Dutch norm data for 13 semantic categories and 338 exemplars

WIM RUTS, SIMON DE DEYNE, EEF AMEEL, WOLF VANPAEMEL,
TIMOTHY VERBEEMEN, and GERT STORMS
University of Leuven, Leuven, Belgium

A data set is described that includes eight variables gathered for 13 common superordinate natural language categories and a representative set of 338 exemplars in Dutch. The category set contains 6 animal categories (*reptiles, amphibians, mammals, birds, fish,* and *insects*), 3 artifact categories (*musical instruments, tools,* and *vehicles*), 2 borderline artifact–natural-kind categories (*vegetables* and *fruit*), and 2 activity categories (*sports* and *professions*). In an exemplar and a feature generation task for the category nouns, frequency data were collected. For each of the 13 categories, a representative sample of 5–30 exemplars was selected. For all exemplars, feature generation frequencies, typicality ratings, pairwise similarity ratings, age-of-acquisition ratings, word frequencies, and word associations were gathered. Reliability estimates and some additional measures are presented. The full set of these norms is available in Excel format at the Psychonomic Society Web archive, www.psychonomic. org/archive/.

Since the mid-1970s, different aspects of lexicalized concepts have been studied extensively by cognitive psychologists. A wide range of questions has been addressed regarding common semantic concepts. For instance, extensive research has been conducted on how people categorize unknown items into well-known categories (e.g., Smits, Storms, Rosseel, & De Boeck, 2002; Storms, De Boeck, & Ruts, 2001), what constitutes the basis of typicality within categories (e.g., Hampton, 1979; Rosch & Mervis, 1975), what determines response times in speeded categorization (e.g., Larochelle & Pineau, 1994; Storms, De Boeck, & Ruts, 2000), what the role of features is in the computation of word meaning (e.g., Cree & McRae, 2003), to what extent features are correlated within categories (e.g., Malt & Smith, 1984; McRae, de Sa, & Seidenberg, 1997), and how people verify simple semantic statements, such as *a bird can fly* or *an ostrich is a bird* (Rips, Shoben, & Smith, 1973).

In many of these studies, data have been gathered, such as generation frequencies of exemplars of categories, the typicality of these exemplars, which features are characteristic for the studied categories or their exemplars, and so on. These data have seldom been published, and even if

they are available in print, it is usually in a summary form in appendices. As a result, many aspects of the data, such as the variability in the data or the distribution of the variables, cannot be consulted by interested readers. Nevertheless, such information can be very useful with regard to model fitting. (For an example of the use of variability and skewness of typicality ratings, see De Wilde, Vanoverberghe, Storms, & De Boeck, 2003, and Heit & Barsalou, 1996.) Moreover, in most studies, only a few concept-related variables have been gathered, and variables from different studies usually cannot be related to each other, because the exemplar sets used in the studies were not identical.

The present article describes an elaborate data set that includes eight different variables, gathered in Dutch, for 13 common semantic categories and a representative set of exemplars from these categories. The data were all gathered in a Flemish population. (Note that Dutch and Flemish are basically the same language, much as American and British English are the same language. Both languages differ only slightly in vocabulary.) A first set of six categories are natural kind categories. More specifically, the set consists of six animal categories defined at the superordinate level, which, taken together, almost completely span the entire animal kingdom. These categories were included to create possibilities for studying hierarchically organized semantic categories. The six animal categories are *amphibians*, *birds*, *fish*, *insects*, *mammals*, and *reptiles*. Some of these categories have been used previously in concept-related studies (e.g., Hampton, 1979; Heit & Barsalou, 1996; Osherson, Smith, Wilkie, López, & Shafir, 1990; Rips, 1975; Ruts, Storms, & Hampton, 2004).

Because of the large number of publications that stress the important differences between natural kind categories and artifact categories (Malt & Johnson, 1992; Springer & Keil, 1991), three different categories of the latter kind

were included—namely, *musical instruments*, *tools*, and *vehicles*. *Vehicles* and *tools* have already been used extensively in the study of semantic categories (e.g., Hampton, 1979; McCloskey & Glucksberg, 1978). The *musical instruments* category was included because of its importance in the study of semantic deficits (e.g., Humphreys & Riddoch, 2003).

Fruits and *vegetables* were included because they form a clear pair of contrasting categories (e.g., Storms et al., 2001; Verbeemen, Vanoverberghe, Storms, & Ruts, 2001) and because they have been important in the debates on category-specific deficits (e.g., Humphreys & Forde, 2001). Note that, although *fruits* and *vegetables* are natural kinds in the sense that they grow naturally, they can also be considered artifacts in the sense that they can be manipulated and shaped through cultivation. They differ from the other natural kinds in the degree to which their human use also determines their classification. It is further important to note that neither of these categories is biologically delineated. Finally, two activity categories were included: *sports* and *professions*. Both have been studied in the context of relative clause descriptions of conjunctive concepts (i.e., combinations of semantic concepts related by an *is–a* link; e.g., Storms, De Boeck, Van Mechelen, & Ruts, 1996).

For each of these 13 categories and for a representational sample of exemplars, eight different variables were gathered. These variables are exemplar generation frequencies, typicality ratings, similarity ratings, age-of-acquisition (AoA) ratings, word frequencies, word association norms, feature generation frequencies for category nouns, and feature generation frequencies for exemplars. We describe these variables and motivate the selection of the eight corresponding tasks below. It is obvious that other variables can also affect results in tasks that require word processing, such as morphological family size (e.g., Coltheart, Davelaar, Jonasson, & Besner, 1977), familiarity (e.g., Gernsbacher, 1984), or imageability (Paivio, Yuille, & Madigan, 1968). We hope to add more variables to the database in the future.

EXEMPLAR GENERATION FREQUENCIES

Description

Frequencies of generated exemplars were gathered on a large scale in English for 43 natural categories, defined at different levels of abstraction in the 1950s by Cohen, Bousfield, and Whitmarsh (1957) and for 56 natural categories by Battig and Montague (1969). These norms have been used in numerous studies on natural language categories over the past 35 years. One might, however, question the adequacy of these data for use nowadays. Due to increased mobility and technical developments, our everyday usage of some of the categories and their examplars has changed drastically. For instance, for categories such as *musical instruments* and *vehicles*, exemplars that did not exist in 1969 (such as *synthesizers* and *F16 airplanes*) are now familiar to everyone. Exemplar generation frequencies have also been used as a dependent variable in Storms et al. (2000).

Method

Participants. Five hundred twenty-seven 1st-year psychology students at the University of Leuven participated for course credit. In the present study (November 2003), 196 of them generated exemplars of four categories. Three hundred thirty-one other students completed the exemplar generation task in January 2001 for the remaining nine categories, together with 46 other categories in the category norms study of Storms (2001). None of the participants, either in this task or in the seven other tasks presented in this article, was involved in any other task in this study.

Materials. All the materials in the experiments were collected and presented in Dutch. The closest English equivalents have been used throughout the present text. For four categories, data were collected in 2003: *reptiles*, *amphibians*, *mammals*, and *tools*. Exemplar generation frequency data for the remaining nine categories (*birds*, *fish*, *insects*, *musical instruments*, *vehicles*, *fruits*, and *vegetables*) were taken from the Storms (2001) category norms study.

Procedure. As in the Storms (2001) category norms study, the procedure for generating exemplars for the four new categories in the present study was the same as the procedure in Battig and Montague (1969). For the 2001 study, participants were divided into five groups that completed the task at different moments with a different randomized order of the presented categories, whereas in the present study all 187 participants performed the task in a collective session. In both studies, all the participants received a sheet of paper with the translated instructions from Battig and Montague's study, with enough free space to write down a list. As the instructions were projected on a screen in front of the room, the experimenter read them aloud:

> The purpose of this experiment is to find out what items or objects people commonly give as belonging to various categories or classes. The procedure goes as follows: First I (the experimenter) will read aloud the name or description of a category, which will be projected on the screen. Then you will be given 30 seconds to write down as many items included in that category as you can, in whatever order they happen to occur to you. Start at the upper left of your page and take a new line for every new item. For example, if you were given the category A SEAFOOD, you might respond with such items as *lobster*, *shrimp*, *clam*, *oyster*, and so on.
>
> When you hear the word "Stop," you are to stop writing, turn around your page and go to the beginning of the next page to be ready for the next category. This procedure will be continued through a total of 55 (4) categories. You are to use a different page for each category. You don't have to write down the name of the category. If these instructions are clear to you, then get ready for the first category.

After making sure that all the participants understood the task, the experimenter read the first category name aloud while it was projected on the screen in the front of the room.

Results

To keep the data comparable with those in the Storms (2001) category norms study, the same procedure was used to summarize the collected data for the four new categories. For 120 randomly chosen participants, all legible responses were tallied, including the rank up to the 10th position in each sequence for a given category. If for a category more than 10 exemplars were generated, all responses with a rank of 11 or higher were coded as having a rank of 11. This was done because most participants did not generate more than 10 responses per category and because the number of generated exemplars varied strongly on a word-by-word basis. As in Battig and Montague (1969), synonyms were not merged. No distinction was made between singular and plural or male and female versions of exemplars. Legible responses that were nonmembers were not removed from the list (e.g., *whale* for the

Table 1
Estimates of Reliability of Collected Data Throughout the
Different Tasks in the Present Study

	Estimates of Reliability				
English (Dutch) Category Name	Exemplar Generation Frequency	Typicality Rating	Similarity Rating	Acquisition Rating	Association Frequency
Reptiles (Reptielen)	.99	.93	.84	.98	.98
Mammals (Zoogdieren)	.99	.90	.92	.95	.99
Birds (Vogels)	.98	.98	.89	.85	.97
Fish (Vissen)	.97	.97	.87	.93	>.99
Insects (Insecten)	.98	.92	.91	.95	.99
Musical instruments (Muziekinstrumenten)	.98	.92	.93	.93	>.99
Tools (Werktuigen)	.98	.92	.87	.98	.98
Vehicles (Voertuigen)	.99	.97	.94	.95	.97
Vegetables (Groenten)	.98	.91	.89	.94	>.99
Fruits (Fruit)	.99	.95	.88	.97	.99
Sports (Sporten)	.98	.96	.89	.96	.99
Professions (Beroepen)	.95	.91	.91	.95	.95

category *fish*, or *spider* for the category *insects*). From the Storms category norms study, we learned that using 120 participants yielded sufficiently reliable results. Reliability of the generation frequency was evaluated by applying the split-half method, followed by the Spearman–Brown formula. First, the correlation is computed between the two halves of a test, which are treated as alternate forms (with halves referring to the frequencies for half of the participants). This *halves reliability* estimate (r_{half}) is then adjusted using the Spearman–Brown prediction formula [$2 * r_{half}/(1 + r_{half})$], resulting in an estimation of the reliability of the full test. Table 1 contains estimates of the reliability of the different tasks for each of the categories, except for *amphibians*. For the latter category, separate reliability values are omitted, since this category consisted of only 5 exemplars, and these exemplars were also included in the *reptiles* category. As can be seen in the sec-

ond column of Table 1, estimates of reliability of newly collected data, as well as for the nine categories taken from the Storms study, were well above .98, except for *professions*, with an estimated reliability of .95.

Table 2 shows some of the comparative measures calculated for the 13 categories for the different tasks in the present study. After the first column, containing the category names in English, columns 2–5 include measures for the exemplar generation task. The second column contains the average number of responses participants gave for the category. The third column presents the correlation between the generation frequency and the mean rank position of the exemplars generated by at least three participants. The fourth and fifth columns show the number of different exemplars generated by all participants and the number of different exemplars generated by at least 10% of the participants, respectively.

Table 2
Some Comparative Measures Calculated for the 13 Categories in the Present Study

English Category Name	Exemplar Generation				Associations		No. of Different Category Features Generated
	Mean No. of Responses	Correlation of Frequency and Mean Rank	No. of Different Exemplars		Mean No. of Different Associations Grouped by Category	Mean No. of Different Associations	
			All Participants	Minimum of 10% of Participants			
Reptiles	3.42	−.65	23	7	363	16	55
Amphibians	3.01	−.79	18	7	74	15	22
Mammals	8.35	−.66	64	19	546	18	46
Birds	7.08	−.62	68	21	501	17	53
Fish	5.93	−.52	59	19	371	16	59
Insects	6.01	−.76	54	14	446	17	42
Musical instruments	7.55	−.64	54	16	417	15	40
Tools	6.49	−.52	105	11	589	20	40
Vehicles	7.49	−.66	65	17	651	22	66
Vegetables	7.66	−.74	47	22	430	14	66
Fruits	7.53	−.82	42	17	374	12	55
Sports	7.89	−.48	100	22	657	22	48
Professions	6.80	−.30	150	15	742	25	42

For each of the 13 categories, a representative sample of exemplars was selected in terms of presumed typicality (based on our intuitive notions of the different categories), ranging from very atypical exemplars to very typical exemplars, but with the restriction that all of the selected exemplars had to be familiar to the vast majority of an adult population in the Dutch-speaking part of Belgium. A sample size of 30 was aimed for, but due to the familiarity restriction, the sample size for some of the categories was considerably smaller. Sampling was done out of the com-

plete exemplar set of the present exemplar generation study and the selected categories of the Storms (2001) study. In Table 3, the 13 categories are summarized, together with their selected exemplars. The second column of Table 3 gives the number of exemplars that were included for each of the studied categories. In the generation task for *amphibians*, participants came up with only three biologically correct exemplars. Considering this small number, two wrong exemplars that were generated within the category were included in the sample for this category, re-

Table 3
Category Names Together With the Number and the Translations of Selected Exemplars

English (Dutch) Category Name	Complete Sample of the Studied Exemplars of the Categories	
	No.	Translated Exemplars Rank Ordered According to Generation Frequency
Reptiles (Reptielen)	22	snake, lizard, crocodile, salamander, tortoise, chameleon, iguana, alligator, dinosaur, monitor lizard, frog, viper, gecko, toad, Komodo dragon, caiman, turtle, boa, iguanodon, blindworm, python, and cobra
Amphibians (Amfibieën)	5	frog, salamander, toad, crocodile, and turtle
Mammals (Zoogdieren)	30	dog, cat, horse, cow, elephant, monkey, lion, pig, mouse, giraffe, sheep, donkey, rabbit, tiger, deer, kangaroo, rhinoceros, hippopotamus, zebra, wolf, polar bear, hamster, bat, fox, hedgehog, squirrel, dromedary, bison, beaver, and llama
Birds (Vogels)	30	sparrow, robin, canary, blackbird, dove, magpie, parrot, crow, seagull, eagle, chickadee, woodpecker, parakeet, swallow, chicken, ostrich, stork, duck, owl, vulture, cuckoo, falcon, heron, penguin, peacock, pelican, pheasant, swan, rooster, and turkey
Fish (Vissen)	23	salmon, goldfish, cod, trout, shark, sole, eel, carp, herring, pike, whale, sardine, plaice, dolphin, ray, squid, flatfish, swordfish, stickleback, piranha, orca, sperm whale, and anchovy
Insects (Insecten)	26	fly, mosquito, ant, bee, wasp, spider, beetle, ladybug, butterfly, cockroach, grasshopper, worm, bumblebee, flea, centipede, dragonfly, cricket, caterpillar, louse, horsefly, earwig, maybug, wood louse, moth, fruit fly, and leech
Musical instruments (Muziekinstrumenten)	27	guitar, piano, violin, flute, drum set, trumpet, saxophone, flute, cello, clarinet, harp, contrabass, bass (guitar), drum, harpsichord, trombone, organ, bassoon, panpipe, harmonica, accordion, triangle, synthesizer, cymbals, bagpipe, banjo, and tambourine
Tools (Werktuigen)	30	hammer, saw, chisel, screwdriver, drill, shovel, tongs, knife, axe, plough, wheelbarrow, file, nail, grinding disc, lawn mower, spanner, wrench, crowbar, paintbrush, level, rope, vacuum cleaner, crowbar, pickaxe, anvil, clamp, wire brush, filling-knife, oilcan, and plane
Vehicles (Voertuigen)	30	car, bicycle, bus, train, airplane, tram, boat, moped, truck, motorbike, subway train, scooter, motorbike, helicopter, cart, truck, taxi, carriage, jeep, skateboard, trailer, go-cart, rocket, delivery van, (hot air) balloon, tractor, sled, Zeppelin, hovercraft, and submarine
Vegetables (Groenten)	30	lettuce, carrot, tomato, leek, cucumber, cauliflower, celery, peas, endive, beans, red cabbage, spinach, Brussels sprouts, pepper, white cabbage, asparagus, onions, potato, radish, parsley, eggplant, mushrooms, zucchini, gherkins, black salsify, corn, beet, chervil, garlic, and (water) cress
Fruits (Fruit)	30	apple, pear, banana, kiwi, orange, grape, pineapple, cherry, lemon, mango, litchi, strawberry, peach, melon, passion fruit, grapefruit, lime, apricot, raspberry, nectarine, plum, blackberries, clementine, dates, coconut, pumpkin, papaya, blueberry, fig, and currant
Sports (Sporten)	30	soccer, tennis, volleyball, swimming, basketball, practice gymnastics, judo, golfing, rugby, ballet, skiing, horseback riding, squash, handball, baseball, table tennis, badminton, chess, bicycle racing, sailing, shot-putting, ice hockey, sport fishing, boxing, walking, surfing, doing the long jump, playing billiards, fencing, and running
Professions (Beroepen)	30	psychologist, doctor, teacher, lawyer, baker, dentist, butcher, secretary, garbage collector, shop-assistant, pilot, cook, physiotherapist, manager, judge, pharmacist, fireman, educator, plumber, postman, architect, information scientist, minister, accountant, police officer, actor, veterinarian, stewardess, archaeologist, and stallholder

sulting in a list of only five exemplars. The third column contains all the translated exemplars that were selected, rank ordered according to their generation frequency.

TYPICALITY RATINGS

Description

Typicality has been shown to be an influential variable in a wide variety of cognitive tasks (Hampton, 1993), such as speeded categorization (Larochelle & Pineau, 1994), inductive inference (Rips, 1975), productive tasks (Hampton & Gardiner, 1983), priming effects (Rosch, 1975), semantic substitutability (Rosch, 1977), and memory interference effects (Keller & Kellas, 1978). Typicalities in the context of semantic categories have been explained both by prototype models (Hampton, 1979; Rosch & Mervis, 1975) and by exemplar models (Heit & Barsalou, 1996; Storms et al., 2000).

Method

Participants. Three hundred fourteen participants were involved, 89 of whom were research assistants and last-year psychology students at the University of Leuven, who participated voluntarily. The remaining 225 participants were first-year psychology students at the University of Leuven, who participated for course credit.

Materials. For each of the 13 categories, a list with 5–30 exemplars was selected as described in the introduction (see Table 3 for the exemplars used).

Procedure. All the participants received a sheet of paper with the item set of one category only. For the categories of *fruits* and *vegetables*, 22 participants rated the typicality of the exemplars. Twenty-four participants judged the typicality of the exemplar set of *vehicles*, and 21 the typicality of the exemplar set of *reptiles*. The item sets of the remaining nine categories were each rated by 25 participants. Standard instructions for typicality ratings were printed on top of the page. A 20-point rating scale was used, ranging from 1 for *very atypical or unrelated items* to 20 for *very typical items*. The participants were asked to rate every item in the list and to circle an item if they did not know the word.

Results

Reliability was evaluated by the split-half correlations corrected with the Spearman–Brown formula. The third column of Table 1 shows that the estimates of reliability of typicality ratings were all above .90.

In the second column of Table 4, the percentages of known items per category are given [i.e., $100 * (1 -$ the number of circled items/the total number of ratings)]. Except for *reptiles*, the proportions of known items are very high.

SIMILARITY RATINGS

Description

Ever since the development of multidimensional scaling and clustering techniques (Borg & Groenen, 1997), pairwise proximity measures have been gathered and analyzed frequently. These measures reflect, for each possible pair of items, how similar the items of the pair are. In most studies, they are gathered directly (by simply asking participants to rate how similar the pairs are) or through sort-

Table 4
Percentages of Known Items for the Typicality Ratings, Similarity Ratings, Age of Acquisition Ratings, and Associations

English Category Name	Percentage of Items Known			
	Typicality Rating	Similarity Rating	AoA Rating	Association
Reptiles	90.9	81.7	90.1	92.9
Mammals	100.0	100.0	99.9	99.7
Birds	100.0	100.0	99.6	99.5
Fish	99.6	91.1	99.5	98.8
Insects	99.6	99.5	98.1	98.6
Musical instruments	98.3	97.0	98.0	98.7
Tools	97.7	91.8	96.7	92.7
Vehicles	99.5	99.5	99.0	98.6
Vegetables	99.3	98.1	99.9	98.8
Fruits	99.5	98.1	96.2	97.5
Sports	100.0	100.0	99.2	99.2
Professions	100.0	100.0	99.9	99.6

Note—AoA, age of acquisition.

ing tasks. Recently, alternative similarity measures have been introduced. One approach is based on large text corpora where similarity between words is inferred through statistical regularities between words and the text samples in which they occur (e.g., the LSA model [Landauer & Dumais, 1997] or the HAL model [Burgess & Lund, 2000]). Yet another way to derive a similarity measure is based on free associations (see the Word Association Norms section).

The resulting scaling and clustering representations have been used extensively to study categorization (e.g., Smits et al., 2002), category-based induction (e.g., Rips, 1975), response times (e.g., Rips et al., 1973), and other category-based measures. In fact, it is very difficult to overestimate the role that similarity plays in the study of concepts and categorization. The notion of similarity is at the very core of both exemplar models (Nosofsky, 1984) and prototype models (Smith & Minda, 2000).

Method

Participants. Forty-two participants, mainly students at the University of Leuven, participated in this task. Each was paid the equivalent of about $10/h.

Materials. The similarity ratings were gathered for exemplar pairs of 12 of the 13 categories. The *amphibians* category was omitted, since its 5 exemplars were also included in the *reptiles* category. The 338 exemplars selected are given in Table 3. If a category contains N exemplars, $N * (N - 1)/2$ different exemplar pairs can be formed (excluding the pairs for self-similarity ratings). The large size of our exemplar set (338) makes it practically impossible to gather all direct similarity ratings. Therefore, we collected similarity ratings only for within-category pairs, and not for between-category pairs. The word order in the pairs was randomized. Each participant rated the exemplar pairs of at least two and at most 7 categories (equivalent to between 760 and 3,045 ratings). The only restriction was that the exemplar pairs of the contrast categories *fruits* and *vegetables* were never rated by the same participant. The pairs were mixed over the categories and presented in a random order.

Procedure. The task was done on a computer. The participants were given standard instructions for direct similarity ratings. They were asked 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 had to enter -1. The participants completed the task in between 1 and 5 h (depending on the number of categories, the size of the categories, and the pace of the participant). They never participated longer than 1 h in a single session and always took a break of at least 2 h before continuing.

Results

All exemplar pairs of all 12 categories were rated by at least 14 and by at most 17 different participants. Reliability was evaluated by the split-half correlations corrected with the Spearman–Brown formula. The results can be found in the fourth column of Table 1. All the estimates of reliability for the similarity ratings varied between .84 and .94. In the third column of Table 4, the percentages of known items for each category can be found. The percentages are high or extremely high, except for the *reptiles* category.

AGE-OF-ACQUISITION RATINGS

Description

Since the mid-1980s, AoA has been shown to be an important variable in numerous tasks that require lexical processing. Several studies have demonstrated that words learned earlier have an advantage in speed and/or accuracy in word recognition (e.g., Brown & Watson, 1987; Gilhooly & Logie, 1981) and in lexical decision (Brysbaert, Lange, & Van Wijnendaele, 2000; Gerhand & Barry, 1999). Recently, however, the age at which words are learned has also been shown to affect semantically related use of words. Evidence comes from word associate generation (van Loon-Vervoorn, 1989) and semantic categorization (Brysbaert, Van Wijnendaele, & De Deyne, 2000; Ghyselinck, Custers, & Brysbaert, 2004) data.

AoA ratings by adult participants have been shown to be sufficiently reliable. Gilhooly and Logie (1980) reported an intergroup reliability of .98. For Dutch words, intraclass correlations (Shrout & Fleiss, 1979) of acquisition ratings varied between .96 and .97 (Ghyselinck, Custers, & Brysbaert, 2003). The validity has also been demonstrated. Objective measures for AoA, where children with varying ages were asked to name pictures, correlated .76 with undergraduate student ratings (Morrison, Chappell, & Ellis, 1997). Similar correlation values between an objective measure of AoA for words and student ratings were found for Dutch words (De Moor, Ghyselinck, & Brysbaert, 2000).

Method

The majority of the AoA ratings for the 338 exemplar words were already available from a study by Ghyselinck et al. (2003). To complete the ratings set, we collected norms for 44 words that were not included in the set used by Ghyselinck et al. (2003).

Participants. The participants were 20 students at the University of Leuven between the ages of 17 and 21 years, who volunteered to participate in this study.

Procedure. The 44 words were randomized in four different permutations. The participants received a sheet with instructions and an answer sheet, which consisted of one of four different randomized permutations of the word list. The instructions were identical to those used in Ghyselinck et al. (2003). Each participant was asked to estimate for each word from which age onward they thought they had known the word. If a word was unknown, they were instructed to underline that word. Contrary to the procedure by Ghyselinck et al., (2003) no computer was used. Instead, a pencil-and-paper task was conceived for practical reasons. The task was completed in less than 5 min.

Results

Reliability. The reliability of the ratings was measured using the Spearman–Brown reliability coefficient. Estimates of reliability for the AoA ratings are given in the fifth column of Table 1 and are in the same range as the reliabilities reported in Ghyselinck et al.'s (2003) study. In column 4 of Table 4, the percentages of known items for each category are presented. The percentages are extremely high, except for *reptiles*, for which the percentage is somewhat lower.

WORD FREQUENCIES

Description

Another variable that has frequently been shown to affect many aspects of word processing is the frequency with which words are used (e.g., Forster & Chambers, 1973; Whaley, 1978). Word frequency has been shown to affect a range of tasks, including semantic classification tasks (Landauer, Ross, & Didner, 1979). Despite the fact that word frequency can differ significantly in written and spoken language (Fromkin & Rodman, 1998), word frequency counts are almost invariably based on written material. For word frequency counts in American English, see Kučera and Francis (1967; Francis & Kučera, 1982). The Cobuild corpus is a corpus of British English word frequency counts that is sampled by the CELEX database (Center for Lexical Information, 1993).

Method

The frequency values reported in this study are logarithmic function values of the lemma counts from the Dutch version of the CELEX database (Baayen, Piepenbrock, & van Rijn, 1993). As in most studies using word frequencies, logarithmic function values are used to correct the skew in the frequency distribution. Some aspects have to be taken into consideration when using these frequency measures. They are based on a corpus of 42,380,000 written words that consists mostly of written adult language up to 1993. Word frequencies from nouns of natural kinds, such as *onion* (frequency count of 11) or *mango* (frequency count of 52), might be underestimated due to a number of factors, such as the specific text corpora used. Nevertheless, the norms are readily available and are considered reliable enough for many practical research contexts.

Results

For all categories for which CELEX frequency counts were available, the log values of the word count frequencies are reported in the database.

WORD ASSOCIATION NORMS

Description

In the Similarity Ratings section, different approaches were mentioned for deriving similarity measures. Besides the use of pairwise proximities (gathered directly by ask-

ing, derived from a sorting task, or based on text corpora), another approach uses free association data. In this approach, free association norms are used to derive similarity measures in a high-dimensional representation. These measures have been successful in predicting semantic similarity ratings of words and in free recall and cued recall tasks (Steyvers, Shiffrin, & Nelson, in press).

Despite a long tradition, recent word association databases are hard to find. A recent extensive study for English words was conducted by Nelson, McEvoy, and Schreiber (1998). However, these associations were gathered at different moments during a period of over 20 years. The most recent extensive word association study for the Dutch language is over 15 years old (de Groot, 1988) and used a discrete free association task (i.e., participants generated only one single association for every stimulus word). The selection of words for this study was determined by their word frequency, imageability ratings, and grammatical form class. In both free association studies (de Groot, 1988; Nelson et al., 1998), the data that were gathered included adjectives and other parts of speech, besides nouns and verbs. The association data we gathered for the exemplars of the 13 semantic categories, however, consisted exclusively of nouns, except for the exemplars of the *sports* category, which were made up of verbs as well as nouns.

Whereas the norms described in the studies of Nelson et al. (1998) and de Groot (1988) were all based on single associations per word, we wanted to gather more associations per word from every participant, to allow a larger variability in associations—for instance, in the case of a very strong first association (e.g., *blood* and *red*). A possible consequence of this procedure is that chaining and retrieval inhibition (McEvoy & Nelson, 1982) may influence the results. However, when participants are asked to generate multiple associations, the first association can always be studied separately whenever chaining and retrieval inhibition are considered undesirable. Besides, retrieval inhibition can be expected to be of minimal influence whenever a limited number of associations per participant are gathered.

Method

Participants. Two thousand one hundred individuals participated in the association study. A large majority of this group consisted of 1st-year students at the University of Leuven and at the University of Ghent. The remaining participants were 17-year-old sixth-grade high school students. Some of the students received credit for their participation. The remaining participants volunteered.

Procedure. The association task was administered collectively. The participants received a sheet with instructions written on top. They were instructed to write down, for every stimulus word, the first three associations that entered their mind. If a word was unknown, they were instructed to underline that word, a procedure that was also applied by Nelson et al. (1998) and de Groot (1988). Every participant wrote down associations for, maximally, 14 different words, each belonging to a different semantic category. (Note that, for this task, a few more semantic categories were added. The results for these additional categories are not further described in this article.) To minimize sequence effects, 10 different permutations were constructed, with the order of the categories randomized. Within a permutation, 30 unique lists were created (since the largest

category size was 30). Some lists were shorter than 14 words because of the varying number of exemplars per category. For every word, three different associations were gathered. Completion of the task took less than 5 min.

Results

An average of 76 participants (minimum, 67; maximum, 85) gave three associations for the words in their list. All nonalphanumeric signs and capitalizations were removed. Reliability was evaluated by the split-half correlations corrected with the Spearman–Brown formula and can be found in the last column of Table 1. The reliability was higher than .95 for all categories. For all 338 exemplars, 60% of the different associations were generated only once. The total number of association judgments for all the exemplars that occurred more than once incorporated 87% of the association data.

Furthermore, the average number of different associations for all the exemplars in the categories that occurred more than once is given in Table 2 in column 6. The average number of different associations per exemplar for the categories occurring more than once is presented in column 7. The last column of Table 4 contains the percentages of known items for each category. They are all very high.

FEATURE GENERATION FREQUENCIES FOR CATEGORY NOUNS

Description

In line with the best-known version of prototype theory, in which it is supposed that an abstract summary representation of categories is stored in semantic memory (Hampton, 1993), Hampton (1979) has shown that *characteristic* features of a category succeed in predicting response times in a speeded categorization task, as well as rated typicality. The characteristic features were gathered in a feature generation task in which the participants were asked to name features that define (technically or loosely speaking) a category and features that determine why some exemplars are typical exemplars of the category. Storms et al. (2000) replicated these findings and further showed that measures based on characteristic features can also predict exemplar generation frequencies and category labeling frequencies.

Method

Participants. There was a total of 62 participants; 39 were research assistants and last-year psychology students at the University of Leuven or other colleagues who participated voluntarily. The remaining 23 participants were 1st-year psychology students at the University of Leuven, who participated for course credit.

Materials. For each of the 13 categories presented in Table 3, features were gathered in this task.

Procedure. The participants were asked to write down 10 or more features for one to three unrelated categories. They received written instructions and a sheet of paper containing the category name and free space to write down a list for every category. Instructions contained an example from a different category (*furniture*) of how the task had to be completed and stated explicitly not to generate word associations. Although no time limit was imposed, the feature generation for one category never lasted longer than 5 min. This

procedure had been successfully applied in previous studies, such as Hampton (1979), Rosch and Mervis (1975), Storms et al. (2000), and Storms, Ruts, and Vandenbroucke (1998). For each of the 13 artifact concepts, 10–30 participants generated features. The presentation order of the categories was randomized over participants. None of the participants from the feature generation task participated in any other task in which the same category name was used.

Results

All legible responses were tallied, including the rank up to the 10th position in a sequence for a given category. Responses with a rank of 11 or higher were coded as having a rank of 11. In order to keep the feature norms for category nouns comparable to the feature norms for exemplars (see the next paragraph), we followed a procedure for the coding of both norms that was analogous to that in McRae et al.'s (1997) Experiment 1; synonyms were merged, and no distinction was made between singular and plural or male and female versions of features. Legible responses that were not features of the category were not removed from the list (e.g., *is hermaphroditic* for the category *amphibians*). Column 8 of Table 2 shows the number of different features generated by the participants. For all the concepts, the frequencies and weighted frequencies of the features that were generated at least once are reported in the database.

FEATURE GENERATION FREQUENCIES FOR EXEMPLARS

Description

In contrast with Hampton's (1993) prototype measure mentioned above, the family resemblance measure used by Rosch and Mervis (1975) and by Malt and Smith (1984) is not based on features generated for the category as a whole, but on features of exemplars of the category. Features that characterize words at the same level of abstraction have also been shown to lead to distinctions between artifacts and living things (McRae et al., 1997) and to predict impairment in patients with category-specific deficits (Cree & McRae, 2003).

Method

Participants. In this feature generation task, 640 participants were involved. Two hundred fourteen 1st-year psychology students at the University of Leuven participated for course credit, and 426 other students at the University of Leuven were each paid the equivalent of $4 for their participation.

Materials. For all the selected exemplars of each of the 13 categories, as presented in Table 3, features were gathered in this task.

Procedure. The participants were asked to write down 10 features for 1–10 exemplars belonging to different semantic categories. They received written instructions and a half sheet of paper for every exemplar, containing the exemplar name and free space to write down a list. Instructions explicitly stated that they should generate different sorts of features:

This bundle contains up to 10 sheets with a word on top of the page. We want you to write down 10 features underneath the word. Try to give different sorts of features such as, for example, physical or perceptual features (what does it look like, what sound does it make, how does it smell, how does it taste?), functional features (what is it used for, when and where is it used?), background information (where does it come from, some historical facts), . . .

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 for the feature generation of one exemplar. The presentation order of the exemplars was randomized over participants and over categories. For every category, five different permutations were constructed of the exemplar list. Within a bundle, five permutations of the categories to which the exemplars belonged were used. For every exemplar, 20 different participants generated features.

Results

The results have been entered into a database, which contains about 67,600 features (10 features, generated by 20 participants, for all 338 words). All legible responses were tallied, including the rank up to the 10th position in a sequence for a given exemplar. Responses with a rank of 11 or higher were coded as having a rank of 11. No distinction was made between singular and plural or male and female versions of features. Using a procedure similar to that in McRae et al.'s (1997) Experiment 1, the feature set was interpreted and organized to ensure that synonymous features were recoded as identical within a category, as well as between categories. Furthermore, redundant quantifiers (e.g., *most of them can fly*) were dropped, and both adjective–noun features (e.g., *has brown fur*) and disjunctive features (e.g., *is used in the garden and in the house*) were broken up if the parts provided different information. Legible responses that were actually wrong features for the exemplar (e.g., *is a reptile* as a feature for *salamander*) were not removed from the list. All features that were generated at least once have been translated into English.

CONCLUSION

Many researchers have collected data regarding different aspects of lexicalized concepts (e.g., Cree & McRae, 2003; Rosch & Mervis, 1975; Smits et al., 2002; Storms et al., 2000). Although most of these data contain interesting information regarding the study of semantic information organization, they are often not as valuable as they could be. Why not? First of all, the data have seldom been published. Second, even if they have been published, it has mostly been in summary form. For a researcher who needs, for example, the variability in the data or the distribution of the variables, these summaries are useless. Finally, in most studies, only a few variables have been collected. Even in a best-case scenario in which all the data have been published, the data set would increase enormously in value were it more elaborate. Turning to other studies is no option, since variables from different studies can be related to each other only when the same exemplar sets have been used in the studies. In sum, a data set within the field of semantic concepts and categories is most valuable when it is made available in all its details and when it is elaborate and gathered in a systematic way.

The present article describes a unique collection of data, gathered in Dutch. The collection includes eight variables

for 13 common semantic categories and 338 exemplars from these categories. This elaborate data set is systematically built up and is made available in all its details. This makes us believe that we can offer a highly valuable data set. For all researchers interested in semantic concepts (not only for Dutch-speaking researchers), it gives numerous possibilities for investigating the structure of semantic information. Realizing the importance of an elaborate and systematically built up data set, we hope to get the opportunity to extend the data set with more variables and more data in the future.

REFERENCES

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

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. (1997). *Modern multidimensional scaling: Theory and applications.* New York: Springer-Verlag.

BROWN, G. D. A., & WATSON, F. L. (1987). First in, first out: Word learning age and spoken word frequency as predictors of word familiarity and word naming latency. *Memory & Cognition*, **15**, 208-216.

BRYSBAERT, M., LANGE, M., & VAN WIJNENDAELE, I. (2000). The effects of age-of-acquisition and frequency-of-occurrence in visual word recognition: Further evidence from the Dutch language. *European Journal of Cognitive Psychology*, **12**, 65-85.

BRYSBAERT, M., VAN WIJNENDAELE, I., & DE DEYNE, S. (2000). Age-of-acquisition effects in semantic processing tasks. *Acta Psychologica*, **104**, 215-226.

BURGESS, C., & LUND, K. (2000). The dynamics of meaning in memory. In E. Dietrich & A. B. Markman (Eds.), *Cognitive dynamics: Conceptual and representational change in humans and machines* (pp. 117-156). Mahwah, NJ: Erlbaum.

CENTER FOR LEXICAL INFORMATION (1993). *The Celex lexical database.* Nijmegen: Author.

COHEN, B. H., BOUSFIELD, W. A., & WHITMARSH, G. A. (1957). *Cultural norms for verbal items in 43 categories* (Tech. Rep. 22). University of Connecticut, Contract No. 631(00), Office of Naval Research.

COLTHEART, M., DAVELAAR, E., JONASSON, J. T., & BESNER, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535-555). London: Academic Press.

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.

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

DE MOOR, W., GHYSELINCK, M., & BRYSBAERT, M. (2000). A validation study of the age-of-acquisition norms collected by Ghyselinck, De Moor, & Brysbaert. *Psychologica Belgica*, **40**, 99-114.

DE WILDE, E., VANOVERBERGHE, V., STORMS, G., & DE BOECK, P. (2003). The instantiation principle re-evaluated. *Memory*, **11**, 533-548.

FORSTER, K. I., & CHAMBERS, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning & Verbal Behavior*, **12**, 627-635.

FRANCIS, W. N., & KUČERA, H. (1982). *Frequency analysis of English usage*. Boston: Houghton Mifflin.

FROMKIN, V. A., & RODMAN, R. (1998). *An introduction to language* (6th ed.). New York: Holt, Rinehart & Winston.

GERHAND, S., & BARRY, C. (1999). Age of acquisition, word frequency, and the role of phonology in the lexical decision task. *Memory & Cognition*, **27**, 592-602.

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*, **43-3**, 181-241.

GHYSELINCK, M., CUSTERS, R., & BRYSBAERT, M. (2004). The effect of age of acquisition in visual word processing: Further evidence for the semantic hypothesis. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **30**, 550-554.

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.

GILHOOLY, K. J., & LOGIE, R. H. (1981). Word age-of-acquisition, reading latencies and auditory recognition. *Current Psychological Research*, **1**, 251-262.

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

HAMPTON, J. A. (1993). Prototype models of concept representations. In I. Van Mechelen, J. A. Hampton, R. S. Michalski, & P. Theuns (Eds.), *Categories and concepts: Theoretical views and inductive data analysis* (pp. 67-95). London: Academic Press.

HAMPTON, J. A., & GARDINER, M. M. (1983). Measures of internal category structure: A correlational analysis of normative data. *British Journal of Psychology*, **74**, 498-516.

HEIT, E., & BARSALOU, L. W. (1996). The instantiation principle in natural language categories. *Memory*, **4**, 413-451.

HUMPHREYS, G. W., & FORDE, E. M. E. (2001). Hierarchies, similarity, and interactivity in object recognition: "Category-specific" neuropsychological deficits. *Behavioral & Brain Sciences*, **24**, 453-509.

HUMPHREYS, G. W., & RIDDOCH, M. J. (2003). A case series analysis of "category-specific" deficits in living things: The HIT account. *Cognitive Neuropsychology*, **20**, 263-306.

KELLER, D., & KELLAS, G. (1978). Typicality as a dimension of encoding. *Journal of Experimental Psychology: Human Learning & Memory*, **4**, 78-85.

KUČERA, H., & FRANCIS, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

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* [Technical 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.

MALT, B. C., & JOHNSON, E. C. (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.

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

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

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.

MORRISON, C. M., CHAPPELL, T. D., & ELLIS, A.W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *Quarterly Journal of Experimental Psychology*, **50A**, 528-559.

NELSON, D. L., MCEVOY, C. L., & SCHREIBER, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms.* Retrieved October 13, 2003, from http://w3.usf.edu/Free Association/.

NOSOFSKY, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 104-114.

Osherson, D. N., Smith, E. E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based induction. *Psychological Review*, **97**, 185-200.

Paivio, A., Yuille, J. C., & Madigan, S. (1968). Concreteness, imagery, and meaningfulness values of 925 nouns. *Journal of Experimental Psychology Monographs*, **76** (1, Pt. 2), 1-25.

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

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.

Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology*, **7**, 532-547.

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.

Ruts, W., Storms, G., & Hampton, J. (2004). Linear separability in superordinate natural language concepts. *Memory & Cognition*, **32**, 83-95.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Use in assessing rater reliability. *Psychological Bulletin*, **86**, 420-428.

Smith, J. D., & Minda, J. P. (2000). Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 3-27.

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.

Springer, K., & Keil, F. C. (1991). Early differentiation of causal mechanisms appropriate to biological and non-biological kinds. *Child Development*, **62**, 767-781.

Steyvers, M., Shiffrin, R. M., & Nelson, D. L. (in press). Word association spaces for predicting semantic similarity effects in episodic memory. In A. Healy (Ed.), *Cognitive psychology and its applications: Festschrift in honor of Lyle Bourne, Walter Kintsch, and Thomas Landauer*.

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., & Ruts, W. (2000). Prototype and exemplar based information in natural language categories. *Journal of Memory & Language*, **42**, 51-73.

Storms, G., De Boeck, P., & Ruts, W. (2001). Categorization of novel stimuli in well-known natural concepts: A case study. *Psychonomic Bulletin & Review*, **8**, 377-384.

Storms, G., De Boeck, P., Van Mechelen, I., & Ruts, W. (1996). The dominance effect in concept conjunctions: Generality and interaction aspects. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 1-15.

Storms, G., Ruts, W., & Vandenbroucke, A. (1998). Dominance, overextensions, and the conjunction effect in different syntactic phrasings of concept conjunctions. *European Journal of Cognitive Psychology*, **10**, 337-372.

van Loon-Vervoorn, W. A. (1989). *Eigenschappen van basiswoorden* [Features of basic words]. Lisse: Swets und Zeitlinger.

Verbeemen, T., Vanoverberghe, V., Storms, G., & Ruts, W. (2001). Contrast categories in natural language concepts. *Journal of Memory & Language*, **44**, 1-26.

Whaley, C. P. (1978). Word–nonword classification time. *Journal of Verbal Learning & Verbal Behavior*, **17**, 143-154.

ARCHIVED MATERIALS

The following materials and links may be accessed through the Psychonomic Society's Norms, Stimuli, and Data archive, http://www.psychonomic.org/archive/.

To access these files or links, search the archive for this article using the journal (*Behavior Research Methods, Instruments, & Computers*), the first author's name (Ruts) and the publication year (2004).

File: Ruts-BRMIC-2004.zip

Description: The compressed archive file contains 18 files:

Ex_gen_freq.xls, containing the exemplar generation frequencies for the 13 semantic concepts. In this and the following files, a sheet called "information" contains an explanation about the measures and column headers.

Typ_ratings.xls, containing the typicality ratings for the 13 semantic concepts.

Sim_ ratings.xls, containing the similarity ratings for 12 of the 13 semantic concepts. The *amphibians* category was omitted, since its five exemplars were also included in the reptiles category (the same holds for the following files).

AoA_and_Logfreq.xls, containing the AoA ratings and the word frequency values adopted from the Dutch version of the CELEX database for 12 of the 13 semantic concepts.

FGF_catnouns.xls, containing the feature generation frequencies for 12 of the 13 categories.

12 files containing feature generation frequencies (abbreviated to FGF) for the instantiations of 12 of the 13. Each file is named after the corresponding concept and contains the feature generation frequencies for the selected exemplars of the concept.

Associations.xls, containing the association data for 12 of the 13 concepts.

Author's e-mail address: gert.storms@psy.kuleuven.ac.be.