Applied Psycholinguistics **33** (2012), 665–689
doi:10.1017/S014271641100052X

Expert knowledge, distinctiveness, and levels of processing in language learning

STEVE BIRD
United Arab Emirates University

Received: February 3, 2010 Accepted for publication: October 14, 2010

ADDRESS FOR CORRESPONDENCE
Steve Bird, Department of Linguistics, College of Humanities and Social Sciences, United Arab Emirates University, P.O. Box 17771, Al Ain, United Arab Emirates. E-mail: sbird@uaeu.ac.ae or birdsa@gmail.com

ABSTRACT
The foreign language vocabulary learning research literature often attributes strong mnemonic potency to the cognitive processing of meaning when learning words. Routinely cited as support for this idea are experiments by Craik and Tulving (C&T) demonstrating superior recognition and recall of studied words following semantic tasks ("deep" encoding) compared to structure-related tasks ("shallow" encoding). However, participants in C&T were not language learners but native speakers of English studying known English nouns. These experiments have never been directly replicated using nonnatives to establish the relevance of the findings to nonnatives and learners. The present study replicated C&T Experiment 5, comparing effects of shallow and deep encoding tasks on subsequent recognition of target words by native and nonnative speakers of English with equivalent short-term memory function. The results showed depth effects similar to C&T for all participants, indicating that C&T's results do generalize to less proficient speakers of the target language. It is crucial, however, that nonnative speakers of English benefited less from semantic encoding than native speakers, suggesting an effect of preexisting knowledge representations on mnemonic effects derived from semantic processing, and hence, a limit to the relevance of C&T for learners. Results are discussed in terms of the constructs of the mental lexicon, expert knowledge, distinctiveness and levels of processing in memory research and language learning.

The vocabulary learning research literature often attributes strong mnemonic potency to the cognitive processing of meaning when learning new words (Coady & Huckin, 1997; Nation, 2001; Schmitt, 2008; Sokman, 1997). This attribution is based in part on studies in experimental psychology demonstrating superior recognition and recall of studied words following semantic study tasks compared to structure-related study tasks (Craik & Lockhart, 1972; for reviews, see Craik, 2002, 2007; Lockhart 2002; Lockhart & Craik, 1990). For example, in a much cited series of experiments, Craik and Tulving (C&T; 1975) tested native speakers of English using English nouns as study items, and demonstrated that subsequent

© Cambridge University Press 2011 0142-7164/11 $15.00

https://doi.org/10.1017/S014271641100052X Published online by Cambridge University Press

recognition and recall of these studied English nouns depended on the type of orienting question that participants had answered about the words during the study phase. Words that had been paired with a semantic orienting question (e.g., "Is the word a type of flower?") during study were recognized much more accurately than words that had been paired with a nonsemantic question (e.g., "Does the word rhyme with train?"; "Does it have the following consonant–vowel pattern?") on a recognition test in which all of the "old" study phase words had been randomly combined with "new" distractor words in a list and participants were asked to circle only the 40 words that had been presented in the study phase.

This semantic processing advantage has been replicated many times in behavioral studies (e.g., Fisher & Craik, 1977; Hannon & Craik, 2001; Moscovitch & Craik, 1976; for review, see Lockhart & Craik, 1990), and related effects have been observed in more recent neuroimaging studies (Guo, Zhu, Ding, Fan, & Paller, 2004; Nyberg, 2002; Otten, Henson, & Rugg; 2001). Together, these studies confirm that semantic–conceptual processing can create a significant mnemonic benefit relative to other kinds of mental engagement with study materials. The term "deep" processing was coined to refer especially to semantic–conceptual processing (Craik, 2002; Craik & Lockhart, 1972); in contrast, "shallow" processing was used to refer to many nonsemantic processing tasks, such as ones that direct attention to structural characteristics of the words (e.g., the consonant–vowel [CV] patterns of words). This continuum from deep to shallow was captured in Craik and Lockhart's (1972) levels of processing (LOP) framework. The general conclusion presented in the LOP model is that, when deciding on the best way to remember information, "smart money will take deep processing" (Lockhart, 2002, p. 399).

LOP has been a very influential model in vocabulary learning research, theory, and practical educational applications. C&T is one of the most cited studies in the literature, often serving as a source of evidence for the merits of meaning-based language learning methods and materials (Coady & Huckin, 1997; Hill & Laufer, 2003; Huckin & Coady, 1999; Hulstijn, 1992; Hulstijn & Laufer, 2001; Meara, 1997; Moonen, De Graaff, & Westhoff, 2006; Nation, 2000, 2001; Schmitt, 2008). One reason for this influence in language learning research might stem from the remarkable magnitude of differences in test scores reported by C&T for semantic encoding conditions. For example, in C&T's Experiment 1, the proportion of words remembered in the shallow encoding condition was 0.14, whereas in the deep encoding condition the proportion was 0.96. Such a large difference might make semantic encoding seem like a holy grail of vocabulary learning techniques if these differences apply to language learners and learning contexts. But do they? Two observations are important here: first, there exists evidence against the LOP as a model that can reliably predict memory performance across diverse encoding and retrieval tasks (e.g., Morris, Bransford & Franks, 1977; Roediger & Gallo, 2002). Roediger and Gallo (2002) argue that the depth effects observed in C&T's experiments may have been achieved because the particular encoding and retrieval task pairings were optimal for a semantic task advantage (for a rebuttal, though, see Craik, 2002). Second, and more important for the present study, the participants in most LOP studies in experimental psychology were native speakers of the target language. That is to say, the participants in those

https://doi.org/10.1017/S014271641100052X Published online by Cambridge University Press

experiments were highly proficient users of the target language and already had robust lexical representations for the target words. C&T tested the effects of various encoding conditions on episodic memory for known words. Episodic memory has been implicated in language learning (e.g., Verfaille, Croce, & Milberg, 1995), and so the impressive results reported in C&T may be relevant to language learning, but there is also evidence to suggest that the what is already stored in long-term memory can affect episodic memory function (Van Overschelde, Rawson, Dunlosky, and Hunt, 2005; see discussion below). Because C&T's participants were native speakers of the target language, it remains unclear whether their successful episodic memory results generalize to situations where the participant has a relatively weak knowledge of the target language. Foreign language learners will almost certainly have weaker knowledge of and processing fluency with respect to target words that are not in their native language (Dijsktra, 2007). How might these relative weaknesses affect episodic memory function of the kind tested in C&T? This question is of relevance both to LOP research and language learning: if proficiency influences LOP effects, we have evidence that it is not just the processing operations carried out during study that matter (as LOP traditionally holds; Craik and Lockhart, 1972), but also the nature of the preexisting knowledge base matters (Craik, 2007). If so, this would reduce the relevance of C&T and the original LOP model to language learners. Again, LOP is one of the most influential and frequently cited models in vocabulary learning research and yet it is unclear whether it should be considered highly relevant.

LOP studies in the applied linguistics literature have reported rather inconsistent results, suggesting perhaps that there may be proficiency-related moderating factors in LOP effects. Although some studies report relative benefits from semantically focused tasks (e.g., Cho & Krashen 1994; Hulstijn, 1992; Hulstijn, Hollander, & Greidanus, 1996; Joe, 1998; Knight, 1994; Luppescu & Day, 1993), others have reported no advantage or negative effects from semantic conditions (e.g., Barcroft, 2002; Pressley, Levin, & Miller, 1982; Sagarra & Alba, 2006; Scribner, 2000). A variety of hypotheses have been put forward by researchers to account for the inconsistent findings (see discussion below). Tracing the causes of these effects is not easy, however, in part because the designs of the experiments have varied considerably among themselves and when compared to C&T and colleagues' designs. There are differences in the language proficiency of the participants, the frequency of the target words, the kinds of deep and shallow tasks used during training and testing, and also the time intervening between study and test. All of these factors have been shown to affect performance on memory tests (Morris et al., 1983; Roediger & Gallo, 2002; Rohrer & Pashler, 2007). It therefore remains unclear to what extent semantic processing of the sort studied by C&T aids foreign language learners. One important way of clarifying the relevance of C&T's work and the LOP model in language learning research is by directly replicating C&T's experiments. This allows clearer comparison of C&T's results to those for nonnative speakers of the target language while holding constant many other potential intervening variables. The present study replicated C&T, Experiment 5 with this aim in mind (see General Design Section).

Theoretically, various differences in lexical representations and processing capabilities in a foreign language learner and a native speaker of the language

https://doi.org/10.1017/S014271641100052X Published online by Cambridge University Press

could have significant effects on memory and learning. For example, there exists evidence from functional magnetic resonance imaging studies demonstrating that cognitive processing in a foreign language requires more computational load as measured by activation of the neural networks underpinning language processing (Hasegawa, Carpenter, & Just, 2002). This processing load increase may be related to the quality of the lexical representations in the native and foreign language and/or the efficiency of processing derived from experience activating and processing lexical information in the first or native language (L1) relative to the second or foreign language (L2). If processing resources are less efficient and overly taxed during a study phase, it seems possible that subsequent retrieval of what was processed may be impaired and subsequent retrieval less accurate. There is evidence for proficiency-related effects of this sort (Sagarra & Alba, 2006; Scribner 2000). Scribner (2000) compared beginner, intermediate, and advanced L2 Spanish speakers' learning of new L2 words using either rote memorization, the keyword mnemonic method (which focuses on the use of visual imagery and L1–L2 phonetic links during encoding), or a semantic processing condition called a "network of conceptual relations." In this condition, participants saw the new foreign language word (e.g., *hebilla*) and a native English language translation (buckle); then, all in Spanish, participants answered a personal question about the target word, answered questions about the target word based on previous experience (e.g., "What are the buckles for?"), and wrote a sentence using the L2 word. The results showed that the network of conceptual relations condition was more effective than the keyword mnemonic method and the rote method for intermediate and advanced students on recall tests. However, for beginners, the network of associations condition was the worst encoding method, and the keyword method was the best. Sagarra and Alba (2006) found essentially the same poor test results for semantic elaboration relative to the keyword method with 778 beginner L2 learners. One explanation for the poorer performance on tests of this sort is that, at a lower language proficiency level, learners may need to focus limited capacity processing resources (e.g., working memory; for a discussion, see Hasegawa et al., 2002) on the form of new words (orthographic–phonological) as much as or more than the meaning, a focus encouraged in the keyword method condition but not in typical semantic processing tasks (e.g., Barcroft, 2002).

Distinguishing between effects of processing demands (working memory, attention, etc.) and effects of knowledge bases on memory performance is not easily achieved experimentally, but several designs (employed in the present experiment; see General Design, below) may be useful. In most learning studies the amount of time allowed for studying new vocabulary was restricted (e.g., Barcroft, 2002), and this restriction would be likely to negatively affect less proficient participants relative to a native speaker of the language if the lower proficiency participants simply cannot perform the task quickly enough. In order to mitigate the possible negative effects of time constraints, study time in the present experiment was not restricted and the target words were known by all participants. Because the words were already known, the task was very simple to perform (as demonstrated by low error rates, for example), and because time was not restricted, it was assumed that working memory and other processing resources would not be taxed in any significant way, and hence, would not be likely to be the source of

https://doi.org/10.1017/S014271641100052X Published online by Cambridge University Press

performance differences between native and nonnative speakers of the target language at test.

Using this approach, the present study examined the effects of preexisting word representations on memory performance. Whatever the precise mechanisms that may be taxed in L2 processing relative to the native language, it seems likely that one contributor is the quality of the lexical representations of words on which processing is based. Various models of lexical knowledge have been proposed that can perhaps provide some theoretical account of how the quality of a word representation in the mental lexicon could affect processing and subsequent memory and learning performance (Aitchison, 1994; Craik, 2007; Meara, 2009; Nation, 2001). For example, Aitchison (1994) uses the metaphor of a gigantic multidimensional cobweb to describe the mental lexicon's structure, with words linked to other features, words, and concepts. This general notion of a web or network is consistent with other current models of lexical representations related to LOP. In an effort to define this network concept in more detail in the context of language learning, Meara (2009) proposes a model of the mental lexicon in which a network of formal and semantic relations can be described not only in terms of a how many words are "known," as measured by a yes/no response on a word/nonword lexical decision task (referred to as vocabulary "size" or "breadth") but also in terms of vocabulary "organization" or "depth," as measured by the number of connections the word has to other words in the mental lexicon (determined using word association techniques, for example). In this model, although both a native and nonnative speaker of a target language might be able to recognize a given letter string as a "word" in that language, hence be said to "know" the word, the native speaker presumably possesses knowledge of the given word that is more connected to other words and features in the network, and represented in a more "organized" or "deeper" fashion in Meara's terminology. Meara does not discuss how this deeper and more organized structure might affect memory function, but reasonable speculation can be put forward (for discussion, see Dijkstra, 2007). Within the memory research literature, Craik (2007) argues that well-established knowledge in long-term memory is represented in a highly organized network (Craik uses the term "schema" for this kind of network; in the present article we use the term "network" for conceptual and terminological consistency with the language learning literature). In this model, individual items are represented in structured networks whereby items are both related to each other but also differentiated from each other through their unique connections to item-specific features in a network. Craik suggests that the increased number of organized and activated connections or nodes available can act as more numerous potential retrieval cues for the given word at test; that is, the activated information provides more potential paths back to the original processing episode. A semantic orienting task is here seen as one that activates more of a word's connections relative to a shallow task. This quantitative difference in cues then results in differences in recognition accuracy. Many language learners will probably have had the phenomenological experience of searching their memory to retrieve a word that was encountered recently (known or new), and finding that it can be brought to mind via some related feature or word involved in the study episode: the sound, the meaning, related words that the learner knows, and so forth. The more cues

https://doi.org/10.1017/S014271641100052X Published online by Cambridge University Press

available for this backward tracking, the more likely it is that the target word will be retrieved correctly. This is the essence of the concept of a *reliable retrieval cue.* In a native speaker of the target language, processing of a given word activates many features that can act as useful cues. The native speaker is a kind of "expert" user of the language relative to a learner, possessing much more mnemonically useful information about target words. If this construal of the mental lexicon is accurate, LOP effects should depend on how well known the target word is. A learner of the language presumably would be at a disadvantage with respect to the breadth and depth of knowledge concerning a given word, offering fewer and less well organized features that could act as reliable retrieval cues.

In support of this approach to LOP effects and mnemonic effectiveness, there is strong evidence in other areas of memory research that implicates better quality of knowledge representation in superior memory function. For example, numerous studies have demonstrated that experts in a given domain, or "high knowledge" (HK) participants, are much better than "low knowledge" participants on verbal–conceptual memory tests when the information to be remembered falls within the field of HK expertise (Einstein & Hunt, 1980; Hunt & Einstein, 1981). Van Overschelde et al. (2005) found that recall of names of American football teams that had been presented during encoding were recalled more accurately during a later recall task by participants who had good knowledge of professional football relative to participants with limited knowledge. This pattern has been replicated using many kinds of task, including chess positions, dinner and drink orders, architectural drawings, city street names, and computer program code (for a review, see Rawson & Van Overschelde, 2008). In this domain of research, the explanation for better memory performance is not related primarily to attention or processing demands but rather to the quality of the existing representation of the information on which processing is based. If native speakers of a language are regarded as experts in this sense, the structure and depth (i.e., in Meara's, 2009, sense described above) of representation of words in the mental lexicon are important for improved memory and for learning (Verfaille et al., 1995). Once again, it is not easy to distinguish the contribution of a lexical representation within a network from the cognitive processes applied to that network of knowledge, and the present focus on representations does not imply that they are primary in LOP effects, only that the nature of these representations should be considered when discussing the causes of LOP effects.

When considering lexical representations as networks of relations, and of retrieval cues as being important for accurate recall and recognition, another potentially useful psychological construct often neglected in the language learning literature is "distinctiveness" (Gallo, Meadow, Johnson, & Foster, 2008; Hunt, 2003; Hunt & Worthen, 2006). Distinctiveness has been defined as "the processing of differences among elements that are similar on some dimension . . . That is, distinctive processing is not the absolute processing of differences but is processing that marks something as different from other things that are related on some dimension" (Hunt, 2003, p. 812). Rawson and Van Overshelde (2008, p. 647) elaborate this definition by offering two candidate cognitive subprocesses involved in distinctive processing: one, termed "organizational processing," involves processing of a word (e.g., "antelope") to slot it into a preexisting category that

https://doi.org/10.1017/S014271641100052X Published online by Cambridge University Press

defines similarities among items (e.g., "hoofed animals"). The other subprocess, termed "item specific processing," involves processing of the properties of an item which uniquely define that item and distinguish it from all other similar items within that category (e.g., features of an antelope that make it different from a horse, cow, etc.). Implicit in these definitions is the idea that an expert possesses a representation of a word or concept that allows certain kinds of processes to occur. A word that is well differentiated and also connected well to other representations permits effective processing and subsequent efficient retrieval because that target word is uniquely specified in the mental lexicon. Knowledge representations are thus employed in processing to render a target word distinctive in memory, hence more easily recognized and recalled on memory tests because the item is less likely to be confused with other similar candidates presented during study. The less detailed and organized the representation is within a network of relations, as in the case of nonexperts, the less it can facilitate accurate retrieval processes. In this connection, it has been well established that in order for a word to be thoroughly learned, it must be met in varying contexts over long periods of time (Nation, 2001; Schmitt, 2008), a demand that should benefit a native speaker relative to a nonnative in terms of the knowledge attained and the organized details available during study and test. On this theoretical basis, memory performance after a study session crucially depends on long-term preexisting lexical representations and a cognitive system that can render studied items distinct in memory. In the case of language learners, presumably the lexical representation for known words will almost always be less detailed and differentiated than it would be for native speakers just on the basis of frequency of exposure (Schmitt, 2008). If so, memory performance in a typical LOP experimental context should be poorer relative to natives, paralleling the results reported in HK–low knowledge studies.

Returning to the LOP framework, Craik (2007, p. 131) hypothesizes that

> . . . knowledge representations are highly organized and differentiated (like a well-ordered library), so that incoming events processed in terms of such knowledge bases will result in a processed record that is *distinctive* relative to many other encoded representations. At the time of remembering, the framework provided by schematic knowledge also facilitates efficient retrieval. A library analogy may again be helpful; if a new acquisition is "encoded deeply" it will be shelved precisely in terms of its topic, author, date, etc., and the structure of the library catalog will later enable precise location of the book. If the new book was simply categorized in terms of its surface features ("blue cover, $8'' \times 10''$, weighs about a pound"), it would be stored with many similar items and be difficult or impossible to retrieve later.

A perceptual feature analysis task (e.g., a CV structure task) does not require that the semantic representation become actively engaged in analysis, even though it may exist in the mental lexicon. If the full lexical representation is not employed to analyze the word, the mnemonic potential of the whole preexisting representation may not be fully realized, and subsequent retrieval less accurate. Distinctiveness and expert knowledge as presented here may offer some insight into a number of experimental findings related to semantic processing in

https://doi.org/10.1017/S014271641100052X Published online by Cambridge University Press

vocabulary learning. For example, there has been considerable debate concerning how best to present new vocabulary sets to learners (for a review, see Ertin & Tekin, 2008). Some researchers have argued that presentation of words in semantically related sets is best because it mirrors the structure of word networks in native speakers (Aitchison, 1994; Hashemi & Gowdasiaei, 2005). Others have argued that presentation in unrelated sets is more beneficial because when words that share numerous common elements and a superordinate concept are introduced at the same time, these words may interfere with each other and have a negative effect on their retention due to cross-association (Finkbeiner & Nicol, 2003; Nation, 2000). In support of this latter claim, Erten and Tekin (2008) report a study of fourth graders who studied words that were presented either in semantically related sets or semantically unrelated sets. They found that words studied in semantically unrelated sets were better recalled than vocabulary that had been presented in semantically related sets. In addition, test completion time was much longer for the semantically related vocabulary items, indicating a slower recall of vocabulary. The authors argue that presenting words in semantically related sets caused interference due to cross association of the new word meanings. In other words, in the absence of well differentiated word networks in the lexicon, the problem for the learner could be a reduced ability to engage in item specific processing that renders similar items unique within a given set. Thus, although expert native speakers can process and discriminate between semantically related pairs because they also know what the differences are between these items, learners may not possess this more detailed knowledge. In order for a semantically related network of words to be useful for memory, a complementary knowledge of differences is also required (item-specific knowledge and processing). On this account, only on the basis of a highly organized and differentiated knowledge resource can semantic–conceptual processing be employed to render information distinctive in memory. Such a knowledge base presumably needs to be developed in a learner before deep semantic processing can be thoroughly effective. The semantically differentiated sets provided in Ertin and Tekin's (2008) study may have been helpful to learners because these sets provided item-specific information for the learner that rendered each item more distinctive.

Similarly, a number of language learning studies comparing different mnemonic strategies for foreign language vocabulary learning have found that semantic processing alone is often less successful than artificial strategies such as the keyword method, which exploits sound, bizarre imagery, orthography, and meaning links between the learner's L1 and the language being learned (e.g., Pressley et al., 1982). This has been especially true when participants are low proficiency learners rather than advanced and experienced in learning the language (Sagarra & Alba, 2006). Distinctiveness may again help explain this tendency. Low proficiency learners must rely on information in the learning environment that renders the new item as distinctive as possible in memory. Without the representation possessed by expert natives, semantic processing may be less effective than mnemonic devices because there is no internal means of making the items distinct. By contrast, natives already possess the expert knowledge of the targets that is needed to render the items distinctive during encoding. In essence, for beginners, item-specific processing provided by the method and materials is supportive of retrieval (Erkin & Tekin,

https://doi.org/10.1017/S014271641100052X Published online by Cambridge University Press

2008); experts do not require this kind of intervention, and may even suffer in performance if it is enforced (Scribner, 2000).

In summary, experts in a given domain not only cluster similar meanings together (organizational processing), they can also render each item distinct (item-specific processing) within that organized network of semantically similar words. Thus, semantic–conceptual processing advantages may depend on the quality of the representation that is activated by the task demands. Learners appear to need support in the environment to establish distinctive memory traces for new words. They may, for example, benefit more from form-related orienting tasks because they do not as yet have well-established formal representations of words (spelling/pronunciation), which are required prior to focusing on other features such as semantic–conceptual similarities and differences between words (Barcroft, 2002).

Assuming nonnatives have weaker representations for all known words (based on lower frequency of exposure in a lifetime), memory should be poorer in nonnatives than natives when all other variables are held constant in a C&T style experimental context. A replication of C&T, extended to include natives and nonnative speakers of the target language, as well as words of different frequencies, allows us to test these hypotheses. If natives perform in a manner different to nonnatives, this provides evidence that the LOP effect can be moderated by the lexical knowledge base available during study, lending support the hypothesis that the interaction of processing tasks and lexical representations affects LOP (Craik, 2007).

One moderating factor in these predictions that should be mentioned is the effect of word frequency on memory tasks. C&T used a recognition memory task to test memory in most of their experiments. Lower frequency words are typically recognized more accurately than high frequency words in recognition memory tests (Gregg, 1976; Malmberg, Steyvers, Stephens, & Shiffrin, 2002). This is often assumed to be because lower frequency words have more noticeable and unusual characteristics that make them more distinctive and draw greater levels of attention at encoding (for a review, see Malmberg, 2008). Note again, however, that participants in such experiments are normally native speakers of the target language. In order to process a word as "unusual" in some respect, knowledge of what is "usual" is presumably required as well. How might word frequency affect lower proficiency participants? In the present study it was difficult to make a prediction because it was unclear exactly how the words were represented in the nonnatives and because there is very little research done in the psychology literature with language learners or proficient nonnatives when testing distinctiveness and word frequency effects on memory.

To briefly summarize the motivations for the present study's general design, C&T (1975) is one of the most frequently cited studies in virtually all applied linguistics vocabulary learning literature, and yet the participants in C&T were not language learners and the target words were taken from their native language. As the research literature summarized above strongly suggests, language proficiency may be an important variable affecting the mnemonic benefits of a semantic task of the sort used in C&T. There is evidence that preexisting knowledge of the target language and associated cognitive processes are moderating variables in LOP

https://doi.org/10.1017/S014271641100052X Published online by Cambridge University Press

effects. However, it is not easy to directly compare among the many studies in the applied linguistics literature and to C&T's seminal experiments because many variables apart from the language proficiency of the participants have differed from C&T's experimental design. No extant experiments have directly tested the effects of proficiency within C&T's original experimental paradigm. The present experiment was design to fill this gap in the literature.

Given the research literature reviewed above, we saw reason to hypothesize that the often-cited C&T study may not be as relevant to language learners as much vocabulary learning research has suggested, and that this might be due to fundamental differences in the knowledge bases of natives and nonnatives in their respective mental lexicons. The present study is relevant both to the LOP model in terms of the role preexisting knowledge might play in LOP effects (Craik, 2007), and to language learning research in contributing to our understanding of the interplay of the mental lexicon and study tasks. The present study was designed to explore main effects and interactions of three factors: the effects of deep and shallow encoding operations (depth of processing) on memory for higher and lower frequency words (frequency) in both native and nonnative speakers (proficiency). These factors were tested through a replication of C&T's Experiment 5.

EXPERIMENT

General design

In designing a replication of C&T's results, we examined the 10 studies for a suitable experiment that could transfer to nonnative speakers of the target language while eliminating historically problematic confounds. Of the potential confounds, time on task during the study phase was identified as important. The reason time on task control is crucial is that, when comparing deep and shallow encoding tasks, it has often been found that the deep task can take significantly longer to perform than the shallow task, hence, study time is confounded with study tasks and we cannot say with confidence whether the task itself or just the time spent processing the words was the cause of the effect. In C&T's study, their first two experiments had a time confound in the study phase, and they reasoned that in order for their claims about the superiority of semantic orienting tasks to be defensible, an experiment in which study time was equivalent over tasks was required. C&T designed their Experiment 5 to address this confound. They designed a shallow but time consuming task that participants needed more time to complete than the deep task (see Design and Procedure Section). The results again showed a semantic task superiority in the test phase comparable to their earlier experiments, and the researchers therefore concluded that time on task alone was not responsible for the reported semantic superiority effect; rather, it was clearly the qualitative nature of the task that caused the improved test phase performance. These results were crucial to C&T's claim for the semantic task superiority found in their other nine experiments.

The time on task confound has also been found in language learning research. For example, Hulstijn and Laufer (2001) compared incidental vocabulary learning scores derived from learning tasks that varied in what they refer to as

https://doi.org/10.1017/S014271641100052X Published online by Cambridge University Press

"involvement load" (*involvement* being construct comprised of a motivational component as well as cognitive processing components; for a discussion, see Hulstijn & Laufer, 2001), a core aspect of which involves encoding operations similar to LOP's concept of elaborated semantic processing. The task was (a) reading comprehension with marginal glosses for target vocabulary words; (b) reading comprehension with the target vocabulary words deleted and listed with their definitions on a separate piece of paper; these then had to be selected for semantic fit in context and inserted into the blanks in the text; or (c) a composition task where participants had to use the target words in a written task. The authors correctly predicted superior retention of target words for the third condition based on their involvement load index. However, the task identified as having the highest involvement load took significantly more time to complete than the other tasks; hence, it is unclear whether the task itself or the time spent doing the task was responsible for the memory improvement. A number of studies have addressed this time confound and found that when time is equated in the study tasks, the effect of greater involvement load disappears (Folse, 2006; Keating, 2008; Webb, 2005). For example, Keating (2008) compared three vocabulary learning tasks that varied in the amount of involvement load (again, where tasks are rated as high or low on a number of cognitive–motivational scales, one being very much like C&T's depth scale): reading comprehension (no effort), reading comprehension with target word supplied (moderate effort), and sentence writing (strong effort). Passive and active knowledge of the target words was assessed immediately after treatment and 2 weeks later. The results showed that retention was highest in the sentence writing task, significantly lower in the reading plus fill-in task, and lowest in the reading comprehension task. However, when time spent on task during the encoding phase was factored in (by converting scores to words learned per minute), the benefit associated with more involving tasks disappeared. That is to say, a task rated as requiring a high degree of semantic–conceptual processing (comparable to C&T's "deep" condition) during study revealed no significantly improved learning relative to tasks with low degree of conceptual processing once time on tasks in the study task was factored in and equated over all tasks. Therefore, as with C&T's experiments, there is a question surrounding the benefits of semantic–conceptual depth of processing because tasks rated as higher on this dimension of involvement load often require more time to complete than others. This is an important issue because, in the applied linguistics literature, one goal is to establish which learning method is better in terms of efficiency, that is, which method yields the best learning outcome in the shortest period of study time. If one task requires twice as long to complete as another task to provide twice as much learning, in terms of efficiency there is no difference between the methods. To the extent that we are asking about task efficiency for learning, these study time confounds must be controlled. C&T's Experiment 5 therefore seemed the most suitable of their ten experiments because it controlled study time in Phase 1. It is interesting that, unlike the applied linguistics studies just reported, the results of C&T's experiment showed the same magnitude of benefits of deep encoding over shallow reported in their other nine experiments, even when study time was controlled. This is another inconsistency between studies testing native rather than nonnative speaker participants. The present experiment was able to examine

https://doi.org/10.1017/S014271641100052X Published online by Cambridge University Press

this difference as well because study time was controlled according to C&T's design.

Replication required that we follow as closely as possible the task conditions and design of the original experiment. Although there are weaknesses in C&T's original design, we opted not to change the tasks and measures of memory because such changes would make it more difficult to directly compare our study's results to the original C&T (but see the discussion of limitations and future research in the Discussion Section). Following C&T's (1975) Experiment 5 exactly, we designed a mixed repeated-measures experiment that compared effects of a shallow and a deep encoding tasks (depth, a within-subjects factor) on subsequent recognition of target words. The important extensions of C&T's experiments were that, first, we recruited both native and nonnative speakers of English as participants (proficiency, the between-subjects factor), and second, we varied the frequency of the target words (frequency, a within-subjects factor) within the depth conditions to explore the possibility that preexisting knowledge might affect memory for previously studied target words.

Participants

Twenty-four native speakers of English and 24 nonnative speakers of English participated. The native speakers ranged in age from 26 to 44 years. The nonnatives were all native speakers of Arabic ranging in age from 22 to 29 years. Their proficiency in English as measured by the International English Language Testing System was within Band 6, roughly characterized as "competent" or "intermediate" users of English relative to native speakers of English, whose score would be 9.0 on the International English Language Testing System scale. All participants were screened using a standard digit span memory test (Miller, 1956) to check that their short-term function was within the normal range (7 ± 2 digits). Both groups were within the normal range, and the two participant groups' scores were not significantly different ($p > .05$; see Results Section), and hence were regarded as having equivalent short-term memory function. The nonnative participants had learned English mainly as an L2 from their late primary and secondary education as a separate course of study in their curriculum, and through exposure to popular Western media such as television programs, movies, and the Internet. As university students they had all been exposed to English as medium of instruction for 2 years. To clarify the abbreviated terminology used below in referring to these two participant groups, we use the term "natives" for native speakers of English, and "nonnatives" for the native Arabic speaker group for whom English was not their mother tongue.

Note that in lexical memory tests, bilingual participants who know two closely related languages in terms of form-meaning relationships (orthography/phonology/semantics; e.g., Dutch and English) can sometimes employ their native language even in a test of their L2 because of the existence of cognate words in the two languages (i.e., similar surface form and meanings in both languages). Cognates can act as mnemonic devices and aid memory in recognition and recall tests (Dijkstra, 2007). In the present experiment, because Arabic and English belong to different language families and have quite dissimilar orthographic and

https://doi.org/10.1017/S014271641100052X Published online by Cambridge University Press

phonological systems, we were confident that when the native speakers of Arabic performed the English lexical decision task and memory test, the Arabic L1 lexicon would be of relatively little help in performing the tasks and test, and that recognition would be based on the L2 representations and the study task episode. This confidence would of course be much reduced if the two relevant languages had been closely related with many overlapping form–meaning relationships, for example in Dutch–English bilinguals.

Materials

Following C&T's basic design, a set of 60 four- to five-syllable relatively high frequency (mean frequency = 245.1) English nouns and 60 comparatively lower frequency (relative to the selected high frequency words) English nouns (mean frequency = 2013.1) was selected from the Thorndike and Lorge (1944) data set on the MRC Psycholinguistic Database (see Appendix A). The lower frequency words were selected at a frequency level that reduced the likelihood that they would not be known by the participants, because the deep study task required at least basic knowledge of the target words. All words were piloted with three participants from the same student population as the participants to confirm that they knew the meanings of the words. Each word was assigned a semantic and a nonsemantic task (replicating C&T's tasks). The nonsemantic task concerned the CV structure of the target word. For example, the target word RIVER conforms to the pattern CVCVC. Two patterns were created for each target word: one that conformed and one that did not. For the semantic condition, a sentence was constructed for each word, with a blank in the sentence (e.g., John went fishing by the _____). One semantic *yes* sentence was created for each word, and an inappropriate sentence from another target word was assigned for the semantic *no* condition. Testing software and a desktop computer were used for presentation of the materials and reaction time recordings in the study phase. For the test phase the targets were randomly combined with distractors and typed in lists on paper, replicating C&T's test phase. Two forms of the recognition test were created, each with the same 120 words randomized differently.

Design and procedure

The pool of 120 target words was randomly partitioned into three blocks of 40 words, 20 higher frequency and 20 lower frequency words in each list, and participants were randomly assigned in equal numbers to each list. For each group of 24 participants the design was as follows: the first 8 participants received one block of 40 words as targets; the second 8 received the second block of 40 words as targets, and the third 8 received the third block of 40, whereas the remaining 80 words in each block served as distracters in the test phase. Within each group of 8 participants who received the same 40 target words, 4 received one form of the recognition test paper and 4 received the other form. This was done to control possible bias derived from the way the words were randomized in a given list. Within each group of 4 participants, each word was rotated around conditions so that it appeared in all four conditions (nonsemantic *yes* and *no* and semantic

https://doi.org/10.1017/S014271641100052X Published online by Cambridge University Press

yes and *no*) once over the four participants' sessions. This was done to control possible word-response confounds (C&T found slightly different results in the *yes* and *no* response conditions). In brief summary, this was a within-subjects design in which the two study task conditions were presented in one session and were randomized together so that any given trial could have been a shallow or deep task. Moreover, all participants received all of the high frequency and lower frequency target words during the study phase.

Participants were tested individually. Each was initially told that the experiment was a perceptual accuracy test in which they would first be presented with a sentence with a gap or else a CV pattern on the computer screen (randomly) concerning the target word that would appear under the sentence of pattern. Participants were instructed to read the information and press the *yes* button on the keyboard (right hand) if the word matched the pattern or fit the sentence, or the *no* button (left hand) if there was not a match or fit. Each participant saw the 40 randomized target words with the assigned task condition during this decision phase of the experiment (plus four buffer trial words at the beginning that were for practice and were excluded from the data analysis). Reaction times and accuracy were recorded by the software program. Following the study phase, there was a 5-min break in which participants filled out a personal information questionnaire (which also asked participants to confirm that they knew all of the target words they had seen; all participants confirmed this). Participants were then given the memory test (all participants later reported in postexperiment discussion that this test was unexpected). In the test phase, participants were given one of the test papers that included the 120 words (40 targets plus 80 distractors). Participants were asked to recognize the 40 words that had been presented during the first phase by circling 40 words only and no more. This was in effect an "old/new" judgment task in which participants had to circle the 40 "old" words and no more than 40. These presentation conditions were identical for both native and nonnative participants and reflect as closely as possible the conditions described in C&T Experiment 5.

RESULTS

Short-term memory test

A t test of the results of the initial short-term memory screening test showed no significant differences between participant groups ($p > .08$) and no outliers, indicating that participant groups had equivalent short-term memory function. Short-term memory function was therefore ruled out as a relevant factor in the Phase 2 test results.

Phase 1

Mean reaction times and accuracy scores for Phase 1 are shown in Table 1. An analysis of variance showed the following: first, for high frequency words, accuracy was near perfect for both groups (native/nonnative) in both the semantic and nonsemantic task, and there was no significant difference between groups

https://doi.org/10.1017/S014271641100052X Published online by Cambridge University Press

Table 1. *Summary table of Phase 1 RTs and Phase 2 recognition accuracy*

	Phase 1 RT (s) (*SD*)				Phase 2 Score/20 (*SD*)			
	HFs	LFs	HFd	LFd	HFs	LFs	HFd	LFd
Native	3.23	3.41	2.34	2.82	6.17	5.50	14.91	16.41
Native	(1.22)	(1.46)	(1.10)	(1.23)	(1.43)	(1.14)	(0.83)	(1.28)
Nonnative	4.66	5.03	4.22	4.33	5.08	4.75	10.92	10.0
Nonnative	(1.68)	(1.73)	(1.14)	(1.11)	(1.38)	(1.29)	(2.35)	(1.58)

Note: RT, reaction time; HF, high frequency; LF, low frequency; s, shallow; d, deep.

($p > .05$). Comparison between proficiency groups showed that the nonnatives required longer than the natives to complete tasks, $F\ (1, 46) = 13.22$, $p < .05$, $\eta_p^2 = 0.232$. For both groups, reaction times in the semantic task and the nonsemantic task were not significantly different; thus, study time was at least as long for the shallow condition as the deep condition. These results largely replicate C&T's Experiment 5 results (see Discussion). Least square difference pairwise comparisons revealed that for lower frequency words in the shallow task there were no differences in accuracy or reaction time between natives and nonnatives; for the semantic task, natives' reaction times were faster ($p < .01$), but there was no difference between the groups in accuracy.

Phase 2

Results for Phase 2 are shown in Table 1 and in Figure 1. There was no significant difference between the semantic and nonsemantic yes–no study conditions for either group, nor was there a difference between the versions of the recognition test paper ($p > .05$). Accuracy scores are therefore collapsed across the Phase 1 yes–no response conditions in the results. A mixed analysis of variance showed a main effect of proficiency, $F\ (1, 46) = 226.04$, $p < .001$, $\eta_p^2 = 0.831$, indicating superior recognition accuracy in the native group; a main effect of depth, $F\ (1, 46) = 918.49$, $p < .001$, $\eta_p^2 = 0.952$, indicating superior recognition accuracy in the deep task condition; and a main effect of frequency, $F\ (1, 46) = 4.93$, $p < .05$, $\eta_p^2 = 0.097$, indicating superior recognition for lower frequency words. There was also a significant interaction of proficiency and depth, $F\ (1, 46) = 95.95$, $p < .001$, $\eta_p^2 = 0.676$, and proficiency and frequency $F\ (1, 46) = 13.99$, $p < .001$, $\eta_p^2 = 0.233$. Finally, there was a three-way interaction of proficiency, depth and frequency, $F\ (1, 46) = 13.09$, $p < .01$, $\eta_p^2 = 0.222$. Separate within-subjects analyses of variance were also performed for native and nonnative groups. For natives there was a main effect of depth, $F\ (1, 23) = 1105.16$, $p < .001$, $\eta_p^2 = 0.980$, and frequency, $F\ (1,23) = 20.66$, $p < .001$, $\eta_p^2 = 0.447$. Planned comparisons also showed that while there was no difference between high and lower frequency words in the natives' shallow condition ($p > .05$), there was a significant difference between high and lower frequency in the deep condition, with lower frequency

https://doi.org/10.1017/S014271641100052X Published online by Cambridge University Press

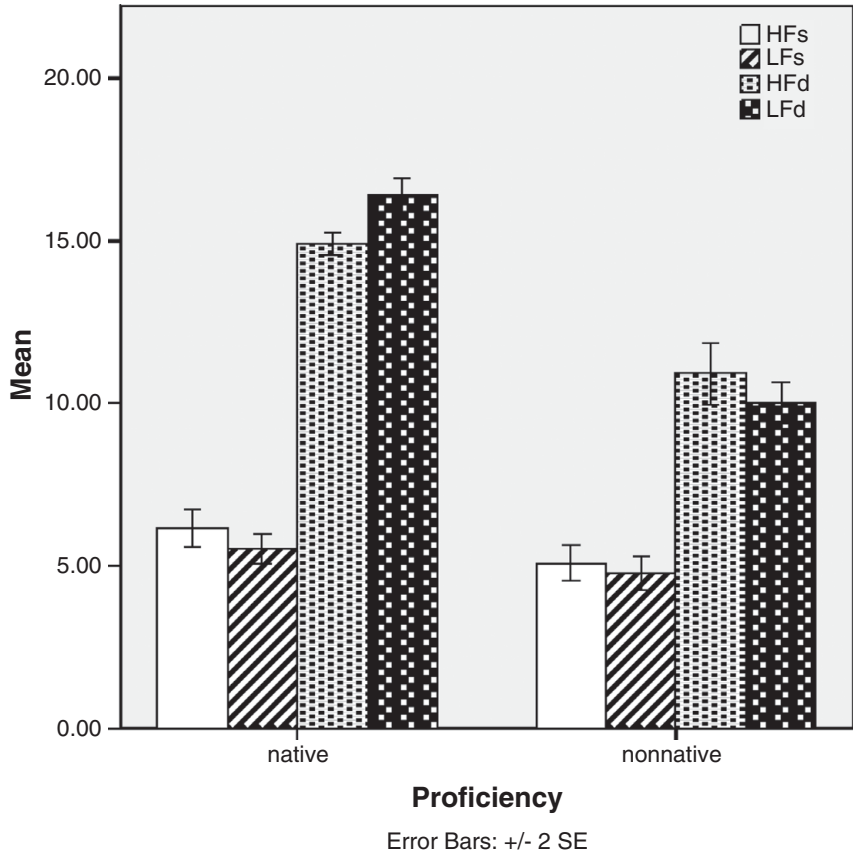


Figure 1. The mean recognition accuracy for natives and nonnatives. Error bars are ±2 *SE*. HF, high frequency words; LF, lower frequency words; s, shallow encoding; d, deep encoding.

words recognized more accurately than high frequency words, t (23) = 3.87, $p < .01$, $d = 1.38$. For nonnatives there was again an overall main effect of depth, F (1,23) = 165.32, $p < .001$, $\eta_p^2 = 0.878$. However, there was no effect of frequency in the nonnative participant group ($p > .05$).

DISCUSSION

The main question addressed in the present study concerned the extent to which language proficiency moderates LOP effects under conditions used by C&T in their influential study and numerous replications (Craik & Tulving, 1975; Lockhart & Craik, 1990). More specifically, the experiment was concerned with how word knowledge representations can affect standard LOP effects. Looking at each proficiency group separately in Figure 1, for the native speaker group the results

https://doi.org/10.1017/S014271641100052X Published online by Cambridge University Press

reported here generally replicate those reported by C&T in their Experiment 5. When time on task was controlled so that the shallow task took at least as long as the deep task to perform in Phase 1, deep processing led to better recognition accuracy for high frequency and lower frequency words relative to the shallow encoding condition. Extending C&T's results, the present study showed that the mnemonic benefit of deep processing was even more pronounced for lower frequency words in natives, indicating that frequency moderated the standard depth effect in native speakers. This difference in recognition of lower frequency words over high frequency words is in line with previous research showing that recognition memory performance is typically better when targets are less frequent words in usage (Gregg, 1976). In summary, the experiment conducted here with native speakers replicates the standard results found in previous studies of depth and frequency factors for native speakers of the target language (Craik & Tulving, 1975; Fisher & Craik, 1977; Hannon & Craik, 2001; Malmberg, 2008; Moscovitch & Craik, 1976) and for words of different frequencies when native speakers are participants. These results act as a benchmark comparison for the results for nonnative group, who underwent identical experimental conditions.

Turning to the nonnative group's results, overall C&T's results were again replicated: as shown graphically in Figure 1, deeper processing led to better recognition than shallow processing both for lower and high frequency words, a pattern similar to that for the natives in Figure 1. This result is in line both with C&T (1975) and related replications in the experimental psychology literature (Lockhart & Craik, 1990). With respect to applied linguistics research, because the present study was investigating effects of processing known words, it did not address learning of *new* vocabulary. The present experiment showed only that semantic processing provided mnemonic benefits for nonnatives when words were already known to some degree. Nevertheless, with respect to the process of acquiring an L2 vocabulary, we argue that the semantic processing benefits reported here for the nonnative participant group should be regarded as highly relevant to research into vocabulary learning. As discussed in the review of literature, vocabulary learning is regarded in most learning models as an ongoing, long-term process built up over multiple exposures to words in various learning context episodes (e.g., Schmitt, 2008). Once a word is basically known (e.g., a learner can reliably identify the given word as a "word" in a lexical decision), further exposure in everyday communication and learning environments extends this learning process through new contexts, yielding an increasingly deeper network of associations in the mental lexicon (Meara, 2009; Nation 2000). In this sense, the fact that nonnatives in the present study already knew the target words does not render the study phase less of a learning episode. If a semantic study episode improves the ability to subsequently recognize known target foreign language words more accurately, this means, first, that the words have been processed in a new context (a learning episode), and second, this context led to more accurate retrieval on a test (evidence of a better mnemonic method). This new processing episode and accurate retrieval are presumably part of the cognitive apparatus by which deeper network representations in the mental lexicon develop over time (Nation, 2001). It is in this sense that the present results can be seen as informative for language learning. Future research would, however, benefit from a design that

https://doi.org/10.1017/S014271641100052X Published online by Cambridge University Press

includes novel words and task learning conditions that more thoroughly explore vocabulary learning within a controlled C&T style study. Research would also benefit from a test phase that includes confidence ratings for each target word that is identified on the test as having been presented in the study phase. This would provide data concerning, for example, the potential relationship between word frequency and recognition accuracy confidence levels. These designs would require considerable changes to C&T's original design, and hence, would not be clear replications of their study, but such designs would certainly yield a richer set of data that could extend the findings reported here.

In terms of investigating preexisting lexical representations through a controlled psycholinguistic experiment, the present study tested whether LOP effects hold the same power as that reported in C&T when the lexical knowledge base underpinning the semantic task was presumably weaker in various respects in the metal lexicon. In this regard, there are some important points to highlight in the present experiment's positive results for deep processing conditions. Looking across the two groups of participants' scores shown in Figure 1, although deep processing benefitted both natives and nonnatives, natives' recognition accuracy both for high and lower frequency words was superior to that of the nonnatives. It is important to note that this result was obtained after a study phase in which the lower proficiency group took significantly longer to complete the tasks than the natives. Thus, even though time on task was greater for the nonnatives, and although their decision accuracy in the study phase was the same as for natives, their memory performance was poorer than for natives in the deep condition. The most plausible explanation for the nonnatives' subsequent weaker performance on the recognition test accuracy is that their preexisting knowledge base was such that it was not capable of benefiting to the same extent as that for natives. Presumably this reflects qualitative and quantitative differences in what natives and nonnatives already know about these words, and possibly what the participants are able to process in relation to these words. The nonnatives' language proficiency in general is presumably weaker both in terms of word representations in the mental lexicon as well as their cognitive processing efficiency (attention, working memory, etc.), any of which could well be involved in these results. As discussed earlier, because language proficiency is a multicomponent capability involving representations and cognitive processes, the present experiment cannot confidently dissociate these two areas of language proficiency and pinpoint exactly which aspects of the nonnatives' language processing mechanisms and representations were responsible. However, as the discussion of expert knowledge research literature suggests, to the extent that expert knowledge representations in a network leads to superior memory, we can tentatively hypothesize that qualitative differences in the two participant groups' mental lexicons were involved in these results (see discussion of future research below). Here it is important to note that in the study phase the task was not one that would be likely to overly tax processing resources because it only involved reading short sentences and evaluating known words. Moreover, because there was no time constraint imposed on the lexical decisions in the study phase, differences in the ability to process the words and sentences would be likely to be offset by offering as long as a participant needed to complete the task. Finally, there were no significant differences between the groups in their

https://doi.org/10.1017/S014271641100052X Published online by Cambridge University Press

study-phase lexical decisions (near perfect), suggesting no difficulty in performing the task. These design factors all lend support to the hypothesis that the benefits of semantic processing appear to vary at least in part with lexical knowledge and not processing difficulty during study. Further research in this area would be required, however, to explore this distinction between processing resources and the mental lexicon's relative contributions to LOP effects. One possible future experiment in this area could be to introduce a number of lexical decision conditions that vary the processing load placed on, for example, working memory during the study phase along with the deep–shallow conditions.

The present results are consistent with research into the effects of expert knowledge on memory in other knowledge domains (Einstein & Hunt 1980; Hunt & Einstein, 1981), and this suggests that we can include high-level language proficiency as a kind of expert knowledge. Assuming preexisting lexical knowledge is a plausible factor in the present results, from the point of view of theories of distinctive processing we can speculate that the nonnatives possessed weaker lexical representations of the target words in some way (Craik, 2007). According to the expert knowledge research literature, a reasonable interpretation is that the activated semantic information, while aiding memory relative to the shallow condition, was not sufficiently organized and differentiated to render memory traces as distinct as those achieved by native speakers, leading to poorer recognition relative to native speakers of the target language. Again, the nonnatives required more time to make their initial decisions, and they were not restricted in how long they took to complete the task, so this result is unlikely to be the consequence of strained attentional or processing resources in the study phase.

We conclude therefore that LOP effects may at least in part be a function of the participant's expertise in the knowledge domain, a finding that supports recent hypothesized role of preexisting representations in standard LOP effects (Craik, 2007) as well as the notion that expertise in a knowledge domain can have a powerful effect on memory for studied information falling within that field of expertise (Rawson & Van Overschelde, 2008).

A second difference between the two participant groups is that word frequency did not moderate the depth effect in nonnatives but did in the natives. That is, nonnatives' recognition memory was not better for lower frequency words in the deep task. This nonsignificant result was obtained even though the nonnative participants took significantly longer than the natives to complete the deep task for lower frequency words. This demonstrates that even when additional time was spent on the study task, poorer performance still occurred. Why would nonnatives not benefit from the frequency effect? It is worth noting that the lower frequency words in the present experiment were not especially rare, but were just less frequent relative to the high frequency words. Word frequency in the present study was therefore a rather fine-grained difference between word groups and presumably it would require very strong knowledge of the target words in terms of item specific characteristics to distinguish the words on the basis of frequency. An analogy here might be the difference between an expert wine taster and an amateur. Upon tasting two different wines, both people may be able to say that they tasted different wines, but the expert would be more likely to know that one of those wines was unusual in some respect (e.g., a rarer vintage; a wine from an unusual location) based

https://doi.org/10.1017/S014271641100052X Published online by Cambridge University Press

on extensive experience. In the expertise literature this additional fine-grained knowledge would render the unusual wine more distinctive and memorable than the others, but of course only for the taster who knows the wine is unusual to begin with. This granularity of analysis of related and distinct features appears to be an important mechanism for achieving mnemonic advantage in memory tests (Einstein & Hunt, 1980). Historically, LOP has assumed that semantic processing provides a "deeper" analysis at encoding than perceptual feature analysis. More recent discussion theorizes that the semantic analysis activates more features of the target words, and this more elaborated activation of known information serves to render the item more *distinctive* during retrieval if that knowledge is highly organized and well differentiated in a lexical network (Craik, 2007). Again, this theory posits a powerful role of preexisting representations in depth effects, and this is supported by the present results.

Finally, it is important to emphasize that there were no proficiency-related differences in recognition accuracy for shallow task target words. The crucial point of divergence in the results for the two proficiency groups appears to have occurred where semantic information was required for the analysis. Presumably the activation of this information drives processing of a greater number of distinctive features of target words, and the differences in knowledge representations between natives and nonnatives were responsible for subsequent differences in retrieval during the recognition task.

A caveat to be repeated here concerns one of the main criticisms made of C&T's original findings: the LOP effect can be reduced and even reversed if the retrieval task is changed to one that does not match well with the encoding task. That is to say, the tasks may have been optimal for producing semantic benefits (Roediger & Gallo, 2002). Future research can explore the interaction of the study-test task pairings at different proficiency levels.

In terms of implications for language learning, although there remains a fundamental problem of generalizing from a C&T style study, where target words are known by the participants, to a language learning study where target words are completely unknown, there are some potentially useful conclusions. It appears that preexisting lexical representations can moderate the semantic processing advantages even when the experimental conditions as biased toward a semantic processing advantage (Barcroft, 2002; Roediger & Gallo, 2002). This may be an important idea for materials developers because it reinforces the conclusion that deep semantic processing tends to have more effective impact on higher proficiency learners (Sagarra & Alba, 2006). For lower level learners, other kinds of task may hold more value, for example, form-related processing tasks that direct attention to orthographic/phonological features rather than meaning (Sagarra & Alba, 2006). This difference suggests that the mnemonic benefit of some task is not simply related to how deeply target words are processed (in the LOP sense), but rather it may be more productively be related to the concept of distinctiveness. For example, in previous studies some authors have argued that because natives appear to store words in related semantic networks, learners should also study words in related semantic networks (Hashemi & Gowdasiaei, 2005). However, this approach seems to assume that words are represented together only in terms of similarity. As the present study has argued, expert knowledge entails that, within a

https://doi.org/10.1017/S014271641100052X Published online by Cambridge University Press

set of semantically related items, there are also encoded differences that make each related item distinct from the otherwise related items in the set (Hunt & Worthen, 2006; Van Overschelde et al., 2005). This lexical network, built of similarities and differences, presumably develops over time and experience of the words in different contexts, each new episode helping learners to develop lexical entries for words in terms of form, meaning, and use (Craik, 2007; Nation, 2001). The point here is that it is not necessarily the semantic factor that is always vital to mnemonic benefits. The historical importance attributed to semantic processing may hold for more advanced learners and natives, but, in contrast to LOP researchers who see deep processing as always the "smart" bet (Lockhart, 2002, p. 399) when choosing a mnemonic method in a language learning context, the best bet may depend on what the learner already knows. Not all participants seem to benefit equally from deep processing conditions in a classical experimental method such as that used by C&T, and, as has been argued here, this appears to be caused to some extent by the nature of the preexisting representations for target words in the mental lexicon.

At the level of a language learning classroom, many factors engage and affect learning outcomes, and so betting on any particular method will always be a difficult endeavor. Some of these factors have been incorporated in recent theoretical accounts of language learning and how best to engage learners with new language (e.g., Hulstijn & Laufer, 2001; Moonen et al., 2006; Schmitt, 2008), and others as yet unidentified will probably emerge as research continues. Future research suggested above can help to determine in more detail how preexisting representations are utilized to render studied information more memorable, and how these processing episodes might contribute to the development of richer lexical representations in the mental lexicon.

https://doi.org/10.1017/S014271641100052X Published online by Cambridge University Press

https://doi.org/10.1017/S014271641100052X Published online by Cambridge University Press

APPENDIX A

Target words and Thorndike and Lorge frequencies from MRC Psycholinguistic Database

Higher	Higher	Higher	Lower	Lower	Lower
ACTOR 240	DOLL 223	MOVIE 285	BABY 1430	HOUR 2485	ROOM 3789
AGENT 209	FENCE 256	NOISE 269	CHAIR 1298	HOUSE 4231	SMILE 2143
ALARM 224	FILM 214	NOON 236	CHILD 1574	IDEA 1824	SOUND 1305
APPLE 220	FLAME 259	ONION 268	CITY 1258	KISS 1027	STORY 1651
BAND 272	FROWN 216	PIANO 212	DOOR 2511	LADY 2114	TABLE 1325
BEAN 265	GOLF 278	PIPE 227	DRESS 1790	LIFE 4804	TALK 3100
BENCH 226	GRASS 253	RICE 203	FACE 3902	LIGHT 2387	THREE 2873
BOWL 268	GRAVE 243	ROPE 244	FACT 2114	LINE 1498	TODAY 1104
BRICK 201	HELL 227	SEED 284	FEET 1540	MIND 2366	TOWN 1607
BRIDE 259	HONEY 274	SHELL 218	FIGHT 1391	MONEY 3089	TRAIN 1019
CHEST 221	HORN 217	SHIRT 264	FIRE 1319	MONTH 1000	VOICE 1998
CHIN 224	JOKE 230	SKIRT 297	FLOOR 1001	NAME 1335	WALL 1122
CLOCK 235	KNIFE 251	SMELL 280	FOOD 1000	NIGHT 3385	WATCH 1722
CLOTH 275	LAMP 270	SOUP 246	FOOL 1140	ORDER 1477	WATER 2067
COPY 276	LOAN 206	STAMP 266	GIRL 4058	PAGE 3610	WEEK 2253
CORN 227	LOCK 277	TAXI 260	GLASS 1001	PAINT 1102	WIFE 1668
CURE 219	LOVER 272	TEAM 286	GREEN 1025	PAPER 1235	WISH 1220
DAWN 238	MENU 216	TOWER 201	HAIR 1183	PARTY 1442	WOMAN 2431
DEVIL 225	MODEL 244	TRAY 203	HAND 3334	PLAN 1873	WORD 2845
DIET 258	MOON 285	WHEEL 265	HEART 2067	PLAY 2606	WORK 2720

ACKNOWLEDGEMENTS

I thank Jan Hulstijn and two anonymous reviewers for their helpful comments on an earlier draft of this article

REFERENCES

Aitchison, J. (1994). *Words in the mind: An introduction to the mental lexicon*. London: Blackwell.

Barcroft, J. (2002). Semantic and structural elaboration in L2 lexical acquisition. *Language Learning, 52*, 323–363.

Cho, K. S., & Krashen, S. (1994). Acquisition of vocabulary from Sweet Valley Kids Series: Adult ESL acquisition. *Journal of Reading, 37*, 662–667.

Coady, J., & T. Huckin, T. (1997). *Second language vocabulary acquisition.* Cambridge: Cambridge University Press.

Craik, F. I. M. (2002). Levels of processing: Past, present . . . and future? *Memory, 10*, 305–318.

Craik, F. I. M. (2007). Encoding: A cognitive perspective. In H. L. Roediger, Y. Dudai, & S. M. Fitzpatrick (Eds.), *Science of memory: Concepts* (pp. 129–135). Oxford: Oxford University Press.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671–684.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*, 268–294.

Dijkstra, T. (2007). The multilingual lexicon. In M. Gareth Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 251–265). Oxford: Oxford University Press.

Einstein, G. O., & Hunt, R. R. (1980). Levels of processing and organization: Additive effects of individual-item and relational processing. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 588–598.

Erten, I. H., & Tekin, M. (2008). Effects on vocabulary acquisition of presenting new words in semantic sets versus semantically unrelated sets. *System, 36*, 407–422.

Finkbeiner, M., & Nicol, J. (2003). Semantic category effects in second language word learning. *Applied Psycholinguistics, 24*, 369–383.

Fisher, R. P., & Craik, F. I. M. (1977). Interaction between encoding and retrieval operations in cued recall. *Journal of Experimental Psychology: Human Learning and Memory, 3*, 701–711.

Folse, K. S. (2006). The effect of type of written exercise on L2 vocabulary retention. *TESOL Quarterly, 40*, 273–293.

Gallo, D. A., Meadow, N. G., Johnson, E. L., & Foster, K. T. (2008). Deep levels of processing elicit a distinctiveness heuristic: Evidence from the criteria recollection task. *Journal of Memory and Language, 58*, 1095–1111.

Gregg, V. H. (1976). Word frequency, recognition, and recall. In J. Brown (Ed.), *Recall and recognition.* London: Wiley.

Guo, C., Zhu, Y., Ding, J., Fan, S., & Paller, K. A. (2004). An electrophysiological investigation in memory encoding, depth of processing, and word frequency in humans. *Neuroscience Letters, 356*, 79–82.

Hannon, B., & Craik, F. I. M (2001). Encoding specificity revisited: The role of semantics. *Canadian Journal of Cognitive Psychology, 55*, 231–243.

Hasegawa, M., Carpenter, P. A., & Just, M. A. (2002). An fMRI study of bilingual sentence comprehension and workload. *NeuroImage, 15*, 647–660.

Hashemi, R. H., & Gowdasiaei, F. (2005). An attribute-treatment interaction study: Lexical-set versus semantically unrelated vocabulary instruction. *RELC Journal, 36*, 341–361.

Hill, M., & Laufer, B. (2003). Type of task, time-on-task and electronic dictionaries in incidental vocabulary acquisition. *IRAL, 41*, 87–106.

Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in second language: A review. *Studies in Second Language Acquisition, 21*, 181–193.

Hulstijn, J. H. (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In P. J. Anaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 113–125). London: Macmillan.

https://doi.org/10.1017/S014271641100052X Published online by Cambridge University Press

Hulstijn, J. H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *Modern Language Journal, 80*, 327–339.
Hulstijn, J. H. & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning, 51*, 539–558.
Hunt, R. R. (2003). Two contributions of distinctive processing to accurate memory. *Journal of Memory and Language, 48*, 811–825.
Hunt, R. R., & Einstein, G. O. (1981). Relational and item specific information in memory. *Journal of Verbal Learning and Verbal Behavior, 20*, 497–514.
Hunt, R. R., & Worthen, J. B. (Eds.). (2006). *Distinctiveness and memory*. New York: Oxford University Press.
Joe, A. (1998). What effects do text-based tasks promoting generation have on incidental vocabulary acquisition? *Applied Linguistics, 19*, 357–377.
Keating, G. D. (2008). Task effectiveness and word learning in a second language: The involvement hypothesis on trial. *Language Teaching Research, 12*, 365–386.
Knight, S. (1994). Dictionary: The tool of last resort in foreign language reading? A new perspective. *Modern Language Journal, 78*, 285–299.
Lockhart, R. S. (2002). Levels of processing, transfer appropriate processing, and the concept of robust encoding. *Memory, 10*, 397–403.
Lockhart, R. S., & Craik, F. I. M. (1990). Levels of processing: A retrospective analysis of a framework for memory research. *Canadian Journal of Psychology, 44*, 87–112.
Luppescu, S., & Day, R. R. (1993). Reading, dictionaries, and vocabulary learning. *Language Learning, 43*, 263–287.
Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology, 57*, 335–384.
Malmberg, K. J., Steyvers, M., Stephens, J. D., & Shiffrin, R. M. (2002). Feature-frequency effects in recognition memory. *Memory & Cognition, 30*, 607–613.
Meara, P. (1997). Towards a new approach to modelling vocabulary acquisition. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 109–121). Cambridge: Cambridge University Press.
Meara, P. (2009). *Connected words: Word associations and second language vocabulary acquisition*. Amsterdam: John Benjamins.
Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81–97.
Moonen, M. L. I., De Graaff, R., & Westhoff, G. J. (2006). Focused tasks, mental actions and second language teaching: Cognitive and connectionist accounts of task effectiveness. *International Journal of Applied Linguistics, 152*, 35–55.
Morris, C., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16*, 519–533.
Moscovitch, M., & Craik, F. I. M. (1976). Depth of processing, retrieval cues, and uniqueness of encoding as factors in recall. *Journal of Verbal Learning and Verbal Behavior, 15*, 447–458.
Nation, I. S. P. (2000). Learning vocabulary in lexical sets: Dangers and guidelines. *TESOL Journal, 9*, 6–10.
Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
Nyberg, L. (2002). Levels of processing: A view from function brain imagining. *Memory, 10*, 345–348.
Otten, L. J., Henson, R. N., & Rugg, M. D. (2001). Depth of processing effects on neural correlates of memory encoding: Relationship between findings from across- and within-task comparisons. *Brain, 124*, 399–412.
Pressley, M., Levin, J. R., & Miller, G E. (1982). The keyword method compared to alternative vocabulary learning strategies. *Contemporary Educational Psychology, 7*, 50–60.
Rawson, K. A., & Van Overschelde, J. P. (2008). How does knowledge promote memory? The distinctiveness theory of skilled memory. *Journal of Memory and Language, 58*, 646–668.
Roediger, H. L., III, & Gallo, D. A. (2002). Levels of processing: Some unanswered questions. In M. Naveh-Benjamin, M. Moscovitch, & H. L. Roediger, III (Eds.), *Perspectives on human memory and cognitive aging: Essays in honour of Fergus Craik* (pp. 28–47). Philadelphia, PA: Psychology Press.

https://doi.org/10.1017/S014271641100052X Published online by Cambridge University Press

Rohrer, D., & Pashler, H. (2007). Increasing retention without increasing study time. *Current Directions in Psychological Science, 16*, 183–186.

Sagarra, N., & Alba, M. (2006). The key is in the keyword: L2 vocabulary learning methods with beginning learners of Spanish. *Modern Language Journal, 90*, 228–243.

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research, 12*, 329–363.

Scribner, G. (2000). *The effects of three methods of vocabulary memorization in second language acquisition: Rote memorization, mnemonics, and network of conceptual relations*. Unpublished master's thesis, Brigham Young University.

Sokman, A. (1997). Current trends in teaching second language vocabulary. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 237–257). Cambridge: Cambridge University Press.

Thorndike, E. L., & Lorge, I. (1944). *The Teacher's word book of 30,000 words*. New York: Teachers College, Columbia University.

Van Overschelde, J. P., Rawson, K. A., Dunlosky, J., & Hunt, R. R. (2005). Distinctive processing underlies skilled memory. *Psychological Science, 16*, 358–361.

Verfaellie, M., Croce, P., & Milberg, W. (1995). The role of episodic memory in semantic learning: An examination of vocabulary acquisition in a patient with amnesia due to encephalitis. *Neurocase, 1*, 291–304.

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition, 27*, 33–52.

https://doi.org/10.1017/S014271641100052X Published online by Cambridge University Press