Spaced Learning and the Lexical Integration of Novel Words

Shane Lindsay (sl574@york.ac.uk)
M. Gareth Gaskell (mgg5@york.ac.uk)
University of York, Department of Psychology
York, YO10 5DD, United Kingdom

Abstract

Learning a new word involves integration with existing lexical knowledge. Previous work has shown that sleep-dependent memory consolidation processes are important for the engagement of novel items in lexical competition. We used spaced learning and testing to investigate memory for novel words and their lexicalization across the course of a single day, compared with a day later. We expected that the benefits of spaced learning would enhance knowledge of the novel words, and may provide the interleaving of new and old information that allows lexicalization to occur. The degree to which a new word engaged in lexical competition with phonological neighbors was employed as a marker for lexicalization. We found evidence for enhancements in memory performance following a time period including sleep, but we also found lexical competition effects emerging within a single day. This suggests that while sleep-dependent memory consolidation may be sufficient for lexicalization, it is not a necessary pre-condition.

Keywords: Psycholinguistics; memory consolidation; lexical competition; speech; sleep

Introduction

After a period of neglect, memory consolidation has been intensively studied in recent years. Research on consolidation has shown that after an initial learning experience, memory processes continue without any further exposure, as unstable memory traces are transformed and strengthened from an initial unstable state to a more permanent form. Recent work has shown that sleep has an important role in memory consolidation (see Walker & Stickgold, 2006, for a review), with increasing evidence for sleep-based consolidation in a variety of different memory tasks, including both implicit learning of motor skills and more explicit declarative knowledge (Gais & Born, 2004).

Studies showing sleep-associated memory consolidation have typically demonstrated that newly acquired memories, such as motor skills or the learning of paired associates, become more resilient to interference or decay following sleep. Strikingly, in some cases performance can actually improve following sleep, despite no further practice or rehearsal (Walker & Stickgold, 2006). Another important aspect of consolidation, which we know comparatively less about, is how new memories are integrated with old memories, and the role for sleep in this integration process (Ellenbogen et al. 2007).

One of the few lines of research looking at the integration of new memories is one that has investigated the time-course of lexicalization of newly learned words in adults. One marker for the integration of new words with old is the development of lexical competition, a central construct for theories of spoken word recognition. When listening to a spoken word, the incoming acoustic information unfolds over time, leading to activation of matching word candidates in the listener's mental lexicon (e.g., Marslen-Wilson, 1993). For example, hearing the first part of the word "*captain*" activates competitors such as "*capsule*" and "*captive*" before acoustic information eventually distinguishes it from its competitors. Consequently, the frequency and the number of phonological neighbors a word has is a predictor of how quickly one can recognize it (Luce, Pisoni & Goldinger, 1990).

Gaskell and Dumay (2003) taught participants fictitious novel words that were derived from real words with early uniqueness points (the point at which a word diverges from competitors). Once a novel word has been fully lexicalized it should engage in lexical competition, which would be demonstrated by slower recognition of the novel word's phonological competitors. Participants in Gaskell and Dumay's experiment were able to recognize these novel words accurately immediately after training. However, there was no immediate evidence for these novel words engaging in lexical competition. Instead, it appeared that the integration of the form of these new words into the lexicon had a time course that spanned several days.

Dumay and Gaskell (2007) investigated the time-course of lexicalization by focusing on the possible role of sleep in the integration process. They trained participants on novel words either in the morning or the evening. Results of a lexical competition test immediately after familiarization were compared with a second test after a 12 hour gap (with an additional 24 hour test to control for time of day confounds). The gap between the first and second test that contained a period of sleep (PM-AM) led to lexical competition effects, yet there was no suggestion of a lexical competition after a 12-hour gap not including sleep (AM-PM). This work supports the idea that nocturnal sleep or its associates are important for lexical integration, and other work in the emergence of lexicalization has not found good evidence for lexical competition effects prior to sleep (e.g. Davis et al., 2009; Tamminen & Gaskell, 2008). While participants are able to recognise new words pre-sleep, their representation does not seem to be in a suitable form for lexical competition effects to emerge. The study reported in this paper was designed to see whether enhanced training before sleep could lead to early lexicalisation. In order to improve learning of the new words, we implemented a spaced training regime.

In spaced or distributed learning, exposure to items during learning is spread out over time. Spaced learning is usually contrasted with massed learning, in which the equivalent amount of exposure is given all at once. The enhancement in memory with spaced learning, known as the spacing effect, has been one of the most studied topics in memory research, and has been shown to occur across different time periods, different learning strategies, materials and across species (see Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006, for a review). Spaced learning studies have demonstrated improved performance in the learning of words in young children (Schwartz & Terrell, 1983). For adults, the benefits of spaced learning in vocabulary acquisition has long been known (Dempster, 1987), and this method is often applied in second language teaching. These results suggest that spaced learning of new words should improve performance in explicit tests of memory (e.g., recall). However, given the evidence of the importance of sleep in the emergence of lexical competition, it remains an open question as to whether spaced learning may accelerate the engagement of novel words in the lexical competition process to within a single day.

Spaced learning is often studied within a single session by varying the number of intervening items in a list. In other cases, the effect is studied with intervals across days, weeks, months or even years. However, there has been relatively little investigation of spaced learning with learning sessions spread across a single day (although cf. Shea, Lai, Black & Park, 2000, in the domain of motor learning). Some theories of memory consolidation suggest that successful memory consolidation may require periods of inactivity or rest in order for consolidation, and therefore additional exposure to a task may disrupt consolidation processes (DeZazzo & Tully, 1995). This would mean spaced learning within a day could harm memory consolidation. However, aside from the general evidence for enhancements following spaced learning, there are reasons to think that spacing of learning within a single day may be beneficial, specifically in regard to the process of lexicalization. One theoretical account of the lexicalization process (Davis et al., 2009) is based on the complementary learning systems model (McClelland, McNaughton & O'Reilly, 1995). This theory proposes that overnight sleep provides an opportunity for the integration of newly acquired information (stored hippocampally) with existing knowledge stored in neocortical networks. This dual-system approach avoids the problem of catastrophic interference that can occur in a single connectionist network, where newly acquired knowledge overwrites distributed representations of existing knowledge. The present study was based on the idea that spaced learning of new words during wake periods, interleaved with samples of known words, provides an alternative means of avoiding catastrophic interference. Spaced learning could therefore potentially alleviate the need for off-line consolidation, and lead to pre-sleep lexical competition effects.

In addition to spaced learning, the current study also made use of a related memory phenomenon: the testing effect. Effortful retrieval involved in testing has been shown to enhance memory (Carrier & Pashler, 1987), and along with the further exposure to the studied words, repeated testing should enhance the learning of the novel words and provide additional opportunity for the interleaving in memory of the novel words with existing words.

In the following experiment, participants were intensively trained and tested on novel words in spaced sessions on a single day, and tested once again on the following day. Each training session involved both familiarization with novel phonological forms and tests of knowledge of these new words. We first exposed participants to the novel forms, then tested both recognition and explicit recall using a cued recall task, where participants had to accurately vocalize the new forms. The extent of participants' knowledge was further tested in a recognition task where they had to distinguish new words from foils, and engagement of the new words in lexical competition was measured by a lexical decision task to existing neighbors of the newly learnt words. The combined use of spaced learning and testing should maximize participants' encoding of the new words, and potentially increase the likelihood of showing pre-sleep lexicalization. Consequently, a primary aim of this study was to address the question of whether sleep is a necessary pre-condition for lexicalization of novel words, or is merely sufficient. If sleep is necessary for lexical integration, then we would not expect participants to show lexical competition effects on the first day of learning, no matter how well they had learnt the new words.

Method

Participants

Thirty-four participants from the University of York were tested. All were native English speakers without visual or auditory impairments. Participants received course credit or were paid £15 for participation. During the course of the study one participant dropped out after the first session, leaving 33 participants completing all sessions.

Stimuli and Design

Forty monomorphemic words were chosen to act as the base words (e.g., "*cathedral*"), which were selected from a larger set used by Davis et al. (2009). All words were bi- or trisyllabic, and contained 6 to 11 phonemes (M = 6.7). Frequencies ranged from 2 occurrences per million to 18. All base words had an early uniqueness point, located before the final vowel. Fictitious novel words were derived from these base words (e.g., "*cathedruke*"). These were constructed so that they diverged from the base words at their final vowel. For each base word/novel competitor pair a foil was also created for use in an explicit recognition test (e.g., "*cathedruce*"). This diverged from the novel word only at the final phoneme.

Filler words and nonwords for the lexical decision task were chosen to have properties similar to those of the experimental items. Nonwords were created by changing one phoneme of a real word to form a nonword. The forty novel words were split into two lists for counterbalancing, with participants learning only 20 items. Participants heard both lists of existing words used to derive the novel words in the lexical decision task, with one list acting as a control. Each list was designed so that no items shared the same initial phoneme, in order to prevent confusion in the stem completion task.

Procedure

The experiment was split into five sessions. Participants started the first session at 09:00, 09:30 or 10:00 and came back to the laboratory for the second, third and fourth sessions at two and half hour intervals; the fifth session took place 24 hours after the final session of day 1.

Four tasks were used in the experiment: two exposure tasks and two test tasks. Phoneme monitoring (PM), served to familiarize participants with the phonological form of the words, and stem completion (SC), tested participants' ability to recognize the initial phoneme, remember the word, and sequence its phonological structure in order to correctly vocalize it, and also provided additional training through feedback. These tasks were primarily used to help participants learn the phonological form of the words. The lexical decision (LD) and familiarity decision (FD) tasks were primarily for assessing participants' representations of the novel words, though we expected these tasks would also provide limited additional exposure. Participants did each task four times in total. The structure of session and order or tasks is shown in Table 1.

Table 1. Example schedule for a 09:00 start.

Session	Time	Day	Tasks (in test order)			
Session 1	09:00	1	PM	SC		
Session 2	11:30	1	LD	FD	PM	SC
Session 3	14:00	1	LD	FD	PM	SC
Session 4	16:30	1	LD	FD		
Session 5	16:30	2	LD	FD	PM	SC

Participants took short breaks between tasks if desired. As we were particularly interested in responses to the LD task, this was always done at the beginning of a session, so that responses would not be affected by immediately preceding tasks. Participants did not perform further repetitions of the PM and SC tasks in Session 4, as we wanted to examine changes in lexical competition as a function of passage of time and/or sleep, which could not be attributable to the additional training and exposure to the novel words in the PM and SC tasks. Participants used a USB gamepad to respond in every task except SC, where participants responded vocally using a headset. Sessions with two tasks took approximately 15 minutes and approximately 30 minutes for all four tasks.

Phoneme Monitoring To provide repeated exposure to the novel words, participants had to pay attention to the phonological form of the novel words and monitor which phonemes they contained. Each trial started with presentation on screen of a target phoneme (1 of 5, in separate blocks), followed by auditory presentation of the novel word. Participants received 10 repetitions of each novel word in each PM task. The participant was asked to indicate whether the target sound was present in the novel word by responding phoneme present or absent as quickly and accurately as possible. To help maintain attention, participants were given visual feedback on correct responses and errors, with additional auditory feedback on errors.

Stem Completion To measure both recognition and explicit recall, participants were given the initial CV cluster of target newly learnt words as a cue for recall. For example, for the novel word "*cathedruke*", they would hear "*ca*". Participants were instructed to vocalize their responses as quickly and accurately as possible, and responses were recorded. Participants were given a visual and auditory signal (a tone) before hearing the stem cue. After each trial regardless of their response, participants heard auditory feedback of the novel word that they should have produced.

Lexical Decision An auditory lexical decision task was used to index the integration of the novel words in lexical competition. Participants made responses to the base words (20), control words (20; base words without a new competitor), filler items (28), and non words (68). Responses were made as quickly and accurately as possible. Participants were given visual feedback if their responses were slower than 1000 ms.

Familiarity Decision To measure recognition of the newly learnt words, participants had to make a speeded decision to auditory presentations of target novel words randomly mixed with the same number of foils (20 for each participant) that differed from target novels words by their final phoneme. Two counterbalanced lists were created so that targets and foils appeared in the first or second halves of each task in different orders.

Results and Discussion

Only the participants who completed all sessions were included in the analyses (33). Following Raaijmakers, Schrijnemakers, and Gremmen (1999), only by-participant analyses were carried out, as by-items analyses are unnecessary in counterbalanced designs. In all analyses reported below, participant group was included as a dummy variable to increase statistical power (Pollatsek & Well, 1995). For LD and FD we also controlled for which order participants saw the stimuli across sessions. Main effects and interactions involving these variables are not reported.

Phoneme Monitoring

While we primarily used the PM task as a method for participants to pay attention to the phonological forms of novel words, the structure of the experiment afforded us the opportunity to examine how response time (RT) and accuracy changed over the four sessions involving this task. Across tasks we expected improvements in performance based on task familiarization and motor learning, which themselves may be subject to consolidation effects. This was particularly true for the PM task, as it less clear that enhancements here occur due to benefits from lexicalization.

Across sessions, participants showed decreased mean RT and reduced errors; Session 1 (RT = 1193 ms, CI = 48 ms, error = 6%, CI =.73%), Session 2 (RT = 954 ms, CI = 24 ms, error = 5.3%, CI=.47%). Session 3 (RT=938 ms, CI=24 ms, error = 4.7%, CI = .35%) and Session 5 (RT = 812 ms, CI = 33 ms, error = 4%, CI = .61%). An ANOVA with session as a within-participants factor was highly significant, $F(3, 32) = 70.59$, $p < .001$. There was a significant drop in RT from S1 to S2, $t(32) = 8.5$, $p < .001$, a nonsignificant drop from S2 to S3 ($t < 1$) and another large drop in RT on the next day, $t(32) = 7.7$, $p < .001$.

Analysis of errors showed a decrease in errors over sessions, $F(3, 32) = 7.4$, $p < .001$, though no pairwise comparison between session was significant at $p < .05$.

Stem Completion

We recorded the onset of vocalizations using manual inspection of the waveform and spectrogram for each utterance. Participant errors mostly involved the final syllable being replaced by the final syllable of another novel word or the final syllable of the base word.

The data from two participants were omitted due to recording error in the last two sessions, leaving 31 participants. Examination of the distribution led to the exclusion of the top two RTs (> 3000 ms). An ANOVA on correct vocalizations showed a highly significant difference in reaction times across sessions, $F(3, 30) = 13.7$, $p < .001$. Paired t-tests showed a significant decrease in RT from S1 to S2, $t(30) = 2.28$, $p = .03$, a slight non-significant increase from S2 to S3 ($t < 1$) and a highly significant drop overnight from S3 to S5, $t(30) = 5.9$, $p < .001$.

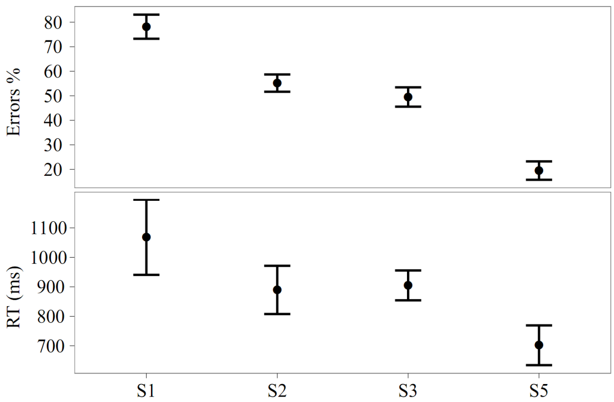

Figure 1: Mean RT and errors across SC sessions. Error bars on figures show 95% confidence intervals calculated as in Loftus and Masson (1994).

Analysis of words correctly remembered and produced showed a highly significant increase in accurate vocalizations across sessions, $F(3, 30) = 114$, $p < .001$. The error data showed a similar pattern to RT scores with large improvements between S1-S2, $t(30) = 6.9$, $p < .001$, a small marginally significant improvement between S2-S3, $t(30) = 2.09$, $p = .058$, and another large enhancement in performance overnight from S3-S5, $t(30) = 12.1$, $p < .001$. RT and error performance for SC is shown in Figure 1.

Familiarity Decision

We examined RTs to correct responses to both target novel words and foils. In an ANOVA with factors of session (4) and target/foil (2), we found main effects of target/foil and session, but no significant interaction. Participants were quicker to respond to novel words compared with foils, $F(1, 32) = 15.9$, $p < .001$, and were quicker across sessions $F(3, 32) = 22.4$, $p < .001$. Planned comparisons showed no significant differences in mean RT between S2-S3 ($t < 1$) or S3-S4 ($t < 1.6$), but there was a significant decrease in RT from S4 to S5, $t(32) = 6.3$, $p < .001$.

Errors showed a reduction across sessions, $F(3, 32) = 10.8$, $p < .001$, and there were slightly fewer errors to targets than to foils, $F(3, 32) = 3.24$, $p = .081$. As with the RT data, the largest changes were for the overnight comparison. Differences between S2-S3 ($t < 1$) and S3-S4 ($t = 1.1$) were not significant, but the S4-S5 change showed a reliable difference, $t(32) = 4.2$, $p < .01$. Reaction time and error performance for SC is shown in Figure 2.

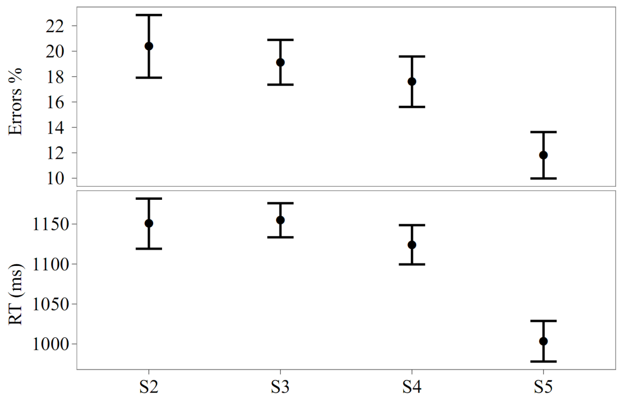

Figure 2: Mean RT and errors across FD sessions.

Lexical Decision

RTs to the base words in the lexical decision tasks were analyzed to examine the extent to which the learning of the novel words led to lexical competition. Our measure of lexical competition was a comparison between the list of 20 base words that were used to derive our novel words, and 20 matched and counterbalanced control words that were not the basis for that participant's novel words. Emergence of lexical competition should be associated with increased RT for the base words relative to the controls.

RT responses less than 300 ms or greater than 2100 ms were excluded (1% of the data). An ANOVA with factors of session and word type (base words vs. control) revealed a main effect of session, with participants showing faster responses across sessions, $F(3, 32) = 9.52$, $p < .001$. This increase was most marked between S2 and S3, and from S4 to S5. There was a main effect of condition, with RTs to base words slower than to control words, $F(1, 32) = 16.6$, $p < .001$, indicating the existence of lexical competition. There was a marginal interaction between session and condition, $F(3,32) = 2.41$, $p = .072$. Examination of Figure 3 shows that while there was a negligible difference between conditions in the first test (S2; $t < 1$). A difference of 23 ms emerged in S3, which was significant $t(32) = 3.4$, $p < .01$. Though the numerical size of the effect was similar in S4 (19 ms), increased variability made it less reliable, $t(32) = 1.5$, $p = .14$[1]. The difference between conditions increased to 45 ms in the final session, $t(32) = 4.5$, $p < .001$.

Analysis of errors showed no significant differences across session or condition, nor an interaction between session and condition (all *F's* < 1).

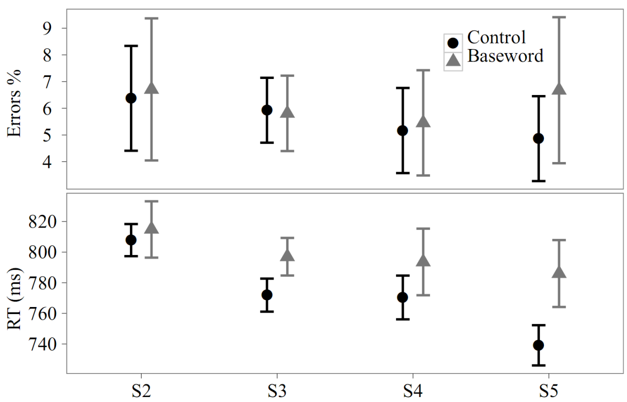

Figure 3: Mean RT and errors across LD sessions.

The reliable effect found in the second lexical decision task (S3) suggests that given the right circumstances a lexical competition effect may emerge without the need for sleep.

General Discussion

Across tasks, in both RT and error data, there was a relatively consistent pattern of improvements in performance shown across sessions. We typically found a pattern of an improvement between the first and the second session, but small or no changes between the second and third sessions for each task. While we found no changes between the second and third sessions, we did find improvements in performance between the third and fourth repetitions of each task (S3-S5 for PM & SC, S4-S5 for LD & FD). This was demonstrated in RT for PM and LD, and in both a reduction in errors and RT for SC and FD. This enhancement is at least consistent with sleep-dependent memory consolidation, but could be attributable to a number of other factors, such as the greater length of time in the interval between the last session for that task on day 1 and the final session on day 2.

More importantly for the aims of this study was whether sleep was a necessary or merely a sufficient precondition for the integration of phonological forms into lexical memory, given results suggesting for the role of sleep in this process (Dumay & Gaskell, 2007). Results from the lexical competition task show evidence for emergence of lexical competition on the second repetition of the LD task on day 1 (S3), which suggests that sleep is not necessary for lexical competition effects following new word learning. We were able to show this pre-sleep lexicalization by using a training and test regime that took advantage of the known benefits that spaced learning and testing has for memory performance.

In this study, the effect emerged on the third session, following two exposure sessions, while most previous studies have used only a single, massed, exposure session (Davis et al. 2008; Gaskell & Dumay, 2003; Dumay & Gaskell, 2007). Part of the rationale for using spaced learning was that it would lead to better developed lexical representations than massed learning. It may be simply the robustness of these representations that underlies the pre-sleep lexical engagement. Alternatively, based on the complementary learning systems approach of McClelland et al. (1995), we hypothesized that via spaced learning and testing, repeated exposure to the novel words and their existing phonological competitors may provide an on-line alternative to the off-line process of consolidation thought to occur during sleep, and hence allow for pre-sleep integration and lexical competition effects. Our results provide tentative support for this hypothesis.

It should be noted that the magnitude of the competition effect appeared to double in the time period containing sleep, which suggests that even in these optimal learning circumstances sleep is still important, and may play a role over and above what is possible within a day. Increases in size of the lexical competition effect are consistent with evidence from the other tasks showing lexical representations were considerably strengthened between the third and fourth repetitions of each task. A striking example of this enhancement is found in the stem completion tasks. Participants in the first session only accurately recalled 22% of items. This improved to just under 50% correctly recalled in the second and third sessions, followed by an increase to 80% correctly recalled in the final session. Nonetheless, any apparent benefits of sleep cannot be secure on the basis of the current experiment alone.

One issue for the design of this study is that in the comparison for PM and SC between S3 and S5, there was a testing session for LD and FD in S4, whereas the

[1] Although we do not report a full by-items analysis, the same pattern of results was found, with the comparison at S4, $t2(39) = 1.9$, $p = .06$.

comparison between S2 and S3 did not have this additional session in between. These additional tests and exposure may be part of the reason for the large improvements in memory from S3 to S5 compared with the absence of significant changes from S2 to S3. However, the changes in LD and FD overnight are not thought to be attributable to additional exposure. Gaskell and Dumay (2003) compared lexical decision followed by a recognition task with lexical decision without a subsequent recognition task, and found the additional exposure involved with the recognition task did not affect the emergence of lexical competition. A more general issue is that this study does not allow us to determine whether the pre-sleep competition effect stemmed from enhancements due to spaced learning, spaced testing, or a combination of the two. We aim to tease apart these possibilities in future work.

Because of the repeated use of lexical decision intermixed with familiarization of the novel forms, participants reported being explicitly aware of the relationship between the base words and novel words, and some participants produced the base word when given the stem cue in the stem completion task. It is unclear to what extent participants' awareness of this relationship impacted upon the nature of the results. One important factor not yet considered is the role of reconsolidation processes (Walker & Stickgold, 2006). It is possible that the repeated activation of the existing base words in close temporal proximity to the phonologically related novel words was important in determining the time course of lexical competition found in this study.

We used a mixture of tasks that tapped into different cognitive processes and memory systems that are involved in the development of lexical representations. We found the time course of performance changes was relatively similar regardless of the type of knowledge tested. While results from other studies show that time and sleep are important for memory consolidation and the learning of new words, the results of this study suggest for the first time that lexical integration can occur without the need for sleep. This study further demonstrates the benefits of spaced learning and testing for memory enhancement, and these results shed valuable light on the processes involved of how we integrate our existing knowledge with new experiences.

Acknowledgments

The research was supported by a UK ESRC research grant RES-063-27-0061 awarded to M.G.G.

References

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory and Cognition,* 20, 633-642.

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.

DeZazzo, J., & Tully, T. (1995). Dissection of memory formation: From behavior pharmacology to molecular genetics. *Trends in Neuroscience, 18,* 212–218.

Dempster, F. N. (1987). Effects of variable encoding and spaced presentations on vocabulary learning. *Journal of Educational Psychology, 79(2),* 162-170.

Dumay, N., & Gaskell, M.G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science, 18(1),* 35-39.

Ellenbogen, J. M., Hu, P. T., Payne, J. D., Titone, D., & Walker, M. P. (2007). Human relational memory requires time and sleep. *Proceedings of the National Academy of Sciences, 104(18),* 7723-7728.

Gais, S., & Born, J. (2004). Declarative memory consolidation: Mechanisms acting during human sleep. *Learning & Memory, 11,* 679-685.

Gaskell, M. G., & Dumay N. (2003). Lexical competition and the acquisition of novel words. *Cognition, 89,* 105-132.

Luce, P. A., Pisoni, D. B., & Goldinger, S. D. (1990). Similarity neighborhoods of spoken words. In G. Altmann (Ed.), *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*. Cambridge, MA: MIT Press.

Loftus, G.R., & Masson, M.E.J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review, 1,* 476-490.

Marslen-Wilson, W. (1993). Issues of process and represenation in lexical access. In G. Altmann & R. Shillcock (Eds.), *Cognitive Models of Language Processes: Second Sperlonga Meeting*. Hove: Erlbaum.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning-systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102,* 419-457.

Pollatsek, A., & Well, A. D. (1995). On the use of counterbalanced designs in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 785-794.

Raaijmakers, J. G. W., Schrijnemakers, J. M. C., & Gremmen, F. (1999). How to deal with "the language-as-fixed-effect fallacy": common misconceptions and alternative solutions. *Journal of Memory and Language, 41,* 416-426.

Schwartz, R. G., & Terrell, B. Y. (1983). The role of input frequency in lexical acquisition. *Journal of Child Language, 10(1),* 57-64.

Shea, C. H., Lai, Q., Black, C., & Park, J. H. (2000). Spacing practice sessions across days benefits the learning of motor skills. *Human Movement Science, 19(5),* 737-760.

Tamminen, J., & Gaskell, M. G. (2008). Newly learned spoken words show long-term lexical competition effects. *Quarterly Journal of Experimental Psychology, 61,* 361-371.

Walker, M. P., & Stickgold, R. (2006). Sleep, Memory, and Plasticity. *Annual Review of Psychology*, *57*, 139-166.