CrossMark

Applied Psycholinguistics **37** (2016), 1477–1499
doi:10.1017/S0142716416000096

Quantifying semantic animacy: How much are words alive?

JELENA RADANOVIĆ
University of Novi Sad

CHRIS WESTBURY
University of Alberta

PETAR MILIN
University of Novi Sad and Eberhard Karls University Tübingen

Received: December 16, 2014 Accepted for publication: January 13, 2016

ADDRESS FOR CORRESPONDENCE
Jelena Radanović, Department of Psychology, Faculty of Philosophy, Dr Zorana Đinđića 2, Novi Sad 21000, Serbia. E-mail: jelena.radanovic@uns.ac.rs

ABSTRACT
The main goal of this study, which comprised two experimental tasks and three normative studies, was to describe the underlying distribution of semantic animacy, with the focus on Serbian and English. Animacy was measured using three normative techniques. The cognitive effects of obtained measures were tested in two experiments conducted in both Serbian and English: a visual lexical decision task and a semantic categorization task. Results suggest that semantic animacy is a graded property. A high correlation between Serbian and English measures suggests that semantic animacy might be language independent, most likely because of its biological grounding. As for its behavioral correlates, animacy does not affect lexical decision times but it does codetermine the categorization speed: the category decision gradually slows as a function of the degree of animacy. These results were consistent across two languages under research scrutiny. We thus conclude that animacy is a continuous aspect of meaning.

Animacy, which is the distinction between animate (living) and inanimate (nonliving) entities, is a widely manipulated feature used to explore how we understand and process semantics (e.g., Cappa, Perani, Schnur, Tettamanti, & Fazio, 1998; Caramazza & Shelton, 1998; Grabowski, Damasio, & Damasio, 1998; Moore & Price, 1999; Mummery, Patterson, Hodges, & Price, 1998; Tyler & Moss, 2001; Tyler, Moss, Durrant-Peatfield, & Levy, 2000). Observations from clinical studies on selective impairments, for example, have raised questions about whether conceptual knowledge is represented by domains or categories and whether this apparent domain- or category-specific organization can emerge from a distributed semantic system. Several patients who had trouble identifying entities from living categories, such as animals or plants, did not exhibit similar difficulties with nonliving classes, such as tools or vehicles (De Renzi & Lucchelli, 1994; Farah, Meyer,

© Cambridge University Press 2016 0142-7164/16

https://doi.org/10.1017/S0142716416000096 Published online by Cambridge University Press

& McMullen, 1996). At the same time, different group of patients showed opposite difficulties, with nonliving but not with living classes (Sacchett & Humphreys, 1992; Warrington & McCarthy, 1987). These findings raised questions related to how conceptual knowledge is organized and used. However, the most influential computational modeling studies on human categorization never attempt to represent discrete concepts, but postulate category features as a continuous system, and then, depending on the tasks like sorting, labeling, and so on, they single out discrete behavioral outcomes (for details, consult Love, Medin, & Gureckis, 2004; Nosofsky, 1986).

Before using animacy as a tool for investigating the organization of the semantic memory, psycholinguistic studies should aim to determine the very nature of animacy. As an inherent property of word referents, animacy is one of the basic semantic features, or the determinants of word meaning. It is therefore of great importance to the studies of lexical access to examine how this feature is represented and activated.

Numerous behavioral and brain imaging studies using different sensory modalities have manipulated animacy. Unfortunately, the overall findings are inconclusive. Some studies have found a processing difference between animate and inanimate entities, at both a behavioral and a neural level (Martin, Wiggs, Ungerleider, & Haxby, 1996; Perani et al., 1995, 1999). Other studies have failed to replicate this finding (Devlin et al., 2002; Ilić, Ković, & Styles, 2013; Pilgrim, Fadili, Fletcher, & Tyler, 2002; Tyler, Bright, et al., 2003).

Devlin et al. (2002) concluded that this inconsistency in results might arise from uncontrolled characteristics of stimuli, such as familiarity, structural similarity among category members, and visual complexity. In addition to uncontrolled and/or confounding effects, variability in findings might also come from inconsistent (even conflicting) operationalizations of animacy. Studies have been using very different exemplars for animate and inanimate categories. The most common animate/inanimate distinction used in previous studies was animals versus tools. Studies that reported differences in processing of animate and inanimate entities used only these two categories (Martin, Wiggs, Ungerleider, & Haxby, 1996; Perani et al., 1995, 1999). Studies that have failed to demonstrate processing differences between animates and inanimates typically have used several categories of objects to instantiate animate and inanimate classes (Devlin et al., 2002; Ilić et al., 2013; Pilgrim et al., 2002; Tyler, Bright, et al., 2003). For example, animals and fruits were treated as living or natural kinds, while tools and vehicles were nonliving, that is, manmade objects (Devlin et al., 2002).

In sum, it seems that only a narrow and specific selection of animate/inanimate categories generates the processing effect, and that wider selection that includes various categories fails to replicate these findings. This suggests that there may be no processing difference between classes of animate and inanimate entities in general, but only between specific categories such as animals and tools. This concern would be alleviated, however, if the assumption that a particular category (e.g., animals) is representative of the whole class of animate entities could be demonstrated empirically. By including various instances or stimuli, it may be possible to assure generalizability of findings about differences in processing

https://doi.org/10.1017/S0142716416000096 Published online by Cambridge University Press

between animate and inanimate entities, which is the approach we used in the present study.

A LINGUISTIC ANIMACY

Previous studies conducted by psychologists have all have assumed that animacy is a binary property, that is, that an entity is either animate or inanimate but never more or less animate. The notion of animacy used by linguists, based on the animacy realization in language, is different, because linguists usually treat animacy as a graded property.

Several linguistic studies demonstrated an influencing role of animacy in different linguistic phenomena in various languages. For example, animacy plays an important role in differential object marking, where direct objects that are positioned high in the animacy scale are overtly marked (Aissen, 2003). In Slavic languages, for example, animate nouns of masculine gender show genitive–accusative syncretism (i.e., an interaction in syntactically marking animacy and gender), which is not present in inanimate nouns (Klenin, 2014). Animacy also helps to predict the dative alternation; in other words, inanimate recipients are more likely to be expressed in a prepositional dative structure (as the word *it* is used in the phrase *give it some thought*) than animate recipients (Bresnan, Cueni, Nikitina, & Baayen, 2007). Similarly, animacy determines the choice between *s*-genitive (i.e., *the dog's fur*) and *of*-genitive (i.e., *the bristles of the toothbrush*) in English; in other words, the more animate the possessor is, the more likely it is to occur in the *s*-genitive than in the *of*-genitive (Rosenbach, 2007). Animate referents are also privileged during production of syntactic structures in a way that they tend to be realized as subjects and in early sentence positions (Branigan, Pickering, & Tanaka, 2008). These preferences, naturally, help in reducing uncertainty (i.e., entropy) in communication (for related discussion, see Dye, Milin, Futrell, & Ramscar, in press).

Overall, in the majority of the linguistic studies, animacy has been treated as being a graded rather than a binary property. It has been proposed that entities are ordered along an animacy continuum or, at least, organized into hierarchical categories (for details, see Aissen, 2003; Dahl & Fraurud, 1996; Dixon, 1979; Silverstein, 1976; Zaenen et al., 2004). Proposed animacy hierarchies differ in the granularity of classes, but they all assume HUMAN > ANIMAL > INANIMATE order; that is, in linguistic animacy scales, human beings are ranked higher than animals, although the latter are no less animate than the former in a biological sense. Investigating its role in linguistic phenomena such as number distinctions, case marking, and word order, Yamamoto (1999) showed that animacy as a linguistic factor cannot solely be defined on the grounds of the biological living/nonliving distinction. What matters is not so much the inherent biological status of an entity, but rather how humans perceive that entity, and to what extent they treat it as being alive (in Yagua language, e.g., rocks and pineapples are animate, as discussed by Frawley, 1992). Therefore, the proposed HUMAN > ANIMAL > INANIMATE hierarchy might be even subject to more fine-grained differentiations, as when dogs are subjectively perceived as more "animate" than bacteria (cf. Yamamoto, 1999).

https://doi.org/10.1017/S0142716416000096 Published online by Cambridge University Press

THE NATURE OF SEMANTIC ANIMACY

Investigations of neural activity have actually suggested a nonbinary classification by showing that some inanimate categories did not evoke maximal response in areas where maximal activation of other inanimate categories has been found, but rather evoke activity in those areas that are the most active for animate categories, and vice versa (Chao, Weisberg, & Martin, 2002; Ishai, Ungerleider, Martin, Schouten, & Haxby, 1999). This finding suggests that the representation of the animacy feature might be distributed across the cortex. The idea of distributed features has received support in recent years in the investigation of organization of semantic system (Tyler, Stamatakis, et al., 2003; Vigliocco, Vinson, Lewis, & Garrett, 2004). This also fits well with computational models of human categorization that we mentioned above (Love et al., 2004; Nosofsky, 1986) and with the gradient approach to grammar: the probabilistic structure of language is very well documented, especially in the study of the internal structure of words (Hay & Baayen, 2005).

A previous study conducted in Serbian tested whether a binary operationalization of animacy could predict word processing (Radanović & Milin, 2011). The nouns were classified as either animate or inanimate using a morphological marker that is present only in masculine Serbian nouns (i.e., genitive–accusative syncretism). In other words, we tested the effect of linguistic animacy marking on cognitive processing in Serbian by comparing the cognitive processing of nouns marked for animacy (masculine nouns) and those that are not (feminine nouns). The main finding was that animacy did not predict response latencies in a lexical decision task, either alone or in interaction with gender, leading to the conclusion that a binary definition of animacy (at least one that is linguistically determined) does not affect word recognition.

Thus far, we have introduced two notions of animacy: semantic animacy and linguistic animacy. Semantic animacy is what we call animacy used in studies investigating semantic knowledge, while linguistic animacy is determined by the role animacy has in linguistic phenomena. Semantic animacy is routinely used in the psychological research but has rarely been explicitly defined. It seems that researchers have been treating animacy as it is defined in biology, but it has not been empirically tested whether semantic animacy can be assumed to be the same as biological animacy.[1] It might happen, for example, that semantic animacy is more similar to the linguistic animacy than to the biological animacy, in being graded, anthropocentric, and language dependent.

THE PRESENT STUDY

In this study we are interested in empirically investigating what counts as semantic animacy and whether the representation of animacy is different in different languages. Thus, we aimed to construct different measures of semantic animacy and to test their cognitive relevance, as well as to compare these findings across Serbian and English.

In order to do so, we aimed to describe the underlying distribution of semantic animacy: to see whether it is a discrete (binary) or a continuous property. We

https://doi.org/10.1017/S0142716416000096 Published online by Cambridge University Press

measured animacy in Serbian using three different normative methodologies. Two of the methods (subjective ratings and pairwise comparison method) are commonly used for judging characteristics of language material, such as words or sentences (Schütze & Sprouse, 2014), while a third method was constructed for the purposes of the study, using the familiar form of a checklist. For the English part of the study, the most often used normative method (subjective ratings) was replicated, and that allowed direct cross-linguistic comparison. We have also included comparisons with results from previous studies, which, by and large, have been conducted in English.

Comparing the distribution of animacy in two languages coming from different language branches within the Indo-European language family, such as Serbian and English, can shed some light on the very nature of semantic animacy by providing evidence about whether semantic animacy is language specific or not. If semantic animacy is determined mainly by inherent biological animacy status of the noun referent, then we should not observe differences between animacy judgments across languages. If, however, animacy judgments in two languages do not correspond, we might suspect that linguistic factors play a role in semantic animacy. To our best knowledge, there are no similar studies comparing the effects of semantic animacy across different languages.

Finally, we were interested in testing the behavioral effects of our measure(s) of semantic animacy, in both Serbian and English. In order to do so, we chose two experimental tasks, frequently used in psycholinguistics studies for investigating cognitive effects of semantic variables: a lexical decision task and a semantic categorization task. These tasks differ with respect to the depth of lexical processing they require; that is, semantic categorization is more semantically demanding, which makes it more sensitive to the effects of the semantic variable in question, in this case animacy.

In the following, we present the Serbian part of the study: three norming studies used to generate measures of semantic animacy, as well as two experiments and the results, which tested cognitive relevance of obtained measures. The replication in English, which follows, consisted of a norming study and two experiments with their results. Finally, we provide a general discussion of the results obtained in the two languages.

EXPERIMENTS 1 AND 2: METHOD

Experiments 1 (lexical decision) and 2 (semantic categorization) were conducted in Serbian. These experiments aimed at testing the cognitive effects of animacy, measured in the normative part of a study, using three different methods. In the following text, we introduce our norming procedures and then provide a description of the experiments we conducted.

Participants

A total of 298 students (18–23 years old) attending classes at the University of Novi Sad participated in the Serbian part of the study. All participants were native speakers of Serbian and had normal or corrected to normal vision. Because the

https://doi.org/10.1017/S0142716416000096 Published online by Cambridge University Press

total sample was randomly divided into five subsamples, no participant was in more than one norming procedure or experiment. Subsample sizes are reported for each norming method/experiment.

Stimuli

We selected 72 Serbian nouns that denote various animate and inanimate entities (humans, animals, plants, tools, etc.). We included exemplars of categories used in previous research as well as some "problematic" exemplars, which, as such, have not been investigated in the previous studies. Motivation for including problematic cases came from work of Yamamoto (1999) on linguistic animacy. For example, we included entities such as AMOEBA, which might be considered as less animate than DOG, due to the anthropocentric nature of linguistic animacy. Although nouns should be concrete because animacy is a property of word referents, some nouns, such as WITCH or GOD, might be less concrete. Therefore, we statistically controlled for concreteness of nouns. Half of the nouns were of masculine gender and half were of feminine gender. Based on the results from our previous study (Radanović & Milin, 2011), we did not expect processing differences between masculine and feminine nouns nor an interaction between animacy and gender.

Measurement of animacy

To quantify animacy, we used three different methods. The first method, *subjective ratings*, provided an absolute measure of animacy of a noun referent. The second method, *a pairwise comparison method*, gave us the relative measure of animacy of a noun referent, within our stimuli set. The third method, *a checklist of biological characteristics of living entities*, showed the degree to which an entity meets biological criteria of animacy.

Subjective ratings. The first method used was subjective ratings. This method makes assessment of psychological attributes simple and straightforward and is commonly used in psycholinguistic studies (Schütze & Sprouse, 2014). Because participants were asked to evaluate stimuli on a single scale, this method gave us an absolute measure of the animacy of the referents of our stimuli. We described this measure as being absolute because each word referent's animacy is rated regardless of other stimuli.

Following standard procedure, 100 participants (82 females, 18 males) were asked to provide a numerical judgment of the animacy of a noun referent. These assessments were given on a 7-point Likert-type scale (1 = *the least animate*, 7 = *the most animate*). The assessment questionnaire was made using the LimeSurvey web application (LimeSurvey Project Team/Carsten Schmitz, 2012). All participants judged all stimuli, and the presentation order was randomized per participant.

We calculated a mean rating across subjects for each of our stimuli. The shape of the ratings' distribution slightly deviated from Gaussian distribution ($W = .96$, $p = .03$).

https://doi.org/10.1017/S0142716416000096 Published online by Cambridge University Press

Pairwise comparison method. In the pairwise comparison method, participants chose between two options by deciding which word referent was more animate. This is a typical two-alternative forced-choice task (Schütze & Sprouse, 2014). We included pairwise comparisons because they provide a relative measure of a particular attribute; that is, this method offers the relative position of a given word (with regard to a referent's animacy), within a given set of stimuli.

Stimuli were presented using a custom-written web application. One hundred participants (76 females, 24 males), none of whom participated in the subjective ratings part, had to choose which of two simultaneously presented words denoted the more animate entity (e.g., SAILOR or DOG). The pairwise comparison method assumes that all possible pairs of the items are presented to every respondent and is usually used with a smaller number of items. Because the number of possible words pairs in our study was 2,556, we were not able to present them all to each participant. We, therefore, divided the pairs of our stimuli in 10 approximately equal groups (9 consisting of 257 pairs of words, and 1 consisting of 243 pairs). Groups of stimuli were presented to 10 groups of participants who had been randomly assigned to a group. Each participant was thereby presented with all words, but not with all pairs that a particular word could form. The sequence of presentation of word pairs and the position of a word in a pair were randomized for each participant.

We calculated the relative frequency (or estimated probability) of choice for each word. Distribution of this measure matched the Gaussian ($W = 0.97$, $p = .10$). Because the shape of the ratings' distribution slightly deviated from normality, we calculated both Pearson product-moment correlation coefficient and a Spearman rank correlation coefficient to determine the association between mean rating and the pairwise measure. Both methods gave a similar result: the correlation between these measures was exceptionally high ($r = .96$, $p < .001$; $\rho = 0.95$, $p < .001$).

A checklist of biological characteristics of living entities. Finally, we created a checklist of biological characteristics of living entities in order to determine whether the animacy estimated by participants corresponds to a rating of biologically determined animacy. The list of attributes was taken from a standard biology textbook (Kadhila, 2005). These characteristics were as follows: nutrition, respiration, movement, excretion, growth, reproduction, and sensitivity. The characteristics were given in the following form: *cow: eats, breathes, moves, etc.* Thirty-two participants (28 females, 4 males), who had not participated in the previous norming methods, were asked to check all characteristics that apply to a word's referent. We formed two versions of a questionnaire with different orders of characteristics and randomly assigned participants to one. Different forms were made to control for possible effect of the order of the characteristics on the participants' answering. In addition, the order of presented stimuli was randomized for each participant. The questionnaire was made using the LimeSurvey web application (LimeSurvey Project Team/Carsten Schmitz, 2012).

The relative frequency of choice (the sum of all checked answers for each item) was calculated for each word. The distribution of the checklist measure deviated from normality ($W = 0.85$, $p < .001$). Thus, we calculated both a Pearson product-moment correlation coefficient and a Spearman rank

https://doi.org/10.1017/S0142716416000096 Published online by Cambridge University Press

correlation coefficient to determine the relationship between the checklist measure and the two previous measures of animacy. The result was similar. The Pearson product-moment correlation coefficient was $r = .91$ ($p < .001$) for both animacy measures. The Spearman rank correlation coefficient was $\rho = 0.90$ ($p < .001$) for the ratings measure and $\rho = 0.92$ ($p < .001$) for the pairwise measure.

From the results of the normative studies, we concluded that any of the three measures could be interchangeably used as a quantitative operationalization of graded semantic animacy. However, instead of making a random choice, we combined them in an optimally weighted composite by applying principal component analysis (see, e.g., Dunteman, 1989; Jolliffe, 2002), and then used component scores in later parts of the study.

Experiment 1: Lexical decision task

The first task chosen for testing cognitive effects of animacy measure was a visual lexical decision. Although it is not a semantic task, lexical decision is sensitive to semantic characteristics of words (Baayen, Feldman, & Schreuder, 2006; Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004), which made it suitable for testing cognitive effects of animacy measures.

Participants. Thirty-two first-year students (30 females, 2 males) from the Department of Psychology at the University of Novi Sad participated in the experiment, in exchange for course credits. All participants were native speakers of Serbian, right-handed, with normal or corrected to normal vision. None had participated in the norming part of the study.

Stimuli. For the purpose of this experiment, pseudowords appeared along with our critical stimuli. There were 144 stimuli in total: 72 words and 72 pseudowords. The pseudowords were constructed using the pseudoword generator Wuggy (Keuleers & Brysbaert, 2010). Pseudowords followed Serbian orthographic and phonotactic rules.

Procedure. Stimuli were presented in a visual lexical decision task using DMDX software (Forster & Forster, 2003), on a standard PC configuration (AMD Sempron 2600+ 1.61-GHz processor with 256 MB RAM, a standard video card, and monitor set to a 70-Hz vertical refresh rate and 1024 × 768 pixel resolution). Participants were aurally instructed to answer as quickly and accurately as possible whether the presented string of letters was a word in the Serbian language or not. They indicated their answers by pressing the left mouse button for "yes" and right mouse button for "no." The presentation of both words and pseudowords was preceded by a 500-ms fixation point. The stimuli remained on the screen until a response, or until 1500 ms had passed. The interstimulus interval (ISI) was 500 ms. Stimuli were written in black, 40-point Times New Roman capital letters and presented on a light gray background in the middle of the screen. The experimental materials were preceded by eight practice trials, with four nouns and four pseudonouns. The presentation sequence was randomized for each participant.

https://doi.org/10.1017/S0142716416000096 Published online by Cambridge University Press

Experiment 2: Semantic categorization task

Participants also performed the semantic categorization task, which assumes deeper processing due to discrimination of specific semantic information (cf. Tyler, Russell, Fadili, & Moss, 2001). By asking our participants to determine whether the referent of a presented noun was an animate or an inanimate entity, we ensured that information regarding semantic animacy would be used in task execution.

Participants. Thirty-four first-year students (32 females, 2 males) from the Department of Psychology at the University of Novi Sad participated in the experiment, in exchange for course credits. All participants were native speakers of Serbian, right-handed, with normal or corrected to normal vision, who had not participated in any of the previous parts of the study.

Stimuli. Stimuli were the same 72 words used in all previous analyses.

Procedure. Stimuli were presented in a semantic categorization task using the same software and equipment as in the previous experiment and had the same physical characteristics. Participants categorized words as denoting animate (living) or inanimate (nonliving) entities by pressing one of two mouse buttons. Half of the participants pressed the left mouse button for animate and the right for inanimate and the other half pressed the left button for inanimate and the right for animate. Participants were randomly assigned to button orientation. In front of the mouse, there was a piece of paper with words animate and inanimate written on the corresponding side, serving as a reminder. As in the lexical decision experiment, each session started with eight practice trials, and the presentation sequence was randomized for each participant.

EXPERIMENTS 1 AND 2: RESULTS

All three animacy measures (mean rating, pairwise measure, and checklist measure) were highly correlated. Therefore, in the first step of the analysis, we applied a principal components analysis (PCA) over those measures, to reparameterize and bypass multicollinearity issues. As expected, the animacy measures were grouped into one factor, which accounted for 95.02% of the total variance. We computed the factor scores of each word and used them in further analyses. The new animacy measure (factor scores) was highly correlated with each of the original animacy measures ($r = .98$, $p < .001$ with the pairwise measure, $r = .98$, $p < .001$ with the ratings measure, and $r = .96$, $p < .001$ with the checklist measure).

Experiment 1: Lexical decision task

To ensure a better approximation of the Gaussian distribution, the predictors and the dependent variable were transformed. Response latencies were transformed using an inverse transformation and then standardized, while log transformation

https://doi.org/10.1017/S0142716416000096 Published online by Cambridge University Press

was applied to word frequencies and word length. In addition, animacy (factor scores from PCA) as well as concreteness were rank transformed. In the case of concreteness, rank transformation was applied because of its restricted range. Concreteness ratings were taken from Brysbaert, Warriner, and Kuperman (2014). The ratings were obtained using English words. However, we used them in Serbian experiments as well because concreteness should not be language dependent. We excluded one item (TANJIR, meaning "plate") because its rank animacy was a clear outlier (−3.49, while the rest of items were in range from −2.17 to 2.43). Visual inspection of the reaction time distribution showed a small number (0.77%) of extreme outliers that were also excluded from further analysis.

We applied generalized additive mixed modeling (GAMM). This procedure was used for analyzing data from all four experiments. The analysis was performed in R (R Core Team, 2013) using the package *mgcv* (Wood, 2006). In principle, linear mixed-effect modeling can be understood as a special case of GAMM. The difference is that GAMM allows for optimal (i.e., parsimonious) nonlinear smoothing and nonlinear interaction of numerical covariates. A number of linear mixed-effect modeling based studies asked for additional nonlinear treatment of variables such as frequency or length (Baayen et al., 2006; Baayen, Milin, Filipović-Đurđević, Hendrix, & Marelli, 2011). Typically, nonlinearity was treated with polynomials, and rarely with restricted cubic splines. For those reasons, using GAMM seemed an obvious choice, providing a principled approach for nonlinearity treatment.

We specified a Gaussian link function, with response latencies as the dependent variable and log frequency, log length, rank concreteness, and rank-transformed animacy as predictor variables, while participants and items were modeled as random effects. We were especially interested in the effect of semantic animacy. Log frequency, log length, and rank concreteness served as control covariates.

The final model was refitted after model criticism (i.e., removing data points with absolute standardized residuals exceeding 2.5; detailed discussion of rationales can be found in Baayen, 2008; Baayen & Milin, 2010). The final model revealed significant effects of the control predictors, in the expected directions: faster decision latencies from higher log frequency and concreteness, and slower from longer log length, although the latter was borderline significant (Table 1). It is important that the model did not show a significant contribution of rank animacy ($t = -0.54$, $p = .59$). The animacy of a noun does not influence the decision latencies over and above the two consistently attested predictors.

Experiment 2: Semantic categorization task

As in the lexical decision experiment, we standardized inverse transformed response latencies and log transformed word frequencies and word length, while animacy (factor scores from PCA) was rank transformed. In addition, one item (TANJIR, meaning "plate") was again excluded due to its large deviation from the animacy distribution. Then, we excluded 0.29% of additional outliers, after visual inspection of the reaction time quantile–quantile plot.

https://doi.org/10.1017/S0142716416000096 Published online by Cambridge University Press

Table 1. *Estimated coefficients and corresponding statistics for the model fitted to response latencies elicited from a visual lexical decision experiment in Serbian*

Parametric Coefficients	Estimate	*SE*	*t*	*p*
Intercept	0.35	0.30	1.19	.236
Log frequency	−0.13	0.02	−5.17	.000
Log length	0.24	0.12	1.94	.053
Rank concreteness	−0.07	0.02	−2.81	.005
Smooth Terms	**E *df***	**Ref. *df***	***F***	***p***
Participants	30.20	31	37.10	.000
Items	51.91	63	4.67	.000

Before analyzing the data, we checked whether response latencies were influenced by the orientation of the animate and inanimate response buttons. The analysis showed that this was not the case ($t = -1.63$, $p = .10$). We also checked whether the distribution of the answers (animate or inanimate) differed between these versions of the task. The test showed that participants' answers are completely independent of the button pressed ($\chi^2 = 0.0007$, $df = 1$, $p = .98$). Thus, we did not consider task version as a factor.

For the main analysis, log frequency, log length, rank concreteness, and spline-smoothed rank animacy, as well as the standardized trial order, were included as predictors, while participants and items were treated as random effects. Nonlinearities were tested for all predictors. Animacy was the only predictor with a significant nonlinear effect on the response latencies. We restricted the number of knots of the animacy smoother to three, to control for overfitting. After considering the number of spline knots in an unrestricted model, the comparison of restricted and unrestricted models and the visual inspection of effects from both models, three turned out to be the optimal number of knots; that is, the fit of the model was not improved significantly by additional knots. Unlike the data in lexical decision, there were no significant effects of the log frequency and log length ($t = -1.34$, $p = .18$ and $t = -1.45$, $p = .15$, respectively), which is not surprising given the nature of the task. However, the refitted model (with extreme outlying residuals removed) revealed a significant effect of the rank animacy smoother, as well as the effect of the concreteness. The effect of the factorial smooth of standardized trial order by participants was also significant (Table 2). The left panel in Figure 1 shows the nonlinear effect of rank animacy; the time required to categorize a word increases as the word referent becomes less animate or more inanimate. We can also see that the words whose referents are clearly animate are categorized faster that the words whose referents are clearly inanimate. The difference in mean reaction time items with extreme animacy values (5% lowest and 5% highest) was 95 ms. The right panel of Figure 1 represents the by-participant factor smooths for experimental trials. Each curve in this figure represents one subject, and curves show how the subject's performance develops in an experimental session. These

https://doi.org/10.1017/S0142716416000096 Published online by Cambridge University Press

Table 2. *Estimated coefficients and corresponding statistics for the model fitted to response latencies elicited from a semantic categorization experiment in Serbian*

Parametric Coefficients	Estimate	*SE*	*t*	*p*
Intercept	0.07	0.10	0.73	.468
Rank concreteness	−0.06	0.02	−3.62	.000
Smooth Terms	**E *df***	**Ref. *df***	***F***	***p***
Rank animacy	1.97	1.98	25.59	.000
Standardized order, participants	60.53	303	3.17	.000
Items	37.41	64	1.44	.000

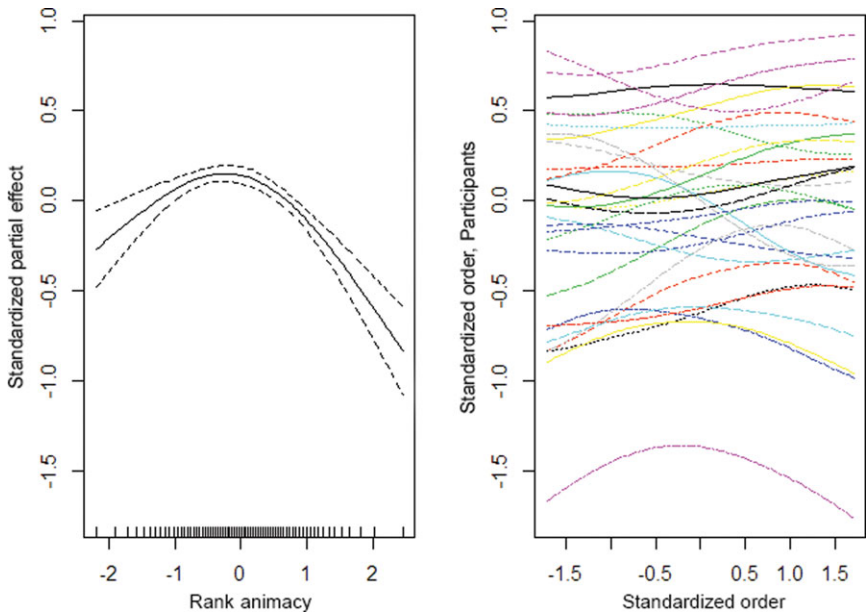


Figure 1. (Color online) (Left) Partial effect of rank-animacy's smoother and (right) factorial spline interaction term of standardized trial order by participants for the model fitted to the response latencies elicited from a semantic categorization experiment in Serbian.

by-participant factor smooths serve for controlling intertrial dependencies in the response latency time series.

EXPERIMENTS 3 AND 4: METHOD

Experiments 3 and 4 were conducted in English. To permit a full comparison, Experiment 3 used a lexical decision task, and Experiment 4 used a semantic

https://doi.org/10.1017/S0142716416000096 Published online by Cambridge University Press

categorization task. The overall procedure for the English experiments was slightly different than the one for Serbian, and thus procedure description for Experiments 3 and 4 is combined, as explained below.

Participants

One hundred twenty-six right-handed University of Alberta undergraduate students (58 males, 68 females; 17–28 years old) participated in order to earn partial course credit as part of a subject pool. All reported themselves to be right-handers with normal (or corrected to normal) vision who had learned to speak English before the age of 6.

Stimuli

The stimuli were the same 72 words that had been rated in Serbian, translated into English. The pseudowords for the lexical decision task were again constructed using the pseudoword generator Wuggy (Keuleers & Brysbaert, 2010) but now followed English orthographic and phonotactic rules. Each participant undertook a rating task and one of the two experimental tasks (either lexical decision or semantic judgment). To ensure that they did not see any words twice, the words were divided into two sets, half of which each participant saw in the rating task and half of which each participant saw in the other task to which s/he was randomly assigned. The forms were counterbalanced between subjects to ensure that all words were seen in all conditions. The order of presentation of the two tasks each subject completed was also counterbalanced.

Procedure

Stimuli were presented using ACTUATE software (Westbury, 2007) running on Apple G4 Macintosh Mini connected to 17-in. LCD monitors. Strings used in both tasks were presented on the screen in 60-point Times font, in the middle of a white rectangle measuring 158×382 pixels centered on the screen against a black background. Both tasks were preceded by the presentation of written instructions that were explained to each participant aurally by the experimenter.

The semantic rating task that was undertaken by all participants asked participants to rate each of the 72 words for animacy, by using the mouse to drag a marker on an unmarked Likert scale labeled *definitely animate* on one side and *definitely inanimate* on the other. The side labeled by each label was counterbalanced between subjects, so that half rated animacy high on the left side of the scale and half rated animacy high on the right side of the scale.

In the lexical decision task, participants were instructed to answer as quickly and accurately as possible whether each string of letters was an English word. They gave their answers by pressing the C key (for correct) for yes and the X key (for incorrect) for no. They were instructed to use the first and second fingers of their right hand to make the respective responses. Each stimulus was preceded by a fixation cross (+), which remained on the screen for a random time between 500

https://doi.org/10.1017/S0142716416000096 Published online by Cambridge University Press

Table 3. *Estimated coefficients and corresponding statistics for the model fitted to response latencies elicited from a visual lexical decision experiment in English*

Parametric Coefficients	Estimate	*SE*	*t*	*p*
Intercept	0.39	0.21	1.89	.059
Log frequency	−0.12	0.02	−6.08	.000
Smooth Terms	**E *df***	**Ref. *df***	***F***	***p***
Participants	58.76	62	18.80	.000
Items	42.75	66	1.90	.000

and 750 ms. Stimuli remained on the screen until a response was given. The ISI was 1 s. The presentation sequence was randomized for each participant.

The semantic decision task required subjects to make responses that varied by trial. The experiment was designed this way to make the "yes" and "no" responses equivalent. Prior to the presentation of each word stimulus, participants were presented with either the word ANIMATE? or the word INANIMATE?. This stayed on the screen for a random time between 1500 and 2500 ms, before being replaced by a fixation cross that remained on the screen for a random time between 500 and 750 ms. The cross was then replaced by the word about which participants had to make a decision. If the word was in the indicated category, they were asked to press the C key (for correct). If it is was not, they were asked to press the X key (for incorrect). As in the lexical decision task, subjects made their response using the first and second fingers of their right hand. The ISI was 1 s.

EXPERIMENTS 3 AND 4: RESULTS

Experiment 3: Lexical decision task

In the preparatory step of the analysis, we transformed all variables in the same manner as we did with the Serbian data set. As in the case of Serbian, one item (BABY), which largely deviated (rank animacy = 3.51) from the animacy distribution (rank animacy = −2.46 to 2.18), was excluded. In addition, a small number (1.46%) of the extreme outliers were excluded from further analysis.

We confirmed that response latencies for the two sets of items did not differ ($t = -0.90$, $p = .37$). As with Serbian data, a Gaussian link function with response latencies as the dependent variable and log frequency, log length, rank concreteness, and rank-transformed animacy as predictor variables was used. Random effects of participants and items were also included.

The refitted model showed a similar pattern of results as was seen in the Serbian data set. Log frequency showed the expected effect: increases speed processing (Table 3). However, the coefficients for log length ($t = -0.006$, $p = .10$) and rank concreteness ($t = -0.55$, $p = .58$) failed to reach significance. As in the Serbian data set, animacy of a word referent did not influence decision latency ($t = 0.79$, $p = .43$).

https://doi.org/10.1017/S0142716416000096 Published online by Cambridge University Press

Table 4. *Estimated coefficients and corresponding statistics for the model fitted to response latencies elicited from a visual lexical decision experiment in English*

Parametric Coefficients	Estimate	*SE*	*t*	*p*
Intercept	0.01	0.07	0.11	.913
Rank concreteness	−0.05	0.02	−2.64	.008
Smooth Terms	**E *df***	**Ref. *df***	***F***	***p***
Rank animacy	1.97	1.98	18.93	.000
Standardized order, participants	112.84	619	2.02	.000
Items	40.10	68	1.49	.000

Experiment 4: Semantic categorization task

All variables, except for the reaction times, were transformed as in the previous analyses. Here, response latencies were transformed using inverse square root transformation and then standardized because this was a more appropriate transformation. The reaction time distribution was mildly trimmed; only 0.09% of the extreme outliers were excluded. Before trimming reaction time distribution, one item (BABY) that was a clear outlier in the animacy distribution was excluded.

As with the lexical decision task, we assured that reaction times for different experimental lists did not differ ($t = -0.61$, $p = .54$). We also checked whether the distribution of the answers (animate or inanimate) differed between these lists. The test showed that this was not the case ($\chi^2 = 1.38$, $df = 1$, $p = .24$).

We used GAMM with log frequency, log length, rank concreteness, rank animacy with a spline smoother, and standardized trial order as predictors and participants and items as random effects. As in Serbian, the model with three knots for the animacy smoother was the most parsimonious. The outcome was the same as in the semantic categorization task in Serbian: there was a significant contribution of the rank-animacy's smoother and a factorial smooth of standardized trial order by participants, as well as the effect of rank concreteness (Table 4), while effects of other control predictors were nonsignificant (log frequency: $t = -1.48$, $p = .14$; log length: $t = 0.91$, $p = .37$). As in Serbian, in English it is easier to categorize a word if its referent is clearly animate or inanimate (Figure 2, left graph). However, in English the overall difference in processing time between animate and inanimate items is not so obvious, perhaps because the method required a more complex motor response (sliding a marker using the mouse as opposed to clicking on the scale). Nevertheless, comparing the mean reaction time for 5% lowest ranked-animacy values and 5% highest ranked values showed a benefit of animacy of similar magnitude (108 ms). The right panel in the Figure 2 shows participants' behavior over trials.

DISCUSSION

In this study we aimed to quantify and operationalize semantic animacy. We determined the underlying distribution of semantic animacy by measuring it with

https://doi.org/10.1017/S0142716416000096 Published online by Cambridge University Press

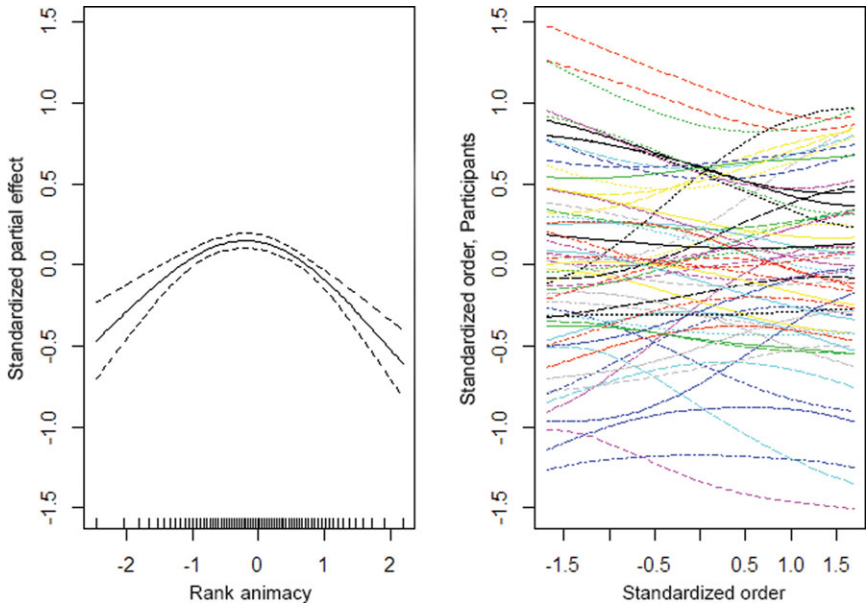


Figure 2. (Color online) (Left) Partial effect of rank-animacy's smoother and (right) factorial spline interaction term of standardized trial order by participants for the model fitted to the response latencies elicited from a semantic categorization experiment in English.

three different normative methodologies in Serbian. Then, the most widely used method of rating scales was replicated in English. We showed high intra- and interlanguage consistency of those measures. To test their cognitive relevance, we ran a lexical decision and a semantic categorization experiment, in both Serbian and English. Results convincingly showed the graded nature of semantic animacy. This effect was robust in a semantic categorization but nonsignificant in a lexical decision task, in both languages.

The primary contribution of our study is empirical, quantitative assessment of animacy, and cross-language experimental evaluation of the relevance of those quantitative measures. In previous studies an implicit assumption was made that animacy is a binary feature, and the authors subjectively determined what counts for animate or inanimate entity. In the present study we let our participants determine the degree to which each item was animate. We probed for the degree of animacy using three different scaling techniques. The subjective ratings and the pairwise comparison measure (or, the absolute and relative measure of animacy of a noun referent) were highly correlated, which indicates that measures of semantic animacy are not confined to a particular method of measurement. The distributions of obtained animacy measures were continual and Gaussian-like (i.e., bell shaped). Furthermore, these measures were highly associated with a measure based on biological properties, which enables us to conclude that most of the *substance* of semantic animacy corresponds highly to its biological determinants.

https://doi.org/10.1017/S0142716416000096 Published online by Cambridge University Press

Thus, semantic animacy, as measured here, and animacy as a linguistic factor, described by the linguists (Aissen, 2003; Dahl & Fraurud, 1996; Dixon, 1979; Silverstein, 1976; Yamamoto, 1999; Zaenen et al., 2004), share the property of gradation, but their respective continuums have somewhat different determinants.

It seems that semantic animacy is based on a biological notion of animacy, while linguistic animacy is more anthropocentric. In particular, linguistic hierarchies are such that humans are placed higher that other animates, while it seems that there is no difference between humans and animals (at least, "higher" ones) in semantic animacy. This can be observed by inspection of animacy ratings of our stimuli that are presented in Appendix A. Table A.1 presents average subjective ratings in Serbian and in English. Humans and "higher animals" are placed high in the animacy distribution, and they are mixed. In contrast, that "lower animals" such as a worm or crab are placed below humans and higher animals reveals that anthropomorphism is also a factor that codetermines semantic animacy. It seems that semantic animacy rests somewhere between pure biological (or evidential) and linguistic animacy. However, it is too soon to draw definite conclusions. A large-scale norming study is needed to confirm this speculation.

Another interesting result of the present study is that Serbian and English subjective ratings were highly correlated ($r = .89$, $p < .001$; $\rho = 0.92$, $p < .001$). Although we cannot make generalizations based on the results obtained with only two languages, bearing in mind that Serbian and English are coming from different language branches within the Indo-European language family, this result is at least quite suggestive in that semantic animacy might be language independent. We assume this could be the case because biological factors are dominant in determination of semantic animacy. However, this hypothesis would also benefit from a large cross-linguistic study.

The results of both Serbian and English lexical decision experiments show that semantic animacy does not affect word recognition latencies. Lexical decision task was selected to test cognitive effects of animacy because it is sensitive to several of the semantic variables (Baayen et al., 2006; Balota et al., 2004). It seems, however, that graded semantic animacy is not one of them. In a previous study in Serbian, the difference was not found in lexical decision times between animacy categories either (Radanović & Milin, 2011). These two results therefore support previous studies that used several categories of objects as stimuli and controlled for relevant variables (Devlin et al., 2002; Pilgrim et al., 2002; Tyler, Bright, et al., 2003). Taken together, these results suggest that semantic animacy does not play a role in the word-recognition process. Whether a noun is clearly animate or inanimate or something in between will not influence its recognition time.

This conclusion is based on basically null results. However, bearing in mind that these studies properly controlled for relevant variables and used numerous categories of stimuli, we argue that this result can be trusted. In addition, the statistical approach we applied here is far more sophisticated than the analysis of variance based analyses that are typically applied. On the one hand, our statistical analysis took into account random variations both within participants and within stimuli, which assured that obtained results could be generalized to both participants and stimuli (Baayen, 2008; Baayen, Davidson, & Bates, 2008). On the other hand, this

https://doi.org/10.1017/S0142716416000096 Published online by Cambridge University Press

kind of analysis is more powerful than traditional analyses in terms of detecting the effects of predictors if they exist (Baayen, 2008; Baayen et al., 2008).

It is crucial that graded semantic animacy does influence response latencies in a semantic categorization task such that decisions are slower as the word referent is less clearly animate or inanimate. This result was again found both for Serbian and for English. We have chosen this task because it required participants to discriminate animacy of word in order to perform appropriately. For lexical decision this was not the case: the animacy information is not required for executing the task, and as we observed, animacy does not emerge as a codeterminant for the word recognition. However, when we force participants to use information regarding animacy, it reveals that this information is represented in a graded manner. Therefore, that it is easier to decide that DOG is animate than is SPIDER might be taken as evidence that semantic animacy is a graded, rather than a binary, property.

Observed findings regarding the graded nature of semantic animacy has both practical and theoretical implications for studies investigating organization of semantic knowledge. At the practical level, it indicates that stimuli have to be selected more carefully with regard to their animacy because entities that seem like members of the same category (e.g., DOG and SPIDER) might differ in terms of semantic animacy. At the theoretical level, present results proved support for recent proposals of distributed representations of semantic features in the brain (Tyler, Stamatakis, et al., 2003; Vigliocco et al., 2004), as well as for the computational models of human categorization proposed by Love et al. (2004) and Nosofsky (1986).

APPENDIX A

Table A.1. *Averaged Serbian and English subjective ratings*

Averaged Serbian Subjective Ratings		Averaged English Subjective Ratings	
Stimulus (English Trans.)	Animacy Rating	Stimulus	Animacy Rating
beba (baby)	6.51	baby	98.29
majka (mother)	6.51	dog	97.92
dečak (boy)	6.43	mother	97.68
brat (brother)	6.34	girlfriend	97.61
pas (dog)	6.34	boy	97.44
drugarica (girlfriend)	6.26	uncle	97.29
konj (horse)	6.20	giraffe	97.29
kuvarica (cook)	5.97	sailor	96.42
ujak (uncle)	5.97	crow	95.79
veverica (squirrel)	5.91	brother	95.53
mornar (sailor)	5.86	professor	95.48
učiteljica (teacher)	5.80	squirrel	95.41
golub (pigeon)	5.69	elephant	95.38
slon (elephant)	5.63	crab	95.27

https://doi.org/10.1017/S0142716416000096 Published online by Cambridge University Press

Table A.1 (*cont.*)

Averaged Serbian Subjective Ratings		Averaged English Subjective Ratings	
Stimulus (English Trans.)	Animacy Rating	Stimulus	Animacy Rating
krava (cow)	5.60	pigeon	95.14
profesor (professor)	5.51	cow	94.67
žirafa (giraffe)	5.46	horse	94.59
kokoška (hen)	5.37	hen	93.92
kraljica (queen)	5.37	spider	93.67
vrana (crow)	5.37	worm	91.53
muva (fly)	5.31	teacher	91.42
pauk (spider)	5.23	prince	91.23
crv (worm)	5.03	fly	89.73
reka (river)	4.89	queen	89.50
kraba (crab)	4.80	amoeba	83.55
princ (prince)	4.74	bacteria	82.59
okean (ocean)	4.71	clam	82.03
bakterija (bacteria)	4.63	dragon	81.98
virus (virus)	4.57	witch	77.06
jabuka (apple)	4.54	cook	74.17
orhideja (orchid)	4.54	giant	73.70
zmaj (dragon)	4.46	virus	69.41
jagoda (strawberry)	4.34	cactus	65.08
limun (lemon)	4.31	fairy	62.86
lubenica (watermelon)	4.31	god	60.69
kaktus (cactus)	4.26	orchid	58.59
školjka (clam)	4.23	cabbage	44.77
ameba (amoeba)	4.20	cucumber	42.35
paradajz (tomato)	4.20	ghost	41.83
paprika (pepper)	4.17	potato	41.48
krastavac (cucumber)	4.14	apple	41.21
krompir (potato)	4.03	watermelon	40.77
kupus (cabbage)	3.94	olive	40.19
maslina (olive)	3.89	lemon	39.39
grom (thunder)	3.74	tomato	38.86
div (giant)	3.63	strawberry	38.72
bog (god)	3.54	robot	33.13
veštica (witch)	3.54	ocean	32.67
vila (fairy)	3.49	pepper	30.71
kiša (rain)	3.43	river	29.85
planina (mountain)	3.34	thunder	28.02
duh (ghost)	3.29	car	16.92
knjiga (book)	3.23	image	15.67
oblak (cloud)	3.20	rain	15.19
slika (image)	3.17	computer	13.92
sneg (snow)	3.00	snow	11.83
duga (rainbow)	2.94	cloud	9.66

https://doi.org/10.1017/S0142716416000096 Published online by Cambridge University Press

Table A.1 (*cont.*)

Averaged Serbian Subjective Ratings		Averaged English Subjective Ratings	
Stimulus (English Trans.)	Animacy Rating	Stimulus	Animacy Rating
robot (robot)	2.91	rainbow	9.00
lopta (ball)	2.86	mountain	8.80
automobil (car)	2.77	book	8.50
kuća (house)	2.69	ball	7.77
kompjuter (computer)	2.49	bracelet	7.11
narukvica (bracelet)	2.06	lighter	6.03
upaljač (lighter)	2.06	coat	5.50
kaput (coat)	2.00	dress	5.08
haljina (dress)	1.94	table	4.70
jastuk (pillow)	1.80	lamp	4.41
prozor (window)	1.77	cup	4.38
lampa (lamp)	1.74	house	4.28
šolja (cup)	1.69	plate	3.89
tanjir (plate)	1.51	window	3.66
sto (table)	1.46	pillow	3.35

ACKNOWLEDGMENTS

This work was supported by the Ministry of Education, Science and Technological Development, Republic of Serbia, under Grant ON179006 and by a Discovery grant from the Natural Sciences and Engineering Research Council of Canada. We thank Laurie Beth Feldman for constructive comments on an earlier version of this manuscript. We also thank the editors and anonymous reviewers for their helpful suggestions.

NOTE

1. The treatment of plants in these studies suggests that researchers use biological criteria for animacy because they usually treat plants as animate. However, from a *naive biology* standpoint, they might be inanimate; for example, young children first categorize plants as nonliving and later, due to everyday experience with plants and learning about them, they recategorize them as living (Inagaki & Hatano, 2003). The terminology used in Caramazza and Shelton (1998) nicely illustrates that plants are problematic. In this study, plants were treated as inanimate, but living (although most of the studies use terms *(in)animate* and *(non)living* interchangeably, as we do here).

REFERENCES

Aissen, J. (2003). Differential object marking: Iconicity vs. economy. *Natural Language & Linguistic Theory, 21*, 435–483.

https://doi.org/10.1017/S0142716416000096 Published online by Cambridge University Press

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412.

Baayen, R. H., Feldman, L. B., & Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language, 53*, 496–512.

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research, 3*, 12–28.

Baayen, R. H., Milin, P., Filipović-Đurđević, D., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review, 118*, 438–481.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*, 283–316.

Branigan, B. H., Pickering, J. M., & Tanaka, M. (2008). Contributions of animacy to grammatical function assignment and word order production. *Lingua, 118*, 172–189.

Bresnan, J., Cueni, A., Nikitina, T., & Baayen, R. H. (2007). Predicting the dative alternation. In G. Bouma, I. Kraemer, & J. Zwarts (Eds.), *Cognitive foundations of interpretation* (pp. 69–94). Amsterdam: Royal Netherlands Academy of Science.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46*, 904–911.

Cappa, S. F., Perani, D., Schnur, T., Tettamanti, M., & Fazio, F. (1998). The effects of semantic category and knowledge type on lexical-semantic access: A PET study. *NeuroImage*, *8*, 350–359.

Caramazza, A., & Shelton, J. R. (1998). Domain-specific knowledge systems in the brain: The animate-inanimate distinction. *Journal of Cognitive Neuroscience*, *10*, 1–34.

Chao, L. L., Weisberg, J., & Martin, A. (2002). Experience-dependent modulation of category-related cortical activity. *Cerebral Cortex, 12*, 545–551.

Dahl, Ö., & Fraurud, K. (1996). Animacy in grammar and discourse. In T. Fretheim & J. K. Gundel (Eds.), *Reference and referent accessibility* (pp. 47–64). Amsterdam: John Benjamins.

De Renzi, E., & Lucchelli, F. (1994). Are semantic systems separately represented in the brain? The case of living category impairment. *Cortex, 30*, 3–25.

Devlin, J. T., Russell, R. P., Davis, M. H., Price, C. J., Moss, H. E., Fadili, M. J., et al. (2002). Is there an anatomical basis for category-specificity? Semantic memory studies in PET and fMRI. *Neuropsychologia, 40*, 54–75.

Dixon, R. M. W. (1979). Ergativity. *Language, 55*, 59–138.

Dunteman, G. (1989). *Principal component analysis*. Newbury Park, CA: Sage.

Dye, M., Milin, P., Futrell, R., & Ramscar, M. (in press). A functional theory of gender paradigms. In F. Kiefer, J. P. Blevins, & H. Bartos (Eds.), *Morphological paradigms and functions*. Leiden: Brill.

Farah, M. J., Meyer, M. M., & McMullen, P. A. (1996). The living/nonliving dissociation is not an artifact: Giving an a priori implausible hypothesis a strong test. *Cognitive Neuropsychology, 13*, 137–154.

Forster, K. I., & Forster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, and Computers, 35*, 116–124.

Frawley, W. (1992). *Linguistic semantics*. New York: Routledge.

Grabowski, T. J., Damasio, H., & Damasio, A. R. (1998). Premotor and prefrontal correlates of category-related lexical retrieval. *NeuroImage*, *7*, 232–243.

Hay, J. B., & Baayen, R. H. (2005). Shifting paradigms: Gradient structure in morphology. *Trends in Cognitive Sciences*, *9*, 342–348.

https://doi.org/10.1017/S0142716416000096 Published online by Cambridge University Press

Ilić, O., Ković, V., & Styles, S. J. (2013). In the absence of animacy: Superordinate category structure affects subordinate label verification. *PLOS ONE, 8*, e83282.

Inagaki, K., & Hatano, G. (2003). Conceptual and linguistic factors in inductive projection: How do young children recognize commonalities between animals and plants? In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 313–333). Cambridge, MA: MIT Press.

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences, 96*, 9379–9384.

Jolliffe, I. (2002). *Principal component analysis*. New York: Wiley.

Kadhila, N. (2005). *NSSC Biology Module 1*. Cambridge: Cambridge University Press.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods, 42*, 627–633.

Klenin, E. (2014). Belebtheit, personalität/Animacy, personhood. In S. Kempgen, P. Kosta, T. Berger, & K. Gutschmidt (Eds.), *Die slavischen Sprachen/The Slavic Languages. Halbband 2* (Vol. 32, pp. 152–161). Chicago: Walter de Gruyter.

LimeSurvey Project Team/Carsten Schmitz. (2012). LimeSurvey: An Open Source survey tool [Computer software]. Retrieved from http://www.limesurvey.org

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review, 111*, 309–332.

Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature, 379*, 649–652.

Moore, C. J., & Price, C. J. (1999). A functional neuroimaging study of the variables that generate category-specific object processing differences. *Brain, 122*, 943–962.

Mummery, C. J., Patterson, K., Hodges, J. R., & Price, C. J. (1998). Functional neuroanatomy of the semantic system: Divisible by what? *Journal of Cognitive Neuroscience*, *10*, 766–777.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*, 39–61.

Perani, D., Cappa, S. F., Bettinardi, V., Bressi, S., Gorno-Tempini, M., Matarrese, M., et al. (1995). Different neural systems for the recognition of animals and man-made tools. *NeuroReport, 6*, 1637–1641.

Perani, D., Schnur, T., Tettamanti, M., Gorno, M., Cappa, S. F., & Fazio, F. (1999). Word and picture matching: A PET study of semantic category effects. *Neuropsychologia, 37*, 293–306.

Pilgrim, L. K., Fadili, J., Fletcher, P., & Tyler, L. K. (2002). Overcoming confounds of stimulus blocking: An event-related fMRI design of semantic processing. *NeuroImage*, *16*, 713–723.

Radanović, J., & Milin, P. (2011). Morpho-semantic properties of Serbian nouns: Animacy and gender pairs. *Psihologija*, *44*, 343–366.

R Core Team. (2013). R: A language and environment for statistical computing [Computer software]. Vienna: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Rosenbach, A. (2007). Animacy and grammatical variation—Findings from English genitive variation. *Lingua, 118*, 151–171.

Sacchett, C., & Humphreys, G. W. (1992). Calling a squirrel a squirrel but a canoe a wigwam: A category-specific deficit for artefactual objects and body parts. *Cognitive Neuropsychology, 9*, 73–86.

Schütze, C. T., & Sprouse, J. (2014). Judgment data. In R. J. Podesva & D. Sharma (Eds.), *Research methods in linguistics* (pp. 27–50). Cambridge: Cambridge University Press.

Silverstein, M. (1976). Hierarchy of features and ergativity. In R. M. W. Dixon (Ed.), *Grammatical categories in Australian languages* (pp. 112–171). Canberra: Australian Institute of Aboriginal Studies.

Tyler, L., & Moss, H. (2001). Towards a distributed account of conceptual knowledge. *Trends in Cognitive Sciences*, *5*, 244–252.

https://doi.org/10.1017/S0142716416000096 Published online by Cambridge University Press

Tyler, L., Russell, R., Fadili, J., & Moss, H. (2001). The neural representation of nouns and verbs: PET studies. *Brain, 124*, 1619–1634.

Tyler, L. K., Bright, P., Dick, E., Tavares, P., Pilgrim, L., Fletcher, P., et al. (2003). Do semantic categories activate distinct cortical regions? Evidence for a distributed neural semantic system. *Cognitive Neuropsychology*, *20*, 541–559.

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 and Language*, *75*, 195–231.

Tyler, L. K., Stamatakis, E. A., Dick, E., Bright, P., Fletcher, P., & Moss, H. (2003). Objects and their actions: Evidence for a neurally distributed semantic system. *Neuroimage, 18*, 542–557.

Vigliocco, G., Vinson, D. P., Lewis, W., & Garrett, M. F. (2004). Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology*, *48*, 422–488.

Warrington, E. K., & McCarthy, R. (1987). Categories of knowledge: Further fractionation and an attempted integration. *Brain, 110*, 1273–1296.

Westbury, C. (2007). ACTUATE: Assessing cases: The University of Alberta Testing Environment [Computer software]. Retrieved from http://www.psych.ualberta.ca/~westburylab/downloads/actuate.download.html

Wood, S. N. (2006). *Generalized additive models: An introduction with R*. Boca Raton, FL: Chapman & Hall/CRC Press.

Yamamoto, M. (1999). *Animacy and reference: A cognitive approach to corpus linguistics*. Amsterdam: John Benjamins.

Zaenen, A., Carletta, J., Garretson, G., Bresnan, J., Koontz-Garboden, A., Nikitina, T., et al. (2004). Animacy encoding in English: Why and how. In D. Byron & B. Webber (Eds.), *Proceedings of the 2004 ACL Workshop on Discourse Annotation* (pp. 118–125). East Stroudsburg, PA: Association for Computational Linguistics.

https://doi.org/10.1017/S0142716416000096 Published online by Cambridge University Press