University of Wollongong

Research Online

Faculty of Arts, Social Sciences and Humanities
- Papers

Faculty of Arts, Social Sciences & Humanities

1-1-2020

Semantic feature effect in verbal short-term memory

Mabel C. Lau

University of Wollongong, bxm1994@uowmail.edu.au

Steven J. Roodenrys

University of Wollongong, steven@uow.edu.au

Leonie M. Miller

University of Wollongong, leoniem@uow.edu.au

Follow this and additional works at: <https://ro.uow.edu.au/asshpapers>

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

Semantic feature effect in verbal short-term memory

Abstract

© 2020, © 2020 Informa UK Limited, trading as Taylor & Francis Group. The present study explored the influences of semantic features in immediate serial recall in order to further examine the involvement of semantic knowledge in short-term memory. The number of semantic features (NoF) was found to have a positive effect on short-term recall where high NoF words were remembered better than low NoF words (Experiment 1). This effect was replicated in a second experiment and was found to persist even after controlling for a potential confound (number of distinguishing features). It was further found that having more distinctive features facilitated recall performance of words whose representation was semantically poorer (Experiment 3). These results provide additional evidence of semantic influences in short-term memory and demonstrate that the organisation of semantic knowledge is reflected in short-term memory performance.

Publication Details

Lau, M. C., Roodenrys, S. & Miller, L. (2020). Semantic feature effect in verbal short-term memory. *Memory*, Online First

Running head: Semantic Feature Effect

Semantic Feature Effect in Verbal Short-term Memory

Mabel C. Lau, Steven Roodenrys, and Leonie M. Miller

University of Wollongong, Wollongong, New South Wales, Australia

Address all correspondence to:

Mabel C. Lau
University of Wollongong
School of Psychology
Wollongong NSW 2522
Australia

E-mail: bxm1994@uowmail.edu.au

Word Count: 10, 334

Abstract

The present study explored the influences of semantic features in immediate serial recall in order to further examine the involvement of semantic knowledge in short-term memory. The number of semantic features (NoF) was found to have a positive effect on short-term recall where high NoF words were remembered better than low NoF words (Experiment 1). This effect was replicated in a second experiment and was found to persist even after controlling for a potential confound (number of distinguishing features). It was further found that having more distinctive features facilitated recall performance of words whose representation was semantically poorer (Experiment 3). These results provide additional evidence of semantic influences in short-term memory and demonstrate that the organisation of semantic knowledge is reflected in short-term memory performance.

Keywords: semantic features, immediate serial recall, semantic knowledge, short-term memory

Semantic Feature Effect in Verbal Short-term Memory

Short-term memory (STM) has been traditionally conceptualised as a separate store from long-term memory (LTM) (e.g., Atkinson & Shiffrin, 1968; Baddeley & Hitch, 1974), where there is a strong reliance on phonological codes in STM, as evidenced in the phonological similarity effect (e.g., Conrad & Hull, 1964), and in the word length effect (e.g., Baddeley, Thomson, & Buchanan, 1975). However, STM has also been found to be influenced by linguistic knowledge in LTM, with STM tasks showing a lexicality effect (e.g., Hulme, Maughan, & Brown, 1991), and a word frequency effect (e.g., Hulme, Roodenrys, Schweickert, Brown, Martin, & Stuart, 1997). Importantly, findings of semantic effects (e.g., concreteness, Walker & Hulme, 1999; semantic relatedness, Poirier & Saint-Aubin, 1995; Wetherick, 1975) suggest the involvement of semantic knowledge in facilitating short-term recall, in addition to phonological influences. The aim of the present study is to further explore the influence of semantics on short-term recall through examining the effects of semantic features.

Lexical-semantic contributions to STM

Immediate serial recall performance has been found to be influenced by lexical variables, such as word frequency (e.g., Hulme et al., 1997) and neighbourhood size (e.g., Roodenrys, Hulme, Lethbridge, Hinton, & Nimmo, 2002). However, it has also been found to be influenced by the semantic properties of words. For instance, serial recall performance is better for lists of concrete words compared to abstract words (e.g., Campoy, Castella, Provencio, Hitch, & Baddeley, 2015; Romani, McAlpine, & Martin, 2008; Walker & Hulme, 1999). Saint-Aubin and colleagues (e.g., Poirier & Saint-Aubin, 1995; Saint-Aubin, Ouellette, & Poirier, 2005; Saint-Aubin & Poirier, 1999; Wetherick, 1975) found better serial recall performance for lists of words from the same category versus multiple categories

(semantic relatedness effect). In addition, the emotional content of words has been found to influence immediate serial recall, where emotional words tend to be remembered better than neutral words (e.g., Majerus & D'Argembeau, 2011; Monnier & Syssau, 2008). These findings provide clear evidence that LTM influences on STM are not limited to lexical or phonological effects. Rather, the recruitment of long-term knowledge to support short-term recall includes the use of semantic representations in LTM.

Theoretical accounts of the interaction between STM and LTM.

The findings of lexical-semantic effects provide evidence for the involvement of the language system in STM tasks, where LTM supports STM in one of two ways. First, LTM is postulated to facilitate STM at the point of retrieval, where long-term representations of words help to reconstruct any degraded phonological traces in STM (the redintegration account; e.g., Schweickert, 1993). Generally, the more accessible these long-term representations are, the more likely it is that they will be successfully reconstructed, and hence recalled. On the other hand, language-based theoretical accounts hypothesise a direct contribution to STM tasks from temporary activation arising from the lexical-semantic knowledge within the language system for the to-be-recalled words (psycholinguistic accounts; e.g., N. Martin & Saffran, 1997; R. C. Martin, Lesch, & Bartha, 1999). In other words, short-term recall is facilitated by feedforward and feedback activation among the various levels (e.g., phonological, lexical, semantic) in the long-term knowledge structure (R. C. Martin et al., 1999) which operates from the time of presentation until recall.

A third account incorporates features from both redintegration theories and language-based models. Thorn, Frankish, and Gathercole (2009) proposed a multiple-mechanism account where the influence of long-term knowledge occurs at multiple time-points. The multiple-mechanism account adds to Schweickert's (1993) redintegration framework by

assuming that long-term knowledge could also exert an influence on the trace integrity.

Assuming that the memory trace is represented by patterns of activation across phonological units, the strength of the memory trace reflects the potential level of activation an item could achieve at the storage stage. Long-term knowledge influences the strength of STM traces via top-down interactive activation amongst the various components in the network. In other words, long-term knowledge facilitates STM at multiple time-points in the memory process, first at the encoding or retention stage (thereby influencing the memory trace's strength and integrity) and also at the retrieval stage (facilitating the reconstruction of degraded traces). In sum, all three approaches indicate that how an item is represented in the language system would have important consequences for its memorability.

Semantic contributions to STM recall

The findings of semantic effects suggest a role for semantic codes in short-term recall. However, within the STM domain, examinations of semantic effects have been largely limited to concreteness and emotional-semantic effects despite the multidimensionality nature of semantic representations. In fact, several theoretical variables exist to capture different aspects of semantics (see McRae & Jones, 2013; Pexman, 2012, for detailed discussions). In other words, they are not limited to concreteness and/or emotional valence, but rather encompass the attributes derived through human experiences about the given concepts (Grondin, Lupker, & McRae, 2009). Hence, the extent to which other semantic variables have an influence on STM remains an important open question.

One semantic variable that has recently received attention in the lexical processing and episodic memory literature is semantic features. Semantic features, which refer to attributes listed for a concept (e.g., features of "cow" would be "*eats grass*", "*has four legs*", etc.), represent a way of conceptualising the organisation of semantic knowledge and have

been considered fundamental to the representations of semantic knowledge (Vinson & Vigliocco, 2008); hence this variable offers a novel potential influence of semantic knowledge on memory tasks. From a feature-based perspective, semantic representations of words consist of a list of descriptive attributes of the target word's referent, with semantic activations of words being represented by the differing patterns of activity of features (McRae, de Sa, & Seidenberg, 1997).

Support for the role of semantic features in enhancing memorability comes from the speech processing and episodic memory literatures. For the former, the number of semantic features (NoF) has been found to affect the speed and accuracy with which words are recognised and produced. High NoF words were found to have faster response times and higher accuracy rates than low NoF words in a picture naming task (Rabovsky, Schad, & Abdel Rahman, 2016), as well as in an auditory lexical decision task (Sajin & Connine, 2014, Experiment 1). The facilitative effects of NoF in speech recognition were evident even after accounting for the influences of a wide array of lexical and semantic variables (Goh, Yap, Lau, Ng, & Tan, 2016).

The properties of features have also been found to influence speech processing. Rabovsky et al. (2016) examined the effects of intercorrelational feature density which is indicative of the extent to which the features of a concept are intercorrelated, and provides a measure of the density of semantic space (Rabovsky et al., 2016). Concepts that inhabit denser parts of semantic space are associated with a high intercorrelational density value. They found words that have a high intercorrelational density were associated with slower response times and higher error rates than words that have a low intercorrelational density; this was attributed to the activation of more competitors in denser parts of semantic space.

Within the memory domain, the NoF effect had been demonstrated using the free recall task in two different studies. The first report of an NoF effect in memory was by Hargreaves, Pexman, Johnson, and Zdrazilova (2012) who compared free recall performance between high and low NoF words and found a high NoF word advantage in delayed recall. Associative chaining processes among studied items were ruled out as an explanation for the high NoF word advantage; the authors found that the NoF of a word that had just been recalled had no influence on the probability of recalling the following word. Instead, the NoF effect was attributed to an item-specific process, where the semantic richness of an item itself contributes to its processing at encoding (i.e., item-specific encoding variability). In support of this view, Hargreaves et al. (2012) found an NoF effect even when participants were unaware that they had to perform a recall test. The finding of an NoF effect in an incidental memory task provided support for the item-specific encoding variability hypothesis, as well as suggesting that the increased elaborative encoding (or processing at the initial learning phase) afforded by high NoF words was not restricted to intentional learning.

Further evidence of the NoF effect in episodic recall came from Lau, Goh, and Yap (2018). They conducted an item-level analysis on both free recall and recognition memory tests using hierarchical multiple regression. After controlling for important lexical variables and other semantic variables, NoF continued to influence free recall performance with increasing number of semantic features associated with better free recall performance.

Taken together, although the effects of semantic features have been the focus of much less research than other lexical-semantic aspects in the broader literature, it has not been examined in relation to STM tasks. While semantic features have been shown to influence speech processing and episodic memory, an effect on short-term recall is yet to be demonstrated. If semantic features represent an organisational principle of semantic knowledge, and other LTM variables have been demonstrated to influence short-term recall,

then it can be hypothesised that STM tasks will show an influence of semantic features. The overall aim of the present study is to determine if there is an NoF effect in the immediate serial recall task.

Further considerations in examining semantic feature effects

Several factors must be considered when examining a potential NoF effect in STM. When contrasting NoF to other semantic factors, such as the semantic relatedness effect, emotionality effect or the concreteness effect, for NoF there is no easily discernible difference between stimuli that might be used by participants as a contextual cue. Taking the semantic relatedness effect as an example, the category from which the words were chosen could be used as a cue to guide learning and retrieval. In contrast, the words in McRae, Cree, Seidenberg, and McNorgan's (2005) semantic feature production norms are all concrete nouns. As such, words drawn from this database and presented to participants in an immediate serial recall task will form lists of unrelated words, with the only difference between lists being the quantity of semantic features (e.g., a high NoF word "ambulance" versus a low NoF word "ball"). In contrast, concreteness offers a distinguishing cue between conditions, although it is not as useful a cue as the categorical relationship amongst words in the list underlying the semantic relatedness effect. An NoF effect would therefore represent an influence of conceptual knowledge on STM in the absence of easily identified categorical cues.

Based on Hargreaves et al.'s (2012) item-specific encoding variability account of the NoF effect, it is possible that high NoF words enjoy greater semantic activation at the point of encoding, which correspondingly facilitates their short-term recall. The finding of an NoF effect in short-term recall would also suggest that a concept's features are activated upon the presentation of the concept, despite the time constraints inherent in an immediate serial recall

task, and support the theoretical position that concepts are represented by their semantic features. Furthermore, an NoF effect in serial recall would provide evidence that the memory system is flexible and dynamic and is able to extract information that is useful for optimising performance (see Lau et al., 2018, for a similar argument in free recall and recognition memory).

The time constraints in serial recall are another factor worth consideration when investigating a possible NoF effect; a point raised by Shulman (1970; 1971) in relation to semantic encoding in the serial recall task. Shulman argued that both phonological and semantic encoding are evident in STM tasks, although they exhibit different temporal courses. He found enhanced STM performance following semantic encoding only when the word stimuli were presented at a slower rate. This led him to conclude that semantic encoding benefits from slower presentation rates because semantic encoding takes a longer time than phonological encoding. In line with this proposal, Campoy et al. (2015) manipulated the presentation rate in their examination of the concreteness effect. They found a larger concreteness effect at a presentation rate of 2 seconds than 1 second per word, suggesting that having a slower presentation rate benefitted semantic processing. Although it should also be noted that the concreteness effect was observed at the standard presentation rate of 1 second per word, hence demonstrating that semantic processing is also possible without the need to extend the presentation rate.

Similarly, a recent study by Kowialiewski and Majerus (2018) demonstrated that the contributions of linguistic knowledge in STM could also occur under a fast presentation rate. They tested a variety of long-term effects (lexicity, word frequency, semantic similarity, and imageability) using a running span procedure under a fast encoding condition, where participants were presented sequences of items at a rate of 2.5 items per second, with an unpredictable list length of 6, 9, or 12 items, and had to recall as many of the last items as

they could in serial order. They replicated the lexicality, frequency, and semantic similarity effects, indicating that certain semantic representations were utilised in a quick and automated manner for recall. However, the imageability effect was found using an immediate serial recall task but not in the running span task. In fact, using the same presentation rate of 1 item per 1.5 seconds, the imageability effect was observed in the immediate serial recall task but remained absent in the running span task. This points to the task-dependent nature of the imageability effect, although it should be noted that this version of a running span task is very different from the immediate serial recall task which has been the prototypical task in examining lexical-semantic effects. Nonetheless, the studies of Shulman (1970; 1971) and Campoy et al. (2015) suggest that presentation rate may moderate the influence of some semantic processes in serial recall. Hence, for the present study, the NoF effect will be examined under 2 presentation rate conditions. Experiment 1 explored the influence of NoF on the immediate serial recall task under standard (1 second per item) presentation rate condition. Experiments 2 and 3 further examined the effects of semantic features by exploring how distinctive features influenced the NoF effect, as well as the recall of high and low NoF words under different presentation rates. In Experiments 2 and 3, word stimuli were presented under fast (1 second) and slow (2 seconds) conditions to determine how rates of presentation influence the semantic features effects.

Experiment 1

Method

Participants. Twenty undergraduates from the University of Wollongong (UOW) took part in this experiment for course credit. All participants had English as their first language with normal or corrected-to-normal vision, and had no speech or hearing

difficulties. This number was chosen as it is fairly typical of STM experiments in the literature demonstrating LTM effects on serial recall¹.

Materials. The word stimuli used in this experiment were selected from McRae et al. (2005) semantic production norms. There were 96 high NoF words (e.g., bus, duck, crown, basement, whistle) and 96 low NoF words (e.g., cello, stone, mirror, panther, cathedral) that were used to create 16 lists of 6 words in each condition. The two sets of words were matched on lexical frequency (log frequency based on subtitles: Brysbaert & New, 2009; New, Brysbaert, Veronis, & Pallier, 2007; and the written corpus of CELEX: Baayen, Piepenbrock, & van Rijn, 1993), number of letters, phonemes, and syllables, number of orthographic neighbours, number of phonological neighbours, and concreteness, all $|t|s \leq 1.63$, $ps \geq .110$, but differed in NoF, $t(181.2) = 28.3$, $p < .001$ (see Table 1 for descriptive statistics of control variables and NoF measure). The low- and high-group were decided by choosing words closer to the lower end and upper end of the scale, respectively, so as to maximise the difference in NoF, while ensuring the items were matched on the other lexical-semantic properties. The stimuli were recorded by a native Australian English speaker and normalised to equate the volume of each word.

[Table 1 near here]

Procedure. The immediate serial recall task was conducted using E-prime (Schneider, Eschman, & Zuccolotto, 2002). Each study list consisted of 6 words, with a total of 32 lists (16 lists for each NoF condition). Words were presented auditorily through an external amplified speaker, at the rate of 1 word per second. Participants had to recall the words in the order they were presented immediately after the last word was played by saying out aloud

¹ Based on the effect size observed in Experiment 1, the results of a power analysis showed that the estimated power to observe significant effects with a sample size of 20 was 99%. The effect size was based on the partial eta-squared obtained from the repeated measures ANOVA analysis which has originally motivated the selection of sample size and design in subsequent experiments.

their responses. In order to maintain the sequential position of words, participants were told to say ‘blank’ when they failed to remember a word. For each participant, the order of presentation of the list-type for each trial was randomised. The 96 words in each set were randomly arranged to create the lists for each participant separately.

Results

A strict correct-in-position scoring was adopted, where a response was scored as correct if the word was correctly recalled in the same serial position as it was presented. Errors were scored as either an item error or order error. Item errors categories consisted of extralist item intrusion, omission, and repetition. An order error occurred when a word from the list was recalled in an incorrect serial position. The proportion of order errors (number of order errors divided by the number of items correctly recalled regardless of serial position) is reported because this measure takes into account the relationship between the number of items correctly recalled and the number of order errors made² (see Saint-Aubin & Poirier, 1999). Unless otherwise stated, an alpha level of .05 was assumed.

Correct serial recall. The data was analysed using a linear mixed-effects model in R software (R Core Team, 2017) using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015); the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2016) was used to obtain the *p*-values for fixed effects. The model included NoF, serial position, and their interaction terms as fixed effects, and participants as random effects³. The results revealed a

² That is, as the number of items correctly recalled increases, so does the number of order errors.

³ This model included random intercepts for participants. We further examined 2 other models: (1) Having random slopes for NoF and serial position fitted, in addition to the inclusion of random intercepts for participants; and (2) having random slopes for NoF, in addition to the inclusion of random intercepts for participants. The first model failed to converge, while the second model did not change the pattern of results reported. The second model was compared to the model that included only random intercepts for participants, and the results showed both models did not significantly differ in terms of fit to the data. For these reasons, subsequent experiments will be analysed using the model that included only random intercepts for participants.

significant main effect of NoF, $F(1, 209) = 9.03, p = .003^4$, where recall rates were higher for high NoF words compared to low NoF words, as well as a significant main effect of serial position, $F(5, 209) = 107.34, p < .001$ (see Figure 1). The interaction term was not statistically significant, $F(5, 209) = .785, p = .561$.

Further analyses of the NoF effect were conducted by calculating the Bayes Factor (BF) to compare the fit of the data under a pair of alternative models. That is, the alternative hypothesis, which contains the target variable, is compared to the null hypothesis, with the BF values indicating the likelihood of one model relative to another. For instance, the alternative hypothesis (containing the NoF term) is compared to the null hypothesis to quantify the likelihood of the model including the NoF term relative to the model excluding the NoF term. The BayesFactor package (Morey, Rouder, Jamil, & Morey, 2015) in R software was used to obtain the BF values; the strength of evidence was interpreted as anecdotal, moderate, strong, very strong, and extreme evidence for a model when the BF value was between 1 and 3, between 3 and 10, between 10 and 30, between 30 and 100, or higher than 100, respectively (Wagenmakers et al., 2018).

In line with the results obtained from the linear mixed-effects analysis, compared to the null model containing the intercept, the model with the highest BF was the model containing the two main effects without the interaction term [NoF + position + participant] ($BF = 3.57E+60$); participant was treated as a random effect. In order to provide an estimate of the NoF effect, the BF of the model [NoF + position + participant] was compared to the BF of the model [position + participant]. Table 2 reveals that the model containing the NoF term was preferred by a factor of 9, hence demonstrating moderate evidence in favour of the model including the NoF term.

⁴ F-values were obtained from the Analysis of Variance Table with Satterthwaite's method that accompanied the linear mixed effects analysis output.

[Figure 1 near here]

Item and order errors. The data were analysed using a linear mixed-effects model. Each error-type was analysed separately. The model included NoF as a fixed effect and participants as random effects. The results revealed a significant NoF effect on the number of item errors (omissions + intrusions), $F(1, 19) = 37.3, p < .001$, where high NoF words had fewer item errors than low NoF words. Further analyses revealed a significant NoF effect on the number of omission errors, $F(1, 19) = 17.96, p < .001$, but no significant NoF effect on the number of intrusion errors, $F(1, 19) = 3.15, p = .092$. There was also no significant NoF effect on the number of repetition errors, $F(1, 38) = .137, p = .714$, and proportion of order errors, $F(1, 19) = 2E-04, p = .989$.

For further analysis of the NoF effect on the errors made, the BF of the model [NoF + participant] was compared to the BF of the model [participant] for each of the error-types reported. Table 2 shows extreme evidence in favour of the model containing the NoF term on the number of item errors, and very strong evidence in favour of the model containing the NoF term on the number of omission errors. The BF also indicated anecdotal level evidence in favour of the NoF effect on the number of intrusion errors, as well as moderate evidence against an NoF effect on the number of repetition (1/.323) and order (1/.304) errors.

[Table 2 near here]

Discussion

The aim of Experiment 1 was to explore potential NoF effects on immediate serial recall. Based on the results of Experiment 1, it was clear that there was a high NoF word advantage over low NoF words. In addition, results from the error analyses suggest that the

high NoF words advantage stemmed from these words having fewer item errors; high NoF words were less likely to be omitted as compared to low NoF words.

The finding of an overall recall advantage for high NoF words paralleled the findings in the speech processing (e.g., Goh et al., 2016; Rabovsky et al., 2016; Sajin & Connine, 2014) and episodic memory literature (Hargreaves et al., 2012; Lau et al., 2018). The finding of a beneficial effect of NoF on immediate serial recall provides additional evidence that short-term recall is sensitive to the influences of semantics. However, the NoF measure is found to be positively correlated with the number of distinguishing features for these stimuli, $r(192) = .504, p < .001$. The number of distinguishing features refers to the number of features the concept has that occurs in one or two other concepts within the norming database. In other words, with an increasing number of semantic features a concept has, there is also an increasing likelihood that some of these features are distinctive. This suggests that the high NoF word advantage seen in Experiment 1 could be attributed, in part, to having more distinguishing features rather than the number of features *per se*. In other words, in addition to having a rich semantic representation, high NoF words could be, in general, more distinctive than low NoF words. In order to better understand the NoF effect, Experiment 2 and 3 were conducted to examine the effects of distinguishing features.

Experiment 2a and 2b

The aim of Experiment 2a and 2b was to explore the feature distinctiveness explanation (Experiment 2b), while at the same time, to replicate the NoF effect using a different list of words (Experiment 2a). Ideally this possibility would be tested in a factorial design with both NoF and number of distinguishing features (NoDF) manipulated. However, due to the constraints of matching these word lists on all control variables, and the confound between NoF and NoDF, it is not possible to obtain a sufficient set of words that are low on

NoF and high on NoDF. As such, the remaining 3 word conditions will be used instead. Two separate comparisons of interest will be analysed: (1) the recall performance of low NoF/low NoDF words with high NoF/low NoDF words to determine whether the NoF effect could be replicated using a different word list (Experiment 2a); and (2) the recall performance of high NoF/low NoDF with high NoF/high NoDF to explore whether the distinctiveness of features makes an independent contribution to the memorability of words (Experiment 2b). It was hypothesised that the high NoF word advantage observed in Experiment 1 would be replicated. However, if the high NoF word advantage was driven primarily by high NoF words having more distinguishing features, then there is a possibility that the NoF effect would not be observed in Experiment 2. Since the comparison is between low NoF/low NoDF words with high NoF/low NoDF words, any advantage that NoDF provides would be eliminated. In addition, if NoDF contributes to short-term recall, then the recall performance for high NoF/high NoDF words would be better than that of high NoF/low NoDF words.

A 2-second presentation rate was included in this experiment in addition to the standard 1-second presentation rate, so as to allow for more time to process semantic information (see also Shulman, 1970; 1971). A similar approach was used by Campoy et al. (2015) who examined the concreteness effect under fast (1-second) and slow (2-second) presentation rates in an attempt to test the relationship between semantic encoding and presentation rate. The basic premise is that phonological and semantic encoding are both possible in an immediate serial recall task, even though they exhibit different temporal courses. Hence, the inclusion of the slow presentation rate condition was meant to allow more time for semantic encoding to take place. The notion that semantic encoding benefitted from a slow presentation rate was demonstrated in the larger concreteness effect when a slower presentation rate was used. This effect was specifically due to an increased recall of concrete words at the slow rate, hence showing the utility of decreasing the presentation rate when

semantic factors are considered. It is possible that featural knowledge (in relation to NoDF) might have more time to be accessed, and therefore produce a stronger effect, should a slower presentation rate be used.

General method

Participants. Forty UOW undergraduates took part in Experiment 2a for course credit; 20 participants were in the 1-second condition, while the remaining 20 participants were in the 2-second condition. A separate 41 UOW undergraduates took part in Experiment 2b for course credit; 20 participants were in each presentation rate condition. One participant's responses were discarded for failing to complete the experiment. All participants had English as their first language with normal or corrected-to-normal vision, and had no speech or hearing difficulties. For both experiments, the sample size was based on a power analysis using data from Experiment 1. The estimated power to observe significant effects with 20 participants in each condition for each experiment was 99%.

Materials and procedure. A different set of word lists from Experiment 1 were used, selected from McRae et al. (2005). There were 24 words in each condition. Some examples of the words in each condition are: low NoF/low NoDF, accordion, seaweed, clamp, veil, yam; high NoF/low NoDF, airplane, marble, nectarine, oven, wheelbarrow; high NoF/high NoDF, raisin, cage, yacht, cucumber, crayon. The sets were matched on lexical frequency (log subtitle frequency and log CELEX), number of letters, phonemes, and syllables, number of orthographic neighbours, number of phonological neighbours, phonological similarity, and concreteness, all $F_s \leq 2.07$, $p_s \geq .135$, but differed in NoF, $F(2, 69) = 165.49$, $p < .001$, and NoDF, $F(2, 69) = 83.5$, $p < .001$ (see Table 3 for descriptive statistics). Post hoc analyses using Bonferroni adjustment indicated that the low NoF/low NoDF set had a lower number of semantic features than the 2 high NoF conditions, $p_s < .001$. The 2 high NoF conditions did

not differ in the number of semantic features, $p = .884$. The high NoF/high NoDF set had a higher number of distinguishing features than the 2 low NoDF conditions, $ps < .001$. The 2 low NoDF conditions did not differ in the number of distinguishing features, $p = 1.00$. The same procedure as Experiment 1 was used, with the exception that each word was presented twice in the current experiment. For each condition a set of 4 trials was created by sampling the stimulus set without replacement. A second set of trials was generated in the same manner to create the 8 trials in each condition. The order of conditions was counterbalanced.

[Table 3 near here]

Results

The same scoring procedure as Experiment 1 was used. Similarly, unless otherwise stated, an alpha level of .05 was assumed.

Experiment 2a: comparing low NoF/low NoDF with high NoF/low NoDF.

Correct serial recall. A linear mixed-effects model was fitted to the data. The model included NoF, serial position, presentation rate, and their interaction terms as fixed effects, and participants as random effects. The results revealed a significant main effect of NoF, $F(1, 418) = 15.56, p < .001$, where high NoF words were remembered better than low NoF words (0.49 versus 0.43), as well as a significant main effect of serial position, $F(5, 418) = 130.20, p < .001$ (see Figure 2). The main effect of presentation rate was not statistically significant, $F(1, 38) = .108, p = .744$. None of the interaction terms were statistically significant, $Fs \leq .832, ps \geq .528$.

Further analyses of the NoF effect were conducted by calculating BF to compare the fit of the data under a pair of alternative models. In line with the results obtained from the linear mixed-effects analysis, compared to the null model containing the intercept, the model with the highest BF was the model containing the two main effects of NoF and position [NoF

+ position + participant] ($BF = 2.58E+86$); participant was treated as a random effect. In order to provide an estimate of the NoF effect, the BF of the model [NoF + position + participant] was compared to the BF of the model [position + participant]. Table 4 reveals that the model containing the NoF term was preferred by a factor of 198⁵, hence demonstrating extreme evidence in favour of the model including the NoF term.

Item and order errors. The data were analysed using a linear mixed-effects model. Each error-type was analysed separately. The model included NoF, presentation rate, and their interaction terms as fixed effects and participants as random effects. The results revealed a significant NoF effect on the number of item errors (omissions + intrusions), $F(1, 38) = 15.11, p < .001$, where high NoF words had fewer numbers of item errors than low NoF words (see Table 4). Further analyses revealed a significant NoF effect on the number of omission errors, $F(1, 38) = 14.11, p < .001$, but no significant NoF effect on the number of intrusion errors, $F(1, 38) = 2.84, p = .100$. There was also no significant NoF effect on the number of repetition errors, $F(1, 38) = .287, p = .595$, and proportion of order errors, $F(1, 38) = 2.26, p = .141$. The main effect of presentation rate for all error-types was not statistically significant, $F_s \leq 2.29, p_s \geq .138$. The NoF x presentation rate interaction term for all error-types was not statistically significant, $F_s \leq 3.86, p_s \geq .057$.

For further analysis of the NoF effect on the errors made, the BF of the model [NoF + presentation + participant] was compared to the BF of the model [presentation + participant] for each of the error-types reported. Table 4 shows very strong evidence in favour of the model containing the NoF term on the number of item errors, as well as on the number of omission errors. The BF also indicated anecdotal level evidence in favour of a null NoF effect

⁵ Similar findings were observed when the BF of the model [NoF + position + presentation + participant] was compared to the BF of the model [position + presentation + participant]; the model containing the NoF term was preferred by a factor of 194.

on the number of intrusion errors (1/.834) and order (1/.552) errors, as well as moderate evidence in favour of a null NoF effect on the number of repetition (1/.258) errors.

[Figure 2 near here]

[Table 4 near here]

Experiment 2b: comparing high NoF/low NoDF with high NoF/high NoDF.

Correct serial recall. A linear mixed-effects model was fitted to the data. The model included NoDF, serial position, presentation rate, and their interaction terms as fixed effects, and participants as random effects. The results revealed no statistically significant main effect of NoDF, $F(1, 411.14) = 1.83, p = .177$, and no statistically significant main effect of presentation rate, $F(1, 74.55) = 1.88, p = .175$ (see Figure 3). The main effect of serial position was, however, statistically significant, $F(5, 411.13) = 123.23, p < .001$. None of the interactions were statistically significant, $F_s \leq 1.34, p_s \geq .249$.

Further analyses of the NoDF effect were conducted by calculating BF to compare the fit of the data under a pair of alternative models. In line with the results obtained from the linear mixed-effects analysis, compared to the null model containing the intercept, the model with the highest BF was the model containing the main effect of serial position [position + participant] ($BF = 3.65E+87$); participant was treated as a random effect. In order to provide an estimate of the NoDF effect, the BF of the model [NoDF + position + participant] was compared to the BF of the model [position + participant]. Table 5 reveals that the model *excluding* the NoDF term was preferred by a factor of 4 (1/.242), hence demonstrating moderate evidence in favour of the model excluding the NoDF term⁶.

⁶ Similar findings were observed when the BF of the model [NoDF + position + presentation + participant] was compared to the BF of the model [position + presentation + participant]; the model excluding the NoDF term was preferred by a factor of 4.

[Figure 3 near here]

Item and order errors. The data were analysed using a linear mixed-effects model. Each error-type was analysed separately. The model included NoDF, presentation rate, and their interaction terms as fixed effects and participants as random effects. The results revealed a significant NoDF effect on the number of item errors (omissions + intrusions), $F(1, 38) = 11.13, p = .002$, where high NoDF words had fewer numbers of item errors than low NoDF words (see Table 5). Further analyses revealed a significant NoDF effect on the number of intrusion errors, $F(1, 38) = 5.64, p = .023$, but no significant NoDF effect on the number of omission errors, $F(1, 38) = 3.86, p = .057$. There was also no significant NoDF effect on the number of repetition errors, $F(1, 38) = .895, p = .350$, and proportion of order errors, $F(1, 38) = .082, p = .776$. The main effect of presentation rate for all error-types was not statistically significant, $F_s \leq .205, p_s \geq .654$. The NoDF x presentation rate interaction term for all error-types was not statistically significant, $F_s \leq 1.27, p_s \geq .266$, except for repetition errors, $F(1, 38) = 4.87, p = .033$. Simple effects revealed a significant NoDF on repetition errors at the 2-second presentation rate, $t(38) = 2.23, p = .032$, but not at the 1-second presentation rate, $t(38) = -.892, p = .378$.

For further analysis of the NoDF effect on the errors made, the BF of the model [NoDF + presentation + participant] was compared to the BF of the model [presentation + participant] for each of the error-types reported. Table 5 shows strong evidence in favour of the model containing the NoDF term on the number of item errors, and anecdotal level evidence in favour of the model containing the NoDF term on omission errors and intrusion errors. The BF also indicated moderate evidence in favour of a null NoDF effect on the number of order errors (1/.242), and anecdotal level evidence in favour of a null NoDF effect on the number of repetition errors (1/.355).

For further analysis of the NoDF x presentation interaction effect on repetition error, the BF of the model [NoDF + presentation + NoDF x presentation + participant] was compared to the BF of the model [NoDF + presentation + participant]. The BF obtained was 2.22, indicating anecdotal level evidence in favour of the model containing the interaction term.

[Table 5 near here]

Discussion

The aim of Experiment 2a and 2b was twofold; first, to replicate the NoF effect using a different set of words, and secondly, to explore the feature distinctiveness explanation. Based on the analyses, there is a recall advantage for high NoF words independent from NoDF; high NoF words are consistently remembered better than low NoF words. Considering that the high NoF word advantage was replicated even after matching the test items on NoDF suggests that the high NoF word advantage observed in Experiment 1 was unlikely to be due to high NoF words having more distinguishing features. In fact, in Experiment 2a the words from the two list-types had an average of no more than 3 distinguishing features. Hence, high NoF words were better remembered because the greater number of semantic features could come together to support the trace's integrity and/or facilitate the reconstruction/deblurring process. Indeed, recall performance for high NoF words associated with more NoDF was comparable to the recall performance of high NoF words associated with fewer NoDF no matter how slow or how fast the to-be-recalled items were presented (Experiment 2b). This point, namely the role of presentation rate and its relation to semantic features effects, will be further considered in the General Discussion. Although it should also be noted that while NoDF did not influence the correct recall of high NoF words, high NoDF words were found to have fewer item errors than low NoDF words.

Nonetheless, the consistent finding of an NoF effect across experiments as a new variable in the STM literature, suggests that this effect is reliable. Replicating the results from Experiment 1, the high NoF word advantage stemmed from high NoF words having fewer item errors than low NoF words. The NoF effect (and the absence of NoDF effect on high NoF words) is compatible with Hargreaves et al.'s (2012) notion that the semantic elaboration afforded by the greater number of semantic features associated with the target item facilitates subsequent retrieval. From this perspective, even if the semantic features of high NoF words were not distinguishing per se (based on the NoDF value), having more features would have allowed for deeper processing which helped to set high NoF words apart from each other. However, at this point it is important to note that the investigation of the feature distinctiveness effect was limited to high NoF words, whose representations were considered to be well-specified. Hence, it was also possible that having more distinguishing features may have marginal influence (if any) on how well these words were remembered. Instead, since the memorability of low NoF words is poorer, these words may benefit to a greater extent should these features be distinctive. This suggests that low NoF words might be a more suitable set of words to examine the feature distinctiveness effect.

Experiment 3

Given that the NoDF effect was not explored using low NoF words in Experiment 2 due to constraints in selecting suitable items as stimuli, Experiment 3 was conducted to determine whether having more distinguishing features would facilitate the memorability of low NoF words. In Experiment 3, the recall performance of low NoDF words was compared with high NoDF words. Similar to Experiment 2, a 2-second presentation rate was also included in addition to the standard 1-second presentation rate, so as to allow for more time to process semantic information (Campoy et al., 2015; Shulman, 1970; 1971).

Method

Participants. Forty-two UOW undergraduates took part in this experiment for course credit. All participants had English as their first language with normal or corrected-to-normal vision, and had no speech or hearing difficulties. Two participants' data were excluded from analysis; one failed to follow the task's protocol, and the other one was due to a computer failure. The sample size was based on a power analysis using data from Experiment 2. The estimated power to observe significant effects with 20 participants in each condition was 99%.

Materials and procedure. There were 42 words that were used to create the lists in each condition. Some examples of the words in each condition are: low NoDF, asparagus, cabinet, pheasant, chain, brick; high NoDF, biscuit, avocado, shield, moth, projector. The sets were matched on lexical frequency (log subtitle frequency and log CELEX), number of letters, phonemes, and syllables, number of orthographic neighbours, number of phonological neighbours, phonological similarity, concreteness, and NoF, all $|t|s \leq .634$, $ps \geq .528$, but differed on NoDF, $t(55.71) = -12.93$, $p < .001$ (see Table 6 for descriptive statistics). The same procedure as Experiment 2a and 2b was used where each word was repeated twice by having 2 blocks of trials for each list-type, resulting in 14 lists of each type, and counterbalancing the order of list-type. There were 20 participants in each presentation rate condition.

[Table 6 near here]

Results

Correct serial recall. A linear mixed-effects model was fitted to the data. The model included NoDF, serial position, presentation rate, and their interaction terms as fixed effects,

and participants as random effects. The results revealed a significant main effect of NoDF, $F(1, 418) = 8.82, p = .003$, where high NoDF words were remembered better than low NoDF words (0.54 versus 0.50), as well as a significant main effect of serial position, $F(5, 418) = 189.36, p < .001$, and presentation rate, $F(1, 38) = 4.57, p = .039$ (see Figure 4). None of the interaction terms were statistically significant, $F_s \leq 3.60, p_s \geq .058$, except for position x presentation rate, $F(5, 418) = 2.61, p = .024$.

Further analyses of the NoDF effect were conducted by calculating BF to compare the fit of the data under a pair of alternative models. Compared to the null model containing the intercept, the model with the highest BF was the model containing all three main effects [NoDF + position + presentation + participant] ($BF = 1.46E+112$); participant was treated as a random effect. The model with the next highest BF was the model containing all three main effects and the position x presentation interaction term [NoDF + position + presentation + position x presentation + participant] ($BF = 1.45E+112$); participant was treated as a random effect. Comparing the BFs of these two models demonstrated anecdotal level evidence in favour of the model *excluding* the position x presentation interaction term (1/.994).

In order to provide an estimate of the NoDF effect, the BF of the model [NoDF + position + presentation + participant] was compared to the BF of the model [position + presentation + participant]. Table 7 reveals that the model containing the NoDF term was preferred by a factor of 6.05, hence demonstrating moderate evidence in favour of the model including the NoDF term.

Item and order errors. The data were analysed using a linear mixed-effects model. Each error-type was analysed separately. The model included NoDF, presentation rate, and their interaction term as fixed effects and participants as random effects. The results revealed a significant NoDF effect on the number of item errors (omissions + intrusions), $F(1, 38) =$

12.58, $p = .001$, and omission errors, $F(1, 38) = 8.75$, $p = .005$. There was no significant NoDF effect on the number of intrusion errors, $F(1, 38) = 2.35$, $p = .133$, repetition errors, $F(1, 38) = .000$, $p = 1.00$, and order errors, $F(1, 38) = .701$, $p = .408$. Neither the main effect of presentation rate, $F_s \leq 2.43$, $p_s \geq .127$, nor the interaction term for all error-types was statistically significant, $F_s \leq 2.22$, $p_s \geq .145$, except for the main effect of presentation rate for order errors, $F(1, 38) = 4.76$, $p = .035$.

For further analysis of the NoDF effect on the errors made, the BF of the model [NoDF + presentation + participant] was compared to the BF of the model [presentation + participant] for each of the error-types reported; participant was treated as a random effect. Table 7 showed strong evidence in favour of the model containing the NoDF term on the number of item errors, and moderate evidence in favour of the model containing the NoDF term on the number of omission errors. On the other hand, the BFs indicated anecdotal level evidence in favour of the model *excluding* the NoDF term on the number of intrusion errors (1/.581), and moderate evidence in favour of the model *excluding* the NoDF term on the number of repetition errors (1/.235) and order errors (1/.280).

To further examine the main effect of presentation rate for order errors, the BF for the model [NoDF + presentation + participant] was compared to the BF for the model [NoDF + participant]. The analysis reveals that the model containing the presentation term was preferred by a factor of 1.96, hence demonstrating anecdotal level evidence in favour of the model including the presentation term.

[Figure 4 near here]

[Table 7 near here]

Discussion

The aim of Experiment 3 was to explore whether NoDF influenced the recall performance of low NoF words. Based on the analyses, high NoDF words were remembered better than low NoDF words. Similar to previous experiments, the high NoDF advantage stemmed from high NoDF words having fewer item errors compared to low NoDF words. This suggested that having more distinguishing features was beneficial when the semantic representation of the to-be-recalled items, based on NoF, was relatively less rich. In other words, when memorability was considerably poorer in general due to the nature of the list items (as in the case for low NoF words), having features that were distinctive could help set apart the target item from the other items. Hence, when compared to low NoF words with fewer numbers of distinctive features, the semantic representation of low NoF words with a greater number of distinctive features would be richer due to the quality of features that word has. This would therefore allow for greater semantic elaboration and processing to occur.

General Discussion

The present study aimed to examine the effects of semantic features on short-term recall. We demonstrated that high NoF words were better remembered than low NoF words (Experiment 1), and this effect was replicated using a different set of words, as well as when the test items were matched on NoDF (Experiment 2a). Based on the error analyses, the locus of the effect is on item recall, and not on order recall. To further explore the NoF effect in immediate recall, the NoDF of each word was also varied. NoDF was found to positively influence the correct recall of low NoF words (Experiment 3) but not high NoF words (Experiment 2b). The finding that NoF and NoDF have the strongest effect on item recall rather than on order recall is compatible with studies demonstrating an influence of semantics on item recall (e.g., Campoy et al., 2015; Majerus & D'Argembeau, 2011). In general, these findings can be positioned in a theoretical framework that assumes a distinction in the memory systems that maintained item and order information (see Majerus, 2009; Nairne &

Kelley, 2004). From this perspective, memory for item information is maintained by the language system (similar to the psycholinguistic accounts) while that of order information is supported by a distinct STM storage system. Hence, the influence of psycholinguistic variables is most evident on item recall.

Presentation rate and its relation to the processing of featural information

Across 3 experiments, the NoF and NoDF effects had been consistently found regardless of how quickly or slowly the items were presented. The manipulation of stimulus presentation rate had previously been used to demonstrate an influence of elaboration on the encoding of semantic information in serial recall. At the same time, researchers have used it to provide some evidence (albeit indirectly) regarding the extent of automaticity of semantic encoding (Campoy et al., 2015). The general premise was that if semantic coding benefitted from slower presentation rate, then these benefits arose from mechanisms that were time-dependent in nature (Campoy et al., 2015). In fact, Campoy et al.'s (2015) finding of a larger concreteness effect when presentation rate was slowed down has been argued to reflect the involvement of time-dependent mechanisms of elaborative encoding and semantic retrieval. However, given that the concreteness effects were observed in both presentation rates (despite the effect being larger in the slow-rate condition), it is challenging to attribute the concreteness effect to strategic processes exclusively. Indeed, in subsequent experiments, Campoy et al. (2015) found that the concreteness effect was still present in a dual-task paradigm; favouring the account that concreteness effects arose from the automatic encoding of the semantic information. In the context of our present study, the examination of semantic effects in different presentation rate conditions provides support for the Campoy et al. (2015) study. In the present study, both NoF and NoDF impacted participants' immediate serial recall performance regardless of the presentation rate. However, unlike Campoy et al.'s (2015) finding of an interaction between concreteness and presentation rate, semantic features

effects were not larger in the slow presentation rate condition. If feature information were utilised in a time-dependent and/or strategic manner, the effects of NoF or NoDF would have been larger in the slow presentation rate condition. Hence, our results suggested that the activations of semantic features stem more from an automatic process.

Theoretical accounts of semantic features effects

Findings of NoF effects in the current studies demonstrate the involvement of the language system in STM. In order to examine the NoF effects, the stimuli used in the present study consisted of only concrete nouns. The consistent finding of semantic features effects showed that there were important differences among concrete words that were matched on a number of key lexical-semantic variables but only differing on NoF or NoDF. Concrete words have been found to have a processing or recall advantage over abstract words as they are more distinctive and have richer semantic representations in LTM (Romani et al., 2008; Walker & Hulme, 1999). The studies reported here demonstrate that even within concrete words, there are important representational differences such that words with more semantic features have memory advantages. The importance of semantic features is further highlighted in Experiments 2 and 3; even if the words do not have a lot of semantic features, as long as these features are distinctive, there will still be a memory advantage for these words. What this suggests is that memory processes utilise this semantic dimension when recalling these words, which were devoid of a clear categorical cue about the lists' composition during the encoding phase⁷.

The findings also directly support the role of featural information in memory (Hargreaves et al., 2012). In Hargreaves et al.'s study, NoF effects were present even when

⁷ After each testing session, participants were asked informally during the debrief session whether they noticed any differences between the word lists. All participants indicated no perceived differences among the words presented.

participants were unaware of an impending memory test. Given the lack of semantic elaboration and perceived need to retrieve (i.e., the intentional learning associated with a typical recall task), the high NoF word advantage could only be attributed to an item-specific process. That is, it was the semantic richness of the words themselves that contributed to their better encoding. In the context of our current study, it is clear that the same semantic richness of the word provides substantial advantages to the memory processes. This is in line with the psycholinguistic accounts (e.g., N. Martin & Saffran, 1997; R. C. Martin et al., 1999) that suggest semantic knowledge is activated at the point of encoding to support the encoding and maintenance of to-be-recalled items. Words associated with richer semantic representations (e.g., high NoF words or high NoDF words) would experience a more robust feedforward and feedback activation among the different levels of linguistic representations (e.g., R. C. Martin et al., 1999). This would increase the probability that the target item will be available for subsequent retrieval.

Importantly, the interaction between NoF and NoDF can also be interpreted from a psycholinguistic perspective, as it arises naturally from the architecture of psycholinguistic models (N. Martin, 2008). Indeed, Miller and Roodenrys (2009) have also found an interactive effect between lexical-semantic variables. Specifically, they found the concreteness effect was diminished at higher levels of word frequency, and the word frequency effect was diminished at higher levels of concreteness. In order to accommodate this finding, it was suggested that there could be an upper bound in terms of the activation level in the lexical layer. This account is congruent with the current finding of an NoDF effect occurring only for low NoF words suggesting that within the context of serial recall there is only so much benefit that can be obtained from the overall lexico-semantic profile of items.

This is not meant to discount the redintegration account. Indeed, Walker and Hulme (1999) have suggested that semantics could be incorporated into the redintegration framework where semantic knowledge is recruited to clean up degraded memory traces. It follows that varying degrees of semantic richness (in terms of semantic features) could influence the availability and ease of access to the relevant semantic knowledge to facilitate the redintegration process. If richer semantic representations are more available it increases the probability of successful restoration of the degraded memory trace. However, as noted by Thorn et al. (2009), the influence of long-term knowledge need not be restricted to one time-point or to one mechanism. Based on the current understanding of the NoF effects on episodic memory (Hargreaves et al., 2012; Lau et al., 2018), where the semantic richness of high NoF words (or high NoDF words) contributes to its encoding strength or how well it is processed and learned, it is possible that the same encoding advantage that occurs during the maintenance of these words for subsequent retrieval in the long-term also occurs in the short-term. Future studies could further examine the viability of the encoding variability hypothesis (Hargreaves et al., 2012) in understanding semantic features effects.

Finally, the results suggest that concepts and their corresponding features are stored in semantic memory, and importantly, the featural information is used for the successful retrieval of the item. Even if the word is not semantically rich (e.g., low NoF words), the memory system is still able to optimise performance by using the to-be-recalled item's distinctive features to facilitate retrieval. This highlights how semantic information could be used in an adaptive manner even when participants were not explicitly told to do so. The results from the current study show that the use of featural information to aid in short-term recall occurred without the need to prompt participants, hence indicating a close interaction between long-term semantic representation and short-term recall.

References

- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The Psychology of Learning and Motivation* (Vol. 2., pp 89 – 195). New York: Academic Press.
- Baayen, R. H., Piepenbrock, R., & Van Rijn, H. (1993). *The Celex Lexical Database*. Philadelphia: Linguistic data consortium, University of Pennsylvania.
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation* (Vol. 8., pp 47 – 89). New York: Academic Press.
- Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 14(6), 575 – 589. doi: 10.1016/S0022-5371(75)80045-4
- Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B.,... & Treiman, R. (2007). The English lexicon project. *Behavior Research Method*, 39(3), 445-459. doi:10.3758/BF03193014
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1 – 48. doi: 10.18637/jss.v067.i01
- Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41(4), 977 – 990. doi: 10.3758/BRM.41.4.977

- Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, 46(3), 904 – 911. doi: 10.3758/s13428-013-0403-5
- Campoy, G., Castellà, J., Provencio, V., Hitch, G. J., & Baddeley, A. D. (2015). Automatic semantic encoding in verbal short-term memory: Evidence from the concreteness effect. *The Quarterly Journal of Experimental Psychology*, 68(4), 759 – 778. doi: 10.1080/17470218.2014.966248
- Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion, and memory span. *British Journal of Psychology*, 55(4), 429 – 432. doi: 10.1111/j.2044-8295.1964.tb00928.x
- Goh, W. D., Yap, M. J., Lau, M. C., Ng, M. M. R., & Tan, L. (2016). Semantic richness effects in spoken word recognition: A lexical decision and semantic categorisation megastudy. *Frontiers in Psychology*, 7:976. doi: 10.3389/fpsyg.2016.00976
- Grondin, R., Lupker, S. J., & McRae, K. (2009). Shared features dominate semantic richness effects for concrete concepts. *Journal of Memory and Language*, 60(1), 1 – 19. doi: 10.1016/j.jml.2008.09.001
- Hall, K. C., Allen, B., Fry, M. Mackie, S., & McAuliffe, M. (2016). *Phonological Corpus Tools, Version 1.2* [Software]. Retrieved from www.github.com/PhonologicalCorpusTools/
- Hargreaves, I. S., Pexman, P. M., Johnson, J. S., & Zdrazilova, L. (2012). Richer concepts are better remembered: Number of features effects in free recall. *Frontiers in Human Neuroscience*, 6(73), 1 – 11. doi: 10.3389/fnhum.2012.00073
- Hulme, C., Maughan, S., & Brown, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span.

Journal of Memory and Language, 30(6), 685 – 701. doi: 10.1016/0749-596X(91)90032-F

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D. A., Martin, S., & Stuart, G. (1997).

Word frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(5), 1217 – 1232. doi: 10.1037/0278-7393.23.5.1217

Kowialiewski, B., & Majerus, S. (2018). The non-strategic nature of linguistic long-term memory effects in verbal short-term memory. *Journal of Memory and Language*, 101, 64 – 83. doi: 10.1016/j.jml.2018.03.005

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13), 1 – 26. doi: 10.18637/jss.v082.i13.

Lau, M. C., Goh, W. D., & Yap, M. J. (2018). An item-level analysis of lexical-semantic effects in free recall and recognition memory using the megastudy approach. *Quarterly Journal of Experimental Psychology*, 71(10), 2207 – 2222. doi: 10.1177/1747021817739834

Majerus, S. (2009). Verbal short-term memory and temporary activation of language representations: The importance of distinguishing item and order information. In A. S. Thorn & M. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain* (pp. 244 – 276). Hove, UK: Psychology Press.

Majerus, S., & D'Argembeau, A. (2011). Verbal short-term memory reflects the organization of long-term memory: Further evidence from short-term memory for emotional

- words. *Journal of Memory and Language*, 64(2), 181 – 197.
doi:10.1016/j.jml.2010.10.003
- Martin, N., & Saffran, E. M. (1997). Language and auditory-verbal short-term memory impairments: Evidence for common underlying processes. *Cognitive Neuropsychology*, 14(5), 641 – 682. doi: 10.1080/026432997381402
- Martin, R. C., Lesch, M. F., & Bartha, M. C. (1999). Independence of input and output phonology in word processing and short-term memory. *Journal of Memory and Language*, 41(1), 3 – 29. doi: 10.1006/jmla.1999.2637
- McRae, K., Cree, G., Seidenberg, M., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, Instruments, & Computers*, 37(4), 547 – 559. doi:10.3758/BF03192726
- McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General*, 126(2), 99 – 130. doi: 10.1037/0096-3445.126.2.99
- McRae, K., & Jones, M. N. (2013). Semantic memory. In D. Reisberg (Ed.) *The Oxford Handbook of Cognitive Psychology*. Oxford University Press.
- Miller, L. M., & Roodenrys, S. (2009). The interaction of word frequency and concreteness in immediate serial recall. *Memory & Cognition*, 37(6), 850 – 865. doi: 10.3758/MC.37.6.850
- Monnier, C., & Syssau, A. (2008). Semantic contributions to verbal short-term memory: Are pleasant words easier to remember than neutral words in serial recall and serial recognition? *Memory & Cognition*, 36(1), 35 – 42. doi:10.3758/MC.36.1.35

- Morey, Richard D, Rouder, Jeffrey N, Jamil, Tahira, & Morey, Maintainer Richard D. (2015). Package ‘BayesFactor’. < <http://cran.r-project.org/web/packages/BayesFactor/BayesFactor.pdf> >
- Nairne, J. S., & Kelley, M. R. (2004). Separating item and order information through process dissociation. *Journal of Memory and Language*, 50(2), 113 – 133. doi: 10.1016/j.jml.2003.09.005
- New, B., Brysbaert, M., Veronis, J., & Pallier, C. (2007). The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics*, 28(4), 661 – 677. doi: 10.1017/S014271640707035X
- Pexman, P. M. (2012). Meaning-level influences on visual word recognition. In J. S. Adelman (Ed.), *Visual word recognition: Vol. 2. Meaning and context, individuals and development* (pp. 24–43). Hove, UK: Psychology Press.
- Poirier, M., & Saint-Aubin, J. (1995). Memory for related and unrelated words: Further evidence on the influence of semantic factors in immediate serial recall. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 48A(2), 384 – 404. doi: 10.1080/14640749508401396
- R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
- Rabovsky, M., Schad, D. J., & Abdel Rahman, R. (2016). Language production is facilitated by semantic richness but inhibited by semantic density: Evidence from picture naming. *Cognition*, 146, 240 – 244. doi: 10.1016/j.cognition.2015.09.016

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. doi:10.1080/17470210601147747

Roodenrys, S., Hulme, C., Lethbridge, A., Hinton, M., & Nimmo, L. M. (2002). Word-frequency and phonological-neighborhood effects on verbal short-term memory.

Journal of Experimental Psychology: Learning, Memory, and Cognition, 28(6), 1019 – 1034. doi: 10.1037/0278-7393.28.6.1019

Saint-Aubin, J., Ouellette, D., & Poirier, M. (2005). Semantic similarity and immediate serial

recall: Is there an effect on all trials? *Psychonomic Bulletin & Review*, 12(1), 171 – 177. doi: 10.3758/BF03196364

Saint-Aubin, J., & Poirier, M. (1999). Semantic similarity and immediate serial recall: Is

there a detrimental effect on order information? *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, 52(2), 367 – 394. doi: 10.1080/713755814

Sajin, S. M., & Connine, C. M. (2014). Semantic richness: The role of semantic features in processing spoken words. *Journal of Memory and Language*, 70, 13 – 35. doi:

10.1016/j.jml.2013.09.006

Schneider, W., Eschmann, A., & Zuccolotto, A. (2002). *E-Prime reference guide*. Pittsburgh,

PA: Psychology Software Tools.

Schweickert, R. (1993). A multinomial processing tree model for degradation and

redintegration in immediate recall. *Memory & Cognition*, 21(2), 168 – 175. doi:

10.3758/BF03202729

- Shulman, H. G. (1970). Encoding and retention of semantic and phonemic information in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 9(5), 499 – 508. doi: 10.1016/S0022-5371(70)80093-7
- Shulman, H. G. (1971). Similarity effects in short-term memory. *Psychological Bulletin*, 75(6), 399 – 415. doi: 10.1037/h0031257
- Thorn, A. S. C., Frankish, C. R., & Gathercole, S. E. (2009). The influence of long-term knowledge on short-term memory: Evidence for multiple mechanisms. In A. Thorn & M. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain* (pp. 198-219). New York: Psychology Press.
- Vinson, D. P., & Vigliocco, G. (2008). Semantic feature production norms for a large set of objects and events. *Behavior Research Methods*, 40(1), 183 – 190. doi: JO.37j8/BRM.40.1.J83
- Wagenmakers, E., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., ...& Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, 25(1), 58 – 76. doi: 10.3758/s13423-017-1323-7
- 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. doi: 10.1037/0278-7393.25.5.1256
- Wetherick, N. E. (1975). The role of semantic information in short-term memory. *Journal of Verbal Learning & Verbal Behavior*, 14(5), 471 – 480. doi: 10.1016/S0022-5371(75)80025-9

Table 1
Lexical-semantic properties of low and high NoF words used in Experiment 1

Measure	<u>Low NoF</u>		<u>High NoF</u>	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Number of letters ¹	5.35	1.65	5.38	1.56
Number of phonemes ¹	4.50	1.52	4.39	1.33
Number of syllables ¹	1.66	.72	1.56	.56
Log subtitle frequency ¹	2.68	.44	2.80	.58
Log Celex	2.27	.45	2.27	.64
Number of orthographic neighbours ¹	6.36	6.96	6.36	7.18
Number of phonological neighbours ¹	12.84	12.62	13.84	14.41
Concreteness ²	4.83	.15	4.86	.13
Number of features ³	9.17	1.52	16.20	1.90

¹ Values were from the English Lexicon Project (Balota et al., 2007).

² Values were from Brysbaert, Warriner, and Kuperman (2014) and were based on participants' ratings on a 5-point scale.

³ Values were from McRae et al. (2005) and were obtained using a feature-listing task.

Table 2

The average number of correct recall, omissions, intrusions, repetitions, and order errors made per list for low and high NoF words in Experiment 1

Measure	Low NoF	High NoF	NoF Effect			
	<i>M(SD)</i>	<i>M(SD)</i>	<i>Regression Coefficient</i> ²	<i>p-value</i>	<i>Confidence Interval</i> ²	<i>Bayes Factor</i>
Correct recall	3.63 (.92)	3.92 (.94)	-.048	.003	[-.079, -.017]	9.21 ³
Item errors ¹	1.74 (.72)	1.41 (.70)	.331	<.001	[.222, .440]	1445 ⁴
Omissions	1.42 (.70)	1.16 (.65)	.263	<.001	[.138, .387]	49.64 ⁴
Intrusions	.32 (.16)	.25 (.15)	.069	.092	[-.009, .147]	1.08 ⁴
Repetitions ⁵	.04 (.05)	.03 (.06)	.006	.714	[-.027, .039]	.323 ⁴
Proportion of order errors	.15 (.09)	.15 (.09)	-.0002	.989	[-.023, .022]	.304 ⁴

¹Item errors = Omissions + Intrusions.

²The regression coefficient and its corresponding confidence intervals were obtained based on the linear mixed-effects model analysis.

³The Bayes Factor was obtained by comparing the model [NoF + position + participant] with the model [position + participant] to provide an estimate of the NoF effect. Participant was treated as a random effect.

⁴The Bayes Factor was obtained by comparing the model [NoF + participant] with the model [participant] to provide an estimate of the NoF effect. Participant was treated as a random effect.

⁵Note that repetition errors have been counted twice – first as repetitions and then as order errors.

Table 3

Lexical-semantic properties of words used in Experiment 2a and Experiment 2b

Measure	<u>Low NoF/low</u>		<u>High NoF/low</u>		<u>High NoF/high</u>	
	<u>NoDF</u>		<u>NoDF</u>		<u>NoDF</u>	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Number of letters	7.29	1.78	7.04	2.14	6.83	1.90
Number of phonemes	5.92	1.67	5.71	1.65	5.58	2.00
Number of syllables	2.29	.86	2.38	.88	2.08	.72
Log subtitle frequency	1.91	.31	1.97	.51	2.17	.61
Log Celex	1.50	.51	1.51	.57	1.62	.66
Number of orthographic neighbours	1.79	3.23	1.29	2.24	1.83	3.17
Number of phonological neighbours	4.38	8.98	3.54	7.94	6.04	7.75
Phonological similarity ¹	30.54	3.71	29.86	3.49	29.42	4.31
Concreteness	4.69	.30	4.79	.30	4.84	.17
Number of features	7.42	1.28	13.42	.93	13.83	1.76
Number of distinguishing features	1.92	1.32	2.29	1.63	7.17	1.74

¹Phonological similarity ratings were derived using the Phonological Corpus Tools (Hall, Allen, Fry, Mackie, & McAuliffe, 2016) with higher values indicating phonological dissimilarity among words. Each word was compared to all the other words in the same list-type, hence obtaining an overall mean value of phonological similarity.

Table 4

The average number of correct recall, omissions, intrusions, repetitions, and order errors made for low NoF and high NoF words in Experiment 2a

	1-second presentation rate		2-second presentation rate		NoF Effect			
	Low NoF	High NoF	Low NoF	High NoF	<i>Regression Coefficients</i> ²	<i>p-value</i>	<i>Confidence Interval</i> ²	<i>Bayes Factor</i>
	<i>M(SD)</i>	<i>M(SD)</i>	<i>M(SD)</i>	<i>M(SD)</i>				
Correct recall	2.53 (.95)	2.89 (1.00)	2.59 (.86)	3.01 (.75)	-.066	<.001	[-.098, -.033]	198.30 ³
Item errors ¹	2.21 (.85)	1.99 (.84)	2.53 (.58)	2.04 (.57)	.356	<.001	[.172, .540]	57.75 ⁴
Omissions	1.84 (.81)	1.67 (.84)	2.17 (.59)	1.75 (.53)	.293	<.001	[.136, .452]	40.4 ⁴
Intrusions	.38 (.34)	.32 (.46)	.36 (.24)	.29 (.27)	.063	.096	[-.010, .135]	.834 ⁴
Repetitions ⁵	.09 (.15)	.04 (.07)	.03 (.06)	.06 (.09)	.009	.608	[-.027, .045]	.258 ⁴
Proportion of order errors	.34 (.15)	.29 (.14)	.26 (.16)	.25 (.13)	.030	.140	[-.010, .071]	.552 ⁴

¹Item errors = Omissions + Intrusions.

²The regression coefficient and its corresponding confidence intervals were obtained based on the linear mixed-effects model analysis.

³The Bayes Factor was obtained by comparing the model [NoF + position + participant] with the model [position + participant] to provide an estimate of the NoF effect. Participant was treated as a random effect.

⁴The Bayes Factor was obtained by comparing the model [NoF + presentation + participant] with the model [presentation + participant] to provide an estimate of the NoF effect. Participant was treated as a random effect.

⁵Note that repetition errors have been counted twice – first as repetitions and then as order errors.

Table 5

The average number of correct recall, omissions, intrusions, repetitions, and order errors made for low NoDF and high NoDF words in Experiment 2b

	1-second presentation rate		2-second presentation rate		NoDF Effect			
	Low NoDF	High NoDF	Low NoDF	High NoDF	<i>Regression Coefficients</i> ²	<i>p-value</i>	<i>Confidence Interval</i> ²	<i>Bayes Factor</i>
	<i>M(SD)</i>	<i>M(SD)</i>	<i>M(SD)</i>	<i>M(SD)</i>				
Correct recall	2.39 (.97)	2.63 (1.14)	2.77 (1.02)	2.79 (.83)	-.021	.174	[-.052, .009]	.242 ³
Item errors ¹	2.46 (.75)	2.22 (.74)	2.33 (.75)	2.14 (.66)	.213	.002	[.088, .337]	19.86 ⁴
Omissions	2.13 (.71)	2.00 (.63)	2.06 (.66)	1.92 (.61)	.134	.054	[.0006, .268]	1.31 ⁴
Intrusions	.33 (.28)	.22 (.25)	.28 (.26)	.23 (.25)	.078	.022	[.013, .143]	2.42 ⁴
Repetitions ⁵	.04 (.07)	.03 (.06)	.01 (.03)	.04 (.06)	-.009	.372	[-.030, .011]	.355 ⁴
Proportion of order errors	.35 (.19)	.31 (.21)	.27 (.17)	.29 (.15)	.008	.777	[-.045, .061]	.242 ⁴

¹Item errors = Omissions + Intrusions.

²The regression coefficient and its corresponding confidence intervals were obtained based on the linear mixed-effects model analysis.

³The Bayes Factor was obtained by comparing the model [NoDF + position + participant] with the model [position + participant] to provide an estimate of the NoDF effect. Participant was treated as a random effect.

⁴The Bayes Factor was obtained by comparing the model [NoDF + presentation + participant] with the model [presentation + participant] to provide an estimate of the NoDF effect. Participant was treated as a random effect.

⁵Note that repetition errors have been counted twice – first as repetitions and then as order errors.

Table 6

Lexical-semantic properties of words used in Experiment 3

Measure	<u>Low NoF/low</u>		<u>Low NoF/high</u>	
	<u>NoDF</u>		<u>NoDF</u>	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Number of letters	6.33	1.78	6.10	1.67
Number of phonemes	5.10	1.46	5.05	1.58
Number of syllables	1.90	.76	1.95	.85
Log subtitle frequency	2.28	.48	2.26	.53
Log Celex	1.81	.67	1.88	.48
Number of orthographic neighbours	3.60	5.30	3.21	5.29
Number of phonological neighbours	8.26	11.53	7.52	10.08
Phonological similarity	26.68	3.27	26.31	3.80
Concreteness	4.77	.23	4.75	.20
Number of features ¹	8.17	1.38	8.29	1.15
Number of distinguishing features	.86	.65	4.12	1.50

¹An independent t-test comparing low NoF words used in Experiment 3 and high NoF words used in Experiment 2 showed that low NoF words had a lower NoF values than high NoF words, $t(130) = 22.6$, $p < .001$.

Table 7

The average number of correct recall, omissions, intrusions, repetitions, and order errors made for low NoDF and high NoDF words in Experiment 3

	1-second presentation rate		2-second presentation rate		NoDF Effect			
	Low NoDF	High NoDF	Low NoDF	High NoDF	<i>Regression Coefficients</i> ²	<i>p-value</i>	<i>Confidence Interval</i> ²	<i>Bayes Factor</i>
	<i>M(SD)</i>	<i>M(SD)</i>	<i>M(SD)</i>	<i>M(SD)</i>				
Correct recall	2.78 (.99)	2.87 (1.02)	3.22 (.80)	3.61 (.78)	.040	.003	[.013, .066]	6.05 ³
Item errors ¹	2.30 (.80)	2.16 (.84)	2.04 (.76)	1.71 (.69)	-.234	.001	[-.367, -.101]	25.8 ⁴
Omissions	2.00 (.76)	1.88 (.71)	1.82 (.74)	1.56 (.64)	-.191	.005	[-.320, -.063]	7.43 ⁴
Intrusions	.30 (.28)	.29 (.33)	.22 (.17)	.15 (.13)	-.043	.132	[-.098, .012]	.581 ⁴
Repetitions ⁵	.05 (.05)	.04 (.04)	.03 (.05)	.04 (.04)	-7.50E-11	1.00	[-.019, .019]	.235 ⁴
Proportion of order errors	.26 (.15)	.26 (.17)	.19 (.11)	.16 (.10)	-.011	.410	[-.039, .016]	.280 ⁴

¹Item errors = Omissions + Intrusions.

²The regression coefficient and its corresponding confidence intervals were obtained based on the linear mixed-effects model analysis.

³The Bayes Factor was obtained by comparing the model [NoDF + position + presentation + participant] with the model [position + presentation + participant] to provide an estimate of the NoDF effect. Participant was treated as a random effect.

⁴The Bayes Factor was obtained by comparing the model [NoDF + presentation + participant] with the model [presentation + participant] to provide an estimate of the NoDF effect. Participant was treated as a random effect.

⁵Note that repetition errors have been counted twice – first as repetitions and then as order errors.

Figure 1. Proportion recalled (+ SEs) of high and low NoF words across serial positions in Experiment 1. The error bars represent the standard error of the mean.

Figure 2. Proportion recalled (+ SEs) of low NoF and high NoF words in 1-second and 2-second presentation rate across serial positions in Experiment 2a. The error bars represent the standard error of the mean.

Figure 3. Proportion recalled (+ SEs) of low NoDF and high NoDF words in 1-second and 2-second presentation rate across serial positions in Experiment 2b. The error bars represent the standard error of the mean.

Figure 4. Proportion recalled (+ SEs) of low NoDF and high NoDF words in 1-second and 2-second presentation rate across serial positions in Experiment 3. The error bars represent the standard error of the mean.

List of Figures

- Figure 1 Proportion recalled (+ SEs) of high and low NoF words across serial positions in Experiment 1. The error bars represent the standard error of the mean.
- Figure 2 Proportion recalled (+ SEs) of low NoF and high NoF words in 1-second and 2-second presentation rate across serial positions in Experiment 2a. The error bars represent the standard error of the mean.
- Figure 3 Proportion recalled (+ SEs) of low NoDF and high NoDF words in 1-second and 2-second presentation rate across serial positions in Experiment 2b. The error bars represent the standard error of the mean.
- Figure 4 Proportion recalled (+ SEs) of low NoDF and high NoDF words in 1-second and 2-second presentation rate across serial positions in Experiment 3. The error bars represent the standard error of the mean.

Author Note

This research is based in part on MCL's PhD thesis. This research has been conducted with the support of the Australian Government Research Training Program Scholarship.

Address correspondence to Mabel Lau, School of Psychology, University of Wollongong, Wollongong, NSW, Australia 2522 (email: bxm1994@uowmail.edu.au).