Exemplar by feature applicability matrices and other Dutch normative data for semantic concepts

SIMON DE DEYNE, STEVEN VERHEYEN, EEF AMEEL, WOLF VANPAEMEL,
MATTHEW J. DRY, WOUTER VOORSPOELS, AND GERT STORMS
University of Leuven, Leuven, Belgium

Features are at the core of many empirical and modeling endeavors in the study of semantic concepts. This article is concerned with the delineation of features that are important in natural language concepts and the use of these features in the study of semantic concept representation. The results of a feature generation task in which the exemplars and labels of 15 semantic categories served as cues are described. The importance of the generated features was assessed by tallying the frequency with which they were generated and by obtaining judgments of their relevance. The generated attributes also featured in extensive exemplar by feature applicability matrices covering the 15 different categories, as well as two large semantic domains (that of animals and artifacts). For all exemplars of the 15 semantic categories, typicality ratings, goodness ratings, goodness rank order, generation frequency, exemplar associative strength, category associative strength, estimated age of acquisition, word frequency, familiarity ratings, imageability ratings, and pairwise similarity ratings are described as well. By making these data easily available to other researchers in the field, we hope to provide ample opportunities for continued investigations into the nature of semantic concept representation. These data may be downloaded from the Psychonomic Society's Archive of Norms, Stimuli, and Data, www.psychonomic.org/archive.

During the past three decades, semantic concepts have been studied extensively in cognitive psychology. Different variables, pertaining to the manner in which the meaning of concepts is stored in semantic memory, have been identified. Examples of such well-studied variables are categorization decisions (e.g., McCloskey & Glucksberg, 1978; Verbeemen, Vanpaemel, Pattyn, Storms, & Verguts, 2007), response times in speeded categorization (Larochelle & Pineau, 1994; Storms, De Boeck, & Ruts, 2000), response times in feature verification (Rips, Shoben, & Smith, 1973), and correlations between concept features (Malt & Smith, 1984; McRae, de Sa, & Seidenberg, 1997).

The connection between the extension of a concept (i.e., the entities that are generally considered to be instances of the item group to which the concept noun refers) and the intension of that concept (i.e., the features that are important in delineating the meaning of the concept) could potentially inform many of these studies. In fact, information about which (potential) exemplars of a category possess which features has been used extensively in the study of semantic concepts to predict such phenomena as typicality (Ameel & Storms, 2006; Hampton, 1979; Rosch & Mervis, 1975), priming (Cree, McRae, & McNorgan, 1999), conceptual coherence (Sloman, Love, & Ahn, 1998), deep dyslexia (Plaut & Shallice, 1993), and so forth.

The information about which features a set of concepts possesses can easily be represented in binary matrix notation, where columns correspond to exemplars and rows represent the attributes or features. This sort of exemplar by feature applicability matrix is the starting point of Rosch and Mervis's (1975) seminal family resemblance notion, which they argued underlies semantic structure. Such a matrix also constitutes the input source for many connectionist models of semantic memory (e.g., Cree & McRae, 2003; Rogers & McClelland, 2004) and for structural data analytic techniques that have been used to study semantic concept representation (Storms & De Boeck, 1997; Storms, Van Mechelen, & De Boeck, 1994; Van Mechelen, Hampton, Michalski, & Theuns, 1993).

Surprisingly, large-scale data sets about which features characterize which concepts are rare in the literature. The data from the landmark article of Rosch and Mervis (1975) on family resemblance have been circulating informally (as Xerox copies), but the data set is limited to a small number of semantic concepts (i.e., the superordinate categories ***clothing***, ***fruit***, ***furniture***, ***vegetables***, ***vehicles***, and ***weapons***), each with a limited number of instances. Moreover, the data are too restricted to allow for a study of the relation between the selected superordinate categories.

In a few studies, such as McRae, Cree, Seidenberg, and McNorgan (2005) and Ruts, De Deyne, et al. (2004), data from large-scale feature generation studies have been presented. In these data sets, features are connected only to those concepts for which they were produced in the generation task. The participants, however, can be as-

G. Storms, gert.storms@psy.kuleuven.be

sumed to know much more about the stimulus concepts than the features they produced in the feature generation task (Tversky & Hemenway, 1984). For instance, although they may know perfectly well that a particular bird has two eyes, participants may omit such an obvious feature because they may try to concentrate on more distinguishing features within the category of **birds** or within the animal kingdom. The main purpose of the present work is to accommodate this void by making a set of exemplar by feature applicability matrices available to the public. Such matrices will be of additional value when they deal with concepts for which other information is already available. Therefore, our approach will be to start from an earlier norming study on semantic concepts, which we will extend with respect to both the concepts included and the measures obtained.

Before proceeding to the description of the concepts and the data, it is important to establish the terminology and notation that will be used throughout the article. We consider the word *concept* to refer to the meaning of either exemplars or categories. The notion of an *exemplar* will be used to denote a lexical entry, such as *tomato*, *airplane*, and so forth, and will be printed in italics. A group of exemplars forms a category, such as **birds**, **musical instruments**, **sports**, and so forth, and will be indicated in the text with a bold italic typeface. Two exemplars belonging to the same category will be referred to as *category coordinates*. In some cases, we will refer to the *category label*, instead of the *category*. This allows us to differentiate between the notion of a collection of different exemplars—that is, the category extension (e.g., **birds**)—and a lexical entry (e.g., *birds*) that points to the category intension. Features or associations that are provided in response to an exemplar or a category label, such as <*green*>, <*fruit*>, and <*sweet*> in response to *apple*, will be printed in an italic typeface between triangular brackets.

OVERVIEW OF CATEGORIES AND SELECTED EXEMPLARS

The present work builds on previous efforts by Ruts, De Deyne, et al. (2004), who made an extensive set of semantic norm data, gathered in the Dutch-speaking community of Flanders, Belgium, available. Ruts and colleagues have provided norm data for 13 superordinate categories, encompassing a total of 338 exemplars. Their data set contained information on 60 natural food instances selected from the categories of **fruit** and **vegetables**, 60 activities taken from the categories of **professions** and **sports**, 131 animals corresponding to 6 animal categories (**amphibians**, **birds**, **fish**, **insects**, **mammals**, and **reptiles**), and 87 artifacts selected from the categories **musical instruments**, **tools**, and **vehicles**. The 6 animal categories together cover the largest part of the (commonly known section of the) animal domain, and because the exemplars within each of the categories were selected so that they were representative for the categories, the complete collection of animal exemplars covers the animal domain quite well.

Numerous authors have argued that natural kind and artifact categories are mentally represented in fundamen-

tally different ways (e.g., Diesendruck & Gelman, 1999; Estes, 2003; Keil, 1989; Malt & Johnson, 1992). Therefore, it is desirable to ensure that the domain of artifacts is covered as well as that of animals. Since the data set of Ruts, De Deyne, et al. (2004) contained only three artifact categories, we decided to enlarge the artifact domain with three additional categories: **clothing**, **kitchen utensils**, and **weapons**. For each of these three categories, a representative sample of 29, 33, and 20 exemplars, respectively, was selected that covered a wide range of exemplar generation frequencies based on the norms reported in Storms (2001). All the categories and corresponding exemplars that will be discussed throughout this article, including the additional artifacts, are displayed in the Appendix. Categories and exemplars are provided in English and in the original Dutch wording that was used throughout the data gathering.

Note that some of the animal categories include exemplars that are not true members of the category according to a biological taxonomy. In the exemplar generation task that informed the selection of exemplars, these instances were frequently generated as exemplars of a biologically incorrect category. For instance, many students generated *dolphin* and *whale* as exemplars of **fish**. We decided to keep these (incorrect) animal exemplars in the category for which they were generated as instances. Many participants also did not appear to know the difference between **reptiles** and **amphibians** and often were not sure as to which of the two categories a particular animal belonged. We therefore decided to include all five exemplars of the **amphibians** category in the exemplar list of the **reptiles** category. As a result, in some of the tasks to be described below, we will make reference only to the category of **reptiles**. The Materials sections of the different tasks hold details on whether the five-item **amphibians** category was included next to the enlarged **reptiles** category or not.

Unlike the animal categories, the artifact categories are not mutually exclusive, and no generally accepted delineation of these categories exists (unlike for the animal categories, for which it is assumed there is a biological taxonomy of nonoverlapping categories). Several objects were generated frequently by participants as instances of more than one of the six artifact categories under study. Accordingly, in three cases, an instance was used in category-related judgments of two different categories: (1) *Knife* was included in the categories **kitchen utensils** and **tools**, and both (2) *axe* and (3) *rope* were included in the categories **weapons** and **tools**.

The data we will describe can be divided into three classes. A first class contains judgments and characteristics that pertain to the categories' exemplars. These measures were either newly gathered or taken from publicly available data sources. We will describe the procedures used to obtain typicality ratings, goodness ratings, goodness rank orders, exemplar generation frequencies, exemplar associative strength, category associative strength, estimated age of acquisition, word frequency, familiarity ratings, imageability ratings, and pairwise similarity ratings.

A second class of data contains the results of a feature generation task. Participants were required to provide features for the categories' exemplars and labels. The gener-

ated features were tallied, which resulted in an indirect measure of their importance: the feature generation frequency. The importance of the features was also assessed in a more direct manner by asking participants to rate how important they thought each of the features was for the category for which they were generated.

Most important, the final class holds large-scale exemplar by feature applicability matrices. Both at the level of superordinate categories (e.g., *birds*, *mammals*, *clothing*, *vehicles*) and at the level of two semantic domains (animals and artifacts), we had participants judge the applicability of features for the comprised exemplars. The features that were to be judged were generated either in response to the comprised exemplars or in response to the category labels. For the domain of animals, this meant having the participants judge the applicability of several hundred features for over 100 animals. For the domain of artifacts, this implied collecting judgments of the applicability of more than a thousand features for 169 artifacts. To our knowledge, such extensive semantic databases do not yet exist, and the availability of such large-scale exemplar by feature matrices provides ample opportunities for experimental and modeling approaches in the study of semantic concepts. To this end, all of the discussed materials, such as exemplars, categories, and generated features, were translated from Dutch into English.[1] Apart from the obvious need for empirical verification of the generalizability of our findings to other languages, we speculate that the collected measures might apply to other language communities. This will especially be the case for measures like the similarity and feature ratings, where the exposure to concepts is of minor importance.

PART I
Exemplar Judgments and Characteristics

In Ruts, De Deyne, et al. (2004), a variety of variables, known to be of importance in semantic processing, were identified. Two of these variables, typicality and exemplar generation frequency, pertain to the internal structure of semantic categories. Given the extensive study of typicality, we will include a discussion of three manners by which one may get at the variable: the rating of the typicality of category exemplars, the rating of the goodness of example of category members, and the ranking of category exemplars according to goodness of example. Furthermore, we will report on exemplar generation for all 16 categories, instead of the 13 discussed previously by Ruts, De Deyne, et al. The exemplar generation frequencies constitute just one measure of the strength with which a category and its exemplars are related. We will provide norms on two alternative measures of this relatedness as well: exemplar and category associative strength, which will be elaborated on following the discussion of the exemplar generation task.

In their norm data, Ruts, De Deyne, et al. (2004) also included variables that can be thought of as measures of availability. Following their onset, we will provide word frequencies and age-of-acquisition ratings for all 16 categories' members. We will extend the set of availability

measures by providing ratings of familiarity and imageability. Ruts, De Deyne, et al. also had participants provide ratings of the similarity of pairs of category members. In a final section of this part of the article, we will extend the existing data by reporting data from additional participants who performed this task. We also will provide pairwise similarities for the 3 artifact categories that were not included in the original norm data.

Typicality Ratings

Description

The study of the internal structure of categories through typicality was (Rips et al., 1973; Rosch & Mervis, 1975) and remains (Ameel & Storms, 2006; Verheyen, Ameel, & Storms, 2007) a frequently undertaken endeavor. Rated typicality has been shown to be an important determinant of behavioral outcomes as diverse as category verification time (Casey, 1992; Chumbley, 1986), priming (Rosch, 1975), categorization probability (Hampton, Dubois, & Yeh, 2006; McCloskey & Glucksberg, 1978), and semantic substitutability (Rosch, 1977), making it a variable well deserving of a place in any norming study of natural language concepts.

Method

Participants. One hundred twelve second-year psychology students at the University of Leuven (89 of them female and 23 male) participated in the typicality-rating task as part of a course requirement. Their ages ranged from 18 to 63 years ($M = 20.5$).

Materials. The materials consisted of the 16 categories, with 5–33 exemplars each, introduced above. The 6 animal categories (*amphibians*, *birds*, *fish*, *insects*, *mammals*, and *reptiles*) and the 6 artifact categories (*clothing*, *kitchen utensils*, *musical instruments*, *tools*, *vehicles*, and *weapons*) were included, as well as the natural food and activity categories (*fruit*, *vegetables*, *professions*, and *sports*).

Procedure. The data collection took place in a large classroom where all the participants were present at the same time. The participants received a booklet with instructions on the first page, followed by four sheets with a semantic category label printed in bold on top. Each of the category labels was followed by a list of 5–33 items, referring to exemplars. The participants were asked to indicate, for every item in the list, how typical it was for the category printed on top of the page. They used a Likert-type rating scale, ranging from 1 for *very atypical items* to 20 for *very typical items*. If they encountered an exemplar they did not know, they were asked to circle it. Every participant completed typicality ratings for four different categories. The assignment of categories to participants was randomized. For every category, four different random permutations of the exemplars were used, and each of these permutations was distributed with an equal frequency among the participants. All the exemplars of a category were rated by 28 different participants. The participants completed the task in less than 10 min.

Results

The reliability of the ratings within each of the 16 categories was evaluated by applying the Spearman–Brown formula to the split-half correlations. All reliability indices were calculated on 10,000 different randomizations of the participants. (Unless explicitly specified, this holds true for all the estimations of reliability reported throughout this article. In the remainder of the article, we will refer to *reliability* or *reliability coefficients* when using the procedure outlined above.) The second column of Table 1 shows

Table 1
Estimations of Reliability of Collected Data for the 16 Categories

Category	Typicality Rating	Goodness Rating	Goodness Rank Order	Exemplar Generation Frequency	Exemplar Associative Strength	Category Associative Strength	Acquisition Rating	Familiarity Rating	Imageability Rating	Similarity Rating
Amphibians	.92	.87	–	.99	.65	.99	.98	.90	.94	–
Birds	.97	.97	.97	.98	.97	.61	.85	.94	.94	.90
Fish	.96	.95	.95	.97	.65	.72	.93	.95	.95	.87
Insects	.87	.90	.89	.98	.71	.86	.95	.96	.95	.88
Mammals	.94	.96	.95	.99	.60	.97	.95	.96	.84	.92
Reptiles	.90	.90	.83	.99	.87	.98	.98	.98	.94	.94
Clothing	.98	.98	.96	.99	.76	.98	.93	.96	.75	.88
Kitchen utensils	.91	.91	.88	.99	<.01	.98	.96	.97	.88	.91
Musical instruments	.94	.94	.93	.98	.91	.95	.93	.95	.93	.94
Tools	.95	.96	.96	.98	.85	.99	.98	.98	.96	.89
Vehicles	.98	.98	.97	.99	.94	.99	.95	.98	.93	.96
Weapons	.98	.97	.96	.99	.85	.97	.80	.94	.88	.94
Fruit	.96	.97	.96	.99	.89	.99	.97	.96	.96	.91
Vegetables	.94	.93	.90	.98	.84	.99	.94	.96	.97	.86
Professions	.92	.90	.90	.95	.50	.75	.95	.96	.95	.91
Sports	.98	.97	.98	.98	.94	.94	.96	.95	.92	.90

the averaged reliability estimates. Typicality was rated very reliably: All the estimates were higher than .90, except for the category of ***insects***, where it reached a value of .87.

Goodness Ratings

Description

Although typicality has often been measured through judgments of the goodness of example of category exemplars (e.g., Barsalou, 1985; Mervis, Catlin, & Rosch, 1976), Kittur, Holyoak, and Hummel (2006) have argued that goodness of example and typicality may not always measure the same thing. They showed that, especially when exemplar familiarity and exemplar idealness differ, goodness-of-example ratings and typicality judgments result in distinct forms of graded structure. Because of the potential difference between both measures of gradedness, goodness ratings were gathered for all the studied category exemplars as well. The procedure employed to gather the goodness ratings was identical to the one used for obtaining typicality. Both tasks differed only with respect to the phrasing of the instructions.

Method

Participants. A group of 112 first-year students at the University of Leuven (80 of them female and 32 male) enrolled in an introductory psychology course participated in the goodness-rating task as part of a course requirement. Their ages ranged from 17 to 25 years ($M = 18.4$).

Materials. The materials were identical to those used in the typicality-rating task.

Procedure. As in the typicality-rating task, data collection took place in a large classroom where all the participants were present at the same time. The participants received a booklet with instructions on the first page, followed by four sheets with a semantic category label printed in bold on top. Each of the category labels was followed by a list of 5–33 items. The participants were asked to indicate, for every item in the list, how good an example it was of the category printed on top of the page. They used a Likert-type rating scale, ranging from 1 for *very bad examples* to 20 for *very good examples*. They were asked to circle items if they did not know what the words meant. Every participant completed goodness ratings for four dif-

ferent categories. The assignment of categories to participants was randomized. Four different random permutations of the exemplars were used for every category, and each of these permutations was distributed with an equal frequency among the participants. All the exemplars of a category were rated by 28 different participants. The participants completed the task in less than 10 min.

Results

The reliability of the ratings within each of the 16 categories is shown in the third column of Table 1. The estimations of reliability were high for the category of ***amphibians*** ($r = .87$) and were very high ($r > .90$) for the other categories. The goodness ratings had a strong correspondence with the typicality ratings. The correlations were .95 for ***amphibians***, .98 for ***birds***, .94 for ***fish***, .89 for ***insects***, .92 for ***mammals***, .84 for ***reptiles***, .98 for ***clothing***, .82 for ***kitchen utensils***, .96 for ***musical instruments***, .96 for ***tools***, .97 for ***vehicles***, .99 for ***weapons***, .95 for ***fruit***, .97 for ***vegetables***, .84 for ***professions***, and .97 for ***sports***.

Goodness Rank Order

Description

Because many of the participants in the goodness-rating task did not differentiate in their judgments between several (mostly good) examples of a category, a new task was administered in which the participants were asked to rank order all the category exemplars according to how good an example they were of the category. In other words, in this task, the participants were forced to differentiate between every exemplar of the category. A similar procedure had previously been employed by Barsalou (1983) and by Janczura and Nelson (1999).

Method

Participants. A group of 100 first-year students of the University of Leuven (76 of them female and 24 male) enrolled in an introductory psychology course participated in the goodness ranking task as part of a course requirement. Their ages ranged from 17 to 31 years ($M = 18.8$).

Materials. The materials were identical to those used in the goodness-rating task, except for the category of *amphibians*, which was not included.

Procedure. The data collection took place in a large classroom where all the participants were present at the same time. The participants received three sheets of paper with a semantic category label printed in bold on top, followed by a list of 20–33 items. The participants were given the following instructions:

> Below you will find a list of items belonging to category *X*. Read through the list carefully and circle the items you do not know. Then we would like you to indicate for the REMAINING words how good of an example you think they are of category *X*. Place a 1 next to the best example, a 2 next to the second best example, a 3 next to the third best example, and work your way through the list in this manner. The best example should get awarded the smallest number, whereas the poorest example should get awarded the highest number. In other words you are required to RANK ORDER the items. Therefore, make sure that all the words that were not initially circled get awarded a different number.

Every participant completed goodness rankings for three different categories. The assignment of categories to participants was randomized. Two different random permutations of the exemplars were used for every category, and each of these permutations was distributed with an equal frequency among the participants. In total, every category's exemplars were rank ordered by 20 different participants. The participants completed the task in less than 15 min.

Results

The reliability of the ratings within each of the 15 categories is shown in the fourth column of Table 1. The estimations of reliability were generally quite high, ranging from .83 for *reptiles* to .98 for *sports*. The goodness rankings had a strong correspondence with the goodness ratings. The correlations were −.94 for *birds*, −.86 for *fish*, −.82 for *insects*, −.78 for *mammals*, −.82 for *reptiles*, −.96 for *clothing*, −.80 for *kitchen utensils*, −.95 for *musical instruments*, −.88 for *tools*, −.94 for *vehicles*, −.96 for *weapons*, −.92 for *fruit*, −.89 for *vegetables*, −.88 for *professions*, and −.92 for *sports*. The correlations are negative because good examples are awarded a high number in the rating task, whereas they were to obtain a small number in the ranking task.

Exemplar Generation Frequency

Description

Production tasks, like the generation of category exemplars task, have provided insight into the extension of categories—concrete (Battig & Montague, 1969; Cohen, Bousfield, & Whitmarsh, 1957) and abstract (Hampton, 1981), as well as ad hoc (Barsalou, 1983) in nature. In addition, the frequency with which exemplars are generated in response to a particular category cue shows a strong relationship with the typicality of the exemplars (Barsalou, 1985; Hampton & Gardiner, 1983; Mervis et al., 1976). Both an unweighted measure (amounting to the number of participants who generated a particular exemplar) and a weighted measure (which takes into account the rank order of the generated exemplars of every participant) can be obtained from an exemplar generation task.

Method

Procedure. Storms (2001) and Ruts, De Deyne, et al. (2004) employed the exemplar generation procedure introduced by Battig and Montague (1969) in a population of Dutch-speaking participants. Details about the data-gathering procedure can be found in either of these articles. Storms and Ruts, De Deyne, et al. give details on the material and the participants involved. For each of the 16 categories under study, two measures of exemplar generation frequency were derived from the earlier published norms: an *unweighted measure*, which simply amounts to the number of participants (out of a total of 120) who generated a particular exemplar, and a *weighted measure*, which takes into account the rank order of the generated exemplars of every participant. In this weighted measure, the exemplar generated first is given a weight of 1, the exemplar generated second receives a weight of 2, and so forth. If a participant generated more than 11 exemplars, which seldom happened, all the generated exemplars beyond the 11th received a weight of 11.

Results

The reliability estimates of the weighted exemplar generation frequency measure can be found in column five of Table 1. They were all well above .90.

Exemplar and Category Associative Strength

Description

Several studies have shown that the association strength between an exemplar and its category is an important determinant of semantic categorization, sentence verification, and typicality rating (Ashcraft, 1978; Barsalou, 1985; Janczura & Nelson, 1999; Larochelle, Richard, & Soulières, 2000; Mervis & Rosch, 1981). Two measures of association strength are particularly important in the study of natural language concepts. The first measure, *exemplar strength*, indicates how frequently an exemplar presented as a cue elicits the category label as a response. The second measure is called *category strength* and measures exemplar dominance: the frequency with which an exemplar is produced in response to the name of a category. The category strength measure has a strong correspondence to the exemplar generation frequencies described earlier. However, the context in which instances are generated is a crucial difference between this task and the exemplar generation task. In the former, there is no category context present, meaning that participants can respond with antonyms, other superordinates, properties, and so forth. In the latter task, only category members are valid responses. At this point, it is unclear how the two generation measures compare in semantic tasks. Although the association strength between category and exemplar has some theoretical appeal, as compared with the more cut-to-the-task exemplar generation frequencies, behavioral evidence on the matter is still to be accumulated. To derive exemplar and category associative strength, the association norms of De Deyne and Storms (2008) were used.

Method

Procedure. Details about participants, materials, and procedure of the association data collection can be found in De Deyne and Storms (2008). This study describes the collection of association norms for 1,424 Dutch words in a continuous task. All 16 category labels, as well as the corresponding category instances, were in-

cluded among the 1,424 cues. In the association task, each participant generated three different associations for a certain cue. For instance, when the word *apple* was presented, a typical response of a single participant could be <*green*>, <*fruit*>, and <*sweet*>. For the following calculations, no distinction was made between first, second, or third associations. Instead, association frequencies were calculated by collapsing response frequencies over response position. Exemplar strength was calculated by counting, for each of the 16 categories' exemplars, the number of times the category label was given as an association response. Category strength was calculated by counting the number of times an exemplar was produced in response to the name of a category presented as cue. All the responses that matched an exemplar or category label's plural, diminutive, or orthographic variants were considered for the calculation of the measures. Both measures of strength were normalized by dividing the response frequency for the matching concepts by the total response frequency for a cue.

Results

The reliability of category associative strength was reasonably high. Only for the categories of *birds* ($r = .61$), *fish* ($r = .72$), *insects* ($r = .86$), and *professions* ($r = .75$) did the reliability not reach .90.

In contrast with the above, we found that the reliability for the exemplar associative strength varied strongly across categories (from $r < .51$ for *kitchen utensils* and *professions* to $r > .94$ for *birds*, *vehicles*, and *sports*). Apart from low agreement between participants, this could also stem from a small number of responses. We calculated the number of category responses across the exemplars of each category. The values were situated between two extremes: At the one extreme, for the category of *kitchen utensils*, for a total of 11,346 responses across all the exemplars, only 8 of the responses (i.e., 0.07%) listed the category label; at the other extreme, for the category of *fish*, 973 out of 6,189 responses (i.e., 16%) referred to the category label. For the remaining categories, this was 15% for *birds*, 5% for *insects*, 0.3% for *mammals*, 3% for *reptiles*, 2% for *clothing*, 5% for *musical instruments*, 1.3% for *tools*, 2% for *vehicles*, 2% for *weapons*, 9% for *fruit*, 9% for *vegetables*, 0.5% for *professions*, and 6% for *sports*.

We will briefly elaborate on this finding, since it has at least two important implications that should be mentioned. First, high exemplar or category association strength for certain categories (e.g., *fruit*) suggests that in categorization tasks, elaborate conceptual processing could be easily bypassed by superficial associative connections. Second, high exemplar association strength can indicate that, as has been found in American college students, certain category labels, such as *fish*, operate as basic-level concepts (Medin & Atran, 2004; Rosch, Mervis, Grey, Johnson, & Boyes-Braem, 1976).

A different use of the category taxonomic structure was also found for *mammals*. When exemplars of this category were presented as cues, the participants tended to respond more frequently with the association <*animal*> than with any of the category labels we have been employing throughout this study. The participants did not respond <*mammal*> to instances like *horse* or *monkey* but, instead, used the less technical term <*animal*>. For this reason, exemplar strengths to the association <*animal*> were also included among the norming data.

Finally, a discussion about important differences and commonalities between the association strength measures and strength derived from category-context-sensitive tasks such as the exemplar generation frequency can be found in De Deyne and Storms (2008).

Estimated Age of Acquisition

Description

The moment in life at which concepts or words are acquired has been found to be an important determinant of word processing. Estimates of this variable are obtained by asking adults at what age they acquired a word. Such estimates have proven to be valid and reliable approximations of the true age of acquisition (Carroll & White, 1973; Ghyselinck, Custers, & Brysbaert, 2003; Morrison & Ellis, 2000). Recently, it has been suggested that the structure of semantic networks is determined by the age at which concepts are learned (Steyvers & Tenenbaum, 2005; van Loon-Vervoorn, 1989). These claims were supported by experimental studies that showed processing advantages for early words in semantic tasks (e.g., Brysbaert, Van Wijnendaele, & De Deyne, 2000; De Deyne & Storms, 2007; van Loon-Vervoorn, 1989) and neuroimaging studies, showing differential processing regions for early and late acquired words (Fiebach, Friederici, Müller, von Cramon, & Hernandez, 2003).

The vast majority of the age-of-acquisition ratings that are presented in the data set were taken from the Dutch norms gathered by Ghyselinck et al. (2003). To complete the ratings for the entire stimulus set, norms were collected for 44 exemplars that were not included in the Ghyselinck et al. norms.

Method

Participants. Twenty students from the University of Leuven volunteered to provide the additional acquisition ratings. Thirteen of them were female, and 7 were male. Their ages ranged from 17 to 21 years ($M = 18.9$).

Procedure. The procedure was identical to that used in the Ghyselinck et al. (2003) study, except that the data were collected with paper and pencil, instead of through a computerized procedure. In the instructions, the participants were asked to indicate in numbers, for every word, the age at which they had first learned it. They were encouraged not to make guesses for words that were unknown to them but to underline these words instead. All the participants rated all 44 words. To minimize sequence effects, four random orders were constructed.

Results

The estimated reliability of the additional ratings was .98, which is comparable to the value of .96 reported by Ghyselinck et al. (2003). Column 8 of Table 1 holds the estimated reliabilities for each of the 16 categories separately.

Word Frequency

Description

One of the most widely studied phenomena in word processing is the effect of word frequency. The variable influences a wide range of different tasks that require access to meaning. Most important for semantic researchers is the influence word frequency exerts on categorization speed (e.g., Monsell, Doyle, & Haggard, 1989), semantic classi-

fication (e.g., Landauer, Ross, & Didner, 1979), and picture naming (e.g., Humphreys, Riddoch, & Quinlan, 1988). Due to the extensiveness of the effect, it cannot be ignored by any modern model or theory of word processing. The study of semantics with the norms presented might also shed light on the origins of availability effects, such as age of acquisition or word frequency. Although theoretical elaborations for the frequency effect abound (see Monsell, 1991, for an overview), recent research suggests that word frequency might be encoded in the semantic structure of language itself (Monaco, Abbott, & Kahana, 2007).

Method

Procedure. The reported frequency values are logarithmic function values of the lemma counts from the Dutch version of the CELEX database (Baayen, Piepenbrock, & van Rijn, 1993) for each of the exemplars. As in most studies in which word frequencies have been used, logarithmic function values were used to correct the skew in the frequency distribution. The frequency count is based on a corpus of 42,380,000 written words that consists mostly of written adult language up to 1993. As a consequence of the number of years that have passed since the material in the corpus was written, word frequencies of some of the stimulus words may be relatively over- or underestimated. (For instance, words like *microwave* may be used more frequently now than 15 years ago.) Nevertheless, the norms can be considered reliable enough for many practical research contexts.

Familiarity Ratings

Description

The prevalence of familiarity effects in many semantic tasks was first demonstrated by McCloskey (1980). McCloskey showed how many of the effects traditionally attributed to semantic variables can also be attributed to stimulus familiarity. McCloskey demonstrated how familiarity influenced category verification latencies in experiments by E. E. Smith, Shoben, and Rips (1974), Lorch (1978), and McCloskey and Glucksberg (1979). Subsequent studies have shown that familiarity also affects typicality ratings (Barsalou, 1985; Hampton & Gardiner, 1983) and lexical decisions (Balota & Chumbley, 1984; Gernsbacher, 1984).

Method

Participants. A group of 112 second-year psychology students at the University of Leuven (93 of them female and 19 male) participated in the familiarity rating task as part of a course requirement. Their age ranged from 19 to 35 years ($M = 20.2$).

Materials. The materials were identical to those used in the typicality- and goodness-rating tasks described above and thus consisted of 16 semantic categories with 5–33 exemplars each.

Procedure. As in the typicality- and goodness-rating tasks, the data collection took place in a large classroom where all the participants were present at the same time. The participants received a booklet with instructions on the first page, followed by four sheets with a semantic category label printed in bold on top. Each of the category labels was followed by a list of 5–33 items. The instructions were similar to those used by Gernsbacher (1984). The participants were asked to indicate, for every item in the list, how familiar they were with the item. The participants used a Likert-type rating scale, ranging from 1 to 5. They were instructed that a 1 meant that *they had never seen, heard, or used the word before*; a 2 meant that *they had seen, heard, or used it just once or twice*; a 3 meant that *they had seen, heard, or used it sometimes*; a 4 meant that *they had seen, heard, or used it often*; and a 5 meant that *they had seen, heard, or used it very often*. Every participant completed familiarity

ratings for four different categories. The assignment of categories to participants was randomized. Four different random permutations of the categories' exemplars were used for every category, and each of these permutations was distributed with an equal frequency among the participants. In total, every list was rated by 28 different participants. The participants completed the task in less than 10 min.

Results

The estimations of reliability are shown in the ninth column of Table 1. For all the categories, a very high reliability was obtained ($r > .90$).

Imageability Ratings

Description

In a variety of tasks, it has been shown that highly imageable words have a processing advantage, as compared with low-imageable words. This is true for tasks such as word association (De Deyne & Storms, 2008; de Groot, 1989), sentence comprehension (e.g., Holcomb, Kounios, Anderson, & West, 1999), and lexical decision (e.g., Coltheart, Patterson, & Marshall, 1980), although it is also worthwhile to note that processing disadvantages have been found (Warrington & Shallice, 1984). Theoretical interpretations of the effect, such as dual coding (Paivio, 1971, 1986) or context availability (Schwanenflugel, Akin, & Luh, 1992), have traditionally pointed toward semantic origins. A recent study has shown that imageability effects can also be explained in terms of the morphophonological properties of the words (Reilly & Kean, 2007).

Method

Participants. Forty-five male and 178 female psychology students at the University of Leuven participated in the rating task for course requirement. Their age varied from 17 to 26 years ($M = 18.5$).

Materials. The materials consisted of a list of 415 words, which included the exemplars of the 16 categories described in the tasks above. Exemplars of various other categories were also included in the list. Ratings for these words were collected for other research purposes.

Procedure. Unlike in the typicality-, goodness-, and familiarity-rating tasks described above, in which the items were presented in lists corresponding to the 16 semantic categories, the exemplars of the semantic categories were merged for the purpose of gathering imageability ratings. Every questionnaire consisted of 63 or 64 stimulus words from different categories. Every stimulus word was rated by 23–38 different participants ($M = 31.6$). Each questionnaire contained a unique randomized list. The first page of the questionnaire contained the following instructions:

> This study is set up to investigate the degree to which words are imageable. Indicate for each word on the following pages the degree to which it evokes a mental image. In other words, we would like to know to what extent a word evokes a sensory representation that refers to the object or the activity that is meant by the word. Indicate your rating on a scale ranging between 1 (*does not evoke a mental image at all*) and 7 (*does easily evoke a mental image*). If you don't know a particular word, please underline the word.

These instructions were similar to those used in previous studies on imageability (e.g., Altarriba, Bauer, & Benvenuto, 1999; Gilhooly & Logie, 1980). The participants completed the task in less than 10 min.

Results

Although ratings were not gathered within the context of semantic categories, the reliability of the ratings was evalu-

ated within each of the 16 categories. The estimations of reliability are shown in column 10 of Table 1. All of them were reasonably high. Only for the categories of **mammals** ($r = .84$), **clothing** ($r = .75$), **kitchen utensils** ($r = .88$), and **weapons** ($r = .88$) did reliability not reach .90.

Pairwise Similarity Ratings

Description

Similarities cannot be overlooked in a study intended to provide norm data on semantic concepts (Goldstone & Son, 2005). It is primarily due to similarity that these concepts can constitute a subject of study, for "*only our sense of similarity allows us to order things into kinds so that these can function as stimulus meanings*" (Quine, 1969, p. 114). Accordingly, many efforts have been tailored toward obtaining suitable measures of similarity (e.g., Borg & Groenen, 2005; Kruskal & Wish, 1978) and adequately representing these similarities (e.g., Pruzansky, Tversky, & Carroll, 1982; Sattath & Tversky, 1977; Shepard & Arabie, 1979; Tenenbaum, 1996; Verheyen et al., 2007). The resulting representations have been exploited in a range of cognitive models, including both exemplar and prototype models (e.g., Lee & Navarro, 2002; J. D. Smith & Minda, 1998; Smits, Storms, Rosseel, & De Boeck, 2002).

Method

Participants. Ninety-two students at the University of Leuven participated in the pairwise similarity rating task. Forty-two of them were paid the equivalent of $10/h. The remaining 50 participants received course credit for their participation.

Materials. Pairwise similarity ratings were collected for all exemplar combinations within each category, except for the category of **amphibians**. The five exemplars from this category were included in the **reptiles** category. Similarities were already available for the exemplars of the categories **birds**, **fish**, **insects**, **mammals**, **reptiles**, **musical instruments**, **tools**, **vehicles**, **professions**, and **sports** from the Ruts, De Deyne, et al. (2004) norms. Additional ratings for these categories' instances were gathered and added to the data set to ensure a minimum estimated reliability of .85. Furthermore, pairwise similarity ratings for the instances of the three artifact categories that were not included in the Ruts, De Deyne, et al. study (**clothing**, **kitchen utensils**, and **weapons**) were gathered as well.

Procedure. The rating task was administered on a PC. The participants were given standard instructions for pairwise similarity ratings. They were asked to judge the similarity of every presented pair and to enter a number between 1 (for *totally dissimilar*) and 20 (for *totally similar*). In case one or two words of an exemplar pair were unknown, they were asked to enter the value −1. Exemplar pairs were presented in a random order. Completion of all the exemplar pairs within a category took less than an hour. Every participant rated the exemplar pairs of at least two and at most seven categories. For the animal categories **birds**, **fish**, **insects**, **mammals**, and **reptiles**, a total of 17, 17, 17, 19, and 20 participants, respectively, participated in the task. For the artifact categories **clothing**, **kitchen utensils**, **musical instruments**, **tools**, **vehicles**, and **weapons**, the respective numbers of participants were 24, 25, 17, 18, 15, and 25. Similarity ratings were provided by 18, 16, 15, and 15 participants, respectively, for the categories **fruit**, **vegetables**, **professions**, and **sports**.

Results

The reliability of the pairwise similarity ratings for each of the categories can be found in the last column of Table 1. The averaged reliability estimates were all higher than the value of .85 that was aimed for.

PART II
Feature Listing, Frequency, and Importance

From the classical view on semantic concept representation (Sutcliffe, 1993), researchers studying natural language concepts have inherited an interest in features. Although it is no longer believed that concepts can be defined by a list of singly necessary and jointly sufficient features, it is commonly held that features, true of a concept, can offer a window into the concept's representation (Cree & McRae, 2003; Garrard, Lambon Ralph, Hodges, & Patterson, 2001; McRae et al., 2005).

Less agreement exists on how to select features to include when semantic phenomena are studied (Martin & Caramazza, 1980; Murphy, 1982). The selection of features has been based either on the intuition of the researchers involved or on the results of feature generation tasks. Whenever the latter method was employed, it was hardly ever explicitly verified whether there were prominent features that the participants did not mention. When asked to generate features for a particular stimulus, participants often fail to mention commonly known and generally considered important features, possibly because these features do not differentiate the stimulus from the others in the task set (McRae et al., 2005). This might become particularly obvious when the stimuli are category coordinates or belong to a meaningful, more general domain (Zannino, Perri, Pasqualetti, Caltagirone, & Carlesimo, 2006). Furthermore, features listed for one stimulus often seem to be equally true for another stimulus for which they were not listed (Tversky & Hemenway, 1984).

The limited set of cues, used in many generation tasks, constitutes an obvious drawback that can be remedied by obtaining a sizeable feature set relevant for a broad semantic domain. In this norming study, we provide such a large collection of features for two semantic domains: animals and artifacts. For other aspects of the feature selection process, it is not that clear whether they constitute disadvantages or opportunities. For instance, one could argue that the finding by Tversky and Hemenway (1984) that a feature, true of two concepts under study but generated only with regard to one of the concepts, is an indication of the feature's importance in the concepts' representations. Similarly, McRae et al. (2005) have suggested that participants' tendency to generate features that distinguish between stimuli in the task set may be exploited when certain hypotheses are tested or experiments are set up.

Here, we will adopt a rather pragmatic attitude toward these issues. In describing the feature generation procedure, we will make reference to the frequencies with which features were generated. These can be taken to indicate the importance that should be allocated to each of the features. In a following step, we will discuss how we assessed the importance of the generated features through participants' ratings. The possible desirability of access to large data sets containing a wide range of relevant features for various categories and larger semantic domains will be addressed in the third and final part of this article. It will include an overview of several exemplar by feature applicability matrices for 15 categories and the domains of artifacts and animals.

Feature Generation Frequencies

Description

The use of feature information in the study of semantic concepts is widespread. Examples include investigations regarding concept coherence (Sloman et al., 1998), similarity judgments (Markman & Gentner, 1993; Markman & Wisniewski, 1997), and cognitive economy in hierarchical taxonomies (Collins & Quillian, 1969; E. E. Smith et al., 1974). In studies aimed at understanding category-specific impairments (McRae & Cree, 2002; Zannino, Perri, Pasqualetti, Di Paola, et al., 2006), feature information is commonly employed as well. Depending on the abstraction level of the cues used in the generation task (i.e., exemplars vs. category labels), the generated features differ (Storms & De Boeck, 1997). Nevertheless, both types of features carry useful information. Rosch and Mervis (1975) succeeded in predicting typicality from features that were generated for the exemplars of a superordinate category, and Hampton (1979) was able to do the same by using features that were generated toward the category label. Ashcraft (1978) also showed how features generated for exemplars and category labels can be used to predict rated typicality. Furthermore, he established that the frequency with which a feature is generated in response to a particular concept is a significant predictor of the time to verify that the feature is true of the concept.

Method

Participants. In the feature generation task, 1,003 students at the University of Leuven participated for course credit (57%) or were paid the equivalent of $10/h (43%).

Materials. The materials consisted of the exemplars listed in the Appendix and corresponded to the items used in the various tasks described in the first part of this article. Apart from exemplars, features were also to be generated for 34 category labels. Among them were the labels of the 15 categories whose exemplars served as cues.

In the remainder of this article, we will use the term *exemplar feature* to refer to a feature that was generated for an exemplar. We will employ the term *category feature* to indicate a feature generated in response to a category label.

Procedure. The participants were asked to write down, preferably, 10 different features for 6 up to 10 different stimulus words. They received written instructions and one sheet of paper for every stimulus word, with space to write down their answers. The instructions were as follows:

> This booklet contains up to 10 sheets with a word written on top of the page. We would like you to write down, preferably, 10 features underneath the word. Try to give different sorts of features, such as, for example, physical or perceptual features (what it looks like, how it smells, how it tastes, . . .), functional features (what it is used for, when and where it is used, . . .), background information (where it comes from, some historical facts, . . .), etc.

The instructions further contained an example of 10 features for *pine tree*. No time limit was imposed. The participants never needed more than 5 min per word for the generation of features.

The presentation order of the stimulus words was randomized over participants and over categories. Within a booklet, one of five permutations of the categories from which the stimulus words were selected was used. The 34 category labels were considered a separate list, besides the exemplar lists that correspond to the 15 semantic categories. For every stimulus word, at least 20 different participants generated features. A minimum of 180 features was gathered for each stimulus word. (Note that not every participant succeeded in writing down 10 different features for every stimulus word presented. They were encouraged to continue until they wrote down 10 but were allowed to stop sooner if they could not think of enough features.)

The feature data were preprocessed by converting all words to lowercase. Minimal stemming was performed manually on the generated features: Synonyms were merged, feminine and masculine versions of the features were grouped, and plural forms were made singular. For both exemplar and category features, the numbers of tokens and types are presented in Table 2. Columns 2 and 3 of Table 2 contain the numbers of exemplar tokens and types summed across a category's exemplars, and columns 5 and 6 contain the counts of category tokens and types.

In a second phase, the features were further processed by removing redundant quantifiers (e.g., <*most of them can . . .*>, <*is often used for . . .*>). Next, both adjective–noun features (e.g., <*has brown fur*>) and disjunctive and conjunctive features (e.g., <*is used in the garden or in the house*>, <*is used in the garden and in the house*>) were broken up if the parts contained different information. Legible responses that were factually wrong (e.g., <*is a reptile*> for *salamander*) were not removed from the list.

Table 2
Number of Tokens and Types for Both Exemplar and Category Features

Category	Exemplar Features			Category Features		
	Tokens	Types	Selected	Tokens	Types	Selected
Birds	5,768	955	225	201	52	28
Fish	4,390	872	156	204	58	32
Insects	5,043	1,066	214	193	70	37
Mammals	5,772	1,202	288	193	60	34
Reptiles	3,877	825	179	193	70	35
Clothing	5,867	1,680	258	204	71	38
Kitchen utensils	6,683	1,210	328	197	73	39
Musical instruments	5,229	1,140	218	205	74	39
Tools	5,802	1,545	285	195	79	37
Vehicles	5,825	1,736	322	201	62	34
Weapons	4,082	1,032	181	203	54	32
Fruit	5,701	741	233	205	52	32
Vegetables	5,741	903	291	193	50	30
Professions	5,767	1,950	370	161	49	21
Sports	5,821	1,310	382	191	56	33

Note—The columns headed "Selected" hold the numbers of features withheld for the Type I applicability matrices in the case of exemplar features and for the Type III applicability matrices in the case of category features. These values also indicate the numbers of features included in the feature importance rating tasks.

Results

For both exemplar and category features, two different counts of the generation frequency were calculated. For exemplar features, the first involved a summation of frequencies across all the exemplars of a semantic category. Since exemplars of 15 categories were included in the feature generation task, 15 lists containing this kind of frequency information are included in the norm data. The second involved a summation of the frequencies across all exemplars of a semantic domain (either animals or artifacts). The generation frequencies of the category features were likewise counted in two different ways. The norm data include the category feature frequencies per semantic category. Furthermore, they include frequencies that were summed over all category labels of a semantic domain (either animals or artifacts). This means that for each feature, four types of summed frequencies are available. For instance, the feature <*can fly*> was generated 20 times as a category feature for *birds* and 203 times as a feature for the 30 exemplars of **birds**. The same feature was generated 31 times in response to all category labels of the animal domain and 341 times in response to all exemplars of the animal domain.

The most frequently generated features were withheld for the feature-rating and applicability tasks to be discussed below. The number of withheld exemplar features per category can be found in column 4 of Table 2. The number of withheld category features per category can be found in column 7 of Table 2.

Feature Importance Ratings: Exemplar Feature Description

In category membership judgments (Barton & Komatsu, 1989; Keil, 1989; Rips, 1989) and typicality ratings (Vanoverberghe & Storms, 2003), different properties can be shown to be of varying influence. Moreover, in production tasks, a particular feature may be listed for one semantic concept but not for another, despite the fact that the listed property can be said to be true of both concepts (Tversky & Hemenway, 1984). Similarly, across a range of participants, some features are more often produced in response to a semantic cue than are others (Ruts, De Deyne, et al., 2004). Findings like these suggest that not all features true of a semantic concept should be regarded as invariants. In other words, the various features listed for a particular category label or this category's exemplars might differ with regard to their relevance for the category (Hahn & Chater, 1997). Only rarely has the relevance or importance of features been assessed through direct judgments (for a notable exception, see Storms, De Boeck, Van Mechelen, & Geeraerts, 1993).

Method

Participants. Eighty-four psychology students at the University of Leuven (77 of them female and 7 male) participated in the rating task. Their age ranged from 18 to 22 years ($M = 18.3$). All of them were paid the equivalent of $10/h for their participation.

Materials. The features used were those generated for the exemplars of the 15 semantic categories included in the generation task. Only features that were generated at least four times across the members of a category were included. (See column 4 of Table 2 for the exact number of withheld exemplar features per category.)

Procedure. The participants were seated in front of a PC. Instructions taken from Storms et al. (1993) informed them that they would see a list of around 200 or 300 features that were generated by different participants for exemplars of a semantic category. The participants were then asked to rate how important or relevant they thought the features to be for the category on a 7-point rating scale, ranging from −3 (for *very unimportant features*) to +3 (for *very important features*). The task was illustrated with five feature examples of the category **furniture**. Feature lists were presented in Excel sheets, with a different line for each feature. Answers other than the ratings on the scale were not accepted (i.e., ratings above the maximum or below the minimum values on the scale were refused). Every participant completed two or three feature lists. Twelve different participants rated the importance of every feature.

Results

The average reliability estimate for each of the categories can be found in column 2 of Table 3. Most estimates were above .80, except those for **fish**, **insects**, **mammals**, and **vehicles**. The estimations of reliability for these categories equaled .78, .74, .79, and .78, respectively.

Feature Importance Ratings: Category Feature Description

Since both exemplar and category features have been shown to capture fundamental aspects of internal category structure (Ashcraft, 1978; Hampton, 1979; Rosch & Mervis, 1975), it is worthwhile not only to determine how relevant each of the various exemplar features is, but also to include the same information for the category features. Interestingly, there is a fair amount of overlap between the features generated in response to the category labels and the features generated for the exemplars of the corresponding categories. The only exception to this finding is the category of **professions**, for which none of the category features were generated in response to the category's instances.

Method

Participants. Twelve young professionals with a university degree, between 22 and 25 years of age ($M = 23.6$), participated in the

Table 3
Estimates of Reliability of Feature Importance and Applicability Ratings for Exemplar and Category Features for the 15 Categories

Category	Feature Importance		Applicability	
	Exemplar Features	Category Features	Exemplar Features	Category Features
Birds	.87	.90	.84	.93
Fish	.78	.85	.85	.87
Insects	.74	.73	.76	.77
Mammals	.79	.87	.87	.92
Reptiles	.82	.83	.83	.85
Clothing	.89	.90	.84	.79
Kitchen utensils	.87	.87	.89	.83
Musical instruments	.90	.89	.87	.83
Tools	.88	.85	.87	.82
Vehicles	.78	.78	.88	.87
Weapons	.92	.90	.86	.86
Fruit	.84	.79	.85	.89
Vegetables	.82	.83	.84	.83
Professions	.80	.92	.83	.78
Sports	.81	.85	.88	.88

rating task. Seven of them were male; 5 of them were female. All of them were paid the equivalent of $10/h for their participation.

Materials. The features that were to be rated were generated in response to the category labels of the 15 categories whose exemplars served as cues too. The category features with a generation frequency of at least 2 were withheld for this task. (See column 7 of Table 2 for the exact number of withheld category features per category.) If they had already been rated in the previous task, the features were not included in the additional judgment task.

Procedure. The procedure was similar to the feature-importance-rating task for exemplar features described above, except for a few changes to the instructions. The participants were informed that they would see a list of features that had been generated by different participants in response to a semantic category label and were asked to rate the importance or relevance of these features for the category. The participants were presented with a feature list for each of the 15 semantic categories. Every participant completed all 15 feature lists.

Results

The average reliability estimates can be found in column 3 of Table 3. Most estimates were above .80, except those for *insects*, *vehicles*, and *fruit*. The estimations of reliability for these categories equaled .73, .78, and .79, respectively.

PART III
Exemplar by Feature Applicability Matrices

A structural data set with information on a category's exemplars and the features that characterize these exemplars is invaluable for many different paradigms in the study of semantic concept representation (Storms & De Boeck, 1997; Van Mechelen et al., 1993). We will elaborate on three applications that are of particular interest.

First, these kinds of structural data sets can be used to derive proximity data. Especially when the stimulus set of interest is quite large and rated pairwise similarities may not be available because of the unpractical large number of stimulus pairs, proximities can be derived from an exemplar by feature matrix by correlating the exemplars' feature vectors. Euclidean distances between the exemplars' feature vectors would constitute another alternative to directly rated similarities. Means of representing similarity, such as multidimensional scaling and clustering techniques, have been shown to be useful in the study of semantic concepts (Henley, 1969; Shoben, 1983). Proximities derived from feature representations can also be used to evaluate the proximities generated from corpus-based approaches—for example, the high-dimensional vector models such as LSA (Landauer & Dumais, 1997) and HAL (Burgess & Lund, 1997). In experimental studies, these kinds of similarity configurations have been used to predict external variables, such as categorization (Smits et al., 2002), response times (Rips et al., 1973; Shoben, 1976), linear separability (Ruts, Storms, & Hampton, 2004), inductive strength (Rips, 1975), and typicality (Ameel & Storms, 2006). These configurations can also be put to use in the study of semantic deficits (Storms, Dirikx, Saerens, Verstraeten, & De Deyn, 2003). In most of these applications, the stimulus sets used have been fairly limited and restricted to a few well-known categories (such as *mammals* or *tools*). Similarity representations of extended stimulus sets that cover the entire domain of animals or artifacts yield many opportunities for further studies of semantic cognition.

Second, connectionist models have been shown to provide good predictions of a broad range of empirical phenomena concerning semantic memory in both normal and abnormal samples. A number of the models employ exemplar by feature matrices of limited size that are said to mimic the characteristics of their empirical counterparts (e.g., Plaut & Shallice, 1993; Rogers et al., 2004; Rogers & McClelland, 2004). Others (Cree & McRae, 2003; Cree et al., 1999; McRae et al., 1997) were built from larger data sets but allowed for the connecting of a feature to an exemplar only if the feature was explicitly generated for the exemplar in the data-gathering procedure. As was argued in the introduction to the feature generation task, this might result in missing connections between exemplars and features. By having participants explicitly judge whether the features that were produced in response to a category's exemplars (or to its label) are true of the category's exemplars, we aim to avoid such lapses. Especially when the categories involved make up broad semantic domains of living (animals) and nonliving (artifacts) things, these data structures will prove useful for many connectionist models, for much of the connectionist modeling endeavors have been aimed at addressing category- or domain-specific deficits through exemplar by feature information (Devlin, Gonnerman, Andersen, & Seidenberg, 1998; Durrant-Peatfield, Tyler, Moss, & Levy, 1997; Greer et al., 2001; Tyler, Moss, Durrant-Peatfield, & Levy, 2000).

Third, the seminal notion of family resemblance (Rosch & Mervis, 1975) is based on exemplar by feature applicability matrices. When Rosch and Mervis introduced their family resemblance notion, they gathered data from six semantic categories, and for each category, they selected a set of 20 representative exemplars. First, they asked a group of participants to generate features for all 20 of the exemplars. Then they merged the features and asked another group of participants to judge the applicability of each of the generated features (regardless of the exemplar for which it was produced) for each of the 20 exemplars. Next, they weighted the features on the basis of the number of exemplars that they characterized. Rosch and Mervis showed that a weighted sum of the features true of an exemplar yields a good prediction of the exemplar's typicality in the category.

Rosch and Mervis (1975) used from 51 to 123 different features for their family resemblance prediction. Although these data have been circulating informally among researchers in the field, they have never been published. Also, the number of exemplars on which the selected feature set was based was rather limited, and only six categories were investigated. Following Rosch and Mervis, several authors have replicated the predictive power of the family resemblance measure, but usually the number of features used in calculating the measure was even smaller (e.g., Malt & Smith, 1984; Maridaki-Kassotaki, 1997). Hampton (1979) has pointed out that a similar measure can be obtained through category features, instead of exemplar features. This method, as well, yields good predictions of exemplar typicality.

As was explained in the section on feature frequencies described above, both types of features were available for the concepts studied here: *exemplar features* that were selected from a feature generation task with the exemplars (e.g., *sparrow*, *dog*, *suit*, *car*, etc.) as cues and *category features* that were selected from a feature generation task with the category labels (e.g., *birds*, *mammals*, *clothing*, *vehicles*, etc.) as cues. These features were used to construct four different types of exemplar by feature matrices.

The first type of matrix (Type I) is defined at the level of the 15 categories whose exemplars served as cues in the feature generation task. Thus, there are 15 matrices of this first type, each constructed by taking all exemplars of a category (e.g., the 30 **mammal** exemplars), on the one hand, and the most frequently generated exemplar features of the particular category, on the other. The feature selection criterion for the Type I matrices was identical to that for the selection of features for the feature-importance-rating task with exemplar features. Only features that were generated at least four times across all category members were included. For the **mammals** category, for instance, this resulted in a matrix holding all 30 exemplars and the 288 most frequently generated exemplar features.

The second type of matrix (Type II) consists of all exemplars of a semantic domain, on the one hand, and those exemplar features generated most frequently across the exemplars of the domain, on the other. The exemplar features included in the Type II matrices were obtained by collapsing across the features of the relevant Type I matrices. The exemplar features included in the **birds**, **fish**, **insects**, **mammals**, and **reptiles** Type I matrices make up the features in the Type II animal matrix. The exemplar features included in the **clothing**, **kitchen utensils**, **musical instruments**, **tools**, **vehicles**, and **weapons** Type I matrices make up the features in the Type II artifact matrix. Features appearing in more than one of the Type I matrices of a semantic domain were included only once in the Type II matrix of the semantic domain. Feature frequencies for such features were adjusted accordingly. In other words, there are two different matrices of this second type: an animal matrix that consists of 129 animal exemplars and 765 animal features and an artifact matrix that consists of 166 artifact exemplars and 1,295 artifact features. The five animal and six artifact matrices of the first type are thus submatrices of those of the second type. Consequently, researchers who are interested in data for the animal or artifact categories (like **mammals** or **weapons**), rather than in those for the complete semantic domains (animals or artifacts), can use the matrices of the first type *and* select the relevant data from the matrices of the second type.

Matrices of the third type (Type III) are similar to those of the first type but contain *category features*, instead of exemplar features. They are defined at the level of the 15 categories for which Type I matrices are available and are constructed by taking all exemplars of a category, on the one hand, and all selected category features, on the other. The feature selection criterion for Type III matrices was identical to that for the selection of features for the feature-importance-rating task with category features: Only features generated at least twice were retained.

Finally, there are two matrices of the fourth type (Type IV). They are similar to the Type II matrices but contain category features, instead of exemplar features. One Type IV matrix has been constructed for the animal domain. It holds 129 animal exemplars, as well as all category features generated in response to the category labels *birds*, *fish*, *insects*, *mammals*, and *reptiles*. Another Type IV matrix has been constructed for the artifact domain. It holds all 166 artifacts exemplars, as well as all category features generated in response to the category labels *clothing*, *kitchen utensils*, *musical instruments*, *tools*, *vehicles*, and *weapons*. Note that no frequency cutoff was used in the construction of these two matrices, and hence, all generated features were included in the matrix. As a result, the five animal and six artifact Type III matrices can be said to be subsets of the Type IV matrices. Consequently, researchers who are interested in data for the animal or artifact categories (such as **mammals** or **weapons**), rather than in those for the complete semantic domains (animals or artifacts), can use the matrices of the third type and select the relevant data from the matrices of the fourth type. A schematic overview of the various exemplar by feature matrices is shown in Figure 1.

Type I
Exemplar-by-Feature Matrices for
15 Categories Based on Exemplar Features

Method

Participants. Fifty-one students and young adults with a university degree participated in the feature applicability judgment task. All the participants were paid the equivalent of $10/h.

Materials. Fifteen exemplar by feature matrices were constructed corresponding to the same 15 categories that were included in the feature importance tasks and in various tasks described in the first part of the article. The columns of the matrices corresponded to the category exemplars (varying in number from 20 to 33), and the rows of the matrices corresponded to the selected exemplar features for these categories (varying in number from 156 for **fish** to 382 for **sports**; see the fourth column of Table 2). The exemplars *Komodo dragon* and *iguanodon* were omitted from the **reptiles** category, since they proved too unfamiliar for the participants judging the feature–exemplar applicability.

Procedure. The participants were instructed to judge, for every feature–exemplar pair, whether the feature characterized the exemplar or not and, accordingly, to write down a 1 or a 0 in the corresponding matrix entry. They performed the task at home and worked on it whenever it suited them. They were given the choice to work on the task row-wise or column-wise, but they were asked not to pause until a row or column was finished. Every matrix was completed by 4 different participants, and every participant completed one, two, or three different matrices. Completing the task took between 2 and 3 h of work per matrix.

Results

The reliability of the exemplar by feature matrices was again evaluated by applying the Spearman–Brown formula to the split-half correlations. There are only three different ways to divide four subjects into two half groups. The three possible estimations of the reliability of the matrices were averaged. These averages can be found in the fourth column of Table 3. All of the estimations were well above .80, except the one for **insects**, which only reached .76.

Animals

Birds	Insects	Fish	Mammals	Reptiles T. III	T. IV
				viper	
				crocodile	
				toad, . . . T. I	T. II

Type	Feature Examples	Description	Type	Feature Examples	Description
T. I	does not fly male is rooster eggs are eaten, . . .	5 matrices based on <u>exemplar</u> features (e.g., *birds*)	T. II	can fly moves has blood, . . .	1 animal matrix based on exemplar features (aggregated across all domain exemplars)
T. III	can fly have wings are animals, . . .	5 matrices based on <u>category</u> features (e.g., *chicken*)	T. IV	has fur builds nest has legs, . . .	1 animal matrix based on category features (aggregated across all 5 category labels)

Artifacts

Clothing	Kitchen Utens.	Music Instr.	Tools	Vehicles	Weapons T. III	T. IV
					hammer	
					anvil	
					drill, . . . T. I	T. II

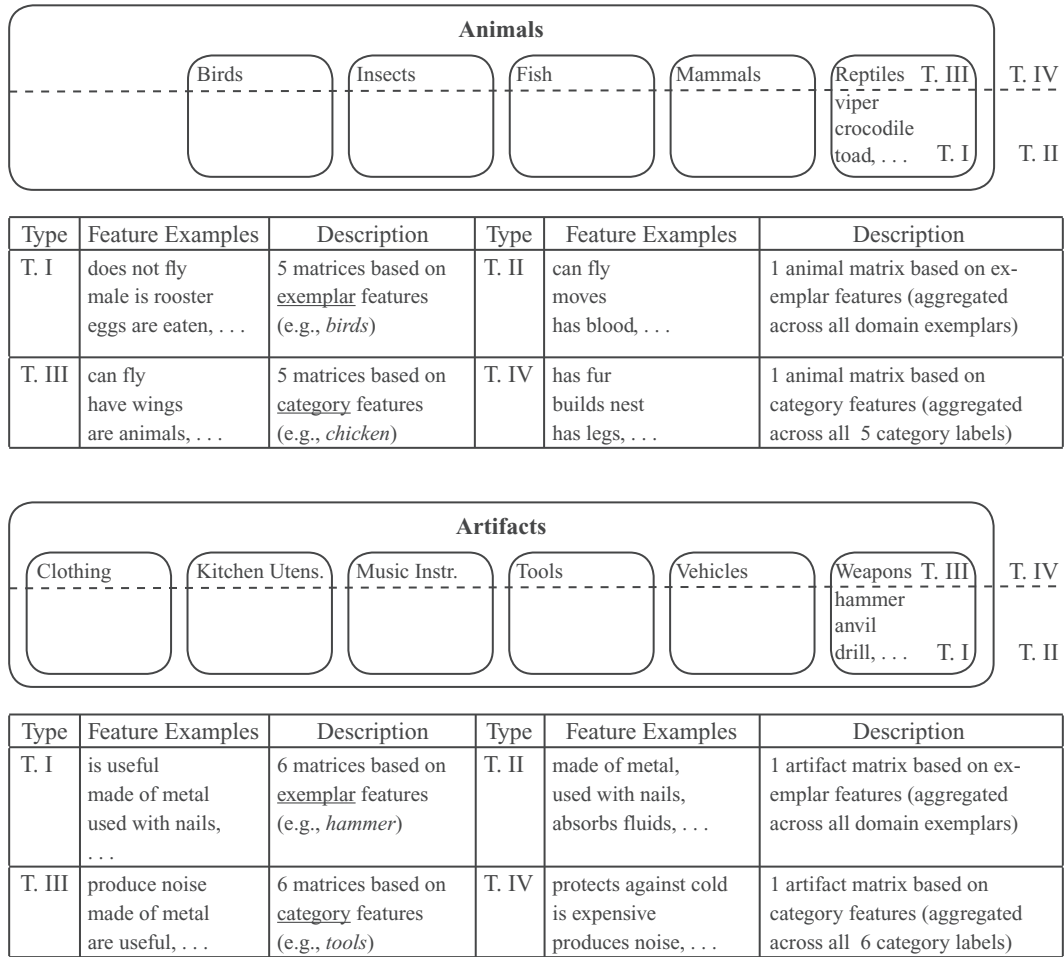| Type	Feature Examples	Description	Type	Feature Examples	Description
T. I	is useful made of metal used with nails, . . .	6 matrices based on <u>exemplar</u> features (e.g., *hammer*)	T. II	made of metal, used with nails, absorbs fluids, . . .	1 artifact matrix based on exemplar features (aggregated across all domain exemplars)
T. III	produce noise made of metal are useful, . . .	6 matrices based on <u>category</u> features (e.g., *tools*)	T. IV	protects against cold is expensive produces noise, . . .	1 artifact matrix based on category features (aggregated across all 6 category labels)

Figure 1. Schematic overview of the different exemplar-by-feature matrices. Type I and Type II matrices contain exemplar features, whereas Type III and Type IV matrices contain category features. Matrices of Types I and III provide information at the level of a single semantic category, whereas matrices of Types II and IV are aggregated across a series of categories to provide information at the level of a semantic domain.

Type II
Exemplar-by-Feature Matrices for the Animal and Artifact Domain Based on Exemplar Features

Method

Participants. Eight volunteers participated in the feature applicability judgment task. Five of the participants were students; the 3 others were young adults with a university degree. All the participants were paid the equivalent of $10/h.

Materials. Two exemplar by feature matrices were constructed: one for the animal domain and one for the artifact domain. The animal matrix contained 129 columns corresponding to animal names that belonged to the categories **birds**, **fish**, **insects**, **mammals**, and **reptiles**. (The animal matrix held 129, instead of the total of 131, animals because the exemplars *Komodo dragon* and *iguanodon* were omitted. They proved too unfamiliar for the participants providing the estimations of applicability.) The rows corresponded to 765 exemplar features of animals. The artifact matrix contained 166 columns corresponding to object names that belong to the categories **clothing**, **kitchen utensils**, **musical instruments**, **tools**, **vehicles**, and **weapons**. The rows of this matrix corresponded to 1,295 exemplar features of artifacts.

Procedure. The procedure of this task was identical to that in the Type I rating task. Four participants completed the animal matrix, and 4 participants completed the artifact matrix. Completing the task took between 40 and 50 h of work.

Results

The reliability of the exemplar by feature matrices was evaluated by applying the Spearman–Brown formula to the split-half correlations. There are only three different ways to divide four subjects into two half groups. The resulting three estimates of the reliability of the animal matrix were .81, .83, and .83. The three estimates of the reliability of the artifact matrix were .80, .80, and .82.

Type III
Exemplar-by-Feature Matrices for the 15 Semantic Categories Based on Category Features

Method

Participants. Four participants completed the feature applicability judgment task. Two of the participants were students; the 2 others

were young adults with a university degree. All the participants were paid the equivalent of $10/h.

Materials. The 15 exemplar by feature matrices were similar to those described for the Type I ratings. The columns corresponded to the category exemplars (varying in number from 20 to 33). In contrast to the Type I matrices, however, the rows corresponded to the selected *category features* for the 15 categories (varying in number from 21 for *professions* to 39 for both *kitchen utensils* and *musical instruments*; see the seventh column of Table 2). As was the case for the corresponding Type I matrix, the exemplars *Komodo dragon* and *iguanodon* were omitted from the *reptiles* Type III matrix.

Procedure. Since the category features partly overlap with the exemplar features, the 15 matrices of Type III overlap partly with the corresponding matrices of Type I. These overlapping parts were copied, and the remaining parts were presented to the participants for further judgment. The procedure was identical to that for the Type I and Type II ratings. Each of the 15 matrices was completed by 4 participants. Completing the task took between 3 and 5 h of work.

Results

The reliability of the exemplar by feature matrices was again evaluated by applying the Spearman–Brown formula to the split-half correlations. There are only three different ways to divide four subjects into two half groups. The three possible estimates of the reliability of the matrices were averaged. These averages can be found in column 5 of Table 3. For three categories, the estimates did not reach .80: *insects* (.77), *clothing* (.79), and *professions* (.78).

Type IV
Exemplar-by-Feature Matrices
for the Animal and Artifact Domains
Based on Category Features

Method

Participants. Four participants completed the feature applicability judgment task. All of them were young adults with a university degree. They were paid the equivalent of $10/h.

Materials. Two exemplar by feature matrices were constructed: one for the animal domain and one for the artifact domain. The animal matrix contained 129 columns corresponding to animal exemplars that belong to the categories *birds*, *fish*, *insects*, *mammals*, and *reptiles*. (As was the case in the Type II animal matrix, the exemplars *Komodo dragon* and *iguanodon* were omitted.) Its rows corresponded to 225 category features that were generated for the five animal category labels (i.e., 52 for *birds*, 58 for *fish*, 70 for *insects*, 60 for *mammals*, and 70 for *reptiles*; but note that some of the features overlapped). The artifact matrix contained 166 columns corresponding to object exemplars that belonged to the categories *clothing*, *kitchen utensils*, *musical instruments*, *tools*, *vehicles*, and *weapons*. The rows of this matrix corresponded to 300 category features that were generated for the six artifact category labels (i.e., 71 for *clothing*, 73 for *kitchen utensils*, 74 for *musical instruments*, 79 for *tools*, 62 for *vehicles*, and 54 for *weapons*; again, with some overlap across the categories).

Procedure. Since the category features partly overlap with the exemplar features, the two Type IV matrices overlap partly with the two matrices of Type II. These overlapping parts were copied, and the remaining parts were presented to the participants for further judgment. The remainder of the procedure was identical to that for the ratings for Types I, II, and III. The participants completed both the animal matrix and the artifact matrix. Completing the task took between 4 and 6 h of work.

Results

The reliability of the exemplar by feature matrices was evaluated by applying the Spearman–Brown formula to the split-half correlations. There are only three different ways to divide four subjects into two half groups. The resulting three estimates of the reliability of the animal matrix were .89, .90, and .87. The three estimates of the reliability of the artifact matrix were .85, .84, and .84.

CONCLUSION

The present data set adds to a long tradition in semantic concept research that has looked at the intension of categories to gain insight into their representation. Traditionally, researchers have relied on feature generation tasks to get at the attributes that might be of importance in establishing the meaning of the concepts under study (Ashcraft, 1978; Garrard et al., 2001; McRae et al., 2005). We have moved beyond simply listing the generated features and their frequencies by (1) providing ratings of the importance of the various features and (2) having participants complete extensive exemplar by feature matrices covering 15 different categories and two large semantic domains. Such data can be employed to validate current (small-scale) modeling efforts addressing the structure of semantic concepts or can inspire new endeavors.

In addition, the normative data include 11 variables that were gathered for the same 15 categories. We expect these normative data to be useful both for the control and selection of stimuli in experimental studies and for theoretical investigations toward a mechanistic explanation of how effects such as age of acquisition, familiarity, word frequency, and imageability come about (e.g., Monaco et al., 2007; Steyvers & Tenenbaum, 2005). Together, these data constitute a unique set that provides ample opportunities for experimental explorations and modeling endeavors aimed at a more elaborate understanding of semantic concept representation.

REFERENCES

ALTARRIBA, J., BAUER, L. M., & BENVENUTO, C. (1999). Concreteness, context availability, and imageability ratings and word associations for abstract, concrete, and emotion words. *Behavior Research Methods, Instruments, & Computers*, **31**, 578-602.

AMEEL, E., & STORMS, G. (2006). From prototypes to caricatures: Geometrical models for concept typicality. *Journal of Memory & Language*, **55**, 402-421.

ASHCRAFT, M. H. (1978). Property dominance and typicality effects in property statement verification. *Journal of Verbal Learning & Verbal Behavior*, **17**, 155-164.

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

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception & Performance*, **10**, 340-357.

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, **11**, 211-227.

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 629-649.

Barton, M. E., & Komatsu, L. K. (1989). Defining features of natural kinds and artifacts. *Journal of Psycholinguistic Research*, **18**, 433-447.

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**(3, Pt. 2), 1-46.

Borg, I., & Groenen, P. J. F. (2005). *Modern multidimensional scaling*. New York: Springer.

Brysbaert, M., Van Wijnendaele, I., & De Deyne, S. (2000). Age-of-acquisition of words is a significant variable in semantic tasks. *Acta Psychologica*, **104**, 215-226.

Burgess, C., & Lund, K. (1997). Modeling parsing constraints with high-dimensional context space. *Language & Cognitive Processes*, **12**, 177-210.

Carroll, J. B., & White, M. N. (1973). Word frequency and age of acquisition as determiners of picture-naming latency. *Quarterly Journal of Experimental Psychology*, **25**, 85-95.

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

Chumbley, J. I. (1986). The roles of typicality, instance dominance, and category dominance in verifying category membership. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 257-267.

Cohen, B. H., Bousfield, W. A., & Whitmarsh, G. A. (1957). *Cultural norms for verbal items in 43 categories* (Tech. Rep. 22). Storrs: University of Connecticut.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning & Verbal Behavior*, **8**, 240-247.

Coltheart, M., Patterson, K., & Marshall, J. (1980). *Deep dyslexia*. London: Routledge & Kegan Paul.

Cree, G. S., & McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of *chipmunk*, *cherry*, *chisel*, *cheese*, and *cello* (and many other such concrete nouns). *Journal of Experimental Psychology: General*, **132**, 163-201.

Cree, G. S., McRae, K., & McNorgan, C. (1999). An attractor model of lexical conceptual processing: Simulating semantic priming. *Cognitive Science*, **23**, 371-414.

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: Norms for 1,424 Dutch words in a continuous task. *Behavior Research Methods*, **40**, 198-205.

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.

Devlin, J. T., Gonnerman, L. M., Andersen, E. S., & Seidenberg, M. S. (1998). Category-specific semantic deficits in focal and widespread brain damage: A computational account. *Journal of Cognitive Neuroscience*, **10**, 77-94.

Diesendruck, G., & Gelman, S. A. (1999). Domain differences in absolute judgments of category membership: Evidence for an essentialist account of categorization. *Psychonomic Bulletin & Review*, **6**, 338-346.

Durrant-Peatfield, M. R., Tyler, L. K., Moss, H. E., & Levy, J. P. (1997). The distinctiveness of form and function in category structure: A connectionist model. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the 19th Annual Conference of the Cognitive Science Society* (pp. 193-198). Mahwah, NJ: Erlbaum.

Estes, Z. (2003). Domain differences in the structure of artifactual and natural categories. *Memory & Cognition*, **31**, 199-214.

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.

Garrard, P., Lambon Ralph, M. A., Hodges, J. R., & Patterson, K. (2001). Prototypicality, distinctiveness, and intercorrelation: Analyses of the semantic attributes of living and nonliving concepts. *Cognitive Neuropsychology*, **18**, 125-174.

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.

Ghyselinck, M., Custers, R., & Brysbaert, M. (2003). Age-of-acquisition ratings for 2332 Dutch words from 49 different semantic categories. *Psychologica Belgica*, **43**, 181-241.

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.

Goldstone, R., & Son, J. Y. (2005). Similarity. In K. Holyoak & R. Morrison (Eds.), *Cambridge handbook of thinking and reasoning* (pp. 13-36). Cambridge: Cambridge University Press.

Greer, M. J., van Casteren, M., McLellan, S. A., Moss, H. E., Rodd, J., Rogers, T. T., & Tyler, L. K. (2001). The emergence of semantic categories from distributed featural representations. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (pp. 358-363). Mahwah, NJ: Erlbaum.

Hahn, U., & Chater, N. (1997). Concepts and similarity. In K. Lamberts & D. Shanks (Eds.), *Knowledge, concepts and categories* (pp. 43-92). Hove, U.K.: Psychology Press.

Hampton, J. A. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning & Verbal Behavior*, **18**, 441-461.

Hampton, J. A. (1981). An investigation of the nature of abstract concepts. *Memory & Cognition*, **9**, 149-156.

Hampton, J. A., Dubois, D., & Yeh, W. (2006). Effects of classification context on categorization in natural categories. *Memory & Cognition*, **34**, 1431-1443.

Hampton, J. A., & Gardiner, M. M. (1983). Measures of internal category structure: A correlational analysis of normative data. *British Journal of Psychology*, **74**, 491-516.

Henley, N. M. (1969). A psychological study of the semantics of animal terms. *Journal of Verbal Learning & Verbal Behavior*, **8**, 176-184.

Holcomb, P. J., Kounios, J., Anderson, J. E., & West, W. C. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: An electrophysiological investigation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 721-742.

Humphreys, G. W., Riddoch, M. J., & Quinlan, P. T. (1988). Cascade processes in picture identification. *Cognitive Neuropsychology*, **5**, 67-103.

Janczura, G. A., & Nelson, D. L. (1999). Concept accessibility as the determinant of typicality judgments. *American Journal of Psychology*, **112**, 1-19.

Keil, F. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.

Kittur, A., Holyoak, K. J., & Hummel, J. E. (2006). Dissociating typicality and goodness judgments: Category-based differences [Abstract]. *Abstracts of the Psychonomic Society 47th Annual Meeting*, **11**, 120.

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling*. Beverly Hills, CA: Sage.

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.

Landauer, T. K., Ross, B. H., & Didner, R. S. (1979). *Processing visually presented single words: A reaction time analysis* [Tech. Memorandum]. Murray Hill, NJ: Bell Laboratories.

Larochelle, S., & Pineau, H. (1994). Determinants of response times in the semantic verification task. *Journal of Memory & Language*, **33**, 796-823.

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.

Lee, M. D., & Navarro, D. J. (2002). Extending the ALCOVE model of category learning to featural stimulus domains. *Psychonomic Bulletin & Review*, **9**, 43-58.

LORCH, R. F. (1978). The role of two types of semantic information in the processing of false sentences. *Journal of Verbal Learning & Verbal Behavior*, **17**, 523-537.

MALT, B. C., & JOHNSON, C. E. (1992). Do artifact concepts have cores? *Journal of Memory & Language*, **31**, 195-217.

MALT, B. C., & SMITH, E. E. (1984). Correlated properties in natural categories. *Journal of Verbal Learning & Verbal Behavior*, **23**, 250-269.

MARIDAKI-KASSOTAKI, K. (1997). Are rating-based procedures reliable for derivation of typicality judgments from children? *Behavior Research Methods, Instruments, & Computers*, **29**, 376-385.

MARKMAN, A. B., & GENTNER, D. (1993). Splitting the differences: A structural alignment view of similarity. *Journal of Memory & Language*, **32**, 517-535.

MARKMAN, A. B., & WISNIEWSKI, E. J. (1997). Similar and different: The differentiation of basic level categories. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 54-70.

MARTIN, R. C., & CARAMAZZA, A. (1980). Classification in well-defined and ill-defined categories: Evidence for common processing strategies. *Journal of Experimental Psychology: General*, **109**, 320-353.

McCLOSKEY, M. [E.] (1980). The stimulus familiarity problem in semantic memory research. *Journal of Verbal Learning & Verbal Behavior*, **19**, 485-502.

McCLOSKEY, M. E., & GLUCKSBERG, S. (1978). Natural categories: Well defined or fuzzy sets? *Memory & Cognition*, **6**, 462-472.

McCLOSKEY, M. E., & GLUCKSBERG, S. (1979). Decision-processes in verifying category membership statements: Implications for models of semantic memory. *Cognitive Psychology*, **11**, 1-37.

McRAE, K., & CREE, G. S. (2002). Factors underlying category-specific semantic deficits. In E. M. E. Forde & G. Humphreys (Eds.), *Category specificity in brain and mind* (pp. 211-249). Hove, U.K.: Psychology Press.

McRAE, K., CREE, G. S., SEIDENBERG, M. S., & McNORGAN, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, **37**, 547-559.

McRAE, K., DE SA, V. R., & SEIDENBERG, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General*, **126**, 99-130.

MEDIN, D. L., & ATRAN, S. (2004). The native mind: Biological categorization and reasoning in development and across cultures. *Psychological Review*, **111**, 960-983.

MERVIS, C. B., CATLIN, J., & ROSCH, E. (1976). Relationships among goodness-of-example, category norms, and word frequency. *Bulletin of the Psychonomic Society*, **7**, 283-284.

MERVIS, C. B., & ROSCH, E. (1981). Categorization of natural objects. *Annual Review of Psychology*, **32**, 89-115.

MONACO, J. D., ABBOTT, L. F., & KAHANA, M. J. (2007). Lexico-semantic structure and the word-frequency effect in recognition memory. *Learning & Memory*, **14**, 204-213.

MONSELL, S. (1991). The nature and locus of word frequency effects in reading. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 148-197). Hillsdale, NJ: Erlbaum.

MONSELL, S., DOYLE, M. C., & HAGGARD, P. N. (1989). Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General*, **118**, 43-71.

MORRISON, C. M., & ELLIS, A. W. (2000). Real age of acquisition effects in word naming and lexical decision. *British Journal of Psychology*, **91**, 167-180.

MURPHY, G. L. (1982). Cue validity and levels of categorization. *Psychological Bulletin*, **91**, 174-177.

PAIVIO, A. U. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart & Winston.

PAIVIO, A. U. (1986). *Mental representations: A dual coding approach*. New York: Oxford University Press.

PLAUT, D. C., & SHALLICE, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology*, **10**, 377-500.

PRUZANSKY, S., TVERSKY, A., & CARROLL, J. D. (1982). Spatial versus tree representations of proximity data. *Psychometrika*, **47**, 3-24.

QUINE, W. V. (1969). *Ontological relativity and other essays*. New York: Columbia University Press.

REILLY, J., & KEAN, J. (2007). Formal distinctiveness of high- and low-imageability nouns: Analyses and theoretical implications. *Cognitive Science*, **31**, 157-168.

RIPS, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning & Verbal Behavior*, **14**, 665-681.

RIPS, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 21-59). New York: Cambridge University Press.

RIPS, L. J., SHOBEN, E. J., & SMITH, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning & Verbal Behavior*, **12**, 1-20.

ROGERS, T. T., LAMBON RALPH, M. A., GARRARD, P., BOZEAT, S., McCLELLAND, J. L., HODGES, J. R., & PATTERSON, K. (2004). The structure and deterioration of semantic memory: A neuropsychological and computational investigation. *Psychological Review*, **111**, 205-235.

ROGERS, T. T., & McCLELLAND, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.

ROSCH, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, **104**, 192-233.

ROSCH, E. (1977). Human categorization. In N. Warren (Ed.), *Studies in cross-cultural psychology* (Vol. 1, pp. 1-49). London: Academic Press.

ROSCH, E., & MERVIS, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, **7**, 573-605.

ROSCH, E., MERVIS, C. B., GREY, W. D., JOHNSON, D. M., & BOYES-BRAEM, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, **8**, 382-439.

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.

RUTS, W., STORMS, G., & HAMPTON, J. (2004). Linear separability in superordinate natural language concepts. *Memory & Cognition*, **32**, 83-95.

SATTATH, S., & TVERSKY, A. (1977). Additive similarity trees. *Psychometrika*, **42**, 319-345.

SCHWANENFLUGEL, P. J., AKIN, C., & LUH, W.-M. (1992). Context availability and the recall of abstract and concrete words. *Memory & Cognition*, **20**, 96-104.

SHEPARD, R. N., & ARABIE, P. (1979). Additive clustering: Representation of similarities as combinations of discrete overlapping properties. *Psychological Review*, **86**, 87-123.

SHOBEN, E. J. (1976). The verification of semantic relations in a same–different paradigm: An asymmetry in semantic memory. *Journal of Verbal Learning & Verbal Behavior*, **15**, 365-379.

SHOBEN, E. J. (1983). Applications of multidimensional scaling in cognitive psychology. *Applied Psychological Measurement*, **7**, 473-490.

SLOMAN, S. A., LOVE, B. C., & AHN, W. K. (1998). Feature centrality and conceptual coherence. *Cognitive Science*, **22**, 189-228.

SMITH, E. E., SHOBEN, E. J., & RIPS, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, **81**, 214-241.

SMITH, J. D., & MINDA, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 1411-1436.

SMITS, T., STORMS, G., ROSSEEL, Y., & DE BOECK, P. (2002). Fruits and vegetables categorized: An application of the generalized context model. *Psychonomic Bulletin & Review*, **9**, 836-844.

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.

STORMS, G., & DE BOECK, P. (1997). Formal models for intracategorical structure that can be used for data-analysis. In K. Lamberts & D. Shanks (Eds.), *Knowledge, concepts, and categories* (pp. 439-459). London: UCL Press.

STORMS, G., DE BOECK, P., & RUTS, W. (2000). Prototype and exemplar based information in natural language categories. *Journal of Memory & Language*, **42**, 51-73.

STORMS, G., DE BOECK, P., VAN MECHELEN, I., & GEERAERTS, D. (1993). Dominance and noncommutativity effects in concept con-

junctions: Extensional or intensional basis? *Memory & Cognition*, **21**, 752-762.

STORMS, G., DIRIKX, T., SAERENS, J., VERSTRAETEN, S., & DE DEYN, P. P. (2003). On the use of scaling and clustering in the study of semantic disruptions. *Neuropsychology*, **17**, 289-301.

STORMS, G., VAN MECHELEN, I., & DE BOECK, P. (1994). Structural analysis of the intension and extension of semantic concepts. *European Journal of Cognitive Psychology*, **6**, 43-75.

SUTCLIFFE, J. P. (1993). Concepts, class, and category in the tradition of Aristotle. In I. Van Mechelen, J. A. Hampton, R. S. Michalski, & P. Theuns (Eds.), *Categories and concepts: Theoretical views and inductive data analysis* (pp. 35-65). London: Academic Press.

TENENBAUM, J. B. (1996). Learning the structure of similarity. In D. S. Touretzky, M. C. Mozer, & M. E. Hasselmo (Eds.), *Neural information processing systems* (Vol. 8, pp. 3-9). Cambridge, MA: MIT Press.

TVERSKY, B., & HEMENWAY, K. (1984). Objects, parts, and categories. *Journal of Experimental Psychology: General*, **113**, 169-193.

TYLER, L. K., MOSS, H. E., DURRANT-PEATFIELD, M. R., & LEVY, J. P. (2000). Conceptual structure and the structure of concepts: A distributed account of category-specific deficits. *Brain & Language*, **75**, 195-231.

VAN LOON-VERVOORN, W. A. (1989). *Eigenschappen van basiswoorden*. Lisse, The Netherlands: Swets & Zeitlinger.

VAN MECHELEN, I., HAMPTON, J. A., MICHALSKI, R. S., & THEUNS, P. (1993). *Categories and concepts: Theoretical views and inductive data analysis*. London: Academic Press.

VANOVERBERGHE, V., & STORMS, G. (2003). Feature importance in feature generation and typicality rating. *European Journal of Cognitive Psychology*, **15**, 1-18.

VERBEEMEN, T., VANPAEMEL, W., PATTYN, S., STORMS, G., & VERGUTS, T. (2007). Beyond exemplars and prototypes as memory representations of natural concepts: A clustering approach. *Journal of Memory & Language*, **56**, 537-554.

VERHEYEN, S., AMEEL, E., & STORMS, G. (2007). Determining the dimensionality in spatial representations of semantic concepts. *Behavior Research Methods*, **39**, 427-438.

WARRINGTON, E. K., & SHALLICE, T. (1984). Category specific semantic impairments. *Brain*, **107**, 829-854.

ZANNINO, G. D., PERRI, R., PASQUALETTI, P., CALTAGIRONE, C., & CARLESIMO, G. A. (2006). Analysis of the semantic representations of living and nonliving concepts: A normative study. *Cognitive Neuropsychology*, **23**, 515-540.

ZANNINO, G. D., PERRI, R., PASQUALETTI, P., DI PAOLA, M., CALTAGIRONE, C., & CARLESIMO, G. A. (2006). The role of semantic distance in category-specific impairments for living things: Evidence from a case of semantic dementia. *Neuropsychologia*, **44**, 1017-1028.

NOTE

1. The data described in this article are available for download from the Psychonomic Society's Archive of Norms, Stimuli, and Data (www.psychonomic.org/archive) as Excel or raw text files. In addition, the data are also available at the authors' Web site (http://ppw.kuleuven.be/concat).

ARCHIVED MATERIALS

The following materials 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 (*Behavior Research Methods*), the first author's name (De Deyne), and the publication year (2008).

FILE: De Deyne-BRM-2008.zip
DESCRIPTION: The compressed archive file contains two directories:
1. cvsdata, containing comma-separated value text files
2. exceldata, containing Microsoft Excel 2003 files

These directories each contain three subdirectories:
1. Exemplar judgments
2. Pairwise similarities
3. Exemplar feature judgments

These subdirectories contain the following files:
1. Exemplar judgments
[exemplarTypicalityRatings], containing exemplar typicality ratings
[exemplarGoodnessRatings], containing exemplar goodness ratings
[exemplarGoodnessRatingsRankOrder], containing exemplar goodness rank order ratings
[exemplarFamiliarityRatings], containing exemplar familiarity ratings
[exemplarImageabilityRatings], containing exemplar imageability ratings
[exemplarWFAoAPercKnownRatings], containing exemplar word frequency age of acquisition and percentage known ratings
[exemplarGenerationFreqAssociationFreq], containing exemplar generation and association frequencies

2. Pairwise similarity, containing 15 pairwise similarity matrices [one for each category; 15 × pairwiseSimilarities(*categoryname*)]

3. Exemplar feature judgments, containing:
[exemplarFeatureImportanceRatings.xls], containing exemplar feature importance ratings
[categoryFeatureImportanceRatings], containing category label feature importance ratings
[TypeI_ExemplarFeatureMatrices/exemplar(*categoryname*)Diagonal Matrices], containing the Type I exemplar feature matrices (including generation frequencies)
[TypeIIAnimalExemplarFeatureMatrix], containing the Type II animal feature matrix (including generation frequencies)
[TypeIIArtifactsExemplarFeatureMatrix], containing the Type II artifact feature matrix (including generation frequencies)
[TypeIII_CategoryLabelFeatureMatrices/categoryLabel(*categoryname*)DiagonalMatrices], containing the Type III category label feature matrices
[TypeIVAnimalCategoryFeaturesMatrix], containing the Type IV animals category label feature matrix
[TypeIVArtifactsCategoryFeaturesMatrix], containing the Type IV artifacts category label feature matrix.

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

APPENDIX
Overview of Categories, Together With the Number and the Translations of Selected Exemplars

English (Dutch) Category	No.	Translated Exemplars in Alphabetical Order
Amphibians (amfibieën)	5	crocodile (krokodil), frog (kikker), salamander (salamander), toad (pad), tortoise (schildpad)
Birds (vogels)	30	blackbird (merel), canary (kanarie), chickadee (mees), chicken (kip), crow (kraai), cuckoo (koekoek), dove (duif), duck (eend), eagle (arend), falcon (valk), heron (reiger), magpie (ekster), ostrich (struisvogel), owl (uil), parakeet (parkiet), parrot (papegaai), peacock (pauw), pelican (pelikaan), penguin (pinguïn), pheasant (fazant), robin (roodborstje), rooster (haan), seagull (meeuw), sparrow (mus), stork (ooievaar), swallow (zwaluw), swan (zwaan), turkey (kalkoen), vulture (gier), woodpecker (specht)
Fish (vissen)	23	anchovy (ansjovis), carp (karper), cod (kabeljauw), dolphin (dolfijn), eel (paling), flatfish (platvis), goldfish (goudvis), herring (haring), orca (orka), pike (snoek), piranha (piranha), plaice (pladijs), ray (rog), salmon (zalm), sardine (sardine), shark (haai), sole (tong), sperm whale (potvis), squid (inktvis), stickleback (stekelbaars), swordfish (zwaardvis), trout (forel), whale (walvis)
Insects (insecten)	26	bee (bij), ant (mier), beetle (kever), bumblebee (hommel), butterfly (vlinder), caterpillar (rups), centipede (duizendpoot), cockchafer (meikever), cockroach (kakkerlak), cricket (krekel), dragonfly (libel), earwig (oorworm), flee (vlo), fly (vlieg), fruit fly (fruitvlieg), grasshopper (sprinkhaan), horsefly (daas), ladybug (lieveheersbeestje), leech (bloedzuiger), louse (luis), mosquito (mug), moth (mot), spider (spin), wasp (wesp), wood louse (pissebed), worm (worm)
Mammals (zoogdieren)	30	bat (vleermuis), beaver (bever), bison (bizon), cat (kat), cow (koe), deer (hert), dog (hond), donkey (ezel), dromedary (dromedaris), elephant (olifant), fox (vos), giraffe (giraf), hamster (hamster), hedgehog (egel), hippopotamus (nijlpaard), horse (paard), kangaroo (kangoeroe), lion (leeuw), llama (lama), monkey (aap), mouse (muis), pig (varken), polar bear (ijsbeer), rabbit (konijn), rhinoceros (neushoorn), sheep (schaap), squirrel (eekhoorn), tiger (tijger), wolf (wolf), zebra (zebra)
Reptiles (reptielen)	22	alligator (alligator), blindworm (hazelworm), boa (boa), caiman (kaaiman), chameleon (kameleon), cobra (cobra), crocodile (krokodil), dinosaur (dinosaurus), frog (kikker), gecko (gekko), iguana (leguaan), iguanodon (iguanodon), Komodo dragon (komodo), lizard (hagedis), monitor lizard (varaan), python (python), salamander (salamander), snake (slang), toad (pad), tortoise (schildpad), turtle (waterschildpad), viper (adder)
Clothing (kleding)	29	bathing suit (zwempak), beanie (muts), belt (riem), blouse (bloes), boots (laarzen), bra (beha), cap (pet), coat (jas), dress (kleed), dungarees (salopet), hat (hoed), jeans (jeans), mittens (wanten), panties (slip), pants (broek), pullover (trui), pyjamas (pyjama), scarf (sjaal), shirt (hemd), shoes (schoenen), shorts (short), skirt (rok), socks (sokken), suit (kostuum), sweater (sweater), tie (das), top (topje), tracksuit (jogging), t-shirt (t-shirt)
Kitchen utensils (keukengerief)	33	apron (schort), bottle (fles), bowl (kom), can opener (blikopener), colander (vergiet), electric kettle (waterkoker), fork (vork), fridge (frigo), glass (glas), grater (rasp), kettle (ketel), knife (mes), microwave oven (microgolf), mixer (mixer), mug (tas), nutcracker (notenkraker), oven (oven), pan (pan), pot (pot), percolator (koffiezet), place mat (onderlegger), plate (bord), scales (weegschaal), scissors (schaar), sieve (zeef), spatula (spatel), spoon (lepel), stove (fornuis), teaspoon (theelepel), toaster (broodrooster), towel (handdoek), whisk (klopper), wok (wok)
Musical instruments (muziekinstrumenten)	27	accordion (accordeon), bagpipe (doedelzak), banjo (banjo), bass guitar (basgitaar), bassoon (fagot), cello (cello), clarinet (klarinet), cymbals (cimbalen), double bass (contrabas), drum (trommel), drum set (drumstel), flute (dwarsfluit), guitar (gitaar), harmonica (mondharmonica), harp (harp), harpsichord (klavecimbel), organ (orgel), pan flute (panfluit), piano (piano), recorder (fluit), saxophone (saxofoon), synthesizer (synthesizer), tambourine (tamboerijn), triangle (triangel), trombone (trombone), trumpet (trompet), violin (viool)

APPENDIX (Continued)

English (Dutch) Category	No.	Translated Exemplars in Alphabetical Order
Tools (werktuigen)	30	adjustable spanner (engelse sleutel), anvil (aambeeld), axe (bijl), chisel (beitel), clamp (houtklem), crowbar (breekijzer), crowbar (koevoet), drill (boormachine), file (vijl), filling knife (plamuurmes), grinding disc (slijpschijf), hammer (hamer), knife (mes), lawn mower (grasmachine), level (waterpas), nail (spijker), oil can (oliespuit), paint brush (verfborstel), pick-axe (houweel), plane (schaaf), plough (ploeg), rope (touw), saw (zaag), screwdriver (schroevendraaier), shovel (schop), tongs (tang), vacuum cleaner (stofzuiger), wheelbarrow (kruiwagen), wire brush (staalborstel), wrench (moersleutel)
Vehicles (voertuigen)	30	airplane (vliegtuig), bicycle (fiets), boat (boot), bus (bus), car (auto), carriage (koets), cart (kar), go-cart (gocart), helicopter (helikopter), hot air balloon (luchtballon), hovercraft (hovercraft), jeep (jeep), kick scooter (step), motorbike (brommer), motorbike (moto), rocket (raket), scooter (bromfiets), skateboard (skateboard), sled (slede), submarine (duikboot), subway train (metrostel), taxi (taxi), tractor (tractor), trailer (caravan), train (trein), tram (tram), truck (camion), truck (vrachtwagen), van (camionette), Zeppelin (zeppelin)
Weapons (wapens)	20	axe (bijl), bazooka (bazooka), bow (boog), cannon (kanon), club (knuppel), dagger (dolk), double barreled shotgun (tweeloop), grenade (granaat), knuckle dusters (boksbeugel), machine gun (mitraillette), pistol (pistool), rifle (geweer), rope (touw), shield (schild), slingshot (katapult), spear (speer), stick (stok), sword (zwaard), tank (tank), whip (zweep)
Fruit (fruit)	30	apple (appel), apricot (abrikoos), banana (banaan), blackberry (braambessen), blueberry (bosbes), cherry (kers), coconut (kokosnoot), dates (dadels), fig (vijg), grape (druif), grapefruit (pompelmoes), kiwi (kiwi), lemon (citroen), lime (limoen), lychee (lychee), mandarine (clementine), mango (mango), melon (meloen), nectarine (nectarine), orange (sinaasappel), papaya (papaya), passion fruit (passievrucht), peach (perzik), pear (peer), pineapple (ananas), plum (pruim), pumpkin (pompoen), raspberry (framboos), red currant (aalbes), strawberry (aardbei)
Vegetables (groenten)	30	asparagus (asperges), beans (bonen), beet (biet), black salsify (schorseneren), Brussels sprouts (spruiten), carrot (wortel), cauliflower (bloemkool), celery (selder), chervil (kervel), corn (maïs), cucumber (komkommer), eggplant (aubergine), endive (witloof), garlic (look), gherkins (augurken), leek (prei), lettuce (sla), mushrooms (champignons), onions (ajuin), parsley (peterselie), peas (erwten), pepper (paprika), potato (aardappel), radish (radijs), red cabbage (rode kool), spinach (spinazie), tomato (tomaat), white cabbage (witte kool), zucchini (courgette), water cress (waterkers)
Professions (beroepen)	30	accountant (boekhouder), actor (acteur), archaeologist (archeoloog), architect (architect), baker (bakker), butcher (slager), cook (kok), dentist (tandarts), doctor (dokter), educator (opvoedster), fireman (brandweerman), garbage collector (vuilnisman), information scientist (informaticus), judge (rechter), lawyer (advocaat), manager (directeur), minister (minister), pharmacist (apotheker), physiotherapist (kinesist), pilot (piloot), plumber (loodgieter), police officer (politieagent), postman (postbode), psychologist (psycholoog), secretary (secretaresse), shop assistant (winkelbediende), stallholder (marktkramer), stewardess (stewardess), teacher (leraar), veterinarian (dierenarts)
Sports (sporten)	30	badminton (badminton), ballet (ballet), baseball (baseball), basketball (basketbal), billiards (biljarten), boxing (boksen), chess (schaken), cycling (wielrennen), fencing (schermen), golfing (golf), gymnastics (turnen), handball (handbal), horseback riding (paardrijden), judo (judo), ice hockey (ijshockey), long jump (verspringen), rugby (rugby), running (hardlopen), sailing (zeilen), shot-put (kogelstoten), skiing (skiën), soccer (voetbal), sport fishing (sportvissen), squash (squash), surfing (surfen), swimming (zwemmen), table tennis (tafeltennis), tennis (tennis), volleyball (volleybal), walking (wandelen)