermore, these structures capture important aspects of meaning or the semantic representation of words (de Groot, 1988; Nelson et al., 2004).

Besides the representation of meaning, the generation of word associations results in probabilities of responses that tell something about the retrieval mechanisms that underlie these responses (de Groot, 1988; Nelson, McEvoy, & Dennis, 2000) and memory performance in other tasks, such as cued recall experiments (Nelson et al., 2000; Nelson, McKinney, Gee, & Janczura, 1998; Steyvers, Shiffrin, & Nelson, 2005). Word associations can also be used as a gold standard, built upon a psychological foundation. This is the case when free association norms are compared with other norms in order to evaluate the validity of different types of procedures for indexing proximity. For instance, Steyvers et al. (2005) compared proximities based on a word association model with proximities derived from a model based on text collocations in the spirit of latent semantic analysis (LSA; Landauer & Dumais, 1997) in cued recall and semantic similarity judgments. Semantic representations based on word associations might be used to evaluate vector models relying on text collocation data such as LSA (Landauer & Dumais, 1997) and HAL (Burgess & Lund, 1997), where it is often unclear how semantic similarity between words comes about. Finally, in empirical studies of word processing, associative norms are often used to avoid confounding word variables such as printed word frequency (Nelson et al., 2004).

Clearly, the widespread use of association data in all these settings creates a considerable attention to the norms used in many previously reported findings. However, the largest sets of association norms in English (Nelson et al., 2004) and Dutch (de Groot, 1988) both employ a discrete version of the free association task. For both association norms, participants provide only one association for each cue. Although there are good reasons to restrict the collection of word associations to one response per cue, there are cases in which the continuous version of the word association task provides a meaningful way to gather these responses. One of the advantages of the continuous approach is that it allows a larger variability in associations—for instance, in the case of a very strong first association (e.g., *blood* and *red*). The collection of multiple responses leads to better estimates of the probabilities of weak associations. A continuous task, in which participants generate more than one

---

S. De Deyne, simon.dedeyne@psy.kuleuven.be

associate, is of particular interest to researchers who try to disentangle priming studies with mediated, associative, or semantic priming. Moreover, the representation of words in vector models based on multiple associations is much richer than vectors based on discrete associations that are the stock data in most studies. Finally, differences in early and late responses in word association tasks might give additional insight in the conceptual structure and the processes that drive the generation of word associates (Barsalou, Santos, Simmons, & Wilson, in press).

In previous large-scale association studies, such a procedure has been shunned since the generation of multiple responses, in comparison with a single, discrete response, evokes chaining and retrieval inhibition (McEvoy & Nelson, 1982). Both these processes can introduce a bias in the resulting norms. Chaining occurs when participants respond with associations based on a previous response rather than responding to the cue. To some extent, this effect can be reduced through instructions. Retrieval inhibition occurs when participants repeatedly sample responses for one cue. Research on retrieval inhibition in memory has shown that when people have to produce subsequent responses to the same cue, additional responses become hindered by a retrieval blockage (e.g., Raaijmakers & Shiffrin, 1981; Roediger, 1973). Again, some measures can limit the effect of response inhibition. First, in contrast to tasks in which participants give exhaustive associations (until they cannot generate any more responses), the task can be restricted to only a limited number of responses. Retrieval inhibition is expected to be of minimal influence whenever a limited number of associations per participant are gathered, in comparison with studies in which the continuous association task was time delimited. Second, it is expected that the effect of retrieval inhibition will be reduced if the stimuli are more diverse.

In this study, we present a Dutch data set of associations for 1,424 words using a continuous association task. Despite a long tradition of studying word associations, recent Dutch word association norms are hard to find. A recent extensive study for English words was conducted by Nelson et al. (2004). However, these associations were gathered at different moments during a period of over 20 years. The most recent extensive word association study for the Dutch language is over 20 years old (de Groot, 1988). In both of these association studies, the data that were gathered included various word classes other than nouns, such as adjectives and verbs. The majority of the association stimuli we selected consists of nouns, since we were primarily interested in the role of the associations between word and semantic representations of concrete concepts that have been documented with a wide range of other research variables, described elsewhere (see below).

DATA COLLECTION

Method

Participants. The study was conducted from 2003 to 2006 and involved 10,292 participating individuals. From this group, 6,329 were female, 3,582 were male, and 381 did not indicate their sexes. The average age was 24 years ($SD = 10.55$) and was indicated for all but 61 participants. The majority of the participants consisted of 1st-year students at the University of Leuven and the University of Ghent. Some of the students received credit for their participation. The remaining participants were 17-year-old sixth-grade high school students and volunteers.

Materials. The entire set of stimulus materials consisted of 1,424 words. This set included the words from the Ruts et al. (2004) corpus and from the Severens, Van Lommel, Ratinckx, and Hartsuiker (2005) study. The study of Ruts et al. contained concepts from various natural categories (*fruit*, *vegetables*, *insects*, *fish*, *birds*, *reptiles*, and *mammals*), artifact categories (*vehicles*, *musical instruments*, and *tools*), and action categories (*sports* and *professions*).[1] The Severens et al. items were used for picture-naming purposes and included a variety of concrete object concepts. The remainder of the items was taken from the semantic categories of *weapons*, *clothing*, *kitchen utensils*, *food*, *drinks*, and *animals*. Furthermore, this set was expanded with words corresponding to superordinate concept nouns such as *mammals* and *vehicles*. Finally, in the course of the data collection study, new words were added in order to provide norms for the most frequent association responses to the cue words described above. In other words, the cue *strawberry* was normed if its most frequent association response, *red*, was also presented as a cue in this phase of data collection. The final set of words that served as cues, from the original set of cues and the set of cues for the most frequent responses, consisted of 1,266 nouns, 77 verbs, 80 adjectives, and 1 numeral.

Procedure. The collection of the data was completed in two phases. During the first phase in 2003, the data for 338 words from Ruts et al. (2004) were collected. Each word in Ruts et al. belonged to a specific semantic category (described earlier). In this phase, the association task was administered collectively. The participants received a sheet with instructions written at the top. They were instructed to write down, for every stimulus word, the first three associations that entered their minds. If a word was unknown, they were instructed to underline that word, a procedure that was also applied by Nelson et al. (2004) and de Groot (1988). Every participant wrote down associations for a maximum of 14 different words, each belonging to a different semantic category. To minimize sequence effects, 10 different permutations were constructed, with the order of the categories randomized. Within a permutation, 30 unique lists were created (since the maximum number of stimulus words per category was 30). Some lists were shorter than 14 words because of the varying number of stimulus words per category. Completion of the task took less than 5 min.

During the second phase, a Web questionnaire was developed that corresponded closely to the written questionnaire. This questionnaire, together with new paper-and-pen questionnaires, was used to collect associations for the Severens et al. (2005) concepts and norms for the items that were already collected. For the Web questionnaire, participants were allowed to write down associations for a minimum of 10 and a maximum of 30 words when they logged into the system. No restrictions were imposed on the composition of the list of cues, since the cues used in the second phase were very diverse in terms of membership in semantic categories. The list content was determined completely at random. The second phase started in 2004 and was completed in 2006.

Results and Discussion

Data from 17 participants who completed less than 50% of the questionnaire were removed. For each word, there were associations given by at least 82 and at most 197 participants. This resulted in totals of a minimum of 246 (3 × 82) and a maximum of 591 (3 × 197) associations for every word. The association responses were preprocessed by converting the responses to lowercase words and replacing numerical strings smaller than 100 (e.g., 21) with their text counterparts (e.g., "twenty-one"). The total number of association responses in which the participant rated the stimulus as unknown was 1,169. These were coded as "x" for first, second, and third association responses. Some participants were not able to generate a second (2,230 participants) or

third (5,067 participants) association response. Adding these responses brings the total of unknown association responses to 8,466, or 2.2% of the data. Next, we collapsed the first, second, and third association responses to calculate the most frequently generated associations for our word cues. The frequencies for these association responses are listed in the Appendix (see the Archived Materials section).

The Appendix contains the complete set of words that served as cues, listed alphabetically in the first column together with the corresponding English translation in the second column. The following columns (3–8) contain the three most frequently associated responses together with their frequencies. These frequencies were calculated over associations generated as a first, second, or third response. Column 9 shows the total number of responses. Column 11 shows the number of unknown responses. Column 11 shows the number of different nonidiosyncratic types (responses with a frequency larger than 1). The final column shows the number of different response types, including idiosyncratic responses.[2] The total number of different types generated over all cues was 30,311, whereas the number of unique cue response combinations was 133,401. The total number of responses was 381,909.

The reliabilities of the association responses were calculated using a procedure similar to that used in Nelson et al. (2000). The association data were split in two parts for the first, second, and third responses. For each of these parts, the frequency of each cue–response pair was calculated for the nonidiosyncratic responses. Then, the split-half reliability was calculated for each pair. The average split-half reliability of the first responses was $r_{\mathrm{splithalf}} = .82$ ($SD = .25$, for an average of 11 associates per cue). The reliability increases when both first and second responses are used ($r_{\mathrm{splithalf}} = .89$, $SD = .19$, for an average of 21 associates) and when all responses are combined, ($r_{\mathrm{splithalf}} = .91$, $SD = .09$, for an average of 31 associates).[3] The first responses are more uniform in comparison with the second and third responses. Since it can be expected that the strongest association will be given as a first response, the variability in responses increases in the second and third responses. However, in contrast to single-response norms, additional responses provide a better estimate of weak associates in the associative set of a word (Nelson et al., 2000). Often, researchers have explicit hypotheses about words that have few or many associates. Words with few associates are said to have a small set size, compared with words with many associates. In the following section, we examine the set size characteristics in the continuous association task.

SET SIZE

One of the purposes of collecting more than one response for each cue is to expand the set size of associates with norms for weak associates. We calculated the set size as the total number of different responses for each cue. Only responses that were generated by at least 2 participants were counted (in other words, no idiosyncratic responses were allowed). The average set size was 11 ($SD = 3.68$) for the first responses, 13 ($SD = 2.99$) for the second responses, and 14 ($SD = 2.91$) for the third responses. In the collapsed set,

the set size was 31 ($SD = 6.46$). These results indicate that second and third responses lead to a substantial increase of different responses (about 20 different associations) compared with the initial set size of 11 responses. The averages also show that the continuous procedure does result in denser representations of the words. As mentioned in the introduction, the density of the representations is important for spatial models of semantic memory.

To get a better idea about the reported set sizes, we calculated the overlap between the Dutch norms and the norms reported by Nelson et al. (2004). The former set contains on average 150 responses to 5,018 words. All Dutch cues were translated to English. On the basis of this translation, we found an overlap between both sets of 1,022 cues. Again, for each cue, the set size was calculated as the number of different responses that were generated by at least 2 participants. The set size of a cue will be larger if the cue is presented more frequently. To take into account differences in presentation rate of the cues, this number was divided by the number of times the cue was presented (i.e., a value between 82 and 197 for the Dutch associations and between 94 and 206 for the English associations).[4] The correlation between the English set size and the Dutch set size was $r = .28$ ($p < .001$) for the first responses and $r = .06$ and $r = .05$ for the second and third response set size, respectively. The correlation with the set size over the collapsed responses was $r = .24$ ($p < .001$). These values show only a weak correlation between the Dutch and English set sizes when the first response is considered, showing the importance of language-specific association norms. The number of different responses in the second and third degrees are not related to the English set size, which was based on a single association only. These results indicate that set size is a very dynamic property that fluctuates depending on the task (discrete or continuous) and the number of responses.

Figure 1 illustrates the effect of a multiple response procedure on set size by plotting the growth of response set size against the total number of responses at different time points. This figure shows that set size increases as the number of responses increases in a nonlinear way. The set size of the first response is less steep than the slope for later responses (R2, R3), indicating that later responses introduce a larger variability in response cues. However, Figure 1 shows that this increase tends to stabilize for later responses. Finally, the curve with the collapsed responses is a lot higher than the separate responses. This shows clearly that later responses are different types, not just responses similar to the first response but with different occurrences. When the set size is plotted for individual cues, curves similar to those shown in Figure 1 are observed, but each curve exhibits a different slope. A possible alternative estimate of set size might be based on the slope instead of on the absolute count of members in the set, since the slope parameter depends less on the number of responses that were collected.

The distributional aspects of word associations are also related to the utility or availability that words have in language. The following section describes the relationship between the frequency of the association responses and commonly used word availability measures.
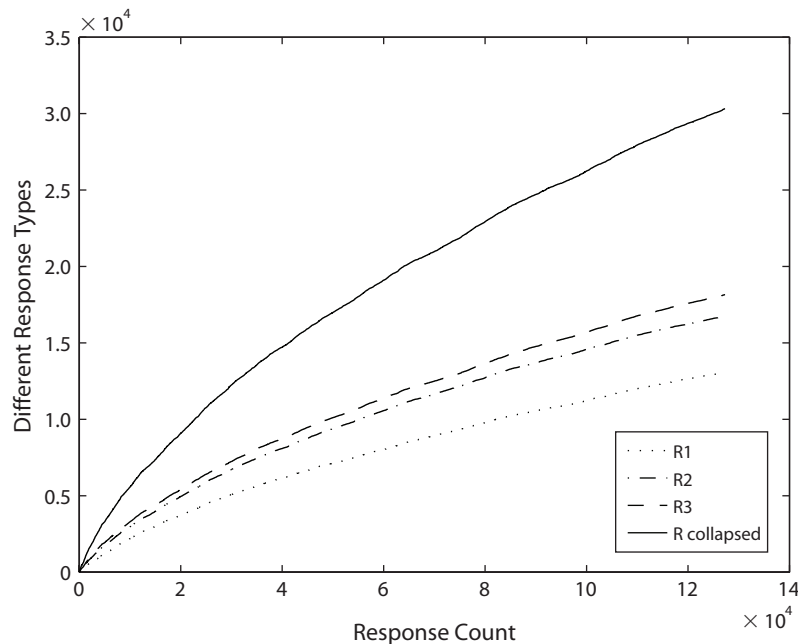
Figure 1. Growth of set size (i.e., different response types) for the first (R1), second (R2), third (R3), and collapsed (R collapsed) responses.

WORD AVAILABILITY EFFECTS

We investigated whether the frequency of each association is related to its frequency in language (CELEX frequency; Baayen, Piepenbrock, & van Rijn, 1993), the age at which a word is acquired (age of acquisition, or AoA), or how well it can be imagined (imageability). Note that the association response frequencies are different from cue set size, since the response frequencies are counted across cues instead of within cues. Apart from identifying the relation between association frequency and our three measures of availability, we also investigated the influence of word class on the relationship between word availability and association frequency of the responses. The responses included a large number of verbs and adjectives besides nouns. However, published AoA and imageability norms for verbs and adjectives are limited. To be able to perform a comparison for all measures of word availability, additional judgments were collected for AoA and imageability, which we will describe first. In contrast to previous studies that investigated the effect of word frequency or imageability (de Groot, 1989), the present study uses a correlational approach instead of an experimental one.

Method

Participants. A total of 48 Dutch-speaking 1st-year students of the faculty of economics at the University of Leuven participated in return for course credit. For the imageability ratings, there were 23 participants (12 males, 10 females) with a mean age of 19 years. For the AoA ratings, there were 25 students (11 males, 14 females) with a mean age of 18 years.

Materials and Procedure. For the entire data set of cue and association responses, the overlap with existing AoA and imageability norms was calculated. The AoA norms consisted of an aggregated set of ratings for 5,790 words taken from Ghyselinck, De Moor,

and Brysbaert (2000), Ghyselinck, Custers, and Brysbaert (2003), Ruts et al. (2004), and De Deyne and Storms (2007). Imageability ratings were taken from Ghyselinck (1999) and van Loon-Vervoorn (1985).

The set of imageability ratings was further expanded with words from the Severens et al. (2005) and Ruts et al. (2004) studies. Despite an extensive set of norms for nouns, the number of ratings for adjectives and verbs was small. To allow a more accurate investigation for these word classes, new norms for AoA and imageability were collected for 500 words. A total of 300 nouns, 121 adjectives, and 129 verbs were selected from the association responses. Only words that occurred more than once were chosen. Within this set, the words were selected randomly and stratified by word frequency. The task used the standard procedures for imageability ratings on a 7-point scale similar to that used by van Loon-Vervoorn (1985). The low end of the labeled scale corresponded to values for words for which a mental image was difficult to produce, whereas the high end corresponded to values for words for which mental images were readily available. For the AoA ratings, participants responded with an estimate of the age at which they had learned the word. The words were presented in random order on a computer. If a word was unknown, participants could indicate this by entering "99."

Results and Discussion

The reliability, established by the Spearman–Brown split-half method, was high for both AoA ($r_{splithalf} = .96$) and imageability ($r_{splithalf} = .92$) ratings. The average AoA rating was 7.31 ($SD = 2.54$) for nouns, 9.48 ($SD = 2.69$) for adjectives, and 7.65 ($SD = 2.52$) for verbs. The average imageability rating was 5.58 ($SD = 1.52$) for nouns, 3.93 ($SD = 1.77$) for adjectives, and 4.63 ($SD = 1.74$) for verbs.

The analyses were performed for nouns, adjectives, and verbs for which there were imageability norms as well as AoA norms and CELEX word or lemma frequencies available. The total number of stimuli for each set is listed

in Table 1. The frequency, AoA, and association count measures were log transformed to reduce their skewness.

We calculated the correlation between the frequency of all associations and three measures of response availability: word frequency, AoA, and imageability. For all variables, the correlations for the number of different response types and tokens (i.e., the total number of responses) were very similar. As can be seen from Table 1, association responses that were generated frequently are learned early, have a high word frequency, and are highly imaginable. The CELEX word form frequency tends to agree better with the association frequencies than the lemma frequencies. When looking at differences between the correlations for nouns, verbs, and adjectives separately, Table 1 indicates a remarkable finding. The correlation of word frequency and association frequency is less than half the size of the correlation of imageability or AoA with this variable. For nouns and adjectives, the strongest relation with association frequency was provided by AoA.

These results can be explained by studies about systematic differences between the semantics of nouns, verbs, and adjectives (Gasser & Smith, 1998; Gentner, 1981; Tyler, Russell, Fadili, & Moss, 2001). The results in these studies indicate that the learning of verbs is harder because of the extra overhaul of relational information that defines the meaning of the verb. The correlations in Table 1 indicate that acquisition age and imageability, rather than word frequency, capture the presumed difficulty and gradual increase of relational structure in adjectives and verbs and seem to fit well with the mechanistic account of Gentner and Kurtz (2005). According to this account, the greater conceptual density of entity concepts contributes to their perceptual availability and, thus, to their early acquisition. The prelinguistic infant has already individuated many entities and has only to attach the word to the referent. But relational terms such as verbs and prepositions pose a greater challenge since their referents are not simply "out there" in the experiential world; they are linguistically selected. To learn what a verb means, the child must discover which aspects of the situation enter into its meaning in his or her language (Gentner & Kurtz, 2005).

Despite this attractive proposal, only a few studies have investigated the effect of word class in AoA and word frequency effects. In one study, the effect of AoA in naming pictures of actions and objects was investigated. Although this variable influenced the naming of both actions and objects, there was no interaction based on the content of the

picture (Bogka et al., 2003). Given the limited amount of research in this matter, an interpretation at this point remains speculative. However, this finding might have important consequences on theories and models of word processing, since accounting for word frequency effects is pivotal.

CONCEPT CENTRALITY IN FREE AND CONTROLLED ASSOCIATION TASKS

A different type of continuous association task controls the semantic context of the responses. For example, Bousfield and Barclay (1951) asked participants to give as many associations that are "birds," "carpenter's tools," or "heavenly bodies" as possible within 18 min. Battig and Montague (1969) used a similar procedure and asked participants to write down as many instances belonging to each of 56 categories in 30 sec. The same study was replicated in Dutch for 39 categories by Storms (2001) and, more recently, for 13 categories by Ruts et al. (2004). All these studies have shown that the greater the belongingness or typicality of a stimulus to a particular concept, the more likely it is to elicit a response that belongs to the concept (Cramer, 1968). This difference in belongingness or typicality reflects the gradedness of categories. Based on our study of Dutch associates, gradedness can be estimated in an unconstrained task. The norms allow two different procedures to estimate this gradedness. First, one can simply look at the occurrence of instances in response to the category noun. Second, one can look at the frequency of producing category coordinates in response to a cue that is an instance. The latter procedure leads to an estimation of the category gradedness by combining the frequencies of all cues that are instances of a certain category. Furthermore, this procedure might reveal a different ranking that better approaches the gradedness of category by some external, theoretically motivated measure.

We investigated the representation of graded category responses in the association norms for each of the 13 categories listed by Ruts et al. (2004) to determine the extent to which free multiple association tasks exhibit the same gradedness found in the exemplar generation tasks described above. In addition to the categories in Ruts et al., three artifact categories (***clothing***, ***weapons***, and ***kitchen utensils***) for which generation frequencies were available in Storms (2001) were added to our analyses.

The first analysis was based on the superordinate cues (e.g., ***weapon***, ***mammal***, ***bird***, etc.) that were present in

Table 1
Correlation Between CELEX Log-Transformed Word Frequency (WF), CELEX Log-Transformed Lemma Frequency (LF), Imageability (IMA), Age of Acquisition (AoA), and the Total Association Responses Token (assoTOF) and Type (assoTYF) Frequencies for Nouns, Adjectives, and Verbs

	Nouns ($n = 2{,}647$)		Adjectives ($n = 160$)		Verbs ($n = 182$)	
	assoTOF	assoTYF	assoTOF	assoTYF	assoTOF	assoTYF
WF	.49**	.54**	.64**	.65**	.22*	.27**
LF	.39**	.45**	.62**	.62**	.12	.17*
IMA	.36**	.32**	.60**	.58**	.53**	.51**
AoA	−.49**	−.50**	−.74**	−.74**	−.56**	−.56**

*$p < .05$ (two-tailed).   **$p < .01$ (two-tailed).

the association study. For each of these superordinate cues, a part of the responses are category instances. To identify category instances, the lemmatized forms of the association responses were matched with the category exemplars in the exemplar generation tasks of Ruts et al. (2004) and Storms (2001). The total number of responses in these exemplar generation tasks was very high: between 361 responses for the smallest category (*amphibians*) and 1,002 responses for the largest category (*mammals*). The association responses that are category instances are thus a subset of these instance generation responses. For the overlapping sets of the 16 categories listed above, the Spearman rank order correlations were calculated.

The second analysis is based on category coordinate association responses such as *lettuce* to members of a certain category (e.g., *tomato* for *vegetables*). The category member cues consisted of the stimuli in Ruts et al. (2004). The coordinates that are associated for all members of a category should reflect the gradedness of these instances as well and provide an alternative measure of the availability of category instances. The frequency of the coordinate responses were collapsed over the category members and compared with the response frequencies in the tasks of Ruts et al. and Storms (2001).

The results of the analyses are shown in Table 2. As can be expected, the correlations for category- and exemplar-based frequencies in Table 2 shows a higher level of agreement for the category-based frequencies and the constrained association tasks (Ruts et al., 2004; Storms, 2001). Furthermore, the number of generated instances is in both tasks reasonable. Only the superordinate cue for the small category of *amphibians* generated fewer than 10 instances. Column 4 of Table 2 shows that category coordinates presented as cues do elicit a larger number of category members than do those of the category label cues. Naturally, the chance of generating more instances that are different is higher, since the total number of responses for the set of cues that make the category is larger than the responses for the corresponding category label. While the average correlation (*rho* = .54) is lower for the procedure based on category coordinates than for those of the superordinate cues (*rho* = .76), they are still moderate. Only the categories *professions* and *birds* have a weak correlation. Closer inspection of the generation frequencies of category coordinates for professions shows that only one profession was generated very often (*doctor*, 77 times), whereas the runner up response (*surgeon*) occurred only 6 times. Looking at the category members in the original Ruts et al. study shows that about a fourth of the members that were studied in the *professions* category were related to a medical profession (*pharmacist*, *veterinarian*, *doctor*, *physiotherapist*, *psychologist*, and *dentist*). The other category members tended to be more diverse. Therefore, the members intended to cover the professions category in Ruts et al. might be somewhat biased. For the *birds* category, the most generated instance based on association coordinate cues was *chicken* (99 times), whereas the constrained tasks have much lower frequencies for *chicken*. Actually, other birds (e.g., *rooster*, *duck*, *goose*) were listed more often as well. This again might indicate a bias in the category representa-

tion of *birds*. It is possible that a response bias effect influences the counts based on the category label. An indication is the smaller proportion of large birds. On the other hand, the opposite might also be true: Large birds are overrepresented in responses based on category coordinates. These biases can only be evaluated by introducing an external measure of category structure.

It must also be noted that the procedure allows that some instances in the association responses are not present in the exemplar generation task and vice versa. However, the focus is on the distribution of the overlapping instances. In other words, does the free association task exhibit the same pattern of belongingness as the controlled association in the exemplar generation task? The results shown above seem to confirm this. This finding opens up an additional use of the association norms as estimates of gradedness in a category instantiated by what is also called exemplar strength in the categorization literature (e.g., Casey, 1992; Hampton, 1997; Larochelle, Richard, & Soulières, 2000; Ward, Patterson, Sifonis, Dodds, & Saunders, 2002). The reverse measure, category strength, is an associative measure of strength between an exemplar and a superordinate category. This index can also be directly derived from the norms. To return to our initial example, *vegetable* was associated with the cue *tomato* by 19 participants, whereas only 1 participant responded with *fruit*. As these examples illustrate, the big advantage of using the word association norms is that norms of exemplar or category strength for arbitrary categories can be selected.

GENERAL DISCUSSION

In this article, we have described a set of free association norms for a large number of concepts. In contrast to most large-scale association sources, three responses were gathered for each association. Our study of the response set size

Table 2
Category- and Exemplar-Based Instance Generation Frequencies Derived From the Free Association Task Spearman Rank Order Correlations (*rho*) With the Generation Instances in the Controlled Association Task of Storms (2001) and Ruts et al. (2004)

Category	Category Based			Exemplar Based		
	rho	p	n	rho	p	n
Amphibians	.90	.01	7	.67	.01	14
Reptiles	.76	.00	12	.83	.00	20
Insects	.89	.00	16	.70	.00	31
Fish	.64	.01	16	.46	.01	28
Birds	.61	.01	16	.19	.23	41
Mammals	.80	.00	20	.61	.00	52
Fruit	.79	.00	17	.74	.00	35
Vegetables	.87	.00	18	.63	.00	36
Musical instruments	.80	.00	20	.61	.00	38
Kitchen utensils	.67	.00	24	.51	.00	59
Tools	.74	.00	27	.49	.00	58
Weapons	.87	.00	13	.43	.00	53
Clothing	.50	.02	22	.74	.00	60
Vehicles	.83	.00	18	.62	.00	38
Sports	.80	.00	22	.33	.01	61
Professions	.69	.00	30	.12	.47	37
$\bar{\chi}$	.76		18	.54		41
SD	.11		6	.20		15

indicates that later (i.e., second and third) responses added associations that would go undetected in a discrete task. The same analysis also points out that care should be taken when properties such as response set size are regarded as static. In general, word associations are very dynamic. Variables such as set size will depend on the amount of data that is collected. Similarly, the number of associations between words might grow during a participant's lifetime or through the evolution of word use over the years. In this respect, some caution for deriving measures is in place. Unavoidable measures such as set size are based on a simplification, which is warranted in some but not all situations.

Effects of word availability have been found in many different tasks and languages (e.g., Alario & Ferrand, 1999; Dell'Acqua, Lotto, & Job, 2000; Ellis & Morrison, 1998) during which a concept has to be retrieved. Generating association responses is heavily constrained by the availability. Previous studies have found that imageability, and not word frequency, puts the largest constraint on association responses (de Groot, 1989). In contrast to these findings, our results indicate that AoA and word frequency are better predictors of association response frequencies. Moreover, we found that this relationship was moderated by word class. The number of studies that have investigated the relationship between word class and word availability measures are limited, but the confirmation of theoretical predictions in the context of verb acquisition (Gentner, 1981; Gentner & Kurtz, 2005) is encouraging for future research directions. The general theoretical foundations of word frequency and AoA effects have been explored in a study by Steyvers and Tenenbaum (2005), who explored the small-world network topology and association model of network growth. The companion paper to this article (De Deyne & Storms, 2008) provides a more detailed treatment of availability as node centrality in an association network and the small-world properties of these networks.

The selection of the cue concepts was partly inspired by the need for norms in semantic tasks. In the context of semantic categorization research, the relationship between instances and the category are often described in terms of exemplar and category strength or dominance (e.g., Casey, 1992; Hampton, 1997; Larochelle et al., 2000; Ward et al., 2002). Controlled associations tasks to estimate these variables show strong agreement with the gradedness or exemplar/category strength measures obtained in a free continuous association task. However, in contrast to these controlled tasks, specific categories can be extracted from the norms ad hoc. When the category gradedness is represented through generation frequencies of a representative set of category instances, the results diverge in a systematic manner and might be used as an additional measure in future empirical research. In this article, superordinate, coordinate, and subordinate information was extracted to investigate the correspondence with other existing procedures. However, this information only captures a fraction of the meaningful relations that are present in the association data. The semantic composition of the association responses can offer additional insight into the question of what information word associations really represent. This topic will also be addressed in the companion article (see pp. 213-231).

The procedure of asking for multiple associations produces denser association representations, which will be especially appropriate for research in semantic memory models that depend on information about weak associations for estimates of semantic proximity. One class of such models consists of high-dimension vector models such as HAL (Burgess & Lund, 1997) and LSA (Landauer & Dumais, 1997). However, the representations of these models are based on text collocates and are often hard to translate to the corresponding psychological representations. Association-based vector models might provide a better yardstick to evaluate these text-based models. Furthermore, most research employs some metric of strong or weak associates to words that is based on a straightforward measure of associative strength or relatedness. Again, an association vector model captures more of the distribution of weak associates in calculating proximity or relatedness between two words. Such a model might be able to capture some empirical data better than the simple measures of relatedness.

Regardless of the concept research orientation in the norms, we do expect them to be suitable in a wide range of studies, such as the study of semantic and associative priming. Again, the benefit resides in the fact that each cue has an extensive set of word associates. This reduces the chance that a very strong word associate absorbs a weaker word associate. Besides priming, associations are also often used in false memory research in which researchers are typically interested in intrusions of associated concepts after tasks such as list learning (e.g., McDermott & Roediger, 1998). The present norms allow for extended ways to study association strength in these and many other experiments.

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.

BAAYEN, R., PIEPENBROCK, R., & VAN RIJN, H. (1993). *The CELEX lexical database* [CD-ROM]. Philadelphia: University of Pennsylvania, Linguistic Data Consortium.

BARSALOU, L. W., SANTOS, A., SIMMONS, W., & WILSON, C. D. (in press). Language and simulation in conceptual processing. In M. De Vega, A. M. Glenberg, & A. C. Graesser (Eds.), *Symbols, embodiment, and meaning*. Oxford: Oxford University Press.

BATTIG, W. F., & MONTAGUE, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monographs*, **80**, 1-45.

BOGKA, N., MASTERSON, J., DRUKS, J., FRAGKIOUDAKI, M., CHATZIPROKOPIOU, E.-S., & ECONOMOU, K. (2003). Object and action picture naming in English and Greek. *European Journal of Cognitive Psychology*, **15**, 371-403.

BORING, E. (1950). *A history of experimental psychology* (2nd ed.). New York: Appleton-Century-Crofts.

BOUSFIELD, W., & BARCLAY, W. (1951). The application of Zipf's analysis of language to sequences of restricted associative responses. *Journal of General Psychology*, **44**, 253-260.

BURGESS, C., & LUND, K. (1997). Modelling parsing constraints with high-dimensional context space. *Language & Cognitive Processes*, **12**, 177-210.

CASEY, P. J. (1992). A reexamination of the roles of typicality and category dominance in verifying category membership. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 823-834.

CRAMER, P. (1968). *Word association*. New York: Academic Press.

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.

DE DEYNE, S., & STORMS, G. (2008). Word associations: Network and semantic properties. *Behavior Research Methods*, **40**, 214-232.

DEESE, J. (1965). *The structure of associations in language and thought*. Baltimore: Johns Hopkins Press.

DE GROOT, A. M. B. (1988). Woordassociatienormen met reactietijden [Word association norms with reaction times]. *Nederlands Tijdschrift voor de Psychologie en Haar Grensgebieden*, **43**, 280-296.

DE GROOT, A. M. B. (1989). Representational aspects of word imageability and word frequency as assessed through word association. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 824-845.

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.

ELLIS, A. W., & MORRISON, C. (1998). Real age-of-acquisition effects in lexical retrieval. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 515-523.

GASSER, M., & SMITH, L. B. (1998). Learning nouns and adjectives: A connectionist account. *Language & Cognitive Processes*, **13**, 269-306.

GENTNER, D. (1981). Some interesting differences between nouns and verbs. *Cognition & Brain Theory*, **4**, 161-178.

GENTNER, D., & KURTZ, K. J. (2005). Relational categories. In W.-K. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman, & P. Wolff (Eds.), *Categorization inside and outside the laboratory: Essays in honor of Douglas L. Medin* (pp. 151-175). Washington, DC: American Psychological Association.

GHYSELINCK, M. (1999). [Imageability responses for Dutch words]. Unpublished raw data.

GHYSELINCK, M., CUSTERS, R., & BRYSBAERT, M. (2003). Age-of-acquisition ratings for 2322 Dutch words from 49 different semantic categories. *Psychologica Belgica*, **43**, 181-214.

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.

HAMPTON, J. (1997). Associative and similarity-based processes in categorization decisions. *Memory & Cognition*, **25**, 625-640.

LANDAUER, T. K., & DUMAIS, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, **104**, 211-240.

LAROCHELLE, S., RICHARD, S., & SOULIÈRES, I. (2000). What some effects might not be: The time to verify membership in "well-defined" categories. *Quarterly Journal of Experimental Psychology*, **53A**, 929-961.

McDERMOTT, K. B., & ROEDIGER, H. L., III (1998). Attempting to avoid illusory memories: Robust false recognition of associates persists under conditions of explicit warnings and immediate testing. *Journal of Memory & Language*, **39**, 508-520.

McEVOY, C. L., & NELSON, D. L. (1982). Category name and instance norms for 106 categories of various sizes. *American Journal of Psychology*, **95**, 581-634.

NELSON, D. L., McEVOY, C. L., & DENNIS, S. (2000). What is free association and what does it measure? *Memory & Cognition*, **28**, 887-899.

NELSON, D. L., McEVOY, C. L., & SCHREIBER, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, **36**, 402-407.

NELSON, D. L., McKINNEY, V. M., GEE, N. R., & JANCZURA, G. A. (1998). Interpreting the influence of implicitly activated memories on recall and recognition. *Psychological Review*, **105**, 299-324.

RAAIJMAKERS, J. G. W., & SHIFFRIN, R. M. (1981). Search of associative memory. *Psychological Review*, **88**, 93-134.

ROEDIGER, H. L., III (1973). Inhibition in recall from cueing with recall targets. *Journal of Verbal Learning & Verbal Behavior*, **12**, 644-657.

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.

SEVERENS, E., VAN LOMMEL, S., RATINCKX, E., & HARTSUIKER, R. J. (2005). Timed picture naming norms for 590 pictures in Dutch. *Acta Psychologica*, **119**, 159-187.

STEYVERS, M., SHIFFRIN, R. M., & NELSON, D. L. (2005). Word association spaces for predicting semantic similarity effects in episodic memory. In A. F. Healy (Ed.), *Experimental cognitive psychology and its applications* (pp. 237-249). Washington, DC: American Psychological Association.

STEYVERS, M., & TENENBAUM, J. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, **29**, 41-78.

STORMS, G. (2001). Flemish category norms for exemplars of 39 categories: A replication of the Battig and Montague (1969) category norms. *Psychologica Belgica*, **41**, 145-168.

TYLER, L. K., RUSSELL, R., FADILI, J., & MOSS, H. E. (2001). The neural representation of nouns and verbs: PET studies. *Brain*, **124**, 1619-1634.

VAN LOON-VERVOORN, W. (1985). *Voorstelbaarheidswaarden van Nederlandse woorden: 4600 substantieven, 1000 verba, 500 adjectieven* [Imageability ratings of Dutch words: 4600 nouns, 1000 verbs, 500 adjectives]. Lisse, Switzerland: Swets & Zeitlinger.

WARD, T. B., PATTERSON, M. J., SIFONIS, C. M., DODDS, R. A., & SAUNDERS, K. N. (2002). The role of graded category structure in imaginative thought. *Memory & Cognition*, **30**, 199-216.

NOTES

1. We mark the category words with bold italic typeface and use a normal typeface for specific category members or instances.

2. All data are available upon request from the author. For each response, the age and sex of the participant, together with information about the procedure (Web or pen-and-paper), are available.

3. The common interpretation of reliabilities has some limitations, since reliability varies strongly among the cues, indicating that some cues have a more diverse set of associate responses. This would not change if more data were collected for cues with low reliability.

4. This is a simplification, since the relationship between cue presentation and set size is nonlinear (see Figure 1).

ARCHIVED MATERIALS

The following materials associated with this article may be accessed through the Psychonomic Society's Norms, Stimuli, and Data archive, www.psychonomic.org/archive.

To access these files, search the archive for this article using the journal name (*Behavior Research Methods*), the first author's name (De Deyne), and the publication year (2008).

FILE: DeDeyne-BRM-2008.zip.

DESCRIPTION: The compressed archive file contains two files:

DeDeyne(2008).txt, containing a cue list and statistics for the three most frequently generated word associations.

DeDeyne(2008).xls, containing the above information in Excel spreadsheet format.

AUTHOR'S E-MAIL ADDRESS: simon.dedeyne@psy.kuleuven.be.