5-2020

The grammatical class effect is separable from the concreteness effect in language learning

Katherine I. Martin
Southern Illinois University Carbondale, martinki@siu.edu

Natasha Tokowicz
University of Pittsburgh, tokowicz@pitt.edu

Recommended Citation
Martin, Katherine I. and Tokowicz, Natasha. "The grammatical class effect is separable from the concreteness effect in language learning." *Bilingualism: Language and Cognition* 23, No. 3 (May 2020): 554-569. doi:10.1017/ S1366728919000233.

The Grammatical Class Effect is Separable from the Concreteness Effect in Language Learning*

Katherine I. Martin[1] and Natasha Tokowicz[2]
[1]Southern Illinois University Carbondale and [2]University of Pittsburgh

*The authors thank Charlie Eddington and members of the PLUM Lab at the University of Pittsburgh for their assistance with this study. During the writing of this manuscript, NT was funded by NIH R 01 HD075800.

Address for correspondence:
Katherine I. Martin
3226 Faner Hall, Mail Code 4517
Department of Linguistics
Southern Illinois University
Carbondale, IL 62901
Email: martinki@siu.edu

Keywords: concreteness, grammatical class, vocabulary learning

Abstract

Typically concrete words are learned better than abstract words (Kaushanskaya & Rechtzigel, 2012), and nouns are learned better than verbs (Kauschke & Stenneken, 2008). However, most studies on concreteness have not manipulated grammatical class (and vice versa), leaving the relationship between the two unclear. Therefore, in two experiments we examined the effects of grammatical class and concreteness simultaneously in foreign language vocabulary learning. In Experiment 1, English speakers learned ‘foreign language’ words (English pseudowords) mapped to concrete and abstract nouns and verbs. In Experiment 2, English speakers learned German words with the same procedure. Overall, the typical advantages for concrete words and nouns were observed. Hierarchical regression analyses provided evidence that the grammatical class effect is separable from the concreteness effect. This result challenges a strict concreteness-based source of noun/verb differences. The results also suggest that the influences of concreteness and grammatical class may vary across language measures and tasks.

The Grammatical Class Effect is Separable from the Concreteness Effect in Language Learning

Substantial research has established the influence of lexical characteristics, such as concreteness and grammatical class, on first and second language (L1 and L2) learning and processing. Concreteness[1] is the degree to which a word refers to a concept or entity that is concrete and easily perceptible (e.g., *pumpkin*) or abstract and difficult to perceive (e.g., *heaven*) (e.g., Tolentino & Tokowicz, 2009), whereas grammatical class refers to whether a word is a noun, verb, etc. (e.g., Vigliocco, Vinson, Druks, Barber & Cappa, 2011). Processing and learning advantages are often found for concrete over abstract words and nouns over verbs on tasks including vocabulary learning, lexical decision, naming, word association, and translation (e.g., de Groot, 2006; de Groot & Keijzer, 2000; Ferré, Sánchez-Casas, Comesaña & Demestre, 2017; Imai et al., 2008; Kaushanskaya & Rechtzigel, 2012; Szekely et al., 2005; van Hell & de Groot, 1998a, 1998b). However, little research has considered how lexical characteristics may interact, how they may differ by task demands (e.g., writing versus speaking; translating into L1 or L2), and what this can reveal about these effects and the nature of language processing. This study therefore investigated the influence of both concreteness and grammatical class on foreign language vocabulary learning.

In the current study, language learning was chosen because of the demonstrated effects of concreteness and grammatical class on word learning, in L1 and L2 (e.g., Bird, Franklin & Howard, 2001; Kaushanskaya & Rechtzigel, 2012; McDonough, Song, Hirsh-Pasek, Golinkoff & Lannon, 2011; van Hell & Candia Mahn, 1997), and because vocabulary learning is a widely used verbal task. For example, concreteness is a significant predictor of age of acquisition for L1 word learning (Bird et al., 2001), and nouns appear earlier and more frequently in children's L1 vocabularies (e.g., Bornstein et al., 2004; Gentner, 1982). Nouns are also learned earlier and more quickly than verbs in lab-based settings (e.g., Childers & Tomasello, 2002; Imai, Haryu & Okada, 2005). For example, van Hell and Candia Mahn (1997) compared the efficacy of rote rehearsal versus the keyword method for foreign language vocabulary learning, and also manipulated the concreteness of the target vocabulary. After training, concrete words were translated more quickly and accurately than abstract words, regardless of learning condition. More recently, Kaushanskaya and Rechtzigel (2012) examined the bilingual word-learning advantage (e.g., Kaushanskaya & Marian, 2009; Papagno & Vallar, 1995) for concrete and abstract nouns. They found a significant advantage for concrete over abstract nouns in both monolinguals and bilinguals, though there was a larger concreteness advantage for bilinguals than monolinguals.

In the current study we also examined the extent to which grammatical class effects are separable from concreteness effects. This question is important because grammatical class effects are often presumed to be driven by differences in concreteness between nouns and verbs (e.g., Bird, Howard & Franklin, 2000; Chiarello, Liu, Shears & Kacinik, 2002): verbs tend to be even more abstract than abstract nouns. We first discuss concreteness effects in nouns then turn to grammatical class effects, finishing with how they may interact.

A number of explanations have been proposed regarding the source of concreteness effects (see Mestres-Missé, Münte & Rodriguez-Fornells, 2014). According to the context availability hypothesis (e.g., Schwanenflugel, Harnishfeger & Stowe, 1988; Schwanenflugel & Shoben, 1983; Schwanenflugel & Stowe, 1989), differences in concrete and abstract words arise because on average, compared to concrete words, it is more difficult to think of a context in which abstract words appear. Supporting this hypothesis, concreteness effects typically disappear when equally supportive contexts are provided (e.g., Schwanenflugel & Shoben, 1983;

Schwanenflugel & Stowe, 1989) or when concrete and abstract words are closely matched on rated context availability (e.g., Schwanenflugel et al., 1988; van Hell & de Groot, 1998b). According to dual-coding theory (e.g., Paivio, 1971; Paivio, 1986, 2006), the difference between concrete and abstract words is more qualitative than quantitative. Specifically, abstract words are represented only in a verbal system, whereas concrete words are also represented in a visual system, and the concreteness advantage derives from this additional representation. The visual representation may confer a particular advantage for memory or other tasks that tap into extended semantic processing (e.g., Tolentino & Tokowicz, 2009). Some researchers have also proposed combining the context availability hypothesis and dual coding theory (the context-extended dual-coding hypothesis) to account for the pattern of event-related potentials (ERPs) found in conjunction with behavioral concreteness advantages. Specifically, although concrete words typically elicit faster responses than abstract words, they also elicit greater N400 effects (typically associated with more effortful semantic processing). To account for this, the context-extended dual-coding hypothesis combines the dual-coding and context availability approaches such that the behavioral concreteness advantage is explained by richer associative networks, and increased N400 effects arise from the use of visual imagery in processing concrete words (e.g., Holcomb, Kounios, Anderson & West, 1999; West & Holcomb, 2000).

Other researchers have hypothesized that concrete words may have more semantic features, and that these features are more consistently accessed, than those of abstract words (de Groot & Keijzer, 2000; Plaut & Shallice, 1993). More recently, Crutch and colleagues have proposed the different organizational frameworks theory (e.g., Crutch, Ridha & Warrington, 2006; Crutch & Warrington, 2005, 2007) to account for verbal behavior in persons with brain lesions and associated conditions, such as semantic refractory access dysphasia. According to this theory, semantic representations for concrete and abstract words are organized in fundamentally different ways: concrete words are organized (more) by similarity (e.g., ‘lake’ and ‘pond’ are semantically similar), whereas abstract words are organized (more) by association (e.g., ‘lake’ and ‘swan’ are associated). This proposal is supported by patient studies showing that semantically similar words interfere more with the processing of concrete words than abstract words, and that semantic associates interfere more with the processing of abstract words than concrete words (e.g., Crutch, 2006; Crutch et al., 2006; Crutch & Warrington, 2005, 2007). However, studies with healthy adults have not replicated these findings (Ferré, Guasch, García-Chico & Sánchez-Casas, 2015; Geng & Schnur, 2015; Zhang, Han & Bi, 2013).

Although concreteness effects are commonly reported, their presence and strength are modulated by other factors. For example, concreteness effects in lexical decision and bilingual word translation disappear when stimuli are tightly controlled to have just a single sense (Tokowicz & Kroll, 2007), and concreteness effects in masked priming are found only with a small range of stimulus-onset-asynchronies (Ferré et al., 2017). Further, the strength of the concreteness effect is also influenced by the number of meanings a word has (Palmer, MacGregor & Havelka, 2013), stimulus order (e.g., Tolentino & Tokowicz, 2009; Zhang et al., 2013), task demands (e.g., Walker & Hulme, 1999; Wang, Conder, Blitzer & Shinkareva, 2010), neuropsychological status (e.g., Breedin, Saffran & Coslett, 1994; Papagno, Capasso & Miceli, 2009), and age and sex (e.g., Dye, Walenski, Prado, Mostofsky & Ullman, 2013; Prado & Ullman, 2009). For example, Romani, McAlpine and Martin (2008) found larger concreteness effects for tasks that require recalling item identity (rather than order) and when phonological representations are less robust. Relatedly, concreteness effects are larger for tasks involving greater semantic processing (James, 1975; West & Holcomb, 2000). Sandberg and Kiran (2014) found greater concreteness effects in more difficult tasks. Other studies have suggested concreteness effects may be more robust for nouns than verbs (Eviatar, Menn & Zaidel, 1990; Zhang, Guo, Ding & Wang, 2006). However, no definitive conclusion has yet been reached regarding the source of the concreteness effect and how or why it varies in such ways.

Concreteness varies across nouns (e.g., Crutch & Jackson, 2011), but also across grammatical classes, such that nouns are generally more concrete than verbs (Bird et al., 2000; Chiarello et al., 2002). This is especially relevant given that grammatical class effects are generally parallel to concreteness effects: nouns typically show advantages over verbs in processing and learning. This grammatical class effect has been documented in various populations, including children, healthy monolingual and bilingual adults, and persons with aphasia. Nouns and object words show an advantage over verbs and action words in picture naming (e.g., Kambanaros, Grohmann, Michaelides & Theodorou, 2014; Mätzig, Druks, Masterson & Vigliocco, 2009; Sheng & McGregor, 2010), and nouns have more associates and are easier to associate to than verbs (van Hell & de Groot, 1998a). Nouns are also easier to learn and are learned earlier in L1 than verbs (e.g., Gentner, 1982, 2006; Waxman et al., 2013). In persons with aphasia and other language disorders, noun-verb dissociations are common, although clean delineations are rare (see Crepaldi, Berlingeri, Paulesu & Luzzatti, 2011; Druks, 2002; Kambanaros & Grohmann, 2015).

The source of grammatical class effects is controversial (Kambanaros & Grohmann, 2015). A number of explanations have been proposed, including that verbs are more syntactically and morphologically complex than nouns (see Druks, 2002; Kauschke & Stenneken, 2008) and that verbs have broader, more context-dependent meanings than nouns (compare ‘*catch* a ball’, ‘*catch* a cold’ vs. ‘throw *a ball*’, ‘bounce *a ball*’) (e.g., Earles & Kersten, 2016; Gentner &

Boroditsky, 2001; Maguire, Hirsh-Pasek & Golinkoff, 2006). A commonly cited explanation for these differences is that nouns are inherently more concrete than verbs (e.g., Barber, Kousta, Otten & Vigliocco, 2010; Bird et al., 2000; Bogka et al., 2003; Bourassa & Besner, 1994; Chiarello et al., 2002; Druks, 2002; Kambanaros & Grohmann, 2015; Kauschke & Stenneken, 2008; Ma, Golinkoff, Hirsh-Pasek, McDonough & Tardif, 2009; Mätzig et al., 2009; McDonough et al., 2011; Vigliocco et al., 2011). Although nouns and verbs vary in their concreteness (Bird et al., 2000; Kambanaros, 2009; Xu, Kang & Guo, 2016), few studies of language processing in healthy adults have considered concreteness and grammatical class simultaneously. Thus, although grammatical class effects are often presumed to be due to concreteness, to date there has been no direct behavioral test of whether grammatical class effects in healthy adults are due to differences in the concreteness of nouns and verbs (though see Berndt, Haendiges, Burton & Mitchum, 2002; Berndt, Mitchum, Haendiges & Sandson, 1997, for relevant studies in persons with aphasia; McDonough et al., 2011, for similar results in child word learning; Bedny & Thompson-Schill, 2006, who found neuroimaging evidence for the separability of grammatical class and concreteness; and Lee & Federmeier, 2008, who found concreteness effects in ERPs for both nouns and verbs). Further, there is little research considering how such effects may vary by task demands, such as recognition versus oral production, despite evidence of variability in concreteness and grammaticality effects across tasks (e.g., Crepaldi et al., 2011; Romani et al., 2008; Wang et al., 2010) and growing recognition of the importance of converging evidence across tasks (e.g., Howard & Gatehouse, 2006; Mätzig et al., 2009).

Given widespread reports of concreteness and grammatical class effects, but limited exploration of whether grammatical class effects go beyond concreteness, the current research was designed to directly address the separability of the grammatical class effect from the concreteness effect. A novel word learning paradigm was used in which the target items were concrete and abstract nouns and verbs. This approach was chosen both because of its prevalence in language learning studies (e.g., de Groot, 2011; de Groot & Keijzer, 2000; Ferré, Ventura, Comesaña & Fraga, 2015) and because of the potential to inform our understanding of the early stages of foreign language vocabulary learning. In Experiment 1, participants learned 'foreign language' words in an unidentified language (actually English pseudowords) and the effects of concreteness and grammatical class were examined during and after training. In Experiment 2 participants learned abstract and concrete German nouns and verbs. To directly test the proposal that differences between nouns and verbs are due to inherent differences in concreteness, in both experiments we used hierarchical regressions to examine the unique contributions of each effect to participants' learning outcomes. Throughout, the separability of the concreteness and grammatical class effects was evaluated in relation to the task demands and measures under consideration: whether participants provided written or spoken translations, and whether this was done from L1 to L2 or vice versa. These variations were considered to determine the extent to which these factors impact the results.

Based on previous research, we expected concrete words and nouns to be learned more quickly and accurately than abstract words and verbs. Although previous research examining both factors simultaneously is limited, findings from persons with aphasia (e.g., Berndt et al., 2002) and other neurophysiological studies (e.g., Bedny & Thompson-Schill, 2006; Lee & Federmeier, 2008) led us to expect no interaction between them. If the grammatical class effect is indeed separable from the concreteness effect, grammatical class should be a significant predictor of accuracy and/or reaction times (RTs) above and beyond concreteness. If, however, the grammatical class effect can be attributed to inherent differences in concreteness between nouns and verbs, grammatical class should not be a significant predictor of learning performance above and beyond concreteness.

Experiment 1

Method

Participants. Twenty-five native English-speaking undergraduates from the University of Pittsburgh contributed data (11 males, 14 females; average age 19.36 years). None had spoken an additional language at home. Data from five additional participants were excluded because they admitted after participating that a language other than English had been spoken at home. All participants reported some experience studying one or more additional languages (most commonly Spanish, French, and German) in a formal classroom setting, beginning at an average age of 13.65 (range: 11-18). Eight participants reported concurrent enrollment in a foreign language, all within the first two years of university study.

Design. A 2 concreteness (concrete vs. abstract) by 2 grammatical class (noun vs. verb) by 3 test (Test 1, Test 2, Test 3) within-subjects design was used.

Materials. Eighty English words were paired with pseudowords taken from Gathercole, Pickering, Hall, and Peaker (2001); these were the target ‘foreign language’ words. All pseudowords were four letters long and phonotactically legal. Items are available in the Supplementary Materials.

There were 20 items each of concrete and abstract nouns and verbs. Part of speech was obtained from Bird et al. (2001) and the MRC psycholinguistic database (Wilson, 1988) and only words whose part of speech was listed as unambiguous were included. Concreteness, age of acquisition, and frequency (from CELEX, Baayen, Piepenbrock & Gulikers, 1995) of the English words were obtained from Bird et al. (2001). Number of letters and number and frequency of orthographic and phonological neighbors were obtained from E-Lexicon (Balota et al., 2007). Number of senses was obtained from WordNet (Princeton, 2010).

Nouns and verbs were matched on age of acquisition, concreteness, frequency, length, and number and frequency of orthographic and phonological neighbors, though verbs had more senses than nouns. Concrete and abstract words were matched on frequency, length, number and frequency of orthographic and phonological neighbors, and number of senses (see Table 1). Because age of acquisition is strongly correlated with concreteness (Bird et al., 2001; Johnston & Barry, 2006) it was not possible to equate concrete and abstract words on this variable. Across conditions, target pseudowords were matched on number of orthographic neighbors and Levenshtein distance between the pseudowords and their English translations. Pseudowords were also matched on bigram sum and frequency across abstract and concrete items. Verb pseudowords had marginally higher bigram sum and frequency than noun pseudowords, but this works against the hypothesized finding of a noun advantage.

<Insert Table 1 about here>

Procedure. Participants were trained and tested on eighty target words. Stimuli were presented and responses recorded using E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA [PST]). Training occurred in three cycles. In each cycle, participants saw one pairing of an English word and its ‘foreign language’ translation side-by-side in the center of the screen, separated by a dash, for five seconds. A one-second blank screen followed, then a fixation cross and a prompt to press a key to continue. The 80 pairs of words were randomized for each participant. After participants had viewed all 80 pairs once, the second presentation was in a new randomized order.

Participants were then tested on their knowledge of the target words. They saw an English word and were asked to type the translation; guessing was encouraged. After they typed their translation and pressed ‘Enter’, participants received feedback in the form “[English word] is [correct translation]”, which remained on the screen for two seconds.

This cycle of two training presentations followed by a typing test was repeated two additional times; participants were encouraged to take a short break between each cycle. The interleaving of study and test trials was based on the testing effect, the finding that including retrieval practice during learning improves retention (e.g., Karpicke & Roediger, 2007; Roediger & Butler, 2011), and previous success using this type of procedure (Palmer et al., 2013). Altogether, participants saw all 80 word pairs six times and were tested three times. After the learning task, participants completed a language history questionnaire (Tokowicz, Michael & Kroll, 2004). The experiment lasted approximately two hours.

Scoring. The accuracy of typed translations was scored in two ways, strict and lenient. Using strict scoring, a translation was considered correct only if it was exactly correct. Using lenient scoring, a ‘percent correct’ for each word was calculated using orthographic similarity between the production and the target (van Orden, 1987). The pattern of results was similar and the scores obtained were highly correlated across the two scoring methods (*r*s .96-.99, *p*s<.001); therefore only the results from strict scoring are presented. Because of the non-speeded nature of the typing task, the RTs (the time between the presentation of the English word and when the participant pressed ‘Enter’ after typing the translation) are less sensitive than traditional RT measures. Results for RTs on the typing task are provided in the Supplementary Materials, for completeness in reporting results.

Analyses. Two sets of analyses were conducted, each to address a specific question. First, 2 concreteness (concrete, abstract) x 2 grammatical class (noun, verb) x 3 test (1, 2, 3) ANOVAs were used to examine the main effects and interactions among these variables[2]. Post-hoc comparisons are reported with Bonferroni-adjusted *p*-values. This step included both subject and item analyses and allowed for a direct examination of the main effects and interactions at each test; these are presented under 'Main effects and interactions'. Second, hierarchical regressions (blockwise entry) were used to determine whether the effect of grammatical class was separable from the effect of concreteness. This involves conducting a regression in several steps; variables on earlier steps account for variance prior to variables entered on subsequent steps, allowing us to determine whether the variables of interest account for additional, independent variance beyond other variables (Cohen, Cohen, West & Aiken, 2003). This second set of analyses focused on the items, and although they drew from the same data as the ANOVA item analyses, hierarchical regression was additionally used because it provides a more stringent control for confounding variables (here, the effects of age of acquisition, length, and frequency). These are presented under 'Separability of the effects of concreteness and grammatical class'.

Results

Main effects and interactions

Descriptive statistics are in Table 2. Concrete words were responded to more accurately than abstract words (54.0% vs. 43.3%), $F_1(1, 24)=99.57$, $MSE=.87$, $p<.001$, $\eta_p^2=.81$, $F_2(1, 76)=15.25$, $MSE=.67$, $p<.001$, $\eta_p^2=.17$. Nouns were responded to more accurately than verbs by participants (50.9% vs. 46.4%), $F_1(1, 24)=9.85$, $MSE=.15$, $p<.01$, $\eta_p^2=.29$, $F_2(1, 76)=2.49$, $MSE=.11$, $p=.12$, $\eta_p^2=.03$. The main effect of test was significant (21.4%, 51.7%, 70.0%), $F_1(2, 48)=124.23$, $MSE=6.07$, $p<.001$, $\eta_p^2=.84$, $F_2(2, 152)=1067.09$, $MSE=4.80$, $p<.001$, $\eta_p^2=.93$.

Paired samples *t*-tests showed that accuracy at Test 2 was higher than at Test 1, $p<.001$, and accuracy at Test 3 was higher than at Test 2, $p<.001$. Grammatical class and test interacted, $F_1(2, 48)=5.68$, $MSE=.03$, $p<.01$, $\eta_p^2=.19$, $F_2(2, 152)=6.27$, $MSE=.03$, $p<.01$, $\eta_p^2=.08$. Simple comparisons showed that accuracy for nouns did not differ from verbs at Test 1, $p=.72$, but responses to nouns were more accurate than to verbs at Test 2, $p<.01$ and Test 3, $p<.05$. No other effects or interactions were significant, *F*s<1.

<Insert Table 2 about here>

Separability of the effects of concreteness and grammatical class

Next, hierarchical regressions were used to determine whether the grammatical class effect was significant after accounting for concreteness differences between nouns and verbs (e.g., Bird et al., 2000; Chiarello et al., 2002; Kauschke & Stenneken, 2008; Ma et al., 2009; McDonough et al., 2011). In these analyses, age of acquisition, length, frequency, and concreteness were continuous variables, and grammatical class was categorical. All effects are in the expected direction unless otherwise noted.

Because of the significant interaction between test and grammatical class, the regression was run on the data from Test 2, the test that showed the greatest difference between nouns and verbs, thus providing the best place to look for a separable grammatical class effect. Table 3 shows the model. Both concreteness and grammatical class showed significant effects regardless of order of entry; the interaction was never significant.

<Insert Table 3 about here>

Summary

The ANOVAs revealed significant main effects of concreteness, grammatical class, and test for accuracy in producing (typing) newly learned foreign language vocabulary. Nouns were produced more accurately than verbs, consistent with past findings. Also consistent with past results, concrete words had an accuracy advantage over abstract words. There was an interaction between grammatical class and test, in which responses to nouns were more accurate than responses to verbs at Tests 2 and 3 (but not Test 1). However, there was no interaction between concreteness and grammatical class.

The regressions provided evidence that both concreteness and grammatical class independently explain variance in learners' performance. Overall, therefore, the results suggest that the effect of grammatical class is separable from concreteness. However, the task of learning highly constrained pseudowords as target vocabulary in an unspecified 'foreign language' was somewhat unnatural, which could raise questions about the external validity and replicability of the results. Therefore, Experiment 2 was conducted to determine whether the results would replicate when a real foreign language, German, was used.

Experiment 2

The core features of Experiment 1 (interleaved training and testing, use of the typing test, scoring method) were maintained in Experiment 2 to facilitate comparisons between the two. However, some changes were made to improve the naturalness of the learning task and expand the scope of inquiry. First, real German words were used as the target items-to-be-learned, and the number of targets was reduced from 80 to 60 to reduce the learning burden, particularly to offset the increased length of the words. Although the lengths and phonotactics of the targets were more variable in Experiment 2, the result was a more realistic vocabulary learning experience, improving the external validity of the results. Learning a real language should also increase participants' motivation for engaging fully in the task. Participants were instructed to say aloud the English-German word pairs twice to make the task more realistic (learning pronunciations and the mappings between written forms and pronunciations is a normal aspect of vocabulary learning, Ehri, 2014; Everson, 1998), and to support the development of new lexical entries for the somewhat longer, more difficult German targets (Ehri, 2014). Further, vocabulary learning studies have demonstrated that saying aloud target words-to-be-learned enhances the learning of semantic representations in addition to orthographic forms (Everson, 1998; Rosenthal & Ehri, 2011; see also Perfetti, 2007; Perfetti & Hart, 2001). Although it is likely that participants did not pronounce the German words completely accurately, research has also demonstrated that pronunciation approximations and non-phonetic, spelling-based pronunciations are equally – or more – effective for supporting vocabulary learning than correct phonetic pronunciations (Drake & Ehri, 1984; Ehri, 1984; Rosenthal & Ehri, 2011; see also Birch, 2015). Finally, the instruction to say aloud the word pair was meant to minimize individual differences in the degree to which participants completely decode (or not) new vocabulary, which substantially impacts vocabulary learning (Rosenthal & Ehri, 2011).

Other changes involved the collection of additional data. Specifically, an L2 to L1 oral translation task was added after the typing task. This backward translation task is generally considered easier than the forward translation required by the typing task, and is the most widely used task to measure vocabulary learning after experimental training (e.g., de Groot, 2011; Ferré, Ventura, et al., 2015; Kaushanskaya & Rechtzigel, 2012). It also has the advantage of providing a more sensitive measure of RTs than the typing task, and was thus expected to provide useful data. The final change was the addition of a second session, after one week, which included the same tasks as the first (typing, L2 to L1 production) and was added to examine consolidated vocabulary retention following sleep. This was of interest based on previous findings that some semantic effects become evident only after consolidation (e.g., Comesaña, Perea, Piñeiro & Fraga, 2009; Davis, Di Betta, Macdonald & Gaskell, 2009; Ferré, Ventura, et al., 2015).

Method

Participants. Forty native English-speaking undergraduates from the University of Pittsburgh contributed data (15 males, 22 females, 3 declined to respond; average age 19.1 years). Data from ten additional participants were excluded because they admitted after participating that they had previous experience with German or Dutch (seven participants) or that English was not their only L1 (three participants). All participants reported some experience studying one or more additional languages (most commonly Spanish and French) in a formal classroom setting, beginning at an average age of 13.45 (range: 10-17). Six participants reported concurrent enrollment in a foreign language, all within the first two years of university study.

Design. A 2 concreteness (concrete vs. abstract) by 2 grammatical class (noun vs. verb) by 2 session (Session 1 vs. Session 2) within-subjects design was used first on each language test, to determine whether analyses should be conducted separately for the two sessions. Following this, a 2 concreteness (concrete vs. abstract) by 2 grammatical class (noun vs. verb) by 3 test (Test 1, Test 2, Test 3) within-subjects design was used to analyze the typing data from Session 1, as in Experiment 1.

Materials. Participants learned 60 German words, 15 each of concrete and abstract nouns and verbs. Concreteness was determined using ratings from Bird et al. (2001) and only words with an unambiguous part of speech according to E-Lexicon (Balota et al., 2007) were included. Number of senses was obtained from WordNet (Princeton, 2010). Across conditions, words did not differ in English or German length, English word frequency (from CELEX, Baayen et al., 1995), English orthographic and phonological neighborhood sizes and the frequency of neighbors, and the sum and mean of their English bigram statistics (from E-Lexicon, Balota et al., 2007). Nouns and verbs were additionally matched on their English age of acquisition and concreteness (see Table 4), though again verbs had more senses than nouns. Concrete and abstract words were matched on number of senses. Items were chosen to avoid cognates, and across conditions the target German words were matched on their Levenshtein distance from the English translations as well as the number of English neighbors and their (English) bigram statistics (from CLEARPOND, Marian, Bartolotti, Chabal & Shook, 2012). Items are available in the Supplementary Materials.

<Insert Table 4 about here>

Learning was assessed using two measures. In the typing task, participants saw an English word and were asked to type its German translation, as in Experiment 1. In the L2 to L1 production task, participants saw a German word and were asked to say its English translation aloud.

Procedure. As in Experiment 1, training occurred in three cycles: two presentations of target words and their English equivalents, followed by a typing production test, repeated three times. Stimuli were presented and responses were recorded using E-Prime 2.0 (PST) and a digital voice recorder. In Session 1, participants began with the vocabulary training. They saw one English word paired with its German translation (e.g., “remit – anweisen”) in the center of the screen and were instructed to say the pair aloud two times. Words were presented in their base form, in all lowercase letters (except for German nouns, which by rule begin with a capital letter). The words remained on the screen for 4250 ms[3], after which they disappeared until the participant manually advanced. They proceeded through all sixty pairs once, then repeated them in a different random order. After these two presentations, participants completed the typing test.

They saw the English word and were asked to type its German translation; guessing was encouraged. After each response, participants received feedback in the form “[English word] is [German word]”. They repeated this sequence two additional times. This training lasted approximately 50 minutes. After the training, participants completed the L2 to L1 production task. Participants’ responses were recorded for off-line accuracy scoring. This concluded Session 1.

In Session 2, participants began by completing the typing task then the L2 to L1 production task again, without review. At the end[4] they completed a language history questionnaire to provide demographic data and information on their language learning experiences (Tokowicz et al., 2004).

Scoring. To facilitate comparability between experiments the accuracy of participants’ typed translations was scored using the strict criteria as described for Experiment 1. For L2 to L1 production, responses were recorded using a digital voice recorder and RTs were measured using a microphone with a voice key trigger. RTs for trials on which participants stuttered, coughed, or otherwise triggered the voice key prior to their response were removed. Responses were also excluded if the RT was greater than three standard deviations from an individual participant’s mean RT; this resulted in excluding 0.59% of the data on Session 1 and 0.30% of the data on Session 2. In total, 2.16% of data were excluded.

Analyses. As with Experiment 1, two sets of analyses were conducted: ANOVAs were used to examine the main effects and interactions of concreteness, grammatical class, and test (for the Session 1 typing data only) on accuracy (for both vocabulary measures) and RTs (for L2 to L1 production). Post-hoc comparisons are reported with Bonferroni-adjusted *p*-values. Hierarchical regressions were used to test for the separability of the grammatical class and concreteness effects. The results are presented by task: data from the typing task, which required written production, are described first; this is the same task as in Experiment 1. Following this are the results for the L2 to L1 production task, new to Experiment 2, which required *oral* production.

Results

Main effects and interactions

Typing task accuracy. Descriptive statistics are in Tables 5 and 6. The three-way interaction between concreteness, grammatical class, and session was not significant, $F_1(1, 30)=1.14$, $MSE<.01$, $p=.29$, $\eta_p^2=.04$, $F_2<1$. The two-way interaction between concreteness and grammatical class was also not significant, $F_1(1, 30)=2.13$, $MSE=.01$, $p=.15$, $\eta_p^2=.07$, $F_2<1$. However, session interacted with both concreteness, $F_1(1, 30)=9.97$, $MSE=.02$, $p<.01$, $\eta_p^2=.25$, $F_2(1, 56)=4.12$, $MSE=.01$, $p<.05$, $\eta_p^2=.07$, and grammatical class, $F_1(1, 31)=9.70$, $MSE=.03$, $p<.01$, $\eta_p^2=.24$, $F_2(1, 56)=5.43$, $MSE=.02$, $p<.05$, $\eta_p^2=.09$. Therefore, separate analyses were conducted for each session, and test was included as a factor in the Session 1 analyses, as in Experiment 1.

<Insert Table 5 about here>

<Insert Table 6 about here>

On Session 1, concrete words were responded to more accurately than abstract words by subjects (17.8% vs. 11.5%), $F_1(1, 30)=49.87$, $MSE=.37$, $p<.001$, $\eta_p^2=.62$, $F_2(1, 56)=2.26$, $MSE=.18$, $p=.14$, $\eta_p^2=.04$. Nouns were responded to more accurately than verbs (20.6% vs. 9.9%), $F_1(1, 30)=64.79$, $MSE=.99$, $p<.001$, $\eta_p^2=.68$, $F_2(1, 56)=6.40$, $MSE=.52$, $p<.05$, $\eta_p^2=.10$. The main effect of test was significant (4.0%, 15.3%, 24.7%), $F_1(2, 60)=82.84$, $MSE=1.33$, $p<.001$, $\eta_p^2=.73$, $F_2(2, 112)=81.47$, $MSE=.70$, $p<.001$, $\eta_p^2=.59$. Paired samples *t*-tests showed that accuracy at Test 2 was higher than at Test 1, p<.001, and accuracy at Test 3 was higher than at Test 2, p<.001. Concreteness and test interacted, $F_1(2, 60)=25.77$, MSE=.10, p<.001, η_p^2=.46, $F_2(2, 112)=5.34$, MSE=.05, p<.01, η_p^2=.09. Simple comparisons showed accuracy for concrete words was not significantly different from abstract words at Test 1, p>.05, but responses to concrete words were more accurate than to abstract words at Test 2 (6.1%), p<.001 and Test 3 (12.0%), p<.001. Grammatical class and test also interacted, $F_1(2, 60)=17.06$, MSE=.09, p<.001, η_p^2=.36, $F_2(2, 112)=5.41$, MSE=.05, p<.01, η_p^2=.09. Simple comparisons revealed that accuracy for nouns was higher than accuracy for verbs at all three tests, ps<.001, and that the size of the noun advantage increased with each test (4.4%, 11.7%, 14.7%). No other effects or interactions were significant, Fs<1.

On Session 2, participants completed the typing task only once, so test was not a factor. Concrete words were responded to more accurately than abstract words by subjects (11.7% vs. 8.9%), $F_1(1, 30)=5.29$, MSE=.02, p<.05, η_p^2=.15, F_2<1. Nouns were responded to more accurately than verbs by subjects (12.9% vs. 6.6%), $F_1(1, 30)=27.57$, MSE=.12, p<.001, η_p^2=.48, $F_2(1, 56)=2.77$, MSE=.06, p=.10, η_p^2=.05. Concreteness and grammatical class did not interact, $F_1(1, 30)=2.25$, MSE=.02, p=.14, F_2<1.

L2 to L1 production accuracy. The three-way interaction between concreteness, grammatical class, and session, as well as all two-way interactions, were not significant, Fs<1. Thus, the data were collapsed across session. Concrete words were responded to more accurately than abstract words (43.8% vs. 29.4%), $F_1(1, 20)=48.53$, MSE=.87, p<.001, η_p^2=.71, $F_2(1, 56)=7.13$, MSE=.61, p<.05, η_p^2=.11. Nouns were responded to more accurately than verbs (42.3% vs. 21.9%), $F_1(1, 20)=84.07$, MSE=2.03, p<.001, η_p^2=.81, $F_2(1, 56)=14.52$, MSE=1.25,

p<.001, η_p^2=.21. Responses were more accurate on Session 1 than Session 2 (37.5% vs. 26.8%), $F_1(1, 20)$=31.54, MSE=.53, p<.001, η_p^2=.61, $F_2(1, 56)$=124.99, MSE=.34, p<.001, η_p^2=.69.

L2 to L1 production RTs. The three-way interaction between concreteness, grammatical class, and session was not significant, $F_1(1, 14)$=1.19, MSE=564,141.14, p=.29, η_p^2=.08, F_2<1. The two-way interaction between concreteness and grammatical class was marginally significant by subjects, $F_1(1, 14)$=3.50, MSE=330,810.46, p=.08, η_p^2=.20, F_2<1. Session interacted with concreteness, $F_1(1, 14)$=7.05, MSE=3,186,285.64, p<.05, η_p^2=.34, $F_2(1, 49)$=9.97, MSE=3,664,710.96, p<.01, η_p^2=.17, and interacted marginally with grammatical class by subjects, $F_1(1, 14)$=4.31, MSE=2,295,169.21, p=.06, η_p^2=.24, $F_2(1, 49)$=1.71, MSE=626,401.60, p=.20, η_p^2=.03. Therefore, separate analyses were conducted for each session.

On Session 1, concrete words were responded to more quickly than abstract words (2384.73 ms vs. 2974.65 ms), $F_1(1, 26)$=8.40, MSE=9,396,335.05, p<.01, η_p^2=.24, $F_2(1, 56)$=12.08, MSE=6,094,199.07, p<.01, η_p^2=.18. There was a trend for nouns to have faster RTs than verbs (2746.50 ms vs. 2958.99 ms), but it was not significant, F_1<1, $F_2(1, 56)$=1.34, p=.25, η_p^2=.02. Concreteness and grammatical class did not interact, Fs<1.

On Session 2, the main effect of concreteness was not significant, F_1<1, $F_2(1, 49)$=1.19, p=.28. The main effect of grammatical class was also not significant, $F_1(1, 14)$=1.43, p=.25, η_p^2=.09, F_2<1. The interaction was not significant, Fs<1.

Separability of the effects of concreteness and grammatical class

Typing task accuracy. For Session 1, regressions were conducted on the data from Test 3 only; this test showed the largest concreteness effect, thus providing the strictest test for a separable grammatical class effect. Grammatical class accounted for unique variance, whereas concreteness and the interaction did not; the same pattern of results held when concreteness was entered after grammatical class. A similar pattern was obtained on Session 2. In this case, grammatical class was only marginally significant, whereas concreteness and the interaction were not significant, regardless of order. Table 7 shows the model details.

<Insert Table 7 about here>

L2 to L1 production accuracy. L2 to L1 production was performed only once per session, thus test was not a factor. Grammatical class accounted for unique variance and concreteness accounted for a marginally significant portion of variance. The same pattern of results held when concreteness was entered after grammatical class. Similar results were found for Session 2. Both concreteness and grammatical class each accounted for significant unique variance, but the interaction did not. The same pattern held when concreteness was entered after grammatical class. Table 8 shows the model details.

<Insert Table 8 about here>

L2 to L1 production RTs. For Session 1 RTs, concreteness accounted for significant variance when entered before grammatical class but only marginally significant variance when entered after it. Grammatical class and the interaction did not account for significant variance. For Session 2, none of the predictors accounted for significant variance. Table 9 shows the model details.

<Insert Table 9 about here>

Summary

The ANOVAs revealed significant main effects of concreteness and grammatical class on accuracy for both tasks (written and oral production) on both sessions, though some effects (particularly Session 2 typing) were not significant by items. Given this finding, it is worth noting that concreteness may be more reliably significant across participants than items (see

Tokowicz & Kroll, 2007, for a similar pattern of results).[5] Regardless, in each case, concrete words and nouns had higher accuracy than abstract words and verbs. These results mirror Experiment 1 and are consistent with previous studies. An interaction with test was replicated: in Experiment 1, there was a noun advantage at Tests 2 and 3, but not Test 1. In Experiment 2 there was a noun advantage at all three tests, as well as an advantage of concrete words over abstract words at Tests 2 and 3 (but not Test 1). The RTs from L2 to L1 production (unique to Experiment 2) also showed a concrete-word advantage on Session 1, though there were no significant effects on Session 2.

The regression results were largely consistent across sessions. Grammatical class accounted for independent variance beyond that accounted for by concreteness on most variables: written production (typing) accuracy and oral (L2 to L1) production accuracy (though not RTs). This is again similar to Experiment 1, in which the grammatical class effect was separable from concreteness for typing accuracy. For L2 to L1 production (only in Experiment 2), concreteness explained unique variance in Session 1 RTs in Session 1 but neither concreteness nor grammatical class was a significant predictor of Session 2 RTs.

Overall the results from Experiment 2 replicate those from Experiment 1: significant effects of both concreteness and grammatical class on vocabulary learning accuracy and separability of the grammatical class effect from the concreteness effect on most measures. Experiment 2 also extended these results to accuracy on oral (L2 to L1) production.

General Discussion

The results of this study demonstrate that both concreteness and grammatical class influence early foreign language vocabulary learning. Consistent with previous research on concreteness and grammatical class effects (e.g., Allen & Hulme, 2006; de Groot & Poot, 1997;

Imai et al., 2008; Kaushanskaya & Rechtzigel, 2012), accuracy was generally higher and RTs faster for concrete than abstract words, and accuracy was higher for nouns than verbs. Additionally, the results showed that both effects appear quite early in the learning process. This is consistent with recent research demonstrating an early emergence of semantic effects in word learning (e.g., Altarriba & Basnight-Brown, 2012; Ferré, Ventura, et al., 2015; Palmer et al., 2013). Thus, evidence from both the current and previous studies suggests that when learning new (foreign language) words for existing concrete or abstract concepts, they are integrated into the semantic system relatively quickly.

The current results also provide important evidence that challenges a strict concreteness-based explanation for the source of noun/verb differences in language tasks. They thus provide support, at least in part, for more grammatically-based accounts (see Kambanaros & Grohmann, 2015, for discussion; we return to this point later). This finding of a grammatical class effect that is clearly separable from concreteness has not been demonstrated previously in behavioral data from healthy adults. Indeed, most previous research examining concreteness or grammatical class effects either confounded these factors or ignored their variance across words, for example by studying concrete words or nouns only (e.g., Crutch & Jackson, 2011; de Groot, 2006; Ferré, Ventura, et al., 2015; Kambanaros et al., 2014; McDonough et al., 2011; van Hell & Candia Mahn, 1997; Zhang et al., 2013).

There are a handful of previous studies with other populations (persons with aphasia, child L1 acquisition) or methodologies (neuroimaging, ERPs) that also provide evidence of the separability of concreteness and grammatical class effects. In a study of persons with aphasia, Berndt et al. (1997) found similar effects of imageability for both nouns and verbs, and found that performance was generally better for abstract nouns than for imageable verbs, suggesting that the noun-verb difference was due to a factor beyond imageability. Similarly, Berndt et al. (2002) found that even when nouns and verbs were matched on imageability, persons with aphasia still processed nouns more accurately than verbs. In a study on child L1 word learning, McDonough et al. (2011) examined the impact of imageability, grammatical class, and word frequency on words' age of acquisition. Grammatical class was a significant predictor of age of acquisition, and imageability accounted for additional unique variance beyond grammatical class. Unfortunately, the authors did not report whether they tested the predictors in the opposite order, making it unclear whether grammatical class would be a significant predictor of age of acquisition beyond imageability. Szekely et al. (2005) examined the differences between object and action naming and concluded that there are "profound" differences (p. 21) between naming actions and objects that cannot be accounted for by factors such as semantic complexity, frequency, and age of acquisition.

Using neuroimaging methods, Bedny and Thompson-Schill (2006) used semantic similarity judgments to examine the neural correlates of imageability in nouns and verbs. They found some brain regions that responded to imageability similarly across both nouns and verbs, as well as anatomically separate effects of imageability and grammatical class. These results suggest that grammatical class effects cannot be attributed solely to differences in imageability. Lee and Federmeier (2008) examined the concreteness effect in ERPs, specifically the semantically-sensitive N400 component, and similarly found concreteness effects for both nouns and verbs. Overall, then, the results from previous studies with different populations and different methodologies are consistent with the current results, adding support to the finding that the grammatical class effect is not due solely to differences in concreteness between nouns and verbs.

It should be noted that although the results were generally consistent across the two experiments reported here, there were some minor differences. For example, the typing test cycle with the largest noun-verb accuracy difference was Test 2 for Experiment 1 but Test 3 for Experiment 2 (a numerical trend, not confirmed statistically). One possible reason for the slightly differing results across experiments stems from the different types of stimuli used and the somewhat different nature of the learning task. In Experiment 1, items were a uniform four letters long; this is a relatively easy, but unrealistic, learning task. In contrast, in Experiment 2 the target items were real German words and varied much more widely in length. Not only is learning authentic German words likely a more difficult task than learning short English pseudowords, the German words were on average over twice as long (8.6 letters) as the pseudowords from Experiment 1. Although the items were matched on English and German word length when comparing concrete versus abstract words and nouns versus verbs, the greater length of the items in Experiment 2 may have led to an overall more difficult task, thus impacting the concreteness and grammatical class effects. This difficulty explanation is supported by the fact that when the target items were real German words, a significant effect of grammatical class appeared already at Test 1 and grew in size for each test cycle. It should also be noted that although the noun-verb difference was numerically smaller at Test 3 than at Test 2 for Experiment 1, both effects were still significant.

Although there were some differences between the two experiments (number and type of target items, instructions to say the items aloud, addition of the L2 to L1 production task and second session), the vocabulary training and typing task were fundamentally the same across both experiments and thus the results from the typing task can be directly compared. In fact, finding the same pattern of results across both experiments, despite the differences between them, demonstrates the robustness of a separable grammatical class effect.

One possible limitation of the current results relates to the difference in the number of senses between the nouns and verbs. As is common (e.g., Gentner, 1981), the verbs used in these experiments typically had more senses than the nouns, raising the possibility that this contributed to the grammatical class effect. The picture is complicated, however, by inconsistent previous findings regarding the impact of number of senses. This variable is modulated by other factors such as task demands, relatedness of meanings, method for defining number of senses, and type of distractors; in some cases, words with more senses actually show an advantage (e.g., Azuma & Van Orden, 1997; Borowsky & Masson, 1996; Haro & Ferré, 2018; Piercey & Joordens, 2000; Rodd et al., 2012; Rodd, Gaskell & Marslen-Wilson, 2002). Thus, it cannot be definitively concluded whether the larger number of senses for verbs in this study influenced the grammatical class effect. Regardless, such an influence is not incompatible with our broader conclusion that the grammatical class effect cannot be attributed solely to concreteness differences. Further, previous research on the concreteness effect has indicated that the effect is attenuated or absent when experimental items are limited to words with a single sense (Basnight-Brown & Altarriba, 2016; Tokowicz & Kroll, 2007). If we had limited the words used in these experiments to only those with a single sense, this would have increased the risk of not finding a significant concreteness effect, thus making it more difficult to clearly demonstrate a grammatical class effect that is significant above and beyond concreteness. Finally, it should be noted that although WordNet is widely used for semantically-related research, it has been criticized for parsing senses in too fine-grained a way (McCarthy, 2006), raising the possibility that the greater number of senses for verbs in these experiments was partially an artifact of the database. Future work is needed to clarify these issues.

Overall, the results from the current study provide strong evidence that, in neurologically healthy adults, the grammatical class effect is distinct from the concreteness effect and cannot be accounted for exclusively by inherent concreteness differences between nouns and verbs. This is consistent with previous research using different populations and methodologies, and also with previous findings demonstrating that, though robust, concreteness effects may be less robust than other word-level effects such as cognate status (Ferré et al., 2017). However, ruling out a strict concreteness-based explanation still leaves a number of possibilities for the source(s) of the grammatical class effect, some of which the current data can address.

At the conceptual/semantic level, other differences between nouns and verbs may be a factor, such as the type of representational features (perceptual versus associative; Druks, 2002; Kambanaros & Grohmann, 2015; Mätzig et al., 2009), type of semantic organization (hierarchical versus matrix-like; Black & Chiat, 2003; Druks et al., 2006; Earles & Kersten, 2016), or number of senses (though this research has focused primarily on proficient bilinguals translating ambiguous words; e.g., Basnight-Brown & Altarriba, 2016; Boada, Sánchez-Casas, Gavilán, García-Albea & Tokowicz, 2012; Laxén & Lavaur, 2009). Although these factors may contribute to the grammatical class effect, the finding that lexical effects are less influential on receptive and comprehension-based tasks (Romani et al., 2008; Szekely et al., 2005; Vigliocco et al., 2011; Wang et al., 2010) suggests that the primary source of the grammatical class effect is *not* at the conceptual level (Howard & Gatehouse, 2006; Kambanaros, 2009; Kambanaros & Grohmann, 2015), because a conceptual-level source should lead to similar difficulties or differences across tasks and modalities. This conclusion, that conceptual differences do not primarily drive the grammatical class effect, is consistent with the present results.

Beyond the conceptual level, many studies of the grammatical class effect have adopted Levelt's speech production framework (Levelt, 1989; Levelt, Roelofs & Meyer, 1999) and used patterns of naming errors to determine whether grammatical class effects were the result of deficits at the lemma or lexeme levels of representation (e.g., Hanley, Hunt, Steed & Jackman, 2013; Kambanaros, 2009; Kambanaros, Grohmann & Michaelides, 2013). Because the lemma level is associated with grammatical encoding, difficulties there should lead to grammatical class errors (Kambanaros, 2009; Kambanaros et al., 2014; Kambanaros & van Steenbrugge, 2006); because the lexeme level is associated with morpho-phonological encoding, difficulties there should lead to morphological and/or phonological errors (Hanley et al., 2013; Kambanaros et al., 2014). However, these expected patterns are frequently not found, suggesting that lemma- and lexeme-level representations are not the source of grammatical class effects (Kambanaros, 2009; Kambanaros & Grohmann, 2015; Kambanaros et al., 2014). This is also consistent with the current results. In these experiments neither the target items nor their English translations were inflected and there was no explicit instruction or focus on phonological representations for the target words. Thus, difficulties with the phonological or morphological representations could not have driven the separable grammatical class effect.

In previous studies, some researchers have concluded that the grammatical class effect may derive from access to, and the connections between, the lemma and lexeme levels, rather than the representation at either level per se (e.g., Hanley et al., 2013; Howard & Gatehouse, 2006; Kambanaros & Grohmann, 2015; Kambanaros et al., 2013; Kambanaros et al., 2014). Though most of these studies have used Levelt's (1989; Levelt et al., 1999) model as their theoretical framework, connectionist computational modeling has also shown that the links between meaning and form can be impacted separately from either representation (Li & Farkaš, 2002; Li, Farkaš & MacWhinney, 2004). Although the current results do not provide direct evidence for this explanation, they are not inconsistent with it. Future research, using converging evidence from multiple tasks with varying levels of language involvement, is needed to evaluate this possibility. At the same time, it is important to remember that the source(s) of the grammatical class effect are likely multiple and variable across individuals (Breedin, 1996; Crepaldi et al., 2011; Howard & Gatehouse, 2006; Kambanaros, 2009; Kambanaros & Grohmann, 2010; Kambanaros & Grohmann, 2015; Kambanaros et al., 2014; Li et al., 2004; Mätzig et al., 2009; Vigliocco et al., 2011).

To conclude, the current research provides the first direct behavioral evidence in healthy adults of the separability of the grammatical class effect from the concreteness differences that exist between nouns and verbs (e.g., Bird et al., 2000; Chiarello et al., 2002; Kauschke & Stenneken, 2008). The results challenge a strict concreteness-based explanation for grammatical class effects and suggest that additional factors, such as the strength of connections between levels of lexical representation (e.g., lemma and lexeme), must be at play. Along with previous studies, the current study suggests that the appearance and relations among these two effects may vary by language measure and task demands, providing an important avenue for future research. In fact, Crepaldi et al. (2011) suggested that the level of processing involved in a task (e.g., semantic processing vs. phonological encoding) may interact with the appearance of grammatical class effects. If this is the case, the grammatical class effects found in the current study (which were often stronger than the concreteness effects, despite the use of semantically-demanding translation tasks) suggest that the involvement of semantics (beyond concreteness) may be a crucial component of the source of grammatical class effects (and would be consistent with the suggestion that the connections between the lemma and lexeme levels of lexical representations are implicated). Examining this issue further could be accomplished, for example, by examining concreteness and grammatical class effects in clinical populations with deficits in varying aspects of lexical processing (e.g., Crutch, Connell & Warrington, 2009; Crutch & Jackson, 2011; Crutch & Warrington, 2007), comparing concreteness and grammatical class effects with items differing in morphological richness (Longe, Randall, Stamatakis & Tyler, 2007), modifying target words to increase or decrease participants' reliance on phonological representations for memory (e.g., Romani et al., 2008), increasing or decreasing the level of semantic access required for task completion (e.g., Borowsky & Masson, 1996), or examining the impact of individual differences in cognitive abilities such as phonological short-term memory on these lexical effects. If task-based or level-of-processing dissociations are indeed confirmed, this may provide important insight into the source and development of grammatical class effects, how they differ from concreteness effects, and the underlying natures of bilingual lexical representations and lexical access that lead to such effects.

References

Allen, R., & Hulme, C. (2006). Speech and language processing mechanisms in verbal serial recall. *Journal of Memory and Language, 55*(1), 64-88. doi:http://dx.doi.org/10.1016/j.jml.2006.02.002

Altarriba, J., & Basnight-Brown, D. M. (2012). The acquisition of concrete, abstract, and emotion words in a second language. *International Journal of Bilingualism, 16*(4), 446-452. doi:10.1177/1367006911429511

Azuma, T., & Van Orden, G. C. (1997). Why SAFE is better than FAST: The relatedness of a word's meanings affects lexical decision times. *Journal of Memory and Language, 36*(4), 484-504. doi:https://doi.org/10.1006/jmla.1997.2502

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*(4), 390-412. doi:https://doi.org/10.1016/j.jml.2007.12.005

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research, 3*(2), 12-28. doi:http://dx.doi.org/10.21500/20112084.807

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX Lexical Database. Release 2 (CD-ROM)*: Linguistic Data Consortium, University of Pennsylvania.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39*(3), 445-459.

Barber, H. A., Kousta, S.-T., Otten, L. J., & Vigliocco, G. (2010). Event-related potentials to event-related words: Grammatical class and semantic attributes in the representation of knowledge. *Brain Research, 1332*, 65-74. doi:http://dx.doi.org/10.1016/j.brainres.2010.03.014

Basnight-Brown, D. M., & Altarriba, J. (2016). Multiple translations in bilingual memory: Processing differences across concrete, abstract, and emotion words. *Journal of Psycholinguistic Research, 45*(5), 1219-1245. doi:10.1007/s10936-015-9400-4

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4 (Version R package version 1.1-11): http://cran.r-project.org/package=lme4.

Bedny, M., & Thompson-Schill, S. L. (2006). Neuroanatomically separable effects of imageability and grammatical class during single-word comprehension. *Brain and Language, 98*(2), 127-139. doi:https://doi.org/10.1016/j.bandl.2006.04.008

Berndt, R. S., Haendiges, A. N., Burton, M. W., & Mitchum, C. C. (2002). Grammatical class and imageability in aphasic word production: Their effects are independent. *Journal of Neurolinguistics, 15*(3–5), 353-371. doi:http://dx.doi.org/10.1016/S0911-6044(01)00030-6

Berndt, R. S., Mitchum, C. C., Haendiges, A. N., & Sandson, J. (1997). Verb retrieval in aphasia. 1. Characterizing single word impairments. *Brain and Language, 56*(1), 68-106. doi:http://dx.doi.org/10.1006/brln.1997.1727

Birch, B. M. (2015). *English L2 reading: Getting to the bottom*. New York, NY: Routledge.

Bird, H., Franklin, S., & Howard, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behavior Research Methods, Instruments & Computers, 33*(1), 73-79.

Bird, H., Howard, D., & Franklin, S. (2000). Why is a verb like an inanimate object? Grammatical category and semantic category deficits. *Brain and Language, 72*(3), 246-309. doi:http://dx.doi.org/10.1006/brln.2000.2292

Black, M., & Chiat, S. (2003). Noun–verb dissociations: A multi-faceted phenomenon. *Journal of Neurolinguistics, 16*(2–3), 231-250. doi:http://dx.doi.org/10.1016/S0911-6044(02)00017-9

Boada, R., Sánchez-Casas, R., Gavilán, J. M., García-Albea, J. E., & Tokowicz, N. (2012). Effect of multiple translations and cognate status on translation recognition performance of balanced bilinguals. *Bilingualism: Language and Cognition, 16*(1), 183-197. doi:10.1017/S1366728912000223

Bogka, N., Masterson, J., Druks, J., Fragkioudaki, M., Chatziprokopiou, E.-S., & Economou, K. (2003). Object and action picture naming in English and Greek. *European Journal of Cognitive Psychology, 15*(3), 371-403. doi:10.1080/09541440303607

Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S.-Y., Pascual, L., . . . Vyt, A. (2004). Cross-linguistic analysis of vocabulary in young children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English. *Child Development, 75*(4), 1115-1139. doi:10.1111/j.1467-8624.2004.00729.x

Borowsky, R., & Masson, M. E. J. (1996). Semantic ambiguity effects in word identification. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 22*(1), 63-85.

Bourassa, D. C., & Besner, D. (1994). Beyond the articulatory loop: A semantic contribution to serial order recall of subspan lists. *Psychonomic Bulletin & Review, 1*(1), 122-125. doi:10.3758/bf03200768

Breedin, S. D. (1996). Patterns of verb impairment in aphasia: An analysis of four cases. *Cognitive Neuropsychology, 13*(1), 51-92. doi:10.1080/026432996382060

Breedin, S. D., Saffran, E. M., & Coslett, H. B. (1994). Reversal of the concreteness effect in a patient with semantic dementia. *Cognitive Neuropsychology, 11*(6), 617-660. doi:10.1080/02643299408251987

Chiarello, C., Liu, S., Shears, C., & Kacinik, N. (2002). Differential asymmetries for recognizing nouns and verbs: Where are they? *Neuropsychology, 16*(1), 35-48. doi:10.1037/0894-4105.16.1.35

Childers, J. B., & Tomasello, M. (2002). Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental Psychology, 38*(6), 967-978.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. New York, NY: Routledge.

Comesaña, M., Perea, M., Piñeiro, A., & Fraga, I. (2009). Vocabulary teaching strategies and conceptual representations of words in L2 in children: Evidence with novice learners. *Journal of Experimental Child Psychology, 104*(1), 22-33. doi:http://dx.doi.org/10.1016/j.jecp.2008.10.004

Crepaldi, D., Berlingeri, M., Paulesu, E., & Luzzatti, C. (2011). A place for nouns and a place for verbs? A critical review of neurocognitive data on grammatical-class effects. *Brain and Language, 116*(1), 33-49. doi:https://doi.org/10.1016/j.bandl.2010.09.005

Crutch, S. J. (2006). Qualitatively different semantic representations for abstract and concrete words: Further evidence from the semantic reading errors of deep dyslexic patients. *Neurocase, 12*(2), 91-97. doi:10.1080/13554790500507172

Crutch, S. J., Connell, S., & Warrington, E. K. (2009). The different representational frameworks underpinning abstract and concrete knowledge: Evidence from odd-one-out judgements.

The Quarterly Journal of Experimental Psychology, 62(7), 1377-1390. doi:10.1080/17470210802483834

Crutch, S. J., & Jackson, E. C. (2011). Contrasting graded effects of semantic similarity and association across the concreteness spectrum. *The Quarterly Journal of Experimental Psychology, 64*(7), 1388-1408. doi:10.1080/17470218.2010.543285

Crutch, S. J., Ridha, B. H., & Warrington, E. K. (2006). The different frameworks underlying abstract and concrete knowledge: Evidence from a bilingual patient with a semantic refractory access dysphasia. *Neurocase, 12*(3), 151-163. doi:10.1080/13554790600598832

Crutch, S. J., & Warrington, E. K. (2005). Abstract and concrete concepts have structurally different representational frameworks. *Brain, 128*(3), 615-627. doi:http://dx.doi.org/10.1093/brain/awh349

Crutch, S. J., & Warrington, E. K. (2007). Semantic priming in deep-phonological dyslexia: Contrasting effects of association and similarity upon abstract and concrete word reading. *Cognitive Neuropsychology, 24*(6), 583-602. doi:10.1080/02643290701577351

Davis, M. H., Di Betta, A. M., Macdonald, M. J. E., & Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *Journal of Cognitive Neuroscience, 21*(4), 803-820. doi:10.1162/jocn.2009.21059

de Groot, A. M. B. (2006). Effects of stimulus characteristics and background music on foreign language vocabulary learning and forgetting. *Language Learning, 56*(3), 463-506. doi:10.1111/j.1467-9922.2006.00374.x

de Groot, A. M. B. (2011). *Language and cognition in bilinguals and multilinguals: An introduction*. New York, NY: Psychology Press.

de Groot, A. M. B., & Keijzer, R. (2000). What is hard to learn is easy to forget: The roles of word concreteness, cognate status, and word frequency in foreign-language vocabulary learning and forgetting. *Language Learning, 50*(1), 1-56.

de Groot, A. M. B., & Poot, R. (1997). Word translation at three levels of proficiency in a second language: The ubiquitous involvement of conceptual memory. *Language Learning, 47*(2), 215-264. doi:10.1111/0023-8333.71997007

Drake, D. A., & Ehri, L. C. (1984). Spelling acquisition: Effects of pronouncing words on memory for their spellings. *Cognition and Instruction, 1*(3), 297-320. doi:10.1207/s1532690xci0103_2

Druks, J. (2002). Verbs and nouns—a review of the literature. *Journal of Neurolinguistics, 15*(3–5), 289-315. doi:10.1016/s0911-6044(01)00029-x

Druks, J., Masterson, J., Kopelman, M., Clare, L., Rose, A., & Rai, G. (2006). Is action naming better preserved (than object naming) in Alzheimer's disease and why should we ask? *Brain and Language, 98*(3), 332-340. doi:http://dx.doi.org/10.1016/j.bandl.2006.06.003

Dye, C. D., Walenski, M., Prado, E. L., Mostofsky, S., & Ullman, M. T. (2013). Children's computation of complex linguistic forms: A study of frequency and imageability effects. *PLoS ONE, 8*(9), e74683. doi:10.1371/journal.pone.0074683

Earles, J. L., & Kersten, A. W. (2016). Why are verbs so hard to remember? Effects of semantic context on memory for verbs and nouns. *Cognitive Science, 41*(S4), 780-807. doi:10.1111/cogs.12374

Ehri, L. C. (1984). How orthography alters spoken language competencies in children learning to read and spell. In J. Downing & R. Valtin (Eds.), *Language awareness and learning to read* (pp. 119-147). New York, NY: Springer New York.

Ehri, L. C. (2014). Orthographic mapping in the acquisition of sight word reading, spelling memory, and vocabulary learning. *Scientific Studies of Reading, 18*(1), 5-21. doi:10.1080/10888438.2013.819356

Everson, M. E. (1998). Word recognition among learners of Chinese as a foreign language: Investigating the relationship between naming and knowing. *The Modern Language Journal, 82*(2), 194-204. doi:10.1111/j.1540-4781.1998.tb01192.x

Eviatar, Z., Menn, L., & Zaidel, E. (1990). Concreteness: Nouns, verbs, and hemispheres. *Cortex, 26*(4), 611-624. doi:https://doi.org/10.1016/S0010-9452(13)80310-3

Ferré, P., Guasch, M., García-Chico, T., & Sánchez-Casas, R. (2015). Are there qualitative differences in the representation of abstract and concrete words? Within-language and cross-language evidence from the semantic priming paradigm. *The Quarterly Journal of Experimental Psychology, 68*(12), 2402-2418. doi:10.1080/17470218.2015.1016980

Ferré, P., Sánchez-Casas, R., Comesaña, M., & Demestre, J. (2017). Masked translation priming with cognates and noncognates: Is there an effect of words' concreteness? *Bilingualism: Language and Cognition, 20*(4), 770-782. doi:10.1017/S1366728916000262

Ferré, P., Ventura, D., Comesaña, M., & Fraga, I. (2015). The role of emotionality in the acquisition of new concrete and abstract words. *Frontiers in Psychology, 6*, 976. doi:10.3389/fpsyg.2015.00976

Gathercole, S. E., Pickering, S. J., Hall, M., & Peaker, S. M. (2001). Dissociable lexical and phonological influences on serial recognition and serial recall. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 54*(1), 1-30.

Geng, J., & Schnur, T. T. (2015). The representation of concrete and abstract concepts: Categorical versus associative relationships. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(1), 22-41. doi:10.1037/a0037430

Gentner, D. (1981). Some interesting differences between verbs and nouns. *Cognition and brain theory, 4*(2), 161-178.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistical relativity versus natural partitioning. In S. A. Kuczaj II (Ed.), *Language development: Vol 2. Language, thought, and culture* (Vol. 301-334). Hillsdale, NJ: Lawrence Erlbaum.

Gentner, D. (2006). Why verbs are hard to learn. In K. Hirsh-Pasek & R. M. Golinkoff (Eds.), *Action meets word: How children learn verbs* (pp. 544-564). Oxford, UK: Oxford University Press.

Gentner, D., & Boroditsky, L. (2001). On individuation, relativity, and early word learning. In M. Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 257-283). Cambridge, UK: Cambridge University Press.

Hanley, J. R., Hunt, R. P., Steed, D. A., & Jackman, S. (2013). Concreteness and word production. *Memory & Cognition, 41*(3), 365-377. doi:10.3758/s13421-012-0266-5

Haro, J., & Ferré, P. (2018). Semantic ambiguity: Do multiple meanings inhibit or facilitate word recognition? *Journal of Psycholinguistic Research, 47*(3), 679-698. doi:10.1007/s10936-017-9554-3

Holcomb, P. J., Kounios, J., Anderson, J. E., & West, W. C. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: An electrophysiological investigation. *Journal of Experimental Psychology-Learning Memory and Cognition, 25*(3), 721-742.

Howard, D., & Gatehouse, C. (2006). Distinguishing semantic and lexical word retrieval deficits in people with aphasia. *Aphasiology, 20*(9), 921-950. doi:10.1080/02687030600782679

Imai, M., Haryu, E., & Okada, H. (2005). Mapping novel nouns and verbs onto dynamic action events: Are verb meanings easier to learn than noun meanings for Japanese children? *Child Development, 76*(2), 340-355. doi:10.1111/j.1467-8624.2005.00849_a.x

Imai, M., Li, L., Haryu, E., Okada, H., Hirsh-Pasek, K., Golinkoff, R. M., & Shigematsu, J. (2008). Novel noun and verb learning in Chinese-, English-, and Japanese-speaking children. *Child Development, 79*(4), 979-1000. doi:10.1111/j.1467-8624.2008.01171.x

James, C. T. (1975). The role of semantic information in lexical decisions. *Journal of Experimental Psychology: Human Perception and Performance, 1*(2), 130-136. doi:10.1037/0096-1523.1.2.130

Johnston, R. A., & Barry, C. (2006). Age of acquisition and lexical processing. *Visual Cognition, 13*(7/8), 789-845. doi:10.1080/13506280544000066

Kambanaros, M. (2009). Investigating grammatical word class distinction in bilingual aphasic individuals. In G. Ibanescu & S. Pescariu (Eds.), *Aphasia: Symptoms, diagnosis, and treatment* (pp. 1-59). Hauppauge, NY: Nova Science Publishers.

Kambanaros, M., & Grohmann, K. K. (2010). Patterns of object and action naming in Cypriot Greek children with SLI and WFDs. *Proceedings of the 34th Boston University Child Language Development—Supplement*.

Kambanaros, M., & Grohmann, K. K. (2015). Grammatical class effects across impaired child and adult populations. *Frontiers in Psychology, 6*, 1670. doi:10.3389/fpsyg.2015.01670

Kambanaros, M., Grohmann, K. K., & Michaelides, M. (2013). Lexical retrieval for nouns and verbs in typically developing bilectal children. *First Language, 33*(2), 182-199. doi:10.1177/0142723713479435

Kambanaros, M., Grohmann, K. K., Michaelides, M., & Theodorou, E. (2014). On the nature of verb–noun dissociations in bilectal SLI: A psycholinguistic perspective from Greek. *Bilingualism: Language and Cognition, 17*(1), 169-188. doi:10.1017/S1366728913000035

Kambanaros, M., & van Steenbrugge, W. (2006). Noun and verb processing in Greek–English bilingual individuals with anomic aphasia and the effect of instrumentality and verb–noun name relation. *Brain and Language, 97*(2), 162-177. doi:http://dx.doi.org/10.1016/j.bandl.2005.10.001

Karpicke, J. D., & Roediger, H. L. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language, 57*(2), 151-162. doi:10.1016/j.jml.2006.09.004

Kauschke, C., & Stenneken, P. (2008). Differences in noun and verb processing in lexical decision cannot be attributed to word form and morphological complexity alone. *Journal of Psycholinguistic Research, 37*(6), 443-452. doi:10.1007/s10936-008-9073-3

Kaushanskaya, M., & Marian, V. (2009). The bilingual advantage in novel word learning. *Psychonomic Bulletin & Review, 16*(4), 705-710. doi:10.3758/pbr.16.4.705

Kaushanskaya, M., & Rechtzigel, K. (2012). Concreteness effects in bilingual and monolingual word learning. *Psychonomic Bulletin & Review, 19*(5), 935-941. doi:10.3758/s13423-012-0271-5

Laxén, J., & Lavaur, J.-M. (2009). The role of semantics in translation recognition: Effects of number of translations, dominance of translations and semantic relatedness of multiple translations. *Bilingualism: Language and Cognition, 13*(2), 157-183. doi:10.1017/S1366728909990472

Lee, C.-L., & Federmeier, K. D. (2008). To watch, to see, and to differ: An event-related potential study of concreteness effects as a function of word class and lexical ambiguity. *Brain and Language, 104*(2), 145-158. doi:https://doi.org/10.1016/j.bandl.2007.06.002

Levelt, W. J. M. (1989). *Speaking*. Cambridge, MA: MIT Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*(1), 1-38.

Li, P., & Farkaš, I. (2002). A self-organizing connectionist model of bilingual processing. In R. R. Heredia & J. Altarriba (Eds.), *Bilingual Sentence Processing* (pp. 59-88). North Holland: Elsevier Science Publisher.

Li, P., Farkaš, I., & MacWhinney, B. (2004). Early lexical development in a self-organizing neural network. *Neural Networks, 17*(8–9), 1345-1362. doi:http://dx.doi.org/10.1016/j.neunet.2004.07.004

Longe, O., Randall, B., Stamatakis, E. A., & Tyler, L. K. (2007). Grammatical categories in the brain: The role of morphological structure. *Cerebral Cortex, 17*(8), 1812-1820. doi:10.1093/cercor/bhl099

Ma, W., Golinkoff, R. M., Hirsh-Pasek, K., McDonough, C., & Tardif, T. (2009). Imageability predicts the age of acquisition of verbs in Chinese children. *Journal of Child Language, 36*(2), 405-423. doi:10.1017/S0305000908009008

Maguire, M. J., Hirsh-Pasek, K., & Golinkoff, R. M. (2006). A unified theory of word learning: Putting verb acquisition in context. In K. Hirsh-Pasek & R. M. Golinkoff (Eds.), *Action meets word: How children learn verbs* (pp. 364-391). Oxford, UK: Oxford University Press.

Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. *PLoS ONE, 7*(8), e43230. doi:10.1371/journal.pone.0043230

Masson, M. E. J. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods, 43*(3), 679-690. doi:10.3758/s13428-010-0049-5

Mätzig, S., Druks, J., Masterson, J., & Vigliocco, G. (2009). Noun and verb differences in picture naming: Past studies and new evidence. *Cortex, 45*(6), 738-758. doi:http://dx.doi.org/10.1016/j.cortex.2008.10.003

McCarthy, D. (2006). Relating WordNet senses for word sense disambiguation. *Proceedings of the Workshop on Making Sense of Sense: Bringing Psycholinguistics and Computational Linguistics Together* (pp. 17-24). Retrieved from http://www.aclweb.org/anthology/W06-2503.

McDonough, C., Song, L., Hirsh-Pasek, K., Golinkoff, R. M., & Lannon, R. (2011). An image is worth a thousand words: Why nouns tend to dominate verbs in early word learning. *Developmental Science*, *14*(2), 181-189. doi:10.1111/j.1467-7687.2010.00968.x

Mestres-Missé, A., Münte, T. F., & Rodriguez-Fornells, A. (2014). Mapping concrete and abstract meanings to new words using verbal contexts. *Second Language Research, 30*(2), 191-223. doi:10.1177/0267658313512668

Paivio, A. (1971). *Imagery and verbal process*. New York, NY: Holt, Rinehart, & Winston.

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford, UK: Oxford University Press.

Paivio, A. (2006). *Mind and its evolution: A dual coding theoretical account*. Mahwah, NJ: Lawrence Erlbaum.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology, 76*(1, Pt.2), 1-25. doi:10.1037/h0025327

Palmer, S. D., MacGregor, L. J., & Havelka, J. (2013). Concreteness effects in single-meaning, multi-meaning and newly acquired words. *Brain Research, 1538,* 135-150. doi:http://dx.doi.org/10.1016/j.brainres.2013.09.015

Papagno, C., Capasso, R., & Miceli, G. (2009). Reversed concreteness effect for nouns in a subject with semantic dementia. *Neuropsychologia, 47*(4), 1138-1148. doi:http://dx.doi.org/10.1016/j.neuropsychologia.2009.01.019

Papagno, C., & Vallar, G. (1995). Verbal short-term memory and vocabulary learning in polyglots. *The Quarterly Journal of Experimental Psychology Section A, 48*(1), 98-107. doi:10.1080/14640749508401378

Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*(4), 357-383. doi:10.1080/10888430701530730

Perfetti, C. A., & Hart, L. (2001). The lexical basis of comprehension skill. In D. S. Gorfein (Ed.), *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity* (pp. 67-86). Washington, DC: American Psychological Association.

Piercey, C. D., & Joordens, S. (2000). Turning an advantage into a disadvantage: Ambiguity effects in lexical decision versus reading tasks. *Memory & Cognition, 28*(4), 657-666. doi:10.3758/bf03201255

Plaut, D. C., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology, 10*(5), 377-500. doi:10.1080/02643299308253469

Prado, E. L., & Ullman, M. T. (2009). Can imageability help us draw the line between storage and composition? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(4), 849-866. doi:10.1037/a0015286

Princeton. (2010). About WordNet. Retrieved from https://wordnet.princeton.edu

PST. E-Prime 2.0. Pittsburgh, PA: Psychology Software Tools.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25*, 111-163.

Rodd, J. M., Berriman, R., Landau, M., Lee, T., Ho, C., Gaskell, M. G., & Davis, M. H. (2012). Learning new meanings for old words: Effects of semantic relatedness. *Memory & Cognition, 40*(7), 1095-1108. doi:10.3758/s13421-012-0209-1

Rodd, J. M., Gaskell, G., & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language, 46*(2), 245-266.

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15*(1), 20-27. doi:10.1016/j.tics.2010.09.003

Romani, C., McAlpine, S., & Martin, R. C. (2008). Concreteness effects in different tasks: Implications for models of short-term memory. *The Quarterly Journal of Experimental Psychology*, *61*(2), 292-323.

Rosenthal, J., & Ehri, L. C. (2011). Pronouncing new words aloud during the silent reading of text enhances fifth graders' memory for vocabulary words and their spellings. *Reading and Writing*, *24*(8), 921-950. doi:10.1007/s11145-010-9239-x

Sandberg, C., & Kiran, S. (2014). Analysis of abstract and concrete word processing in persons with aphasia and age-matched neurologically healthy adults using fMRI. *Neurocase*, *20*(4), 361-388. doi:10.1080/13554794.2013.770881

Schwanenflugel, P. J., Harnishfeger, K. K., & Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language*, *27*(5), 499-520. doi:http://dx.doi.org/10.1016/0749-596X(88)90022-8

Schwanenflugel, P. J., & Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(1), 82-102. doi:10.1037/0278-7393.9.1.82

Schwanenflugel, P. J., & Stowe, R. W. (1989). Context availability and the processing of abstract and concrete words in sentences. *Reading Research Quarterly*, *24*(1), 114-126.

Sheng, L., & McGregor, K. K. (2010). Object and action naming in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *53*(6), 1704-1719. doi:10.1044/1092-4388(2010/09-0180)

Szekely, A., D'Amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., . . . Bates, E. (2005). Timed action and object naming. *Cortex*, *41*(1), 7-25. doi:10.1016/s0010-9452(08)70174-6

Tokowicz, N., & Kroll, J. F. (2007). Number of meanings and concreteness: Consequences of ambiguity within and across languages. *Language and Cognitive Processes, 22*(5), 727-779.

Tokowicz, N., Michael, E. B., & Kroll, J. F. (2004). The roles of study-abroad experience and working-memory capacity in the types of errors made during translation. *Bilingualism: Language and Cognition*, *7*(3), 255-272. doi:10.1017/S1366728904001634

Tolentino, L. C., & Tokowicz, N. (2009). Are pumpkins better than heaven? An ERP investigation of order effects in the concrete-word advantage. *Brain and Language, 110*(1), 12-22.

van Hell, J. G., & Candia Mahn, A. (1997). Keyword mnemonics versus rote rehearsal: Learning concrete and abstract foreign words by experienced and inexperienced learners. *Language Learning, 47*(3), 507-546. doi:10.1111/0023-8333.00018

van Hell, J. G., & de Groot, A. M. B. (1998a). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition, 1*(3), 193-211.

van Hell, J. G., & de Groot, A. M. B. (1998b). Disentangling context availability and concreteness in lexical decision and word translation. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 51A*(1), 41-63.

van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound, and reading. *Memory & Cognition, 15*(3), 181-198. doi:10.3758/bf03197716

Vigliocco, G., Vinson, D. P., Druks, J., Barber, H., & Cappa, S. F. (2011). Nouns and verbs in the brain: A review of behavioural, electrophysiological, neuropsychological and imaging studies. *Neuroscience & Biobehavioral Reviews, 35*(3), 407-426. doi:http://dx.doi.org/10.1016/j.neubiorev.2010.04.007

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review, 14*(5), 779-804. doi:10.3758/bf03194105

Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(5), 1256-1271.

Wang, J., Conder, J. A., Blitzer, D. N., & Shinkareva, S. V. (2010). Neural representation of abstract and concrete concepts: A meta-analysis of neuroimaging studies. *Human Brain Mapping, 31*(10), 1459-1468. doi:10.1002/hbm.20950

Waxman, S., Fu, X., Arunachalam, S., Leddon, E., Geraghty, K., & Song, H.-j. (2013). Are nouns learned before verbs? Infants provide insight into a long-standing debate. *Child Development Perspectives, 7*(3), 155-159. doi:10.1111/cdep.12032

West, W. C., & Holcomb, P. J. (2000). Imaginal, semantic, and surface-level processing of concrete and abstract words: An electrophysiological investigation. *Journal of Cognitive Neuroscience, 12*(6), 1024-1037.

Wilson, M. D. (1988). The MRC psycholinguistic database: Machine readable dictionary, Version 2. *Behavior Research Methods, 20*(1), 6-11.

Xu, X., Kang, C., & Guo, T. (2016). Imageability and semantic association in the representation and processing of event verbs. *Cognitive Processing*, *17*(2), 175-184. doi:10.1007/s10339-015-0747-0

Zhang, Q., Guo, C.-Y., Ding, J.-H., & Wang, Z.-Y. (2006). Concreteness effects in the processing of Chinese words. *Brain and Language, 96*(1), 59-68. doi:https://doi.org/10.1016/j.bandl.2005.04.004

Zhang, X., Han, Z., & Bi, Y. (2013). Are abstract and concrete concepts organized differently? Evidence from the blocked translation paradigm. *Applied Psycholinguistics, 34*(5), 1059-1092. doi:10.1017/S0142716412000124

[1] We use ‘concreteness’ to refer to both concreteness and imageability, which, although not exactly the same (e.g., Paivio, Yuille & Madigan, 1968), are very highly correlated (e.g., de Groot & Poot, 1997; Hanley et al., 2013; Howard & Gatehouse, 2006; Maguire et al., 2006; Mestres-Missé et al., 2014; Palmer et al., 2013; Romani et al., 2008).

[2] The data from both experiments were also analyzed using linear mixed effects models (LMEs), implemented via the lme4 package in R (Bates, Maechler, Bolker & Walker, 2014), given the growing preference for this approach (Baayen, Davidson & Bates, 2008; Baayen & Milin, 2010). However, we report here the results from the ANOVAs and linear hierarchical regressions because the analyses with LMEs required the use of multiple different optimizers and random effects structures (e.g., including or excluding random slopes for interaction effects) to achieve model convergence. This means different models are difficult to directly compare (e.g., across test cycles or across sessions). In all cases the pattern of results from the LMEs was the same as from the ANOVAs and linear hierarchical regressions and support the same interpretations and conclusions.

[3] This length was determined during pilot testing; it was the average amount of time needed to repeat the pair two times, plus a 250 ms buffer.

[4] After completing the language tasks, participants completed measures of cognitive individual differences; these results are not analyzed here.

[5] To address this issue further, we calculated the Bayesian posterior probabilities of the null and alternative hypotheses (regarding the presence of a concreteness effect) using the BIC approximation described in Wagenmakers (2007) and Masson (2011). According to the interpretations suggested by Raftery (1995), the results suggest strong or very strong evidence for a concreteness effect for all measures on both tasks on Session 1, and for L2 to L1 production

accuracy and RTs on Session 2 (though only weak evidence for a concreteness effect in typing accuracy on Session 2). We thank Montserrat Comesaña for this suggestion.